Volume 13, Number 2 June 2017

IJCB INTERNATIONAL JOURNAL OF CENTRAL BANKING Volume 13, Number 2 June 2017

IJCB

INTERNATIONAL JOURNAL OF CENTRAL BANKING

International Journal of Central Banking

Volume 13, Number 2 June 2017

*The contents of this journal, together with additional materials provided by article authors, are available without charge at **www.ijcb.org**.*

Copyright © 2017 by the Association of the International Journal of Central Banking. All rights reserved. Brief excerpts may be reproduced or translated provided the source is cited. Consult www.ijcb.org for further information.

The *International Journal of Central Banking* is published quarterly (ISSN: 1815-4654). Online access to the publication is available free of charge at **www.ijcb.org**. Individual print subscriptions are available. Orders may be placed by phone (001 415 974 2035), via fax (001 415 974 2168), or by e-mail (editor@ijcb.org).

Renewals, claims, address changes, and requests for permission to reprint material from this journal should be addressed to:

International Journal of Central Banking
Economic Research Department
Federal Reserve Bank of San Francisco
101 Market Street
San Francisco, CA 94105
USA

Phone: 001 415 974 2035
Fax: 001 415 974 2168
E-mail: editor@ijcb.org

The views expressed in this journal do not necessarily represent the views of the Association of the International Journal of Central Banking or any of its members.

ISSN: 1815-4654

International Journal of Central Banking

Board of Directors

Chairman

Claudio Borio, *Bank for International Settlements*

Board Members

Q. Farooq Akram, *Norges Bank*
Abdulaziz Al-Furaih, *Saudi Arabian Monetary Agency*
David E. Altig, *Federal Reserve Bank of Atlanta*
Carlos Hamilton Vasconcelos Araujo, *Central Bank of Brazil*
Jan Marc Berk, *The Nederlandsche Bank*
Mohamed Tahar Bouhouche, *Bank of Algeria*
Lillian Cheung, *Hong Kong Monetary Authority*
Laurent Clerc, *Bank of France*
Francisco G. Dakila Jr., *Central Bank of the Philippines*
Mary Daly, *Federal Reserve Bank of San Francisco*
Michael Dotsey, *Federal Reserve Bank of Philadelphia*
William English, *Federal Reserve Board*
Gabriel Fagan, *Central Bank of Ireland*
Jiang Feng, *People's Bank of China*
Manuel Ramos Francia, *Bank of Mexico*
Jeffrey C. Fuhrer, *Federal Reserve Bank of Boston*
Kamil Galuscak, *Czech National Bank*
Niels Lynggård Hansen, *Danmarks Nationalbank*
Philipp Hartmann, *European Central Bank*
Elena Iorga, *National Bank of Romania*
Seung-Cheol Jeon, *Bank of Korea*
Juan F. Jimeno, *Bank of Spain*
George A. Kahn, *Federal Reserve Bank of Kansas City*
Sujit Kapadia, *Bank of England*
Ali Hakan Kara, *Central Bank of Turkey*
Christopher Kent, *Reserve Bank of Australia*
Evan Koenig, *Federal Reserve Bank of Dallas*
Ana Christina Leal, *Bank of Portugal*
Carlos Lenz, *Swiss National Bank*
Jesper Lindé, *Sveriges Riksbank*
Choy Keen Meng, *Monetary Authority of Singapore*
John McDermott, *Reserve Bank of New Zealand*
Emanuel Moench, *Deutsche Bundesbank*
Alberto Naudon, *Central Bank of Chile*
Edward Offenbacher, *Bank of Israel*
Fabio Panetta, *Bank of Italy*
Michael D. Patra, *Reserve Bank of India*
Thórarinn G. Pétursson, *Central Bank of Iceland*
Ivan Ribnikar, *Bank of Slovenia*
Lawrence Schembri, *Bank of Canada*
Sam Schulhofer-Wohl, *Federal Reserve Bank of Minneapolis*
Mark Schweitzer, *Federal Reserve Bank of Cleveland*
Shigenori Shiratsuka, *Bank of Japan*
Jan Smets, *National Bank of Belgium*
Daniel Sullivan, *Federal Reserve Bank of Chicago*
Juha Tarkka, *Bank of Finland*
George Tavlas, *Bank of Greece*
Joseph Tracy, *Federal Reserve Bank of New York*
Dobieslaw Tymoczko, *National Bank of Poland*
Hernando Vargas Herrera, *Banco de la República*
Christopher Waller, *Federal Reserve Bank of St. Louis*
John Weinberg, *Federal Reserve Bank of Richmond*
Ksenia Yudaeva, *Central Bank of Russian Federation*

Editorial Board

Managing Editor

Loretta Mester
Federal Reserve Bank of Cleveland

Co-editors

Tobias Adrian
International Monetary Fund

Boragan Aruoba
University of Maryland

Pierpaolo Benigno
LUISS Guido Carli

Harrison Hong
Columbia University

Luc Laeven
European Central Bank

Lucrezia Reichlin
London Business School

Rafael Repullo
CEMFI

Barbara Rossi
Universitat Pompeu Fabra

Associate Editors

Patrick Bolton
Columbia University

Michael D. Bordo
Rutgers University

Mark Carey
Federal Reserve Board

Pierre Collin-Dufresne
Columbia University

Guy Debelle
Reserve Bank of Australia

Douglas W. Diamond
University of Chicago Graduate School of Business

Francis Diebold
University of Pennsylvania

Michael Dotsey
Federal Reserve Bank of Philadelphia

Darrell Duffie
Stanford University

Jordi Galí
Centre de Recerca en Economia Internacional (CREI)

Marvin Goodfriend
Carnegie Mellon University

Michael B. Gordy
Federal Reserve Board

Luigi Guiso
European University Institute

Andrew G. Haldane
Bank of England

Takatoshi Ito
Columbia University

David Lando
Copenhagen Business School

Philip Lane
Trinity College Dublin

Francesco Lippi
University of Sassari

Carmen M. Reinhart
Harvard Kennedy School

Eli M. Remolona
Bank for International Settlements

Hélène Rey
London Business School

Jean-Charles Rochet
University of Zurich

Andrew K. Rose
University of California, Berkeley

Klaus Schmidt-Hebbel
Organisation for Economic Co-operation and Development (OECD)

Lars E.O. Svensson
Stockholm School of Economics

Jürgen von Hagen
University of Bonn

Ernst-Ludwig von Thadden
University of Mannheim

Tsutomu Watanabe
University of Tokyo

Advisory Board

Franklin Allen
The Wharton School of the University of Pennsylvania

Charles Goodhart
London School of Economics

Hyun Shin
Bank for International Settlements

Kazuo Ueda
University of Tokyo

Carl E. Walsh
University of California

Michael Woodford
Columbia University

John Taylor
Stanford University

Inferring Inflation Expectations from Fixed-Event Forecasts*

Diego Winkelried
Universidad del Pacífico (Lima, Perú)

Often, expected inflation measured by surveys is available only as fixed-event forecasts. Even though these surveys do contain information of a complete term structure of expectations, direct inferences about them are troublesome. Records of fixed-event forecasts through time are associated with time-varying forecast horizons, and there is no straightforward way to interpolate such figures. This paper proposes an adaptation of the measurement model of Kozicki and Tinsley (2012) to suit the intricacies of fixed-event data. Using the Latin American Consensus Forecasts, the model is estimated to study the behavior of inflation expectations in four inflation targeters (Chile, Colombia, Mexico, and Peru). For these countries, the results suggest that the announcement of credible inflation targets has been instrumental in anchoring long-run expectations.

JEL Codes: C32, E37, E52.

1. Introduction

Expectations about future inflationary outcomes play a key role in macroeconomic analysis. For instance, the determination of aggregate prices in modern macroeconomic models is often summarized

*I am indebted to Carlos Barrera, Carlos Casas, Eduardo Morón, Javier Torres, and Marco Vega for insightful discussion, and to participants at the XXXII Economists Meeting of the Central Reserve Bank of Peru and at the Viernes CIUP seminar of Universidad del Pacífico for their comments. I also gratefully acknowledge the suggestions of the associate editor and an anonymous referee for improving the content and clarity of this article. I would also like to thank Natalia Guerrero for her research assistance. I alone am responsible for any errors that may remain and for the views expressed in this article. Author contact: Faculty of Economics and Finance, Universidad del Pacífico, Av. Salaverry 2020, Lima 11, Perú. E-mail: winkelried_dm@up.edu.pe.

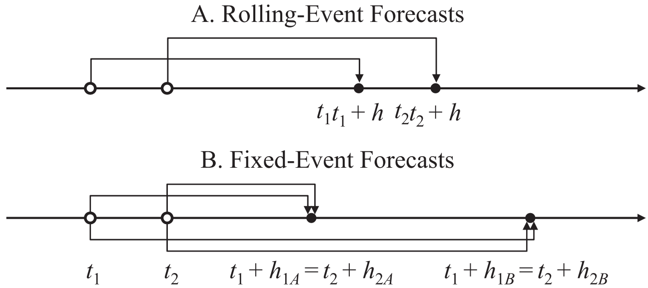
in expectations-augmented Phillips curves. Also, real interest rates, which provide a rough measure of the monetary policy stance once compared with suitable reference levels, can be computed by deducting inflationary expectations from nominal interest rates. Finally, many countries conduct their monetary policy within an explicit inflation-targeting regime, and the pillar of such a monetary arrangement is to maintain expectations anchored to a predetermined target. Being able to track and anticipate the behavior of expectations, thus, seems to be an important ingredient for the successful implementation of such a policy regime.

Economic surveys, where a number of participants are asked to produce forecasts of future inflation, provide a direct source of information on expected inflation.[1] As illustrated in figure 1, two types of data structures emerge.[2] The first one corresponds to "rolling-event" forecasts (REFs, henceforth), such as those recorded in the Livingston Survey for the United States or the Gallop Poll for the United Kingdom. Here, the survey collects h-period-ahead forecasts, so for every new release of the survey, the "event" to be forecast "rolls" forward. In other words, in an REF the horizon h is fixed, and the target date is always separated h periods from the forecast origin. The bulk of the literature that uses empirical measures of expectations from surveys, either to assess their rationality (e.g.,

[1]An alternative approach is to deduce expected inflation from the difference between nominal and indexed bond yields. Such an approximation, nonetheless, is often biased and more volatile than survey measures, as it also captures factors that are not directly linked to inflation expectations, such as movements of risk or liquidity premia. Consistently with this critique, Ang, Bekaert, and Wei (2007) and Chernov and Mueller (2012) find that survey measures categorically outperform financial measures in forecasting inflation.

[2]We focus our analysis on *quantitative* (i.e., point) forecasts. A large literature, in contrast, favors using *qualitative* or *directional* forecasts ("would inflation be higher or lower?"), arguing that survey respondents are unable to produce trustworthy numerical predictions. Thus, the mapping from a qualitative perception to a precise quantitative figure is left to the econometrician, using either the so-called probability approach or the regression approach (see Pesaran and Weale 2006 for a comprehensive survey). This assumes, of course, that the econometrician is better suited to perform such mapping than the survey respondent. This may be true for consumer surveys, but the argument is weakened when the respondents are experts or informed professionals, as with the surveys used in this study.

Figure 1. Rolling-Event and Fixed-Event Forecasts



Davies and Lahiri 1999) or to use them in econometric models (e.g., Mehra and Herrington 2008; Mavroeidis, Plagborg-Møller, and Stock 2014), is based on this type of measure. This is so because expectations in theoretical macroeconomic models are formulated as REFs, and also because of the widespread popularity in applied work of sources such as the Livingston Survey.

The second data structure corresponds to "fixed-event" forecasts (FEFs, henceforth) which, in contrast to REFs, have received scant attention despite the fact that FEF data are available for a much larger number of countries. The "event" to be forecast is often an annual rate, for current and subsequent years. Specifically, following figure 1B, in period t_1 survey participants are asked to forecast inflation for period $t_1 + h_1$, an h_1-period-ahead prediction; later on, in period $t_2 > t_1$ they are asked for a forecast of inflation *for the same date*, which now corresponds to an h_2-period-ahead forecast, where $h_2 = h_1 - (t_2 - t_1) < h_1$. The forecast event is kept fixed throughout, while the forecasting horizon *shrinks* as the timeline approaches the event. The highly reputed company Consensus Economics Inc. conducts the Consensus Forecasts monthly poll among forecasters working in the private sector (in more than seventy countries), compiling their predictions of a range of economic variables and reporting them in a fixed-event format. Other widespread sources of FEFs of global and country-specific economic variables are the International Monetary Fund's (IMF's) World Economic Outlook, the World Bank's Global Economic Prospects, the OECD's Economic Outlook, and polls conducted by central banks.

Notwithstanding their availability, FEFs are an unexploited resource to describe the evolution of expected inflation.[3] We conjecture that this is so because the time-varying nature of the forecast horizons makes comparisons over time troublesome. As suggested, it seems more natural to conceptualize expectations in macroeconomic models as REFs (for instance, to estimate an expectations-augmented Phillips curve or to compute real interest rates). And even though a survey that registers FEFs for horizons h_A and h_B (see figure 1) does contain information for expectations at any intermediate horizon $h_A \leq h \leq h_B$ (for instance, twelve-month-ahead expectations are implicitly contained in current- and next-year forecasts), there is no obvious way to interpolate such figures. The purpose of this paper is to develop an empirical model to explicitly infer expectations from data on actual inflation and FEFs.

The model is a version of the *shifting-endpoint* model advanced in Kozicki and Tinsley (1998, 2001, 2012), suitably modified to deal with the intricacies of the FEF data structure. We reckon that our modifications to Kozicki and Tinsley's framework widens its scope and should be useful in practice, especially to analyze expectations in countries for which only FEFs are available. Importantly, even though the model is fitted to irregularly sampled observations associated with time-varying forecast horizons, it is capable of producing a coherent and complete term structure of inflation expectations. Thus, the model provides an interpolation method that is internally consistent with inflation dynamics and survey information.

As stressed in Kozicki and Tinsley (1998, 2001), predictions derived from time-series models of inflation are likely to be poor proxies of expected inflation. Such predictions are based on historical data only and fail to accommodate structural changes in inflation dynamics that may be well reflected in survey measures.[4] On the

[3]Since the seminal contribution of Nordhaus (1987), FEFs have become popular in testing forecasts' rationality or optimality (e.g., Clements 1997; Bakhshi, Kapetanios, and Yates 2005; Timmermann 2007). FEFs are in fact ideal for such purposes. Their evolution over time provides a stream of revisions whose correlation against various information sets can be directly assessed. A further advantage is that rationality can be tested without having the data of the target variable.

[4]This is indeed the case in our empirical application below. Survey measures in all the Latin American countries of our sample correctly anticipated the trend decline in inflation during the 1990s. This is remarkable, since purely backward-looking or adaptive expectations would tend to systematically overpredict inflation in such a context.

other hand, forecasts from surveys are often responsive to the latest inflation observations for short-term horizons and relatively unresponsive for long-term horizons. This pattern is difficult to reconcile with model-based multi-period forecasts. If the model fitted to the inflation process exhibits strong mean reversion to a fixed level, forecasts for all horizons would be too insensitive to recent inflation, whereas if the inflation model is persistent (say, because it contains a unit root), then forecasts would be excessively sensitive to inflation news.

The shifting-endpoint model is a compromise whose multi-period forecasts reflect the main dynamic features of the inflation process and also account for the behavior of available survey data. The endpoint is an unobservable variable that measures the perceptions of long-run inflation held by economic agents at a given point in time, and that directly affects actual inflation and its forecasts for arbitrary horizons. Survey records are taken as error-ridden versions of such forecasts, and so the inferred inflation expectations would exploit all available information from inflation and surveys.

The remainder of the document is organized as follows. Section 2 presents the shifting-endpoint model, discusses its main properties, and describes some variants. The model can be represented in state-space form, and the moving nature of the forecast horizons in FEFs implies that this representation is time varying. The Kalman filter can be used for estimation and inference, even when survey expectations are sampled less frequently than inflation. Section 3 presents empirical results for four Latin American countries, using data from the Latin American Consensus Forecasts survey. The evolution of the predicted endpoints reveals the key role that expectations have played in these countries, first to reduce inflation to single-digit levels and then to keep it stable, as well as the importance of the central bank's credibility in anchoring long-run expectations. This section also illustrates the potential usage of the proposed shifting-endpoint model to infer about inflation expectations in real time, an empirical exercise that is likely to be valuable for the routine conduct of monetary policy in an inflation-targeting framework. Finally, section 4 presents concluding remarks and questions for future research.

2. Econometric Framework

Next, we describe the workings of the shifting-endpoint model for inflation and discuss its statistical treatment. We also present model variants that can be evaluated within the same analytical framework. In what follows, $x_{t\,|\,s}$ denotes the expectation of a random variable x_t, formed by economic agents conditional on the information up to and including period s.

2.1 Shifting-Endpoint Model for Inflation

The limiting conditional forecast of inflation is given by

$$\mu_t = \lim_{h \to \infty} \pi_{h\,|\,t}\,, \tag{1}$$

where t denotes the time subscript of the information set on which expectations are conditioned. Thus, given information up to and including period t, μ_t measures the perceived level at which inflation would eventually stabilize, i.e., the *endpoint*.

We assume that long-run expectations are formed in a weakly rational manner, in the sense of Nordhaus (1987): changes in perception are unpredictable or, more formally, μ_t is a martingale with respect to its own past. In other words, if agents can anticipate future changes to their long-run perceptions, then such changes should be immediately incorporated in their current perceptions, as in the law of iterated expectations. This behavior can be satisfactorily modeled by assuming that the endpoint follows a random walk,

$$\mu_t = \mu_{t-1} + \nu_t\,, \tag{2}$$

where ν_t is an innovation satisfying $\nu_{t\,|\,s} = 0$ for $s < t$. The endpoint μ_t is treated as an unobservable variable, and the main purpose of the analysis is to infer about its state using inflation and survey data.

The notion of a varying endpoint for expected inflation can be easily accommodated in a parametric forecasting model. Let π_t denote inflation at period t and suppose that the expectations in period t are formed using information up to and including period $t-1$. Define also $\Phi(L) = \phi_1 + \phi_2 L + \cdots + \phi_{p-1} L^{p-2} + \phi_p L^{p-1}$ as a polynomial in the lag operator ($L^k \pi_t = \pi_{t-k}$), and assume that the roots of $1 - \Phi(z)z = 0$ lie outside the unit circle (we make some allowance for a unit root in $1 - \Phi(z)z$ later). Inflation dynamics are captured by

$$\pi_t = \Phi(L)\pi_{t-1} + (\,1 - \Phi(1)\,)\mu_{t-1} + \epsilon_t\,, \tag{3}$$

where ϵ_t is an inflation shock, assumed to be uncorrelated with ν_t at all lags and leads. Under the aforementioned assumptions, $\pi_t - \mu_{t-1}$, the deviation of inflation from its latest perceived long-run level, is a zero-mean stationary process. By construction, given the information of period $t-1$, inflation is expected to converge to μ_{t-1} as the forecasting horizon increases. Thus, as is further discussed in section 2.5, the dynamic specification (3) allows us to disentangle the effects of the shifting endpoint from short-run fluctuations in inflation.

2.2 Incorporating Survey Data

It is convenient to write the forecasting model in companion form. Let

$$z_t = \begin{bmatrix} \pi_t \\ \pi_{t-1} \\ \vdots \\ \pi_{t-p+1} \end{bmatrix} \qquad \text{and} \qquad C = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{4}$$

where z_t is a p-vector with current and past inflation data and C is the companion matrix. Let also I_p be the $p \times p$ identity matrix, i_p be a p-vector of ones, and e_1 be a unit p-vector with a one in the first element and remaining elements zero. Then, (3) corresponds to the first row of the companion form

$$z_t = Cz_{t-1} + (I_p - C)i_p\mu_{t-1} + e_1\epsilon_t\,. \tag{5}$$

Multi-step forecasts of inflation based on this model are $\pi_{t+h\,|\,t-1} = e_1{}'z_{t+h\,|\,t-1}$, where

$$z_{t+h\,|\,t-1} = C^{h+1}z_{t-n} + (I_p - C^{h+1})i_p\mu_{t-1}\,. \tag{6}$$

Survey data registered in period t represent participants' forecasts conditional on inflation up to and including period $t-1$. Moreover, consider that the survey collects information for m different forecasts F_{it}, each associated with a different horizon that may be time varying h_{it} $(i = 1, 2, \ldots, m)$. Taking the inflation forecast from the shifting-endpoint model as an approximation of the survey expectation, we may write

$$F_{it} = \pi_{t+h_{it}\,|\,t-1} + \varepsilon_{it}\,, \tag{7}$$

where ε_{it} is an approximation error that captures discrepancies between the implicit (and unknown) forecasting model of survey participants and the shifting-endpoint model, as well as possible measurement errors in survey data. It is important to note that the errors in (7) do not reflect differences between actual inflation and forecasts. This is because both the model-based forecasts and survey records are conditioned on $t-1$, which implies that $\mathrm{cov}(\epsilon_t, \varepsilon_{it}) = \mathrm{cov}(\nu_t, \varepsilon_{it}) = 0$ for $i = 1, 2, \ldots, m$. Thus, unlike forecast errors, there is no reason to expect them to be serially correlated; hence, $\mathrm{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0$ for $t \neq s$ and any i or j.

On the other hand, revisions of forecasts of different target dates made at the same time are likely to be highly correlated (Clements 1997; Bakhshi, Kapetanios, and Yates 2005). News at period t that leads to a revision in the forecast at horizon h_{it} is also likely to produce a revision in the forecast at horizon h_{jt} $(i \neq j)$. Following (7), even though an important part of these co-movements would be accounted for by changes in the endpoint, we expect the approximation errors to be contemporaneously correlated $\sigma_{ij} = \mathrm{cov}(\varepsilon_{it}, \varepsilon_{jt}) \neq 0$. Note that whether σ_{ij} is zero or not is a straightforward testable hypothesis that can be formally addressed.

2.3 State-Space Form

The statistical treatment of the model is based on its state-space representation (see Harvey 1989, ch. 5). The Kalman filter yields linear least-squares predictions of the unobserved endpoint μ_t based on current and past observations, along with their corresponding mean square errors.[5] Moreover, given the parameters of the model ($\phi_1, \ldots, \phi_p$, the variances of ν_t and ϵ_t, and the covariances among $\varepsilon_{1t}, \ldots, \varepsilon_{mt}$), a Gaussian likelihood function can be evaluated from the one-step-ahead prediction errors produced by the Kalman filter. This function can be maximized numerically, thereby providing (quasi) maximum-likelihood estimates of the unknown parameters.

The law of motion of the unobservable endpoint (2) constitutes the scalar transition (or state) equation. Upon stacking the shifting-endpoint model (3), the conditional forecasts (6), and their relationship with various data points from surveys (7), we obtain the measurement equations

$$y_t = A_t z_{t-1} + B_t \mu_{t-1} + w_t \,, \tag{8}$$

where y_t is an $(m+1)$-vector that contains current date information on inflation and surveys; A_t and B_t are, respectively, an $(m+1) \times p$ matrix and $(m+1)$-vector of coefficients; and w_t is an $(m+1)$-vector of measurement errors. The entries of A_t and B_t depend on h_{it} and thus may be time varying, capturing the shrinking nature of the forecasting horizon for FEFs. This is the most important difference with respect to the model analyzed by Kozicki and Tinsley (2001, 2012).

More explicitly, let $a(h) = e_1{}' C^{h+1}$ (a p-row vector) and $b(h) = 1 - a(h) i_p$ (a scalar). Then,

$$\begin{bmatrix} \pi_t \\ F_{1t} \\ F_{2t} \\ \vdots \\ F_{mt} \end{bmatrix} = \begin{bmatrix} a(0) \\ a(h_{1t}) \\ a(h_{2t}) \\ \vdots \\ a(h_{mt}) \end{bmatrix} z_{t-1} + \begin{bmatrix} b(0) \\ b(h_{1t}) \\ b(h_{2t}) \\ \vdots \\ b(h_{mt}) \end{bmatrix} \mu_{t-1} + \begin{bmatrix} \epsilon_t \\ \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{bmatrix} , \tag{9}$$

where each vector and system matrix in (8) is implicitly defined.

It is worth mentioning that although measurement equations (9) are designed to accommodate m different survey responses at different horizons, they are flexible enough to adapt to many other

[5] If desired, a refined prediction conditioned on all the available information in the sample can be obtained by means of a smoothing algorithm that uses the output of the Kalman filter.

structures. One case would be if the surveys contain both REFs and FEFs: for REFs the horizons are fixed $h_{it} = h_i$, whereas for FEFs the forecasting horizons will vary with t in a deterministic fashion. For instance, in surveys such as the Consensus Forecasts, the data from month M_t refer to forecasts by the end of year i, with $i = 1$ being the current year, so the forecasting horizons (in months) evolve deterministically as $h_{it} = 12i - (M_t - 1)$.

Surveys may also report *averages* of forecasts over R different horizons. In this case, the corresponding measurement equation would read $F_t = (1/R)\sum_{r=1}^{R}[\,a(h_{rt})z_{t-1} + b(h_{rt})\mu_{t-1}\,] + \varepsilon_t$. We did attempt to incorporate information of this kind in our empirical application below, but with marginal effects on the results. Thus, we ignore such forecasts and stick to the model (9) as it stands.

On the other hand, we have implicitly interpreted $F_{1t}, \ldots, F_{mt}$ as aggregate responses for different forecasting horizons. This is just for expositional convenience, since (9) can handle responses made by different agents, by simply allowing subindex i to denote a forecaster/horizon pair. For instance, for two forecasters and two horizons, F_{1t}, F_{2t} can represent the results for the first participant, and F_{3t}, F_{4t} for the second, with $h_{1t} = h_{3t}$ and $h_{2t} = h_{4t}$ indicating that there are only two target horizons. In this case, more structure to the measurement error covariances, in the spirit of Davies and Lahiri (1999), may be appropriate—for instance, by letting $\varepsilon_{it} = R_i + H_t + \text{error}_{it}$, where R_i and H_t are, respectively, respondent and horizon effects.[6]

2.4 Missing Data

Generally, survey data are available less frequently than inflation data (which are assumed to be available for all periods in the sample), so for certain periods some of the observations in the y_t vector will be missing. The treatment of missing observations is perhaps one of the clearest advantages of using the Kalman filter to process

[6] A further extension of the model allows expectations to be formed at different moments. The measurement equation for an expectation formed with information up to period $t - n$ ($n \geq 1$) is $F_t = a(h_t + n - 1)z_{t-n} + b(h_t + n - 1)\mu_{t-n} + \varepsilon_t$. Thus (8) becomes $y_t = A_t Z_t + B_t \bar{\mu}_t + w_t$, where Z_t and $\bar{\mu}_t$ stack, respectively, the lags of z_t and μ_t associated with each information set. The number of columns of the now sparse matrices A_t and B_t should vary accordingly.

time-series models, as the filter requires only minor modifications to deal with such a problem (see Harvey 1989, p. 144).

Consider the case when only $\bar{m}_t < m$ forecasts in y_t are available at period t, and let D_t be the $(\bar{m}_t + 1) \times (m + 1)$ selection matrix that collects non-missing elements; so, for instance, $y_t^* = D_t y_t$ is an $(\bar{m}_t + 1)$-vector that contains observable data. Note that D_t is formed by $\bar{m}_t + 1$ rows of the identity matrix of order $m + 1$. Then, upon pre-multiplying the measurement equations (8) by D_t,

$$y_t^* = A_t^* z_{t-1} + B_t^* \mu_{t-1} + w_t^*, \tag{10}$$

where A_t^*, B_t^*, and w_t^* contain the rows of A_t, B_t, and w_t, respectively, that correspond to the available observations in y_t^*. The state variable μ_t is not affected by the transformation, so the Kalman filter can be applied normally to the observable system (10) in period t. The missing observations contribute neither to the state predictions nor to the likelihood function.

2.5 Signal Extraction and the Term Structure of Expectations

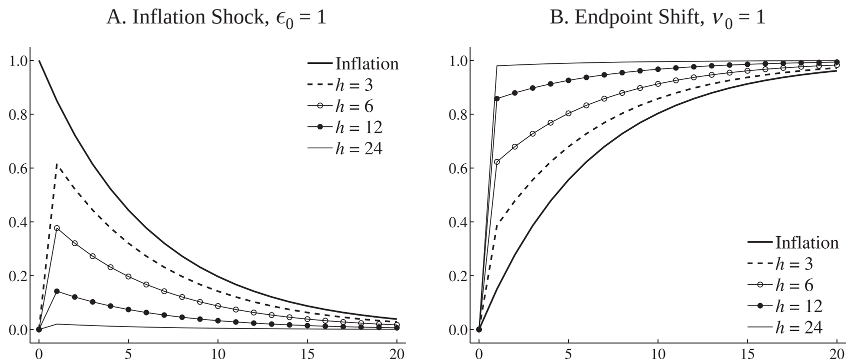
From (6) and (7), we can readily verify that

$$\pi_{t+h\,|\,t-1} = a(h) z_{t-1} + b(h) \mu_{t-1}\,, \tag{11}$$

so that expected inflation is estimated as a linear combination of the latest realizations of inflation z_{t-1} and the perceived long-run level μ_{t-1}. Note $\lim_{h\to\infty} a(h) = 0$ and $\lim_{h\to\infty} b(h) = 1$. Thus, as h increases, expectations converge from a short-term forecast dominated by recent history to the endpoint (1). By the same token, equation (7) implies that (roughly) $F_{it} \simeq \pi_{t-1} + \varepsilon_{it}$ for small h_{it}, and $F_{it} \simeq \mu_{t-1} + \varepsilon_{it}$ for large h_{it}. Therefore, in the process of extracting the signal μ_t, more importance is given to long-term survey expectations (when available), even though inflation data and shorter-term expectations contain also information about this hidden state. Kozicki and Tinsley (2012) provide a more detailed discussion on the information content of the observable variables. See also the discussion in section 3.3 below.

These properties of the model are illustrated in figure 2. For $\Phi(L) = \Phi(1) = 0.85$, the figure shows the response of inflation and expectations formed with information up to period $t - 1$, to the two

Figure 2. Inflation and Expectations Responses to Shocks



Note: The horizontal axis shows the number of periods (months) after the shock occurs at period 0. The vertical axis shows the response of inflation and expectations in percent terms.

shocks of the system. An inflation shock $\epsilon_0 = 1$, panel A, produces a persistent deviation in the inflation rate from the endpoint. Given the transitory nature of the shock, for short-run horizons (say, $h \leq 6$) expected inflation reflects the temporarily higher current inflation; in contrast, expectations for longer horizons (say, $h \geq 12$) are virtually unaffected, as agents correctly anticipate that the influence of the shock would vanish almost completely by the end of the forecast horizon. On the other hand, expectations are sensitive to shifts in the endpoint $\nu_0 = 1$, panel B. Such a shock causes a sluggish response in inflation, which displays a smooth transition towards its new long-run level, that in turn affects short-run expectations; on the other hand, the shift passes through immediately and almost completely to long-run expectations, as agents can anticipate that the change in the endpoint will be completely transferred to inflation by the end of the forecast horizon.

The state-space form includes error-ridden versions of (11) for time-varying horizons where survey data is available. Thus, even though survey data may be limited to infrequently sampled observations and to selected horizons, once the model is estimated, it can be used to construct a complete term structure of expected inflation. Put differently, in practice we only observe some points of the responses shown in figure 2, but the model allows us to interpolate the entire profile of responses in a coherent, data-consistent way.

2.6 Model Variants

In our empirical exploration, we also analyze three variants of the model within the same state-space framework, in order to highlight the most important features of the shifting-endpoint model that uses both inflation and survey data. We are particularly interested in assessing the model's ability to simultaneously explain inflation and expectations data under competing specifications.

The first variant assumes a constant endpoint (CE), $\mu_t = \mu$. This can be achieved by setting $\text{var}(\nu_t) = 0$ and treating the initial condition $\mu_0 = \mu$ as an additional parameter to be estimated. Provided that $\Phi(1) < 1$, inflation is treated as a stationary, mean-reverting process.

In contrast, the second variant imposes a unit root (UR) to the autoregressive polynomial in (3), $\Phi(1) = 1$. In this case $C^h i_p = i_p$ for any $h > 0$, and the companion form (5) reduces to $z_t = Cz_{t-1} + e_1\epsilon_t$. Since it turns out that $B_t = 0$ in all measurement equations, μ_t is not identified and cannot be computed with the Kalman filter. However, μ_t can be reformulated to be the limiting forecast which continues to exist: it can be verified that $\lim_{h\to\infty} C^h = \bar{C}$, where $C\bar{C} = \bar{C}$, and so $\mu_t = e_1{}'\bar{C}z_t$. The endpoint is a moving average of the most recent inflation observations, and Kozicki and Tinsley (1998) show that, following the definition in (1), it corresponds to the permanent component of the Beveridge-Nelson decomposition of the inflation equation. The Kalman filter can still be used to evaluate the likelihood function in this case, under $\text{var}(\nu_t) = 0$.

The last variant is a model of inflation that ignores the information from surveys. The result is a univariate generalization of the local level (LL) model described in Harvey (1989, ch. 2), where inflation is decomposed as the sum of a random walk (the endpoint) and a zero-mean stationary component. This variant can be easily treated by considering a single measurement equation, the first row in (9), or by making the variances of the approximation errors arbitrarily large, $\text{var}(\varepsilon_{it}) = \kappa \to \infty$ for all i.

3. Application to Latin American Countries

After a long history of high inflation, during the 1990s many Latin American countries adopted a series of reforms that would eventually bring inflation down to single-digit levels (see, inter alia, Mishkin and Savastano 2001). Even though experiences may differ in the detail, many commonalities across countries can be identified (see Corbo and Schmidt-Hebbel 2001; Quispe-Agnoli 2001). First, to facilitate the reduction of inflation, and to isolate monetary policy from political pressures, these countries granted independence to their central banks at an early stage of the stabilization effort. Then, economic authorities would adopt a rudimentary form of inflation targeting, typically by simply announcing numerical targets (or official forecasts), as a first attempt to anchor market expectations. The process of disinflation would be gradual as the central banks improved their credibility and built a reputation as inflation targeters. Once stable levels of inflation were reached, the central banks would adopt a fully fledged inflation-targeting regime, characterized by the announcement of long-run targets and the abandonment of any other nominal anchor (typically, currency depreciation or money growth).

We estimate the shifting-endpoint model of section 2 using data from four successful Latin American inflation targeters: Chile, Colombia, Mexico (all of which adopted the regime in 1999), and Peru (which adopted the regime in 2002).[7] Following Vega and Winkelried (2005), the central banks began announcing numerical targets long before the definite adoption date: 1991 in Chile, 1995 in Colombia and Mexico, and 1994 in Peru. Thus, by analyzing the evolution of the estimated endpoints, our empirical application aims to assess the role of expectations in the initial disinflation and the subsequent periods of price stability.

3.1 Data

Monthly inflation corresponds to the percent variation of the officially targeted consumer price index. These data and the numerical targets come from each central bank's website.

Survey data, on the other hand, are extracted from the well-known Latin American Consensus Forecasts reports. These reports

[7] Brazil is also an important Latin American inflation targeter. Its inflationary experience is different from those of the listed countries, enough to deserve special treatment. Hence, for the sake of brevity, the results for the Brazilian case (which are available upon request) are not reported.

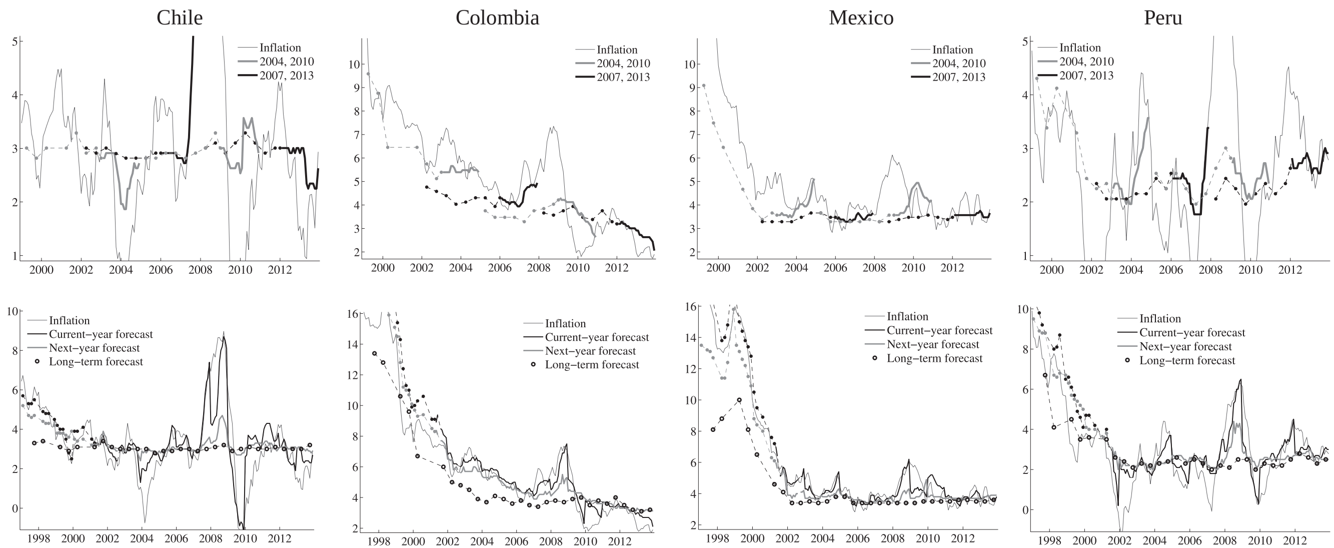
are available bi-monthly (alternate, even months) between April 1993 and April 2001, and monthly thereafter (since May 2001). Each report surveys a number of prominent financial and economic analysts and publishes their individual forecasts as well as simple descriptive statistics. The focus of our analysis is on the mean (i.e., the "consensus") forecast F_{it}. The surveys are conducted by the middle of the month, after the inflation figure for the previous month is released. Thus, in agreement with the shifting-endpoint model assumptions, in period t expectations are formed with information up to period $t-1$.

The typical issue of the Latin American Consensus Forecasts provides forecasts for current-year and next-year inflation. Denoting the year of the survey as year 1, these forecasts correspond to the measurements F_{1t} and F_{2t}. We refer to these variables jointly as "short-term" forecasts. In addition, the April and October issues include also "long-term' (up to ten years ahead) forecasts. Forecasts for years 3 to 6 $(F_{3t}, F_{4t}, F_{5t}, F_{6t})$ are explicitly published, whereas forecasts for longer horizons (from year 7 to year 11) are reported as an average. In practice, such averages show little variation with respect to F_{6t} and are thus discarded from our analysis with no consequential effects on our results. Thus, depending on the month, the measurement equations have a minimum of $m = 2$ horizons and a maximum of $m = 6$. As mentioned earlier, the forecasting horizons evolve deterministically as $h_{it} = 12i - (M_t - 1)$, where M_t denotes the month of the survey.

Our sample spans from February 1997 to December 2013. Thus, for each country, the estimations are based on 203 observations on inflation, 178 observations on "short-term" forecasts, and 34 observations on "long-term" forecasts.

Figure 3 displays some of the data. The first row plots the evolution of the FEFs as they approach the inflation outturn (to avoid clutter, only the events 2004, 2007, 2010, and 2013 are displayed). For a given event, each point corresponds to a different survey. The second row presents the data as they enter the regression model: the short-term forecasts F_{1t} and F_{2t}, and the longest-term forecast F_{6t}, all of them reported at the same time t. Each month is associated with a different forecast horizon, so the strong swings in F_{1t} and F_{2t} are due to a change in the event to be forecast (say, $h_{1t} = 1$ in December and then $h_{1t} = 12$ in January).

Figure 3. Inflation, Selected Fixed-Event Forecasts, and Moving-Horizon Forecasts



Note: The dots in the figure represent forecasts that are irregularly sampled (long-term forecasts in all of the sample period, and short-term forecasts until April 2001). To ease visualization, these dots are connected with linearly interpolated values represented by discontinuous lines (the interpolations are not used in the estimation). Within each row, the scale of the vertical axis is the same for Chile/Peru and Colombia/Mexico.

A preliminary analysis of these surveys provides prima facie evidence of the adequacy of the shifting-endpoint model to explain the behavior of measured expectations. An important implication of the model is that the deviation of inflation from the endpoint is expected to be a mean-reverting process around zero. Also, short-term forecasts would be influenced by recent news in inflation, whereas long-term forecasts should be determined by the endpoint. Consider the regression equation

$$F_{it} - F_{6t} = \alpha_i + \beta_i(\pi_{t-1} - F_{6t}) + \text{error}_t\,, \tag{12}$$

where π_{t-1} denotes the latest inflation figure (the forecast origin), and F_{6t} is the the forecast corresponding to the longest horizon available, which serves as a rough proxy of the moving endpoint. The shifting-endpoint model is to be regarded as a reasonable description of how expectations are formed if both α_i and β_i approach zero as i increases. This is exactly the pattern that emerges in table 1. The estimates decrease with i in all cases and often lose statistical significance for $i \geq 3$ onwards (note that the sample size for these estimations is constrained by the thirty-four observations available for long-term forecasts, so the results should be taken as indicative rather than conclusive).

3.2 Estimation Results

Next, we present the estimation results of the shifting-endpoint model (SE), and its variants (CE, UR, and LL). We use a diffuse prior (i.e., setting the initial state variance to a very large number) to initialize the Kalman filter. In all cases, the lag length of the autoregressive model for inflation is set to $p = 13$, which is the value that minimizes the Schwarz information criterion. Also, the null hypothesis that the approximation errors in (7) are not contemporaneously correlated was contrasted with a likelihood-ratio test, and categorically rejected in all instances. Hence, these covariances are estimated unrestrictedly.

For each model, table 2 presents the sum of the estimated autoregressive coefficients $\Phi(1)$ as a measure of persistence of inflation around the endpoint, which is restricted to $\Phi(1) = 1$ in the UR model; the estimated standard deviation of the inflation shock, $\text{std}(\epsilon_t)$; the estimated standard deviation of the endpoint shock,

Table 1. Mean Reversion of Forecasts

		F_1–F_6	F_2–F_6	F_3–F_6	F_4–F_6	F_5–F_6
Chile	α	0.257 (0.147)*	0.181 (0.070)*	0.083 (0.047)*	0.055 (0.039)	0.013 (0.023)
	β	0.565 (0.119)*	0.116 (0.035)*	0.026 (0.027)	0.007 (0.021)	0.004 (0.016)
Colombia	α	0.525 (0.201)*	0.451 (0.194)*	0.276 (0.180)	0.166 (0.163)	0.036 (0.096)
	β	0.696 (0.090)*	0.413 (0.110)*	0.262 (0.128)*	0.159 (0.112)	0.066 (0.048)
Mexico	α	0.101 (0.113)	0.005 (0.085)	0.028 (0.054)	0.019 (0.037)	0.016 (0.034)
	β	0.732 (0.058)*	0.395 (0.032)*	0.204 (0.020)*	0.096 (0.016)*	0.037 (0.017)*
Peru	α	0.422 (0.156)*	0.357 (0.143)*	0.201 (0.099)*	0.077 (0.053)	0.015 (0.046)
	β	0.544 (0.111)*	0.191 (0.095)*	0.059 (0.062)	0.045 (0.032)	0.015 (0.032)

Notes: The table shows least-squares estimates of equation (12), using the thirty-four available observations for long-term forecasts. HAC standard errors are shown in parentheses. * denotes coefficients different from zero at a 5 percent significance level.

Table 2. Estimation Results

		Model Specification				RMSE	RMSE Survey Forecasts			
		$\Phi(1)$	std(ε_t)	std($\Delta\mu_t$)	std(ε_t)	Inflation	1	2	3	$\geq$4
Chile	CE	0.931 (0.022)	0.379	0.000	0.473	0.38	0.88	0.93	0.36	0.24
	UR	1.000 (0.000)	0.387	0.000	1.417	0.39	1.07	1.37	1.62	1.64
	LL	0.544 (0.043)	0.311	0.286	1.038	0.46	0.64	1.06	1.17	1.24
	SE	0.677 (0.040)	0.405	0.175	0.844	0.43	0.57	0.78	0.56	0.54
Colombia	CE	0.979 (0.011)	0.309	0.000	1.322	0.33	0.99	0.98	1.50	1.62
	UR	1.000 (0.000)	0.333	0.000	1.570	0.34	1.07	1.03	1.44	1.69
	LL	0.480 (0.054)	0.242	0.257	1.122	0.43	0.41	0.83	1.11	1.45
	SE	0.713 (0.051)	0.324	0.193	0.925	0.36	0.53	0.62	0.77	1.06
Mexico	CE	0.981 (0.007)	0.278	0.000	0.798	0.28	0.60	0.72	0.60	0.80
	UR	1.000 (0.000)	0.292	0.000	1.750	0.29	0.82	1.29	1.80	2.21
	LL	0.564 (0.051)	0.219	0.241	1.263	0.34	0.58	1.01	1.68	2.08
	SE	0.790 (0.031)	0.291	0.161	0.987	0.30	0.51	0.66	0.76	1.01
Peru	CE	0.949 (0.015)	0.334	0.000	0.745	0.34	0.94	0.91	0.76	0.59
	UR	1.000 (0.000)	0.346	0.000	1.037	0.35	0.97	0.96	1.26	1.25
	LL	0.511 (0.041)	0.268	0.234	0.819	0.41	0.50	0.82	1.05	1.11
	SE	0.693 (0.038)	0.369	0.143	0.606	0.39	0.50	0.60	0.78	0.74

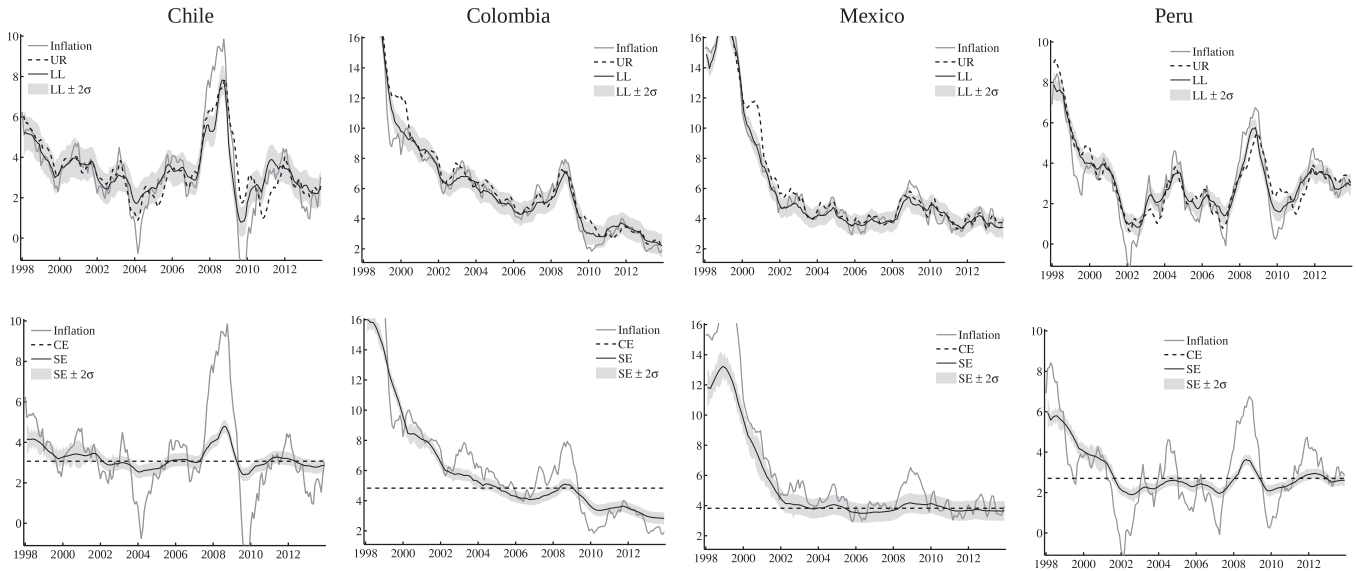
Notes: The table shows maximum-likelihood estimates, using data from February 1997 to December 2013. All models use $p = 13$. CE: constant-endpoint model; UR: unit-root model; LL: local-level model; SE: shifting-endpoint model. $\Phi(1)$: sum of the autoregressive coefficients, $\Phi(1) = \sum_{i=1}^{p} \phi_i$, (robust standard errors in parentheses) restricted to $\Phi(1) = 1$ in the UR model; std(ε_t): standard deviation of the inflation shock; std($\Delta\mu_t$): standard deviation of the endpoint shock, restricted to std($\Delta\mu_t$) $= 0$ in the CE and UR models; std(ε_t): *average* standard deviation of the approximation errors of the $m = 6$ survey measures; "RMSE" is the root mean square error of the one-step-ahead predictions produced by the Kalman filter for inflation and survey forecasts for years 1 (current), 2 (next), 3, and $\geq$4 (subsequent) (the column labeled "$\geq$4" shows the average RMSE for the remaining three long-term forecasts).

std($\Delta\mu_t$), which is restricted to std($\Delta\mu_t$) $= 0$ in the CE and UR models; and, to save space, the *average* standard deviation of the approximation errors of the $m = 6$ survey measures, std(ε_t). In addition, as measures of fit, the table presents the root mean square error (RMSE) of the one-step-ahead predictions produced by the Kalman filter for the observed data, i.e., inflation and survey forecasts. In the case of the LL model, we proceed in two steps: first, the inflation parameters were estimated using only the first measurement equation; second, the remaining output in the table was obtained *conditional on the first-stage estimates.* Figure 4 shows the predicted endpoints, evaluated at the maximum-likelihood estimates.

There are various similarities in the estimations across countries. In all cases, the UR model tracks actual inflation quite closely, but it performs rather poorly when it comes to survey data. The RMSEs for inflation are amongst the lowest across models (ranging from 0.29 in Mexico to 0.39 in Chile), whereas the RMSEs for the survey forecasts are the highest (ranging, for long-term forecasts, from 1.64 in Chile to 2.21 in Mexico).

This finding is related to the mean-reverting properties of expectations discussed in table 1 which, by construction, the UR model completely overlooks. In this respect, the LL model constitutes an improvement upon the UR model. Figure 4 reveals that the predicted endpoint in the LL model can be regarded as a de-noised and smoother version of the moving-average endpoint of the UR model; inflation expectations revert quickly from the observed data to such an estimated trend (the sum of autoregressive coefficients is $\Phi(1) \simeq 0.5$ in all cases), which improves the model's performance to account for the variability of short-term forecasts. Compared with the UR model, the RMSE of the LL model for the current-year inflation forecast (i.e., year 1) reduces sizably from 1.07 to 0.64 in the case of Chile, from 1.07 to 0.41 in Colombia, from 0.82 to 0.58 in Mexico, and from 0.97 to 0.50 in Peru. Long-term forecasts, nonetheless, are also poorly predicted in the LL model, as survey data appear to be much less sensitive to inflation news than what is implied in the LL model. Thus, for long-term forecasts, the reductions in RMSE compared with the UR model are quite modest: from 1.64 to 1.24 in the case of Chile, from 1.69 to 1.45 in Colombia, from 2.21 to 2.08 in Mexico, and from 1.25 to 1.11 in Peru.

Figure 4. Predicted Endpoints by Model Variant



Note: CE: constant-endpoint model; UR: unit-root model; LL: local-level model; SE: shifting-endpoint model. For the CE, the figure displays the estimated value of μ. For the UR, the moving-average endpoint, $\mu_t = e_1{}'\bar{C}z_t$ (see section 2.6). Only point estimates are reported for these models in order to avoid overloading the graphs. For the LL and SE models, the figure shows the smoothed predictions along with a 95 percent confidence interval. The scale of the vertical axis is the same for Chile/Peru and Colombia/Mexico.

On the other hand, in the CE model the sum of autoregressive coefficients $\Phi(1)$ is significantly higher than in the LL and SE models (where this quantity is estimated unrestrictedly). In the latter specifications, some of the persistence in observed inflation is attributed to the dynamics of a time-varying mean, and some to short-run deviations from this mean. Since inflation in the CE model is assumed to revert to an imposed constant level, the process is unsurprisingly estimated as highly persistent. Thus, short-term forecasts (especially for the current year) are similar to those of the UR model: the RMSEs for the current-year forecast range from 0.60 to 0.99 in the CE model, as compared with 0.82 to 1.07 in the UR model. However, with the exception of Colombia, the performance of the CE in fitting survey data improves rapidly and dramatically for long-term forecasts. A unit increase in i (the year index) implies an increase of $h_{i+1,t} - h_{it}$ = twelve months in the associated forecast horizon, so despite the large values of $\Phi(1)$, current inflation becomes less influential even for small values of i. As a result, the CE estimates turn out to be closer to the sample average of long-term forecasts than to the sample average of observed inflation (see figure 4), and the CE model does remarkably well in explaining long-term forecasts, three years ahead and beyond, in Chile (with an RMSE of 0.36 for the forecast three years ahead and of 0.24 for the "$\geq$4" forecast), Mexico (0.60 and 0.80), and Peru (0.76 and 0.59). For these countries, inflation and expectations have been fluctuating (albeit, persistently) within a relatively narrow range for most of the sample period. In contrast, inflation trends downwards during all of the sample period in the Colombian case, making the CE model unsuitable.

The SE model is a compromise between the LL and the CE models. Survey information produces a smoother endpoint than the LL model, as shown in figure 4, which is manifested in a reduction of the standard deviation of the endpoint shocks from $\text{std}(\Delta\mu_t) \simeq 0.25$ to $\text{std}(\Delta\mu_t) \simeq 0.17$ and, more importantly, to significant increases in the noise-to-signal ratios $q = \text{var}(\epsilon_t)/\text{var}(\Delta\mu_t)$ from $q \simeq 1$ to $q \in [2.8, 6.7]$. Given the actual inflation persistence, a smoother endpoint is traded with an increase in $\Phi(1)$, from $\Phi(1) \simeq 0.5$ to $\Phi(1) \simeq 0.7$. With this, the SE model accounts for short-term forecasts' variability as much as the LL model, while clearly outperforming the LL model to explain long-term forecasts. When compared with the CE model, the SE model performs considerably better when it comes to short-term forecasts (years 1 and 2 in the table) but, except in the Colombian case, it is outperformed when predicting long-term forecasts. However, the improvement in the short-term fit of the SE model seems to more than compensate for the moderate deterioration in fitting longer-term expectations. As we shall discuss below, the improvement brought by the SE model is on the forecasting horizons that can be regarded as relevant for monetary policy analysis.

3.3 Signal Extraction

It has been argued that the SE model is able to disentangle short-run inflation shocks from long-run expectation shocks. Furthermore, it was argued that the SE would interpolate a complete profile of expectations from sparsely sampled survey information. Next, we present details on the computations involved in the prediction of the endpoint in the SE model.

From the Kalman filter, the endpoint is determined following the recursion $\mu_t = \mu_{t-1} + K_t v_t$, where K_t is the Kalman gain and v_t is the vector of prediction errors of the observed data. Thus, K_t measures how these errors are weighted to arrive at an updated inference about the unobserved state μ_t. To get a better understanding of how the endpoint is finally determined in our application, table 3 presents the implied Kalman gains in the SE model for selected dates, and for the case of Peru (an identical pattern is found for the remaining countries).

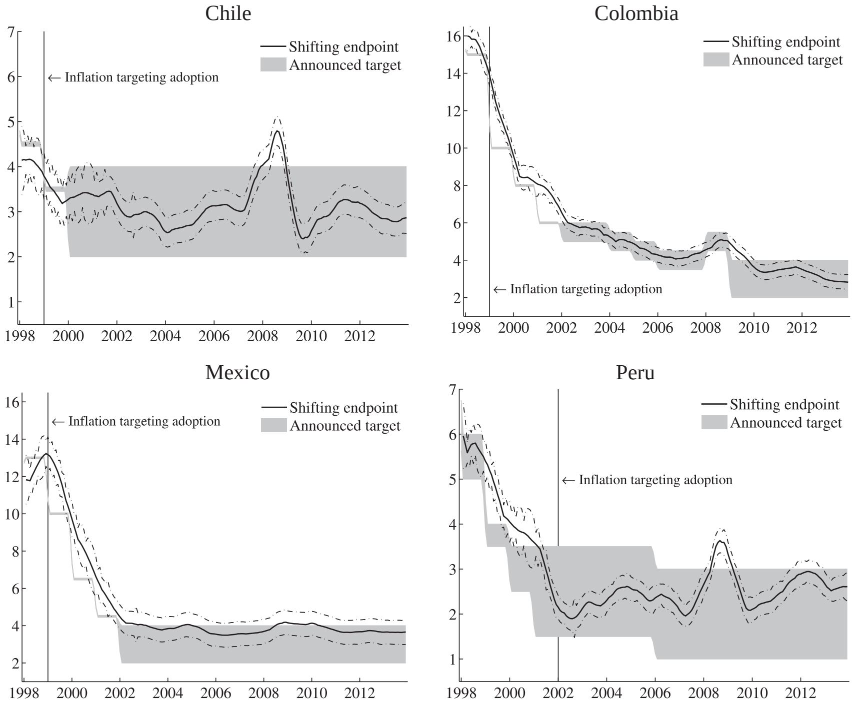
By construction, the gain is exactly zero for missing data (all survey data in March 2001 and long-term forecasts in February 2001, September 2006, and March 2013), which follows from the adjustments described in section 2.4. On the other hand, upon comparing the results from February to April 2001, September to October 2006, or March to April 2013, it can be seen that whenever long-term forecasts are available, the Kalman filter gives more weight to them, and less to actual inflation, to predict the endpoint. This is exactly what one would want: long-horizon expectations should provide more information about the endpoint and, thus, should receive more weight in the filtering process. When only short-term forecasts are available, the gains for F_1 and F_2 are of comparable magnitude and always higher than that of inflation. When long-term forecasts

Table 3. Kalman Gains in the Peruvian SE Model

	Inflation	F_1	F_2	F_3	F_4	F_5	F_6
February 2001	0.1036	0.1767	0.1661	0	0	0	0
March 2001	0.1187	0	0	0	0	0	0
April 2001	0.0654	0.1410	0.0720	0.0167	0.0364	0.0669	0.1364
September 2006	0.0640	0.0985	0.0947	0	0	0	0
October 2006	0.0528	0.0992	0.0550	0.0160	0.0220	0.0568	0.1034
March 2013	0.0621	0.1047	0.0993	0	0	0	0
April 2013	0.0492	0.1062	0.0542	0.0126	0.0274	0.0504	0.1028

Notes: The Kalman gain measures how prediction errors of the observed data are weighted in order to get an updated prediction of μ_t. The table shows these weights for selected dates. The gain is set to zero for missing values. Survey data are not available for March 2001, whereas long-term forecasts are only available in April and October.

Figure 5. Predicted Endpoints and the Announcement of Inflation Targets



Note: The scale of the vertical axis is the same for Chile/Peru and Colombia/Mexico.

become available, the gain of F_2 halves while the gain of F_6, the furthest forecast, increases to a level close to the gain of F_1. Interestingly, since the prediction errors are typically larger for F_6 than for F_1 (see the RMSE results in table 2), the absolute magnitude of the correction would also be larger for F_6 than for F_1. Thus, when F_6 is observed, the Kalman filter updates the endpoint prediction towards it.

3.4 Inflation Targeting and Disinflation

Figure 5 shows the evolution of the predicted endpoints along with the inflation targets for each country *shifted forward* one year. For instance, Chile's inflation target was 3.5 percent for 2000, and this value appears during 1999 in the figure.[8] The purpose is to illustrate the joint evolution of the endpoint and the *announcements* of the inflation targets, and thus we are assuming that the central banks have announced their targets one year in advance. This was the actual practice during disinflation; however, all central banks have progressively increased the horizon for the targets as inflation has reached long-run levels.

During disinflation, roughly until 2002, the announcements have clearly served as a benchmark for private economic agents' long-run forecasts. Shocks to the endpoints bear an almost one-to-one correspondence to changes in the announced targets.

From 2002 onwards, long-run expectations have lain within the target ranges in almost all instances. The remarkable exception is year 2008, especially for Chile and Peru. During 2007 and most of 2008, these economies where hit by a sequence of large food price and oil shocks that deviated inflation from its target. The shocks were persistent enough that they affected long-run perceptions. However, the endpoint returned rapidly to the inflation targets by late 2008 as the result of both aggressive increases in the monetary policy rates and the start of the global financial crisis. It has remained within the target range ever since.

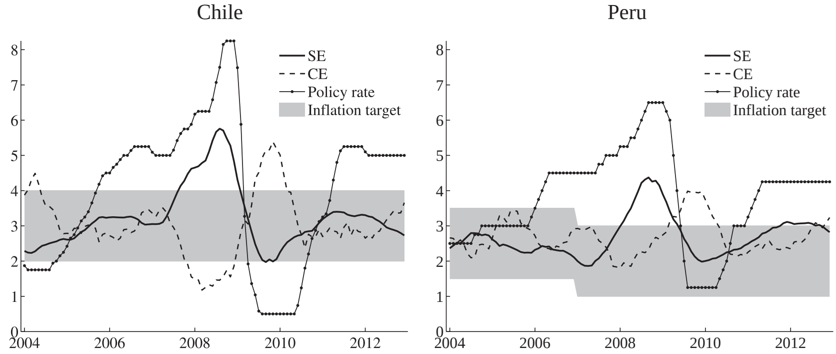
Given these dynamics, our results support the conclusion that the countries in our sample are good examples of successful inflation-targeting experiences (see Corbo and Schmidt-Hebbel 2001). Their central banks have managed to establish a credible regime of stable inflation with anchored expectations.

3.5 The Shifting-Endpoint Model as an Early-Warning Device

The SE model seems useful to better understand how inflation expectations evolve over time by exploiting available survey information. In this section we explore the real-time properties of this model and how it summarizes relevant information for the day-to-day conduct of monetary policy. To this end, we focus on the cases of

[8]Chile announced point targets until 2000; thereafter, the inflation targets have been published as ranges (actually, central values surrounded by a symmetric tolerance level). The same occurred in Colombia and Mexico until 2002. Peru has announced target ranges during all of the sample period.

Figure 6. Eighteen-Month-Ahead Inflation Expectations and Monetary Policy Interest Rates



Chile and Peru, with emphasis on the period from 2007 to 2009. Even though these two cases are generally regarded as successful inflation-targeting experiences, our previous findings indicate that expectations were temporarily unanchored and moved away from their target ranges by late 2007/early 2008 (cf. figure 5). As mentioned, before the international financial crisis in late 2008, both economies experienced, on one hand, a series of persistent food and fuel price shocks and, on the other hand, a sequence of positive demand shocks (a byproduct of extremely favorable terms of trade) that eventually passed through inflation and its expectations (see Cespedes, Chang, and Velasco 2014 for a general review).

Based on the (plausible) assumption that the relevant monetary policy forecasting horizon lies between one and two years, figure 6 presents the evolution of the eighteen-month-ahead inflation expectations implied by the SE and CE models, computed as in equation (11).[9] From the estimation results in table 2, these are the two models that fare best in accounting for the behavior of long-run expectations. Besides, on a priori grounds, the CE model seems to

[9]Here, we seek to mimic the real-time inference that could be made by a researcher whose information set ends in period t. Thus, unlike the predictions presented in figures 4 or 5 that correspond to *smoothed* predictions, i.e., made with all information in our sample, the interpolated expectations in figure 6 are based on *filtered* or *updated* predictions on the endpoint, i.e., made with information up to period t.

be an adequate alternative to the SE model for the cases and samples selected in this analysis, where inflation has been fluctuating around the target ranges.

Yet, it is quite apparent that the SE predictions lead the CE predictions with a considerable lag. Recall that whereas the endpoint is able to respond to expectation news in the SE model, it is held constant in the CE model. Hence, before the financial crisis, medium-term inflation expectations seemed to have increased substantially, a fact recorded by the surveys and thereby reflected in the dynamics of the endpoint predicted by the SE model and its expectations at various horizons. On the contrary, the expectations of the CE model reflect the transition of current inflation to a given steady-state level, which is determined by very long-run forecasts, and so for the relevant horizons for monetary policy analysis the CE model expectations resemble actual inflation too much. Similar dynamics are recorded after late 2008, where the endpoint and expectations rapidly return to the target ranges in the SE model, and more sluggishly in the CE model. This differentiated behavior can also be observed in table 2, where the RMSE of the CE model is substantially higher than that of the SE model for short-term forecasts (year 1: respectively, 0.88 versus 0.57 in the case of Chile, and 0.94 versus 0.50 for Peru) and medium-term forecasts (year 2: respectively, 0.98 versus 0.78 for Chile, and 0.91 versus 0.60 for Peru).

Figure 6 also shows the path of the monetary policy interest rates. Being inflation targeters, it is reasonable to suppose that the movements of the policy rate reflect the perceptions about future inflation of the central bank, using all available information. In both countries, the strong correlation (or at least the directional co-movements) between the actual policy rates and the expectations inferred from the SE model is quite noticeable. There are many ways to rationalize such co-movements. For instance, private expectations may be responding to a high (low) inflation forecast made by the central bank; such a forecast may be driving an increase (decrease) in the interest rate and, as long as its effects are recorded in the surveys, may also be signaling an increase (decrease) in the inferred endpoint. On the other hand, the central bank may be directly responding to private expectations, especially when it suspects that they may have been affected by long-lived shocks. In either way, this exercise illustrates the extent to which the SE model is capable of summarizing, in a satisfactory way, real-time information regarded as relevant for monetary policy decisions. This information, of course, is captured by the FEFs of inflation.

4. Closing Remarks

Fixed-event forecasts provide a widespread, yet unexplored, source of inflation expectations in many countries. The main difficulty is that the very structure of the FEFs, especially the fact that they correspond to moving forecast horizons, hinders their direct applicability in empirical work. To overcome this hindrance, and to infer about the term structure of inflation expectations from FEFs, we have proposed an extended version of the shifting-endpoint model of Kozicki and Tinsley (2012). Even though the resulting model is time varying, it can be easily handled with the Kalman filter, even for irregularly sampled survey expectations. By fitting the shifting-endpoint model to Latin American data, we conclude that it is able to jointly account for the behavior of actual inflation and survey records, often outperforming competing specifications. Our empirical exploration also suggests that survey FEFs provide a valuable source of information on expected inflation, complementary to that contained in historical records of inflation. Finally, the extended shifting-endpoint model may be a useful tool to track the behavior of inflation and its expectations, as reflected in the FEF records, in real time.

Given the availability of FEFs, exploring alternative methods to readily and effectively use such data in econometric models is likely to have important practical implications. Moreover, it would be of interest to compare the performance of FEFs versus REFs, or to assess the informational content of one type of forecast relative to the other, for countries where both are available. We hope our analysis proves to be a meaningful contribution to this promising research agenda.

References

Ang, A., G. Bekaert, and M. Wei. 2007. "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics* 54 (4): 1163–1212.

Bakhshi, H., G. Kapetanios, and T. Yates. 2005. "Rational Expectations and Fixed-Event Forecasts: An Application to UK Inflation." *Empirical Economics* 30 (3): 539–53.

Cespedes, L. F., R. Chang, and A. Velasco. 2014. "Is Inflation Targeting Still on Target? The Recent Experience of Latin America." *International Finance* 17 (2): 185–208.

Chernov, M., and P. Mueller. 2012. "The Term Structure of Inflation Expectations." *Journal of Financial Economics* 106 (2): 367–94.

Clements, M. P. 1997. "Evaluating the Rationality of Fixed-Event Forecasts." *Journal of Forecasting* 16 (4): 225–39.

Corbo, V., and K. Schmidt-Hebbel. 2001. "Inflation Targeting in Latin America." Working Paper No. 105, Central Bank of Chile.

Davies, A., and K. Lahiri. 1999. "Re-examining the Rational Expectations Hypothesis Using Panel Data on Multi-period Forecasts." In *Analysis of Panels and Limited Dependent Variable Models*, ed. C. Hsiao, K. Lahiri, L. F. Lee, and M. H. Pesaran, 226–54 (chapter 9). Cambridge University Press.

Harvey, A. C. 1989. *Forecasting, Structural Time Series and the Kalman Filter.* Cambridge University Press.

Kozicki, S., and P. A. Tinsley. 1998. "Moving Endpoints and the Internal Consistency of Agents' Ex Ante Forecasts." *Computational Economics* 11 (1–2): 21–40.

———. 2001. "Shifting Endpoints in the Term Structure of Interest Rates." *Journal of Monetary Economics* 47 (3): 613–52.

———. 2012. "Effective Use of Survey Information in Estimating the Evolution of Expected Inflation." *Journal of Money, Credit and Banking* 44 (1): 145–69.

Mavroeidis, S., M. Plagborg-Møller, and J. H. Stock. 2014. "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52 (1): 124–88.

Mehra, Y. P., and C. Herrington. 2008. "On the Sources of Movements in Inflation Expectations: A Few Insights from a VAR Model." *Economic Quarterly* (Federal Reserve Bank of Richmond) 94 (2, Spring): 121–46.

Mishkin, F. S., and M. A. Savastano. 2001. "Monetary Policy Strategies for Latin America." *Journal of Development Economics* 66 (2): 415–44.

Nordhaus, W. D. 1987. "Forecasting Efficiency: Concepts and Applications." *Review of Economics and Statistics* 69 (4): 667–74.

Pesaran, M. H., and M. Weale. 2006. "Survey Expectations." In *Handbook of Economic Forecasting,* Vol. 1, ed. G. Elliot, C. W. J. Granger, and A. Timmermann, 715–76 (chapter 14). Elsevier.

Quispe-Agnoli, M. 2001. "Monetary Policy Alternatives for Latin America." *Economic Review* (Federal Reserve Bank of Atlanta) 86 (3): 43–53.

Timmermann, A. 2007. "An Evaluation of the World Economic Outlook Forecasts." *IMF Staff Papers* 54 (1): 1–33.

Vega, M., and D. Winkelried. 2005. "Inflation Targeting and Inflation Behavior: A Successful Story?" *International Journal of Central Banking* 1 (3): 153–75.

Fixed Prices and Regulatory Discretion as Triggers for Contingent Capital Conversion: An Experimental Examination*

Douglas Davis[a] and Edward Simpson Prescott[b]
[a]Virginia Commonwealth University
[b]Federal Reserve Bank of Cleveland

We report a laboratory experiment that evaluates two price-based mechanisms for triggering the conversion of contingent-capital bonds into equity: a regulator who decides based on observed prices and a mechanistic fixed-price trigger. We find that when conversion decreases incumbent equity value, the regulator mechanism generates fewer conversion errors, particularly in environments where incentives bias a regulator against conversion and where a regulator receives his own signal. In contrast, when conversion increases incumbent equity value, a fixed-price trigger generates fewer conversion errors in these environments as well as when the regulator has the option to delay conversion.

JEL Codes: C92, G14, G28.

1. Introduction

Following the 2008 financial crisis, regulators and banking scholars devoted considerable attention to improving bank capital regulation

*We thank the editor Raphael Repullo, an anonymous referee, and Glenn Harrison for helpful comments. We also thank Aaron Lazar for z-Tree programming and assistance with experiment administration. Experiment instructions, experimental data, and a series of unpublished appendixes are available at http://www.ijcb.org. This paper was presented at the Federal Reserve Bank of Atlanta, the University of Virginia, Virginia Commonwealth University, the 2014 meeting of the Southern Economics Association, and the 2015 SAET conference. The usual disclaimer applies. Financial assistance from the National Science Foundation (SES 1426980) is gratefully acknowledged. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Corresponding author's e-mail: dddavis@vcu.edu; tel.: (804) 828–7140; fax: (804) 828–9103.

in order to increase the stability of the banking system. One regulation that has been proposed is to require systemically important banks to issue a new class of "contingent-capital" (CoCo) bonds, which convert to equity when a bank passes a predefined triggering condition. One advantage of these bonds is that in times of financial distress they reduce a bank's debt-overhang problem by raising equity at a predetermined price just when raising equity in the capital markets is most expensive. A second advantage is that they automatically recapitalize a bank, thus reducing the chance of a costly resolution.[1]

Critical to the effective implementation of CoCos is the trigger used for converting debt into equity. Existing CoCo bonds use an accounting trigger, such as regulatory capital levels, along with some regulatory discretion. Unfortunately, accounting numbers often lag economic values, so CoCo bonds that are based primarily on an accounting trigger may not convert until it is too late.[2] A more promising approach is to use a trigger that incorporates information contained in market prices. Unlike most accounting ratios, market prices incorporate expectations and so are forward looking.[3] One such trigger is a mechanistic price-based rule, with conversion occurring automatically whenever a pre-specified threshold, such as the price of a bank's equity shares, is breached (e.g., Calomiris and Herring 2013; Flannery 2009; McDonald 2013; Pennacchi 2011; Pennacchi, Vermaelen, and Wolff 2013; and Raviv 2004). An advantage of this kind of rule is that it is transparent regarding the conditions for and magnitude of a conversion.

Alternatively, regulators may use price information to make conversion decisions. Regulations encourage this since CoCo issues are

---

[1]See Prescott (2012) for a discussion of these and other possible benefits.

[2]The U.S. experience with accounting triggers is not promising. Since 1992, U.S. bank regulators have been required to follow prompt corrective action (PCA) rules. These rules require regulators to force banks to increase capital and to take other preventative actions when regulatory capital ratios are breached. Furthermore, they require regulators to put the bank into receivership when capital is below 2 percent. Nevertheless, despite failed banks having on average positive book capital at time of failure, average losses to the deposit insurance fund from commercial bank failures were about 25 percent of failed bank assets. See Balla, Prescott, and Walter (2015) for details. Possibly related, accounting-based measures can sometimes be manipulated. See Calomiris and Herring (2012).

[3]For a classic example, see Roll (1984).

required to allow some regulatory discretion to qualify for tier 1 or tier 2 capital under the Basel III rules (Avdjiev et al. 2015).[4] And while a regulator-based triggering rule is subject to the potential of forbearance, it has the advantage of allowing regulators to incorporate the additional information that they gather from their supervision of the bank. Relative to a mechanistic trigger, however, it creates uncertainty about how the regulator will interpret and react to prices.

Unfortunately, recent theoretical and experimental work has identified costs to using either type of price-based mechanism. Theoretical work on rational expectations models by Birchler and Facchinetti (2007), Bond, Goldstein, and Prescott (2010), and Sundaresan and Wang (2015) show that both fixed-trigger and regulator-based mechanisms are subject to equilibrium non-existence and multiple equilibria.[5] Experimental work by Davis, Korenok, and Prescott "DKP" (2014) finds that these mechanisms induce variability in prices, misallocations, and conversion errors. Nevertheless, accounting-based alternatives have severe limitations, so while imperfect, price-based triggering mechanisms are worth evaluating. Furthermore, there may be ways to improve the performance of these mechanisms.

This paper reports an experiment that measures the effects of three changes to the regulatory price-based mechanism studied in DKP. The first is to change the rewards the regulator receives for making conversion decisions. The second is to give the regulator information in addition to market prices. The third is to give the regulator the option of delaying the conversion decision in return for better information in the future. All three changes correspond to realistic features of the regulatory environment. The results of this experiment and the one reported in DKP are then used to evaluate the relative performance of a mechanistic fixed trigger versus a regulator with discretion. We find that the relative desirability of the two classes of mechanisms depends on whether conversion decreases or increases the value of incumbent equity. The regulator usually performs better when the conversion is value decreasing, which is

[4]Regulators do currently use market data in their assessment of banks, though not as a formal part of security conversion. See Feldman and Schmidt (2003).

[5]For recent theoretical work on these issues, see Siemroth (2015).

the case most relevant for the recent policy debate. However, the fixed trigger performs better when the conversion is value increasing, which is relevant for some types of CoCo conversions.

1.1 Literature

Sundaresan and Wang (2015) studied the rational expectations equilibria of a price-based, fixed-trigger mechanism that converted debt to equity when the price of equity dropped below a trigger. They found that a fixed trigger can undermine the informational content of the prices on which it relies. The way it undermines informational content depends on the *conversion rule*, which determines whether and by how much a conversion affects incumbent equity values.

A conversion reduces the value of incumbent equity if the bond-to-equity exchange rate is sufficiently dilutive. However, a conversion rule can be value increasing to incumbent equity owners if the value of retiring debt outweighs the dilutive effect of increasing the number of equity shares. The effects of triggering mechanisms on both value-increasing and value-decreasing conversion rules are of interest. Most academic proposals for the design of CoCos advocate implementing a value-decreasing conversion rule as a means of tempering risk-taking incentives by bank managers. Furthermore, the proposals for debt that can be "bailed in" often require that equity be written down and debt converted to equity during resolution. As a practical matter, however, value-increasing conversions are common. For example, roughly half of recent or planned CoCo issues specify a value-increasing conversion.[6]

In the case of a value-decreasing conversion, Sundaresan and Wang (2015) showed that a fixed trigger creates multiple equilibria for a range of market fundamentals above the trigger price. The intuition driving this result is straightforward. Suppose that traders are risk neutral and have rational expectations, so that the market

[6] Avdjiev et al. (2015) report that 55 percent of CoCo issues between 2009 and March 2015 involve either a partial or complete principal write-down of the CoCo debt. Such bonds represent an extreme case of a value-increasing conversion, because in the event of a conversion, debt is simply retired, with no associated change to existing equity. But more generally, even if conversion dilutes equity—and the trigger is the price of equity—then there are situations in which conversion increases the value of incumbent equity.

price of equity equals the expected payout for bank equity. For specificity, suppose also that an equity price of \$5.00 or less triggers a conversion that reduces the value of equity by \$2.00 per share. Consider now a market fundamental of \$5.50. If traders believe that there will not be a conversion, then the price is \$5.50 and no conversion will occur, which is consistent with traders' beliefs. However, if traders believe that conversion will occur, they will all incorporate the value of the conversion into their price and trade equity at \$3.50 per share. This price is below the trigger, so conversion will occur, which is also consistent with traders' beliefs. Following this reasoning, multiple equilibria exist for fundamental realizations between \$5.00 and \$7.00.

In the case of a value-increasing conversion, Sundaresan and Wang (2015) showed that under a fixed-trigger rule there is instead a problem of equilibrium non-existence. The intuition is again straightforward. Using the parameters of the above example, suppose now that in the case of a conversion, the value of equity increases by \$2.00 per share. Consider in this case how traders with rational expectations would treat a fundamental value of, say, \$3.50. If they believe conversion will occur, then a share is worth \$5.50, but that is above the trigger, so conversion will not occur. Conversely, if they believe that conversion will not occur, then a share is worth \$3.50, but that is below the trigger, so conversion will occur. For such a realization (and for any fundamental realization between \$3.00 and \$5.00) no equilibrium price exists.

Birchler and Facchinetti (2007) and Bond, Goldstein, and Prescott (2010) studied the rational expectations equilibria of a related model in which traders had information that a regulator did not (such as knowing the market fundamental) and where the regulator used the market price to decide whether to intervene in the bank, along the lines of a CoCo bond conversion. In their models, they did not find multiple equilibria for value-decreasing conversions, but they did find non-existence of equilibria for value-increasing conversions. Furthermore, they found it for a broader range of fundamentals than under fixed-trigger regimes.[7]

[7]The intuition for the value-decreasing case is straightforward. Unlike a fixed trigger, the regulator has the option to convert at any price, so were the price between \$3.00 and \$5.00, the regulator would know not to convert because the

These theoretical results raise concerns about how this kind of triggering mechanism will work in practice. Furthermore, data from existing markets provide little guidance. Although issues of CoCo-type bonds have increased substantially in the last several years, no instance of a triggering condition being breached has yet occurred.[8] Furthermore, all the existing issuances use accounting ratios (sometimes supplemented by regulatory discretion) rather than market prices as the trigger. Given the paucity of empirical data, experimental methods are a particularly useful source of evidence. DKP report an experiment that evaluates the relevance of the predicted imprecisions with price-based triggering mechanisms. They find that in both fixed-trigger and regulator regimes the theoretically predicted problems of multiple equilibria and equilibrium non-existence manifest themselves as variability in realized prices, prices deviating from realized values, resource misallocations, and conversion errors. Further, and contrary to theoretical predictions, in the case of a value-decreasing conversion, errors occurred in the regulator regime as well as in the fixed-trigger regime. As a consequence, for some ranges of fundamentals, frequent conversion errors occur in markets using either rule.

1.2 Overview

An important question not decisively addressed by DKP regards the relative performance of the two triggering mechanisms.[9]

fundamental must exceed \$5.00. Traders recognize this and incorporate it into their beliefs. For a value-increasing conversion, the non-existence problem arises because the price as a function of the fundamentals has to be non-monotonic, but then for some values of the fundamental the regulator cannot tell from the price whether the fundamental is below or above the \$5.00 threshold. See Prescott (2012) for a more detailed explanation of the source of these problems as well as those in the fixed-trigger regime.

[8] Avdjiev et al. (2015) report \$280 billion in CoCo issuances between January 2009 and March 2015. The vast majority of issues were from European banks, but substantial issues were also made by banks in Australia, Brazil, and Russia. Issues differ widely with respect to conversion ratios and triggering conditions, and while no triggering condition has ever been breached, there is considerable concern in the investment community about the lack of standardization in conditions. See, e.g., Hayden (2014).

[9] Observing that the fixed-trigger mechanism was more readily understood by traders and also eliminated uncertainty regarding the regulator's actions, DKP offers some guarded support for the fixed-trigger mechanism. However,

Furthermore, a variety of features relevant to the environment in which a regulator acts might affect a regulator's performance. In this paper, we study three such variations from the baseline regulatory experiment. The first variation changes the regulator's incentive to act, so that there is a bias away from conversion. This bias may affect not only the regulator's behavior but also what traders expect the regulator will do, and thus the informational content of prices.

The second variation from the baseline environment supplements the regulator's price observations with non-market information. Bank regulators have special legal powers to conduct bank examinations, and via this process they observe information beyond what is reported in financial statements. Several studies report that such examinations do give bank regulators access to some information before the markets.[10] While current and accurate information should clearly help a regulator make decisions, timely non-market information is not always available. We model this by showing the regulator the market fundamental with some probability.

Finally, the third variation we study allows the regulator to wait for non-market information before acting. The time between bank examinations suggests this treatment.[11] Immediately following an examination, regulators may have better information about

---

as reviewed below, support for the fixed-trigger mechanism in the DKP experiment is far from unqualified, and in an important respect the regulator regime generated superior results.

[10]The survey in Flannery (1998) discusses the literature on market prices and bank condition. He reports numerous studies that find that market prices contain information that predicts changes in bank condition as measured by bank regulators. However, he also reports on some studies, such as Berger and Davies (1998), that find that regulators also have some information that the market does not have. See also Berger, Davies, and Flannery (2000) and DeYoung et al. (2001). Furthermore, Bond, Goldstein, and Prescott (2010) show that if the regulator is given his own signal drawn from a uniform distribution, the range of fundamentals for which an equilibrium does not exist decreases as the quality of the signal improves.

[11]For all but the largest banks, bank examinations happen at regular intervals, usually annually, and for that reason, regulator information at any particular point in time may be outdated. Furthermore, as Flannery (1998, p. 293) observes, bank examiners are interested in different information than markets; examiners are most concerned with information about default risk, while equity owners are most concerned with information about the value of a bank in non-default states. For this reason, the market may have information that examiners do not have.

a bank's financial condition than does the market. However, as the time following an examination increases, the quality of non-market information deteriorates, making market-based information a valuable source of information until the next examination. The possibility of delaying action until the next bank examination may affect a regulator's willingness to rely on price-based market signals.

We study these questions by reporting an experiment that builds upon DKP in order to evaluate the relative accuracy of fixed-trigger and regulator-based conversion rules. In addition to more systematically analyzing conversion error rates in the thirty-four market sessions previously reported in DKP, we report eighteen new market sessions conducted in variants of a regulator regime that incorporate inaction bias, non-market information, and the opportunity to delay action.

We find that the conversion rule affects the relative accuracy of fixed-price and regulator-based triggering mechanisms. In the case of a value-increasing conversion, results support economists' predisposition toward the use of a fixed trigger: Previously reported fixed-trigger mechanisms are marginally superior to regulator mechanisms. In the case of a value-decreasing conversion, the reverse is true. Furthermore, we find that both inaction bias and non-market information strengthen this result. In the case of a value-increasing conversion the fixed-trigger rule generates fewer conversion errors than a regulator-based regime, while in the case of a value-decreasing conversion the regulator-based regime is superior.

In the delay treatment, however, the fixed-trigger mechanism is generally preferable. In the case of a value-increasing conversion, a fixed-trigger rule is vastly superior to a regulator-based rule, largely because in this case some regulators appear to ignore altogether the informational content of prices. In the case of a value-decreasing conversion, the fixed-trigger and regulator mechanisms are statistically indistinguishable.

On net we conclude that in the case of a value-increasing conversion, a fixed-trigger rule is strongly preferable. In the perhaps more policy-relevant case of a value-decreasing conversion, however, the regulatory mechanism is usually better, though there are some environments where that is not the case.

2. Experiment Design and Procedures

All the market sessions reported in this paper are variants of the common market structure described in subsection 2.1. Subsection 2.2 reviews the relevant results of the DKP experiment. Afterwards, subsections 2.3, 2.4, and 2.5, respectively, describe the regulator bias, non-market information, and delay treatments that are new to this paper.

2.1 Market Design and General Procedures

In each session, there is a cohort of ten traders. In sessions with regulatory treatments, there are also three monitors, who correspond to regulators.[12] Each session consists of a sequence of twenty to twenty-five trading periods. In each period, each trader is endowed with two perishable asset units and a working capital loan of \$16.00 that can be used to start buying assets.[13] Asset units expire at the end of each period and pay out an amount that is determined by three factors: (i) the market fundamental, (ii) a trader's idiosyncratic valuation of the asset, and (iii) a conversion, if it occurs. For sessions with a regulatory treatment, monitors are motivated to make a decision to "convert" (intervene) if the underlying market fundamental is below \$5.00. If the selected monitor chooses to convert, all assets either increase in value by \$2.00 (in the case of a value-increasing conversion) or fall in value by \$2.00 (in the case of a value-decreasing conversion).

To induce trade, we give the traders different values from holding the asset at the end of the period. Six traders (who are endowed with twelve units in total) value the asset at θ, which is taken from the interval [\$2.00,\$8.00]. We refer to θ as the market fundamental. The remaining four traders (who are endowed with eight units in total) value the asset at $\theta - \$0.60$. A trader knows his own valuation of the asset, but does not know whether it is the high or low valuation.

[12] In the experiments, we referred to the people who could make the conversion decision as monitors.

[13] Unless otherwise noted, dollar amounts refer to lab dollars. Subjects' performance was kept track of in terms of lab dollars and then at the end of a session, lab dollars were converted to U.S. dollars to pay the subjects. The conversion ratio is reported at the end of section 2.1.

A trader also knows the aggregate number of high- and low-value units each period, and thus knows that there are more high-value traders than low-value traders. Once shown their asset value for the period, traders are given 110 seconds to buy and sell assets.

Trading follows standard open-book double-auction rules, with traders seeing all bids, offers, and the history of contract prices in the period. Period earnings for each trader are the sum of revenues from asset sales net of asset purchases and the value of assets held at the period's end, or

$$Payoff = \sum_{i=1}^{m_s} p_i - \sum_{j=1}^{n_b} p_j + (\nu - c) \times (n_b - m_s + 2), \tag{1}$$

where m_s units are sold at prices $p_i, i = \{1, \ldots, m_s\}$, and n_b units are bought at prices $p_j, j = \{1, \ldots, n_b\}$. On the right side of the equation, ν is the trader's valuation of holding an asset, and $c \in \{-\$2.00, \$0.00, \$2.00\}$ reflects the effect on the asset value of the conversion decision. Finally, the working capital that traders start with is paid back at the end of the period, so it nets out in (1).

Conversion is determined by the monitors' decisions following the close of trade. Monitors do *not* observe the market fundamental, but they do see the median price of contracts. Based solely on the observed price, monitors guess the underlying market fundamental and make a conversion decision. Monitors earn \$3.00 if their guess is within \$0.20 of the actual θ for the period, and \$1.00 if their guess is between \$0.21 and \$0.50 of the period's θ. For conversion decisions, monitors earn \$12.00 from a correct decision, where a "correct" decision is either to convert if the underlying fundamental is less than \$5.00 or to not convert if the fundamental is \$5.00 or more. To prevent insurance strategies, monitors are obligated to make a conversion decision consistent with their guess of θ.

Once all monitor decisions are complete, the market fundamental is displayed and earnings are determined. For a monitor, earnings are \$0.00, \$1.00, \$3.00, \$12.00, \$13.00, or \$15.00, depending on the accuracy of his guess about the fundamental and whether or not he made the correct conversion decision. For traders, one of the three monitor decisions is selected at random and implemented.[14] If the selected monitor chose conversion, then the asset payoff either increases by \$2.00 per unit (in the case of a value-increasing conversion) or decreases by \$2.00 per unit (in the case of a value-decreasing conversion).

The fixed-trigger treatment is identical to the above process except that the three monitors are eliminated, and in their place a conversion mechanically occurs whenever the period's median price is less than \$5.00.

All sessions followed the same general procedures. At the start of each session a cohort of student volunteers (ten or thirteen) was randomly seated at visually isolated personal computers. An experiment administrator then read aloud a common set of instructions, which explained trader and monitor incentives as well as how to make decisions on the computer interface used in the experiment.[15] Participants followed along on printed copies of their own, while the administrator read instructions from a version projected onto a wall at the front of the lab. Following the instructions, participants completed a short quiz of understanding, which the experiment administrator reviewed publicly. At any time during the instructions and quiz, participants were encouraged to ask questions by raising their hands. Questions were answered privately.[16] Following the quiz, the sessions commenced. The experiment was programmed and conducted with the software z-Tree (Fischbacher 2007).

To facilitate understanding, instructions were presented in parts. Initial instructions were provided for a simple condition in which traders buy and sell asset units in the absence of any possible conversion.[17] In this condition, monitors are told the period's median contract price following the close of trade and are asked to guess the underlying market fundamental. After several periods in

[14]We used three monitors in each session in order to generate more data points on conversion decisions per session.

[15]Instructions are available in unpublished appendix 1, available at http://www.ijcb.org.

[16]Quiz responses were not formally recorded. Participants, however, had few questions and seemed comfortable with the instructions.

[17]In DKP, these periods are referred to as the *BASE* condition. In the *BASE* condition, prices were close to the market fundamental θ, and allocations were very close to fully efficient.

this condition, the session was paused and additional instructions describing a conversion treatment were distributed. An experiment administrator then read aloud these instructions and administered another short quiz of understanding, after which a second segment commenced for ten to fifteen periods. In the second segment, some sessions had value-decreasing conversions and others had value-increasing conversions. In most sessions, this protocol was repeated in a third segment, which was identical to the second except that the conversion rule was switched to value decreasing (value increasing).[18] Following the conclusion of the final segment, the experiment ended and participants were privately paid and dismissed from the lab.

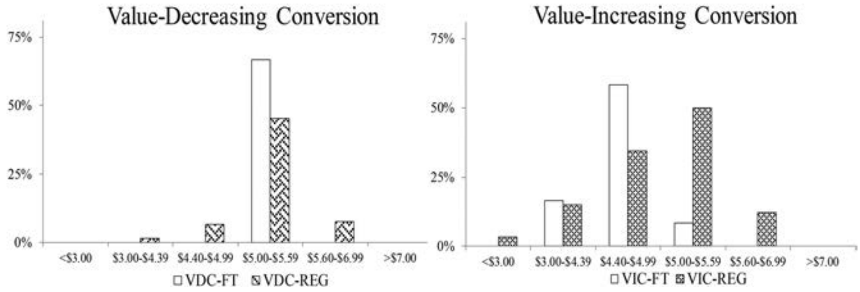
Participants were upper-level math, science, engineering, and business students enrolled in courses at Virginia Commonwealth University in the spring semesters of 2010 and 2011 (for the previously reported DKP sessions) and the spring semester of 2014 and the fall semester of 2015 (for the new sessions). No one participated in more than one session. Lab earnings were converted to U.S. currency at \$12.00 lab = \$1.00 U.S. rate. In U.S. dollars, participant earnings for the 90–120 minute sessions ranged from \$14.00 to \$32.25 and averaged \$23.25 (inclusive of a \$6.00 appearance fee).

2.2 DKP Results

DKP reports a series of thirty-four market sessions involving 424 participants. Sessions included both regulator and fixed-trigger regimes, under both value-increasing and value-decreasing conversion conditions. Figure 1 summarizes the conversion error results. In each panel of the figure, notice that market fundamentals are clustered into six ranges: the $<$\$3.00 and $\geq$\$7.00 ranges, where neither triggering mechanism should create price ambiguity, as well as partitions of the \$3.00–\$4.99 and \$5.00–\$6.99 ranges, where, depending on the triggering mechanism and conversion rule, either multiple equilibria or equilibrium non-existence arise in the rational expectations benchmarks. The partitions separate out the ranges \$0.60 above and below the \$5.00 efficient conversion limit. Although the

[18] As explained in DKP, in some treatments value-increasing and value-decreasing regimes were not varied within sessions.

Figure 1. Conversion Error Rates for Regulator (*REG*) and Fixed-Trigger (*FT*) Treatments



Notes: The horizontal axes list the market fundamentals broken up into different ranges. The vertical axes list the fraction of the time a conversion error is made for fundamentals in each range. A decision to convert when the fundamental is ≥$5.00 is counted as an error, as is a decision not to convert when the fundamental is <$5.00.

proximity of a market fundamental realization to $5.00 is not pertinent in the Bond, Goldstein, and Prescott (2010) or Sundaresan and Wang (2015) models (both of which use a rational expectations framework), it is relevant to the market used in DKP and this paper because they include value heterogeneity in order to generate trade. The $0.60 range around the conversion cutoff is used because for market fundamental realizations in this range, some traders must infer from the trading process whether the market fundamental is above or below the $5.00 trigger.[19]

Looking first at the case of value-decreasing conversions, observe that errors occur frequently in the fixed-trigger (*FT*) regime for fundamental realizations just above the $5.00 cutoff, as predicted by the Sundaresan and Wang (2015) analysis. Not consistent with theoretical predictions are the conversion errors under a regulator (*REG*) regime. Here again, when market fundamentals were slightly

[19]For example, if the market fundamental is $5.20, then six traders (with twelve asset units) see a fundamental above $5.00, while four other traders (with eight asset units) see a fundamental of $4.60. The traders with the low value do not know for sure whether the market fundamental is above $5.00 and hence no conversion should occur. A similar situation arises for a fundamental of $4.80, except that in this case the high-value traders do not know for sure whether the fundamental is above or below $5.00.

above \$5.00, traders often partially incorporated the value of a conversion into prices and generated median prices in the \$3.00–\$4.99 range. These downward price adjustments yield prices that would not be observed in an equilibrium for the *REG* regime—traders should either fully incorporate the value of a conversion, generating market prices of \$2.99 or less, or realize that the market fundamental exceeds \$5.00 and no conversion will occur, generating prices of \$5.00 or more. However, upon seeing these prices, the monitors often errantly concluded that the market fundamental was below \$5.00 and decided to convert.

In the case of value-increasing conversions, conversion errors tend to cluster around fundamental realizations close to the \$5.00 cutoff under both triggering mechanisms. However, in the *REG* treatment, errors are distributed both above and below the \$5.00 cutoff, while in the *FT* treatment, errors cluster in the \$3.00 to \$4.99 range. These results are consistent with theoretical predictions: In the *REG* regime market, fundamentals of, say, \$4.60 and \$5.40 may generate very similar transaction prices, both somewhere in excess of \$5.00, leaving monitors confused as to the desirability of conversion and roughly as likely to err by converting as by not converting. On the other hand, in the *FT* regime, conversion errors occur predominantly when the market fundamental is slightly below \$5.00 and traders are unable to keep the median transaction price below the automatic conversion cutoff.

Combined, these experimental results indicate that the problems of discerning fundamental information from market prices suggested in the theoretical work may frequently lead to conversion errors. Inspecting figure 1 further, however, notice that the relative desirability of the two triggering mechanisms is not obvious.

In the case of a value-decreasing conversion, overall incidence of conversion errors is slightly higher in the *FT* regime. Perhaps more problematically, the incidence of conversion errors for fundamentals above \$5.00 was also higher for the *FT* regime, meaning they were socially undesirable conversions (e.g., type II errors of commission). McDonald (2013) argues that type II errors are likely more problematic than type I errors, so under his criterion these results would make the *REG* mechanism more appealing. On the other hand, errors occur over a considerably wider range of fundamental realizations in the *REG* mechanism.

In the case of a value-increasing conversion, the overall incidence of errors is roughly the same under either mechanism. Nevertheless, the type of error differs. Under the *REG* mechanism, the bulk of errors were type II errors of commission, occuring for market fundamentals above \$5.00. In contrast, under the *FT* mechanism, errors are concentrated in the ranges of market fundamentals below \$5.00, meaning that socially desirable conversions failed to occur (type I, errors of omission). Under these criteria, results would make the *FT* mechanism more appealing.

As discussed in the introduction, a more complete evaluation of the relative desirablity of fixed-trigger and regulator-based triggering mechanisms requires evaluation of regulator behavior in environments enrichened by features that may importantly affect the frequency and accuracy of regulator conversion decisions. The following three subsections describe procedures for treatments that explore the effects of three such environmental alterations on market outcomes: inaction bias, probabilistically supplied non-market information, and the opportunity to delay action.

2.3 Inaction Bias

The distortionary effects of external pressures on the timing and magnitude of conversion decisions is potentially an important factor in assessing the value of allowing a regulator to make a conversion decision. There are two distinct reasons to consider introducing bias as a treatment. First, in both the Great Recession and the Savings and Loan Crisis of the 1980s, regulators were criticized for not intervening with sufficient speed. In the Savings and Loan Crisis, in particular, it is well documented that many S&Ls were insolvent but were allowed to continue to operate in the hope they would recover (White 1991). Introducing a bias toward not acting can be viewed as a more realistic description of regulatory incentives in such contexts. Second, in the previously reported sessions, where the regulator had symmetric preferences over taking actions, undesirable conversions were the most frequent error. This suggests that adjusting the regulator's incentives toward inaction may be a way to reduce these errors.

To assess the effect of a bias against making a conversion decision, we conduct a *REGB* treatment, which replicates the *REG* treatment in DKP, except that we alter monitors' incentives by imposing penalties for an incorrect decision to convert. As in the *REG* treatment, monitors were told the median contract price for the periods and then they both guessed the underlying market fundamental and made a conversion decision. As before, they earned a small amount (up to \$3.00) for guessing with sufficient accuracy the market fundamental, and then a larger amount (\$12.00) for making a correct conversion decision (e.g., converting when socially desirable). However, in *REGB* they were also punished if they converted when they should not have. In particular, relative to the *REG* treatment, payment falls from \$0.00 to –\$12.00 if they convert when the fundamental is above \$5.00. As in the *REG* treatment, monitors earn nothing from an error of omission, or a decision to not convert when the fundamental is below \$5.00. Also as before, after all monitors make decisions, one of their choices is selected at random and implemented in the market (e.g., increasing or decreasing the market fundamental by \$2.00 in the case conversion is selected). We evaluate both value-increasing and value-decreasing conversions.

In the literature, the commitment problem that encourages inaction is widely regarded as a reason to prefer a fixed-trigger mechanism to a regulator-based mechanism. The analysis in DKP focuses on a different trade-off, namely, how allowing a regulator to react to prices changes the price's informational content. In this respect a bias toward inaction could actually improve performance. As can be seen in figure 1, for the *REG* regime, conversion errors concentrate in the \$5.00–\$5.59 range of fundamentals, as traders incorporate the value of a socially undesirable conversion when the market fundamental is slightly above the \$5.00 cutoff. A bias toward inaction should make monitors more reluctant to make a decision to convert for the subsequent prices in the \$3.00–\$4.99 range, and in this way reduce the incidence of such type II errors.

In the case of a value-increasing conversion, the effect is unclear. Inaction bias may weaken the willingness of regulators to convert when facing trading prices in the uninformative \$5.00–\$6.99 range. Incentives against action thus may increase the incidence of forgone desirable conversions that generate such uninformative prices (e.g., for market fundamentals in the \$3.00–\$4.99 range), but at the same time they should reduce the incidence of socially undesirable conversions that generate roughly the same prices (e.g., for market fundamentals between \$5.00 and \$6.99). Referring back to figure 1, observe that in the value-increasing *REG* sessions reported in DKP, the bulk of conversion errors occurred for market fundamentals between \$5.00 and \$5.59. Inaction bias would be beneficial on net if it reduces the incidence of these socially undesirable and arguably more costly errors.

2.4 Non-market Information

Our second treatment assesses the effects of non-market information. One argument for not adopting a fixed-price trigger is that bank regulators have information about bank quality that is not available to market participants and that information should be used in making decisions. Indeed, Bond, Goldstein, and Prescott (2010) established in their rational expectations model that if the quality of a regulator's signal was sufficiently high, a unique equilibrium exists for a value-increasing conversion, so regulators could learn from market prices. Nevertheless, the quality of a regulator's signal is not always accurate, and the market response to a perception that regulators know the market fundamental may dampen the information transmitted by prices. For this reason, it is worth studying the case in which a regulator has a signal that is not always of sufficient quality to eliminate the equilibrium existence problems found in the theoretical literature.

In this second (*REGI*) treatment, we evaluate the effects of probabilistically providing regulators with accurate non-market information. Procedurally, the *REGI* treatment is identical to the *REG* treatment in DKP except that in addition to being shown the median transactions price after the close of trade each period, there is a 50 percent probability that the monitors are also shown the market fundamental. Traders know the probability that monitors are informed, but they do not know whether the monitor is informed in any particular period. We ran the *REGI* treatment under both value-increasing and value-decreasing conversion rules.

2.5 Action Delay

Our third treatment (*REGD*) evaluates the capacity of non-market information to endogenously induce an inaction bias.[20] Specifically,

[20]We are grateful to a referee for suggesting this treatment.

the availability of accurate non-market information via regular but infrequent bank examinations may, in the intervals between examinations, induce regulators to discount or ignore the information conveyed in market prices and delay making conversion decisions.

Procedures for the *REGD* treatment parallel those for the *REG* treatment with the following differences. Following the close of trade, and after being shown the period's median contract price, the monitor makes a decision to either convert or wait. If the monitor decides to convert, he earns \$12.00 from a correct decision (e.g., if the market fundamental is less than \$5.00) and \$0.00 from an incorrect decision. If the monitor decides to wait, he is shown the market fundamental and is then obligated to make the correct conversion decision. In either case, the monitor earns \$6.00 from a decision to wait.[21] We conducted sessions under both value-increasing and value-decreasing conversion rules.

2.6 Summary

In total, the new experiment consists of a series of eighteen market sessions, with six sessions conducted in each of *REGB, REGI*, and *REGD* treatments. Table 1 summarizes the experiment design. In total, 234 undergraduate student volunteers participated in the new experiment. With only minor differences, the previously reported markets were conducted under the same conditions as the sessions reported here.

3. Experiment Results

An assessment of the relative merits of the triggering mechanisms in terms of conversion error rates involves comparisons across multiple dimensions. To quantify relative performance differences, we use four criteria. First, and perhaps most obviously, we compare the overall

[21]As in the *REG* treatment, monitors also guessed the market fundamental each period and were compensated for the accuracy of their decisions. One additional minor difference from the *REG* treatment is that we relaxed the requirement for consistency between the fundamental guess and the conversion decision. In the *REGD* treatment, a monitor could choose to wait even if he guessed that the fundamental was below \$5.00. As in the *REG* treatment, he could not convert if he guessed that the fundamental exceeded \$5.00.

Table 1. Experiment Session Structure

Session	No. of Sessions	Session Structure, by Periods		
		1–2	3–12	13–25
REGB-VIC/REGB-VDC	3	*BASE*	*REGB-VIC*	*REGB-VDC*
REGB-VDC/REGB-VIC	3	*BASE*	*REGB-VDC*	*REGB-VIC*
REGI-VIC/REGI-VDC	3	*BASE*	*REGI-VIC*	*REGI-VDC*
REGI-VDC/REGI-VIC	3	*BASE*	*REGI-VDC*	*REGI-VIC*
REGD-VIC/REGD-VDC	3	*BASE*	*REGD-VIC*	*REGD-VDC*
REGD-VDC/REGD-VIC	3	*BASE*	*REGD-VDC*	*REGD-VIC*

Notes: *VDC* refers to value-decreasing conversion. *VIC* refers to value-increasing conversion. *BASE* refers to the treatment in which no conversion decision was allowed. *REGB* refers to the regulatory-inaction-bias treatment. *REGI* refers to the regulatory-information treatment. *REGD* refers to the regulatory-delay treatment.

incidence of conversion errors across triggering mechanisms. Second, following McDonald's suggestion that socially undesirable conversions (i.e., errors of commission) are generally more problematic than failures to convert, we evaluate the incidence of conversion errors for market fundamentals of \$5.00 and above. Third, we consider the extent to which a triggering mechanism facilitates the discovery of the market fundamental via the trading process by evaluating the incidence of conversion errors in the ranges of \$0.60 below and \$0.60 above the \$5.00 efficient-conversion cutoff. Fourth, we consider the extent to which a triggering rule results in "gross" errors, or errors for fundamentals that deviate by more than \$0.60 from \$5.00.

A series of simple bivariate linear probability regressions allows evaluation of performance in terms of these four criteria.[22] Labeling the six fundamental ranges <\$3.00, \$3.00–\$4.39, \$4.40–\$4.99, \$5.00–\$5.59, \$5.60–\$6.99, and >\$7.00 alphabetically as a, b, c, d, e, and f, respectively, we estimate the incremental effect of the *REG* treatment relative to the *FT* benchmark. A first regression estimates the incremental effect of the regulator treatment relative to the fixed-trigger rule for all fundamental ranges a though f combined. A

[22] We report linear probability estimates for expositional ease. Comparable probit regression estimates, reported in unpublished appendix 2, yield substantially identical results.

second regression assesses the incremental effect of a regulator treatment on the incidence of errors of omission by restricting the set of observations to those where the market fundamental exceeds \$5.00, or is in the d, e, or f ranges. A third regression assesses the incremental effect of the regulator-based triggering mechanism when the market fundamental is close to the \$5.00 efficient-conversion cutoff by restricting the set of observations to market fundamentals in the c and d ranges. The fourth set of regressions estimates the incremental effect of the regulator-based triggering mechanism on the incidence of gross errors where market fundamentals are in the ranges of below \$4.39 or above \$5.60, namely, the a, b, e, and f ranges.

Formally, for the set of observations in each value range $j \in \{abcdef,\ def,\ cd,\ abef\}$ we estimate

$$ce_{it} = \beta_0 + \beta_{REGj} D_{REGit} + \mu_i + \varepsilon_{it}, \qquad (2)$$

where the dependent variable ce_{it} takes on a value of 1 if monitor i committed a conversion error in period t and is 0 otherwise, $i \in \{1, 2, \ldots, \#\ monitors\}$, and $t \in \{1, 2, \ldots, \#\ periods\}$. D_{REGit} is an indicator variable that takes on a value of 1 if the regulator treatment is in effect and 0 otherwise. All regressions use the monitor as the unit of observation and model repeated measures on monitors as random effects.[23] We further cluster data by session and use a robust (White "sandwich") estimator to control for possible unspecified autocorrelation or heteroskedasticity.[24]

3.1 DKP Regulator

Table 2 summarizes the performance of the *REG* sessions relative to those using a fixed-trigger mechanism. Looking at estimates for the

[23] In the fixed-trigger sessions, the unit of observation is a market.

[24] In unpublished appendix 3 we follow this same general method to evaluate trading efficiency, which we calculate as the percentage of available gains extracted via the trading process. As is often the case in experiments involving the trade of exogenously valued assets, trader earnings may be quite high even absent any active exchange, because traders may realize a sizable portion of the available earnings by simply holding on to their asset endowment (something participants in fact tend not to do). Our trading efficiency measure more narrowly assesses the extent to which assets flow from low- to high-value holders of assets. Trading efficiency results echo in all respects the conversion-error-frequency results reported in the text.

Table 2. Linear Regressions that Describe the Incremental Effect of the Baseline Regulatory (*REG*) Treatment Relative to the Fixed-Trigger (*FT*) Treatment

	Value-Decreasing Conversion				Value-Increasing Conversion			
β_0	13.3*** (2.0)	26.7*** (4.0)	33.3*** (5.0)	0.0 —	16.7*** (4.0)	3.3 (3.1)	33.3*** (9.9)	5.6* (3.3)
β_{REG}	−5.6** (2.4)				2.0 (5.1)			
$\beta_{REG\text{-}def}$		−12.9*** (4.8)				14.3*** (4.5)		
$\beta_{REG\text{-}cd}$			−14.4** (5.8)				7.1 (11.9)	
$\beta_{REG\text{-}abef}$				3.4* (1.9)				3.0 (3.8)
Wald χ^2	5.15**	7.2***	6.17**	3.35*	0.15	10.36***	0.36	0.59
N	510	234	158	354	510	234	156	354

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REG} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REG\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REG\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REG\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

value-decreasing conversion, shown in the left panel, observe that the regulator rule outperforms the fixed trigger on most criteria. The *REG* markets reduce the overall incidence of conversion errors, $\beta_{REG} = -5.6\%$ ($p < 0.05$), and the incidence of type II errors, $\beta_{REG\text{-}def} = -12.9\%$ ($p < 0.01$), and help with the discovery of the market fundamental, $\beta_{REG\text{-}cd} = -14.4\%$ ($p < 0.05$). The regulator-based trigger does suffer relative to the fixed-trigger rule in the sense that it generates more gross errors, $\beta_{FT\text{-}abef} = 3.4\%$ ($p < 0.10$).

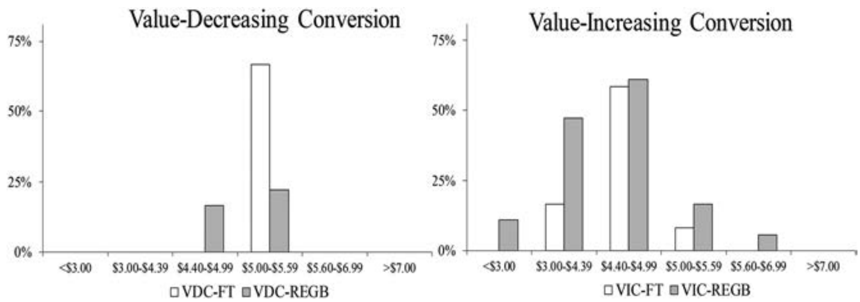
On the other hand, in the case of a value-increasing conversion, the comparison is closer. The small and statistically insignificant coefficient estimate on β_{REG} suggests no overall difference between the fixed-trigger and regulator regimes. The *REG* and *FT* triggering mechanisms are further indistinguishable for realizations close to \$5.00 and for realizations far from \$5.00, as indicated by the statistically insignificant coefficient estimates on $\beta_{REG\text{-}cd}$ and $\beta_{FT\text{-}abef}$, respectively. The baseline *REG* markets, however, do raise the incidence of type II errors over the *FT* markets by 14.3 percent, as indicated by the coefficient estimate on $\beta_{REG\text{-}def}$ ($p < 0.01$).

Collectively, these baseline results suggest a sensitivity of the triggering mechanism to the conversion rule. In the case of a value-increasing conversion, a fixed-trigger mechanism is superior particularly in the sense that it results in significantly fewer type II errors (i.e., socially unnecessary conversions). In the case of a value-decreasing conversion, the regulator regime generates fewer errors in most respects.

3.2 Inaction Bias

Figure 2 plots conversion error rates for the *REGB* treatment relative to the counterpart *FT* treatment. In the graph, we cluster observations into the same six ranges of fundamental realizations reported in figure 1. Notice in the figure that in the case of a value-decreasing conversion, the regulator-based rule is clearly superior. As seen in the left panel of figure 2, the *VDC-REGB* treatment cuts the incidence of undesirable conversions in the \$5.00–\$5.59 range of fundamental realizations by roughly two-thirds relative to the *VDC-FT* treatment. This reduction in conversion error rates for market fundamentals just above \$5.00 comes at the cost of some increase in the incidence of forgone socially desirable conversions for market

Figure 2. Conversion Error Rates for Regulatory-Inaction-Bias (*REGB*) and Fixed-Trigger (*FT*) Treatments



Notes: The horizontal axes list the market fundamentals broken up into different ranges. The vertical axes list the fraction of the time a conversion error is made for fundamentals in each range. A decision to convert when the fundamental is ≥$5.00 is counted as an error, as is a decision not to convert when the fundamental is <$5.00. In contrast, in the case of a value-decreasing conversion, inaction bias improves the relative performance of a regulator-based triggering rule.

fundamentals in the $4.40–$4.99 range. Nevertheless, both the overall incidence of errors and the incidence of errors of commission are considerably lower in the *VDC-REGB* treatment than in the *VDC-FT* treatment.

On the other hand, in the case of value-increasing conversions, inaction bias shifts the incidence of conversion errors from the roughly symmetric dispersion around the $5.00 cutoff generated in the *VIC-REG* treatment to a dispersion more like that in the *VIC-FT* treatment, with the bulk of errors occurring for market fundamental realizations below $5.00 (compare figures 1 and 2).[25] However, the reduction in type II errors of commission in the *VIC-REGB* treatment does not compensate for the increased incidence of conversion errors for fundamentals below $5.00, and relative to the *VIC-FT* treatment, the *VIC-REGB* condition is unreservedly less accurate: For every range of fundamental values over which conversion errors occurred, the incidence of errors is higher in the

[25] Unpublished appendix 4 directly compares the *REGB, REGI*, and *REGD* treatments with the *REG* treatment.

VIC-REGB condition than in the *VIC-FT* treatment. Thus, in the case of a value-increasing conversion, a fixed-trigger rule yields more desirable results than a monitor who faces pressures to not make a conversion.

Table 3, formatted as table 2, allows a quantitative assessment of the effects of inaction bias by estimating the incremental effects of the *REGB* treatment relative to the *FT* regime. As can be seen in the table, comparison of the *REGB* and *FT* treatments is more dependent on the conversion type than was true for the baseline comparisons. In the case of a value-decreasing conversion, inaction bias makes the regulator-based rule a more clearly superior alternative. Incentives inducing inaction eliminate gross errors in a regulator regime (i.e., errors outside the c and d ranges), making the *REGB-VDC* and *FT-VDC* treatments identical in this respect. In the other three dimensions, the regulator regime results in fewer errors overall, $\beta_{REGB} = -5.6\%$ ($p < 0.10$), fewer socially undesirable conversions, $\beta_{REGB\text{-}def} = -17.8$ ($p < 0.01$), and fewer errors when the market fundamental is close to the \$5.00 cutoff, $\beta_{REGB\text{-}cd} = -13.9$ ($p < 0.10$).

On the other hand, in the case of a value-increasing conversion, the regulator-based regime raises the overall incidence of conversion errors, $\beta_{REGB} = 10.6\%$ ($p < 0.10$), and the incidence of gross errors, $\beta_{REGB\text{-}abef} = 13.9$ ($p < 0.05$).

3.3 Regulatory Information

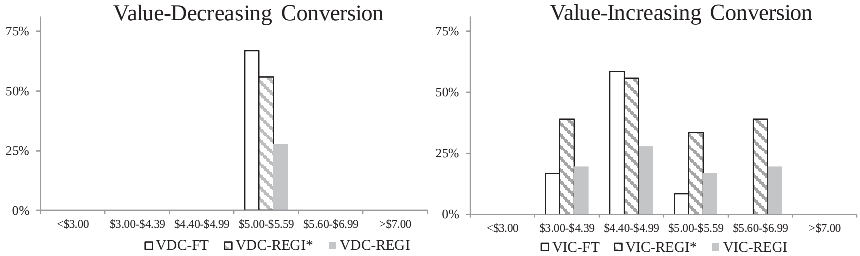
In the *REGI* treatment, our primary interest regards the indirect market response to monitors probabilistically knowing the underlying fundamental, because the direct effects of such information must necessarily reduce the incidence of conversion errors. We evaluate these indirect effects by examining the incidence of conversion errors in the periods in which the market fundamental was *not* revealed to monitors. The striped bars in figure 3 illustrate the incidence of conversion errors in these *REGI** periods relative to the *FT* regime. Looking first at the case of a value-decreasing conversion, errors are confined to the \$5.00–\$5.59 range of fundamentals in the *REGI** periods and roughly parallel the incidence of conversion errors in the *FT* treatment. The failure of probabilistically providing information to monitors to clarify traders' pricing incentives was a surprise to us,

Table 3. Linear Regressions that Describe the Incremental Effect of the Regulatory-Inaction-Bias (*REGB*) Treatment Relative to the Fixed-Trigger (*FT*) Treatment

	Value-Decreasing Conversion				Value-Increasing Conversion			
β_0	13.3***	26.7***	33.3***	†	16.7***	3.3	33.3***	5.6*
	(2.0)	(4.0)	(5.1)		(4.0)	(3.2)	(10.1)	(3.4)
β_{REGB}	−5.6*				10.6*			
	(3.0)				(6.1)			
$\beta_{REGB\text{-}def}$		−17.8***				5.6		
		(4.3)				(5.1)		
$\beta_{REGB\text{-}cd}$			−13.9*				5.6	
			(7.6)				(11.4)	
$\beta_{REGB\text{-}abef}$				†				13.9**
								(5.6)
Wald χ^2	3.34*	17.45***	3.22*		3.03*	1.21	0.24	6.11**
N	240	120	96		240	120	96	144

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGB} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGB\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REGB\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REGB\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).
†No conversion errors were observed.

Figure 3. Conversion Error Rates for the Regulatory-Information Treatment Overall (*REGI*), for Treatment Periods when the Monitor Does Not Observe the Fundamental (*REGI), and for the Fixed-Trigger (*FT*) Treatment**



Notes: The horizontal axes list the market fundamentals broken up into different ranges. The vertical axes list the fraction of the time a conversion error is made for fundamentals in each range. A decision to convert when the fundamental is $\geq$\$5.00 is counted as an error, as is a decision not to convert when the fundamental is <\$5.00.

because we expected that the possibility of such information would temper traders' tendencies to pessimistically incorporate the value of a conversion for market fundamentals slightly above \$5.00. This did not occur to any important extent. Nevertheless, the regulator regime still does slightly better (or at least no worse) than the fixed-trigger regime in the case of a value-decreasing conversion.

In the case of a value-increasing conversion, observe that the market response to monitors possibly knowing the market fundamental significantly undermines the accuracy of a regulator-based trigger mechanism when monitors are not informed. Compared with the *FT* regime, the incidence of conversion errors in the *REGI** periods is much worse. Except for the \$4.40–\$4.99 range, where conversion error rates for the two treatments are quite similar, conversion errors occur with a markedly higher frequency in the *REGI** periods than in the *FT* treatment.

The overall incidence of conversion errors in the *REGI* treatment, illustrated as the solid gray bars in figure 3, combines the beneficial direct effects of informed monitors with the indirect market responses reflected in the *REGI** periods and allows assessment of the net effect of probabilistically revealing the market fundamental to monitors. As seen in the left panel of the figure, in the case of a value-decreasing conversion, incorporation of the direct effects of market information on conversion errors makes the regulator in the *REGI-VDC* treatment unquestionably superior to the *FT-VDC* counterpart.

In the case of a value-increasing conversion, however, the direct effects of informed monitors, which scale down the overall incidence of conversion errors, are offset by the increased range of fundamentals for which conversion errors occur. Thus, while the overall incidence of errors in the *REGI-VIC* treatment is not noticeably higher than in the *FT-VIC* counterpart, the increased dispersion in errors causes monitors in the *REGI-VIC* treatment to both err more frequently when the market fundamental is not close to the $5.00 cutoff and commit more type II errors of commission (socially undesirable conversions).

The estimates of incremental effects reported in tables 4A and 4B provide a quantitative evaluation of the effects of probabilistically revealing the market fundamental to monitors. Table 4A summarizes indirect effects by estimating incremental error rates in the uninformed *REGI** periods relative to the *FT* regime. As shown on the left side of table 4A, in the case of a value-decreasing conversion, the accuracy of a regulator in the *REGI** periods is statistically indistinguishable from a fixed-trigger rule in all dimensions. In stark contrast, in the case of a value-increasing conversion, the indirect effects of probabilistically informed monitors unquestionably weakens a regulator-based triggering rule relative to a fixed-trigger mechanism. Over the four dimensions we evaluate, the regulator makes between 11.1 percent and 23.6 percent more errors in the *REGI** periods than does a fixed trigger, with the differences significant at $p < 0.01$ in all dimensions but one.

Table 4B summarizes the incremental effect of the *REGI* treatment relative to the fixed-trigger mechanism. For the value-decreasing conversion, the regulator mechanism performed significantly better under the *REGI* treatment than under the fixed-trigger treatment. There was a reduced overall rate of conversion errors, $\beta_{REGI} = -7.8\%$ ($p < 0.01$), a lower incidence of socially undesirable conversions, $\beta_{REGI\text{-}def} = -15.6\%$ ($p < 0.01$), and a lower incidence of "close" errors, $\beta_{REGI\text{-}cd} = -19.4\%$ ($p < 0.01$).

Table 4A. Linear Regressions that Describe the Incremental Effect of the Regulatory-Information Treatment when the Monitor Does Not Observe the Fundamental (*REGI) Relative to the Fixed-Trigger (*FT*) Treatment**

	Value-Decreasing Conversion				Value-Increasing Conversion			
β_0	13.3***	26.7***	33.3***	†	16.7***	3.3	33.3***	5.6*
	(2.0)	(4.0)	(5.1)		(4.0)	(3.2)	(10.1)	(3.4)
β_{REGI^*}	−2.2				13.3**			
	(2.9)				(5.4)			
$\beta_{REGI^*\text{-}def}$		−2.9				23.6***		
		(6.7)				(4.3)		
$\beta_{REGI^*\text{-}cd}$			−5.6				11.1	
			(7.4)				(12.2)	
$\beta_{REGI^*\text{-}abef}$				†				14.8***
								(4.7)
Wald χ^2	0.58	0.18	0.57		6.1**	30.8***	0.83	9.94***
N	150	72	60		150	78	60	90

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGI^*} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGI^*\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REGI^*\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REGI^*\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).
†No conversion errors were observed.

Table 4B. Linear Regressions that Describe the Incremental Effect of the Regulatory-Information (*REGI*) Treatment Relative to the Fixed-Trigger (*FT*) Treatment

	Value-Decreasing Conversion				Value-Increasing Conversion			
β_0	13.3*** (2.0)	26.7*** (4.0)	33.3*** (5.1)	†	16.7*** (4.0)	3.3 (3.2)	33.3*** (10.1)	5.6* (3.4)
β_{REGI}	−7.8*** (2.3)				0.0 (4.6)			
$\beta_{REGI\text{-}def}$		−15.6*** (4.6)				11.1*** (3.7)		
$\beta_{REGI\text{-}cd}$			−19.4*** (5.7)				−11.1 (10.7)	
$\beta_{REGI\text{-}abef}$				†				7.4* (4.0)
Wald χ^2	11.67***	11.62***	11.59***		0	8.8***	1.09	3.36*
N	240	120	96		240	120	96	144

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGI} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGI\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REGI\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REGI\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).
†No conversion errors were observed.

Remarkably, for the value-increasing conversion, the fixed-trigger mechanism performs better than *REGI* treatment despite the monitors' better information. There are significantly higher incidences of socially undesirable conversions ($\beta_{REGI\text{-}def} = 11.1\%,\ p < 0.01$) and gross errors ($\beta_{REGI\text{-}abef} = 7.4\%,\ p < 0.10$) in *REGI* than in the fixed-trigger treatment. The desultory effects of better information on the accuracy of conversion decisions is noteworthy: The possibility of monitors knowing the underlying fundamental reduces the informational content of prices, as traders, now less reluctant to incorporate the value of a conversion into their trading prices, generate prices in the ambiguous \$5.00–\$6.99 range with an increased frequency.[26]

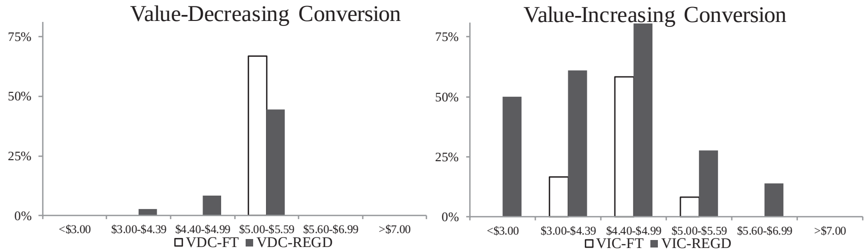
3.4 Optional Regulatory Delay

Finally, consider the conversion error rates for the *REGD* treatment relative to the *FT* treatment. Here, we count as a conversion error the decision to wait if the fundamental is below \$5.00, along with the decision to convert if the fundamental is above \$5.00. Figure 4 illustrates the incidence of conversion errors in the *REGD* treatment relative to the *FT* treatment.

In the case of a value-decreasing conversion, shown in the left panel of figure 4, notice that the incidence of conversion errors for fundamental realizations above \$5.00 in the *REGD* treatment is about two-thirds of that for the *FT* regime. For this reason, the presence of a monitor results in somewhat fewer errors of "commission" (although the improvement is smaller than observed in the *REGB* and *REGI* treatments). Nevertheless, in the case of a value-decreasing conversion, the *REGD* treatment does not obviously reduce the overall level of conversion errors relative to the *FT* regime. In fact, in the case of a value-decreasing conversion, the overall incidence of conversion errors closely parallels the levels observed in the *REGI** periods of the probabilistic information treatment in which monitors did not know the market fundamental prior to making conversion decisions.

[26]For example, for fundamentals in the \$4.40–\$4.99 range, mean price deviations from the ex post efficient level are much smaller in the *REGI* treatment than in either the *FT* or the *REG* treatments, as can be seen in figure A3.1 of unpublished appendix 3.

Figure 4. Conversion Error Rates for the Regulatory-Delay (*REGD*) and Fixed-Trigger (*FT*) Treatments



Notes: The horizontal axes list the market fundamentals broken up into different ranges. The vertical axes list the fraction of the time a conversion error is made for fundamentals in each range. A decision to convert when the fundamental is ≥$5.00 is counted as an error, as is a decision to wait when the fundamental is <$5.00.

On the other hand, as shown in the right panel of figure 4, in the case of a value-increasing conversion, allowing the monitor to delay acting until the arrival of additional information results in a uniformly higher incidence of conversion errors for every range of fundamental realizations. Particularly noticeable is the spectacularly high incidence of conversion errors in the *REGD* treatment for fundamental realizations below $5.00. Error rates are over 80 percent in the $4.40–$4.99 range and reach 50 percent even for fundamentals below $3.00. The high rate of conversion errors for fundamentals below $3.00 is particularly remarkable. Not only do the monitors defer action when prices fail to clearly convey accurate information regarding whether conversion is warranted, but surprisingly, they also defer acting even when prices clearly indicate that the fundamental is below $5.00. Of course, the cost to monitors of this conversion error is smaller than in the earlier treatments because the monitors still receive $6.00 if they wait, so that explains some of this high error rate. Overall, it appears that some monitors frequently choose to wait rather than make the effort and take the risk of making a decision.

Table 5 summarizes the incremental effects of the regulator-based rule relative to the fixed-trigger mechanism when the regulator has

Table 5. Linear Regressions that Describe the Incremental Effect of the Regulatory-Delay (*REGD*) Treatment Relative to the Fixed-Trigger (*FT*) Treatment

	Value-Decreasing Conversion				Value-Increasing Conversion			
β_0	13.3***	26.7***	33.3***	0.0	16.7***	3.3	33.3	5.6
	(2.0)	(4.0)	(5.1)	—	(4.0)	(3.2)	(10.1)	(3.4)
β_{REGD}	−0.6				26.7***			
	(4.0)				(6.7)			
$\beta_{REGD\text{-}def}$		−7.8				26.7**		
		(5.8)				(11.6)		
$\beta_{REGD\text{-}cd}$			−5.6				20.8*	
			(8.4)				(12.2)	
$\beta_{REGD\text{-}abef}$				2.8				30.6***
				(1.8)				(8.7)
Wald χ^2	0.02	1.79	0.44	2.34	15.70***	5.29**	2.94*	12.47***
N	240	120	96	144	240	120	96	144

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGD} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGD\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above $5.00. The coefficient $\beta_{REGD\text{-}cd}$ measures incremental errors for fundamentals within $0.60 of the $5.00. The coefficient $\beta_{REGD\text{-}abef}$ measures incremental errors for fundamentals not within $0.60 of $5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

the option to wait for the arrival of fully accurate non-market information. As can be seen in the left panel of table 5, in the case of a value-decreasing conversion, the combination of non-market information and an endogenously induced inaction bias does not exert any significant effects. Although the *REGD* treatment in this case generates fewer errors than the *FT* treatment in every dimension except one (the incidence of gross errors, $\beta_{REGI\text{-}abef} = 2.8$), none of the differences differ from zero at conventionally accepted significance levels.[27]

In contrast, in the case of a value-increasing conversion, the *REGD* treatment generates both sizably and significantly more conversion errors than the *FT* treatment by every measure, as can be seen by the large and uniformly positive coefficients in the right side of table 5. In fact, in the case of a value-increasing conversion, this inaction bias endogenously induced by the future availability of non-market information generates significantly higher conversion error rates than either inaction bias or the probabilistic availability of non-market information.[28]

As discussed earlier, in this treatment we count as a conversion error a decision to wait when conversion is socially desirable, even though conversion ultimately occurs with a delay. For this reason, the opportunity costs of inaction are lower than in the *REGB* treatment and, unlike the *REGI* treatment, regulators always have access to non-market information (although they must always make a wait/convert decision prior to seeing the fundamental).

The consequences of allowing monitors to delay action suggest two comments. First, the possibility of waiting for further information translated into a strong desire to wait, particularly when the prices were not fully revealing. These results provide one reason for why regulators may sometimes wait to act in practice.

[27] Note however, that a prior expectation that the delay treatment would result in reduced errors (e.g., using a one-tailed test) allows rejection of the null that $\beta_{REGD\text{-}def} = 0$ at $p < 0.09$.

[28] For the case of a value-increasing conversion, the *REGD* treatment yields significantly higher conversion error rates in every dimension than either the *REGB* or the *REGI* treatment, with the exception of an insignificant difference in the incidence of errors of commission with respect to the *REGI* treatment. Even this single insignificant difference is fairly large ($\beta_{REGD|I\text{-}def} = 15.6$, $p < 0.17$). Results appear in tables A4.5 and A4.6 of unpublished appendix 4.

Second, in the case of a value-decreasing conversion, the opportunity to wait fails to further enhance the performance of the regulator relative to the *REGB* and *REGI* regimes. Price deviations for the *REGD* and *REGI* treatment (not shown) are again similar in each case, but unlike the *REGI* treatment, monitors in the *REGD* treatment must uniformly make wait/convert decisions prior to seeing the fundamental.[29]

By way of contrast, in the case of a value-increasing conversion, the performance of a regulator in the *REGD* regime is worse than in the other treatments. Not only does the opportunity to delay action induce traders to more fully incorporate the value of a conversion into prices (as in the *REGI* regime), but the opportunity to receive fundamental information following a delay induces some monitors to ignore the informational content of prices altogether.[30]

3.5 Summary of Results

Table 6 provides an overall summary of results. Each row of the table summarizes instances in which conversion error rates differed significantly across the regulator and fixed-trigger regimes, where significance means $p < 0.10$ (two-tailed test). Capitalized entries R or F indicate a large difference, where the error rate for the regulator (fixed-trigger) regime exceeds that for the fixed-trigger (regulator) regime by at least 10 percentage points. Lowercase entries r or f

[29] In the case of a value-decreasing conversion, the *REGD* treatment generates significantly more errors of commission than either the *REGB* or the *REGI* treatments, as can be seen in tables A4.5 and A4.6 of unpublished appendix 4. In this case, the incidence of conversion errors for the *REGD* treatment is statistically indistinguishable from that in the *REGI** periods, as shown in table A4.7. In general, errors of commission occur in the value-decreasing case, as traders pessimistically incorporate the value of a conversion into prices when the fundamental is close to $5.00. The possibility of non-market information tends to make traders more pessimistic than inaction bias. Unlike the *REGI* treatment, however, monitors never see the underlying fundamental prior to making a conversion decision, and hence generate outcomes like those in the *REGI** periods.

[30] Figure A3.1 in unpublished appendix 3 illustrates the similarity of price deviations in the *REGI* and *REGD* regimes for case of a value-increasing conversion. The additional tendency of some monitors to ignore the informational content of prices altogether in the *REGD* regime can be formally seen in table A4.7 of unpublished appendix 4. Here, the overall incidence of conversion errors in the *REGD* treatment significantly exceeds that in the *REGI** periods because of a higher incidence of gross errors.

Table 6. Summary of Significant Differences in Conversion Rate Errors

	Value-Decreasing Conversion				Value-Increasing Conversion			
Error Type	***REG***	***REGB***	***REGI***	***REGD***	***REG***	***REGB***	***REGI***	***REGD***
Overall	f	f	f	—	—	R	—	R
Errors of Commission	F	F	F	—	R	—	R	R
Close Errors	F	F	F	—	—	—	—	R
Gross Errors	r	—	—	—	—	R	r	R

Notes: The “Overall” row is the comparison made for the entire range of fundamentals. The “Errors of Commission” row is for values of the fundamental greater than \$5.00. The “Close Errors” row is for values of fundamentals within \$0.60 of \$5.00. The “Gross Errors” row is for values of fundamentals that are more than \$0.60 away from \$5.00. R (r) indicates that the corresponding regulator regime generates >10 (<10) percentage points more conversion errors of the type indicated in the row than the fixed-trigger treatment. F (f) indicates that the fixed-trigger regime generates >10 (<10) percentage points more conversion errors than the corresponding regulator regime. All differences are significant at a minimum $p < 0.10$ (two-tailed test).

indicate smaller but still significant differences in conversion error rates. In the value-decreasing regime, shown in the left panel, the fixed-trigger rule usually generates a higher rate of conversion errors, although none of these differences persist in the *REGD* treatment. In contrast, for the value-increasing conversion case, shown in the right panel, the regulator regime usually yields more conversion errors. Differences tend to be large, and in the *REGD* treatment, they are large in every range.

4. Conclusion

CoCos have justifiably received a great deal of attention as a way to improve the stability of systemically significant financial institutions. An important unresolved issue regarding the implementation of CoCo bonds regards the selection of a triggering mechanism. Many academic proposals suggest use of a mechanistic fixed-triggering mechanism, which provides increased certainty regarding the timing and magnitude of a conversion relative to a regulator-based mechanism, and is less susceptible to manipulation. In practice, however, the triggering condition for many CoCo issues includes regulatory discretion. Theoretical analysis provides no clear guidance as to which type of mechanism is preferable. Furthermore, due to their novelty, no naturally occurring evidence exists to guide policymakers in their mechanism choice.

Our experimental results indicate that the relative superiority of mechanistic fixed-trigger and regulator-based triggering mechanisms depends on the conversion rule. In the case of a value-decreasing conversion, the regulator-based mechanism performs better in many of the cases, but not all of them. In particular, it is no better than the fixed-trigger mechanism in the regulatory-delay regime. In the case of a value-increasing conversion, a fixed-trigger mechanism robustly outperforms a regulator-based mechanism.

The usually better performance of the regulator regime under value-decreasing conversion is of interest because many policy proposals advocate for a fixed-trigger mechanism. However, it is important to note that the basis for this recommendation is often to prevent forbearance by supervisors, that is, to prevent supervisors not acting when a bank is in trouble because they fear political repercussions. Our experiment addresses a different regulatory decision issue, namely, whether regulators can effectively use information in prices to make decisions when the market takes that into account. Consequently, our results should not be viewed as a critique of forbearance arguments.

Nevertheless, our results do illuminate an aspect of decision making that is related to forbearance and to our knowledge has not been considered in the bank regulation literature. In our regulatory-delay treatment, we found that knowing that better information would arrive in the future made the monitors reluctant to act. This type of behavior looks like forbearance even though the motivations may well be different than those driving regulatory behavior during past banking crises.

Another result of interest was the perverse effect of giving the regulator more information in the *REGI* and *REGD* treatments. Improving a regulator's information or access to information does not necessarily improve the accuracy of conversion decisions. Here, apparently, if a regulator either may know or may have access to the underlying fundamental, the market acts as if the regulator in fact knows it, which is fine if the regulator actually has either good information or easy access to that information. But in states where the regulator is uninformed, prices convey even less useful information to the regulator. This result can be viewed as a variation on results in information economics, where more information is not necessarily better for welfare.

Contingent capital and the related idea of bail-in debt require an effective trigger if conversions are to be made before a bank reaches insolvency. The experiment reported in this paper indicates that the effectiveness of mechanisms that depend on price depend on the details of the conversion as well as the information and choices that a regulator has.

More generally, our experiment suggests that experimental methods can be a useful tool for evaluating mechanisms designed to improve the capital structure of a bank, prior to implementing them in practice. The results reported here identified ways in which the trigger design affects the frequency of conversion errors. If price-based triggers were to become closer to implementation, then it would be useful to know to what degree the forces we identified here vary with parameters such as the amount of delay or the asymmetry of regulatory incentives. Investigating these questions would require additional experiments with different parameterizations. Similarly, it may be desirable to evaluate the effects of other features of banking such as actions of other banks or the ability of a bank to influence its fundamental. Again, additional experimentation using variants of our design could also help answer these questions.

References

Avdjiev, S., P. Bolton, W. Jiang, A. Kartasheva, and B. Bogdanova. 2015. "CoCo Bond Issuance and Bank Funding Costs." Manuscript, Columbia Business School.

Balla, E., E. Prescott, and J. Walter. 2015. "Did the Financial Reforms of the Early 1990s Fail? A Comparison of Bank Failures and FDIC Losses in the 1986–92 and 2007–13 Periods." Working Paper No 15-05, Federal Reserve Bank of Richmond.

Berger, A., and S. Davies. 1998. "The Information Content of Bank Examinations." *Journal of Financial Services Research* 14 (2): 117–45.

Berger, A., S. Davies, and M. Flannery. 2000. "Comparing Market and Regulatory Assessments of Bank Performance: Who Knows What When?" *Journal of Money, Credit and Banking* 32 (3): 641–67.

Birchler, U., and M. Facchinetti. 2007. "Self-Destroying Prophecies? The Endogeneity Pitfall in Using Market Signals for Prompt Corrective Action." Manuscript, University of Zurich.

Bond, P., I. Goldstein, and E. Prescott. 2010. "Market-Based Corrective Actions." *Review of Financial Studies* 23 (2): 781–820.

Calomiris, C., and R. Herring. 2012. "Why and How to Design a Contingent Convertible Debt Requirement." In *Rocky Times: New Perspectives on Financial Stability*, ed. Y. Fuchita, R. J. Herring, and R. E. Litan (chapter 5). Washington: Brookings/NICMR Press.

———. 2013. "How to Design a Contingent Convertible Debt Requirement That Helps Solve Our Too-Big-to-Fail Problem." *Journal of Applied Corporate Finance* 25 (2): 39–62.

Davis, D., O. Korenok, and E. Prescott. 2014. "An Experimental Analysis of Contingent Capital with Market-Price Triggers." *Journal of Money, Credit and Banking* 46 (5): 999–1032.

DeYoung, R., M. Flannery, W. Lang, and S. Sorescu. 2001. "The Information Content of Bank Exam Ratings and Subordinated Debt Prices." *Journal of Money, Credit and Banking* 33 (4): 900–925.

Feldman, R., and J. Schmidt. 2003. "Supervisory Use of Market Data in the Federal Reserve System." Manuscript, Federal Reserve Bank of Minneapolis.

Fischbacher, U. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10 (2): 171–78.

Flannery, M. 1998. "Using Market Information in Prudential Bank Supervision: A Review of the U.S. Empirical Evidence." *Journal of Money, Credit and Banking* 30 (3): 273–305.

——. 2009. "Stabilizing Large Financial Institutions with Contingent Capital Certificates." Manuscript, University of Florida.

Hayden, E. 2014. "Coco Helping Investors Sleep?" *Economia* (Institute of Chartered Accountants in England and Wales, ICAEW) April 11.

McDonald, R. 2013. "Contingent Capital with a Dual Price Trigger." *Journal of Financial Stability* 9 (2): 230–41.

Pennacchi, G. 2011. "A Structural Model of Contingent Bank Capital." Manuscript, University of Illinois.

Pennacchi, G., T. Vermaelen, and C. Wolff. 2013. "Contingent Capital: The Case for COERCs." Manuscript, University of Illinois.

Prescott, E. 2012. "Contingent Capital: The Trigger Problem." *Economic Quarterly* (Federal Reserve Bank of Richmond) 98 (1): 33–50.

Raviv, A. 2004. "Bank Stability and Market Discipline: Debt-for-Equity Swap versus Subordinated Notes." Available at http://EconPapers.repec.org/RePEc:wpa:wuwpfi:0408003.

Roll, R. 1984. "Orange Juice and Weather." *American Economic Review* 74 (5): 861–80.

Siemroth, C. 2015. "The Informational Content of Prices when Policy Makers React to Financial Markets." Manuscript, University of Mannheim (November 14).

Sundaresan, S., and Z. Wang. 2015. "On the Design of Contingent Capital with a Market Trigger." *Journal of Finance* 70 (2): 881–920.

White, L. 1991. *The S&L Debacle.* New York: Oxford University Press.

Online Appendixes to Fixed Prices and Regulatory Discretion as Triggers for Contingent Capital Conversion: An Experimental Examination

Douglas Davis[a] and Edward Simpson Prescott[b]
[a]Virginia Commonwealth University
[b]Federal Reserve Bank of Cleveland

Appendix A1. Sample Instructions

Note: The following contains instructions for initial *BASE* periods that present the environment and pertinent incentives for traders and monitors. After two *BASE* periods, participants are given instructions for the *REGB, REGI*, or *REGD* treatment condition. Here we present instructions for the *REGB* condition in the case of a value-increasing conversion (termed in the instructions a "positive corrective action"). In each session, ten periods in one conversion condition are followed by ten periods in the other condition.

REGI instructions parallel the *REGB* instructions, with two differences. First, monitors in the *REGI* treatment suffer no penalty for making an unnecessary conversion. Second, in the *REGI* treatment monitors were told the underlying market fundamental each period with a 50 percent probability. Similarly, the *REGD* treatment parallels the *REGB* treatment, with two differences. First, monitors in the *REGD* treatment suffer no penalty for making an unnecessary conversion. Second, following the close of trade each period, the monitor chooses to either intervene (convert) or wait. If the monitor chooses to intervene, she earns either \$12.00 or \$0, depending on whether or not the decision is correct. If the monitor elects to wait, she earns \$6.00. In this case she is shown the underlying market fundamental and is obligated to make the correct decision.

Overview

Welcome! Thank you for coming to today's session. This is an experiment in the economics of decision making. Various foundations have provided funds for this research. The instructions are simple, and if you follow them carefully and make good decisions, you may earn a considerable amount of money that will be paid to you in cash at the end of the experiment. Your earnings will be determined partly by your decisions and partly by the decisions of others.

General Description

Today's experiment consists of two types of people, *traders* and *monitors*. *Traders* earn money from buying and selling units of an abstract stock we'll call an "asset." *Monitors* earn money from correctly guessing the asset's value. There will be ten *traders* and three *monitors*. The session consists of twenty-one *trading periods* in which traders buy and sell assets. The trading portion of each period will last 110 seconds. During each trading period, traders may buy and/or sell assets, and monitors observe contract prices and guess the asset's underlying value.

Actions and Incentives for Traders

At the outset of each period, traders are given a *portfolio* consisting of two assets. Traders are also given a $16.00 (lab) loan to purchase assets. Traders repay this loan at the end of each trading period, without interest.

The value of each asset to a trader is determined by the trader's *dividend*. The dividend is each asset's *intrinsic value*—that is, each asset held by a trader at the end of a period will be converted into this dividend. Traders' dividends for each period are determined as follows:

- The program takes (draws) a number over the range [$2.00 $8.00]; each number is equally likely to be selected. The draws were made by a computer prior to the experiment.
- For six randomly selected traders, their dividend equals this draw. We call these high-value traders.

- For the remaining four traders, their dividend value will be 60¢ below this draw. We call these low-value traders.
- Traders will know only their own dividend in a period. They will not know if they are a high-value or a low-value trader that period. That is, they won't know whether their dividend is high or low relative to the other traders in that period.

The dividend of an asset depends on whether it is held by a high-value or a low-value trader. Thus, the same unit may have different dividends for different traders.

At the end of each period each trader's portfolio net of their \$16.00 loan is converted to lab dollars: That is, they earn the sum of any cash on hand in excess of the \$16.00 which they have to pay back, plus their dividends for all assets held.

Actions and Incentives for Monitors

At the end of the trading period, monitors observe the *median* of all contract prices. The median price is the price that divides evenly the higher and the lower prices. Then monitors guess the high dividend. After all monitors make their guesses, the correct answer will be revealed.

Monitor payoffs are determined by the accuracy of their guesses. Specifically, monitors will earn

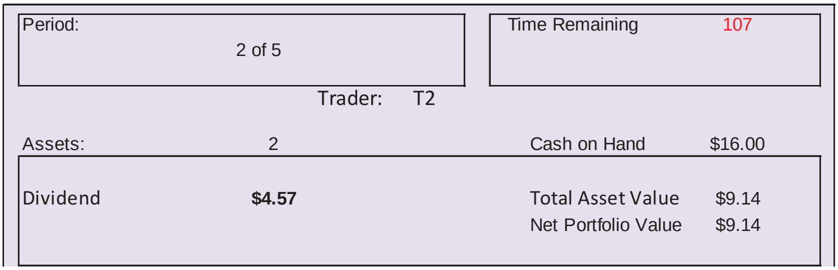
- \$3.00 (lab) if their guess is within 20¢ of the correct answer;
- \$1.00 (lab) if their guess is within 50¢ of the correct answer;
- \$0 otherwise.

Specific Instructions/Screen Displays

Trader's Screen Display

Upper Portion of Screen

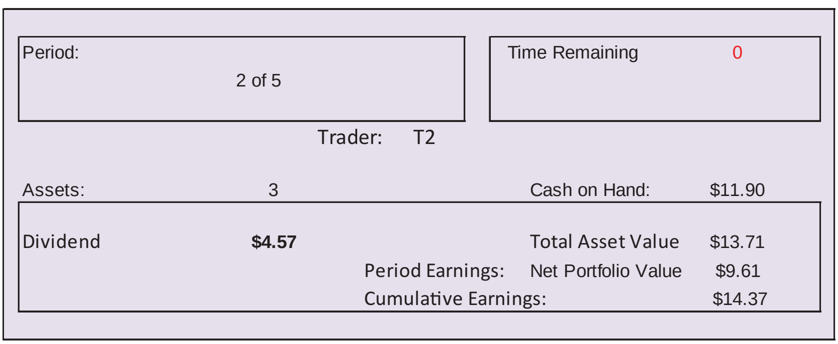
The upper portion of trader T2's screen at the beginning of a period is shown below. This screen conveys information regarding trader identity, the trading period, and the trader's portfolio.



Question: *Observe that the trader's dividend in the above example is \$4.57. What are the possible dividend values for the other traders? Why?*

Question: *Observe that the net portfolio value is \$9.14 despite the trader having cash on hand of \$16.00. Why?*

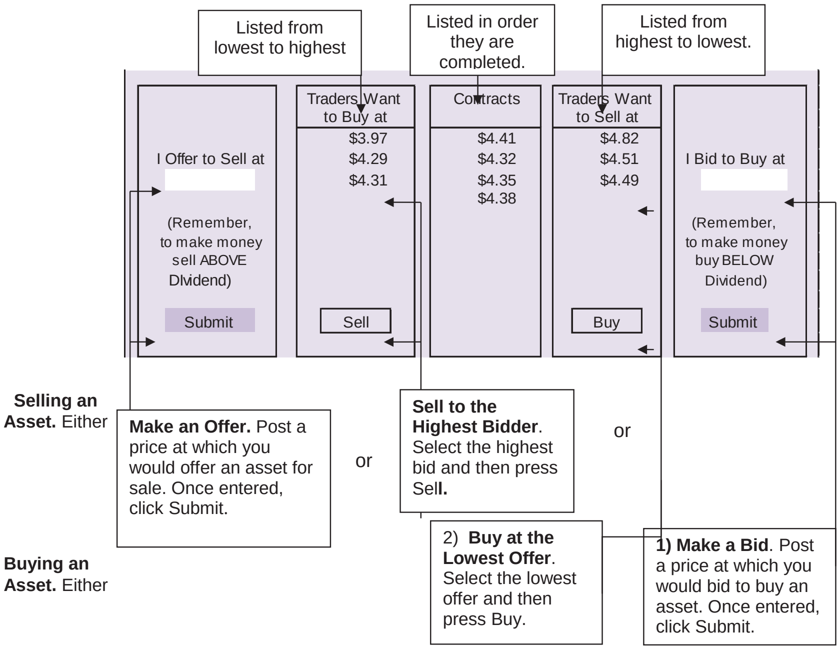
At the end of the period, the upper portion of the trader's screen reflects trader earnings, as seen below. Notice below that in period 2, trader T2 acquired one asset. Her net portfolio value is the sum of the total asset value and her remaining cash less the initial \$16.00 working capital loan. The trader's cumulative earnings for the session are also displayed. Once you've reviewed your earnings, press Continue.



Question: *Notice that trader T2 finished the period with a portfolio value of \$9.61. She started the period with only \$9.14. How did she increase her earnings?*

Trading Assets

A trader increases her net portfolio value by buying and selling assets. She uses the lower portion of the screen for this, as shown below.



Note: Traders may submit new offers and bids as often as they like.

Question: *Suppose a trader's dividend is $4.80.*

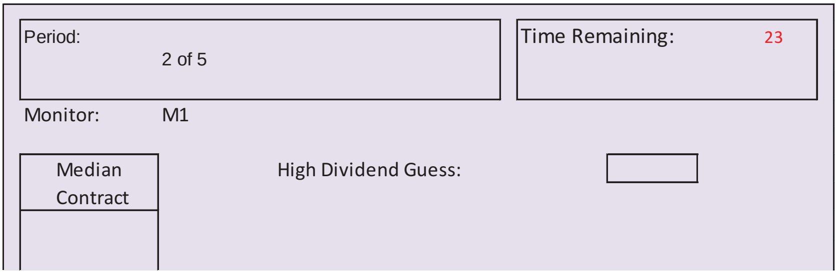
- *If she saw other traders offering to sell for $4.49 and bidding to buy for $4.31 (as shown above), should she consider buying or selling?*
- *How much could she earn in this case?*

Question: *Suppose a trader's dividend is $4.20.*

- *If other traders are offering to sell for $4.49 and bidding to buy for $4.31 (as shown above), should she consider buying or selling?*
- *What is the maximum amount she could earn in this case?*

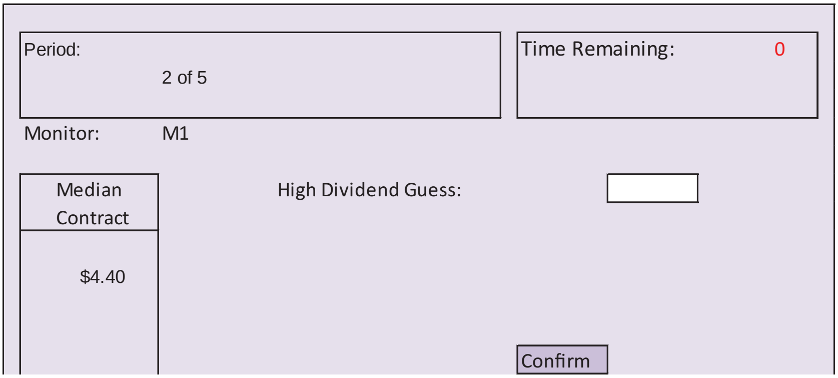
Monitor Screen Displays

The monitor does nothing until the trading period concludes. When the period ends, the monitor sees the median contract price.



Monitor Actions

When the trading period concludes, the median contract price is displayed. The monitor then guesses dividend for high-value traders and presses the Confirm button.



Monitor Earnings

When all monitors have submitted their guesses, the correct answer is revealed and earnings are calculated, as shown below.

High Redempton Value			
	Your Guess:	$4.41	
	Actual:	$4.47	
	Guess Earnings:		$3.00
Period Earnings:			$3.00
Cumulative Earnings			$3.50

After reviewing earnings, press Continue.

Question: *Why does the monitor in the illustration earn $3.00?*

Question: *Suppose a monitor guesses $7.82 and the correct answer turns out to be $7.60. How much does the monitor earn? Why?*

Question: *Suppose a monitor guesses $2.48 and the correct answer turns out to be $4.50. How much does the monitor earn? Why?*

Quiz of Understanding

1. Suppose you are a trader and your dividend is $7.31. What are the possible dividends for the other traders?

2. If the two dividends are $7.31 and $7.91 in a period, how many traders will have the $7.31 dividend and how many will have the $7.91 dividend?

3. Suppose a trader has a dividend of $3.47 and sees a bid to buy of $3.63. Should the trader consider buying or selling his asset? How much can he earn?

4. Suppose a trader has a dividend of $3.47 and sees an offer of $3.20. Should the trader consider buying or selling his asset? How much can he earn?

5. Suppose a monitor guesses that the high dividend value is $2.32 and it turns out that the high dividend is $2.58. How much does the monitor earn?

Final Details

1. Your identity as a monitor or as a trader will be revealed to you once the experiment starts. Other participants will not know your identity. Your role as a monitor or trader will remain fixed throughout today's session. However, it is important that you DO NOT publicly disclose your identity.

2. To ensure that you understand how the market proceeds, we will conduct one practice period. You will not be paid for your decisions in this period. During this practice period, please feel free to raise your hand and ask any questions you might have.

3. Any questions?

(following the practice period)

Thank you again for coming to today's session and bearing with us as we read through the instructions. Now we will begin the session.

1. The first portion of today's session consists of five trading periods under the conditions described above. After that we will stop and explain a second condition.

2. Your lab earnings will be converted to U.S. currency at a rate of 12 lab dollars = $1 U.S. Your total earnings for participating in today's session will be the sum of your earnings from trades or guesses plus the $6.00 appearance fee.

3. Any final questions? Please don't ask questions or talk to each other during the next five trading periods.

Summary Sheet
Baseline

Traders make money by buying and selling assets.

Buying and Selling Assets: To increase portfolio value,

- buy cheaply (at prices below dividend);
- sell dearly (at prices above dividend).

Dividends:

- Six traders have the high dividend.
- Four traders have the low dividend (60¢ below the high dividend).

Monitors make money by guessing the high dividend.

Guess Accuracy	Earnings (in lab dollars)
Within 20¢	$3.00
Within 50¢	$1.00
More than 50¢	$0.00

Treatment Condition (Positive Corrective Action—*REGB*)

Introduction

We now modify the market in one respect: In addition to guessing the high dividend each period, the monitors also make a decision to *intervene* or to *not intervene*.

Changes in Monitors' Incentives

In addition to guessing the high dividend, monitors not make an intervention decision. They may either earn or lose money from this decision.

If they *intervene*,

- they **earn** \$12.00 (lab) if the high dividend (before intervention) turns out to be less than \$5.00.
- they **lose** \$12.00 (lab) if the high dividend (before intervention) turns out to be more than \$5.00.

If they *do not intervene*,

- they **earn** \$12.00 (lab) if the high dividend (before intervention) turns out to be more than \$5.00.
- they lose \$0 if the high dividend (before intervention) turns out to be less than \$5.00.

After all monitors make their decisions, the choice of one of the three monitors will be randomly selected and implemented in the market.

Changes in Traders' Incentives

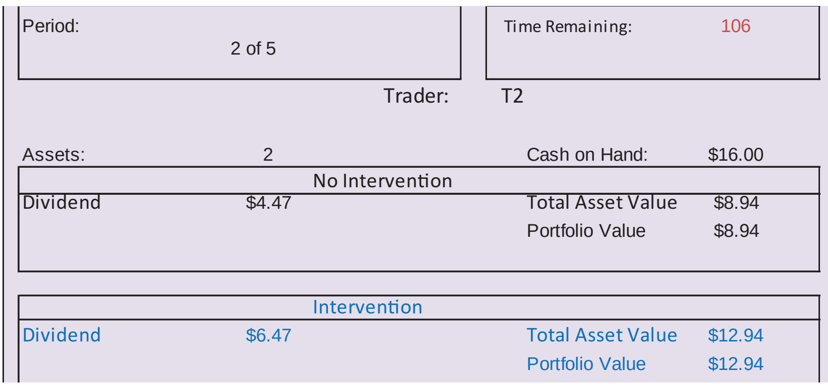
If the chosen monitor picks "*intervention*," all dividends increase by \$2.00. If the chosen monitor picks "*no intervention*," dividends do not change.

Specific Instructions: Changes Relative to the Baseline

Changes in Trader Screens

Upper Portion of Screen

The upper portion of the trader screen shown below is identical to that shown previously except now a new (blue) row of entries appears. The blue row lists the dividend, value of assets, and portfolio value in case the monitor intervenes.



Notice that the difference between black and blue lines is that the dividend increases by $2.00 per unit in the case of intervention.

Question: *When do the BLUE numbers determine dividends? What sort of contract prices would make the blue numbers more likely to be relevant (e.g., high or low)?*

End of Period

After trading concludes and monitors make intervention decisions, one of these decisions is implemented in the market. If the selected monitor does not intervene, the no-intervention part of the screen is bolded to emphasize the choice, as indicated below. Also, period and cumulative earnings appear.

Period:	2 of 5		Time Remaining:	0
Monitor Signal:	$5.24	Trader	T2	
Assets:	**4**		**Cash on Hand:**	**$7.92**
		No Intervention		
Dividend	**$4.47**		**Total Asset Value**	**$17.88**
		Period Earnings:	**Portfolio Value**	**$9.80**
		Cumulative Earnings:		**$15.32**
		Intervention		
Dividend	$6.47		Total Asset Value	$25.88
			Portfolio Value	$17.80

If the selected monitor does intervene, entries in the lower part are bolded, as shown below, and period and cumulative earnings appear.

Assets:	4		Cash on Hand:	$7.92
		No Intervention		
Dividend	$4.47		Total Asset Value	$17.88
			Portfolio Value	$9.80
		Intervention		
Dividend	**$6.47**		**Total Asset Value**	**$25.88**
		Period Earnings:	**Portfolio Value**	**$17.80**
		Cumulative Earnings:		**$19.32**

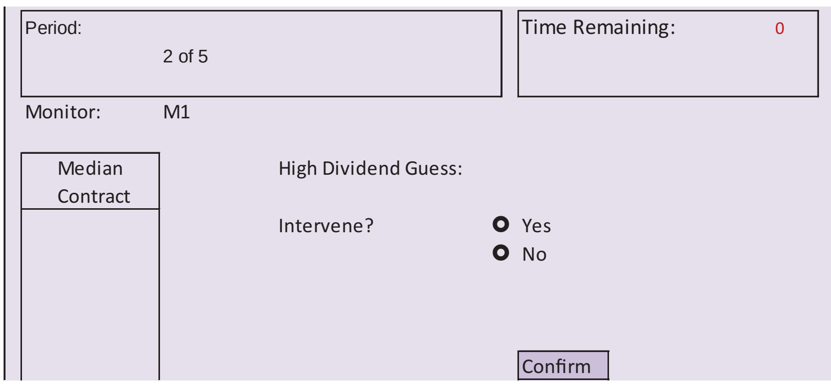
Question: *Suppose a trader T2 has a "no intervention" dividend of $4.47 and an "intervention" dividend of $6.47, as shown above.*

- *Trading starts and the trader sees contract prices of $5.23 and $5.35. Can she increase her portfolio by buying an asset for $5.50?*
- *Do these contract prices present any possible problems for traders and/or for the monitor?*

Question: *Suppose a trader T2 has a "no intervention" dividend of $2.47 and an "intervention" dividend of $4.47. Trading starts and* *she sees initial contract prices of \$2.73 and \$2.86. Can she increase her portfolio by buying an asset for \$3.50?*

Changes in Monitor's Screens

As the screen below shows, in addition to submitting a high dividend guess, the monitor makes an intervention decision.



The screen below shows that in addition to a return from guessing the high dividend, the monitor earns a return from the intervention decision.

In this example, the monitor intervened. This decision turned out to be the correct one because the actual high dividend was below \$5.00. Thus the monitor earns \$12.00 (lab) dollars from her intervention decision. The intervention part of monitor earnings would be \$0 had she chosen not to intervene in this case.

Suppose, however, in the above example that the high dividend turned out to be \$5.22. In this case the monitor would earn nothing from his dividend guess and would lose \$12.00 from incorrectly choosing to not intervene, as shown below. Losses will be deducted from the monitor's cumulative earnings.

High Dividend Value Guess		\$4.41
Intervene?	◉ Yes ○ No	

High Dividend Value		
Your Guess:	\$4.41	
Actual:	\$5.22	
Guess Earnings:		\$0.00
Intervention Earnings:		-\$12.00
Period Earnings:		(\$12.00)
Cumulative Earnings		\$1.00

Note: You must make an intervention decision that is consistent with your guess. That is, you must intervene if your high dividend guess is less than \$5.00 and you may not choose to intervene if your high dividend value is greater than or equal to \$5.00.

Question: *Suppose a monitor saw a median contract price of \$3.32 in a period. Would a monitor likely find an intervention to be profitable in this case? Why or Why not?*

Question: *Suppose a monitor saw a median contract price of \$5.32 in a period. Would a monitor likely find an intervention to be profitable in this case? Why or why not?*

Question: *In general, why is it riskier to intervene than to not intervene?*

Question: *Suppose a monitor saw a median contract price of \$7.32 in a period. Would a monitor likely find an intervention to be profitable in this case? Why or why not?*

Quiz of Understanding

1. Suppose a trader is given a dividend of \$4.37. What is the maximum possible value of that unit to the trader?
2. Consider a period where the high dividend value is \$2.63. In this case:
   a. What trading prices might a monitor observe?

   b. What could the monitor infer from contract prices about intervention? Why?

   c. Suppose a trader with a high dividend value sees another trader offering to sell an asset for \$3.20. Could the trader increase her portfolio value by buying this unit? Is this likely? Why or why not?
3. Consider a period where the high dividend value is \$7.89. In this case:
   a. What median contract prices might a monitor observe?

   b. What could the monitor infer from these prices about intervention? Why?

   c. Suppose a trader with a high dividend value sees another trader offering to sell an asset for \$9.00. Could the trader increase her portfolio value by buying this unit? Is this likely? Why or why not?
4. Suppose a monitor decides to intervene, but the high dividend turns out to be \$5.18.
   a. How much do traders with high dividend values earn from each asset?

   b. How much do traders with low dividends earn from each asset?

   c. What does the monitor earn from her decision to intervene?

5. Suppose a monitor decides to not intervene, but the high dividend turns out to be \$4.54.
   a. How much do traders with high dividend values earn from each asset?

   b. How much do traders with low dividends earn from each asset?

   c. What does the monitor earn from her decision to not intervene?

Final Details

1. There will be ten periods in this treatment. At the conclusion of this treatment the experiment will end, and you will be paid.

2. Your earnings will be the sum of your appearance fee, your earnings from the first part, and your earnings from this second part.

3. Questions? If not, we will begin. Again, thank you for your participation!

Summary Supplement
Positive Corrective Action

Traders make money by buying and selling assets.

To increase portfolio value,

- buy cheaply (at prices below the dividend);
- sell dearly (at prices above the dividend).

Dividends:

- Six traders have the high dividend.
- Four traders have the low dividend (60¢ below the high dividend).

Each period, traders know whether their dividend draw is high or low.

If the selected monitor intervenes, ALL dividend values increase by $2.00.

Monitors make money by guessing the high dividend and by making an intervention decision.

Earnings from guessing the high dividend:

Guess Accuracy	Earnings (in lab dollars)
Within 20¢	$3.00
Within 50¢	$1.00
More than 50¢	$0.00

If the monitor chooses to intervene,

- she earns $12.00 if the high dividend is below $5.00, and
- she loses $12.00 if the high dividend is greater than or equal to $5.00.

If the monitor chooses to not intervene,

- she earns $12.00 if the high dividend is greater than or equal to $5.00, and
- she loses $12.00 if the high dividend is less than $5.00.

Appendix A2. Probit Estimates of Conversion Error Rates

This appendix parallels the linear probability estimates reported in tables 2–5 in the text, using a probit estimation technique. Although coefficient estimates are not directly comparable, observe that instances of significant interactions largely overlap across the linear probability and probit estimates.

Table A2.1. Error Rates Comparisons of *REG* Relative to *FT* (probit regression coefficients)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−1.11***	−0.62***	−0.43***	−6.77***	−1.00***	−1.88***	−0.46	−1.59***
	(0.09)	(0.12)	(0.14)	(0.28)	(0.17)	(0.44)	(0.30)	(0.30)
β_{REG}	−0.31**				0.05			
	(0.14)				(0.21)			
$\beta_{REG\text{-}def}$		−0.47***				0.92**		
		(0.17)				(0.47)		
$\beta_{REG\text{-}cd}$			−0.45**				0.19	
			(0.19)				(0.36)	
$\beta_{REG\text{-}abef}$				4.74***				0.22
				(0.39)				(0.32)
Wald χ^2	1.95	3.05*	2.35	0	0.04	3.65*	0.31	0.38
N	510	234	156	318	510	234	156	354

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A2.2. Error Rates Comparisons of *REGB* Relative to *FT* (probit regression coefficients)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−1.11***	−0.62	−0.43***	†	−0.97***	−1.83***	−0.43	−1.59***
	(0.09)	(0.12)	(0.14)		(0.16)	(0.44)	(0.28)	(0.30)
β_{REGB}	−0.31*				0.36*			
	(0.18)				(0.21)			
$\beta_{REGB\text{-}def}$		−0.72***				0.49		
		(0.15)				(0.49)		
$\beta_{REGB\text{-}cd}$			−0.43*				0.15	
			(0.24)				(0.31)	
$\beta_{REGB\text{-}abef}$				†				0.73***
								(0.34)
Wald χ^2	1.58	5.52**	1.88		2.62	1.03	0.24	3.96**
N	240	120	96		240	120	96	144

Notes: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

Table A2.3(a). Error Rates Comparisons of *REGI Relative to *FT* (probit regression coefficients)**

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−1.11 (0.09)	−0.62 (0.12)	−0.43 (0.14)	†	−0.97 (1.94)	−2.24*** (0.57)	−0.43 (0.33)	−1.59*** (0.30)
β_{REGI^*}	−0.11 (0.15)				0.44 (1.08)			
$\beta_{REGI^*\text{-}def}$		−0.09 (0.21)				1.46** (0.58)		
$\beta_{REGI^*\text{-}cd}$			−0.16 (0.21)				0.29 (0.35)	
$\beta_{REGI^*\text{-}abef}$				†				0.76** (0.32)
Wald χ^2	0.17	0.08	0.21		3.48*	4.01**	0.74	3.81*
N	150	72	60		150	78	60	90

Notes: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

Table A2.3(b). Error Rates Comparisons of *REGI* Relative to *FT* (probit regression coefficients)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	-1.11^{***} (0.09)	-0.62^{***} (0.12)	-0.43^{***} (0.14)	†	-0.97^{***} (0.21)	-1.97^{***} (0.47)	-0.43 (0.28)	-1.59^{***} (0.30)
β_{REGI}	-0.48^{***} (0.13)				0.00 (0.18)			
$\beta_{REGI\text{-}def}$		-0.60^{***} (0.17)				0.82^{*} (0.48)		
$\beta_{REGI\text{-}cd}$			-0.65^{***} (0.18)				-0.33 (0.30)	
$\beta_{REGI\text{-}abef}$				†				0.47 (0.32)
Wald χ^2	3.59^{***}	3.93^{**}	4.12^{**}		0	2.28	1.15	1.55
N	240	120	96		240	120	96	144

Notes: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

Table A2.4. Error Rates Comparisons of *REGD* Relative to *FT* (probit regression coefficients)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−1.11*** (0.09)	−0.62*** (0.12)	−0.43*** (0.14)	†	−0.97*** (0.16)	−2.07*** (0.54)	−0.45 (0.31)	−1.60*** (0.30)
β_{REGD}	0.03 (0.18)				0.86*** (0.21)			
$\beta_{REGD\text{-}def}$		−0.18 (0.22)				1.54** (0.66)		
$\beta_{REGD\text{-}cd}$			−0.08 (0.26)				0.64* (0.38)	
$\beta_{REGD\text{-}abef}$				†				1.29*** (0.37)
Wald χ^2	0.01	0.39	0.06		15.96***	6.67**	3.32*	12.12***
N	240	120	96		240	120	96	144

Notes: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

Appendix A3. Efficiency Comparisons and Price Deviations

This appendix reports a series of bivariate linear regressions conducted to assess relative trading efficiency performance in the *REG*, *REGB, REGI*, and *REGD* treatments relative to the *FT* treatment. Also, at the end of this appendix we report average price deviations for each treatment, and to facilitate understanding of notes 26 and 30 in the text, we illustrate those deviations for some select treatments in the case of a value-increasing conversion.

We measure trading efficiency as a percentage of the eight asset units held by low-value traders at the beginning of a period that were held by high-value traders at the period's end. We use the same criteria for assessing relative efficiency performance used for assessing conversion error rates discussed in the text. Also, the following regressions use the market rather than the monitor as the unit of observation (thus explaining the difference in the number of observations for the regressions reported here relative to the comparable tables in the text). We control from repeated measures on markets across periods by modeling them as random effects. Finally, we do not assess separately the *REGI** periods relative to the *FT* treatment. For the purposes of trading efficiency, periods where the market fundamental is or is not revealed to monitors are indistinguishable, since such information is revealed only after the close of trade.

Figure A3.1. Mean Price Deviations from the Ex Post Efficient Price for the Fixed-Trigger (*FT*), Regulator (*REG*), Regulatory-Information (*REGI*), and Regulatory-Delay (*REGD*) Treatments under a Value-Increasing Conversion Rule

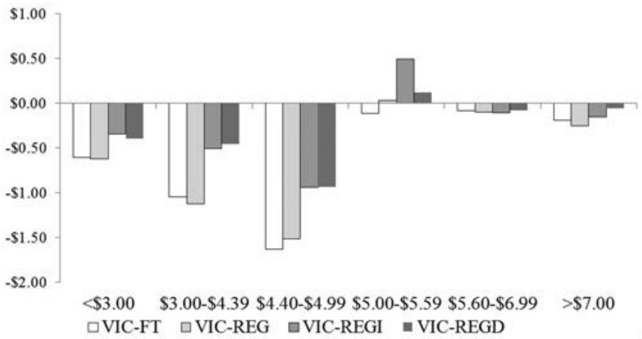


Table A3.1. Trading Efficiencies for *REG* Relative to *FT* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	0.79***	0.85***	0.79***	0.79***	0.79***	0.83***	0.77***	0.81***
	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)	(0.05)	(0.03)	(0.03)
β_{REG}	0.00				−0.03			
	(0.03)				(0.03)			
$\beta_{REG\text{-}def}$		−0.02				−0.03		
		(0.04)				(0.06)		
$\beta_{REG\text{-}cd}$			0.03				−0.11***	
			(0.04)				(0.04)	
$\beta_{REG\text{-}abef}$				−0.01				0.00
				(0.04)				(0.04)
Wald χ^2	0.00	0.21	0.44	0.15	1.02	0.31	6.95***	0.00
N	210	98	68	142	210	98	68	142

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REG} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REG\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above $5.00. The coefficient $\beta_{REG\text{-}cd}$ measures incremental errors for fundamentals within $0.60 of the $5.00. The coefficient $\beta_{REG\text{-}abef}$ measures incremental errors for fundamentals not within $0.60 of $5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A3.2. Trading Efficiencies for *REGB* Relative to *FT* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	0.79***	0.85***	0.79***	0.79***	0.79***	0.83***	0.77***	0.81***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
β_{REGB}	0.08***				−0.06**			
	(0.03)				(0.03)			
$\beta_{REGB\text{-}def}$		0.01				−0.11***		
		(0.04)				(0.04)		
$\beta_{REGB\text{-}cd}$			0.09**				−0.15***	
			(0.03)				(0.04)	
$\beta_{REGB\text{-}abef}$				0.07**				0.00
				(0.04)				(0.05)
Wald χ^2	6.82***	0.03	5.11***	3.99**	3.82***	7.05***	11.10***	0.01
N	120	60	48	72	120	60	48	72

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGB} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGB\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REGB\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REGB\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A3.3. Trading Efficiencies for *REGI* Relative to *FT* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	0.79***	0.85***	0.79***	0.79***	0.79***	0.83***	0.77***	0.81***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
β_{REGI}	0.07*				−0.06*			
	(0.04)				(0.03)			
$\beta_{REGI\text{-}def}$		0.05				−0.14***		
		(0.03)				(0.03)		
$\beta_{REGI\text{-}cd}$			0.08*				−0.23***	
			(0.04)				(0.04)	
$\beta_{REGI\text{-}abef}$				0.06				0.05
				(0.04)				(0.03)
Wald χ^2	3.45*	2.38	3.69*	1.83	3.76*	9.11***	28.69***	2.03
N	120	60	48	72	120	60	48	72

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGI} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGI\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REGI\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REGI\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A3.4. Trading Efficiencies for *REGD* Relative to *FT* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	0.79***	0.85***	0.79***	0.79***	0.79***	0.83***	0.77***	0.81***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
β_{REGD}	0.08***				−0.07**			
	(0.03)				(0.03)			
$\beta_{REGD\text{-}def}$		0.01				−0.11***		
		(0.04)				(0.04)		
$\beta_{REGD\text{-}cd}$			0.10**				−0.17***	
			(0.04)				(0.05)	
$\beta_{REGD\text{-}abef}$				0.06*				0.00
				(0.03)				(0.04)
Wald χ^2	8.42***	0.02	4.58**	3.43*	5.64**	7.88***	14.32***	0.00
N	120	60	48	72	120	60	48	72

Notes: Each column corresponds to a regression run on a different range of fundamentals. The coefficient β_{REGD} measures the incremental effect for the full range of fundamentals. The coefficient $\beta_{REGD\text{-}def}$ measures the incremental effect for errors of omission, that is, for the range above \$5.00. The coefficient $\beta_{REGD\text{-}cd}$ measures incremental errors for fundamentals within \$0.60 of the \$5.00. The coefficient $\beta_{REGD\text{-}abef}$ measures incremental errors for fundamentals not within \$0.60 of \$5.00. The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A3.5. Mean Price Deviations by Treatment

	Value-Decreasing Conversions					Value-Increasing Conversions				
Range	***FT***	***REG***	***REGB***	***REGI***	***REGD***	***FT***	***REG***	***REGB***	***REGI***	***REGD***
<$3.00	0.15	0.57	−0.09	0.15	−0.08	−0.61	−0.63	−0.33	−0.35	−0.40
$3.00–$4.39	0.19	0.58	0.01	−0.04	−0.11	−1.04	−1.12	−0.59	−0.51	−0.46
$4.40–$4.99	0.14	0.53	0.12	−0.06	0.08	−1.63	−1.51	−1.34	−0.94	−0.93
$5.00–$5.59	−0.67	−0.94	−0.92	−0.74	−1.18	−0.11	0.03	0.24	0.49	0.12
$5.60–$6.99	−0.39	−0.59	−0.35	−0.20	−0.39	−0.09	−0.10	0.03	−0.11	−0.08
≥$7.00	−0.37	−0.43	−0.31	−0.39	−0.61	−0.19	−0.25	−0.21	−0.16	−0.06

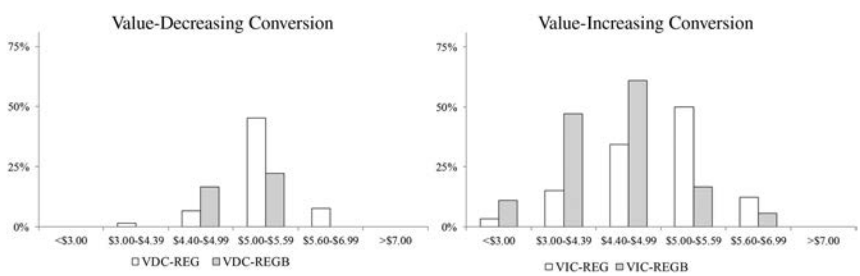
Appendix A4. Additional Error Rate Comparisons

In the text we report only the incremental effects on conversion errors of the various *REG* treatments relative to the *FT* treatment for purposes of conciseness. This restricted focus makes sense, because our primary interest is in the superiority or inferiority of each regulator treatment relative to the *FT* treatment. The interested reader, however, may also wish to see other cross-treatment effects. This appendix reports error rate comparisons of the *REGB*, *REGI*, and *REGD* treatments relative to the *REG* treatment. Also, due to the substantially larger conversion error rates in the *REGD* treatment relative to the *REGB* and *REGI* treatments, we compare across treatment conversion error rates for these treatments as well. In each case, our presentation parallels the comparisons in the text of the *REG* treatments relative to the *FT* treatment.

Conversion Error Rates Comparison, REGB to REG

Figure A4.1 illustrates the incidence of conversion error rates for the *REGB* and *REG* treatments. For the case of a value-decreasing conversion, shown in the left panel of figure A4.1, we observe that the introduction of inaction bias reduces the propensity for monitors to make socially undesirable conversions. The overall incidence of conversion errors, however, decreases only slightly. Moreover, the incidence of "gross" errors (e.g., conversion errors outside the range of fundamentals extending from 60¢ below to 60¢ above the \$5.00 cutoff for efficient intervention) falls.

Figure A4.1. Conversion Error Rates for *REGB* and *REG* Treatments



For the case of a value-increasing conversion, shown in the right panel, observe that the introduction of a regulator bias toward inaction shifts reduces the propensity of monitors to intervene when socially undesirable (e.g., for values above \$5.00), but at the cost of a substantially increased propensity for monitors to forgo socially desirable conversions for fundamental realizations in the \$4.00–\$4.39 and \$4.40–\$4.99 ranges. Also, in the case of value-increasing conversions, while fewer "errors of commission" occur in the *REGB* treatment, the overall incidence of conversion errors appears to rise.

Linear probability estimates of conversion error rates by the four performance criteria assessed in tables 2–5 in the text provide some statistical support for these observations.[1] In the case of a value-decreasing conversion, shown in the left panel of table A4.1, observe that inaction bias may modestly improve performance of the regulator treatment. Inaction bias significantly reduces the incidence of "gross" errors ($\beta_{REGB\text{-}abef} = -3.4$, $p < 0.10$). It also reduces the incidence of errors of commission but just misses significance at the 10 percent level ($\beta_{REGB} = -4.8$, $p < 0.113$).

On the other hand, looking at the case of a value-increasing conversion, shown in the right panel of table A4.1, observe that in the case of a value-increasing conversion, the introduction of inaction bias significantly reduces the incidence of errors of commission ($\beta_{REGB\text{-}def} = -8.7$, $p < 0.10$), but at the cost of an increase in "gross" errors ($\beta_{REGB\text{-}abef} = 10.9$, $p < 0.05$). Also, the overall incidence of conversion errors rises in the bias treatment but just misses the cutoff for significance ($\beta_{REGB} = 8.7$, $p < 0.113$).

Overall, a comparison of conversion error rates in the *REGB* and *REG* treatments supports the pertinent discussion in section 3.2 in the text. In the case of a value-decreasing conversion, inaction bias improves the relative performance of the regulator regime because it eliminates the incidence of "gross" errors—the single dimension for which the fixed-trigger regime was superior to the baseline regulator

[1]As in the text, we report linear probability estimates for expositional ease. Comparable probit estimates appear as tables A3.4 and A3.5 and at the end of this appendix. All regressions use the monitor as the unit of observation, and we control for repeated measures on the monitors by modeling them as random effects. Further to control for possible correlations across monitors within markets, we cluster data by markets.

Table A4.1. Error Rates Comparisons of *REGB* Relative to *REG* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	7.9*** (1.5)	13.7*** (2.8)	18.9*** (3.0)	3.4* (1.8)	18.5*** (3.2)	17.5*** (3.1)	40.1*** (6.4)	8.5*** (1.9)
β_{REGB}	−0.1 (2.7)				8.7 (5.5)			
$\beta_{REGB\text{-}def}$		−4.8 (3.1)				−8.7* (5.0)		
$\beta_{REGB\text{-}cd}$			0.5 (6.4)				−1.2 (8.2)	
$\beta_{REGB\text{-}abef}$				−3.4* (1.8)				10.9** (4.8)
Wald χ^2	0	2.5	0.01	3.37*	2.52	3.03*	0.02	5.21**
N	630	294	204	426	630	294	204	426

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

regime—while it also reduces the incidence of type II errors of commission. In the case of a value-increasing conversion, inaction bias weakens the performance of the regulator regime to a fixed-trigger regime because it significantly increases the incidence of "gross" errors outside the range extending 60¢ above and below the efficient conversion cutoff and, while not quite significant, may also increase the incidence of conversion errors overall.

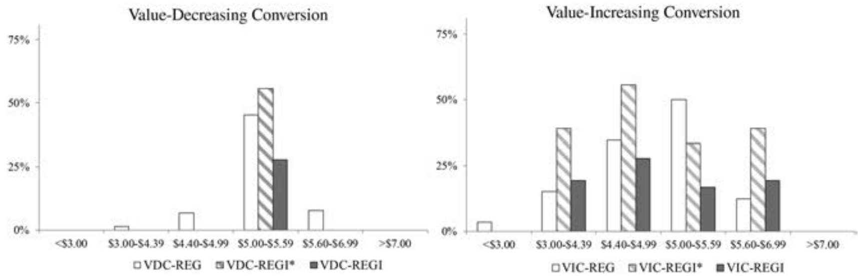
Conversion Error Rates Comparison, REGI to REG*

As discussed in the text, probabilistically revealing to monitors the underlying market fundamental has the necessarily ameliorative effect of reducing the incidence of conversion errors in those periods where the monitor is informed. Counterbalancing this effect, however, is an indirect effect of traders more fully incorporating the value of a conversion into prices on the belief that monitors are informed when in fact they are not. To assess the magnitude of this indirect effect, we assess the incidence of conversion errors in those periods of the *REGI* treatment where the market fundamental is not revealed to monitors. Label the pertinent periods *REGI**.

The striped and white bars in figure A4.2 plot the incidence of conversion errors in the *REGI** periods and the *REG* treatment, respectively. Looking first at the case of a value-decreasing conversion, shown in the left panel, conversion error rates in the *REGI** treatment concentrate in the \$5.00–\$5.59 range but occur at a somewhat higher frequency than in the *REG* regime. In the case of a value-increasing conversion, shown in the right panel, observe that error rates in the *REGI** regime exceed those for the *REG* regime in every range of fundamental realizations except one, the \$5.00–\$5.59 range. Even here, however, the relatively modest reduction in conversion errors of commission for fundamentals just above \$5.00 are more than offset by an increase in conversion errors in the \$5.60–\$6.99 range.

The linear probability estimates shown in table A4.2(a) verify the significance of the differences suggested by inspection of figure A4.2. As can be seen in the right panel of the table, in the case of a value-increasing conversion, monitors in the *REGI** periods perform worse in every respect than in the *REG* regime, and these differences are significant in every case but one. Turning to the left panel of the

Figure A4.2. Conversion Error Rates for *REGI and *REG* Treatments**



table, observe that in the *REGI** periods, the incidence of "gross" errors falls significantly ($\beta_{REGB\text{-}abef} = -3.2$, $p < 0.10$), but at the cost of a 10 percent significant increase in the incidence of errors of commission ($\beta_{REGB\text{-}def} = 10.1$, $p < 0.10$).

Overall, a comparison of the *REGI** periods with the *REG* treatment complements the comments in the text comparing the *REGI** and *FT* treatments: probabilistically revealing the market fundamental to monitors exerts a strong indirect effect. In the case of a value-decreasing conversion, the *REGI** periods increase the incidence of type II errors of commission relative to the *REG* treatment. At the same time, however, the incidence of "gross" conversion errors in the *REGI** periods falls relative to the *REG* treatment. In the case of a value-increasing conversion, this indirect effect weakens the performance of the regulator regime in each of the four dimensions assessed, and significantly so in three of four comparisons.

Conversion Error Rates Comparison, REGI to REG

The solid gray bars in figure A4.2 illustrate conversion error rates for the *REGI* treatment (including both states where monitors are and are not informed). Since the market fundamental was revealed to monitors in half the periods, combining informed and uninformed periods cuts by half the incidence of conversion errors for each range of fundamental realizations. Looking first at the case of a value-decreasing conversion, shown in the left panel of the figure, observe that the incidence of conversion errors in the *REGI* treatment is lower overall than for the *REG* regime and occurs over a narrower range of fundamental realizations.

In the case of a value-increasing conversion, shown in the right panel, notice that while revealing to monitors the market fundamental in half of the periods clearly reduces the incidence of "close" conversion errors, this reduction is offset by small increases in the incidence of "gross" conversion errors both above and below the efficient conversion cutoff in the *REGI* regime. In this way the indirect effect of traders incorporating the value of a conversion into prices when monitors are not informed importantly dampens the necessarily ameliorative effects of revealing fundamental information to monitors.

Results of the linear probability estimates shown in table A4.2(b) verify the significance of the differences suggested by inspection of figure A4.2. As can be seen in the left panel of the table, for the case of a value-decreasing conversion, the *REGI* regime yields fewer conversion errors by every measure than the *REG* regime, although only the reduction in the incidence of "gross" conversion errors is significant ($\beta_{REGI\text{-}abef} = -3.4$, $p < 0.10$). In the case of a value-increasing conversion, the *REGI* regime generates significantly fewer "close" errors ($\beta_{REGI\text{-}cd} = -17.8$, $p < 0.10$). At the same time, the overall incidence of conversion errors and the incidence of "gross" errors rise, albeit by insignificant amounts.

Overall, comparison of the *REGI* treatment with the *REG* treatment supports the results in section 3.3 of the text, that in the *REGI* treatment the indirect effects of traders incorporating the value of conversions into prices when monitors are uninformed leave a regulator-based regime inferior to a fixed-trigger mechanism in the case of a value-increasing conversion rule, but superior in the case of a value-decreasing conversion rule. As seen in the table, in the case of a value-decreasing conversion, the *REGI* treatment on net marginally reduces the incidence of conversion errors in all dimensions, but with the difference significant only for reductions in the incidence of "gross" conversion errors. In the case of a value-increasing conversion, the net effect of probabilistically revealing the market fundamental to monitors reduces the incidence of conversion errors relative to the baseline *REG* regime in only a single dimension and causes small (but statistically insignificant) increases in incidence of the conversion errors in both overall and "gross" conversion errors.

Table A4.2(a). Error Rates Comparisons of *REGI Relative to *REG* (%)**

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	7.8***	13.7***	18.9***	3.2*	18.4***	18.4***	39.5***	8.5***
	(1.4)	(2.8)	(3.0)	(1.7)	(3.1)	(3.3)	(6.1)	(1.8)
β_{REGI^*}	3.3				11.6**			
	(2.5)				(4.7)			
$\beta_{REGI^*\text{-}def}$		10.1*				8.6**		
		(5.9)				(4.3)		
$\beta_{REGI^*\text{-}cd}$			8.8				5.0	
			(6.0)				(9.0)	
$\beta_{REGI^*\text{-}abef}$				−3.2*				11.8**
				(1.7)				(3.7)
Wald χ^2	1.73	2.95*	2.15	3.53*	6.08**	4.13**	0.30	10.25***
N	540	246	168	372	540	252	168	372

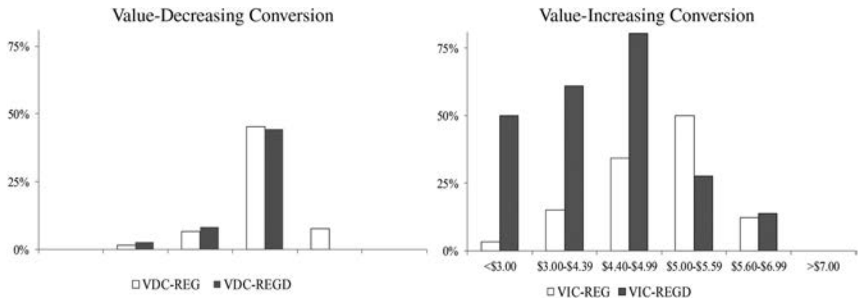
Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.2(b). Error Rates Comparisons of *REGI* Relative to *REG* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	7.8***	13.7***	18.9***	3.4*	18.4***	17.7***	40.0***	8.5***
	(1.5)	(2.8)	(3.0)	(1.8)	(3.1)	(3.2)	(6.3)	(1.9)
β_{REGI}	−2.2				−1.7			
	(1.8)				(3.8)			
$\beta_{REGI\text{-}def}$		−2.6				−3.2		
		(3.5)				(3.7)		
$\beta_{REGI\text{-}cd}$			−5.1				−17.8**	
			(4.0)				(7.2)	
$\beta_{REGI\text{-}abef}$				−3.4*				4.5
				(1.8)				(2.9)
Wald χ^2	1.57	0.57	1.61	3.37*	0.2	0.76	6.18**	2.38
N	630	294	204	426	630	294	204	426

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Figure A4.3. Conversion Error Rates for *REGD* and *REG* Treatments



Conversion Error Rates Comparison, REGD to REG

Figure A4.3 illustrates conversion error rates for the *REG* treatment and for the *REGD* treatment. Looking first at the left panel of the figure, observe that in the case of a value-decreasing conversion, the incidence of conversion errors in the *REGD* treatment parallels that for the *REG* treatment almost exactly. On the other hand, in the case of a value-increasing conversion, the opportunity to delay action spectacularly increases the incidence of conversion errors of omission relative to the baseline *REG* treatment, as monitors frequently elect to wait rather than make a conversion decision. In fact, as shown by the high frequency of conversion errors for fundamentals less than \$3.00, monitors frequently choose to wait even when prices unambiguously indicate that a conversion is not warranted.

Results of the linear probability estimates shown in table A4.3 verify the significance of the differences suggested by inspection of figure A4.3. As can be seen in the left panel of the table, in the case of a value-decreasing conversion, the *REG* and *REGD* treatments are statistically indistinguishable. On the other hand, in the case of a value-increasing conversion, the *REGD* periods generate significantly more errors overall ($\beta_{REGD} = 24.8$, $p < 0.01$), as well as significantly more "gross" errors ($\beta_{REGD\text{-}abef} = 27.6$, $p < 0.01$).

Finally, we observe that although the *REGD* treatment does not generate a significantly higher incidence of conversion errors than the *REG* treatment in the case of a value-decreasing conversion, it is in

Table A4.3. Error Rates Comparisons of *REGD* Relative to *REG* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	7.9***	13.7***	18.9***	3.3*	18.5***	18.3***	40.4***	8.5***
	(1.5)	(2.8)	(3.0)	(1.8)	(3.2)	(3.3)	(6.5)	(1.8)
β_{REGD}	4.9				24.8***			
	(3.7)				(6.2)			
$\beta_{REGD\text{-}def}$		5.2				11.7		
		(4.9)				(11.4)		
$\beta_{REGD\text{-}cd}$			8.8				13.8	
			(7.3)				(9.3)	
$\beta_{REGD\text{-}abef}$				−0.5				27.6***
				(2.5)				(8.0)
Wald χ^2	1.79	1.1	1.49	0.05	16.21***	1.06	2.21	11.79
N	630	294	204	426	630	294	204	426

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

this case that the *REGD* treatment is, in important respects, more prone to conversion error than either the *REGB* or *REGI* treatment. The comparisons of the *REGD* treatment with the *REGB* and *REGI* treatments in tables A4.4 and A4.5(a) illustrate. As seen below in the left panel of table A4.4, the *REGD* treatment generates 10 percent more errors of commission than the *REGB* treatment. Similarly, as seen in the left panel of table A4.5(a), the *REGD* treatment generates a significantly higher incidence of conversion errors than the *REGI* in every respect except the incidence of "gross" errors.

Looking at the right panels of tables A4.4 and A4.5(a), observe further that in the case of a value-increasing conversion, the *REGD* treatment performs significantly worse in every respect except the incidence of errors of commission relative to the *REGI* treatment. Even in this case the coefficient is fairly large ($\beta_{REGD|I\text{-}def} = 15.6$). However, the high variability of outcomes within treatments leaves this difference insignificant at standard significance levels ($p < 0.16$).

The generally weaker performance of the *REGD* treatment relative to both the *REGB* and *REGI* treatments merits some discussion, because in an important sense the delay option in the *REGD* treatment is a combination of the inaction bias and non-market information conditions that define the *REGB* and *REGI* treatments. Comparison of the *REGD* treatment with the uninformed periods of the *REGI* treatment (*REGI**) shown in table A4.5(b) provides some insight as to the generally increased incidence of conversion errors in the *REGD* treatment relative to the *REGB* and *REGI* treatments. The *REGD* treatment parallels the uninformed *REGI** periods in the sense that the monitor never sees the underlying fundamental prior to making conversion decision. In the case of a value-decreasing conversion, shown in the left panel of table A4.5(b), this results in an overall incidence of conversion errors that does not differ significantly in any respect from those observed in the *REGI** periods.

In the case of a value-increasing conversion, the *REGD* treatment generates significantly more conversion errors overall than the *REGI** periods ($\beta_{REGD|I*} = 13.3, p < 0.05$), because of an increased incidence of "gross" conversion errors ($\beta_{REGD|I*\text{-}abef} = 15.7$, $p < 0.10$), many in instances where monitors could unambiguously from price information determine that conversion would be desirable. We

Table A4.4. Error Rates Comparisons of *REGD* Relative to *REGB* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	7.8***	8.9***	19.4***	0.00	27.2***	8.9**	38.9***	19.4***
	(2.3)	(1.3)	(5.7)	—	(4.5)	(3.9)	(5.3)	(4.5)
$\beta_{REGD\|B}$	5.0				16.1*			
	(4.1)				(7.0)			
$\beta_{REGD\|B\text{-}def}$		10.0*				21.1*		
		(4.4)				(11.8)		
$\beta_{REGD\|B\text{-}cd}$			8.3				15.3*	
			(8.8)				(8.6)	
$\beta_{REGD\|B\text{-}abef}$				2.8				16.7*
				(1.8)				(9.1)
Wald χ^2	1.48	5.21*	0.89	2.35	5.24**	3.2*	3.16*	3.32*
N	360	180	144	216	360	180	144	216

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.5(a). Error Rates Comparisons of *REGD* Relative to *REGI* (%)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	5.6***	11.1***	13.9***	0.0	16.7***	14.4***	22.2***	13.0***
	(1.1)	(2.1)	(2.7)	—	(2.2)	(2.0)	(3.4)	(2.2)
$\beta_{REGD\|I}$	7.2**				26.7***			
	(3.6)				(5.8)			
$\beta_{REGD\|I\text{-}def}$		7.8*				15.6		
		(4.7)				(11.3)		
$\beta_{REGD\|I\text{-}cd}$			13.9*				31.9***	
			(7.2)				(7.5)	
$\beta_{REGD\|I\text{-}abef}$				2.8				23.1***
				(1.8)				(8.3)
Wald χ^2	4.05**	2.76*	3.69*	2.35	21.06***	1.89	17.95***	7.83***
N	360	180	144	216	360	180	144	216

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.5(b). Error Rates Comparisons of *REGD* Relative to *REGI (%)**

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	11.1	23.8	27.8	0.00	30.0***	27.0***	44.4***	20.4***
	(2.1)	(5.3)	(5.3)	—	(3.6)	(2.8)	(6.7)	(3.3)
$\beta_{REGD\|I^*}$	1.7				13.3**			
	(4.0)				(6.5)			
$\beta_{REGD\|I^*\text{-}def}$		−4.9				3.0		
		(6.7)				(11.5)		
$\beta_{REGD\|I^*\text{-}cd}$			0.0				9.7	
			(8.6)				(9.5)	
$\beta_{REGD\|I^*\text{-}abef}$				2.8				15.7*
				(1.8)				(8.6)
Wald χ^2	0.17	0.53	0	2.34	4.24**	0.07	1.04	3.34*
N	270	132	108	162	270	138	108	162

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

attribute this increased incidence of conversion errors in this circumstance to some monitors, given the opportunity to wait, choosing to ignore the information content of prices altogether.

Probit Estimates

Here, we provide the probit regressions that correspond to the linear regression results reported in this appendix. They are consistent with the linear results, so we provide them without any discussion.

Table A4.6. Error Rates Comparisons of *REGB* Relative to *REG* (probit estimates)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−142.0***	−109.3***	−88.0***	−202.7***	−94.8***	−95.8***	−26.9***	−137.3***
	(9.8)	(12.3)	(11.8)	(28.2)	(12.8)	(12.4)	(16.2)	(12.7)
β_{REGB}	0.0				31.8			
	(18.7)				(18.7)			
$\beta_{REGB\text{-}def}$		−25.5*				−41.2		
		(15.1)				(27.8)		
$\beta_{REGB\text{-}cd}$			1.8				−1.4	
			(22.4)				(21.3)	
$\beta_{REGB\text{-}abef}$				−475.1***				51.1***
				(39.4)				(20.2)
Wald χ^2	0	1.39	0.01	0	4.50**	3.34	0.01	8.93***
N	630	294	204	426	630	294	204	426

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.7(a). Error Rates Comparisons of *REGI Relative to *REG* (probit estimates)**

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−142.0	−109.3	−88.0	−202.7***	−94.4***	−98.2***	−26.9*	−137.3***
	(9.8)	(12.3)	(16.2)	(27.9)	(12.6)	(13.3)	(16.1)	(12.9)
β_{REGI^*}	20.0				40.7*			
	(15.3)				(16.2)			
$\beta_{REGI^*\text{-}def}$		38.0				32.5*		
		(20.9)				(17.6)		
$\beta_{REGI^*\text{-}cd}$			29.1				12.9	
			(19.0)				(23.3)	
$\beta_{REGI^*\text{-}abef}$				−471.7***				54.4***
				(39.4)				(16.9)
Wald χ^2	1.04	2.53	1.29	0	5.52**	1.67	0.3	6.22**
N	540	246	168	372	540	252	168	372

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.7(b). Error Rates Comparisons of *REGI* Relative to *REG* (probit estimates)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−142.0***	−109.3***	−88.0***	−202.7***	−94.0***	−97.3***	−26.9*	−137.3***
	(9.8)	(12.3)	(22.3)	(27.9)	(12.3)	(12.3)	(16.2)	(12.9)
β_{REGI}	−17.3				−4.1			
	(14.4)				(15.1)			
$\beta_{REGI\text{-}def}$		−12.8				−13.9		
		(16.8)				(18.7)		
$\beta_{REGI\text{-}cd}$			−20.5				−49.6*	
			(15.0)				(20.2)	
$\beta_{REGI\text{-}abef}$				−475.1***				24.5
				(39.4)				(16.2)
Wald χ^2	0.98	0.38	0.85	0	0.08	0.4	6.25	1.78
N	630	294	204	426	630	294	204	426

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.8. Error Rates Comparisons of *REGD* Relative to *REG* (probit estimates)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−142.0***	−109.3***	−88.0***	−196.1***	−94.5***	−98.9***	−27.1	−138.7***
	(9.9)	(12.4)	(11.5)	(25.8)	(12.7)	(13.1)	(18.7)	(12.7)
β_{REGD}	33.5*				82.9***			
	(18.7)				(18.5)			
$\beta_{REGD\text{-}def}$		29.0				48.6		
		(22.0)				(33.2)		
$\beta_{REGD\text{-}cd}$			37.2				45.8*	
			(23.7)				(27.1)	
$\beta_{REGD\text{-}abef}$				−5.6				107.7***
				(40.0)				(23.6)
Wald χ^2	5.32**	2.46	3.47*	0.03	34.92***	5.12**	4.37**	40.42***
N	630	294	204	426	630	294	204	426

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests).

Table A4.9. Error Rates Comparisons of *REGD* Relative to *REGB* (probit estimates)

	Value-Decreasing Conversions			Value-Increasing Conversions			
β_0	-1.42^{***} (0.16)	-1.35^{**} (0.09)	-0.86^{***} (0.20)	†	-0.61^{***} (0.14)	-1.46^{***} (0.28)	-0.28^{**} (0.14)	-0.86^{***} (0.16)
$\beta_{REGD\|B}$	0.33 (0.23)				0.50^{***} (0.19)			
$\beta_{REGD\|B\text{-}def}$		0.55^{***} (0.20)				0.95^{**} (0.43)		
$\beta_{REGD\|B\text{-}cd}$			0.35 (0.30)				0.46^{*} (0.24)	
$\beta_{REGD\|B\text{-}abef}$				†				0.56^{**} (0.26)
Wald χ^2	3.46^{*}	5.22^{**}	2.37	†	11.47^{***}	10.48^{***}	4.69^{**}	9.02^{***}
N	360	180	144		360	180	144	216

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

Table A4.10(a). Error Rates Comparisons of *REGD* Relative to *REGI* (probit estimates)

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−1.59***	−1.22***	−1.09***	†	−0.97***	−1.20***	−0.76***	−1.13***
	(0.10)	(0.11)	(0.12)		(0.09)	(0.17)	(0.12)	(0.11)
$\beta_{REGD\|I}$	0.51***				0.86***			
	(0.19)				(0.16)			
$\beta_{REGD\|I\text{-}def}$		0.42*				0.68*		
		(0.21)				(0.38)		
$\beta_{REGD\|I\text{-}cd}$			0.58**				0.94***	
			(0.25)				(0.23)	
$\beta_{REGD\|I\text{-}abef}$				†				0.82***
								(0.23)
Wald χ^2	7.02***	3.31*	5.75**	†	34.67***	5.16**	17.97***	17.54***
N	360	180	144		360	180	144	216

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

Table A4.10(b). Error Rates Comparisons of *REGD* Relative to *REGI (probit estimates)**

	Value-Decreasing Conversions				Value-Increasing Conversions			
β_0	−1.22***	−0.71***	−0.59***	†	−0.52***	−0.74***	−0.14	−0.83***
	(0.11)	(0.17)	(0.16)		(0.10)	(0.23)	(0.17)	(0.12)
$\beta_{REGD\|I^*}$	0.14				0.41**			
	(0.19)				(0.17)			
$\beta_{REGD\|I^*\text{-}def}$		−0.09				0.20		
		(0.25)				(0.40)		
$\beta_{REGD\|I^*\text{-}cd}$			0.08				0.31	
			(0.27)				(0.26)	
$\beta_{REGD\|I^*\text{-}abef}$				†				0.52**
								(0.23)
Wald χ^2	0.41	0.12	0.09		6.07**	0.33	1.5	5.19**
N	270	132	108		270	138	108	162

Note: The symbols ***, **, and * denote rejection of the null hypothesis at $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively (two-tailed tests). †No observations.

To Respond or Not to Respond: Measures of the Output Gap in Theory and in Practice*

Guy Segal
Bank of Israel

This paper analyzes the implications of responding to either the model-based New Keynesian output gap or to its estimates, and in particular, a Hodrick-Prescott-filtered output gap or a linearly detrended output gap. Responding to these estimates instead of to the "true" unobserved output gap generates long-lasting business cycles and lower welfare. Furthermore, correlations between the estimates and the theoretical output gap depend on the stochastic structure of the shocks affecting the economy. In particular, productivity shocks generate a negative such correlation. Hence, the output gap estimates may provide poor guidance to monetary policy.

JEL Codes: E32, E52.

1. Introduction

Over the last two decades, monetary policy analysis has been conducted mainly within the framework of the New Keynesian (NK) model, both by central bankers and by academia. A key variable in the NK model is the output gap, which reflects inflationary pressures from the demand side and explains inflation dynamics—as reflected in the New Keynesian Phillips curve (NKPC). The output gap is defined as the deviation of actual from potential output. Assuming that the monopolistic distortion remains,[1] potential output is

*I am greatly indebted to Zvi Hercowitz for his guidance and support. I am grateful to Alex Cukierman, Assaf Razin, John C. Williams, and Joseph Zeira for useful comments and discussions. I also thank Akiva Offenbacher, Alex Ilek, Eyal Argov, participants in the seminars in the Bank of Israel and in Tel-Aviv University, and an anonymous referee for helpful comments. Author e-mail: guy.segal@boi.org.il.

[1]Distortion stemming from monopolistic power of firms may be eliminated by the fiscal authority by a subsidy to production.

defined as output that prevails under flexible prices, while price markups are held constant at their steady-state values. As such, potential output is unobserved, and thus practical policy advice from the model requires the use of an estimate for potential output and output gap.

The various estimates for the output gap can be divided into univariate statistic methods (e.g., Hodrick-Prescott filter and detrended output methods), unobserved-components methods (e.g., Kalman filter), or the production function approach. In this paper we compare NK-model-based output gaps with their Hodrick-Prescott (HP) and linearly detrended output gap estimates, and we analyze the implications of responding to these estimates instead of responding to the "true" unobserved NK-model-based output gap. It should be noted that the different estimates of the output gap are both model and definition sensitive. However, the consensus is that (short term) potential output should be more volatile than traditional smoothed estimates of it, as potential output responds to real shocks in the economy (Justiniano and Primiceri 2008, Neiss and Nelson 2005, and Smets and Wouters 2003b).

We follow McCallum (2001) and choose an analytic rather than historical perspective. We conduct Monte Carlo simulations of an extension of the basic NK model introduced in Galí (2008), and derive the theoretical output gap as well as different estimates of it as endogenous variables, by calculating them within the model in "real time." Cúrdia et al. (2015) also calculate the Hodrick-Prescott-filtered output gap as an endogenous variable, by applying Christiano and Fitzgerald's (2003) approximation of ideal band-pass filters to Baxter and King's (1999) ideal HP-filtered output gap. The advantage of Cúrdia et al.'s calculation is that it does not dramatically augment the dimension of the model's state vector. We calculate the HP and linearly detrended output gap using their exact formulation. The advantage of our calculation is that it sheds light on the analysis of the formation of the HP-filtered series.

In the historical perspective the analysis uses actual data, and hence it is exposed to the problem of identifying the true shocks. Here, this problem is inexistent because the shocks are known. By choosing an analytic perspective, we can identify the effect of each "true" structural shock on the economy, and particularly on the output gaps, and conduct a normative analysis.

We assume that the central bank (CB) sets the monetary interest rate using a Taylor rule which responds to an output gap estimate, rather than the true output gap, while actual inflation responds to the true unobserved output gap.[2] We find that a response of the central bank to an estimate of the output gap—instead of to the "true" unobserved output gap—generates long-lasting business cycles as well as lower welfare. The business cycles are created by the fact that both the HP-filtered output and the linearly detrended output identify potential output by smoothing actual output series regardless of the source of the shock affecting the economy. Hence, they fail to identify the "true" potential output.

Furthermore, following a technology shock, these estimates provide completely misleading indicators of potential output. A positive technology shock raises both output and potential output, with the latter increasing by more, so that the New Keynesian output gap *decreases*. However, the estimates of the output gap translate the observed rise in output to a lower rise in potential output, and as a result the output gap estimates *increase*.

Consequently, the paper analytically explains the early literature, which focused on the technology shock/trend productivity as the main source for measurement error of the estimates of the output gap. Using an analytic perspective within the canonical New Keynesian model, McCallum (2001) assumes that a measurement error in the output gap estimate stems from the CB's ignorance of the effect of a technology shock on the output gap. He shows that when the CB responds to the output gap estimate, it leads to higher volatility in inflation, output, and interest rate than when the CB responds to the true output gap. Furthermore, as the CB's response to the estimate of the output gap increases, volatilities increase. Hence, McCallum (2001) concludes that the CB should not respond to an output gap estimate. Orphanides et al. (2000) choose a mixture of historical and

[2]Our assumption implies that the private sector has full information while the CB does not, as in Aoki (2003) and Svensson and Woodford (2004). The logic behind this assumption is that at the micro level, the single firm which sets its prices knows its own relevant information—for example, its current production and cost conditions—while the CB has only a macro estimate. An aggregation of all firms then implies a wider set of information about the macro state of the supply side than the information set held by the CB. Furthermore, as argued by Svensson and Woodford (2004), this assumption allows the implications of a partial information assumption to be tractable.

analytic perspectives, using the Federal Reserve's model of the U.S. economy, and examine the implications of measurement error in the estimate of the output gap for the design of monetary policy. They show, similar to McCallum (2001), that when the measurement error is large, such as when there is a significant structural change in the economy, the response of the CB to the measure of the output gap should decrease, although not completely. A similar result is shown in Orphanides (2003) and Smets (2002), where the latter interprets this as an explanation for the small empirical output gap coefficients in the interest rate rules reported in the literature. The long duration of the business cycle following a response of the central bank to one of the output gap estimates tested in this paper is explained by the fact that these estimates are derived by smoothing historical output series. This smoothing gives rise to sluggishness in the estimates, in contrast to the rapidly responding, forward-looking nature of the New Keynesian output gap. The implications of responding to a linearly detrended output gap are found to be the most pronounced, given the higher weights attached to historical output far in the past, compared with the weights in the Hodrick-Prescott filter. We find that correlations between the output gap and its estimates are sensitive to the variances of the different shocks (cost-push, demand, and technology shocks), and that the different estimates may provide misleading guidance to monetary policy, as has been shown empirically (Adolfson et al. 2014; Coenen, Smets, and Vetlov 2008; Neiss and Nelson 2005).

Another finding is that the volatility of the New Keynesian output gap and its estimates depends on monetary policy. If the central bank responds to the New Keynesian output gap, the volatilities of the true output gap and its estimates are about the same. However, when the central bank responds to an output gap estimate, the volatility of the New Keynesian output gap rises dramatically, while the volatility of the estimate is only slightly higher than when the central bank responds to the true output gap. This results from the fact that the output gap estimates examined in the paper are a moving average of actual output with an irregular inertia.

Finally, we follow the welfare analysis in Orphanides et al. (2000), where the inflation coefficient in the CB's interest rate rule is fixed, and the output gap's coefficient is optimal. Similar to Orphanides et al. (2000), we find that the optimal output gap coefficients exhibit attenuation, that is, the coefficients are small, although not zero. However, the price of responding to an output gap measure is higher output volatility. Like Orphanides et al. (2000), we also test the welfare implications of the CB's response to output growth rather than to output gap measure. We find that while responding to output growth is welfare improving in terms of the welfare function, it also leads to higher output volatility than in the alternative optimal rules, and hence it is preferable for the CB not to respond to output growth either.

The paper is organized as follows. Section 2 surveys the literature of the various definitions used for output gap within the NK model. Section 3 summarizes the NK model. Section 4 describes the calibration of the model. Section 5 analyzes the dynamic impulse response function and Monte Carlo simulations of the price markup shock, technology shock, and demand shock separately. Section 6 compares the output gap measures based on Monte Carlo simulations when all three shocks simultaneously hit the economy. Section 7 examines the welfare implications of responding to the various output gaps or to the output growth, and section 8 concludes.

2. Literature Review of Output Gap Definitions

In this section we summarize the various definitions used for the output gap in the New Keynesian framework. In this framework the output gap reflects inflationary pressures from the demand side and explains inflation dynamics.

As we assume that the monopolistic distortion is not offset by a subsidy and remains, we follow Adolfson et al. (2014), Sala, Söderström, and Trigari (2010), Smets and Wouters (2003a, 2003b), and Woodford (2003), and define potential output as that prevailing under flexible prices (and wages[3]), while price (and wage) markups are held constant at their steady-state values. Hence, potential output is an unobserved variable. Smets and Wouters (2003b) show that the above potential output definition[4] narrows the gap between

[3]For simplicity, the model used here does not include wage markups.

[4]There is no definite consensus regarding the different output gap definitions. Different papers may use different notions for the same definitions mentioned in the literature review.

traditional measures of the output gap and the model-based definition in comparison with natural output, in terms of volatility. They do, however, mention that potential output remains quite different from the traditional measures of the output gap.

Two other frequently used definitions in the NK framework are natural output and efficient output. Both, as well as the above definition of the potential output, relate to theoretical outputs when prices are flexible. The difference is that natural output (appendix 1) also responds to markup shocks, and it is relevant when the monopolistic distortion is eliminated. Efficient output (appendix 1) assumes perfect competition, and hence it is the relevant notion for welfare analysis.

Like Blanchard and Galí (2007) and Justiniano and Primiceri (2008), we derive analytically both potential and efficient model-based outputs. We show that natural output cannot be higher than efficient output, and that the gap between the two is a function of the changes in the monopolistic power of the firms (section 7). In contrast to Justiniano and Primiceri (2008), efficient output in our model is higher than potential output by a constant, as in Adolfson et al. (2014) and Sala, Söderström, and Trigari (2010). While this outcome is similar to Blanchard and Galí (2007), it stems from different assumptions and definitions. Blanchard and Galí assume constant demand elasticity and use natural output (as we define it) as efficient output, while we assume stochastic demand elasticity.

We find that the economy is sensitive to monetary policy response to different output gaps, a finding that is in line with Adolfson et al. (2014). They show that different output gaps (detrended or potential output gaps) in the central bank objective function imply different transmission mechanisms in the economy. Specifically, they show that when the theoretical output gap enters the objective function, the trade-off between stabilizing inflation and stabilizing output is lower than when it is the traditional estimate of potential output gap which enters the loss function. However, while Adolfson et al. (2014) use an ad hoc welfare loss function and a historical perspective, we use Galí's (2008) model-based welfare loss function and an analytical perspective to evaluate the welfare implications of responding to different estimates of the output gap. The model-based welfare loss function includes a negative linear term on the efficient output gap, apart from the squared deviations of inflation and efficient output gap from steady state. This negative term reflects the positive effect of an increase in output on welfare—as output is below its efficient level due to the monopolistic power of the firms. In sections 3–6 we focus on the output deviation from potential output, referring to it as output gap.

3. The New Keynesian Model and Model-Based Output Gaps

We use an extension of the basic NK model introduced in Galí (2008, ch. 3 and 5), where we assume the following: (i) A stochastic demand elasticity which reflects price markup shocks (following Adolfson et al. 2014, Christoffel, Coenen, and Warne 2008, Sala, Söderström, and Trigari 2010, and Smets and Wouters 2003a). In order to analyze monetary policy, a key feature of the model has to be the Taylor frontier—that is, the trade-off between inflation stabilization and output (gap) stabilization. As Clarida, Galí, and Gertler (1999) and Woodford (2003) show, the Taylor frontier may be introduced by including a shock to the Phillips curve. The stochastic demand elasticity reflects price markup shocks and induces the Taylor frontier. A decrease in the elasticity of demand implies an increase in market power of the monopolistic firms which raise prices. As prices rise, monetary interest rate is increased and output (gap) decreases. (ii) A preference shock to consumption in the utility function of the household, which reflects demand shocks (e.g., Cristoffel, Coenen, and Warne 2008 and Smets and Wouters 2003a).

3.1 The Model's Main Structural Equations

The model's two main structural log-linear equations are as follows (see appendix 1 for derivation):

$$\pi_t = \beta E_t\{\pi_{t+1}\} + \kappa\tilde{y}_t + \lambda\hat{\varphi}_t, \tag{1}$$

$$\tilde{y}_t = E_t(\tilde{y}_{t+1}) - \frac{1}{\sigma}(i_t - E_t(\pi_{t+1}) - r_t^p) + \frac{1}{\sigma}(1-\rho^c)\xi_t. \tag{2}$$

Equation (1) is the NK Phillips curve (NKPC) derived from firms' optimal price setting à la Calvo. $\pi_t \equiv p_t - p_{t-1}$ is the inflation rate; $\tilde{y}_t \equiv y_t - y_t^p$ is the potential output gap—that is, the gap between output, y_t, and potential output, y_t^p, which is the output that would prevail in a flexible price economy when price markup shocks are held constant at steady-state values. Note that the potential output gap is not a deviation of output from its steady state. The potential output gap reflects inflationary pressures from the demand side. Potential output increases with positive consumption preference shocks and positive technology shocks but decreases as real distortion, stemming from the economy's monopolistic structure, increases (see appendix 1 for derivation). $\hat{\varphi}$ is the log-deviation of a markup shock from its steady state, which is assumed to follow an AR(1) process, $\hat{\varphi}_t = \rho^{\varphi}\hat{\varphi}_{t-1} + v_t^{\varphi}$, where v_t^{φ} is white noise. β is the discount factor, and κ and λ are structural parameters. Small letters denote the log of a variable, $\wedge$ denotes the log-deviation of a variable from steady state, and the operator E_t denotes the mathematical unconditional expectation.

Equation (2) is the dynamic IS equation, which is derived from optimal consumer allocation (the Euler equation). i_t is the nominal interest rate and r_t^p is the potential real interest rate (see appendix 1 for derivation). ξ_t is a consumption preference shock, a demand shock, which follows an AR(1) process $\xi_t = \rho^{\xi}\xi_{t-1} + v_t^{\xi}$, where v_t^{ξ} is white noise. σ is the inverse of the consumption intertemporal elasticity of substitution.

3.2 Monetary Policy

In order to close the model, we assume that the central bank (CB) uses a standard Taylor rule to set the monetary policy rate:

$$i_t = \rho + \mu_\pi \cdot \pi_t + \mu_x \cdot gap_t. \tag{3}$$

Although the Taylor rule is not a structural equation, in contrast to the NKPC and the IS equations (equations (1) and (2)), it is the most commonly used rule to model monetary policy. μ_π and μ_x are the Taylor-rule parameters and are set at the standard values $\mu_\pi = 1.5$ and $\mu_x = 0.5/4$ (Taylor 1993). As potential output is unobserved, we calculate within the model the HP-filtered output gap (HPG) and the linearly detrended output gap (LTG), which are computed using the simulated output, y_t, in real time (appendix 3).[5,6] Finally, we assume that the steady state of the potential interest rate, ρ, is known.

To analyze the implication of the Taylor rule responding to an output gap estimate instead of the "true" output gap, we compare four variants of the model, which differ by the CB policy rule: (i) a Taylor rule which responds to the "true" potential output gap and is denoted by NKTR (New Keynesian gap Taylor rule), (ii) a Taylor rule which responds to the Hodrick-Prescott-filtered output gap (HPG), denoted by HPTR, (iii) a Taylor rule which responds to a linearly detrended output gap (LTG), denoted by LTTR, and (iv) a Taylor rule which responds only to inflation, denoted by SITR (strict inflation Taylor rule). The welfare analysis (section 7) also compares the implication of responding to output growth rather than to an estimate of output gap, a Taylor rule denoted by GTR (growth Taylor rule).

4. Calibration

Most parameters are calibrated based on Galí (2008, p. 52) (see table 1).

As for the markup shocks, in a basic model with exogenous cost-push shock, Galí (2008) calibrates the autoregressive coefficient of the cost-push shock $\rho^{\varphi} = 0.5$, while Sala, Söderström, and Trigari (2010) obtain the same value using U.S. data. Christoffel, Coenen, and Warne (2008) estimate ρ^{φ} in the range (0.255, 0.554) using European data, while Rabanal and Rubio-Ramírez (2008) estimate $\rho^{\varphi} = 0.97$ using U.S. data. In line with Christoffel, Coenen, and Warne (2008), we set $\rho^{\varphi} = 0.3$, as it yields a response similar to the estimated impulse response of inflation to a markup shock, as in Argov et al. (2012) and Christoffel, Coenen, and Warne (2008).

[5] Cúrdia et al. (2015) find that estimating a DSGE model with a Taylor rule, in which the HP-filtered output gap replaces the theoretical output gap, significantly improves the fit of the model to U.S. data. However, the main result of Cúrdia et al. (2015) is that including the efficient interest rate as the indicator of real activity in the policy rule fits the U.S. data better than alternative indicators do.

[6] The output trends are based on the last forty observations of simulated output (ten years' history).

Table 1. Calibrated Parameters

Parameter	Value	Description
α	1/3	Output elasticity with respect to labor
β	0.99	Discount factor
ε	6	Elasticity of substitution among goods
η	1	Frisch elasticity of labor supply
θ	2/3	Calvo price parameter
μ_π	1.5	Taylor-rule coefficient of inflation
μ_x	0.5/4	Taylor-rule coefficient of output gap
σ	1	Consumption intertemporal elasticity of substitution (inverse)
ρ^a	0.9	Persistence of technology process
ρ^φ	0.3	Persistence of price markup process
ρ^ξ	0.85	Persistence of consumption preferences shock process
σ^a	0.03	Standard deviation of technology process
σ^φ	0.0152	Standard deviation of price markup process
σ^ξ	0.03	Standard deviation of consumption preferences shock process

As for the demand shocks, we set the autoregressive coefficient at $\rho^\xi = 0.85$, following Rabanal and Rubio-Ramírez (2008) and Smets and Wouters (2003a).

We set the standard deviations of the shocks as follows. A standard calibration of a 20 percent price markup in steady state implies $\hat{\varphi}_t \in (-0.182, \infty)$ (section 7). Assuming normal distribution, $\bar{\sigma}^\varphi = 0.182/3 = 0.061$ is the upper limit of σ^φ (with a probability of 99.7 percent). Finally, we calibrate $\sigma^a = \sigma^\xi = 0.5\bar{\sigma}^\varphi$ and $\sigma^\varphi = 0.0152 \approx \bar{\sigma}^\varphi/4$, as this calibration yields plausible simulated measures of the output gaps (section 5).

5. Impulse Response Functions

In this section we analyze the impulse response functions to a positive price markup, technology and demand shocks, which are later used for simulated data.

5.1 Price Markup Shock—NKPC Shock

A positive shock to the NKPC reflects a higher monopolistic power that stems from a decrease in the elasticity of substitution among differentiated goods. We normalize the markup shock in the NKPC and denote it by $u_t \equiv \lambda\hat{\varphi}_t$.

If the central bank responds either to the "true" output gap (NKTR) or only to inflation (SITR), the impulse response functions to a markup shock are standard (figure 1). A positive markup shock leads to a higher inflation rate (π), while potential output (y^p) and potential real interest rate (r^p), by definition, do not change. Hence, in the presence of markup shocks only, potential output gap (output gap onwards) equals output. Inflation increases by more than the markup shock through the rational expectations channel—the expected increase in next-period inflation increases inflation today. To offset the markup shock, the interest rate (i) is increased and output (y) decreases roughly by 50 percent more than inflation. Thus, a trade-off between stabilizing inflation and stabilizing output gap emerges with a markup shock, reflecting the Taylor frontier. As the markup shock fades according to its data-generating process, the economy converges back to steady state in about six quarters.

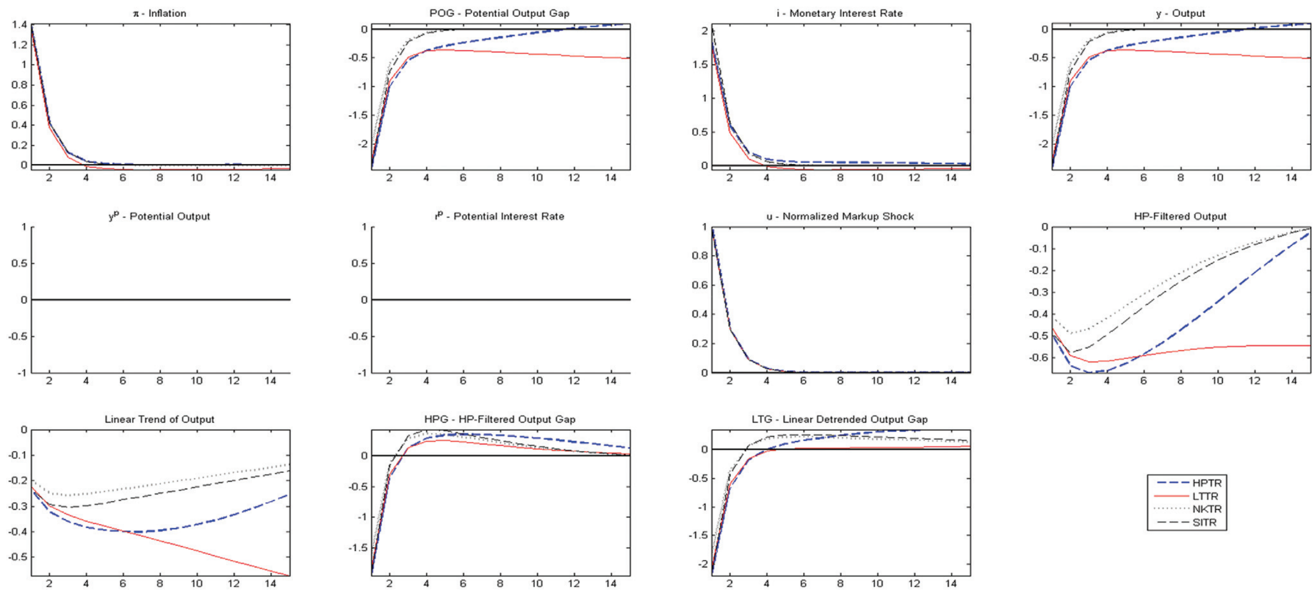
Note that when the central bank responds either to the "true" output gap (NKTR) or only to inflation (SITR), but the econometrician detrends output using the HP filter or a linear trend, both estimates of the output gap reflect the Taylor frontier in the first two quarters but show a false hump-shaped evolution of the gap afterwards. This is in line with Harvey and Jaeger (1993), who show that an HP-filter detrending may lead to spurious business cycles.

If instead the central bank itself responds to the Hodrick-Prescott-filtered output gap (HPG), the impulse response function (IRF) to the markup shock is similar to the IRFs analyzed above, but with two major differences: (i) The response of the CB to the HPG "contaminates" actual output, y, and as a result the potential output gap (POG) is also "contaminated," that is, the hump-shaped evolution in the economy is now a real one. (ii) As a result, steady state is reached only after fifty periods (figure 2).[7] Figure 2 differs from figure 1 only in tracing the longer-run response.[8]

[7]While the long duration of the convergence to steady state implied by the IRF may look unreasonable at first glance, its plausibility may be tested only by simulating all shocks simultaneously as in reality (section 4). Indeed, the simulated smoothed output gap estimates in that case resemble actual output gap estimates.

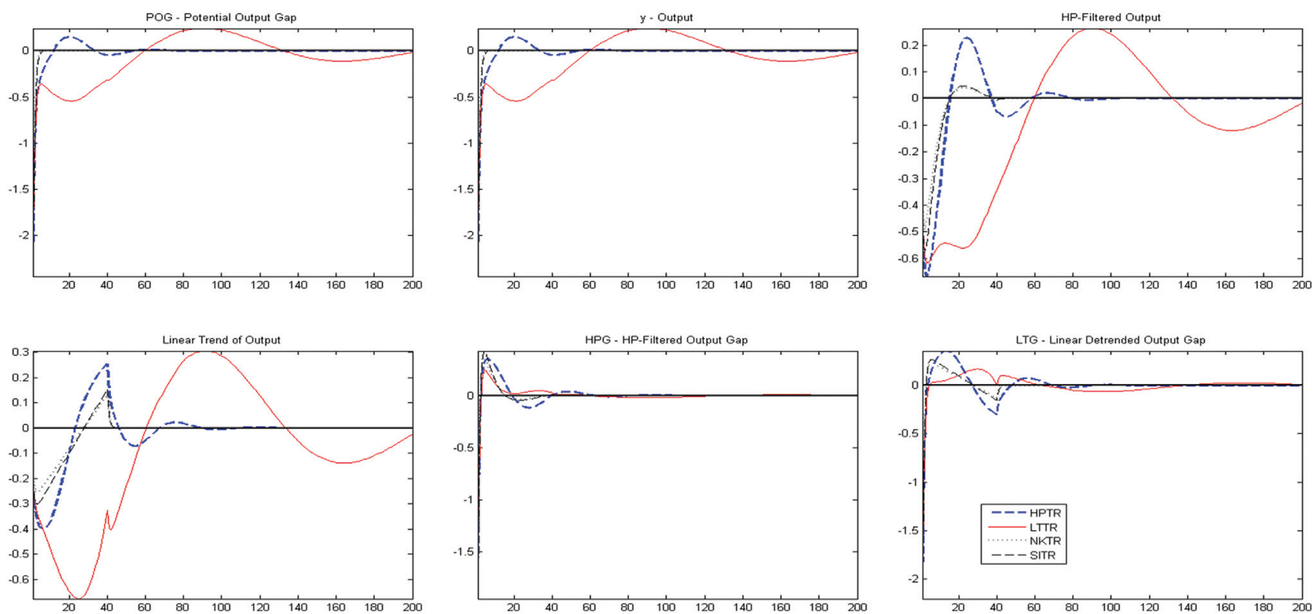
[8]We drop variables π, i, y^p, r^p, and u from figure 2, as they converge rapidly to steady state.

Figure 1. Impulse Response Function to a Positive 1 Percent Normalized Markup Shock—Short Horizon



Notes: Impulse response function to a positive 1 percent normalized markup shock under various output gaps in the policy rule: Hodrick-Prescott-filtered output gap (HPTR), linearly detrended output gap (LTTR), New Keynesian output gap (NKTR), and no output gap in the policy rule (SITR).

Figure 2. Impulse Response Function to a Positive 1 Percent Normalized Markup Shock—Long Horizon



Notes: Impulse response function to a positive 1 percent normalized markup shock under various output gaps in the policy rule: Hodrick-Prescott-filtered output gap (HPTR), linearly detrended output gap (LTTR), New Keynesian output gap (NKTR), and no output gap in the policy rule (SITR).

The implications of responding to the linearly detrended output gap (LTG) after a markup shock are striking (LTTR); output remains under its steady state for about seventy quarters (figure 2). The response to the LTG creates a business cycle which is four times longer than the business cycle created after responding to the HPG; POG and output does not converge to steady state even after 200 quarters.

The business cycles created by a response to either HPG or LTG are explained by the fact that both estimates fail to identify the "true" output gap. These estimates smooth actual output series in a way that creates an artificial inertia in the estimated output gap. The poor indicative power of the Hodrick-Prescott-filtered output and the linearly detrended output results from (i) the statistical nature of these estimates and (ii) the fact that both estimates have no economic link to the definition of the New Keynesian output gap, as these estimates do not use any information other than actual output. As both HPG and LTG are based on a weighted historical average of output (of forty observations in the simulations), their convergence is therefore a slow one. The longer business cycle under LTTR in comparison with HPTR is explained by the smoother weights attached to historical outputs in the calculation of LTG.[9]

5.2 Understanding the Source of the Created Business Cycles

In order to understand how these business cycles are created, we analyze the IRF presented above (markup shock only) and concentrate on the dynamic IS equation, the Phillips curve, and the policy rule:

$$\hat{x}_t = E_t\{\hat{x}_{t+1}\} - \frac{1}{\sigma}(i_t - E_t\{\pi_{t+1}\} - \rho),$$
$$\pi_t = \beta E_t\{\pi_{t+1}\} + \kappa \hat{x}_t + \lambda \hat{\varphi}_t,$$
$$i_t = \rho + \mu_\pi \cdot \pi_t + \mu_x \cdot gap_t.$$

We denote $gap_t \equiv y_t - \bar{y}_t$ as the measure of the output gap, where $\bar{y}_t$ is either the linear trend of output or its HP-filtered trend. $\bar{y}_t$ is

[9] We do not deal with the well-known endpoint bias of the HP-filtered output trend.

a function of the sample size used for the calculation of the output trend, T (equal to forty quarters in the simulations), and the historical output series. Specifically, the estimated endpoint of the output trend is calculated as $\bar{y}_t = \sum_{i=0}^{T-1} \chi_i y_{t-i}$, where y is the output up-to-date series in the last T periods and χ_i is derived from the first-order conditions of smoothing the target series (appendix 3). Note that χ_i has different signs for different i's. For the Hodrick-Prescott trend, with $T =$ forty quarters, χ_i is positive for $i = 0, 1, \ldots, 14$ and $i = 37$, 38, 39 and negative for the rest of the i's. Plugging the policy rule, the Phillips curve, and the output trend expression into the dynamic IS equation yields the next expression for the output gap:

$$\hat{x}_t = \left(1 + \frac{\mu_x}{\sigma} - \frac{\kappa}{\beta\sigma} - \frac{\mu_x}{\sigma}\chi_0\right)^{-1} \times \left[E_t\{\hat{x}_{t+1}\} + \frac{1}{\beta\sigma}(1 + \beta\mu_\pi)\pi_t - \frac{\lambda}{\beta\sigma}\varphi_t - \frac{\mu_x}{\sigma}\sum_{i=1}^{T-1}\chi_i y_{t-i}\right]. \quad (4)$$

Note that when the CB responds to the "true" output gap, which in the case of only markup shock equals output, then $\chi_i = 0\ \forall i = 0, 1, 2, \ldots, T-1$. According to equation (4), when the CB responds to the "true" output gap, the output gap is linked to its expectations, the rate of inflation, and the markup shock. When instead the CB responds to the deviation of the output from a trend $\bar{y}_t$, then the output gap is also a function of a weighted average of the last T observations of output, captured by the last term in equation (4):[10]

$$\left(1 + \frac{\mu_x}{\sigma} - \frac{\kappa}{\beta\sigma} - \frac{\mu_x}{\sigma}\chi_0\right)^{-1} \left[-\frac{\mu_x}{\sigma}\sum_{i=1}^{T-1}\chi_i y_{t-i}\right]. \quad (5)$$

Expression (5) explains the evolution of the output (and the "true" output gap) as shown in figures 1 and 2. Comparing the dynamics of the output gap when the CB responds to the "true" output gap

[10]Note that the result of long-duration business cycles is also obtained in the more general case of inertia in the Phillips curve. To see this, assume that the Phillips curve is given by $\pi_t = \beta[hE_t\{\pi_{t+1}\} + (1-h)\pi_{t-1}] + \kappa\hat{x}_t + \lambda\hat{\varphi}_t$, where h is the weight of expected inflation and $1-h$ is the weight of lagged inflation. In this case, following the methodology in section 5.2, the output gap is given by $\hat{x}_t = (1 + \frac{\mu_x}{\sigma} - \frac{\kappa}{\beta\sigma h} - \frac{\mu_x}{\sigma}\chi_0)^{-1}[E_t\{\hat{x}_{t+1}\} + \frac{1}{\beta\sigma h}(1 + \beta h\mu_\pi)\pi_t - \frac{1-h}{\sigma h}\pi_{t-1} - \frac{\lambda}{\beta\sigma h}\varphi_t - \frac{\mu_x}{\sigma}\sum_{i=1}^{T-1}\chi_i y_{t-i}]$.

(case A) or, instead, to its measure (case B), in the first period the additional influence on the output gap in case B in comparison with case A is $-(1+\frac{\mu_x}{\sigma}-\frac{\kappa}{\beta\sigma}-\frac{\mu_x}{\sigma}\chi_0)^{-1}\frac{\mu_x}{\sigma}\chi_0 y_{t-1}$, a minor influence of less than 1 percent of output as $\mu_x = 0.125$, $\sigma = 1$, and $\chi_0 = 0.0963$ in the case of a linear trend.[11] Thus, in the first period the IRFs of the two cases are quite similar. However, over time, there is an additional influence of lagged output. This influence is reflected by inertia in the output, which delays the convergence to steady state in case B in comparison with case A. Because χ_i is positive for $i = 0, 1, \ldots, 14$ and $i = 37$, 38, 39, and negative for the rest of the i's, and given the fact that in each period there is an additional influence of a different range of lagged output series, in the first twenty periods the output gap in case B is lower than in case A. Twenty periods after the markup shock hits the economy, the influence of the lagged output in case B becomes positive and causes a higher rate of change in output in comparison with case A. Over time, the (positive) influence of the expected output gap and the (negative) influence of inflation[12] decline as the shock fades, and hence output volatility becomes more moderate.

As the model's potential output (and potential interest rate) is not affected by markup shocks, it has no correlation with the HP-filtered output, the linearly detrended output, or any other univariate filtered output trend. However, the relevant question is what the correlation between the true output gap (POG) and its estimates, HPG and LTG, is. In order to answer this question and to better understand the implications of the markup shocks, we conduct Monte Carlo simulations. In each period a random AR(1) markup shock drawn from $\hat{\varphi}_t \sim N(0, \sigma_\varphi^2)$ hits the economy. The Monte Carlo simulation starts from steady state. In order to have a more reasonable comparison, we drop the first 100 simulation periods and calculate the different output gap estimates and statistics using the rest of the simulated data. We simulate 400,000 periods and calculate the mean of the 1,000 consecutive samples of 400 periods (each reflecting a sample of 100 years).

[11] In the baseline calibration, $(1+\frac{\mu_x}{\sigma}+\frac{\kappa}{\beta\sigma}+\frac{\mu_x}{\sigma}\chi_0)^{-1}$ is 0.886 in case A and 0.876 in case B.

[12] The influence of inflation is similar in both cases because inflation is similar, as shown in figures 1 and 2. The analysis of the expected output gap influence, however, is not complete. A complete analytic analysis is not trivial due to the influence of lagged output on the output gap in case B.

Figure 3 illustrates POG, HPG, and LTG when only markup shocks hit the economy and the CB responds to the true output gap (when using NKTR) in a simulated period of 100 years. When only markup shocks hit the economy, the correlation between POG and HPG is very high and ranges between (0.82, 090),[13] while that between POG and LTG is even higher, between (0.87, 0.95). However, although correlation is high, the different measures of the output gap differ from the true guidance reflected by the POG with about 17 percent of the cases of the HPG (the upper bar plot) and about 13 percent of the cases of the LTG (the lower bar plot).

5.3 Technology Shock

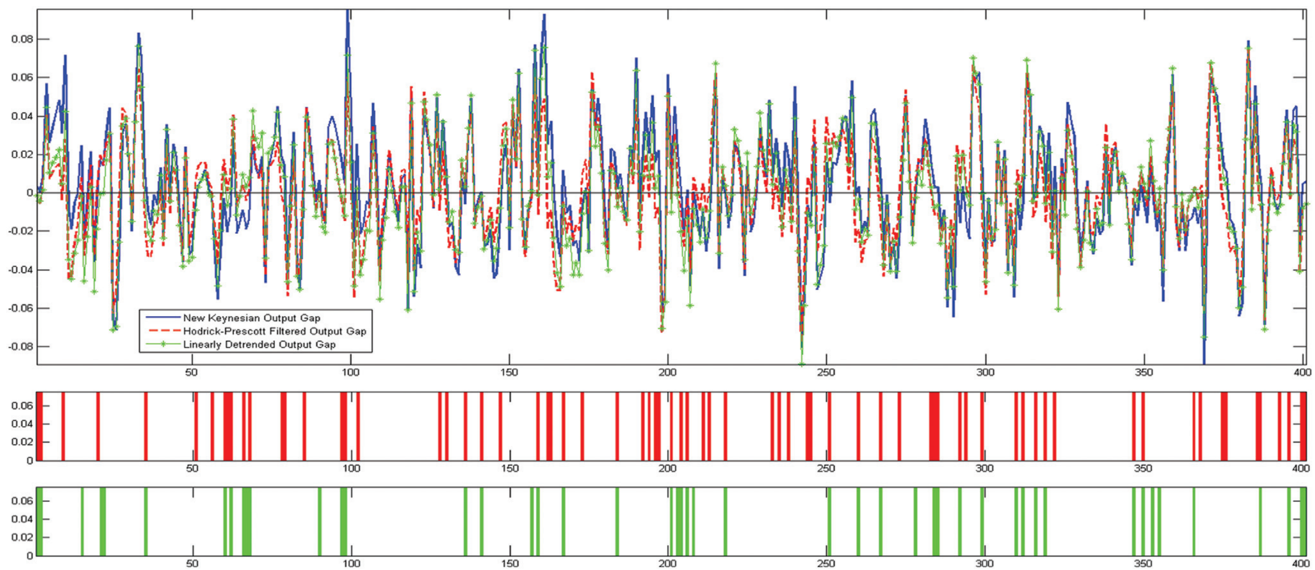
The impulse response function to a positive 1 percent technology shock (a) is shown in figure 4. First, we analyze the IRF under NKTR and SITR. A positive shock to technology level leads to higher levels of outputs—both actual and potential, although not by the same magnitude. Due to the fading technology shock, expected potential output growth after the first quarter is negative, implying that the potential interest rate declines on impact. Actual output rises by less than potential output. Thus, the output gap (POG) decreases after a positive technology shock. To complete the analysis, following a positive technology shock, (marginal) costs decrease and inflation decreases. Hence, a decrease in the nominal interest rate (i) follows. As the technology shock fades, the economy converges gradually back to steady state.

As in the analysis of the markup shock, the IRFs assuming either NKTR or SITR are monotonic and converge to steady state rapidly, while the IRFs under HPTR or LTTR create long-lasting business cycles, which are much more significant under LTTR (figure 5).

Recall that a positive technology shock raises both output and potential output, with the latter increasing by more, so that the New Keynesian output gap decreases. However, the statistical estimates of the output gap translate the observed rise in output to a lower rise in potential output, and as a result the output gap estimates rise. Hence, the output gap estimates provide misleading indicators about the output gap. Note that the monetary interest rate has a different path under each different policy rule, in contrast with the case of a markup shock only. These different paths contribute further

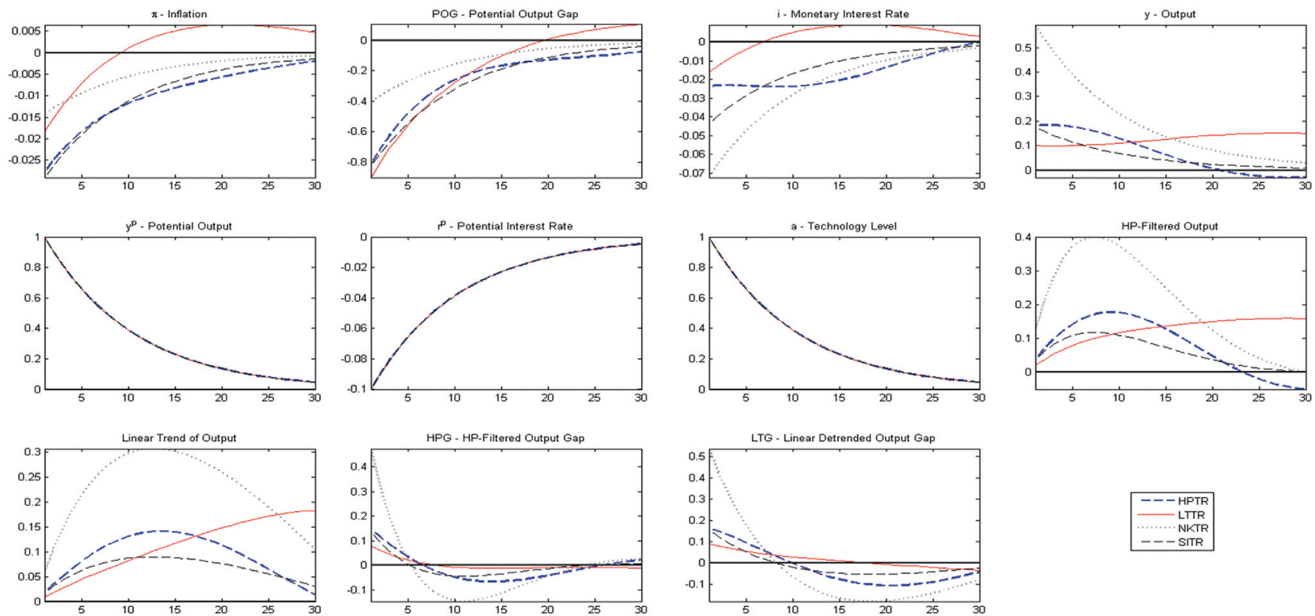
[13]This range reflects a 95 percent confidence interval of the Monte Carlo simulations.

Figure 3. Simulated Output Gaps Driven by Randomized Price Markup Shocks; CB Response to the "True" New Keynesian Output Gap



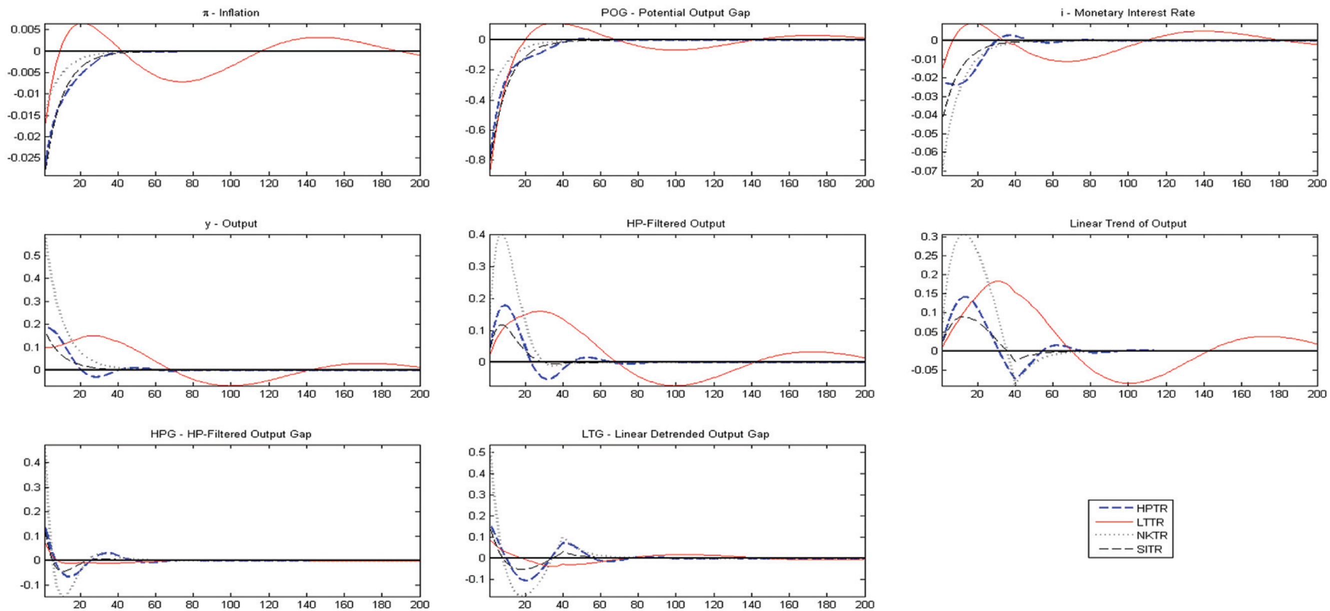
Notes: Simulated output gaps driven by randomized price markup shocks when the CB responds to the "true" New Keynesian output gap: New Keynesian output gap, Hodrick-Prescott-filtered output gap, and linearly detrended output gap. The upper bar plot signals when the "true" output gap and the HP-filtered output gap have different signs. The lower bar plot signals when the "true" output gap and the linearly detrended output gap have different signs.

Figure 4. Impulse Response Function to a Positive 1 Percent Normalized Technology Shock—Short Horizon



Notes: Impulse response function to a positive 1 percent normalized technology shock under various output gaps in the policy rule: Hodrick-Prescott-filtered output gap (HPTR), linearly detrended output gap (LTTR), New Keynesian output gap (NKTR), and no output gap in the policy rule (SITR).

Figure 5. Impulse Response Function to a Positive 1 Percent Normalized Technology Shock—Long Horizon



Notes: Impulse response function relating to a positive 1 percent normalized technology shock under various output gaps in the policy rule: Hodrick-Prescott-filtered output gap (HPTR), linearly detrended output gap (LTTR), New Keynesian output gap (NKTR), and no output gap in the policy rule (SITR).

to the different dynamics of the economy under the four alternative policy rules.

Similar to the exercise depicted in figure 3, figure 6 illustrates Monte Carlo simulation of POG, HPG, and LTG when only technology shocks hit the economy. In that case, correlation between POG and HPG is negative and ranges between (–0.45, –0.16)[13] and correlation between POG and LTG ranges between (–0.62, –0.26). In more than half of the simulated data, the output gap estimates show an opposite sign to that of the true output gap (in about 60 percent of the cases with the HPG and 65 percent with the LTG).

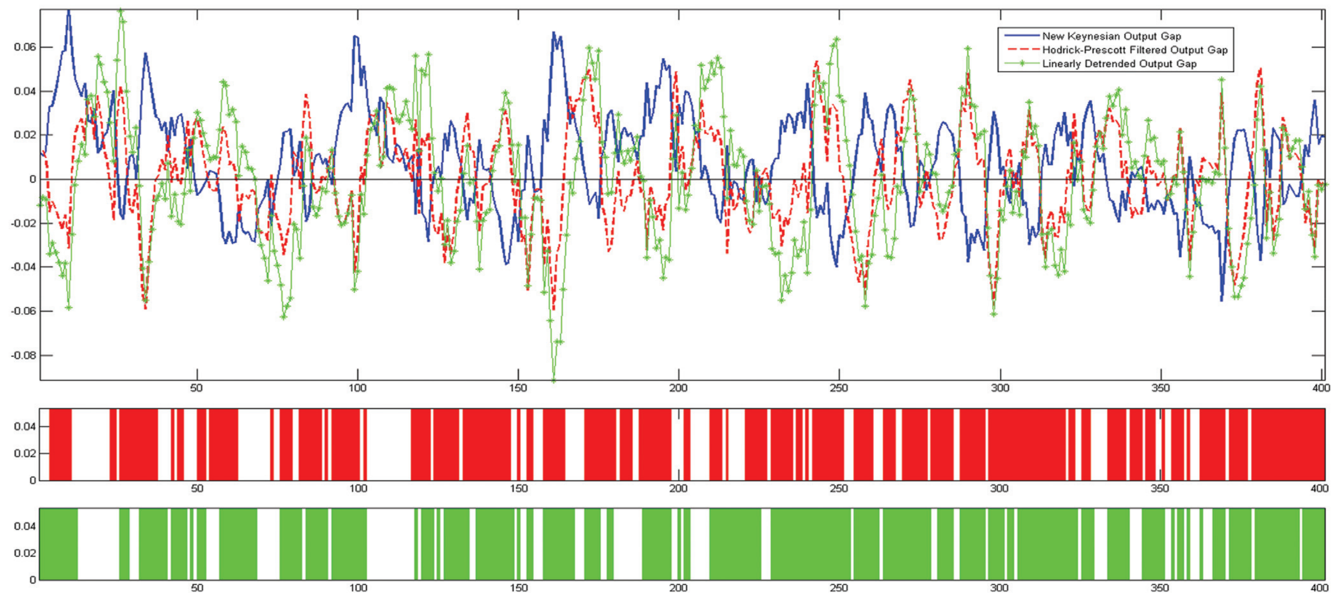
5.4 Demand Shock

A preference shock to consumption in the utility function translates to a shock to the dynamic IS equation—a demand-side shock. The impulse response function to a positive normalized 1 percent demand shock (triggered by ξ) is shown in figure 7. When the CB uses either NKTR or SITR, a positive demand shock leads to higher potential and actual output, although, again, not by the same magnitude. Actual output increases more than potential output (as opposed to the case of a technology shock). Hence, POG increases as well as inflation. The demand shock, similar to the technology shock, does not create a trade-off between stabilizing policy targets, and thus an increase in the nominal interest rate follows. The minor increase in inflation is due to both the small output gap coefficient in the Taylor rule and the small slope of the NKPC.

Note that the expected potential output growth after the first quarter is negative due to the fading demand shock, and hence the potential interest rate becomes negative on impact. As the demand shock fades, the economy converges gradually but rapidly back to steady state. Comparing NKTR to SITR, in the latter, monetary policy is less contractionary, and hence output and POG are higher.[14]

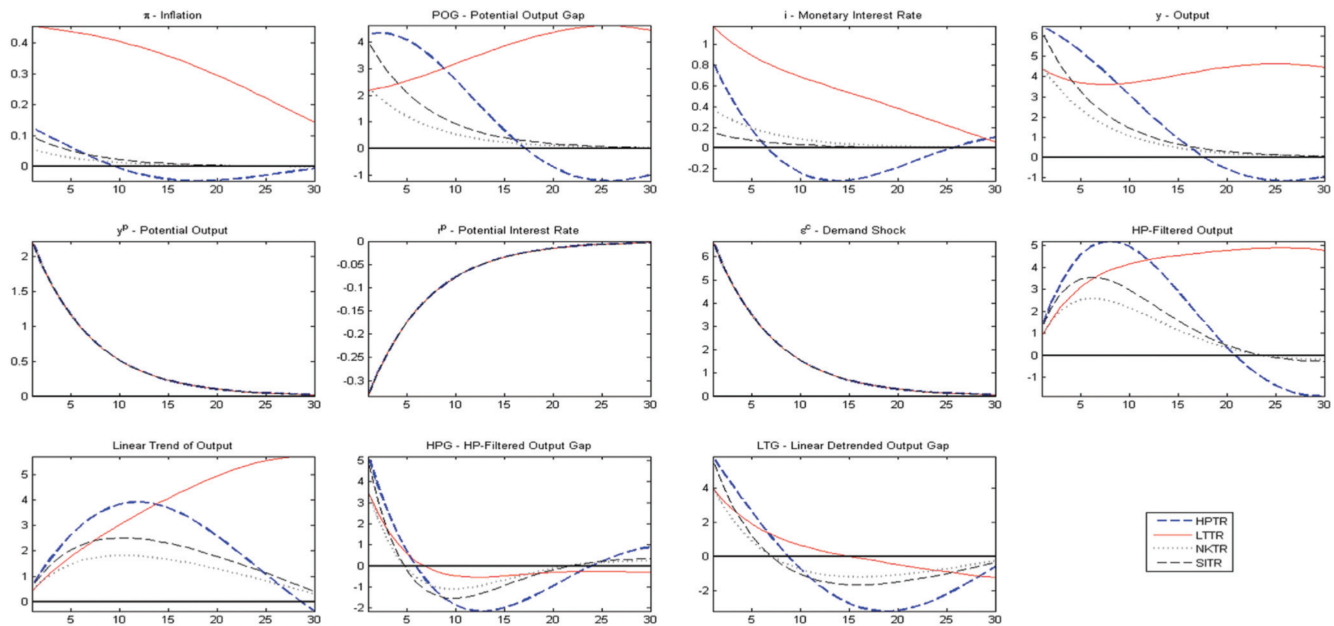
[14]Note that in the LTTR case, the interest rate is highest in the first twenty-nine periods, while output and potential output gap are higher in comparison with the NKTR case—and from the tenth period onward are the highest (figure 7). This may seem puzzling, as one can expect according to the IS equation a lower output gap along with a higher interest rate. The puzzle is resolved by realizing that the IS equation relates the monetary interest rate to the New Keynesian potential output gap, while in the LTTR and HPTR cases the central bank responds to the statistical measure of the output gap.

Figure 6. Simulated Output Gaps Driven by Randomized Technology Shocks Only; CB Response to the "True" New Keynesian Output Gap



Notes: Simulated output gaps driven by randomized technology shocks only when the CB responds to the "true" New Keynesian output gap: New Keynesian output gap, Hodrick-Prescott-filtered output gap, and linearly detrended output gap. The upper bar plot signals when the "true" output gap and the HP-filtered output gap have different signs. The lower bar plot signals when the "true" output gap and the linearly detrended output gap have different signs.

Figure 7. Impulse Response Function to a Positive 1 Percent Normalized Demand Shock—Short Horizon



Notes: Impulse response function to a positive 1 percent normalized demand shock under various output gaps in the policy rule: Hodrick-Prescott-filtered output gap (HPTR), linearly detrended output gap (LTTR), New Keynesian output gap (NKTR), and no output gap in the policy rule (SITR).

Similar to previous results, when the CB responds to a measure of the output gap (using either HPTR or LTTR, figure 8), a demand shock generates long-lasting business cycles. These business cycles are, again, the most pronounced in the LTTR case, and as a result the interest rate, output gap, and potential output gap are the most volatile. When only demand shocks hit the economy (figure 9), if the CB responds to the POG, the correlation between POG and its estimates ranges between (0.30, 056) for the HPG and slightly higher (0.42, 0.72) for the LTG.[13] In about one-third of the cases the estimate of the output gap signals an erroneous direction of inflationary pressures (36 percent and 31 percent in HPG and LTG, respectively).

The analysis above shows that different shocks affect the economy differently. In section 6 we conduct Monte Carlo simulations in which we let all shocks hit the economy simultaneously, as in reality.

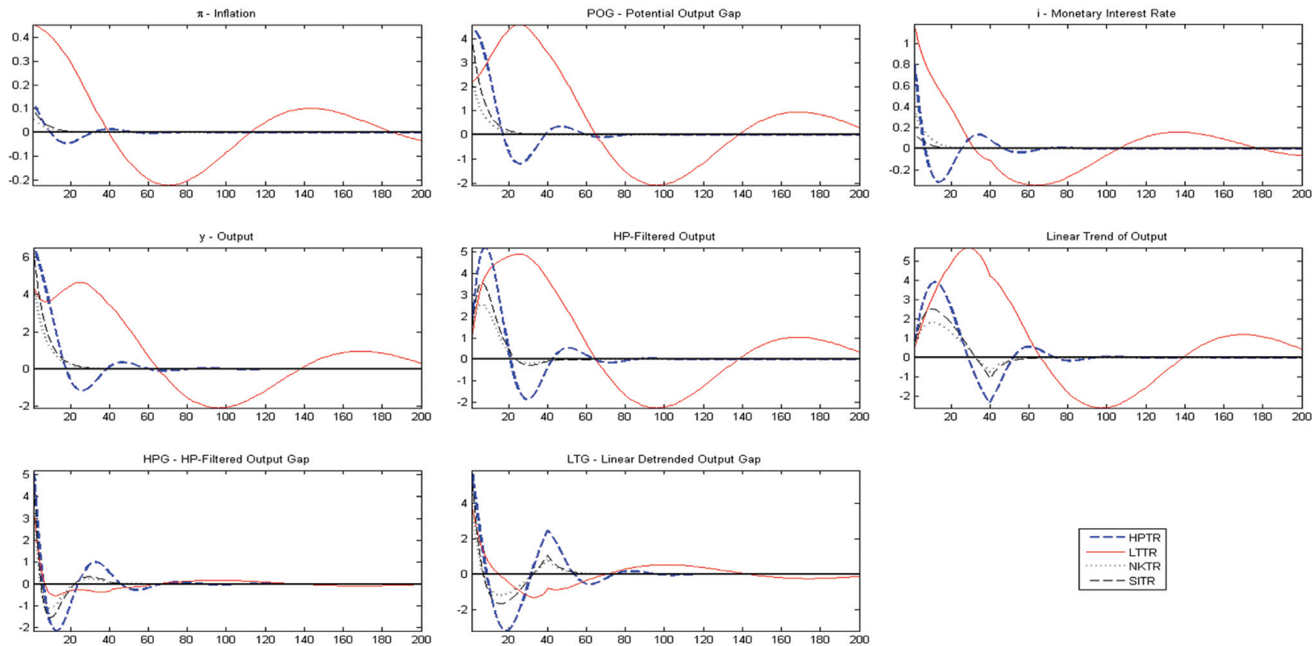
6. Monte Carlo Simulations—All Shocks Simultaneously

We showed that each shock affects the economy differently, and particularly on inflation, output, and output gaps. Hence, when all shocks hit the economy simultaneously, correlations between the output gap and its estimates depend on monetary policy as well as on the composition of the shocks.

Figure 10 illustrates a Monte Carlo simulation of the output gap and its estimates when the CB responds to the true output gap (when using NKTR) and all shocks hit the economy. In this case, the high correlation between the output gap and its estimates when only markup shocks hit the economy is offset by the negative correlation when technology shocks hit the economy (table 2, first row).

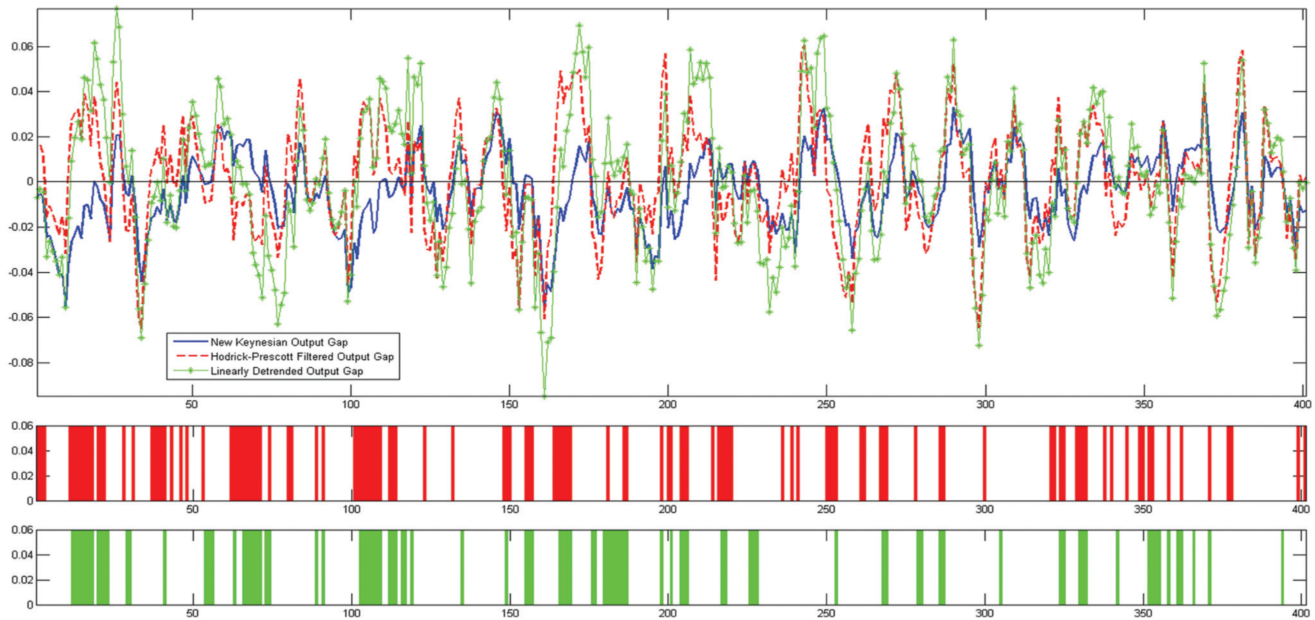
Table 2 shows that the correlations between the POG and its estimates vary under the different policy rules. When the CB uses either NKTR or SITR, the correlation between the output gap and its estimates is about 0.4. When the CB responds to the HPG, the correlation between the output gap and its estimate drops to 0.25, and it becomes negative when the CB responds to the LTG. When policy responds only to inflation, the correlations between the output gap and its estimates are the highest (table 2, last row). The correlation between the two estimates of the output gap is high in all policy rules.

Figure 8. Impulse Response Function to a Positive 1 Percent Normalized Demand Shock—Long Horizon



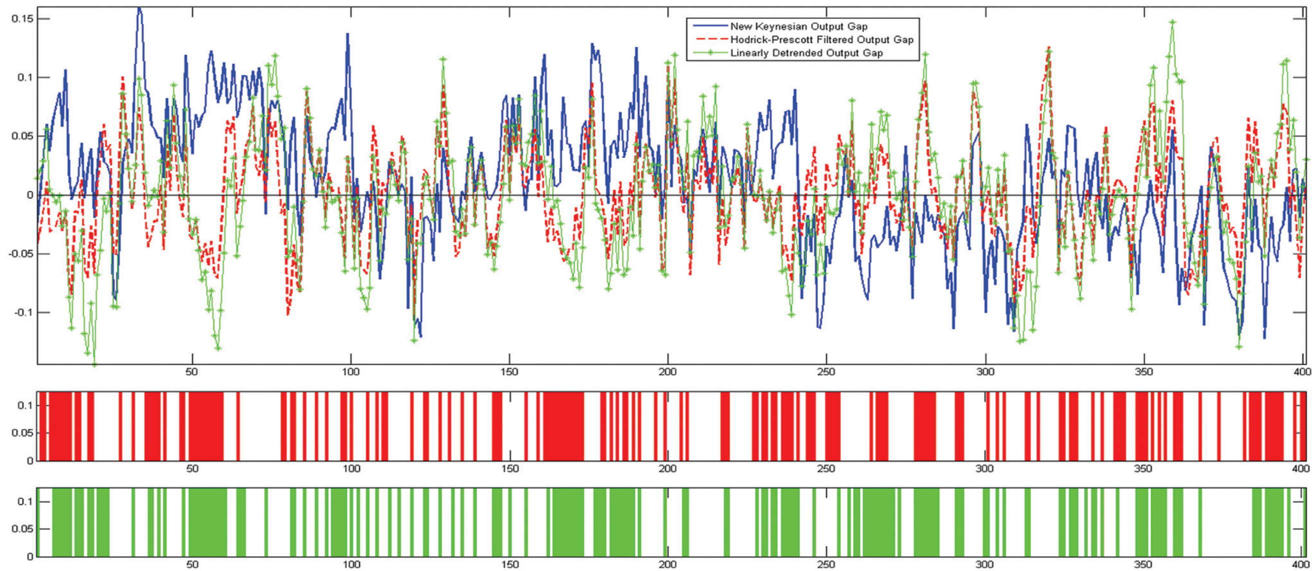
Notes: Impulse response function relating to a positive 1 percent normalized demand shock under various output gaps in the policy rule: Hodrick-Prescott-filtered output gap (HPTR), linearly detrended output gap (LTTR), New Keynesian output gap (NKTR), and no output gap in the policy rule (SITR).

Figure 9. Simulated Output Gaps Driven by Randomized Demand Shocks Only; CB Response to the "True" New Keynesian Output Gap



Notes: Simulated output gaps driven by randomized demand shocks only when the CB responds to the "true" New Keynesian output gap: New Keynesian output gap, Hodrick-Prescott-filtered output gap, and linearly detrended output gap. The upper bar plot signals when the "true" output gap and the HP-filtered output gap have different signs. The lower bar plot signals when the "true" output gap and the linearly detrended output gap have different signs.

Figure 10. Simulated Output Gaps Driven by Randomized Composition of All Shocks Simultaneously; CB Response to the "True" New Keynesian Output Gap



Notes: Simulated output gaps driven by randomized composition of all shocks simultaneously when the CB responds to the "true" New Keynesian output gap: New Keynesian output gap, Hodrick-Prescott-filtered output gap, and linearly detrended output gap. The upper bar plot signals when the "true" output gap and the HP-filtered output gap have different signs. The lower bar plot signals when the "true" output gap and the linearly detrended output gap have different signs.

Table 2. Correlations among Different Output Gaps under Different Taylor Rules in the Presence of All Shocks—Simulated Data

	Corr(POG,HPG)	**Corr(POG,LTG)**	**Corr(HPG,LTG)**
NKTR	0.38	0.33	0.84
HPTR	0.25		0.76
LTTR		−0.25	0.84
SITR	0.40	0.42	0.86

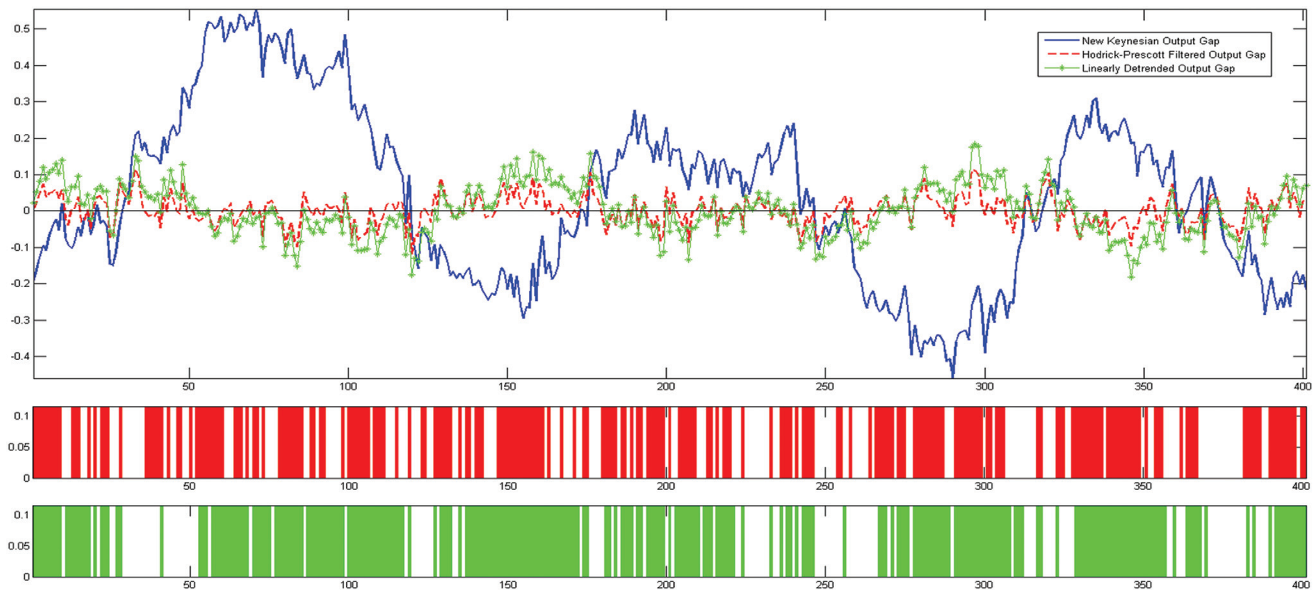
As the lower bar plots of figure 10 imply, both estimates may provide opposite guidance to monetary policy compared with the "true" output gap, as has been shown empirically (Adolfson et al. 2014; Coenen, Smets, and Vetlov 2008). Thus, the traditional univariate estimates of the New Keynesian output gap are poor guidance indicators for monetary policy.

Figure 11 illustrates a Monte Carlo simulation of the output gap and its estimates when all shocks hit the economy simultaneously if the CB responds to the LTG. The simulation shows that a response to the LTG instead of to the output gap (figure 10) creates a significant business cycle. This outcome, namely the significant output gap, stands in interesting contrast with Harvey and Jaeger's (1993) result. While responding to the "true" output gap may be accompanied by *spurious* statistical measures of business cycles (in line with Harvey and Jaeger 1993), when the CB responds either to the HPG or to the LTG it creates *real* significant business cycles, particularly in the latter case.

Table 3 summarizes the implications of responding to each of the different output gaps, as reflected by the opposite signs of the output gap estimates from the output gap (shown in the bar figures) and by the root mean square deviation (RMSD) among them. This is based on simulated data when all shocks hit the economy. When the CB responds to output gap estimates, the gaps between these estimates and the "true" output gap increase dramatically. Furthermore, these estimates provide misleading guidance about half of the time.

The RMSD of the output gap from its estimate is twofold higher when the CB responds to the HPG and is fourfold higher when the CB responds to the LTG. In contrast, when the CB responds either

Figure 11. Simulated Output Gaps Driven by Randomized Composition of All Shocks Simultaneously; CB Response to the Linearly Detrended Output Gap



Notes: Simulated output gaps driven by randomized composition of all shocks simultaneously when the CB responds to the linearly detrended output gap: New Keynesian output gap, Hodrick-Prescott-filtered output gap, and linearly detrended output gap. The upper bar plot signals when the "true" output gap and the HP-filtered output gap have different signs. The lower bar plot signals when the "true" output gap and the linearly detrended output gap have different signs.

Table 3. Monetary Policy Guidance—Simulated Data—All Shocks

	Percentage of Opposite Sign Between:		RMSD[a] Between:	
	POG,HPG	**POG,LTG**	**POG,HPG**	**POG,LTG**
NKTR	37.6%	39.3%	5.1%	6.1%
HPTR	41.9%		9.4%	9.3%
LTTR		58.2%	16.5%	18.5%
SITR	37.0%	36.1%	7.3%	7.5%

[a]Root mean square deviation.

to the output gap as well as inflation (NKTR) or only to inflation (SITR), output gap estimates provide misleading guidance about 37 percent of the time, but as the CB does not respond to these estimates, no harm is done.

7. Welfare Implications

The analysis of the IRFs (section 5) and the simulated data (section 6) suggests that the CB response to output gap estimates causes a welfare loss, which is reflected in long-lasting business cycles. Following Rotemberg and Woodford (1997) and Woodford (2003)'s methodology of using a second-order approximation to the expected utility of the representative household, we use the model-based welfare loss function of Galí (2008, pp. 111–12) to evaluate these losses[15]:

$$W = -\frac{1}{2}(1-\Delta)\frac{\varepsilon}{\lambda}E_0\sum_{t=0}^{\infty}\beta^t[\pi_t^2 + \alpha_1\hat{x}_t^2 - \alpha_2\hat{x}_t] + t.i.p.^{16} \quad (6)$$

[15]Although we assume stochastic demand elasticity in contrast with Galí's (2008), the model-based welfare loss function is the same under the assumption of zero inflation in the steady state.

[16]A derivation of an optimal discretionary policy rule from a loss function in the case of small distortions in the steady state gives rise to the classical inflation bias, e.g., Galí (2008). Instead, we assume that the CB follows a Taylor rule, and thus avoid this bias.

We define the efficient output gap as the deviation of actual from efficient output:

$$x_t \equiv y_t - y_t^e. \tag{7}$$

We follow Blanchard and Galí (2007) and Sala, Söderström, and Trigari (2010) and define efficient output as output that would prevail in perfect competition and a flexible-price economy, that is, output in an economy with no distortions—neither nominal nor real. Hence, the efficient output is the relevant notion for welfare analysis (see appendix 1 for derivation).

The gap between efficient output and potential output is given by

$$y_t^e - y_t^p = \psi_{y\varphi}^p \varphi = \frac{1-\alpha}{\sigma(1-\alpha)+\eta+\alpha}\varphi. \tag{8}$$

When firms have monopolistic power, efficient output is higher than potential output by a constant term, which is a function of the real distortion in the economy resulting from the monopolistic power of the firms. In perfect competition, both potential and efficient outputs equal each other. Using standard calibration, we can quantify the loss of output due to the monopolistic power of the firms. A 20 percent price markup in the steady state implies, using equation (6), a loss of 6.1 percent of output. As $\hat{\varphi}_t \equiv \log(\Phi_t/\Phi)$, it also implies that $\hat{\varphi}_t \in (-0.182, \infty)$—that is, the price markup shock is bounded from below. Note that when firms have monopolistic power, efficient output is also always higher than natural output, as expected. The gap between the two is not constant, and it is proportional to the monopolistic power captured by φ_t (appendix 1).

$\alpha_1 \equiv \kappa/\varepsilon, \alpha_2 \equiv 2\lambda\Delta/\varepsilon$, and *t.i.p.* is terms independent of policy. Δ denotes distortion stemming from monopolistic competition in steady state, and is defined as the distortion that equates the marginal product of labor to the marginal rate of substitution: $-U_n/U_c = MPN(1-\Delta)$, implying $\Delta = \varepsilon^{-1}$. Apart from welfare losses captured by the squared deviations of inflation and efficient output gap (due to price dispersion as prices are staggered and distortions in the allocation of consumption due to relative price distortions), in the case of a distorted steady state, the welfare loss function also includes a negative linear term of the efficient output gap. This term reflects the positive effect of an increase in output on welfare as output is below its efficient level due to the monopolistic power of the firms.

The welfare loss function is expressed in terms of a fraction of steady-state consumption.

7.1 The Basic Calibration—Taylor (1993) Coefficients

Table 4 summarizes the welfare implications of responding to the different output gaps, as expressed by the loss function (equation (6)), using the coefficients in the original Taylor rule (Taylor 1993) $\mu_\pi = 1.5$ and $\mu_x = 0.5/4$. Following Orphanides et al. (2000), we also examine the welfare implications of responding to the last four quarters' output growth—GTR (table 4, last row), rather than to the output gap or one of its measures.

When the CB responds to the LTG, the welfare loss is about 20 percent higher than the loss under the other alternatives. However, the loss function does not capture the further welfare loss which occurs when the CB responds to the HPG. The reason is the minor weight of the efficient output gap in the loss function stemming from the small slope of the Phillips curve—the inflation elasticity with respect to output gap. However, the variance of output is doubled under the HPTR, and almost doubled under the GTR compared with the NKTR and the SITR alternatives—reflecting the long-duration business cycles created under the HPTR and the GTR. When the CB responds to the LTG, the output variance is six times higher than in the case of using either NKTR or SITR.

Table 4. Welfare Implications of Using Different Taylor Rules—Welfare Loss in Terms of Discounted Fraction of Steady-State Consumption (average annual loss, percent)

	Welfare Loss	**Loss ($\hat{x}$)**	**Loss (π)**	**Var(y)**	**Var(π)**
NKTR	−0.976%	−0.003%	−0.973%	0.42%	0.05%
HPTR	−0.986%	−0.010%	−0.976%	0.80%	0.05%
LTTR	−1.187%	−0.032%	−1.155%	2.41%	0.06%
SITR	−0.984%	−0.007%	−0.977%	0.45%	0.05%
GTR	−0.973%	−0.010%	−0.963%	0.72%	0.05%

Table 5. Variance of the Different Output Gaps under Different Monetary Rules (percent)

	VAR (POG)	VAR (HPG)	VAR (LTG)
NKTR	0.22%	0.19%	0.32%
HPTR	0.82%	0.34%	
LTTR	2.61%		0.34%
SITR	0.60%	0.23%	0.37%
GTR	0.79%	0.26%	0.54%

While GTR yields the lowest welfare loss—although quite similar to the NKTR and slightly less than the SITR—the variance of output is almost doubled under the GTR compared with the NKTR and the SITR alternatives, and only slightly below that under the HPTR.

While the volatility of the inflation rate is not sensitive to monetary policy (table 4), the volatilities of the New Keynesian output gap and its estimates do depend on monetary policy (table 5). If the CB responds to the New Keynesian output gap, volatilities are about the same. However, when the CB responds to an output gap estimate, the volatility of the New Keynesian output gap rises dramatically, while the volatility of the estimate is slightly higher than in the case when the CB responds to the New Keynesian output gap.

When the CB responds only to inflation, efficient output gap volatility is about three times higher than when it responds to the efficient output gap; it is about four times higher when it responds to the HPG or to output growth; and it is about twelve times higher when it responds to the LTG. Note that the volatility of the LTG estimate of the output gap is similar among the different monetary policy rules, but under GTR it is almost doubled.

7.2 Optimal Output Gap Coefficient in the Taylor Rule

In this subsection, we follow the welfare analysis conducted in Orphanides et al. (2000). Specifically, we hold the coefficient of the inflation rate fixed ($\mu_\pi = 1.5$) and use a grid search for $\mu_x = \{[0, \ldots, 0.25] \; by \; 0.025\}$, which minimizes the welfare loss function (equation (6)). As above, we simulate 400,000 periods and

Table 6. Optimal Output Gap Coefficient and the Corresponding Welfare Loss in Terms of Discounted Fraction of Steady-State Consumption (average annual loss, percent)

	Optimal μ_x	Welfare Loss	Loss ($\hat{x}$)	Loss (π)	Var(y)	Var(π)
NKTR	0.2–0.25	−0.975%	−0.002%	−0.973%	0.42%	0.050%
HPTR	0.050	−0.983%	−0.008%	−0.975%	0.53%	0.051%
LTTR	0.025	−0.982%	−0.008%	−0.974%	0.53%	0.051%
SITR	0	−0.984%	−0.007%	−0.977%	0.44%	0.050%
GTR	0.2–0.225	−0.968%	−0.013%	−0.955%	0.93%	0.049%

calculate the mean of the 1,000 consecutive samples of 400 periods (each reflecting a sample of 100 years). Table 6 summarizes the results.

Table 6 shows that each of the various optimal responses of the CB to the output gap or to its measures leads to similar welfare losses. The optimal output gap coefficients reflect attenuation—that is, the optimal coefficients of the output gap measures are only slightly above zero, similar to Orphanides et al. (2000) findings and as is observed empirically by Smets (2002). In contrast, the optimal coefficient under NKTR, as well as under GTR, is in the range of 0.2–0.25, similar to the updated Taylor rule (Taylor 1999). However, if the CB responds to output growth, the variance of output is more than doubled compared with the NKTR and the SITR alternatives. Hence, all in all, it is preferable that the central bank responds only to inflation.

8. Summary and Conclusion

The heart of the research question in this paper is the identification of the unobserved output gap using traditional statistical filters and the implications of using such measures. We use Monte Carlo simulations of the canonical New Keynesian model to analyze the positive and the welfare implications of responding to either a Hodrick-Prescott-filtered output gap or a linearly detrended output gap—instead of to the model-based New Keynesian output gap.

We find that responding to output gap estimates creates long-lasting business cycles as well as welfare loss. These business cycles are created due to the fact that both the Hodrick-Prescott-filtered output and the linearly detrended output measure potential output by smoothing the output series regardless of the source of the shock hitting the economy and hence fail to identify the "true" output gap. The particular combination of the various shocks, as implied by the standard deviation and inertia of each one, determines the implications of responding to a statistical measure of the output gap on the economy. Moreover, following a technology shock, these estimators provide misleading guidance to monetary policy, a result which is also present in the cases of price shocks (cost-push shocks) and demand shocks, although to a smaller extent.

The long duration of the business cycles is explained by the fact that the estimates are derived from historical output data, which introduce inertia into the estimated output gap series. The implications of responding to a linearly detrended output gap are found to be the most pronounced due to the smoother weights attached to historical output, compared with the weights in the Hodrick-Prescott filter.

As the output gap is unobserved, we find that it is preferable that the central bank responds only to inflation and not to the output gap estimates.

Appendix 1. The Model's Derivation

We rely heavily on the methodology of Galí (2008) to derive the model's equations.

The Households Problem

A representative infinitely lived household maximizes its expected discounted lifetime utility:

$$E_0 \sum_{t=0}^{\infty} \beta^t U(C_t, N_t), \tag{9}$$

where the constant relative risk aversion (CRRA) utility function is given by

$$U(C_t, N_t) = \Xi_t \frac{C_t^{1-\sigma}}{1-\sigma} - \frac{N_t^{1+\eta}}{1+\eta},$$

and where C_t is an elasticity of substitution consumption index of a continuum goods $i \in (0, 1]$, given by

$$C_t \equiv \left(\int_0^1 C_t(i)^{\frac{1}{\Phi_t}} di \right)^{\Phi_t}. \tag{10}$$

$C_t(i)$ denotes the quantity of good i consumed by the household, N_t denotes its labor supply, and $\Phi_t > 1$ is inversely related to the elasticity of substitution among goods—$\varepsilon_t \equiv \Phi_t/(\Phi_t - 1)$. Ξ_t is a preference shock to consumption (e.g., Cristoffel, Coenen, and Warne 2008 and Smets and Wouters 2003a).

The household budget constraint is given by

$$\int_0^1 P_t(i)C_t(i)di + Q_t B_t \leq B_{t-1} + W_t N_t + TAX_t, \tag{11}$$

where $P_t(i)$ is the price of good i. B_t denotes one-period-ahead bond units purchased by the household in the price of Q_t per bond, W_t is nominal wage, and TAX_t denotes lump-sum taxes. The solution of the optimal consumption allocation for given expenditures yields the demand function:

$$C_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\frac{\Phi_t}{\Phi_t - 1}} C_t, \tag{12}$$

where aggregate price index is given by $P_t \equiv \left(P_t(i)^{\frac{1}{1-\Phi_t}} di \right)^{1-\Phi_t}$.

Hence, $\int_0^1 P_t(i)C_t(i)di = P_t C_t$, implying that the budget constraint can be written as $P_t C_t + Q_t B_t \leq B_{t-1} + W_t N_t + TAX_t$.

The first-order conditions with regard to consumption, labor, and bonds yield the standard intratemporal optimal condition, the optimal labor supply condition which equates real wage to marginal rate of substitution (MRS),

$$\frac{W_t}{P_t} = \frac{C_t^{\sigma} N_t^{\eta}}{\Xi_t}, \tag{13}$$

and the intertemporal condition, the Euler, which is given by

$$Q_t = \beta E_t \left\{ \left(\frac{\Xi_{t+1}}{\Xi_t} \right) \left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} \left(\frac{P_t}{P_{t+1}} \right) \right\}. \tag{14}$$

Log-linearizing the first-order conditions—(13) and (14)—yields

$$w_t - p_t = \sigma c_t + \eta n_t - \xi_t \tag{15}$$

and

$$c_t = E_t(c_{t+1}) - \frac{1}{\sigma}(i_t - E_t(\pi_{t+1}) - \rho) + \frac{1}{\sigma}(1 - \rho^{\xi})\xi_t, \tag{16}$$

where $i_t \equiv -\log Q_t$ is the nominal interest rate, $\rho \equiv -\log \beta$ is the time discount rate, $\pi_t \equiv p_t - p_{t-1}$ is the CPI inflation rate, and ρ^{ξ} is the preference shock to consumption persistence. $\xi_t \equiv \log(\Xi_t)$ denotes the preference shock, which follows an AR(1) process $\xi_t = \rho^{\xi}\xi_{t-1} + \nu_t^{\xi}$, where ν_t^{ξ} is white noise.

As the model is a closed-economy model without investment and government, output, y_t, equals consumption. Hence, the consumer's Euler equation (16) yields the equilibrium condition:

$$y_t = E_t(y_{t+1}) - \frac{1}{\sigma}(i_t - E_t(\pi_{t+1}) - \rho) + \frac{1}{\sigma}(1 - \rho^{\xi})\xi_t. \tag{17}$$

Technology

Each firm $i \in (0, 1]$ uses the production function

$$Y_t(i) = A_t N_t(i)^{1-\alpha} \tag{18}$$

to produce its differentiated good, where $N_t(i)$ denotes labor in firm i and A_t is the productivity level affecting all firms. Productivity evolves according to the AR(1) process $a_t = \rho^a a_{t-1} + v_t^a$, where $a_t \equiv \log(A_t)$ and v_t^a is white noise.

Derivation of the New Keynesian Phillips Curve

The optimal price P_t^* is set to maximize the discounted flow of profits:

$$\underset{P_t^*}{Max} \sum_{k=0}^{\infty} \theta^k E_t \{ Q_{t,t+k} (P_t^* Y_{t+k|t} - \Psi_{t+k}(Y_{t+k|t}) \}$$

subject to the demand constraint

$$Y_{t+k|t} = \left(\frac{P_t^*}{P_{t+k}} \right)^{-\frac{\Phi_{t+k}}{\Phi_{t+k}-1}} Y_{t+k}. \tag{19}$$

$Y_{t+k|t}$ is output in period $t+k$ for a firm that last optimized its price in period t, and $Q_{t,t+k} \equiv \beta^k (C_{t+k}/C_t)^{-\sigma} (P_t/P_{t+k})$ is the stochastic discount factor for nominal payoffs, derived from iterating the Euler equation (16) forward. $\Psi(\cdot)$ is the cost function of the firm.

The first-order condition of the firm's problem is given by

$$\sum_{k=0}^{\infty} \theta^k E_t \left\{ Q_{t,t+k} \left[Y_{t+k|t} + P_t^* \frac{\partial Y_{t+k|t}}{\partial P_t^*} - \psi_{t+k|t} \frac{\partial Y_{t+k|t}}{\partial P_t^*} \right] \right\} = 0, \tag{20}$$

where $\psi_{t+k|t} \equiv \Psi'_{t+k}(Y_{t+k|t})$ is the nominal marginal cost in period $t+k$.

As $\frac{\partial Y_{t+k|t}}{\partial P_t^*} = -\frac{\Phi_{t+k}}{\Phi_{t+k}-1} \frac{1}{P_t^*} Y_{t+k|t}$, we get

$$\sum_{k=0}^{\infty} \theta^k E_t \left\{ Q_{t,t+k} Y_{t+k|t} \left[\left(1 - \frac{\Phi_{t+k}}{\Phi_{t+k}-1} \right) + \frac{1}{P_t^*} \frac{\Phi_{t+k}}{\Phi_{t+k}-1} \psi_{t+k|t} \right) \right] \right\} = 0. \tag{21}$$

Multiplying by P_t^* and dividing by $1 - \frac{\Phi_{t+k}}{\Phi_{t+k}-1}$, we can express this as

$$\sum_{k=0}^{\infty} \theta^k E_t \{ Q_{t,t+k} Y_{t+k|t} (P_t^* - \Phi_{t+k} \psi_{t+k|t}) \} = 0, \tag{22}$$

where $\Phi_{t+k} = \frac{\varepsilon_{t+k}}{\varepsilon_{t+k}-1}$ is interpreted as the desired markup charged by the firm; when demand elasticity decreases, the monopolists can charge a higher markup, which reflects a higher market power of the firm.

Note that when prices are flexible, the first-order condition in (22) implies that the optimal price is set as a desired markup over the firm's nominal marginal cost.

The first-order condition (22) can be written as

$$\sum_{k=0}^{\infty} \theta^k E_t \left\{ Q_{t,t+k} Y_{t+k|t} \left(\frac{P_t^*}{P_{t-1}} - \Phi_{t+k} MC_{t+k|t} \Pi_{t-1,t+k} \right) \right\} = 0, \tag{23}$$

where $\Pi_{t,t+k} \equiv P_{t+k}/P_t$ and $MC_{t+k|t} \equiv \psi_{t+k|t}/P_{t+k}$ is the real marginal cost in period $t+k$ for a firm which set its price in period t.

In a zero-inflation steady state, $P_t/P_{t-1} = 1$ and $\Pi_{t-1,t+k} = 1$. Hence $Y_{t+k|t} = Y$ and $MC_{t+k|t} \equiv MC$, as all firms produce the same quantity of output in the steady state, and $Q_{t,t+k} = \beta^k$. Accordingly, $MC = 1/\Phi$. Thus, a first-order Taylor expansion of (23) around the zero-inflation steady state yields

$$p_t^* - p_{t-1} = (1-\beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k E_t\{\hat{mc}_{t+k|t} + (p_{t+k} - p_{t-1}) + \hat{\varphi}_{t+k}\}, \tag{24}$$

where $\hat{mc}_{t+k|t} \equiv mc_{t+k|t} - mc$ is the log-deviation of real marginal cost from its steady-state value $mc = -\varphi$, where $\varphi \equiv \log(\Phi)$. Following Galí (2008), we assume an exogenous AR(1) process for $\hat{\varphi}_t$, $\hat{\varphi}_t = \rho^{\varphi}\hat{\varphi}_{t-1} + v_t^{\varphi}$, where v_t^{φ} is white noise.

The economy's real marginal cost is defined as the ratio of real wage to marginal product of labor:

$$MC_t \equiv \frac{1}{P_t}\frac{\partial TC_t}{\partial Y_t} = \frac{1}{P_t}\frac{\partial TC_t}{\partial Y_t}\frac{\partial N_t}{\partial N_t} = \frac{1}{P_t}\frac{\partial TC_t}{\partial N_t} \Bigg/ \frac{\partial Y_t}{\partial N_t} = \frac{W_t}{P_t} \Bigg/ MPN_t. \tag{25}$$

Using the production function (18), the economy's average real marginal cost is

$$mc_t = (w_t - p_t) - (1-\alpha)^{-1}(a_t - \alpha y_t) - \log(1-\alpha). \tag{26}$$

Hence, $mc_{t+k|t} = mc_{t+k} + \frac{\alpha}{1-\alpha}(y_{t+k|t} - y_{t+k})$.

Next, using the demand function (19) and market clearing condition,

$$\hat{mc}_{t+k|t} = \hat{mc}_{t+k} - \frac{\alpha}{1-\alpha}\frac{\varphi}{\varphi-1}(p_t^* - p_{t+k}). \qquad (27)$$

Substituting (27) into (24) yields

$$p_t^* - p_{t-1} = \beta\theta E_t\{p_{t+1}^* - p_t\} + (1-\beta\theta)\Theta(\hat{mc}_t + \hat{\varphi}_t) + \pi_t, \qquad (28)$$

where $\Theta \equiv \frac{\varphi-1+\alpha-\alpha\varphi}{\varphi-1+\alpha} \leq 1$.

Log-linearizing the aggregate price index around the zero steady-state inflation implies that $\pi_t = (1-\theta)(p_t^* - p_{t-1})$, which together with (28) yields the Phillips curve in terms of log-deviation of marginal cost and the desired markup from steady state:

$$\pi_t = \beta E_t\{\pi_{t+1}\} + \lambda(\hat{mc}_t + \hat{\varphi}_t), \qquad (29)$$

where $\lambda \equiv \frac{(1-\beta)(1-\beta\theta)}{\theta}\Theta$.

Potential Output and Potential Real Interest Rate

The average real marginal cost, the gap between the real wage and the marginal product of labor (26), along with the optimal labor supply condition (15) and the production function (18), yield the next relation between the economy's real marginal cost and aggregate activity:

$$mc_t = \left(\sigma + \frac{\eta+\alpha}{1-\alpha}\right)y_t - \frac{1+\eta}{1-\alpha}a_t - \log(1-\alpha) - \xi_t. \qquad (30)$$

Using this expression, and defining potential output, y_t^p, as output which would prevail in a flexible-price economy when price markup shocks are held constant at steady-state values,[17] real marginal cost is given by $mc_t = -\varphi$.[18] Hence, potential output satisfies the equation

$$-\varphi = \left(\sigma + \frac{\eta+\alpha}{1-\alpha}\right)y_t^p - \frac{1+\eta}{1-\alpha}a_t - \log(1-\alpha) - \xi_t. \qquad (31)$$

[17]Following Adolfson et al. (2014), Sala, Söderström, and Trigari (2010), Smets and Wouters (2003a, 2003b), and Woodford (2003).

[18]This relation is derived from equation (22) when prices are flexible.

Thus, potential output is given by

$$y_t^p = \psi_{ya}^p a_t - \psi_{y\varphi}^p \varphi + \psi_{y\varphi}^p \xi_t + \vartheta_y^p, \tag{32}$$

where $\psi_{ya}^p \equiv \frac{1+\eta}{\sigma(1-\alpha)+\eta+\alpha}$, $\psi_{y\varphi}^p \equiv \frac{1-\alpha}{\sigma(1-\alpha)+\eta+\alpha}$, and $\vartheta_y^p \equiv \frac{(1-\alpha)\log(1-\alpha)}{\sigma(1-\alpha)+\eta+\alpha}$.

Using (30) and (31), the log-deviation of the economy's real marginal cost from steady state of potential economy, $\hat{mc}_t$, is given by

$$\hat{mc}_t = \left(\sigma + \frac{\eta+\alpha}{1-\alpha}\right)(y_t - y_t^p). \tag{33}$$

Denoting the potential output gap as $\tilde{y}_t \equiv y_t - y_t^p$ and plugging (33) in (29) yields the NK Phillips curve (NKPC) in terms of the potential output gap (equation (1)), where $\kappa \equiv \lambda(\sigma + \frac{\eta+\alpha}{1-\alpha})$.

Substituting the potential output gap expression into the consumer's Euler equation yields the dynamic IS equation in terms of the potential output gap (equation (2)), where the potential real interest rate is given by

$$r_t^p \equiv \rho + \sigma E_t\{\Delta y_{t+1}^p\}. \tag{34}$$

Natural Output

Using (30), and as under flexible prices real marginal cost is given by $mc_t = -\varphi_t$,[18] the flexible-price output, natural output, y_t^f, satisfies

$$-\varphi_t = \left(\sigma + \frac{\eta+\alpha}{1-\alpha}\right) y_t^f - \frac{1+\eta}{1-\alpha} a_t - \log(1-\alpha) - \xi_t. \tag{35}$$

Thus, natural output is given by

$$y_t^f = \psi_{ya}^p a_t - \psi_{y\varphi}^p \varphi_t + \psi_{y\varphi}^p \varepsilon_t^c + \vartheta_y^p, \tag{36}$$

where $\psi_{ya}^p \equiv \frac{1+\eta}{\sigma(1-\alpha)+\eta+\alpha}, \psi_{y\varphi}^p \equiv \frac{1-\alpha}{\sigma(1-\alpha)+\eta+\alpha}$, and $\vartheta_y^p \equiv \frac{(1-\alpha)\log(1-\alpha)}{\sigma(1-\alpha)+\eta+\alpha}$.

Efficient Output and Efficient Real Interest Rate

We follow Blanchard and Galí (2007) and Sala, Söderström, and Trigari (2010) and define efficient output as output that would prevail in perfect competition and a flexible-price economy, that is, output in an economy with no distortions—neither nominal nor real. Hence, the efficient output is the relevant notion for welfare analysis. The efficient output is derived as follows.

The maximization problem of a competitive firm is given by

$$M\underset{Y_t,N_t}{A}X\ \ \{P_tY_t - W_tN_t\},$$

subject to the production function (18).

The standard first-order condition implies that the firm sets its production such that real wage equals marginal product of labor:

$$\frac{W_t}{P_t} = (1-\alpha)A_tN_t^{-\alpha} \Rightarrow w_t - p_t = \log(1-\alpha) + a_t - \alpha n_t. \quad (37)$$

Combining (37) and (15), and using aggregate production function to express labor in terms of output and technology, the efficient output is given by (in log terms)

$$y_t^e = \psi_{ya}^p a_t + \psi_{y\varphi}^p \xi_t + \vartheta_y^p. \quad (38)$$

Note that the gap between efficient output and potential output is given by

$$y_t^e - y_t^p = \psi_{y\varphi}^p \varphi = \frac{1-\alpha}{\sigma(1-\alpha)+\eta+\alpha}\varphi. \quad (39)$$

As $\psi_{y\varphi}^p > 0$ and $\varphi \equiv \log(\Phi) > 0$ where $\Phi > 1$, when firms have monopolistic power, efficient output is higher than potential output by a constant term, which is a function of the real distortion in the economy resulting from the monopolistic power of the firms. In perfect competition, both potential and efficient outputs equal each other.

Next, we will find the relation between the potential output gap and the efficient output gap. Note that

$$\tilde{y}_t \equiv y_t - y_t^p. \quad (40)$$

As output equals potential output in the steady state, we get

$$\tilde{y}_t = y_t - y_t^e + y_t^e - y_t^p = (y_t - y_t^e) - (y - y^e) + (y_t^e - y_t^p) - (y^e - y^p). \tag{41}$$

Defining the efficient output gap as

$$x_t \equiv y_t - y_t^e, \tag{42}$$

and as the gap between efficient and potential output is constant (39), the last two terms in (41) cancel each other and $\tilde{y}_t = \hat{x}_t$. Hence, the NKPC in terms of log-deviation of the efficient output gap from its steady state is given by

$$\pi_t = \beta E_t\{\pi_{t+1}\} + \kappa \hat{x}_t + \lambda \hat{\varphi}_t. \tag{43}$$

Using the consumer's Euler equation (17), the dynamic IS equation in terms of log-deviation of efficient output gap from its steady state is given by

$$\hat{x}_t = E_t(\hat{x}_{t+1}) - \frac{1}{\sigma}(i_t - E_t(\pi_{t+1}) - r_t^e) + \frac{1}{\sigma}(1 - \rho^c)\xi_t, \tag{44}$$

where the efficient real interest rate is given by

$$r_t^e \equiv \rho + \sigma E_t\{\Delta y_{t+1}^e\} = \rho + \sigma \psi_{ya}^p E_t\{\Delta a_{t+1}\} + \sigma \psi_{y\varphi}^p E_t\{\Delta \xi_{t+1}\}. \tag{45}$$

Hence, the gap between efficient output and flexible-price output is

$$y_t^e - y_t^f = \psi_{y\varphi}^p \varphi_t = \frac{1-\alpha}{\sigma(1-\alpha) + \eta + \alpha}\varphi_t. \tag{46}$$

Appendix 2. The Model's Log-linear Main Equations

NKPC – Inflation Equation:

$$\pi_t = \beta E_t\{\pi_{t+1}\} + \kappa \hat{x}_t + \lambda \hat{\varphi}_t. \tag{47}$$

Dynamic IS Equation:

$$\hat{x}_t = E_t(\hat{x}_{t+1}) - \frac{1}{\sigma}(i_t - E_t(\pi_{t+1}) - r_t^e) + \frac{1}{\sigma}(1 - \rho^\xi)\xi_t. \tag{48}$$

Monetary Policy Taylor Rule:

$$i_t = \rho + \mu_\pi \cdot \pi_t + \mu_x \cdot gap_t. \tag{49}$$

Potential Output:

$$y_t^p = \psi_{ya}^p a_t - \psi_{y\varphi}^p \varphi + \psi_{y\varphi}^p \xi_t + \vartheta_y^p, \tag{50}$$

where $\psi_{ya}^p \equiv \frac{1+\eta}{\sigma(1-\alpha)+\eta+\alpha}$, $\psi_{y\varphi}^p \equiv \frac{1-\alpha}{\sigma(1-\alpha)+\eta+\alpha}$, and $\vartheta_y^p \equiv \frac{(1-\alpha)\log(1-\alpha)}{\sigma(1-\alpha)+\eta+\alpha}$.

Potential Interest Rate:

$$r_t^p \equiv \rho + \sigma E_t\{\Delta y_{t+1}^p\} = \rho + \sigma \psi_{ya}^p E_t\{\Delta a_{t+1}\}. \tag{51}$$

Efficient Output:

$$y_t^e = \psi_{ya}^p a_t + \psi_{y\varphi}^p \xi_t + \vartheta_y^p. \tag{52}$$

Efficient Interest Rate:

$$r_t^e \equiv \rho + \sigma E_t\{\Delta y_{t+1}^e\} = \rho + \sigma \psi_{ya}^p E_t\{\Delta a_{t+1}\} + \sigma \psi_{y\varphi}^p E_t\{\Delta \xi_{t+1}\}. \tag{53}$$

Natural Output – Flexible-Price Output:

$$y_t^f = \psi_{ya}^p a_t - \psi_{y\varphi}^p \varphi_t + \psi_{y\varphi}^p \xi_t + \vartheta_y^p. \tag{54}$$

Appendix 3. Calculation of Model-Based Hodrick-Prescott-Filtered Output and Linearly Detrended Output which Enter the Taylor Rules

Hodrick-Prescott-Filtered Output Trend

The Hodrick-Prescott filter (HP filter) derives a smoothed series, τ_t, of an s_t series by minimizing the sum of the squared difference between the two series and a penalty term related to the smoothed degree of the smoothed series. Formally,

$$\underset{s_t}{Min} \sum_{t=1}^{T} (s_t - \tau_t)^2 + \lambda^{HP} \sum_{t=2}^{T-1} ((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2,$$

where λ^{HP} sets the smoothness degree. For quarterly data, λ^{HP} is usually set to 1,600.

First-order conditions can be expressed in matrix form as

$$\tau = (\lambda^{HP} F + I)^{-1} s,$$

where I is a $T \times T$ identity matrix and F is a $T \times T$ matrix given by

$$F = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & & & & & & & & & & 0 \\ -2 & 5 & -4 & 1 & 0 & \dots & & & & & & & & & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & \dots & & & & & & & & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \dots & & & & & & & 0 \\ 0 & 0 & 1 & - & 4 & 6 & -4 & 1 & 0 & \dots & & & & & 0 \\ \vdots & & & & & & & & & \ddots & & & & & \\ & & & & & & & & & & \ddots & & & & \\ 0 & \dots & & & & & & & \dots & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \dots & & & & & & & & \dots & 0 & 1 & -4 & 5 & -2 \\ 0 & \dots & & & & & & & & & \dots & 0 & 1 & -2 & 1 \end{bmatrix}$$

The Taylor rule used in the model responds to the endpoint output gap estimate. Therefore, the output HP-filtered trend which constructs the output gap the CB responds to is calculated as

$$\tau_T = (last\ \ line\ \ of\ \ matrix\ \ H) \cdot s, \text{ where } H = (\lambda^{HP} F + I)^{-1}.$$

We calculate the trend using last forty observations (ten years) of simulated output.

Linear Output Trend

The linear trend of output is calculated by regressing the output series on a constant and a time trend index, that is, $\chi_i = (Z'Z)^{-1} Z'y$, where $Z' \equiv \begin{bmatrix} 1 & 1 & \dots & & 1 \\ 0 & 1 & 2 & \dots & T-1 \end{bmatrix}$, and the linear trend is given by $\chi_i \cdot Z'$. Alternatively, the linear trend of output can also be calculated as a special case of the HP-filtered trend, by setting $\lambda^{HP} = \infty$. For practicality, we set $\lambda^{HP} = 10^9$.

References

Adolfson, M., S. Laseen, J. Linde, and L. E. Svensson. 2014. "Monetary Policy Trade-Offs in an Estimated Open-Economy DSGE Model." *Journal of Economic Dynamics and Control* 42: 33–49.

Aoki, K. 2003. "On the Optimal Monetary Policy Response to Noisy Indicators." *Journal of Monetary Economics* 50 (3): 501–23.

Argov, E., A. Barnea, A. Binyamini, E. Borenstein, D. Elkayam, and I. Rozenshtrom. 2012. "MOISE: A DSGE Model for the Israeli Economy." Discussion Paper No. 2012.06, Bank of Israel.

Baxter, M., and R. G. King. 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." *Review of Economics and Statistics* 81 (4): 575–93.

Blanchard, O., and J. Galí. 2007. "Real Wage Rigidities and the New Keynesian Model." *Journal of Money, Credit and Banking* 39 (s1): 35–65.

Christiano, L. J., and T. J. Fitzgerald. 2003. "The Band Pass Filter." *International Economic Review* 44 (2): 435–65.

Christoffel, K., G. Coenen, and A. Warne. 2008. "The New Area-Wide Model of the Euro Area: A Micro-Founded Open-Economy Model for Forecasting and Policy Analysis." ECB Working Paper No. 944.

Clarida, R., J. Galí, and M. Gertler. 1999. "The Science of Monetary Policy: A New Keynesian Perspective." *Journal of Economic Literature* 37 (4): 1661–1707.

Coenen, G., F. Smets, and I. Vetlov. 2008. "Estimation of Euro Area Output Gap Using the NAWM." Working Paper No. 5/2009, Bank of Lithuania.

Cúrdia, V., A. Ferrero, G. C. Ng, and A. Tambalotti. 2015. "Has U.S. Monetary Policy Tracked the Efficient Interest Rate?" *Journal of Monetary Economics* 70: 72–83.

Galí, J. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework.* Princeton, NJ: Princeton University Press.

Harvey, A. C., and A. Jaeger. 1993. "Detrending, Stylized Facts and the Business Cycle." *Journal of Applied Econometrics* 8 (3): 231–47.

Justiniano, A., and G. E. Primiceri. 2008. "Potential and Natural Output." Unpublished manuscript, Federal Reserve Bank of Chicago.

McCallum, B. T. 2001. "Should Monetary Policy Respond Strongly to Output Gaps?" *American Economic Review: Papers and Proceedings* 91 (2): 258–62.

Neiss, K. S., and E. Nelson. 2005. "Inflation Dynamics, Marginal Costs, and the Output Gap: Evidence from Three Countries." *Journal of Money, Credit and Banking* 37 (6, December): 1019–45.

Orphanides, A. 2003. "Monetary Policy Evaluation with Noisy Information." *Journal of Monetary Economics* 50 (3): 605–31.

Orphanides, A., R. D. Porter, D. Reifschneider, R. Tetlow, and F. Finan. 2000. "Errors in the Measurement of the Output Gap and the Design of Monetary Policy." *Journal of Economics and Business* 52 (1–2): 117–41.

Rabanal, P., and J. F. Rubio-Ramírez. 2008. "Comparing New Keynesian Models in the Euro Area: A Bayesian Approach." *Spanish Economic Review* 10 (1): 23–40.

Rotemberg, J. J., and M. Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." *NBER Macroeconomics Annual 1997*, ed. B. S. Bernanke and J. Rotemberg, 297–361. MIT Press.

Sala, L., U. Söderström, and A. Trigari. 2010. "The Output Gap, the Labor Wedge, and the Dynamic Behavior of Hours." CEPR Discussion Paper No. 8005.

Smets, F. 2002. "Output Gap Uncertainty: Does It Matter for the Taylor Rule?" *Empirical Economics* 27 (1): 113–29.

Smets, F., and R. Wouters. 2003a. "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1123–75.

———. 2003b. "Output Gaps: Theory versus Practice." Paper presented at the 9th International Conference of the Society for Computational Economics, Seattle, Washington, July 10–13.

Svensson, L. E., and M. Woodford. 2004. "Indicator Variables for Optimal Policy under Asymmetric Information." *Journal of Economic Dynamics and Control* 28 (4): 661–90.

Taylor, J. B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie Rochester Conference Series on Public Policy* 39 (1): 195–214.

———. 1999. "A Historical Analysis of Monetary Policy Rules." In *Monetary Policy Rules*, ed. J. B. Taylor, 319–48. University of Chicago Press.

Woodford, M. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ: Princeton University Press.

When the Walk Is Not Random: Commodity Prices and Exchange Rates*

Emanuel Kohlscheen, Fernando Avalos, and Andreas Schrimpf
Bank for International Settlements

We show that there is a distinct commodity-related driver of exchange rate movements, even at fairly high frequencies. Commodity prices predict exchange rate movements of eleven commodity-exporting countries in an in-sample panel setting for horizons up to two months. We also find evidence of systematic (pseudo) out-of-sample predictability, overturning the results of Meese and Rogoff (1983): information embedded in our country-specific commodity price indexes clearly helps to improve upon the predictive accuracy of the random walk in the majority of countries. We further show that the link between commodity prices and exchange rates is not driven by changes in global risk appetite or carry.

JEL Codes: F10, F31, G12.

1. Introduction

Recent developments in the oil market have brought the connection between commodity prices and the exchange rates of a number of countries back to the forefront of the policy debate. By affecting prospective inflows, substantial changes to the terms of trade of a given country are thought to exert a significant influence on exchange rates.

*We are grateful to three anonymous referees, Jonathan Kearns, Marco Lombardi, Robert McCauley, Ramon Moreno, Hyun Song Shin, and seminar participants at the BIS for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank for International Settlements. We thank Anamaria Illes and Diego Urbina for excellent research assistance. Author contact: Monetary and Economic Department, Bank for International Settlements, Centralbahnplatz 2, 4051 Basel, Switzerland. Author e-mail addresses: emanuel.kohlscheen@bis.org; fernando.avalos@bis.org; and andreas.schrimpf@bis.org.

In the short run, higher commodity prices lead to an increased supply of foreign exchange in the markets of commodity exporters, as a result of increased export revenues—causing an appreciation of the domestic currency. In the medium to long run, this effect might then be compounded by ensuing foreign direct investment, as a result of more attractive investment prospects in the local commodity sector.[1]

The above mechanisms tend to be fairly evident in the economies of commodity exporters. For such countries, price variation of key export commodities is often seen as a reasonably good *proxy* for terms-of-trade movements, as export price variation typically trumps the variation in import prices—which tends to be more dependent on more rigidly priced manufactured goods. Hence, changes in the prices of key exports may well bear a close link with exchange rate movements.[2]

In this paper we analyze the relation between commodity prices and the exchange rates of key commodity exporters in a systematic way. We base our analysis on a more timely proxy for terms of trade, which is based on granular three-digit UN Comtrade data as well as market price information of eighty-three associated proxy commodities, which were used to construct country-specific commodity export price indexes (CXPIs) at daily frequency for eleven commodity-exporting countries.[3]

The daily CXPIs allow us to analyze the relation between commodity prices and exchange rates with greater precision at different frequencies, as well as to tease out the extent to which this relation is independent of variations in global risk appetite or carry. We show how the information that is contained in these indexes clearly

[1] Over time, increased income due to improved terms of trade also tends to raise the demand for non-tradable goods, pushing their relative prices up, causing further real exchange rate appreciation.

[2] While it is of course possible that commodity price movements also affect the exchange rates of commodity importers, the link in these cases is likely to be less clear-cut, as there may be greater symmetry in the effects of import price fluctuations on different countries.

[3] As we show in the paper, the volatility of the export indexes (expressed in U.S. dollars) is lower than that of the oil price. It is, however, typically much larger than that of the aggregate Commodity Research Bureau (CRB) commodity price index for all eleven countries. Indeed, for eight of them the standard deviation of the commodity export price indexes is more than double that of the CRB.

improves the predictive performance of exchange rate models for all eleven of the commodity exporters that we study. In addition, the indexes provide more prompt information about the direction in which equilibrium exchange rates may be moving. They could thus prove to be useful for the evaluation of central bank or sovereign wealth fund actions in foreign exchange (FX) markets.

We find that commodity prices predict exchange rate movements of commodity exporters up to two months ahead when the analysis is based on in-sample panel regressions. Out-of-sample estimations also show that simple linear predictive models based on our commodity price indexes tend to have superior predictive performance for exchange rates when compared with random-walk benchmarks. These findings hold true for the three advanced economies and eight emerging markets in our sample. They hold for bilateral variations against the U.S. dollar (USD) and the Japanese yen (JPY), as well as for the nominal effective exchange rate (NEER) variations.

The key finding that commodity price models dominate random-walk models is based on the usual approach of utilizing realized variables as predictors (so-called pseudo out-of-sample tests), as pioneered by Meese and Rogoff (1983). As we show, evidence of out-of-sample predictability using only lagged predictors is clearly weaker, possibly as a consequence of the fact that commodity prices themselves are hard to predict.

We further show that variation in commodity prices has an effect on nominal exchange rates at high frequency that goes beyond the impact of global risk appetite. Daily variations in the Chicago Board Options Exchange (CBOE) volatility index (VIX) also explain a share of the nominal exchange rate variation. But, commodity prices explain a significant part of the variation of the exchange rate that is orthogonal to risk.[4] In other words, the high-frequency relation that exists between commodity prices and exchange rates goes beyond what is driven by the simultaneous movement of investors into (out of) commodity markets and high-yielding currencies during risk-on

[4]The finding that variations in global risk and risk appetite influence currency movements—in particular, those that feature strongly as funding or investment currencies in carry trades—is in line with recent studies by Adrian, Etula, and Shin (2009), Lustig, Roussanov, and Verdelhan (2011), Menkhoff et al. (2012), Gourio, Siemer, and Verdelhan (2013), and Farhi and Gabaix (2014), among others.

(risk-off) episodes. Our results are also found to be robust to the incorporation of information on short-term government bond yields differentials.

All in all, we provide extensive evidence that there is a distinct commodity-related driver of exchange rate movements, even at relatively high frequencies. For commodity exporters, variation in exchange rates is not random, but is tightly linked to movements in commodity prices.

1.1 Relation to the Literature

Several prior studies have established a low-frequency relation between commodity export prices and real exchange rates, including the seminal papers of Chen and Rogoff (2003) and Cashin, Cespedes, and Sahay (2004). Along the same lines, MacDonald and Ricci (2004) found strong evidence of cointegration between the real value of the South African rand and real commodity export prices.[5] In contrast to this literature, our study focuses on the high-frequency relation by drawing on a very rich data set that allows us to examine the relation between daily variations in nominal variables in a systematic way for eleven major commodity-exporting countries. More specifically, we make use of much more granular data on commodity export prices and export volumes. This much broader coverage ensures that the constructed country-specific export indexes are a better *proxy* of terms-of-trade shocks, measured at high frequency. Furthermore, incorporating price information for eighty-three commodity groups allows us to investigate the link between commodity prices and exchange rates also for countries that have a more diversified base of commodity exports.

In terms of the empirical strategy and methodology, our paper is closely related to that of Ferraro, Rogoff, and Rossi (2015), which

[5]Sidek and Yussof (2009) as well as Kohlscheen (2014) report similar findings for the Malayan ringitt and the Brazilian real, respectively. Also Hambur et al. (2015) report a strong relationship between Australia's terms of trade and the real exchange rate. Equally, terms-of-trade shocks affect the Chilean peso in a very significant way, according to estimates presented in de Gregorio and Labbe (2011). These authors find that effects on the exchange rate under inflation targeting are more immediate but of smaller magnitude in the long run when compared with the pre-inflation targeting period.

focuses mostly on the relationship between oil prices and the nominal value of the Canadian dollar. We go beyond by studying a much wider array of currencies and commodity prices. By finding that a key economic variable—commodity prices—consistently helps improve upon the predictive accuracy of the random walk, we overturn the well-known negative results of Meese and Rogoff (1983) and of Cheung, Chinn, and Pascual (2005). These two papers had established that models based on macroeconomic fundamentals are unable to outperform a simple random walk.[6] We also show that these findings are not driven by changes in uncertainty and global risk appetite, as proxied for by the VIX, which generally tend to be correlated with commodity price movements.

1.2 Outline

The article proceeds as follows. Section 2 describes the construction of the country-specific commodity export price indexes (the CXPIs). Section 3 shows how the high-frequency variation in these indexes is tightly related to the nominal exchange rate movements of commodity exporters. Section 4 shows that short-term yield differentials tend to perform relatively poorly as exchange rate predictors (with the notable exceptions of Australia and Canada), while adding information on commodity prices greatly improves forecasting performance. Section 5 presents several robustness tests. We conclude by discussing some possible directions for future research.

2. Constructing Country-Specific Commodity Export Price Indexes (CXPIs)

To study the link between commodity prices and exchange rates, we construct a daily commodity export price index (CXPI) for each major commodity-exporting country based on market price data of key commodities. We were able to associate quoted prices at daily frequency with a total of eighty-three UN Comtrade three-digit

[6]While many studies claimed better long-horizon predictability for models based on monetary fundamentals, Kilian (1999) argued that these findings were mostly due to size distortions.

commodity groups. Twenty-six referred to metals, thirty-six to agricultural commodities, eleven to livestock, and ten to energy. Price information was collected from Datastream and from Bloomberg. The main original source of data is the London Metal Exchange (LME) and the Chicago Mercantile Exchange (CME), but data from a number of alternative sources were also used. Iron ore prices, for instance, were based on data from the Shanghai Metal Exchange. For oil we used the Brent reference price.

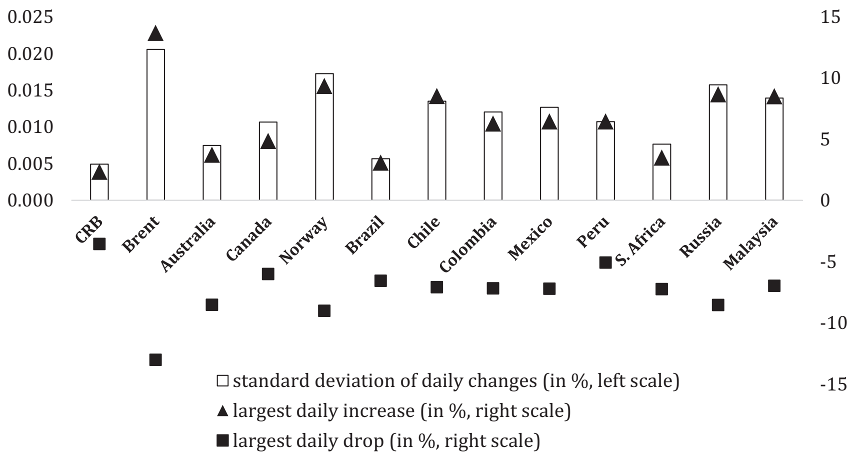
The country-specific commodity export price indexes were constructed as Laspeyres indexes. The weight of each commodity in each country basket was chosen so as to match the share of export revenues in total commodity export revenues in the respective country between 2004 and 2013.[7]

The weight of commodity groups for which good *proxy* market prices were not available was assumed to be zero. The underlying baskets of the CXPI indexes cover 98 percent of commodity exports of the countries considered in this study. The ten most important commodity segments for each country, according to their share in total export revenues, and their respective weights can be seen in table 9 in the appendix. The resulting indexes give a measure of the price of the exported commodity index in U.S. dollars. Note that this refers to nominal terms, as no correction for inflation was made. The sample period covers the time span between January 2, 2004 and February 28, 2015.

Figure 1 shows the variation of the country-specific commodity export price indexes for the eleven countries we analyze, as well as for the oil price (Brent) and the CRB commodity price index. The bars show that the Norwegian and the Russian CXPIs are clearly the most volatile indexes—which is a reflection of the very large contribution of oil in the commodity exports of these two countries. Even though the standard deviation of the Norwegian CXPI is 3.5 times larger than that of the CRB (1.73 percent versus 0.49 percent),

[7]Technically, because the basket weights are taken from an average over the entire period, the CXPI index is a Lowe index, which also belongs to the family of Laspeyres indexes. Triplett (1981) offers a more complete discussion. Strictly speaking, our index is not a pure Laspeyres index because the basket weight is not the weight measured at one specific instance of time. The value of a Lowe index tends to lie between the value given by the pure Laspeyres index and a Paasche index.

Figure 1. Variability of Commodity Export Price Indexes (CXPIs)

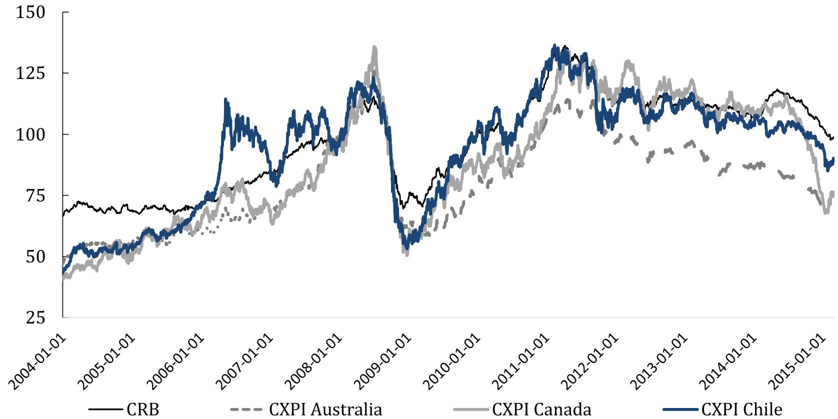


its volatility is somewhat lower than the volatility of the oil price because of diversification. On the other hand, Brazil has the least volatile index (with a standard deviation of 0.57 percent). That said, there have been episodes of basket price drops in excess of 5 percent within a day for all countries during our sample period (with one case of a drop in excess of 9 percent for the Norwegian CXPI). On the upside, Chile, Malaysia, Russia, and Norway have witnessed basket price increases above 8 percent within a single day.

Figure 2 plots the evolution of the CXPIs for each of the three selected countries, as well as for the CRB. The graph shows, for instance, how the sharp increase in commodity export prices in the second half of the 2000s in Chile predated similar movements for the Australian and Canadian indexes. The end of the sample captures the sharp oil price fall in late 2014 and the temporary partial rebound in early 2015, which is reflected in a discernible way in the evolution of the Canadian CXPI.

We compared the evolution of the Australian CXPI with the monthly index published by the Reserve Bank of Australia (RBA), which explains 75 percent of the variation in Australian exports according to Robinson and Wang (2013). At monthly frequency, the

Figure 2. Evolution of the CXPIs and the CRB



Note: Average of 2008 = 100. Own computation.

correlation of our index with that of the RBA is 0.904. While movements are broadly similar, the amplitude of the variations of the RBA index during the sample period is slightly larger than that of the corresponding CXPI, which is a reflection of the fact that the RBA index is rebased from time to time, whereas we did not rebase our daily index during our sample period.

The correlation matrix in table 1 shows that pairwise correlations vary quite substantially between countries. Commodity indexes for Colombia and Mexico, for instance, are highly correlated (0.971), again reflecting the predominance of oil in the commodity baskets of these countries. On the other hand, the cross-country correlations tend to be much lower for Chile (a large copper exporter). Of course, correlation of oil price variations with the changes in the values of the other commodity baskets creates the possibility that oil prices alone may actually predict exchange rate movements of countries that barely export any oil (see Ferraro, Rogoff, and Rossi 2015).

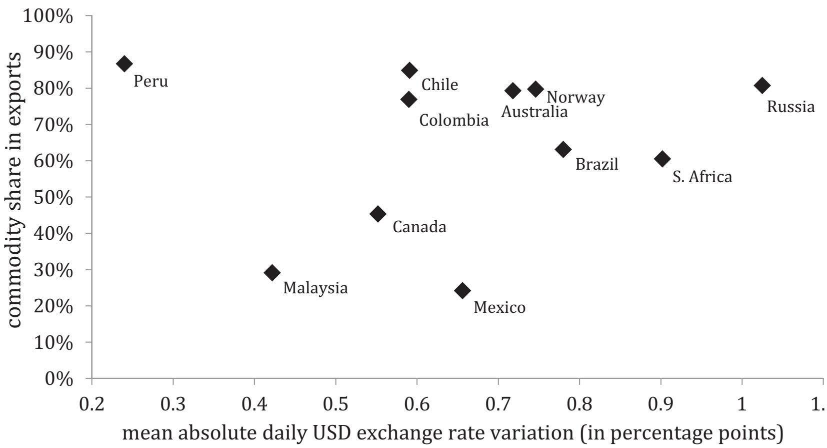
Lastly, figure 3 illustrates that the mean absolute daily exchange rate variation tends to be larger in countries in which the share of commodities in total exports is larger. An exception is Peru, possibly due to active intervention in FX markets (see Fuentes et al. 2013 and Blanchard, Adler, and Carvalho Filho 2015). This suggests

Table 1. Correlations of Commodity Price Indexes

	Australia	Canada	Norway	Brazil	Chile	Colombia	Mexico	Peru	S. Africa	Russia	Malaysia	CRB	Brent
CXPI Australia	1												
CXPI Canada	0.794	1											
CXPI Norway	0.670	0.937	1										
CXPI Brazil	0.720	0.687	0.538	1									
CXPI Chile	0.665	0.599	0.411	0.518	1								
CXPI Colombia	0.678	0.871	0.789	0.716	0.484	1							
CXPI Mexico	0.664	0.896	0.796	0.713	0.565	0.971	1						
CXPI Peru	0.763	0.730	0.543	0.607	0.919	0.631	0.698	1					
CXPI S. Africa	0.810	0.676	0.496	0.661	0.669	0.635	0.662	0.806	1				
CXPI Russia	0.691	0.964	0.969	0.618	0.453	0.894	0.903	0.587	0.552	1			
CXPI Malaysia	0.687	0.917	0.955	0.546	0.422	0.740	0.744	0.551	0.519	0.927	1		
CRB	0.491	0.475	0.329	0.477	0.517	0.425	0.473	0.543	0.507	0.375	0.367	1	
Brent	0.550	0.839	0.784	0.622	0.428	0.957	0.971	0.546	0.509	0.898	0.714	0.364	1

Note: Data at daily frequency, in log changes.

Figure 3. Commodity Share in Exports and Exchange Rate Volatility



that there could be a direct relation between commodity price and exchange rate movements. As we show in the sections that follow, this is indeed the case.

3. Commodity Prices as Drivers of Exchange Rate Movements

3.1 In-Sample Fit: Contemporaneous Correlations

As a first step, we run some simple panel regressions to explore the contemporaneous relation between exchange rates and commodities prices for our panel of eleven commodity exporters. These first-pass regressions suggest a clear association of nominal exchange rates with daily commodity price index variations in sample for all countries. More specifically, we estimate

$$\Delta s_{i,t} = \alpha + \beta \cdot \Delta CXPI_{i,t} + \gamma_i + \theta_t + \varepsilon_{i,t}, \tag{1}$$

where $s_{i,t}$ stands for the log of the (nominal) exchange rate of country i vis-à-vis the USD on day t, $CXPI_{i,t}$ for the log of the country-specific commodity export price index on the same day,

α for the constant term, γ_i for country fixed effects, θ_t for a vector of year dummies, and $\varepsilon_{i,t}$ for the error term.[8] The choice of a first-difference approach appears natural, as we are focusing on high-frequency variations and the variables in question typically contain stochastic trends.

The exercise was based on 30,294 country-day observations. The sample period goes from January 2004 to February 2015. For Malaysia, however, the sample starts only in August 2005, after the country abandoned its peg against the USD, whereas for Russia the sample period starts only in February 2009 (i.e., after the very substantial widening of the dual currency board (Bank for International Settlements 2013).

We obtain the following panel estimation results:

$$\Delta s_{i,t} = const. - \underset{[6.19]}{0.21} \cdot \Delta CXPI_{i,t} + Fixed\ Effects,\ \mathrm{R}^2 = 0.104, \tag{2}$$

where the t-statistics in the brackets below the coefficient estimate were based on cluster-robust standard errors. The estimated coefficient indicates that a 10 percent increase in the price of the commodities that are exported by a country in our sample is associated with a 2.1 percent appreciation of the respective currency, on average.

On a country-by-country basis (not tabulated here), the information of commodity price variation alone explains more than 23 percent of the variation in the USD exchange rate in the cases of Australia and Canada, on an ex post basis. On the other hand, this explanatory power was only about 3 percent for Peru.

3.2 Predictive Regression: In-Sample

To assess whether our commodity price indexes are able to predict nominal exchange rate variations, we estimated a generalized version of equation (1):

$$\Delta s_{i,t+k} = \alpha + \beta \cdot \Delta CXPI_{i,t} + \gamma_i + \theta_{t+k} + \varepsilon_{i,t+k}, \tag{3}$$

[8] Daily variations in real terms—if they were available—would tend to follow nominal variables quite closely, as s_t and $CXPI_t$ are based on market prices which adjust rapidly to news.

Table 2. Exchange Rate Predictability (in-sample) in a Panel Setting (dependent variable: log change of bilateral exchange rate)

	Prediction Horizon in Days				
	k = 1	**k = 5**	**k = 22**	**k = 44**	**k = 66**
CXPI	−0.020***	−0.016*	−0.044***	−0.047***	−0.018
t-stat	3.51	1.89	5.19	2.65	0.88
R2 Overall	0.0032	0.0113	0.0540	0.0934	0.1228
R2 Within	0.0032	0.0111	0.0530	0.0917	0.1206
R2 Between	0.5714	0.5608	0.6062	0.6143	0.6055
Observations	30,294	30,283	30,096	29,854	29,612
Groups	11	11	11	11	11
Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the results of the panel regression $\Delta s_{i,t+k} = \alpha + \Delta CXPI_{i,t} + \gamma_i + \theta_{t+k} + \varepsilon_{i,t+k}$, where k stands for the length of the prediction horizon. t-statistics are based on clustered standard errors. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively. The estimation is based on information from January 2004 to February 2015.

where k denotes the forecasting horizon (in working days). A significant β coefficient indicates that the commodity price information that is available at day t is indeed useful for predicting the variation of the exchange rate between t and $t + k$. We estimated this panel regression for horizons of one day up to three months. Estimations were based on country fixed effects γ_i and time effects θ_{t+k} as well as clustered standard errors.[9] For a discussion of pooled panel data and their merits for studying exchange rate predictability, the reader is referred to Mark and Sul (2001, 2012).

The results that are reported in table 2 show that commodity prices are indeed significant predictors of future exchange rates for horizons of up to two months. The R^2 statistics indicate that the explanatory power in this in-sample forecasting exercise is larger for variation between countries than within.

[9] Only yearly dummy variables were used as time effects, so as to avoid having more than 2,500 daily time dummy variables.

Up to now we have imposed a common coefficient for all eleven countries of our study. Of course, coefficients may vary a great deal between countries due, for instance, to the differing weight of commodities in total export revenues or to differences in the volatility of these indexes. The exchange rate may react less to price changes in countries where price indexes are very volatile, as these changes may be perceived as having only temporary effects on export revenues.

Table 10 in the appendix reports the results country by country. Even though the much smaller sample implies greater variation in coefficients over different horizons and countries, the monthly horizon (i.e., twenty-two days) stands out as being the one in which forecasting performance is more robust across countries. Commodity prices emerge as significant in-sample predictors for ten of the eleven countries. With the notable exception of South Africa, in-sample exercises suggest that exchange rates are at least to some extent predictable.

3.3 An Out-of-Sample Forecasting Experiment

A natural question is whether exchange rates of commodity exporters are also predictable out of sample. To evaluate the out-of-sample (OOS) performance of exchange rate models, we rely on the classical pseudo-OOS prediction framework, pioneered by Meese and Rogoff (1983). To this end, we run the following regression equation based on a rolling window:

$$\Delta s_t = \widehat{\alpha}_{t-T,t-1} + \widehat{\beta}_{t-T,t-1} \cdot \Delta CXPI_{i,t} + \varepsilon_t.$$

The estimated parameters $\widehat{\alpha}_{t-T,t-1}$ and $\widehat{\beta}_{t-T,t-1}$ capture drifts and the magnitude of the exchange rate response to commodity price changes. In other words, our procedure is able to capture the long-term variations in the sensitivity of exchange rates to commodity prices that may result, for instance, from secular changes in the share of commodities in the total exports of a country or changes in FX intervention policies. The use of out-of-sample forecasts for performance evaluation also diminishes the risks associated with data mining.[10]

[10]See Cheung, Chinn, and Pascual (2005) and Inoue and Kilian (2005).

We follow the convention of the literature (Meese and Rogoff 1983; Cheung, Chinn, and Pascual 2005) in that we use a rolling window of fixed length T to estimate the parameters $\widehat{\alpha}_{t-T,t-1}$ and $\widehat{\beta}_{t-T,t-1}$, which are then used to produce an out-of-sample forecast. The window is then rolled forward one period at a time to produce the coefficient estimates for the subsequent period. This procedure has also been dubbed a pseudo out-of-sample experiment, as only contemporaneous (and not lagged) realizations of the predictors are used. Yet, even in this very basic framework it has proven very challenging for any economic models to outperform a random-walk forecast (Rossi 2013).

In our baseline specification, we use a five-year estimation window (roughly half of the sample size), which leaves an evaluation period of 1,607 working days.[11] In other words, we estimate the set of coefficients 1,607 times for each country and then evaluate the performance of the model between January 2009 and February 2015.[12]

To evaluate the predictive performance of the exchange rate models, we compare the mean square prediction error (MSPE) of the baseline model with that of a pure random walk, as well as that of a random walk with a time-varying drift (which is obtained from the estimation of the model $\Delta s_t = \alpha_{t-T,t-1} + \varepsilon_t$ for each period).

The statistical significance of the difference between the squared error losses of the models is evaluated based on a long-run estimate of its variance, following the methodology proposed by Diebold and Mariano (1995). The Diebold-Mariano (DM) test is known to be asymptotically valid also for nested models when the size of the prediction sample grows, while the length of the estimation window is held fixed (see Giacomini and White 2006).

DM test statistics for each country and benchmark are reported in table 3. A statistically significant negative DM statistic indicates that the CXPI-based model has superior forecast accuracy relative to random-walk benchmarks.

[11]Because of the shorter time series, a three-year window is used in the case of Russia.

[12]In the robustness section we reduce the length of the estimation window. As we show, this does not lead to any substantive change to our conclusions.

Table 3. Exchange Rate Predictability by Commodity Prices (out-of-sample analysis)

		Forecasting Performance vs. RW Benchmark					
		RW without Drift			RW with (Time-Varying) Drift		
Currency	**Observations**	**RMSE Ratio**	**DM Stat**	**p-value**	**RMSE Ratio**	**DM Stat**	**p-value**
AUD	1,607	0.858	−5.30***	0.000	0.858	−5.29***	0.000
CAD	1,607	0.850	−4.16***	0.000	0.850	−4.16***	0.000
NOK	1,607	0.908	−3.79***	0.000	0.908	−3.78***	0.000
BRL	1,607	0.936	−3.51***	0.001	0.937	−3.45***	0.001
CLP	1,607	0.918	−4.64***	0.000	0.918	−4.63***	0.000
COP	1,607	0.953	−3.83***	0.000	0.953	−3.82***	0.000
MXN	1,607	0.910	−4.29***	0.000	0.910	−4.32***	0.000
PEN	1,607	0.965	−4.12***	0.000	0.966	−4.09***	0.000
ZAR	1,607	0.905	−5.36***	0.000	0.905	−5.34***	0.000
RUB	802	0.958	−2.04**	0.041	0.960	−1.99**	0.046
MYR	1,195	0.972	−1.95*	0.052	0.972	−1.95*	0.051

Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark. The coefficients were estimated with a five-year rolling window, following the Meese-Rogoff approach. The RMSE ratio refers to the root mean square error of the model based on commodities divided by the RMSE of the random-walk (RW) benchmark in question. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

The results show that the information on commodity price variation clearly leads to a one-step-ahead prediction performance that beats both benchmark random-walk models. The null of equal forecast accuracy is rejected with p-values below 1 percent for ten of the eleven currencies (Australia, Canada, Norway, Brazil, Chile, Colombia, Mexico, Peru, Russia, and South Africa). The two cases where the DM statistics are less significant are those of the two countries for which the sample size is smaller (Russia and Malaysia).[13]

Note also that the countries in which the reduction in the relative RMSE ratio is larger are advanced economies (in particular, the RMSE ratio is 0.850 for Canada and 0.858 for Australia). Countries that tend to perform larger interventions in terms of the average turnover of the respective FX market show considerably lower MSE reductions.

Since in most cases the absolute value of the DM statistic is smaller in the case of the random walk with time-varying drift, this model proves to be (slightly) more difficult to beat. For this reason we adopt the random walk with drift as the benchmark to beat in the sections that follow.

3.4 OOS Predictability over Longer Horizons

Given the strong relationship between commodity price developments captured by the CXPI indexes and exchange rates, we checked whether this link is also evident at lower frequencies.[14]

The results in table 4—and plotted in the associated figure 4—show that for most cases the relation is also found to be important at lower frequencies. In most cases, lengthening the window in which price variations are measured has the effect of weakening the relation somewhat in this out-of-sample exercise. Lengthening the window could have the effect of including additional sources of shocks that end up affecting the measured relation.

[13]Overall, we find that adding a drift component to the random-walk benchmark has very minor effects on forecast accuracy, as the estimated drifts are very small and generally not statistically significant.

[14]We thank an anonymous referee for making this suggestion.

Table 4. Exchange Rate Predictability by Commodities over Longer Horizons

		DM Stats [p-value] Length of Window			
Currency	**Observations**	**One Day**	**One Week**	**One Month**	**Six Months**
AUD	1,611	−5.28*** [0.000]	−4.81*** [0.000]	−3.27*** [0.001]	−2.83*** [0.005]
CAD	1,611	−4.16*** [0.000]	−4.43*** [0.000]	−3.62*** [0.000]	−3.08*** [0.002]
NOK	1,611	−4.63*** [0.000]	−3.84*** [0.000]	−2.80*** [0.005]	−2.93*** [0.003]
BRL	1,611	−3.45*** [0.001]	−3.06*** [0.002]	−2.62*** [0.009]	−2.65*** [0.008]
CLP	1,611	−4.63*** [0.000]	−3.75*** [0.000]	−2.72*** [0.007]	−4.98*** [0.000]
COP	1,611	−3.81*** [0.000]	−3.17*** [0.002]	−1.55 [0.121]	−1.37 [0.168]
MXN	1,611	−4.32*** [0.000]	−3.51*** [0.000]	−2.53** [0.011]	−2.48** [0.013]
PEN	1,611	−4.08*** [0.000]	−2.50** [0.012]	−1.76* [0.077]	−2.01** [0.044]
ZAR	1,611	−5.34*** [0.000]	−4.12*** [0.000]	−2.18** [0.029]	−3.55*** [0.000]
RUB	806	−1.99** [0.046]	−1.90* [0.057]	−1.54 [0.123]	−1.53 [0.126]
MYR	1,199	−1.95* [0.051]	−2.15** [0.031]	−1.52 [0.127]	−1.42 [0.154]

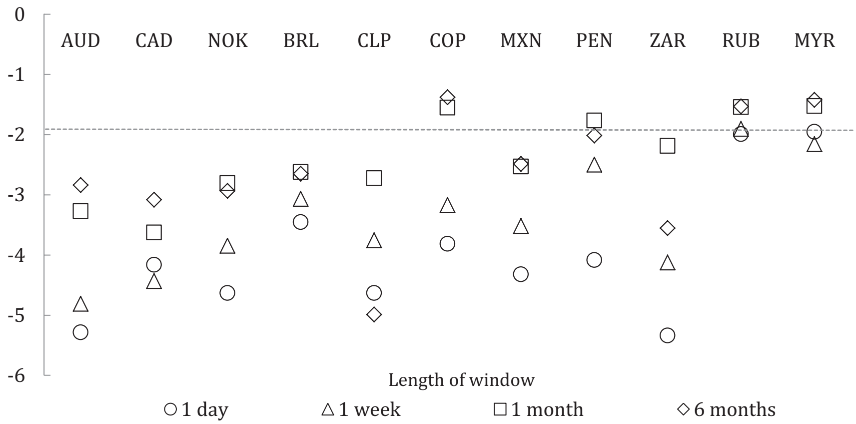
Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. The benchmark that is used as a reference is the random-walk model with a time-varying drift. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark. The coefficients were estimated with a five-year rolling window, following the Meese-Rogoff approach. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

3.5 OOS Predictability with Alternative Commodity Prices

As we have already pointed out above, the correlation of oil price variations with changes in the values of the other commodity baskets creates the possibility that oil prices alone may actually predict exchange rate movements of countries that barely export any oil (Ferraro, Rogoff, and Rossi 2015).

Indeed, the results shown in the third column of table 5 show that daily Brent price variations predict the exchange rates of many

Figure 4. Exchange Rate Predictability by Commodities: DM Test Statistics



of the countries in question in a way that is superior to the random walk at the 5 percent confidence level at daily frequency. Overall, however, the performance of the CXPI model is superior at daily frequency in all these cases, with the exception of the Russian ruble. Contrary to the case of the CXPI models, however, this relation disappears completely once the frequency is lowered to six months.

Similar results obtain for the CRB commodity price index. Again, the performance of the CXPI is superior, even though the CRB clearly does convey information that is relevant for exchange rate prediction at daily frequency.

3.6 OOS Predictability with Lagged Commodity Prices

An important consideration is that the out-of-sample exercises above are based on the well-established Meese and Rogoff (1983) benchmark of utilizing realized economic variables which are not known ex ante. This implies that the relations that were found are not necessarily useful for true forecasting or for making profitable investment bets, as the Meese and Rogoff methodology is based on information which is only available ex post.

Table 5. Exchange Rate Predictability by Oil Prices and the CRB Index (out-of sample analysis)

		DM Stats [p-value]			
		Model Based on Brent Price		Model Based on CRB	
Currency	Observations	One Day	Six Months	One Month	Six Months
AUD	1,611	−2.86*** [0.004]	−1.28 [0.200]	−2.83*** [0.005]	−0.38 [0.698]
CAD	1,611	−3.25*** [0.001]	−0.74 [0.462]	−2.81*** [0.005]	0.98 [0.329]
NOK	1,611	−3.69*** [0.000]	−1.32 [0.185]	−3.32*** [0.001]	1.11 [0.267]
BRL	1,611	−2.52** [0.012]	−0.76 [0.449]	−2.29** [0.022]	0.13 [0.896]
CLP	1,611	−2.99*** [0.003]	−0.29 [0.770]	−2.65*** [0.008]	0.97 [0.331]
COP	1,611	−2.70*** [0.007]	−0.02 [0.986]	−1.75 [0.079]	1.46 [0.143]
MXN	1,611	−2.83*** [0.005]	−0.23 [0.821]	−2.99*** [0.003]	0.67 [0.501]
PEN	1,611	−1.86* [0.062]	0.72 [0.473]	−1.04 [0.296]	2.52 [0.988]
ZAR	1,611	−3.25*** [0.001]	−1.25 [0.208]	−3.28*** [0.001]	−1.18 [0.235]
RUB	806	−2.15** [0.031]	−1.35 [0.177]	−1.99** [0.046]	0.88 [0.377]
MYR	1,199	−0.47 [0.639]	0.42 [0.672]	−1.21 [0.225]	−0.74 [0.457]

Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. The benchmark that is used as a reference is the random-walk model with a time-varying drift. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark, while positive values indicate that the random-walk benchmark is superior. The coefficients were estimated with a five-year rolling window, following the Meese-Rogoff approach. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

While forecasting actual exchange rate movements is clearly beyond the scope of this paper, we reestimated the model using only information on lagged commodity prices. Results of this exercise are shown in table 11 in the appendix. Forecasts that use exclusively lags of CXPIs beat the random-walk benchmarks in only a few countries (at the 10 percent level), with Chile standing out as the case with the greatest success. This result is in line with earlier findings in the literature, which have concluded that success in forecasting future exchange rate movements is often only detectable in certain instances and sample periods (Rossi 2013).[15,16]

3.7 A Reverse Link: Do Exchange Rates Predict CXPIs?

As our commodity price indexes are country specific, there could also be a reverse link—going from exchange rates to commodity prices measured in U.S. dollars. One rationale is that by increasing production costs, an appreciation of the currency of a commodity exporter might push up the U.S. dollar prices of the commodities produced in these exporters.

Indeed, there is the possibility of emergence of feedback loops between countries that produce the same type of commodities. An appreciation of the Brazilian real, for instance, could increase the costs of iron ore production, increasing the international price of this commodity. This in turn may well exert upward pressure on the value of the Australian dollar, which then pushes the price of iron ore further up, leading to a new round of appreciation of the Brazilian real and amplifying initial shocks. Mechanisms of this kind are explored in greater detail in Clements and Fry (2008).

To test for the possibility of this reverse link, we estimated the inverse model

$$\Delta CXPI_{i,t+k} = \alpha + \beta \cdot \Delta s_{i,t} + \gamma_i + \theta_{t+k} + \varepsilon_{i,t+k}, \tag{4}$$

using the panel approach that was outlined in the previous section.

Contrary to what is the case for the direct link, any indication of a reverse link disappears as soon as the six (or twelve) months after the collapse of Lehman Brothers are excluded from the analysis (table 6).

---

[15]It is commonly accepted that forecasting exchange rates in the time-series dimension is very hard, especially at shorter horizons. Engel and West (2005) show that the weak predictive relation between exchange rates and economic fundamentals can be reconciled within a standard present-value model when discount factors are close to unity and fundamentals are non-stationary.

[16]More success in predicting actual exchange rates has been obtained in the microstructure literature (Evans and Lyons 2005; Rime, Sarno, and Sojli 2010; and Menkhoff et al. 2016).

Table 6. The Reverse Link: Commodity Price Predictability by Exchange Rates (dependent variable: log change of commodity prices)

	Prediction Horizon in Days				
	k = 1	**k = 5**	**k = 22**	**k = 44**	**k = 66**
Bilateral Exchange Rate	−0.017	−0.005	−0.030	−0.037	−0.077
t-stat	1.41	0.46	0.71	0.94	0.75
R2 Overall	0.0066	0.0228	0.0912	0.1700	0.1802
R2 Within	0.0065	0.0226	0.0907	0.1690	0.1794
R2 Between	0.5762	0.5413	0.5957	0.6280	0.5290
Observations	28,941	28,897	28,710	28,468	28,226
Groups	11	11	11	11	11
Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the results of the fixed effects regression $\Delta CXPI_{i,t+k} = \alpha + \Delta s_{i,t} + \gamma_i + \theta_{t+k} + \varepsilon_{i,t+k}$, where k stands for the length of the prediction horizon. t-statistics are based on clustered standard errors. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively. The estimation is based on information from January 2004 to February 2015, except for observations of the six months after the collapse of Lehman Brothers.

4. Commodity Prices vs. Carry

To check for the robustness of the results of the previous section, we also compared the performance of forecasting models based on commodities with that of models based on interest rate differentials vis-à-vis the United States (known as “carry”). The literature on the forward premium puzzle (Fama 1984) has generally established that interest rate differentials have predictive power for exchange rates, yet in a manner that is inconsistent with uncovered interest parity (UIP).[17]

Our carry indicator is based on the difference between the one-year government bonds yield for the country in question and the United States. These data were obtained from Datastream. Overall,

[17]Verdelhan (2015) also presents evidence that carry is an important driver of variation in bilateral exchange rates. Akram and Mumtaz (2016) on the other hand show that, in the case of Norway, the correlations between money market rates and nominal exchange rates have fallen towards zero.

the results in table 7 show that the pure yield differential models only outperform the random-walk benchmarks in the cases of the Australian and Canadian dollar. When the information of the CXPIs is added to the yield differential model, the expanded model beats the random-walk benchmarks in nine cases at the 5 percent confidence level (and in ten cases at the 10 percent level).

What is clear is that for many commodity exporters, information of commodity prices appears to be more important than that of government bond yields. The DM statistics when commodity prices are used as predictors are systematically below those obtained when relying on carry—also in the cases of Australia and Canada.[18] In the latter two cases, however, the model that combines information of both factors tends to be the superior one.

5. Robustness

We performed a number of robustness checks, which largely confirmed the conclusions drawn above. In the following, we summarize a few main take-aways.

5.1 Changes in Uncertainty and Global Risk Appetite

At least in principle, there could be the possibility that the relations that were highlighted in the previous sections are mainly due to changes in global risk appetite. These may cause global investors to move into or out of commodity markets and foreign exposures in a synchronized way, with consequent effects on exchange rates. Daily variation in risk perceptions can be proxied by the CBOE VIX, as in Adrian, Etula, and Shin (2009), McCauley (2012), or Bock and Carvalho Filho (2015).[19] The latter studies suggest that the VIX is indeed a good indicator to flag risk-off episodes in global financial markets.

[18]Note that this does not mean that trading strategies based on interest differentials (so-called carry trades) are unprofitable. See Hassan and Mano (2014) for evidence that the informational content of carry is mostly cross-sectional rather than in the time-series dimension.

[19]As discussed in Bekaert, Hoerova, and Lo Duca (2013), the VIX can be thought of as a measure of stock market uncertainty and the reward investors require for taking on risk.

Table 7. Exchange Rate Predictability: Carry vs. Commodities

		Carry Model			Carry + CXPI Model		
Currency	**Observations**	**RMSE Ratio**	**DM Stat**	**p-value**	**RMSE Ratio**	**DM Stat**	**p-value**
AUD	1,607	0.951	−4.66***	0.000	0.819	−6.15***	0.000
CAD	1,607	0.992	−2.85***	0.004	0.841	−4.42***	0.000
NOK	1,607	1.000	0.23	0.817	0.908	−3.80***	0.000
BRL	1,607	1.036	2.62	0.009	0.958	−2.66***	0.008
CLP	1,607	1.000	0.74	0.460	0.918	−4.69***	0.000
COP	1,607	1.009	1.64	0.100	0.961	−3.31***	0.001
MXN	1,607	0.998	−0.72	0.474	0.909	−4.29***	0.000
PEN	1,607	1.003	0.90	0.368	0.961	−3.42***	0.001
ZAR	1,607	1.023	0.61	0.544	0.919	−2.20**	0.028
RUB	802	0.808	−1.12	0.261	0.776	−1.25	0.210
MYR	1,195	1.000	1.702	0.089	0.972	−1.90*	0.057

Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. The benchmark that is used as a reference is the random-walk model with a time-varying drift. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark, while positive values indicate that the random-walk benchmark is superior. The coefficients were estimated with a five-year rolling window, following the Meese-Rogoff approach. The RMSE ratio refers to the RMSE of the model based on commodities divided by the RMSE of the RW benchmark in question. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

To ensure that the relation which we found above is not just a side effect of variations in global risk (appetite), we checked whether commodity price variations are able to explain the component of exchange rate variation that is orthogonal to changes in the VIX. The results in table 8 show that in all eleven cases that were listed before, daily variations in the commodity price index indeed explain the exchange rate movements that are unrelated to changes in the VIX. The explanatory power attains its maximum in the cases of Australia and Canada, but is economically and statistically significant at 1 percent in all eleven cases.

5.2 Changing the Base Currency

Results in the previous sections were based on bilateral USD exchange rates. Table 12 in the appendix shows that results weaken only marginally when the nominal exchange rate vis-à-vis the JPY is used instead.[20] This can be explained by the fact that most commodities are actually priced in U.S. dollars. A change in the value of the U.S. dollar—the invoicing currency—tends to lead to some change in the final USD price of commodities. Indeed, periods of U.S. dollar weakness tend to be associated with higher oil prices (see Akram 2009). Nevertheless, linear JPY exchange rate models based on commodities still outperform the random-walk benchmark for nine of the eleven currencies at the 5 percent significance level.

The last column of the table shows that very similar results are also obtained when daily exchange rate variations are measured in terms of an effective nominal exchange rate. Here the effective nominal exchange rates were computed against a basket of the five major global currencies (the U.S. dollar, the euro, the Japanese yen, the British pound, and the Chinese yuan). The weight of each currency in the country-specific basket was based on the total trade relation of the country in question with the United States, Japan, the twelve first members of the euro area, the United Kingdom, and China.

[20] The JPY is the global currency that is least correlated with the USD during our sample period.

Table 8. Global-Risk-Adjusted Exchange Rates and Commodity Prices

	Australia	**Canada**	**Norway**	**Brazil**	**Chile**	**Colombia**	**Mexico**	**Peru**	**S. Africa**	**Russia**	**Malaysia**
*First-Stage Regression (D.V.: 100*Log Diff of Exchange Rate)*											
VIX	0.130***	0.088***	0.088***	0.158***	0.100***	0.088***	0.120***	0.028***	0.152***	0.165***	0.026***
	(7.24)	(7.45)	(6.12)	(8.46)	(8.36)	(7.75)	(7.99)	(6.36)	(8.46)	(6.37)	(5.24)
R2	0.0673	0.0597	0.0368	0.0896	0.0695	0.0482	0.1011	0.0285	0.0655	0.0725	0.0166
Second-Stage Regression (D.V.: Residual of First-Stage Regression)											
CXPI	−0.518***	−0.257***	−0.167***	−0.440***	−0.154***	−0.131***	−0.163***	−0.039***	−0.493***	−0.221***	−0.061***
	(11.46)	(17.56)	(14.49)	(9.94)	(13.03)	(9.64)	(12.64)	(5.43)	(12.20)	(11.94)	(10.07)
R2	0.1967	0.1928	0.1301	0.0768	0.0999	0.0515	0.1042	0.0207	0.1330	0.0896	0.0461
Obs.	2,912	2,912	2,912	2,912	2,912	2,912	2,912	2,912	2,912	1,584	2,499

Notes: This table shows regression of the residual of the first-stage regression on the log change of the commodity price index at daily frequency. The sample period is January 2004 to February 2015. Constants are not shown, as they were not significant in any case. t-statistics based on Newey-West standard errors are reported in parentheses. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

5.3 Clark and West Tests

So far we have reported the outcome of Diebold and Mariano tests, which is appropriate given the setup of our out-of-sample exercise (Giacomini and White 2006). Clark and West (2007), however, point out that, for nested models, the mean square prediction errors should be adjusted to account for the possibility that less parsimonious models might introduce noise by estimating a parameter whose value might actually be zero in the population. The statistic proposed by Clark and West properly takes this possibility of model degeneration into account. The adjusted MSPE is

$$\frac{\Sigma\left(y_{t+\tau}-\widehat{y}_{1t,t+\tau}\right)^2-\left[\Sigma\left(y_{t+\tau}-\widehat{y}_{2t,t+\tau}\right)^2-\Sigma\left(\widehat{y}_{1t,t+\tau}-\widehat{y}_{2t,t+\tau}\right)^2\right]}{N},$$

where $\widehat{y}_{1t,t+\tau}$ is the predicted value of the parsimonious model, $\widehat{y}_{2t,t+\tau}$ is the predicted value of the model that nests the parsimonious specification, $y_{t+\tau}$ is the actual outcome, and N is the number of predictions. The last term on the numerator represents the adjustment to the estimated variance of the nesting model.

Table 13 in the appendix shows that the use of the Clark and West statistic tends to strengthen the results of the previous sections. In all cases the p-values of the null of equal forecast accuracy diminishes relative to the one obtained from the Diebold and Mariano comparison of MSPEs.

5.4 Shorter Estimation Windows

Finally, to be sure that our results are not driven by our particular selection of the window length, we also replicated the estimation of the previous section using a different window length. The original five-year choice had been made so that the estimation window was roughly half the sample size of eleven years. Table 14 shows that our conclusions are not changed in any material way if we use a three-year window instead. We find that the commodity price model outperforms the random-walk models at the 5 percent significance level for all eleven countries.

6. Conclusion

This paper shows evidence of a distinct commodity-related driver in currency movements. The link between commodity prices and exchange rates is economically and statistically significant even at high frequency. Further, the commodity price–exchange rate nexus is largely unaffected when changes in uncertainty and global risk appetite are taken into account: models incorporating commodity prices explain the component of exchange rate variations that is purely orthogonal to changes in risk and risk appetite. They also tend to deliver better predictive accuracy than standard models based on interest rate differentials (carry).

Our intention in this paper was not to provide daily forecasts of movements in actual exchange rates. Following the usual practice of the literature, we utilized realized variables in the exchange rate prediction. Based on this standard setting, we show that even high-frequency movements of the exchange rates of commodity exporters have a strong relationship with the market value of their exports.

Our finding of a distinct commodity-related driver of exchange rates suggests that currency movements are not purely random. There is a factor related to commodities that helps explain movements in exchange rates which goes beyond the information embedded in carry, global uncertainty, and risk appetite.

Finally, the connection between export commodity prices and the exchange rates of resource rich countries raises a number of more fundamental questions. For instance, several commodity exporters intervene in their FX markets with some regularity. It would be interesting to establish the degree to which these interventions are affected by commodity price developments or take these into account. Still other countries seek some degree of stabilization via operations of sovereign wealth or oil funds. To the extent that these operations shift inflows intertemporally and generate expectations of future inflows, they may well have an effect on the exchange rate and possibly other macroeconomic variables. Identifying such effects could be an additional interesting avenue for future research.

Appendix

Table 9. Shares of Commodity Groups in Exports (2004–13)

	Commodity Exports/Total Exports	Share in Commodity Exports	Description of Group
Australia	0.793	0.227	Iron ore, concentrates
		0.212	Coal, not agglomerated
		0.076	Gold, non-monetary excl. ores
		0.054	Petroleum oils, crude
		0.053	Natural gas
		0.033	Aluminum ore, conctr., etc.
		0.029	Bovine meat
		0.028	Aluminum
		0.028	Wheat, meslin, unmilled
		0.026	Copper ores, concentrates
Brazil	0.631	0.172	Iron ore, concentrates
		0.103	Petroleum oils, crude
		0.100	Oilseed (sft. fix veg. oil)
		0.074	Sugars, molasses, honey
		0.058	Other meat, meat offal
		0.042	Coffee, coffee substitute
		0.041	Animal feed stuff
		0.033	Petroleum products
		0.032	Pulp and waste paper
		0.032	Bovine meat
Canada	0.453	0.271	Petroleum oils, crude
		0.107	Naural gas
		0.076	Petroleum products
		0.054	Gold, non-monetary excl. ores
		0.042	Aluminum
		0.037	Wood, simply worked
		0.035	Pulp and waste paper
		0.026	Wheat, meslin, unmilled
		0.025	Coal, not agglomerated
		0.023	Oilseed (sft. fix veg. oil)
Chile	0.849	0.402	Copper
		0.229	Copper ores, concentrates
		0.066	Fruit, nuts excl. oil nuts
		0.049	Fish, fresh, chilled, frozen

(*continued*)

Table 9. (Continued)

	Commodity Exports/Total Exports	Share in Commodity Exports	Description of Group
		0.040	Pulp and waste paper
		0.034	Ore, concentr. base metals
		0.017	Wood in chips, particles
		0.017	Petroleum products
		0.017	Gold, non-monetary excl. ores
		0.015	Iron ore, concentrates
Colombia	0.769	0.429	Petroleum oils, crude
		0.162	Coal, not agglomerated
		0.094	Petroleum products
		0.067	Coffee, coffee substitute
		0.055	Gold, non-monetary excl. ores
		0.038	Crude veg. materials, nes
		0.031	Pig iron, spiegeleisn, etc.
		0.025	Fruit, nuts excl. oil nuts
		0.012	Sugars, molasses, honey
		0.011	Coke, semi-coke, ret. carbn.
Malaysia	0.291	0.206	Natural gas
		0.205	Fixed veg. fat, oils, other
		0.175	Petroleum oils, crude
		0.160	Petroleum products
		0.025	Copper
		0.020	Wood simply worked
		0.017	Petroleum Gases, nes
		0.015	Cocoa
		0.015	Aluminum
		0.011	Wood rough, rough squared
Mexico	0.242	0.532	Petroleum oils, crude
		0.065	Petroleum products
		0.058	Vegetables
		0.055	Gold, non-monetary excl. ores
		0.032	Silver, platinum, etc.
		0.030	Fruit, nuts excl. oil nuts
		0.019	Coffee
		0.016	Ore, concentr. base metals
		0.016	Ingots etc. iron or steel
		0.013	Non-ferrous waste, scrap

(*continued*)

Table 9. (Continued)

	Commodity Exports/Total Exports	Share in Commodity Exports	Description of Group
Norway	0.797	0.475	Petroleum oils, crude
		0.267	Natural gas
		0.059	Petroleum products
		0.056	Fish, fresh, chilled, frozen
		0.042	Aluminum
		0.026	Liquified propane, butane
		0.017	Nickel
		0.008	Fish, dried, salted, smoked
		0.006	Pig iron, spiegeleisn, etc.
		0.004	Petroleum gases, nes
Peru	0.867	0.229	Gold, non-monetary excl. ores
		0.183	Copper ores, concentrates
		0.108	Ore, conctr. base metals
		0.101	Copper
		0.077	Petroleum products
		0.055	Animal feed stuff
		0.026	Coffee, coffee substitute
		0.018	Iron ore, concentrates
		0.017	Fruit, nuts excl. oil nuts
		0.017	Petroleum oils, crude
Russia	0.807	0.407	Petroleum oils, crude
		0.212	Petroleum products
		0.164	Natural gas
		0.025	Coal, not agglomerated
		0.022	Ingots etc. iron or steel
		0.022	Aluminum
		0.016	Nickel
		0.014	Flat-rolled iron etc.
		0.014	Copper
		0.012	Pig iron, spiegeleisn, etc.
South Africa	0.605	0.192	Silver, platinum, etc.
		0.110	Coal, not agglomerated
		0.096	Iron ore, concentrates
		0.092	Pig iron, spiegeleisn, etc.
		0.067	Gold, non-monetary excl. ores
		0.063	Ore, concentr. base metals
		0.047	Petroleum products
		0.043	Aluminum
		0.042	Fruit, nuts excl. oil nuts
		0.023	Flat-rolled, alloy steel

Note: Compiled based on three-digit UN Comtrade data.

Table 10. Exchange Rate Predictability by Commodities by Country

	Prediction Horizon in Days					
	k = 0	**k = 1**	**k = 5**	**k = 22**	**k = 44**	**k = 66**
AUD						
CXPI	−0.579***	−0.044	−0.091**	−0.074***	−0.026	0.002
t-stat	12.12	1.30	2.20	3.07	1.43	0.18
R2	0.2286	0.0013	0.0068	0.0083	0.0012	0.0000
CAD						
CXPI	−0.287***	−0.019	−0.014	−0.057***	−0.089***	−0.071***
t-stat	18.96	1.33	0.83	4.22	7.51	8.53
R2	0.2248	0.0010	0.0006	0.0136	0.0399	0.0275
NOK						
CXPI	−0.180***	−0.018*	0.001	−0.074***	−0.060***	−0.025**
t-stat	15.31	1.78	0.14	8.89	5.40	2.51
R2	0.1446	0.0015	0.0000	0.0326	0.0196	0.0034
BRL						
CXPI	−0.545***	−0.060	−0.074	−0.244***	−0.261***	−0.154***
t-stat	11.86	1.25	1.55	8.54	11.69	9.17
R2	0.1073	0.0013	0.0027	0.0503	0.0718	0.0267
CLP						
CXPI	−0.190***	−0.027**	−0.058***	−0.027**	0.027**	0.074***
t-stat	14.66	2.23	3.93	2.23	2.56	9.21
R2	0.1405	0.0029	0.0105	0.0026	0.0034	0.0263
COP						
CXPI	−0.152***	−0.025*	0.000	−0.046***	−0.099***	−0.094***
t-stat	10.43	1.74	0.02	3.71	7.45	9.76
R2	0.0657	0.0017	0.0000	0.0071	0.0329	0.0310
MXN						
CXPI	−0.195***	0.002	−0.032**	−0.076***	−0.143***	−0.121***
t-stat	13.70	0.15	2.04	6.90	13.00	13.97
R2	0.1336	0.0000	0.0039	0.0274	0.1054	0.0807
PEN						
CPXI	−0.050***	−0.004	−0.016**	−0.058***	−0.076***	−0.065***
t-stat	6.64	0.55	2.05	9.89	16.44	16.08
R2	0.0319	0.0002	0.0029	0.0445	0.0796	0.0610
ZAR						
CXPI	−0.558***	0.036	−0.015	−0.033	−0.034	0.014
t-stat	13.75	1.08	0.38	1.22	1.50	1.16
R2	0.1592	0.0007	0.0001	0.0012	0.0016	0.0004
RUB						
CXPI	−0.277***	−0.075***	−0.050*	−0.170***	−0.315***	−0.348***
t-stat	14.44	3.52	1.87	7.23	10.28	9.55
R2	0.1293	0.0097	0.0044	0.0569	0.1202	0.1165
MYR						
CXPI	−0.067***	−0.028***	−0.005	−0.041***	−0.041***	−0.039***
t-stat	10.88	5.15	0.89	7.89	7.62	9.66
R2	0.0549	0.0096	0.0003	0.0240	0.0245	0.0257

Notes: This table shows the results of the OLS regression $\Delta s_{t+k} = \alpha + \Delta \mathrm{CXPI}_t + \varepsilon_{t+k}$, where k stands for the length of the prediction horizon. Constants are not tabulated. t-statistics, which are reported below the coefficients, are based on Newey-West standard errors. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively. The estimation is based on information from the full sample period (i.e., January 2004–February 2015), except for Russia and Malaysia, where the sample period starts in February 2009 and August 2005, respectively.

Table 11. Exchange Rate Predictability with Lagged Commodity Prices (out-of-sample analysis)

		Forecasting Performace vs. RW Benchmark							
		Prediction of Exchange Rate vs. USD				Prediction of Exchange Rate vs. JPY			
		RW without Drift		RW with Drift		RW without Drift		RW with Drift	
Currency	**Obs.**	**DM Stat**	**p-value**	**DM Stat**	**p-value**	**DM Stat**	**p-value**	**DM Stat**	**p-value**
AUD	1,607	−0.01	0.994	−0.07	0.944	1.24	0.212	0.69	0.486
CAD	1,607	−0.10	0.922	−0.07	0.937	1.31	0.190	1.22	0.221
NOK	1,607	2.35**	0.019	2.58***	0.010	2.58***	0.010	2.27**	0.023
BRL	1,607	−1.66*	0.095	−1.12	0.248	0.08	0.938	−0.12	0.902
CLP	1,607	−2.13**	0.034	−2.37**	0.018	−2.78***	0.005	−2.70***	0.007
COP	1,607	−1.62	0.105	−1.62	0.104	−1.58	0.113	−1.68*	0.092
MXN	1,607	1.62	0.105	1.20	0.231	1.23	0.218	0.66	0.505
PEN	1,607	−1.01	0.311	−0.46	0.643	0.14	0.893	−0.51	0.608
ZAR	1,607	0.36	0.722	−0.23	0.818	−0.87	0.384	−0.91	0.359
RUB	802	0.21	0.836	0.08	0.933	−1.69*	0.091	−1.80*	0.071
MYR	1,195	−1.71*	0.087	−1.84*	0.065	−2.05**	0.040	−2.20**	0.028

Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark. The coefficients were estimated with a five-year rolling window. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

Table 12. Exchange Rate Predictability by Commodities for Alternative Base Currencies

Currency	Observations	Diebold-Mariano Statistics [p-value] vs. JPY	Diebold-Mariano Statistics [p-value] vs. NEER
AUD	1,611	−4.09*** [0.000]	−5.23*** [0.000]
CAD	1,611	−3.09*** [0.002]	−4.16*** [0.000]
NOK	1,611	−3.16*** [0.002]	−3.82*** [0.000]
BRL	1,611	−2.70*** [0.007]	−3.43*** [0.006]
CLP	1,611	−4.24*** [0.000]	−4.62*** [0.000]
COP	1,611	−2.51** [0.012]	−3.80*** [0.001]
MXN	1,611	−3.14*** [0.002]	−4.32*** [0.000]
PEN	1,611	0.11 [0.909]	−3.97*** [0.000]
ZAR	1,611	−3.92*** [0.000]	−5.30*** [0.000]
RUB	806	−2.06** [0.039]	−2.00** [0.045]
MYR	1,199	−0.95 [0.340]	−1.89* [0.058]

Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. The benchmark that is used as a reference is the random-walk model with a time-varying drift. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark, while positive values indicate that the random-walk benchmark is superior. The coefficients were estimated with a five-year rolling window, following the Meese-Rogoff approach. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

Table 13. Exchange Rate Predictability by CXPI Model—Clark and West Statistics

		RW with (Time-Varying) Drift		
Currency	**Observations**	**RMSE Ratio**	**CW Stat**	**p-value**
AUD	1,611	0.913	11.77***	0.000
CAD	1,611	0.918	11.72***	0.000
NOK	1,611	0.920	12.90***	0.000
BRL	1,611	0.952	9.83***	0.000
CLP	1,611	0.952	9.53***	0.000
COP	1,611	0.965	8.63***	0.000
MXN	1,611	0.942	10.35***	0.000
PEN	1,611	1.002	2.35**	0.019
ZAR	1,611	0.938	9.74***	0.000
RUB	806	0.964	6.08***	0.000
MYR	1,199	0.991	4.63***	0.000

Notes: The null hypothesis of the Clark West (2007) test is that forecast accuracy is equal. Coefficients were estimated with a five-year rolling window, following the Meese-Rogoff approach. Significant statistics indicate that the tested model has superior predictive performance when compared with the benchmark. The RMSE ratio refers to the RMSE of the model based on commodities divided by the RMSE of the random walk with time-varying drift benchmark. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

Table 14. Exchange Rate Predictability by CXPI Model Based on Estimation with Three-Year Rolling Windows

		Forecasting Performance vs. RW Benchmark					
		RW without Drift			RW with (Time-Varying) Drift		
Currency	**Observations**	**RMSE Ratio**	**DM Stat**	**p-value**	**RMSE Ratio**	**DM Stat**	**p-value**
AUD	1,607	0.856	−4.97***	0.000	0.856	−4.97***	0.000
CAD	1,607	0.831	−4.36***	0.000	0.831	−4.36***	0.000
NOK	1,607	0.891	−3.84***	0.000	0.891	−3.85***	0.000
BRL	1,607	0.937	−3.13***	0.002	0.937	−3.11***	0.002
CLP	1,607	0.919	−4.34***	0.000	0.918	−4.34***	0.000
COP	1,607	0.949	−3.84***	0.000	0.949	−3.82***	0.000
MXN	1,607	0.899	−4.05***	0.000	0.898	−4.09***	0.000
PEN	1,607	0.963	−3.74***	0.000	0.964	−3.68***	0.000
ZAR	1,607	0.907	−5.22***	0.000	0.906	−5.16***	0.000
RUB	802	0.958	−2.04**	0.041	0.960	−1.99**	0.046
MYR	1,607	0.964	−2.58***	0.010	0.964	−2.59***	0.009

Notes: The null hypothesis of the Diebold-Mariano test is that forecast accuracy is equal. Negative DM statistics indicate that the tested model has superior predictive performance when compared with the benchmark. The coefficients were estimated with a three-year rolling window, following the Meese-Rogoff approach. The RMSE ratio refers to the RMSE of the model based on commodities divided by the RMSE of the RW benchmark in question. *, **, and *** denote statistical significance at 10 percent, 5 percent, and 1 percent, respectively.

References

Adrian, T., E. Etula, and H. S. Shin. 2009. "Risk Appetite and Exchange Rates." Staff Report No. 361, Federal Reserve Bank of New York.

Akram, Q. F. 2009. "Commodity Prices, Interest Rates and the Dollar." *Energy Economics* 31 (6): 838–51.

Akram, Q. F., and H. Mumtaz. 2016. "The Role of Oil Prices and Monetary Policy in the Norwegian Economy since the 1980s." Working Paper No. 1/2016, Norges Bank.

Bank for International Settlements. 2013. "The History of the Bank of Russia's Exchange Rate Policy." *BIS Papers* 73 (October): 293–99.

Bekaert, G., M. Hoerova, and M. Lo Duca. 2013. "Risk, Uncertainty, and Monetary Policy." *Journal of Monetary Economics* 60 (7): 771–88.

Blanchard, O., G. Adler, and I. Carvalho Filho. 2015. "Can Foreign Exchange Intervention Stem Exchange Rate Pressures from Global Capital Flow Shocks?" NBER Working Paper No. 21427.

Bock, R., and I. Carvalho Filho. 2015. "The Behavior of Currencies during Risk-off Episodes." *Journal of International Money and Finance* 53: 218–34.

Cashin, P., L. F. Cespedes, and R. Sahay. 2004. "Commodity Currencies and the Real Exchange Rate." *Journal of Development Economics* 75 (1): 239–68.

Chen, Y., and K. Rogoff. 2003. "Commodity Currencies." *Journal of International Economics* 60 (1): 133–60.

Cheung, Y. W., M. D. Chinn, and A. G. Pascual. 2005. "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" *Journal of International Money and Finance* 24 (7): 1150–75.

Clark, T., and K. West. 2007. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics* 138 (1): 291–311.

Clements, K. W., and R. Fry. 2008. "Commodity Currencies and Currency Commodities." *Resources Policy* 33 (2): 55–73.

de Gregorio, J., and F. Labbe. 2011. "Copper, the Real Exchange Rate and Macroeconomic Fluctuations in Chile." Working Paper No. 640, Central Bank of Chile.

Diebold, F. X., and R. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13 (3): 253–63.

Engel, C., and K. D. West. 2005. "Exchange Rates and Fundamentals." *Journal of Political Economy* 113 (3): 485–517.

Evans, M. D., and R. Lyons. 2005. "Meese-Rogoff Redux: Micro-based Exchange-Rate Forecasting." *American Economic Review: Papers and Proceedings* 95 (2): 405–14.

Fama, E. F. 1984. "Forward and Spot Exchange Rates." *Journal of Monetary Economics* 14 (3): 319–38.

Farhi, E., and X. Gabaix. 2014. "Rare Disasters and Exchange Rates." Mimeo, New York University.

Ferraro, D., K. Rogoff, and B. Rossi. 2015. "Can Oil Prices Forecast Exchange Rates? An Empirical Analysis of the Relationship between Commodity Prices and Exchange Rates." *Journal of International Money and Finance* 54: 116–41.

Fuentes, M., P. Pincheira, J. M. Julio, H. Rincon, S. G. Verdu, M. Zerecero, M. Vega, E. Lahura, and R. Moreno. 2013. "The Effects of Intraday Foreign Exchange Market Operations in Latin America: Results for Chile, Colombia, Mexico and Peru." BIS Working Paper No. 462.

Giacomini, R., and H. White. 2006. "Tests of Conditional Predictive Ability." *Econometrica* 74 (6): 1545–78.

Gourio, F., M. Siemer, and A. Verdelhan. 2013. "International Risk Cycles." *Journal of International Economics* 89 (2): 471–84.

Hambur, J., L. Cockerell, C. Potter, P. Smith, and M. Wright. 2015. "Modelling the Australian Dollar." RBA Research Discussion Paper No. 2015-12.

Hassan, T., and R. C. Mano. 2014. "Forward and Spot Exchange Rates in a Multi-currency World." NBER Working Paper No. 20294.

Inoue, A., and L. Kilian. 2005. "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews* 23 (4): 371–402.

Kilian, L. 1999. "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics* 14 (5): 491–510.

Kohlscheen, E. 2014. "Long-Run Determinants of the Brazilian Real: A Closer Look at Commodities." *International Journal of Finance and Economics* 19 (4): 239–50.

Lustig, H., N. Roussanov, and A. Verdelhan. 2011. "Common Risk Factors in Currency Markets." *Review of Financial Studies* 24 (11): 3731–77.

MacDonald, R., and L. Ricci. 2004. "Estimation of the Equilibrium Real Exchange Rate for South Africa." *South African Journal of Economics* 72 (2): 282–304.

Mark, N. C., and D. Sul. 2001. "Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Panel." *Journal of International Economics* 53 (1): 29–52.

———. 2012. "When Are Pooled Panel-Data Regression Forecasts of Exchange Rates More Accurate than the Time-Series Regression Forecasts?" In *Handbook of Exchange Rates,* ed. J. James, I. W. Marsh, and L. Sarno, 265–82 (chapter 9). John Wiley & Sons.

McCauley, R. N. 2012. "Risk-on/Risk-off, Capital Flows, Leverage and Safe Assets." BIS Working Paper No. 382.

Meese, R., and K. S. Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14 (1–2): 3–24.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf. 2012. "Carry Trades and Global Foreign Exchange Volatility." *Journal of Finance* 67 (2): 681–718.

———. 2016. "Information Flows in Foreign Exchange Markets: Dissecting Customer Order Flow." Working Paper.

Rime, D., L. Sarno, and E. Sojli. 2010. "Exchange Rates, Order Flow and Macroeconomic Information." *Journal of International Economics* 80 (1): 72–88.

Robinson, T., and H. Wang. 2013. "Changes to the RBA Index of Commodity Prices." *RBA Bulletin* (March): 23–28.

Rossi, B. 2013. "Exchange Rate Predictability." *Journal of Economic Literature* 51 (4): 1063–1119.

Sidek, N. Z. M., and M. B. Yussof. 2009. "An Empirical Analysis of the Malayan Ringitt Equilibrium Exchange Rate and Misalignment." *Global Economy and Finance Journal* 2 (2): 104–26.

Triplett, J. 1981. "Reconciling the CPI and the PCE Deflator." *Monthly Labor Review* (September): 3–15.

Verdelhan, A. 2015. "The Share of Systematic Risk in Bilateral Exchange Rates." Mimeo, Massachussetts Institute of Technology.

The Effects of Monetary Policy Announcements at the Zero Lower Bound*

Natsuki Arai
National Chengchi University

This paper investigates the effects of monetary policy announcements at the zero lower bound using Japanese data from 1998 to 2013. I find that the effect of expansionary monetary policy shocks is directly passed on to corporate bond yields, notably for high-grade corporate bond yields. However, the magnitude of estimated pass-through to stock prices and the exchange rate is substantially smaller than in the United States, and not statistically significant in most cases.

JEL Codes: E43, E44, E52, E58.

1. Introduction

The effect of unconventional monetary policy at the zero lower bound (ZLB), which includes forward guidance and asset purchases, has been a centerpiece of the debate in macro finance since many advanced economies reached the ZLB after the financial crisis of 2008. Since the crisis, a number of important contributions have

*This paper is based on chapter 1 of my dissertation submitted to Johns Hopkins University. I am grateful to Johns Williams, Jonathan Wright, and an anonymous referee for their extremely helpful comments and suggestions. I thank Larry Ball, Chris Carroll, Varanya Chaubey, Jon Faust, Yoichi Goto, Takashi Kano, Matt Raskin, Katsuhiro Oshima, Yuya Sasaki, Rodrigo Sekkel, Richard Spady, Ruli Xiao, Egon Zakrajsek, and seminar participants at Asian Development Bank, Bank of Canada, Central Bank of Philippines, Hitotsubashi University, International Monetary Fund, National Chengchi University, National Taiwan University, National Tsing Hua University, CFE, SNDE, and GTD conferences for their helpful comments. I gratefully acknowledge financial support from Taiwanese Ministry of Science and Technology's grant 104-2410-H-004-208. All errors are the sole responsibility of the author. Author e-mail: narai@nccu.edu.tw.

been made regarding this topic.[1] However, the analysis in the literature primarily focuses on the U.S. economy after the crisis, which is restricted by a short sample, and researchers are not sure about the effect of unconventional monetary policy in a different environment.

In this paper, I study the effects of unconventional monetary policy in Japan, which has experienced a substantially longer period at the ZLB, from 1995 to the present.[2] Specifically, I use the method of identification through heteroskedasticity, which was originally proposed by Rigobon (2003) and Rigobon and Sack (2003, 2004) and has been widely used in the recent literature,[3] to estimate the pass-through of monetary policy shocks. Identification is based on the assumption that the variance of monetary policy shocks is particularly high on important announcement days, whereas nothing unusual happens to other shocks on these days.

To assess the stimulative effect of monetary policy on aggregate demand, I focus on the pass-through of monetary policy shocks to three financial assets, which are of interest to central banks: corporate bonds, stocks, and the exchange rate. First, I study the pass-through to corporate bond yields, because the reduction in the borrowing cost of firms is a key channel through which monetary policy could stimulate aggregate demand. Second, I analyze the pass-through to stock prices, which is relevant because the response of stock prices could increase consumption through the wealth effect. Finally, I evaluate the pass-through to the exchange rate, through which aggregate demand can be boosted via the trade balance.

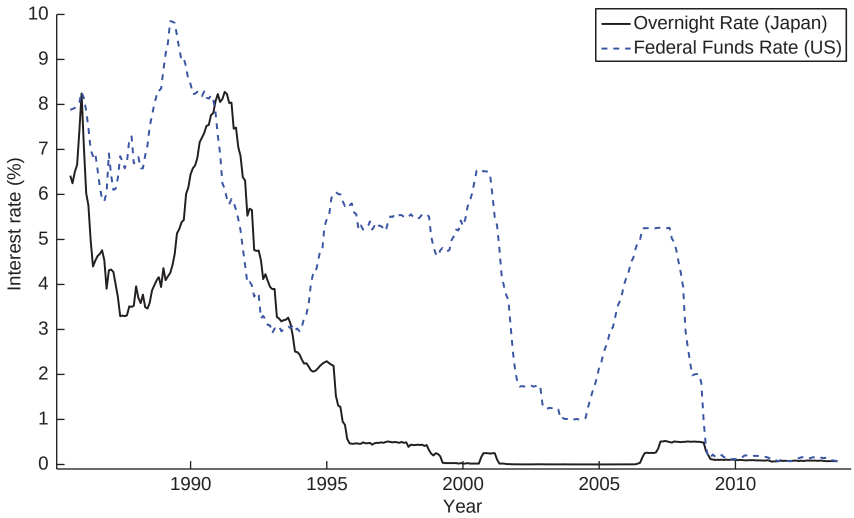
The results show that there is a stark contrast between the pass-through to corporate bond yields and the pass-through to stock prices and the exchange rate. For corporate bond yields, there is a statistically significant and about one-to-one pass-through, notably

[1]For example, see D'Amico and King (2013), Gagnon et al. (2011), Krishnamurthy and Vissing-Jorgensen (2011), and Wright (2012) for the Federal Reserve's large-scale asset purchase (LSAP) programs, and see Joyce et al. (2012) for the Bank of England's asset purchases. For more comprehensive review, see Bernanke (2012). For the comparison across advanced economies, see Rogers, Scotti, and Wright (2014).

[2]Figure 1 shows the comparison between policy rates in Japan and the United States.

[3]For example, see Gilchrist and Zakrajsek (2013) and Raskin (2013).

Figure 1. BOJ's Overnight Policy Rate and Federal Reserve's Federal Funds Rate from 1985 to 2013



for high-grade corporate bond yields. On the other hand, the pass-through to stock prices and the exchange rate is not statistically significant in most cases. However, these estimates in Japan are markedly different from the estimates in the United States based on the data from 2008 to the present. While I find one-to-one pass-through to corporate bond yields in the United States, which is broadly similar to Japan, the U.S. pass-through to stock prices and the exchange rate is statistically significant, and its magnitude is substantially larger than that of Japan.

In addition, I use a simple event study to analyze the effects of announcements in 2013, to show that these announcements have substantial effects even on stock prices. The announcements in 2013 are associated with the regime change of the Bank of Japan's (BOJ) monetary policy to commit to the 2 percent inflation target by 2015. Unlike the previous announcements, these announcements had substantial effects, not only on corporate bonds but also on stock prices. This difference may be due to the different nature of the BOJ's commitment after 2013; the commitment is open ended and the BOJ announces that it will do whatever it takes to achieve the target.

Lastly, I provide several robustness checks, to which the main results are generally robust. First, I analyze the pass-through to other financial assets: (i) real estate investment trusts (REITs), (ii) credit default swaps (CDSs), and (iii) the exchange rate of the OECD and Asian economies. Second, I estimate the pass-through based on a more selected set of announcements. Third, I consider subsamples focusing on different programs: 2001–06, 2006–10, and 2010–13. Fourth, I provide the analysis based on the principal component of government bond yields with different maturities. Last, I use alternative sets of non-announcement days.

The remainder of the paper is organized as follows: section 2 describes the methodology used in the paper, section 3 explains the data and background of Japanese monetary policy, section 4 presents the results, and section 5 concludes.

2. Method

This section describes an analytical framework to estimate the pass-through of monetary policy shocks to various financial assets. Based on the standard setup of two simultaneous equations, I present a simple event study and the framework of identification through heteroskedasticity.

2.1 Setup

Consider the system of two simultaneous equations between the change in the interest rate and the growth rate of the asset price, Δi_t and Δs_t. The notation follows Rigobon and Sack (2004):

$$\Delta i_t = \beta \Delta s_t + \gamma X_t + \varepsilon_t, \tag{1}$$

$$\Delta s_t = \alpha \Delta i_t + \delta X_t + \eta_t, \tag{2}$$

where X_t is a common exogenous shock that simultaneously affects both the interest rate and the asset price, ε_t is a monetary policy shock, and η_t is a shock to the asset price.

In this system, I primarily focus on estimating the parameter α because it indicates how much monetary policy shocks affect asset prices through the changes in the interest rate. However, the OLS estimate of the pass-through, α, is biased since both variables, Δi_t and Δs_t, are simultaneously determined in the system.[4]

2.2 Event Study

An event study is a simple way to estimate the pass-through using a directly measured monetary policy surprise. By picking the important announcements and regarding them as a complete surprise, we can use the corresponding changes in the asset prices to estimate the effect of monetary policy shocks. Gagnon et al. (2011) directly measured the effect of the Federal Reserve's large-scale asset purchases (LSAPs) by assuming that the announcements about the LSAP were complete surprises, and added up the changes on the announcement days. Though the event study is based on the strong assumption that no other material news came within the announcement window, it can provide useful benchmark results.

2.3 Identification through Heteroskedasticity

To obtain a consistent estimate of the pass-through under weaker assumptions, we employ a scheme called identification through heteroskedasticity, proposed by Rigobon (2003) and Rigobon and Sack (2003, 2004). Essentially, it uses the shift of the variances of endogenous variables between the announcement days and non-announcement days as instruments for the identification.

I introduce some notation and assumptions to describe this scheme of identification. First, I denote a subset of the policy announcement days as A and a subset of the non-announcement days as $\bar{A}$.[5] Second, I denote the number of announcement days and non-announcement days as T_A and $T_{\bar{A}}$, and thus the total number of days as $T \equiv T_A + T_{\bar{A}}$. Finally, I assume that the variance of monetary policy shocks is larger on the announcement days than on the non-announcement days, but the variances of other shocks are the same across these two sets of days. Under this assumption, the difference

[4] For the derivation of the OLS estimate and its bias, see the appendix.

[5] Unlike Rigobon and Sack (2004), all business days that do not belong to A are treated as the non-announcement days. To make this point clear, I denote a subset of the non-announcement days as $\bar{A}$. The results using alternative sets of the non-announcement days are presented in section 4.4.5.

of the conditional variance-covariance matrixes in these two sets of days, $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_{\bar{A}}$, only depends on the variance of monetary policy shocks. Specifically, we can compute the difference of the variances, $\Delta\mathbf{\Omega}$, as follows:

$$\Delta\mathbf{\Omega} \equiv \mathbf{\Omega}_A - \mathbf{\Omega}_{\bar{A}} = \frac{\sigma^2_{\varepsilon|A} - \sigma^2_{\varepsilon|\bar{A}}}{(1-\alpha\beta)^2} \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}, \tag{3}$$

where $\sigma^2_{\varepsilon|A}$ and $\sigma^2_{\varepsilon|\bar{A}}$ are the conditional variances of monetary policy shocks on the announcement days and the non-announcement days, respectively.[6] This is because the effect of other shocks cancels out by taking the difference between the announcement days and non-announcement days.

As discussed in Rigobon and Sack (2004), α can be estimated by using $\Delta\mathbf{\Omega}$ as the instruments for identification. To formalize the instruments, we first define endogenous variables. Let $\mathbf{\Delta i}_A$ and $\mathbf{\Delta s}_A$ be $T_A \times 1$ vectors of variables on the announcement days, and $\mathbf{\Delta i}_{\bar{A}}$ and $\mathbf{\Delta s}_{\bar{A}}$ be $T_{\bar{A}} \times 1$ vectors of variables on the non-announcement days. Then, we can combine these two vectors into $T \times 1$ vectors of endogenous variables:

$$\mathbf{\Delta i} \equiv [\,\mathbf{\Delta i}_A, \quad \mathbf{\Delta i}_{\bar{A}}\,]', \tag{4}$$

$$\mathbf{\Delta s} \equiv [\,\mathbf{\Delta s}_A, \quad \mathbf{\Delta s}_{\bar{A}}\,]'. \tag{5}$$

Given these endogenous variables, $\mathbf{\Delta i}$ and $\mathbf{\Delta s}$, instruments are constructed by normalizing with the number of days in each subset of days and by flipping the signs of the variables on the non-announcement days:

$$\mathbf{z_i} \equiv \left[\frac{1}{T_A}\mathbf{\Delta i}_A, \quad -\frac{1}{T_{\bar{A}}}\mathbf{\Delta i}_{\bar{A}}\right]', \tag{6}$$

$$\mathbf{z_s} \equiv \left[\frac{1}{T_A}\mathbf{\Delta s}_A, \quad -\frac{1}{T_{\bar{A}}}\mathbf{\Delta s}_{\bar{A}}\right]'. \tag{7}$$

It is easy to see that $\mathbf{z_i}$ and $\mathbf{z_s}$ are relevant and valid instruments to identify the pass-through in equation (2). First, these instruments

[6] For the derivation, see the appendix.

are correlated with the endogenous variables as long as the variances on the announcement days and non-announcement days are different. On the other hand, these instruments are uncorrelated with the shocks to the asset prices as presented in equation (3).[7]

In this paper, I use the orthogonality of both instruments as the moment conditions for GMM estimation. Though I can estimate the pass-through by IV estimation using just one instrument, as implemented in Rigobon and Sack (2004), GMM estimation should provide more efficient estimates. The moment conditions are described as follows:

$$E[f_t(\alpha)] = 0, \tag{8}$$

where

$$\begin{aligned} f_t(\alpha) &= Z_t \cdot e_t, \\ Z_t &= [z_{i,t},\, z_{s,t}]', \\ e_t &= \Delta s_t - \alpha \Delta i_t. \end{aligned}$$

The GMM estimate of α can be obtained by solving the minimum distance problem:

$$\alpha_{GMM} = \arg\min_{\alpha} f_T(\alpha)'\, W\, f_T(\alpha), \tag{9}$$

where $f_T(\alpha) = \sum_{t=1}^{T} f_t(\alpha)$ and W is an appropriate 2×2 weighting matrix. I use the two-step GMM for the estimation, in which I first use the identity matrix as a weighting matrix to solve the minimization problem, and then use the inverse of estimated variance-covariance matrix of the moment conditions in the first step as a weighting matrix in the second step. Inference for the GMM estimation is based on heteroskedasticity-robust standard errors.

3. Data and Background

This section describes the data and background to show that there is enough variation to identify the pass-through of monetary policy shocks. After describing the data, I provide a brief summary of

[7]For the proof of the orthogonality, see the appendix.

Japanese monetary policy. Then, I explain the selection of important announcements and provide a statistical analysis showing that the selection is valid in terms of identification.

3.1 Data

I estimate the pass-through of monetary policy shocks to three financial assets: (i) corporate bond yields, (ii) stock prices, and (iii) the exchange rate. The analysis is based on daily data from April 1998 to December 2013, all of which are obtained from Bloomberg.

The details of the series are described as follows:

(i) Japanese government bond (JGB) yield: generic yield with a maturity of five, ten, and twenty years;

(ii) Corporate bond yield: the Bloomberg Fair Value (BFV) indexes of AA and BBB corporate bond yields for the industrial sector, with a maturity of five and ten years;[8]

(iii) Stock prices: Nikkei 225;

(iv) Exchange rates: spot exchange rates of the U.S. dollar, measured by the Japanese yen.

I compute the daily changes in levels for the JGB and corporate bond yields, and the continuously compounding rate of daily change for stock prices and the exchange rate. The data between March 11 and March 18, 2011 are excluded from the analysis to eliminate the effect of the earthquake in March 2011.

3.2 Brief Summary of Japanese Monetary Policy

The BOJ's overnight interest rate has been reduced to nearly zero for approximately two decades, and three different programs of unconventional monetary policies have been implemented. Table 1 summarizes the timeline of the important events in Japanese monetary policy, and figure 2 shows the volume of monetary base.[9]

---

[8]The AA index and the BBB index are available from June 8, 1999 and January 30, 2003, respectively. The BBB index with a maturity of ten years is not available from February 6, 2012.

[9]For more comprehensive accounts of the BOJ's monetary policy, see Ito and Mishkin (2006) and Ueda (2012a).

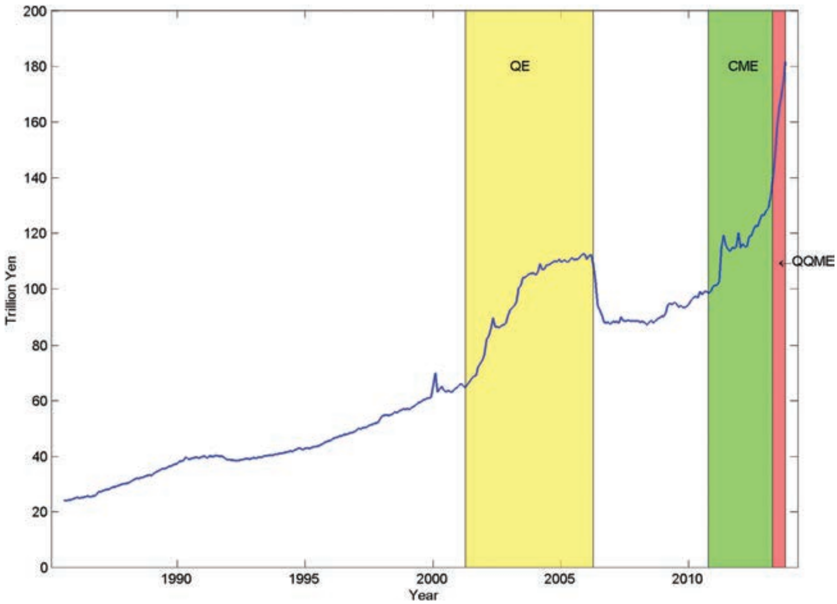
Table 1. Timeline of BOJ's Monetary Policy

Date	Event	Policy Rate (%)	Governor
4/14/1995		1.0	Matsushita
9/8/1995		0.5	
4/1/1998	Revised BOJ Act came into effect	0.5	Hayami
2/12/1999		0	
8/11/2000		0.25	
3/19/2001	Quantitative easing (QE) launched	0	
3/9/2006	QE terminated	0	Fukui
7/14/2006		0.25	
2/21/2007		0.5	
10/31/2008		0.3	Shirakawa
12/19/2008	JGB purchase increased	0	
10/5/2010	Comprehensive monetary easing (CME) launched	0	
4/4/2013	CME terminated, and quantitative and qualitative monetary easing (QQME) launched	0	Kuroda

The first program was called "quantitative easing" (QE) and was implemented from March 2001 to March 2006 under Governors Hayami and Fukui. Under the QE program, the BOJ set its current account balance as the main policy target and purchased long-term JGBs to achieve this target.[10] The QE program was terminated in March 2006, and the overnight interest rate was gradually raised to 0.5 percent. However, in response to the global financial crisis in 2008, the BOJ reduced the overnight interest rate to zero again.

[10]For an analysis of the QE program, see Shiratsuka (2010) and Ugai (2007). More recently, Shibamoto and Tachibana (2013) analyze the effects of the QE program using the method of identification through heteroskedasticity. Kimura and Nakajima (2013) use a variant of regime-switching structural VAR using the data from 1981 to 2012, with ad hoc shrinkage in certain parameters.

Figure 2. Monetary Base from 1985 to 2013



The second program was called "comprehensive monetary easing" (CME) and was implemented from October 2010 to April 2013 under Governor Shirakawa. Under the CME program, the BOJ purchased not only JGBs but also commercial paper and risky assets such as exchange-traded funds (ETFs) and REITs, while also making a commitment to keep the policy rate at zero. This is the policy prescription proposed by Eggertsson and Woodford (2003). In addition, the BOJ provided various forms of lending programs to financial institutions.

Most recently, the BOJ substantially expanded its asset purchase program and launched it as "quantitative and qualitative monetary easing" (QQME). The QQME program has been implemented since April 2013 under Governor Kuroda. Under the QQME program, the BOJ commits to achieving 2 percent inflation by 2015. To achieve this goal, the BOJ announced that it would increase the monetary base at a pace of 60 to 70 trillion yen per year and extend the average maturity of its JGB holdings from three years to seven years.[11]

[11]For details, see Kuroda (2013).

3.3 Selection of Monetary Policy Announcements

The selection of the important announcements is crucial for the analysis in this paper, and I select forty-one announcement days from April 1998 to July 2013 based on Ueda (2012b). The exact dates and overview are listed in table 2. These dates are associated with the BOJ's "official" change in its monetary policy.[12]

In addition, I include other dates when strong signals concerning future changes in BOJ's monetary policy were made. For example, I include the days when the new BOJ governor was nominated and the confirmation hearing was held by the National Diet in 2013. On the other hand, I do not include other meeting days or the days of speeches by the BOJ governor or other board members. This is because including trivial or indirect news announcement days will make the distinction between the announcement days and the non-announcement days unclear, which undermines the identification as discussed in Wright (2012).

3.4 Standard Deviations of the Series

To see whether the actual data are consistent with the assumptions for the identification, table 3 compares the standard deviations of the daily changes on the announcement days and non-announcement days. Consistent with the assumptions, the standard deviations are higher on the announcement days than on the non-announcement days for almost all series, except for the BBB yield with a maturity of ten years.

To test whether the variances in these two sets of days are significantly different, I conduct three statistical tests: the F-test, the block bootstrap, and the stationary bootstrap. All tests are based on the null hypothesis that the population variance in these two sets of days are equal. Since the F-test assumes that each observation

[12]I exclude two dates from the list in Ueda (2012b)—April 13, 1999 and March 14, 2011. First, I exclude April 13, 1999 because the commitment to keep the policy rate at zero made by the BOJ governor is not entirely clear. Ueda (2012b) also notes that the market reacted to this event very slowly. However, including this date does not materially change the results. Second, I exclude March 14, 2011 since this is the meeting right after the earthquake.

Table 2. Dates of BOJ's Monetary Policy Announcements

Date	Event	Summary
9/9/1998	BOJ Statement	Policy rate reduced to 0.25 percent
2/12/1999	*BOJ Statement*	*Policy rate reduced close to zero*
8/11/2000	BOJ Statement	Policy rate raised to 0.25 percent
*3/19/2001**	*BOJ Statement*	*Quantitative easing (QE) launched (policy rate close to zero)*
*8/14/2001**	*BOJ Statement*	*QE expanded*
*12/19/2001**	*BOJ Statement*	*QE expanded*
*2/28/2002**	*BOJ Statement*	*JGB purchase increased*
9/18/2002*	BOJ Statement	Stock purchase announced
*10/30/2002**	*BOJ Statement*	*QE expanded*
3/25/2003*	BOJ Statement	Stock purchase expanded
4/8/2003*	BOJ Statement	ABS purchased announced
*4/30/2003**	*BOJ Statement*	*QE expanded*
*5/20/2003**	*BOJ Statement*	*QE expanded*
*10/10/2003**	*BOJ Statement*	*QE expanded*
*1/20/2004**	*BOJ Statement*	*QE expanded*
*3/9/2006**	*BOJ Statement*	*QE terminated*
7/14/2006	BOJ Statement	Policy rate raised to 0.25 percent
2/21/2007	BOJ Statement	Policy rate raised to 0.5 percent
9/18/2008*	BOJ Statement	Dollar swap
10/31/2008	BOJ Statement	Policy rate reduced to 0.3 percent
12/2/2008*	BOJ Statement	Facilitation of corporate finance
*12/19/2008**	*BOJ Statement*	*JGB purchase increased (policy rate reduced close to zero)*
2/3/2009*	BOJ Statement	Stock purchase restarted
*3/18/2009**	*BOJ Statement*	*JGB purchase increased*
12/1/2009*	BOJ Statement	Fixed-rate three-month operation
12/18/2009*	BOJ Statement	"Inflation target" clarified
3/17/2010*	BOJ Statement	Fixed-rate operation expanded
5/21/2010*	BOJ Statement	Growth-enhancing operation
8/30/2010*	BOJ Statement	Fixed-rate six-month operation
*10/5/2010**	*BOJ Statement*	*Comprehensive monetary easing (CME) launched*
8/4/2011	*BOJ Statement*	*CME expanded*
10/27/2011	*BOJ Statement*	*CME expanded*
2/14/2012	*BOJ Statement*	*CME expanded*
4/10/2012	*BOJ Statement*	*CME expanded*
4/27/2012	*BOJ Statement*	*CME expanded*
10/30/2012	*BOJ Statement*	*CME expanded and the joint statement with the government issued*
12/20/2012	*BOJ Statement*	*CME expanded*
1/22/2013	*BOJ Statement*	*CME expanded and inflation target clarified*
2/25/2013	*Nomination of new governor*	
3/4/2013	*Confirmation hearing from new governor at National Diet*	
4/4/2013	*BOJ Statement*	*Quantitative and qualitative monetary easing (QQME) launched*

Notes: Dates with * are listed in table 2 of Ueda (2012b). The announcement days that are directly related to the JGB purchase and the QE programs are in italics.

Table 3. Standard Deviations and Variance Ratio on Announcement Days and Non-announcement Days

Series	Announcement	Non-announcement	Variance Ratio
JGB			
5 Year	3.18	2.85	1.25
10 Year	4.28	3.28	1.70**†
20 Year	4.88	3.55	1.89**†
Corporate Bond Yield			
AA, 5 Year	3.50	2.68	1.71**‡
AA, 10 Year	3.92	2.88	1.85**‡
BBB, 5 Year	3.64	2.74	1.76**‡
BBB, 10 Year	3.21	3.13	1.06
Stock Prices			
Nikkei 225	2.19	1.56	1.96**†
TOPIX	2.02	1.41	2.07**†
Exchange Rates			
U.S. Dollar	1.06	0.71	2.22**‡
Euro	1.24	0.84	2.17**‡

Notes: This table compares the standard deviation of daily changes on announcement days and non-announcement days. Daily changes of the yield in basis points and daily percent changes of the stock prices and exchange rates are used. * and ** denote significance at the 10 percent and 5 percent level, respectively, based on the F-test. † and ‡ denote significance at the 10 percent and 5 percent level, respectively, based on the block bootstrap with 10,000 replications. The stationary bootstrap leads to similar results as the block bootstrap.

is independent, I use the block bootstrap and the stationary bootstrap to take the heteroskedasticity observed in the daily data into account. In the block bootstrap, I construct an artificial sample by resampling the block of ten days to preserve heteroskedasticity. On the other hand, in the stationary bootstrap, the length of the block is randomly determined.[13] After constructing an artificial sample, I compute the variance ratio in the artificial sample. By repeating this exercise, I can form the bootstrap distribution of the variance

[13]More specifically, an artificial sample is constructed by the Bernoulli trial, either to pick a random sample or the sample on the next day. I set the probability of the latter as 0.9 to make the expected length of the block ten days.

ratio and compute the p-values based on the percentile of the sample variance ratio.

The null hypothesis is rejected for most series, with the sample variance ratio larger than one, except for the ten-year BBB corporate bond yields. Therefore, we could conclude that the variance is significantly larger on the announcement days. The F-test tends to reject the null hypothesis more often than the other two tests based on the bootstrap.

4. Results

In this section, I present GMM estimates showing that the pass-through of monetary policy shocks is one-to-one for corporate bond yields, but it is smaller than the U.S. estimates for stock prices and the exchange rate. In addition, I provide an event study showing that the announcements in 2013 have substantial effects on asset prices. Lastly, I present several robustness checks using additional variables and subperiods to confirm the main results.

4.1 Event Study

Table 4 presents the results of the event study focusing on the QQME program in 2013. The results show that the announcements of the QQME program led to a substantial decline in the long-term JGB and corporate bond yields. Furthermore, stock prices substantially increased on these announcement days, but the effect on exchange rates was mixed.

The results show that the long-term JGB and corporate bond yields substantially declined responding to the announcements. Specifically, the JGB yields declined 11.4 (ten years) and 17.7 (twenty years) basis points on April 4, and the cumulative decline on all announcements was 14.0 (ten years) and 23.3 (twenty years) basis points, respectively. Corporate bond yields also declined on these days, but the magnitude is smaller than JGB yields. The ten-year AA bond yield declined 9.72 basis points on April 4, and the cumulative decline was 10.65 basis points.[14]

[14] The effects for yields with shorter maturities are trivial, primarily because they are all stuck at the zero lower bound.

Table 4. Effects of the QQME Announcements on JGB, Corporate Bond Yields, Stock Prices, and Exchange Rates

Date	Event	JGB 5 Yr.	JGB 7 Yr.	JGB 10 Yr.	JGB 20 Yr.	AA 5 Yr.	AA 10 Yr.	BBB 5 Yr.	Nikkei 225	TOPIX	U.S. Dollar	Euro
2/25/2013	*Nomination*	−1.30	−2.00	−2.20	−2.00	−1.37	−1.71	−0.73	2.40	1.77	−1.73	−2.71
3/4/2013	*Confirmation Hearing*	−1.90	−2.30	−4.40	−7.30	−0.69	0.88	−0.96	0.30	0.11	0.79	1.87
4/4/2013	*BOJ Statement*	−1.20	−5.50	−11.40	−17.70	−2.08	−9.72	−1.65	2.18	2.67	3.49	4.16
4/26/2013	BOJ Statement	1.50	1.40	0.80	−0.70	3.02	−0.71	1.83	−0.30	−0.99	−1.23	−1.10
5/22/2013	BOJ Statement	−1.20	0.50	1.20	1.70	−1.04	1.27	−1.02	1.59	0.44	0.64	0.26
6/11/2013	BOJ Statement	4.20	6.30	4.70	4.50	5.39	5.57	5.19	−1.47	−0.98	−2.80	−2.39
7/11/2013	BOJ Statement	−0.80	−3.80	−2.70	−1.80	−3.94	−2.61	−2.06	0.39	−0.04	−0.72	0.20
Baseline Events		−4.40	−9.80	−18.00	−27.00	−4.14	−14.17	−3.76	4.97	5.24	1.64	1.37
Cumulative		−0.70	−5.40	−14.00	−23.30	−0.71	−10.65	0.18	5.18	3.68	−2.47	−1.65
Average before 2013 (Negative Surprise)		−2.31	−2.48	−2.12	−1.70	−2.14	−1.75	−2.11	−0.61	−0.54	−0.15	0.11
Average before 2013 (Positive Surprise)		2.13	3.45	2.73	2.77	2.03	2.52	1.59	0.24	0.29	0.12	0.36

Notes: This table shows the daily changes of the yields and the daily percent changes of stock prices and exchange rates on the days listed above. Baseline events are listed in table 2 and are in italics. The table also shows the average effects using the pre-2013 data when there are positive/negative changes in the five-year JGB yields as a reference.

The results also show that stock prices substantially increased on the announcement days. The cumulative increase of the Nikkei 225 and TOPIX indexes was 5.18 percent and 3.68 percent, respectively. On the other hand, the effect on exchange rates was mixed. Even though the announcement on April 4 led to a substantial depreciation of the Japanese yen, 3.49 percent for the U.S. dollar and 4.16 percent for the euro, this effect was quickly offset on subsequent announcements.

4.2 GMM Estimates of Pass-Through

Table 5 presents the pass-through to corporate bond yields, stock prices, and the exchange rate. The results show that there is a statistically significant and nearly one-for-one pass-through to corporate bond yields, notably for high-grade bond yields. On the other hand, the pass-through to stock prices and the exchange rate is negative, but the estimated magnitude is quite small and not statistically significant in most cases. It is important to note that this analysis estimates the *average* effects of unconventional policies, and these estimates cannot be interpreted as the effects of individual policies, forward guidance, or asset purchases.

I present the estimates of the pass-through to different asset prices based on an individual JGB yield. This is because monetary policy shocks primarily influence the overall level of the JGB yields, and the change in JGB yield with any maturity should reflect such shocks.

4.2.1 Corporate Bond Yields

For the AA corporate bond yields, most of the pass-through is statistically significant, with a magnitude from 0.84 to 0.99. It implies that an expansionary monetary policy shock, which lowers the twenty-year JGB yield by 100 basis points, will lower the ten-year AA corporate bond yield 84 basis points. All estimates are smaller than one and statistically significant.

On the other hand, the estimates of the pass-through to the medium-grade corporate bond yields varies; the estimated magnitude is from –2.90 to 1.12. The estimates are positive in most cases, which suggests that monetary policy shocks are passed on to BBB

Table 5. GMM Estimates of the Pass-Through to Corporate Bond Yields, Stock Prices, and the Exchange Rate

	Japan		United States	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
A. Pass-Through from 5-Year Government Bonds				
AA Corporate Bond	0.99**	(0.25)	0.70**	(0.07)
BBB Corporate Bond	1.12	(1.13)	0.61**	(0.09)
Stock Prices	−1.19	(1.59)	−8.07**	(3.81)
Exchange Rate	−0.40	(0.83)	−7.52**	(0.93)
B. Pass-Through from 10-Year Government Bonds				
AA Corporate Bond	0.94**	(0.05)	0.63**	(0.07)
BBB Corporate Bond	1.09	(1.24)	0.58**	(0.07)
Stock Prices	−0.20	(0.17)	−7.19*	(3.74)
Exchange Rate	−0.20	(0.16)	−6.57**	(0.87)
C. Pass-Through from 20-Year Government Bonds				
AA Corporate Bond	0.84**	(0.21)	0.73**	(0.33)
BBB Corporate Bond	−2.90	(14.96)	0.63**	(0.31)
Stock Prices	−0.12	(0.08)	−10.80*	(6.15)
Exchange Rate	−0.20**	(0.09)	−9.27**	(3.09)

Notes: This table shows the pass-through to the corporate bond yields, stock prices, and exchange rates. * and ** denote significance at the 10 percent and 5 percent level, respectively. Heteroskedasticity-robust standard errors are presented in parentheses. For Japan, the corporate bond yield with the maturity of five years is used for the pass-through from five-year government bonds, and the corporate bond yield with the maturity of ten years is used for the pass-through from ten- and twenty-year government bonds. The Nikkei 225 index is used for stock prices, and the spot exchange rate against the U.S. dollar is used for the exchange rate. For the United States, the effective corporate bond yield is used for all maturities. The Dow Jones Industrial Average index is used for stock prices, and the spot exchange rate against the euro is used for the exchange rate.

corporate bond yields to some extent. However, no estimate is statistically significant because of the large standard errors.

4.2.2 Stock Prices

The estimates of the pass-through to stock prices are mostly negative, ranging between –1.19 and –0.12. These estimates imply that an expansionary monetary policy shock, which lowers JGB yields by 100 basis points, increases stock prices by 0.12 to 1.19 percent. However, few of them are statistically significant. In other words, the estimated magnitude of the pass-through to stock prices is so small that the estimates are not significantly different from zero in most cases.

4.2.3 Exchange Rates

The results show that the pass-through to the exchange rate is mostly negative, with the magnitude between –0.4 and –0.2. These estimates imply that an expansionary monetary policy shock, which lowers JGB yields by 100 basis points, leads a depreciation of the Japanese yen by 0.2 percent to 0.4 percent.[15] However, only the pass-through from the twenty-year JGB yield, –0.2, is statistically significant.

These results are consistent with the prediction of conventional interest rate parity, which suggests that the expected return on domestic assets should be the same as the exchange-rate-adjusted expected return on foreign assets. In other words, a decline in the JGB yield should be adjusted by a depreciation of the Japanese yen, which would boost the stimulative effect of expansionary monetary policy shocks. However, the estimated magnitude of the pass-through is quite small and few estimates are statistically significant.

4.3 Discussion of the Results

Given these results, I first provide the comparison between Japanese and U.S. results. Though Japanese results are similar to the U.S. results in their signs, the estimated magnitude of pass-through is considerably smaller for stock prices and the exchange rate.

4.3.1 Comparison with the U.S. Estimates

Table 5 also presents the estimates of pass-through in the United States using the same methodology. For corporate bond yields,

[15]Since I use the exchange rate that measures the value of non-Japanese currency by the Japanese yen, a rise in the exchange rate implies depreciation of the Japanese yen.

Japanese and U.S. estimates are similar in their signs and magnitudes. On the other hand, the magnitude of the Japanese estimates for stock prices and the exchange rate is substantially smaller than that of the U.S. estimates, though the sign of Japanese estimates is consistent with the predictions of economic theories.

To measure the monetary policy shock in the United States, I used the daily data of zero-coupon bond yields computed by Gürkaynak, Sack, and Wright (2007) with the maturities of five, ten, and twenty years. To estimate the pass-through to corporate bond yields, stock prices, and the exchange rate, I used Bank of America Merrill Lynch effective corporate bond yields with AA and BBB ratings, the Dow Jones Industrial Average, and the spot exchange rate against the euro. I used the sample from December 2008 to April 2015, when the U.S. economy was at the ZLB.

For corporate bond yields, the pass-through to corporate bond yields is broadly similar in Japan and in the United States. Similar to the Japanese case, the pass-through to high-grade bond yields with the rating of AA is statistically significant, and its magnitude is close to one, which is consistent with Raskin's (2013) findings.[16] In addition, I found a significantly positive pass-through to the BBB bond yields in the United States.

For stock prices, the pass-through is negative in both Japan and the United States, but the estimated magnitudes are substantially smaller in Japan. While the Japanese estimates are marginally negative and statistically insignificant, the U.S. estimates are between –7.19 and –10.80 and are statistically significant.[17] Though these estimates are slightly larger than the estimates Bernanke and Kuttner (2005) and Gürkaynak, Sack, and Swanson (2005) obtained

[16]To explain the limited pass-through to the medium-grade bond yields found in Krishnamurthy and Vissing-Jorgensen (2011), Krishnamurthy and Vissing-Jorgensen (2012) propose the safety premium, which is the premium investors pay to satisfy their unique demand for safe long-term assets. Since the medium-grade corporate bonds are not regarded as safe assets, the decline of the government bond yields does not affect their yields. Krishnamurthy and Vissing-Jorgensen (2011) argue that the effect of the Federal Reserve's LSAP announcements is primarily due to reduction in the safety premium.

[17]These estimates are larger than Kiley's (2013) estimates at the ZLB, which range from –3.0 to –1.5. This difference is likely due to a different methodology of identification, since Kiley (2013) uses intraday data and focuses on a much narrower window around the announcements.

before the Great Recession, they are consistent with recent findings in Claus, Claus, and Krippner (2014): Using the estimated shadow short rate, they found that the monetary policy surprise has a larger impact on asset prices at the ZLB period and suggest some kind of structural change in the monetary transmission mechanism at the ZLB. Similar to the Japanese situation, Rosa (2012) documents that the Bank of England's announcements about gilt purchases do not have statistically significant effects on stock prices.

For the exchange rate, the magnitude of Japanese estimates, –0.2, is also substantially smaller than the U.S. estimates, even though the sign of pass-through is negative in both countries. The U.S. estimates are between –6.57 and –9.27 and highly statistically significant. These estimates are largely consistent with the findings in Glick and Leduc (2013) and Neely (2013), which report that the pass-through to the exchange rate during the Great Recession is around –3 to –4. The difference in the magnitude could be due to the difference in the methodology and the usage of intraday data.

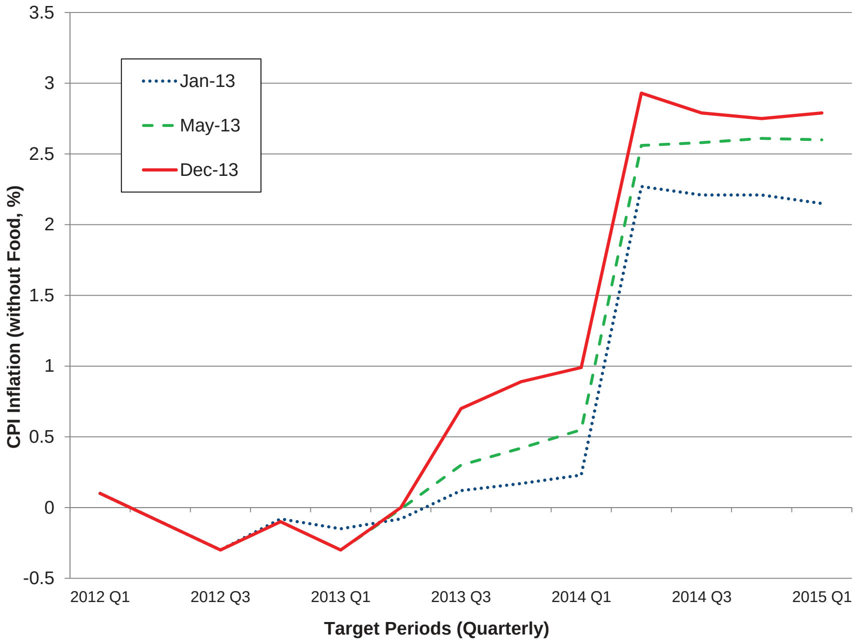
4.3.2 Comparison of Announcements Before and After 2013

The comparison of the event study and GMM show that the announcements in 2013 had substantial effects, not only on corporate bonds but also on stock prices. It may be because the announcements in 2013 were associated with the BOJ's regime change, in which the BOJ made an open-ended commitment to achieve the inflation target, creating a huge surprise.

Before 2013, a number of papers, including Ito and Mishkin (2006), are critical of the passive stance of the BOJ.[18] However, Romer (2014) describes the policy change in 2013 as "an honest-to-goodness regime shift" and says, "(the BOJ) took dramatic actions and pledged convincingly to do whatever it takes to end deflation." Consistent with her argument, some survey measures show that inflation expectations have substantially increased in 2013. For example, figure 3 shows the upward shift of the inflation expectation

[18]Based on a New Keynesian model with two different regimes (a targeted-inflation regime and a deflation regime), Aruoba, Cuba-Borda, and Schorfheide (2014) provide an interpretation that the BOJ's passive stance has triggered an adverse shock that moves the economy into the deflation regime.

Figure 3. ESP Inflation Forecast in 2013 (target periods of 2013–15)



in the mean of the ESP Forecast (survey of professional forecasters in Japan) in 2013, which could be up to 0.8 percent. Hausman and Wieland (2014, 2015) also provide similar analysis and conclude that the set of policy package called "Abenomics," which includes the introduction of the QQME program in 2013, raised long-run inflation expectation.[19]

However, even though the announcements in 2013 have had substantial effects on stock prices, the GMM estimates in this paper are not statistically significant, since I have estimated an average effect over the full sample from 1998 to 2013. As presented in table 4, the average effects of monetary policy announcements based on the pre-2013 event study have been much smaller than those in 2013. This

[19]On the other hand, Fujiwara, Nakazono, and Ueda (2014) argue that there is no significant increase in inflation expectation at the ten-year horizon after the introduction of the QQME, and the effect of monetary policy is quite limited.

finding is consistent with those of other studies in the literature. For example, Ueda (2012b) conducted an event study of monetary policy announcements between 1999 and 2011 and showed that the announcements regarding the QE programs lowered the JGB and corporate bond yields but did not significantly affect stock prices or exchange rates. Lam (2011) conducted an event study from 2008 to 2011, which covered part of the CME program, and obtained similar results as Ueda (2012b).

4.4 Robustness Checks

In this section, I provide several robustness checks, to which the main results are generally robust. First, I analyze the pass-through to other financial assets. Second, I conduct the analysis based on the set of selected announcements. Third, I consider subsamples focusing on different programs. Fourth, I provide the analysis based on the principal component of JGB yields. Last, I use alternative sets of non-announcement days.

4.4.1 Additional Variables

I extend the analysis using three kinds of additional variables: (i) REITs, (ii) CDS index, and (iii) exchange rates in other OECD and Asian economies.[20] These variables are of interest for slightly different reasons. First, REITs are of interest because the BOJ purchased them in the QE programs. Second, the analysis of the CDS index will shed some light on the effects of the QE programs on credit default risk as discussed in Gilchrist and Zakrajsek (2013). Lastly, I focus on exchange rates relative to these countries because the so-called yen carry trade, in which investors borrow money in the Japanese yen and invest in high-interest rate currencies, has been prominent since the late 1990s. These high-yield currencies include the Australian and New Zealand dollars and other emerging Asian currencies.[21]

[20]The REIT index is available from April 2003.

[21]For details, see Hattori and Shin (2008).

Table 6. Standard Deviation and Variance Ratio of the Series on Announcement Days and Non-announcement Days

Series	Announcement	Non-announcement	Variance Ratio
Financial Assets			
REIT Index	2.14	1.52	1.99**†
CDS Index	5.30	4.19	1.60*
Exchange Rates:			
OECD Countries			
Australian Dollar	1.26	1.12	1.26
Canadian Dollar	1.22	0.94	1.67**†
Korean Won	1.34	1.07	1.56**
New Zealand Dollar	1.29	1.10	1.38*
U.K. Pound	1.20	0.85	2.01**†
Exchange Rates:			
Asian Economies			
Hong Kong Dollar	1.06	0.71	2.26**‡
Singapore Dollar	0.90	0.69	1.71**‡
Taiwanese Dollar	1.01	0.75	1.82**‡
Thai Baht	0.97	0.80	1.49**

Notes: This table compares the standard deviation of daily changes on announcement days and non-announcement days for additional variables: REITs, CDSs, and the OECD and Asian exchange rates. Technical notes are the same as table 3.

Table 6 presents the standard deviations and variance ratios of these variables. Consistent with the assumptions of the identification, the standard deviations are higher on the announcement days than on the non-announcement days for all series. In addition, most of these differences are statistically significant.

Table 7 shows the GMM estimates of the pass-through to the REITs, the CDS index, and exchange rates using the baseline announcement days. For the REITs, the pass-through is mostly negative and some of them are statistically significant. The negative pass-through implies that an expansionary monetary policy shock led to the increase in the price of these financial assets. The magnitude of the pass-through to the REITs is between –0.46 and –0.16. The pass-through to the CDS index is positive, but the magnitude

Table 7. GMM Estimates of the Pass-Through to Additional Variables

	5-Year JGB		10-Year JGB		20-Year JGB	
Pass-Through from	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
REIT Index	−0.43	(0.28)	−0.36*	(0.19)	−0.30**	(0.12)
CDS Index	1.09	(0.82)	0.77	(0.53)	0.30	(0.23)
Australian Dollar	0.07	(0.46)	−0.16	(0.14)	−0.14*	(0.08)
Canadian Dollar	−0.84	(1.44)	−0.24	(0.20)	−0.21**	(0.11)
Korean Won	−0.78	(1.23)	−0.18	(0.17)	−0.18**	(0.11)
New Zealand Dollar	−0.34	(0.75)	−0.21	(0.19)	−0.17**	(0.09)
U.K. Pound	0.03	(0.36)	−0.15	(0.18)	−0.20*	(0.11)
Hong Kong Dollar	−0.51	(1.02)	−0.21	(0.17)	−0.20**	(0.10)
Singapore Dollar	−0.04	(0.19)	−0.15	(0.13)	−0.15**	(0.07)
Taiwanese Dollar	0.02	(0.23)	−0.10	(0.12)	−0.15*	(0.08)
Thai Baht	−0.07	(0.28)	−0.15	(0.15)	−0.15**	(0.08)

Notes: This table shows the pass-through to additional variables: REITs, CDSs, and the OECD and Asian-Pacific exchange rates. * and ** denote significance at the 10 percent and 5 percent level, respectively. Heteroskedasticity-robust standard errors are presented in parentheses.

varies substantially with the maturity of the JGB yields and none are statistically significant. The pass-through to exchange rates in the OECD and Asian economies is negative in almost all cases, with a magnitude between –0.14 and –0.22. The estimates are similar to the estimates for the U.S. dollar and the euro, which suggests that the expansionary monetary policy shocks led to a depreciation of the Japanese yen relative to all other currencies. However, the magnitude of the depreciation is quite small.

4.4.2 Selected Announcements

In order to shed some light on the effects of the asset purchases, I have conducted an analysis using a selected set of announcements explicitly related to asset purchase programs, both in Japan and the United States.[22] Though estimating the effects of individual policies is extremely difficult because the announcements often contain information about different policies, the results based on the selected set of announcements could still serve as a useful benchmark.

As presented in table 8, the results are broadly similar to the results of the baseline case: The pass-through to corporate bond yields is statistically significant, and its magnitude is close to one in most cases. On the other hand, the pass-through to stock prices and the exchange rate is mostly negative, but its magnitude is small and not statistically significant. In addition, the estimated magnitude of the U.S. pass-through is substantially larger than that of Japan for stock prices and the exchange rate and is statistically significant in most cases.

I also report the weak-identification robust confidence sets in the appendix to determine if the differences in the variances on the announcement and non-announcement days are large enough for identification.[23] While the weak-instrument robust confidence sets often fail to identify the relevant confidence sets in the baseline case, they do identify the relevant confidence sets in many cases for

[22]The selection of U.S. monetary policy announcements is based on Rogers, Scotti, and Wright (2014).

[23]I use two statistics presented in the appendix to derive weak-instrument robust confidence sets: the S statistic in Stock and Wright (2000), an extension of Anderson and Rubin's (1949) statistic to GMM, and the K statistic in Kleibergen (2005).

Table 8. GMM Estimates of the Pass-Through to Corporate Bond Yields, Stock Prices, and the Exchange Rate (selected announcements)

	Japan		United States	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
A. Pass-Through from 5-Year Government Bonds				
AA Corporate Bond	0.41	(0.26)	0.73**	(0.05)
BBB Corporate Bond	0.12	(0.22)	0.70**	(0.04)
Stock Prices	0.06	(0.26)	−7.77*	(4.12)
Exchange Rate	−0.14	(0.11)	−8.13**	(1.01)
B. Pass-Through from 10-Year Government Bonds				
AA Corporate Bond	0.78**	(0.18)	0.69**	(0.03)
BBB Corporate Bond	0.86**	(0.21)	0.65**	(0.02)
Stock Prices	−0.24	(0.35)	−7.48*	(4.23)
Exchange Rate	0.09	(0.08)	−7.27**	(0.90)
C. Pass-Through from 20-Year Government Bonds				
AA Corporate Bond	0.78**	(0.19)	1.23**	(0.19)
BBB Corporate Bond	0.81**	(0.15)	1.17**	(0.18)
Stock Prices	−0.39	(0.25)	−13.21*	(7.99)
Exchange Rate	0.05	(0.07)	−11.31**	(2.05)

Notes: This table shows the pass-through to the corporate bond yields, stock prices, and exchange rates, based on a selected set of announcements. Technical notes are the same as table 5.

the selected announcements. This result suggests that eliminating less-relevant announcements greatly improves identification.

4.4.3 Analysis of Subsamples

To analyze how different QE programs affect corporate bond yields and asset prices, I conduct the same analysis in three subsamples focusing on the different QE programs: (i) the QE program (2001–06), (ii) the period between the QE and the CME programs (2006–10), and (iii) the CME program (2010–13). The results are presented in table 9. The period before the QE program (1998–2001) and the QQME program (2013–) are excluded since the number of announcements are too small.

Table 9. GMM Estimates of the Pass-Through in Subperiods

	2001–06 (QE)	**2006–10**	**2010–13 (CME)**
A. Pass-Through from 5-Year JGB			
AA Corporate Bond	1.02**	1.03**	1.99
BBB Corporate Bond	1.04**	1.13**	1.81
Nikkei 225	−0.66	−0.43	0.52
U.S. Dollar	−0.04	−0.05	8.19
B. Pass-Through from 10-Year JGB			
AA Corporate Bond	0.98**	0.98**	2.26**
BBB Corporate Bond	0.60*	0.77**	1.98**
Nikkei 225	−0.05	1.59	0.41
U.S. Dollar	−0.03**	−0.20	−0.03
C. Pass-Through from 20-Year JGB			
AA Corporate Bond	0.97**	0.98**	0.71**
BBB Corporate Bond	0.44**	1.22**	1.08**
Nikkei 225	−0.06	0.89	0.12
U.S. Dollar	−0.02	−0.19	0.06
Number of Announcements	13	13	9

Notes: This table shows the pass-through to the coprorate bond yields, stock prices, and exchange rates in the subsamples: 2001–06 (the QE program), 2006–10, and 2010–13 (the CME program). * and ** denote significance at the 10 percent and 5 percent level, respectively. The corporate bond yield with the maturity of five years is used for the pass-through from five-year government bonds, and the corporate bond yield with the maturity of ten years is used for the pass-through from ten- and twenty-year government bonds.

The findings are broadly similar to the main results. The QE programs have statistically significant pass-through to corporate bond yields, mostly one-to-one, but the pass-through to the stock prices or exchange rates is not statistically significant. For the high-grade corporate bond yields, pass-through is mostly one-to-one with a magnitude between 0.36 and 1.87. Unlike the main results, the pass-through to the medium-grade bond yields is statistically significant in some cases, with the magnitude between 0.44 and 1.98. Such a wide range of estimates may reflect their volatility due to small samples in the subperiods. The signs of pass-through to stock prices are mixed, but none are statistically significant. The pass-through to exchange rates is negative in most cases, but only a few estimates are statistically significant.

Furthermore, there is no obvious difference in the pass-through in different subperiods, which suggests that the effects of different QE programs are broadly similar over time. One evident difference is that the pass-through from the five-year JGB yield is not statistically significant during the CME program between 2010 and 2013. But this is because the five-year JGB yield is extremely low in this period.

4.4.4 Analysis Using the Principal Component of JBG Yields

To jointly analyze the pass-through from the JGB yields with different maturities, I conduct the same analysis using a principal component of these JGB yields. Table 10 presents the results for main and additional variables. The results are broadly similar to the main results. The pass-through to the high-grade corporate bond yields is positive and statistically significant, whereas the pass-through to the stock prices and exchange rates is negative but not statistically significant in most cases.

4.4.5 Alternative Set of Non-Announcement Days

I use two alternative definitions of non-announcement days to estimate the pass-through of monetary policy shocks. First, following Rigobon and Sack (2004), a non-announcement day is defined as one business day prior to the announcement day. Accordingly, the number of announcement days and non-announcement days are the same. Second, following Gilchrist and Zakrajsek (2013), I exclude all the BOJ meeting days from the set of non-announcement days because some trivial or indirect news released on these meeting days may contaminate the identification. The estimates using these alternative definitions are not listed in order to conserve space, but the results are essentially the same as the baseline set of non-announcement days.

Table 10. GMM Estimates of the Pass-Through Based on the Principal Component of JGB Yields

	Estimate	**Std. Error**
A. Pass-Through to Main Variables		
AA, 5 Year	0.57**	(0.32)
AA, 10 Year	0.76**	(0.25)
BBB, 5 Year	1.41	(0.80)
BBB, 10 Year	0.58	(0.50)
Nikkei 225	−0.09	(0.22)
TOPIX	−0.05	(0.19)
U.S. Dollar	−0.06	(0.11)
B. Pass-Through to Additional Variables		
Euro	−0.00	(0.07)
REIT Index	−1.18	(1.52)
CDS Index	1.77	(2.54)
Australian Dollar	−0.05	(0.10)
Canadian Dollar	−0.15	(0.23)
Korean Won	−0.17	(0.26)
New Zealand Dollar	−0.08	(0.14)
U.K. Pound	−0.00	(0.08)
Hong Kong Dollar	−0.08	(0.14)
Singapore Dollar	−0.04	(0.08)
Taiwanese Dollar	−0.01	(0.05)
Thai Baht	−0.04	(0.08)

Notes: This table presents the pass-through based on the principal component of JGB yields with a maturity of five, ten, and twenty years. ** denotes significance at the 5 percent level. Heteroskedasticity-robust standard errors are presented in parentheses.

5. Conclusion

This paper analyzes the effects of unconventional monetary policy in Japan, which has been stuck at the ZLB for a substantially longer period than any other economy. To discuss if we could learn any lessons from Japanese experience, this paper compares the estimates of the pass-through of monetary policy shocks with the U.S. estimates.

By using the tools of an event study and identification through heteroskedasticity, I find that the effects of expansionary monetary policy shocks are directly passed on to corporate bond yields, notably for high-grade bond yields. However, the pass-through to stock prices and the exchange rate is not statistically significant in most cases. These results contrast with the U.S. results, where the pass-through to all assets is statistically significant.

To interpret such differences in the effects of unconventional monetary policy, one may focus on the characteristics of the financial markets. For example, because of various institutional reasons, the Japanese financial markets may be more segmented than the U.S. markets and may not be responsive to monetary policy shocks. On the other hand, one could focus on the different economic environment in Japan and the United States to explain different pass-through. More specifically, since the Japanese economy has been stuck at the ZLB for two decades, it would be extremely difficult for any announcement to change expectations about future inflation or short rates. Investigating the reason for such differences across countries is critical for the evaluation of the effects of unconventional monetary policy and is a question for future research.

Appendix

Derivation of the OLS Estimate

Describe the system of equations (1) and (2) in matrix form:

$$\begin{pmatrix} 1 & -\beta \\ -\alpha & 1 \end{pmatrix} \begin{pmatrix} \Delta i_t \\ \Delta s_t \end{pmatrix} = \begin{pmatrix} \gamma \\ \delta \end{pmatrix} X_t + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}. \tag{10}$$

By solving this equation, we obtain the reduced-form solution of the system,

$$\begin{pmatrix} \Delta i_t \\ \Delta s_t \end{pmatrix} = \frac{1}{1-\alpha\beta} \left[\begin{pmatrix} \beta\delta + \gamma \\ \alpha\gamma + \delta \end{pmatrix} X_t + \begin{pmatrix} 1 \\ \alpha \end{pmatrix} \varepsilon_t + \begin{pmatrix} \beta \\ 1 \end{pmatrix} \eta_t \right]. \tag{11}$$

Let σ_X^2, σ_ε^2, and σ_η^2 denote the variance of each shock. Then, the OLS estimate of α in equation (2) is

$$\hat{\alpha}_{OLS} = \frac{Cov(\Delta i_t, \Delta s_t)}{Var(\Delta i_t)}, \tag{12}$$

$$= \frac{(\beta\delta + \gamma)(\alpha\gamma + \delta)\sigma_X^2 + \alpha\sigma_\varepsilon^2 + \beta\sigma_\eta^2}{(\beta\delta + \gamma)^2\sigma_X^2 + \sigma_\varepsilon^2 + \beta^2\sigma_\eta^2}. \tag{13}$$

Accordingly, the bias of the OLS estimate is

$$\hat{\alpha}_{OLS} - \alpha = \frac{(1 - \alpha\beta)[\delta(\beta\delta + \gamma)\sigma_X^2 + \beta\sigma_\eta^2]}{(\beta\delta + \gamma)^2\sigma_X^2 + \sigma_\varepsilon^2 + \beta^2\sigma_\eta^2}. \tag{14}$$

Equation (14) indicates that the OLS estimate has a non-zero bias, unless there is a certain restriction on parameters (such as $\alpha\beta = 1$). For the discussion about the signs of the bias in the OLS estimates, see Gilchrist and Zakrajsek (2013).

Derivation of Conditional Variance

Denote the conditional variances of the shocks on the announcement days as $\sigma_{X|A}^2$, $\sigma_{\varepsilon|A}^2$, and $\sigma_{\eta|A}^2$. Similarly, denote the conditional variances of the shocks on the non-announcement days as $\sigma_{X|\bar{A}}^2$, $\sigma_{\varepsilon|\bar{A}}^2$, and $\sigma_{\eta|\bar{A}}^2$. Given the reduced-from solution of the system in equation (11), the conditional variance-covariance matrixes of the system, $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_{\bar{A}}$, are computed as follows:

$$\mathbf{\Omega}_A = \frac{1}{(1-\alpha\beta)^2} \left(\begin{array}{c|c} (\beta\delta + \gamma)^2\sigma_{X|A}^2 + \sigma_{\varepsilon|A}^2 + \beta^2\sigma_{\eta|A}^2 & (\beta\delta + \gamma)(\alpha\gamma + \delta)\sigma_{X|A}^2 + \alpha\sigma_{\varepsilon|A}^2 + \beta\sigma_{\eta|A}^2 \\ \hline \cdot & (\alpha\gamma + \delta)^2\sigma_{X|A}^2 + \alpha^2\sigma_{\varepsilon|A}^2 + \sigma_{\eta|A}^2 \end{array}\right), \tag{15}$$

$$\mathbf{\Omega}_A = \frac{1}{(1-\alpha\beta)^2} \left(\begin{array}{c|c} (\beta\delta + \gamma)^2\sigma_{X|A}^2 + \sigma_{\varepsilon|A}^2 + \beta^2\sigma_{\eta|A}^2 & (\beta\delta + \gamma)(\alpha\gamma + \delta)\sigma_{X|A}^2 + \alpha\sigma_{\varepsilon|A}^2 + \beta\sigma_{\eta|A}^2 \\ \hline \cdot & (\alpha\gamma + \delta)^2\sigma_{X|A}^2 + \alpha^2\sigma_{\varepsilon|A}^2 + \sigma_{\eta|A}^2 \end{array}\right), \tag{16}$$

Assume that the variance of the monetary policy shock is larger on the announcement days than on the non-announcement days, but the variances of the other shocks are the same across these two sets of days. Namely, we assume that

$$\sigma^2_{\varepsilon|A} > \sigma^2_{\varepsilon|\bar{A}}, \tag{17}$$

$$\sigma^2_{X|A} = \sigma^2_{X|\bar{A}}, \tag{18}$$

$$\sigma^2_{\eta|A} = \sigma^2_{\eta|\bar{A}}. \tag{19}$$

When taking the difference between $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_{\bar{A}}$ in equations (15) and (16), only the variances of the monetary policy shock remain and the variances of other shocks cancel out. Thus we obtain

$$\mathbf{\Omega}_A - \mathbf{\Omega}_{\bar{A}} = \frac{\sigma^2_{\varepsilon|A} - \sigma^2_{\varepsilon|\bar{A}}}{(1-\alpha\beta)^2} \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}. \tag{20}$$

Orthogonality of Instruments

To see that two instruments, $\mathbf{z_i}$ and $\mathbf{z_s}$, are orthogonal to the residuals $\mathbf{e}$, compute $\sum_{t=1}^{T} Z_t \cdot e_t$:

$$\sum_{t=1}^{T} Z_t \cdot e_t$$

$$= \mathbf{Z}' \cdot \mathbf{e} \tag{21}$$

$$= \begin{pmatrix} \mathbf{z}'_i \\ \mathbf{z}'_s \end{pmatrix} (\mathbf{\Delta s} - \alpha \mathbf{\Delta i}) \tag{22}$$

$$= \begin{bmatrix} \frac{1}{T_A}\mathbf{\Delta i}'_A, & -\frac{1}{T_{\bar{A}}}\mathbf{\Delta i}'_{\bar{A}} \\ \frac{1}{T_A}\mathbf{\Delta s}'_A, & -\frac{1}{T_{\bar{A}}}\mathbf{\Delta s}'_{\bar{A}} \end{bmatrix} [\mathbf{\Delta s}_A - \alpha \mathbf{\Delta} \boldsymbol{i}_A,\ \mathbf{\Delta s}_{\bar{A}} - \alpha \mathbf{\Delta} \boldsymbol{i}_{\bar{A}}] \tag{23}$$

$$= \begin{bmatrix} \frac{1}{T_A}\mathbf{\Delta i}'_A \cdot (\mathbf{\Delta s}_A - \alpha\mathbf{\Delta i}_A) - \frac{1}{T_{\bar{A}}}\mathbf{\Delta i}'_{\bar{A}} \cdot (\mathbf{\Delta s}_{\bar{A}} - \alpha\mathbf{\Delta i}_{\bar{A}}) \\ \frac{1}{T_A}\mathbf{\Delta s}'_A \cdot (\mathbf{\Delta s}_A - \alpha\mathbf{\Delta i}_A) - \frac{1}{T_{\bar{A}}}\mathbf{\Delta s}'_{\bar{A}} \cdot (\mathbf{\Delta s}_{\bar{A}} - \alpha\mathbf{\Delta i}_{\bar{A}}) \end{bmatrix} \tag{24}$$

$$= \begin{bmatrix} (\hat{\Omega}_{A,12} - \alpha\hat{\Omega}_{A,11}) - (\hat{\Omega}_{\bar{A},12} - \alpha\hat{\Omega}_{\bar{A},11}) \\ (\hat{\Omega}_{A,22} - \alpha\hat{\Omega}_{A,21}) - (\hat{\Omega}_{\bar{A},22} - \alpha\hat{\Omega}_{\bar{A},21}) \end{bmatrix} \tag{25}$$

$$= \begin{bmatrix} \Delta\hat{\Omega}_{12} - \alpha\Delta\hat{\Omega}_{11} \\ \Delta\hat{\Omega}_{22} - \alpha\Delta\hat{\Omega}_{21} \end{bmatrix}. \tag{26}$$

Since $\mathbf{\Delta\Omega} = C \begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{pmatrix}$, we have

$$\sum_{t=1}^{T} Z_t \cdot e_t \longrightarrow_p 0. \tag{27}$$

Accordingly, we have moment condition $E[Z_t \cdot e_t] = 0$.

Weak-Identification Robust Confidence Set

One concern of identification through heteroskedasticity is that the difference between the announcement days and non-announcement days may not be large enough for strong identification. In order to address the issue of weak identification, I employ the two statistics that could derive weak-instrument robust confidence sets: the S statistic in Stock and Wright (2000), an extension of Anderson and Rubin's (1949) statistic to GMM, and the K statistic in Kleibergen (2005).[24] These statistics test the null hypothesis for a hypothesized value of the parameter, based on the moment conditions evaluated at the hypothesized value. The confidence set is derived as the set of parameter values for which the test accepts the null hypothesis.

The S statistic is defined as follows:

$$S(\alpha_0) = \left[\sqrt{\frac{1}{T}} f_T(\alpha_0)\right]' \hat{V}_{ff}(\alpha_0)^{-1} \left[\sqrt{\frac{1}{T}} f_T(\alpha_0)\right], \tag{28}$$

where

$$\hat{V}_{ff}(\alpha) = Var\left(\sqrt{\frac{1}{T}} f_T(\alpha)\right), \tag{29}$$

[24]For details, see Stock, Wright, and Yogo (2002).

which is an estimate of the asymptotic variance-covariance matrix of the moment conditions. Note that the S statistic is based on the objective function in the minimization problem in equation (9), but the value is evaluated at the hypothesized value of the parameter, α_0. The S statistic has a chi-square null limiting distribution, with the number of moment conditions as the degrees of freedom.

The K statistic is defined as follows:

$$K(\alpha_0) = \frac{1}{4T}\left(\left.\frac{\partial S(\alpha)}{\partial \alpha}\right|_{\alpha_0}\right)[\hat{D}_T'(\alpha_0)\hat{V}_{ff}(\alpha_0)^{-1}\hat{D}_T(\alpha_0)]^{-1} \times \left(\left.\frac{\partial S(\alpha)}{\partial \alpha}\right|_{\alpha_0}\right)', \quad (30)$$

$$= \left[\sqrt{\frac{1}{T}} f_T(\alpha_0)\right]' \left[\hat{V}_{ff}(\alpha_0)^{-\frac{1}{2}} P_{\hat{V}_{ff}(\alpha_0)^{-\frac{1}{2}}\hat{D}_T(\alpha_0)} \hat{V}_{ff}(\alpha_0)^{-\frac{1}{2}}\right] \times \left[\sqrt{\frac{1}{T}} f_T(\alpha_0)\right], \quad (31)$$

where

$$\hat{D}_T(\alpha) = q_T(\alpha) - \hat{V}_{\alpha f}(\alpha)\hat{V}_{ff}(\alpha)^{-1} f_T(\alpha),$$

$$q_T(\alpha) = \frac{\partial f_T(\alpha)}{\partial \alpha},$$

$$\hat{V}_{\alpha f}(\alpha) = Cov\left(\sqrt{\frac{1}{T}} f_T(\alpha), \sqrt{\frac{1}{T}} q_T(\alpha)\right), \quad \text{and}$$

$$P_{\hat{V}_{ff}(\alpha_0)^{-\frac{1}{2}}\hat{D}_T(\alpha_0)} = \hat{V}_{ff}(\alpha_0)^{-\frac{1}{2}}\hat{D}_T(\alpha_0)[\hat{D}_T(\alpha_0)'\hat{V}_{ff}(\alpha_0)^{-1}\hat{D}_T(\alpha_0)]^{-1} \times \hat{D}_T(\alpha_0)'\hat{V}_{ff}(\alpha_0)^{-\frac{1}{2}}.$$

Essentially, the K statistic uses an optimal subset of moment conditions to improve the power of the tests. In other words, by using a subset of more relevant moment conditions, we could improve the efficiency of the test statistic, which leads to the higher power of the tests. The only difference between the S statistic in equation (28) and the K statistic in equation (31) is that the K statistic uses the variance-covariance matrix adjusted by the projection matrix based on $\hat{D}_T(\alpha)$. $\hat{D}_T(\alpha)$ is a residual of the gradient of moment conditions, $q_T(\alpha)$, after projecting it on the level of the moment conditions, $f_T(\alpha)$. By construction, $\hat{D}_T(\alpha)$ is orthogonal to $f_T(\alpha)$, and we use this orthogonality to improve the efficiency. The K statistic also has a chi-square null limiting distribution, with the number of parameters as the degrees of freedom.

References

Anderson, T. W., and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 20 (1): 46–63.

Aruoba, S. B., P. Cuba-Borda, and F. Schorfheide. 2014. "Macroeconomic Dynamics Near the ZLB: A Tale of Two Countries." Working Paper.

Bernanke, B. S. 2012. "Monetary Policy since the Onset of the Crisis." Speech at the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming, August 31.

Bernanke, B. S., and K. N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60 (3): 1221–57.

Claus, E., I. Claus, and L. Krippner. 2014. "Asset Markets and Monetary Policy Shocks at the Zero Lower Bound." Working Paper No. 2014-42, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, Australian National University.

D'Amico, S., and T. B. King. 2013. "Flow and Stock Effects of Large-scale Treasury Purchases: Evidence on the Importance of Local Supply." *Journal of Financial Economics* 108 (2): 425–48.

Eggertsson, G. B., and M. Woodford. 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity* (Spring): 139–235.

Fujiwara, I., Y. Nakazono, and K. Ueda. 2014. "Policy Regime Change against Chronic Deflation? Policy Option under a *Long-Term Liquidity Trap*." RIETI Discussion Paper No.14-E-019.

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2011. "The Financial Market Effects of the Federal Reserve's Large-Scale Asset

Purchases." *International Journal of Central Banking* 7 (1): 3–43.

Gilchrist, S., and E. Zakrajsek. 2013. "The Impact of the Federal Reserve's Large-Scale Asset Purchase Programs on Corporate Credit Risk." NBER Working Paper No. 19337 (August).

Glick, R., and S. Leduc. 2013. "The Effects of Unconventional and Conventional U.S. Monetary Policy on the Dollar." Working Paper No. 2013-11, Federal Reserve Bank of San Francisco.

Gürkaynak, R. S., B. Sack, and E. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1): 55–93.

Gürkaynak, R. S., B. Sack, and J. H. Wright. 2007. "The U.S. Treasury Yield Curve: 1961 to the Present." *Journal of Monetary Economics* 54 (8): 2291–2304.

Hattori, M., and H. S. Shin. 2008. "Yen Carry Trade and the Subprime Crisis." Working Paper (August).

Hausman, J. K., and J. Wieland. 2014. "Abenomics: Preliminary Analysis and Outlook." *Brookings Papers on Economic Activity* (Spring): 1–76.

———. 2015. "Overcoming the Lost Decades? Abenomics after Three Years." *Brookings Papers on Economic Activity* (Fall): 385–413.

Ito, T., and F. S. Mishkin. 2006. "Two Decades of Japanese Monetary Policy and the Deflation Problem." In *Monetary Policy under Very Low Inflation in the Pacific Rim*, Vol. 15 of NBER-EASE, ed. I. Takatoshi and A. K. Rose, 131–202 (chapter 4). University of Chicago Press.

Joyce, M., D. Miles, A. Scott, and D. Vayanos. 2012. "Quantitative Easing and Unconventional Monetary Policy — An Introduction." *Economic Journal* 122 (564): 271–88.

Kiley, M. T. 2013. "The Response of Equity Prices to Movements in Long-term Interest Rates Associated with Monetary Policy Statements: Before and After the Zero Lower Bound." FEDS Paper No. 2013-15, Board of Governors of the Federal Reserve System.

Kimura, T., and J. Nakajima. 2013. "Identifying Conventional and Unconventional Monetary Policy Shocks: A Latent Threshold Approach." Working Paper No. 13-E-7, Bank of Japan.

Kleibergen, F. 2005. "Testing Parameters in GMM Without Assuming that They Are Identified." *Econometrica* 73 (4): 1103–23.

Krishnamurthy, A., and A. Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity* (Fall): 215–65.

———. 2012. "The Aggregate Demand for Treasury Debt." *Journal of Political Economy* 120 (2): 233–67.

Kuroda, H. 2013. "Quantitative and Qualitative Monetary Easing." Speech at a meeting held by the Yomiuri International Economic Society in Tokyo, April 12.

Lam, W. R. 2011. "Bank of Japan's Monetary Easing Measures: Are They Powerful and Comprehensive?" IMF Working Paper No. 11/264.

Neely, C. J. 2013. "Unconventional Monetary Policy Had Large International Effects." Working Paper, Federal Reserve Bank of St. Louis (July).

Raskin, M. 2013. "The Effects of Monetary Policy on Corporate Bond Yield." Working Paper.

Rigobon, R. 2003. "Identification through Heteroskedasticity." *Review of Economics and Statistics* 85 (4): 777–92.

Rigobon, R., and B. Sack. 2003. "Measuring the Reaction of Monetary Policy to the Stock Market." *Quarterly Journal of Economics* 118 (2): 639–69.

———. 2004. "The Impact of Monetary Policy on Asset Prices." *Journal of Monetary Economics* 51 (8): 1553–75.

Rogers, J. H., C. Scotti, and J. H. Wright. 2014. "Evaluating Asset-Market Effects of Unconventional Monetary Policy: A Multi-Country Review." *Economic Policy* 29 (80): 749–99.

Romer, C. 2014. "It Takes a Regime Shift: Recent Developments in Japanese Monetary Policy through the Lens of the Great Depression." *NBER Macroeconomics Annual 2013*, ed. J. A. Parker and M. Woodford, 383–400. University of Chicago Press.

Rosa, C. 2012. "How 'Unconventional' Are Large-Scale Asset Purchases? The Impact of Monetary Policy on Asset Prices." Staff Report No. 560, Federal Reserve Bank of New York.

Shibamoto, M., and M. Tachibana. 2013. "The Effect of Unconventional Monetary Policy on the Macro Economy: Evidence from

Japan's Quantitative Easing Policy Period." Discussion Paper No. 2013-12, Kobe University (April).

Shiratsuka, S. 2010. "Size and Composition of the Central Bank Balance Sheet: Revisiting Japan's Experience of the Quantitative Easing Policy." *Monetary and Economic Studies* 28 (November): 79–106.

Stock, J. H., and J. H. Wright. 2000. "GMM with Weak Identification." *Econometrica* 68 (5): 1055–96.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business and Economic Statistics* 20 (4): 518–29.

Ueda, K. 2012a. "Japan's Deflation and the Bank of Japan's Experience with Nontraditional Monetary Policy." *Journal of Money, Credit and Banking* 44 (s1): 175–90.

———. 2012b. "The Effectiveness of Non-Traditional Monetary Policy Measures: The Case of the Bank of Japan." *Japanese Economic Review* 63 (1): 1–22.

Ugai, H. 2007. "Effects of the Quantitative Easing Policy: A Survey of Empirical Analyses." *Monetary and Economic Studies* 25 (1): 1–48.

Wright, J. H. 2012. "What Does Monetary Policy Do to Long-Term Interest Rates at the Zero Lower Bound?" *Economic Journal* 122 (564): 447–66.

Assessing the Sources of Credit Supply Tightening: Was the Sovereign Debt Crisis Different from Lehman?*

Paolo Del Giovane, Andrea Nobili, and Federico M. Signoretti
Bank of Italy

We estimate a structural econometric model for the credit market in Italy, using bank-level data on lending and interest rates and identifying shifts in demand and supply based on the responses of Italian banks to the Eurosystem's Bank Lending Survey. We distinguish supply restrictions due to increased borrowers' riskiness from those due to banks' balance sheet constraints, and test for the presence of credit rationing. We assess whether the effects of supply tightening differed during the sovereign debt crisis compared with the global financial crisis. We find that the effects of supply shocks transmit to loan quantities via an increase in lending rates and are larger

*A previous version of the paper has been published as a working paper under the title "Supply Tightening or Lack in Demand? An Analysis of Credit Developments during the Lehman and the Sovereign Debt Crises," Temi di discussione (Economic Working Paper) No. 942, Bank of Italy, 2013. We thank Paolo Angelini, Hans Dewachter, Enzo Dia, Luc Dress, Giovanni Ferri, Eugenio Gaiotti, Cristina Gualdani, Alberto Franco Pozzolo, Glenn Schepens, Gabriele Sene, Stefano Siviero, and two anonymous referees for helpful comments and discussions. We also thank participants at seminars held at the Bank of Italy and Banque de France, the workshop "The Sovereign Debt Crisis and the Euro Area" organized by the Bank of Italy (Rome, February 15, 2013), the European System of Central Banks workshop in preparation for the 2013 Surveillance Report (Frankfurt am Main, June 24–25, 2013), the 28th Annual Congress of the European Economic Association (Gothenburg, August 26–30, 2013), and the Macro Banking and Finance Workshop organized by the University of Milan Bicocca (Milan, September 19–20, 2013). Ginette Eramo and Stefano Piersanti provided help with the data set used in this analysis. The views expressed in the paper do not necessarily reflect those of the Banca d'Italia. All errors are the sole responsibility of the authors. Author contact: paolo.delgiovane@bancaditalia.it (Tel.: +39 064792 3219); andrea.nobili@bancaditalia.it (Tel.: +39 064792 4508); federicomaria.signoretti@bancaditalia.it (Tel.: +39 064792 3754).

when they reflect banks' funding difficulties as opposed to a deterioration of borrowers' riskiness. During phases of acute financial tensions, there is evidence of credit-rationing phenomena, related to banks' assessment of the constraints on their capital position. Based on a counterfactual exercise, the effects of the supply restriction on the cost and amount of credit were larger during the sovereign debt crisis than the global crisis, mostly reflecting the larger contribution of banks' funding conditions.

JEL Codes: E30, E32, E51.

1. Introduction

Understanding the role of supply factors in credit market developments has become a crucial issue since the onset of the great financial crisis. A growing body of literature uses indicators of banks' credit standards based on survey data to identify the relative contributions of demand and supply shocks.[1] An important limitation of these studies is that they have typically exploited only the overall supply indicators, and survey data have not been used to determine the *sources* of supply shocks. The literature on the credit channel of monetary policy (e.g., Bernanke and Gertler 1995) has instead emphasized how a restriction due to a deterioration in the creditworthiness of borrowers and one reflecting a worsening in banks' balance sheet conditions may have different implications and thus require different policy responses.

In this paper we study the effects of credit supply restrictions on the loan market and assess whether the effects of supply tightening differed during the sovereign debt crisis compared with the global financial crisis. To this end, we estimate a structural model of the Italian business loans market. The sources of supply restriction are likely to have been different in the two phases of the crisis: during the

[1]Works based on the Eurosystem's Bank Lending Survey include Hempell (2004), Berg et al. (2005), de Bondt et al. (2010), Hempell and Kok Sorensen (2010), Del Giovane, Eramo, and Nobili (2011), and Ciccarelli, Maddaloni, and Peydró (2015). Other analyses based on the Federal Reserve's Senior Loan Officer Opinion Survey are Lown, Morgan, and Rohatgi (2000), Cunningham (2006), Lown and Morgan (2006), Bayoumi and Melander (2008), Swiston (2008), and Basset et al. (2014).

global crisis, economic activity harshly contracted in all euro-area countries, borrowers' creditworthiness deteriorated, and banks' risk aversion sharply increased. In the sovereign debt crisis, banks' liquidity positions and access to wholesale funding were severely strained in the vulnerable euro-area countries, where the banks' creditworthiness was treated as equal to that of the respective governments. Italy is a well-suited case to conduct this analysis, as both phases of the crisis had important repercussions on its economy.

We use micro data and match individual banks' interest rates and loan amounts with the replies provided by the banks participating in the Eurosystem's quarterly Bank Lending Survey (BLS). We use the information in the BLS not only to disentangle supply and demand but also—crucially for the aim of our analysis—to disentangle the effect of the different *sources* of supply restrictions, by considering the individual factors that banks report behind changes in lending standards. In particular, we distinguish between (i) factors related to the banks' balance sheets (capital position, ability to access market financing, liquidity position) and (ii) factors related to perceived borrowers' riskiness (connected with the general economic conditions or with the specific industry or firm-specific outlook). In addition, we test for the presence of rationing, by estimating a "disequilibrium" relation in the spirit of Fair and Jaffee (1972) and Quandt (1978), which allows us to better capture the impact of the different supply factors.

As in other empirical works on lending demand and supply, a very important issue is the reliability of the identification strategy. In our case, a crucial challenge is the possibility that the indicators of demand and supply as reported by the banks in the BLS are not orthogonal to each other, which might reflect either common factors leading to changes in both demand and supply (e.g., the deterioration of macroeconomic and financial conditions) or a causal link from changes in supply to changes in demand (e.g., demand might be discouraged by higher borrowing costs).[2] We tackle this issue in various ways. First, we detect no or very low simultaneous correlation between bank-level indicators of demand and supply. Second, in our baseline regression we control for a number

[2]For example, Basset et al. (2014) showed that this is the case in the United States.

of bank-level characteristics and macro variables, which attenuates potential endogeneity concerns. Third, we replicate the methodology proposed by Basset et al. (2014; hereafter BCDZ), in which the (bank-specific) changes in demand are used to partial out, in a first-stage regression, changes in standards that are related to the reported changes in loan demand at the same bank. We find that the results are very similar to those obtained with the unadjusted supply indicators. All in all, this evidence is reassuring as to the possibility to identify supply and demand with reasonable confidence. Nonetheless, we acknowledge that we cannot exclude the possibility that our results are affected to some extent by some residual endogeneity.

We contribute to the existing literature along a number of dimensions. Differently from most previous studies, we use bank-level data as opposed to aggregate data both for survey information and credit developments. This type of information has been used before by Del Giovane, Eramo, and Nobili (2011; hereafter DEN) and by BCDZ (2014): DEN (2011) use micro data for Italy in reduced-form equations for lending quantities during the global crisis; BCDZ (2014) use data from the U.S. Senior Loan Officer Opinion Survey (SLOOS) to estimate the macroeconomic effects of exogenous changes in bank loan supply. Compared with these works, we use bank-level data also for the cost of credit and explicitly estimate a structural model for demand and supply; we investigate the effect of different sources of supply shocks, using the BLS factors rather than the overall indicator of credit standards; we allow for supply effects consistent with credit-rationing phenomena; and by including the sovereign debt crisis period, we are able to compare two phases with different sources of supply shocks.

The main results are the following. First, the effects of supply shocks are significantly larger when they reflect banks' balance sheet constraints as compared with a deterioration of borrowers' riskiness. We estimate that a tightening of lending standards due to funding constraints reported by all banks in the panel is associated with a widening of the loan spread by 70–80 basis points, depending on the specification. Considering the interest rate elasticity of credit demand, the consequent effect on quantities amounts to a reduction of 1.6–1.8 percentage points in the quarter-on-quarter lending growth rate. A comparable tightening due to an increase in risk perception is associated with a 15 basis point increase in the spread;

the impact is stronger when banks reported risk to have contributed "considerably" to the tightening, but this is a very rare occurrence in the sample. We also find evidence of episodes of credit rationing occurring when banks report capital constraints as a factor behind changes in credit standards:[3] in this case, we estimate that the effect of a tightening in banks' capital position is a 2 percentage point reduction in loan growth.

Second, a counterfactual exercise—assuming that supply indicators remained unchanged at their pre-crisis levels—suggests that the effects of credit tightening on loan rates were stronger in the sovereign debt crisis than in the global crisis: the cumulative effect is 190 basis points through the second quarter of 2012, of which one-third came during the global crisis and two-thirds during the sovereign debt crisis. Moreover, during the global crisis supply effects were mostly related to the banks' risk perception, while funding conditions became predominant during the sovereign crisis. In addition, we estimate that at the end of the sample period supply factors had a cumulative negative impact on the stock of loans of around 4 percent during the global crisis and 5 percent during the sovereign debt crisis. These effects can be attributed in about the same proportion to the adjustment of loan demand to the increase in the cost of credit and to credit rationing.

In addition to considering the endogeneity issue, we test for the robustness of our results by using alternative definitions of the loan spread, considering the level, rather than the variation, of the loan interest rate as one of the dependent variables, and including the sovereign spread as one of the regressors in our estimates.[4] In all these cases the estimated elasticity of both demand and supply curves remains broadly unchanged.

The rest of the paper is organized as follows. Section 2 describes the BLS data used in the estimation. Section 3 illustrates the methodology used for the identification of loan demand and supply curves. Section 4 discusses the empirical findings, for both the specification that assumes equilibrium in the credit market and the

[3]This result is in line with the classical findings on the role of capital constraints in contributing to "credit crunches" (Bernanke and Lown 1991; Peek and Rosengren 2005).

[4]During the sovereign debt crisis, the sovereign spread was often regarded as a sort of "sufficient statistic" to measure the intensity of tensions (see, e.g., Bofondi, Carpinelli, and Sette 2013; Albertazzi et al. 2014).

model with credit rationing. Section 5 illustrates the counterfactual exercises, comparing the importance of supply factors during the sovereign debt crisis with those during the global financial crisis. Section 6 provides a number of robustness checks, and section 7 concludes.

2. Data and Descriptive Evidence

The study is carried out on data for the panel of Italian banking groups (henceforth "banks") participating in the Eurosystem's quarterly Bank Lending Survey (BLS),[5] which represent roughly 60 percent of total outstanding amount of loans to enterprises in Italy. The number of banks changes somewhat over time, due to mergers and new survey participants. The data set consists of an unbalanced panel of eleven banks (with a maximum of eight per quarter) over thirty-nine quarters (2002:Q4 to 2012:Q2), providing a total of 287 observations.

As endogenous variables we use the bank-level quarter-on-quarter growth rate of loans to enterprises and the margin between the average interest rate on new bank loans to firms and the EONIA rate, which rules out the effects of monetary policy.[6] Loan data are computed using consolidated bank-level data from the Bank of Italy supervisory reports. They include repos and non-performing loans and are adjusted for the effects of securitization, for reclassifications, and for other variations not due to transactions, notably mergers and takeovers.[7]

[5]Detailed information on the BLS and the complete questionnaire can be found on the websites of the European Central Bank and Bank of Italy.

[6]Section 6 shows the estimates obtained with alternative measures of the cost of credit.

[7]Loan data include the drawn amount on credit lines. In principle, this may be a confounding factor; for example, during the early stages of the financial crisis in the United States, commercial and industrial (C&I) loans on banks' books grew rapidly as firms drew down their credit lines. In our analysis, however, this does not seem to be an important concern. First, considering the aggregate of all Italian banks, in our sample period credit lines account for less than one-third of total loans to non-financial corporations, and the share fell quite steadily during the crisis. Second, since 2009—the shorter time frame for which credit lines data for the panel of BLS banks are available—the dynamics of total loans to firms and that of an aggregate where drawn credit lines are excluded are very similar, except for the higher volatility of the latter series.

The sample is highly representative of the aggregate evolution of loan amount and interest rates. Figure 1 shows that—both for the entire banking system and for our sample—there were two phases of sharp slowdown in lending to firms: in 2008–09, coinciding with the "global crisis," and since the second half of 2011, when the sovereign debt crisis hit Italy severely. During both episodes the cost of credit also raised sharply.

Survey data are taken from the banks' responses to the BLS. For demand, we use the answers to the specific question on a bank's perception on loan demand from firms.[8] For supply, we use the question on the contribution of the different factors to lending standards.[9] For each question and supply factor, banks can indicate whether supply (demand) conditions have been tightened/eased (increased/decreased) considerably/somewhat or remained unchanged. As DEN (2011) show, in constructing indicators based on the BLS replies, it is important to take into account potential non-linear effects, which may be especially large in the case of supply conditions. We accordingly construct a set of dummy variables $\{BLS\_D^{c}_{it},\ BLS\_S^{f,c}_{it}\}$, where $\{D, S\}$ refer to demand and supply indicators, respectively; $c = \{increased/tightened, decreased/eased\}$ and $f = \{capital\ position,\ funding\ conditions,\ perception\ of\ risk\}$.[10] For risk perception, we construct separate dummy variables for the cases of *considerable* and *somewhat* easing/tightening.

[8]The question is: "Over the past three months, how has the demand for loans or credit lines to enterprises changed at your bank, apart from normal seasonal fluctuations?"

[9]The question reads: "Over the past three months, how have the following factors affected your bank's credit standards as applied to the approval of loans or credit lines to enterprises?"

[10]*Capital position* takes a value of 1 if banks reported tightening/easing in the factor "costs related to bank's capital position." For the other supply factors, the replies are aggregated in order to avoid collinearity problems: *Funding conditions* takes a value of 1 if banks reported tightening/easing in at least one of the two factors "bank's ability to access market financing" or "bank's liquidity position"; *risk perception* takes a value of 1 if banks reported tightening/easing in at least one of the three factors "perception of risk related to expectations regarding general economic activity," "perception of risk related to industry or firm-specific outlook," or "risk on collateral demanded." In the questionnaire, possible replies also include factors related to competition; we ignore them because they have been reported very rarely in the sample period that we consider.

Figure 1. Representativeness of the BLS Panel

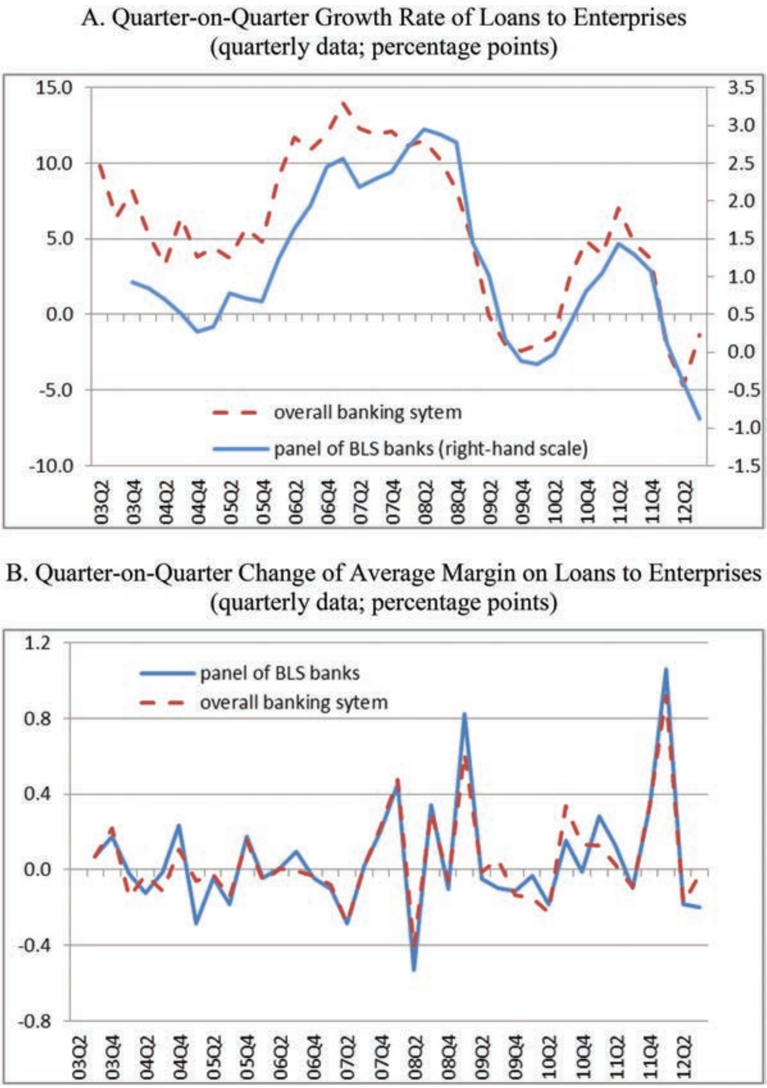
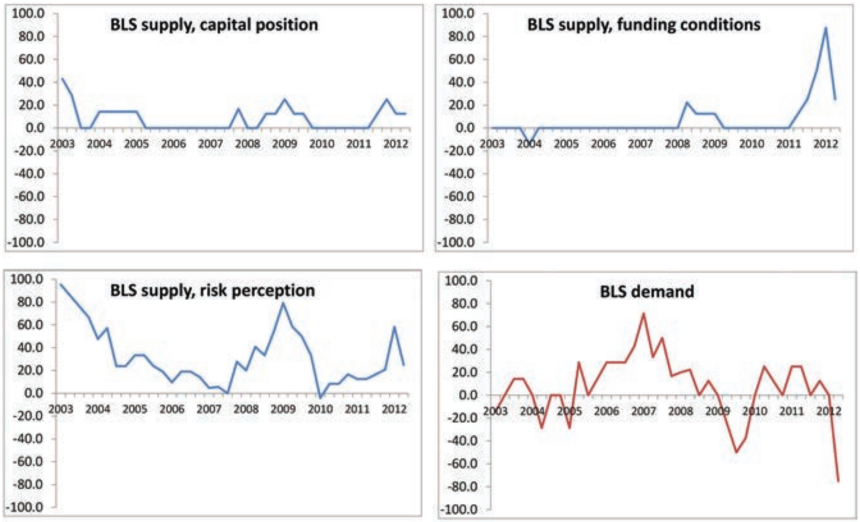


Table 1 reports descriptive statistics on the survey variables; figure 2 shows their evolution for the entire sample. First, the changes in lending standards are highly asymmetrical, with almost all the reported changes on the "tightening" side, whereas

Table 1. BLS Supply Factors and Demand Conditions for Loans to Enterprises: Descriptive Statistics (frequency of responses and, in parentheses, percentages with respect to total in each period)

	Factors Affecting Bank's Credit Standards				
	Bank's Capital Position	**Bank's Funding Conditions**	**Bank's Risk Perception**		**Bank's Demand Conditions**
1 = "contributed considerably to easing of credit standards"	0 (0.0)	0 (0.0)	0 (0.0)	1 = "decreased considerably"	0 (0.0)
2 = "contributed somewhat to easing of credit standards"	1 (0.3)	1 (0.3)	10 (3.5)	2 = "decreased somewhat	60 (20.9)
3 = "contributed to basically unchanged credit standards"	262 (91.3)	265 (92.3)	121 (42.2)	3 = "basically unchanged"	179 (62.4)
4 = "contributed somewhat to tightening of credit standards"	24 (8.4)	19 (6.6)	142 (49.5)	4 = "increased somewhat"	47 (16.4)
5 = "contributed considerably to tightening of credit standards"	0 (0.0)	2 (0.7)	14 (4.9)	5 = "increased considerably"	1 (0.3)
Total Observations	287 (100.0)	287 (100.0)	287 (100.0)	Total Observations	287 (100.0)

Figure 2. BLS Supply and Demand Indicators for Loans to Enterprises in Italy (quarterly data; "net percentages")



Source: Authors' calculation based on Bank of Italy data.

Notes: For BLS supply factors, "net percentages" are constructed as the difference between the share of banks reporting that the factor contributed to a tightening (considerably or somewhat) and the share of banks reporting that the factor contributed to an easing (considerably or somewhat). Positive (negative) values indicate a supply restriction (easing) compared with the previous quarter. Likewise, for BLS demand indicators, "net percentages" are constructed as the difference between the share of banks that reported an increase in demand (considerably or somewhat) and the share of banks that reported a decrease in demand (considerably or somewhat). Positive (negative) values indicate an increase (decrease) in loan demand compared with the previous quarter. In all cases, the range of variation of the indicator is from –100 to 100.

changes in demand are quite evenly balanced between "increase" and "decrease." Second, "considerable" tightening of standards is reported rarely, and only for risk related to firm-specific outlook. Third, the relative importance of the factors affecting credit standards differed in the two periods considered: in the aftermath of the Lehman Brothers collapse, the tightening mostly reflected increased perception of risk, with capital constraints also playing a role in some quarters; during the sovereign debt crisis, banks' funding (and liquidity) conditions played a much larger role. The two waves of supply restriction went along with a reported fall in loan demand.

Overall, developments in demand and supply conditions mirror the behavior of lending growth and cost reported in figure 1. In both phases of the crisis, the slowdown in lending corresponded to a fall in the BLS demand indicator and a tightening of the supply factors.

The endogenous variables and the BLS indicators are combined with a set of bank-level and macroeconomic controls. In particular, we include the *quarter-on-quarter* percentage change in nominal GDP; firms' financing needs, measured by the absolute change in the ratio between the corporate sector's investments and gross operating profit; bank size, measured by the logarithm of total assets; the share of core loans, i.e., loans to the private sector over total assets, as a control for the bank's business model; the ratio of nonperforming to total loans as a proxy of bank loan quality; the bank funding gap, measured as customer loans less customer deposits over customer loans; the capital-to-asset ratio, i.e., the bank's equity over total assets; the marginal cost of funding, computed as the difference between the weighted average of the interest rates paid by the bank on its sources of funding (customer deposits and debt securities); and the EONIA rate, with the weights reflecting the relative importance of each type of liability.[11] Given that the sovereign spread has been considered as a sufficient statistic for the tensions in the banking sector (see Albertazzi et al. 2014, among others), in the robustness section we also include the change in the Italian ten-year sovereign spread with respect to the ten-year German bund as an additional control variable.

[11] The inclusion of this variable allows us to address the concern that the BLS indicator of funding conditions, as a dummy variable, may only partially capture the banks' funding difficulties and that the EONIA rate used to compute the bank markup may also depend on other bank-specific variables not included in the equation. Affinito (2011) and Angelini, Nobili, and Picillo (2011) showed that the interbank rates at longer maturities faced by Italian banks during the crisis depended significantly on some specific characteristics of borrowers and lenders and that some of the estimated relationships increased dramatically after the outbreak of the 2007–08 crisis. This might also be the case for overnight interbank rates.

3. Methodology and Identification

We estimate the following system of two simultaneous equations:

$$\Delta spread_{it} = \alpha_{1i} + \sum_{f,c} \beta_1^{f,c}(L) BLS\_S_{i,t}^{f,c} + \theta_1 \cdot \Delta loans_{it} + \gamma_1 X_{it}^1 + \mu_{it}^S \quad (1)$$

$$\Delta loans_{it} = \alpha_{2i} + \sum_{c} \beta_2^{c}(L) BLS\_D_{i,t}^{c} + \theta_2 \cdot \Delta spread_{it} + \gamma_2 X_{it}^2 + \mu_{it}^D, \quad (2)$$

where the endogenous variables $\Delta spread_{it}$ and $\Delta loans_{it}$ are, respectively, the first difference of the spread between bank i's average interest rate on new loans and EONIA in quarter t and the quarter-on-quarter rate of lending growth.[12] The variables $\{BLS\_D_{it}^c,\ BLS\_S_{it}^{f,c}\}$ are the bank-level demand and supply factor indicators, as described in section 2, with $c = \{$*increased/tightened*, *decreased/eased*$\}$ and $f = \{$*capital position, funding conditions, perception of risk*$\}$.[13] These variables may enter contemporaneously and/or with lags. The lag order for each variable is selected by trying a range of lags from 0 to 4 and judging on the basis of the fit of the regression and the indications derived from standard information criteria. The vector $X_{it}^k (k = 1, 2)$ includes the bank-level and macroeconomic controls, as described above. We also add bank-specific fixed effects $(\alpha_{ki}, k = 1, 2)$ to control for unobserved bank-specific factors that might be correlated with the BLS variables and could result in inconsistent estimates. Moreover, in order to allow for possible non-linearity in the impact of the BLS indicators during specific periods of tension, we include a crisis-period dummy (taking a value of 1 from 2008:Q3 to the end of the sample period) and allow for interactions between this variable and the BLS indicators. Finally, we include lagged values of the dependent variable, if

[12] In what follows, we will sometimes loosely refer to the *spread* as *markup* or *margin*.

[13] As mentioned in section 2, in the case of *perception of risk* we construct separate dummy variables for the cases of *considerable* and *somewhat* easing/tightening, i.e., $c = \{$*tightened considerably*, *tightened somewhat*, *eased somewhat*, *eased considerably*$\}$.

statistically significant,[14] and dummies to capture seasonal effects. The system is consistently estimated using the two-step efficient generalized method of moments (GMM) estimator.[15] The selection of the model is based on a general-to-specific approach, where non-significant variables and/or lags are removed sequentially. Standard errors are clustered over time periods.

Our identification strategy for the system (1)–(2) relies on exclusion restrictions on the BLS indicators: the demand dummies $BLS\_D_{it}^{c}$ are excluded from equation (1) and the supply factor dummies $BLS\_S_{it}^{f,c}$ are excluded from equation (2). If θ_1 and θ_2 are, respectively, non-negative and non-positive, then equations (1) and (2) can be interpreted as, respectively, a credit supply and a credit demand curve.[16]

A crucial issue for the identification strategy is whether BLS replies are reliable indicators of changes in firm demand and/or bank lending standards. In this regard, BCDZ (2014) show that, in the United States, changes in credit standards as captured by bank-level responses to the Federal Reserve's SLOOS are negatively correlated with changes in the demand indicators. Therefore, as a proxy of loan supply shocks, they use the residual of a regression of

[14]The lagged dependent variables also act as instruments and are also reciprocally excluded. In a system of simultaneous equations, these variables are, by definition, predetermined. In principle, they can be used as regressors in both structural equations. In our system, the exclusion restrictions on predetermined variables are based upon the statistical significance of their coefficients and the indications provided by the Sargan test.

[15]The GMM estimator is more efficient than the traditional IV/2SLS estimator for over-identified systems of equations and when the residuals present heteroskedasticity and arbitrary intragroup correlation (see Hayashi 2000). BCDZ (2014) contrast ordinary least squares estimates of loan demand with instrumental-variables estimates and show that the latter is more (negatively) sloped, corroborating the validity of the changes in standards series as a potential loan supply shock measure. In a(n unreported) similar experiment we obtain similar results: the semi-elasticity of loan demand estimated with OLS is about –0.4 and statistically not significant.

[16]A special case of this structural model is $\theta_1 = 0$, i.e., credit supply is flat and the intermediaries set loan interest rates and fully accommodate credit demand, consistently with the standard imperfect competition framework of credit market (Freixas and Rochet 2008; Degryse, Kim, and Ongena 2009). It is also to be noted that, since in the sample period considered there are very few instances of easing of credit standards, only a portion of the demand curve can actually be estimated.

Table 2. Estimated Correlation between Demand Conditions and Supply Factors

	Demand Decrease	Demand Increase
BLS Capital Position, Tightening	0.20	−0.03
BLS Funding Conditions, Tightening	0.20	0.09
BLS Risk Perception, Tightening Considerably	−0.01	0.04
BLS Risk Perception, Tightening Somewhat	0.21	−0.05

changes in credit standards on the demand indicator and a number of macroeconomic and bank-specific controls.

It is important to check whether similar concerns apply to our case. Visual inspection of figure 2 provides preliminary evidence about the presence of unconditional correlation between changes of the BLS supply factors and those of the demand indicator. The contemporaneous correlation is very low for all supply factors over the entire sample period; for risk perception, it increases somewhat in the phase of most acute tensions of the global crisis. A more useful investigation assesses the empirical correlation in the panel data, rather than for the aggregate series over time, by considering separately the indicators of demand decreases and demand increases. The results, reported in table 2, confirm that the correlation is generally low, albeit not trivial. It is to be noted that these estimates do not control for any macroeconomic variables or bank-specific features, which are instead included, together with bank-fixed effects, in all our regressions. The inclusion of control variables further limits the potential endogeneity problems in the use of the BLS indicators.

As a further test for our identification strategy, we use statistical tools to assess the validity of our excluding restrictions. In particular, as we deal with *over-identified* supply and demand equations, we perform the Sargan-Hansen test for each structural equation separately, and separate tests for the exclusion of each individual instrument with the "difference-in-Sargan" statistic. We also report the Wald version of the Kleibergen-Paap statistic, which tests whether the excluded instruments are correlated with the endogenous regressors.[17]

[17] See Kleinbergen and Paap (2006).

Finally, as a further robustness check, in section 5 we rerun our main regressions using the same methodology as in BCDZ (2014). We find that all our results hold also in this case.

The system of equations (1) and (2) describes a framework in which the market for loans clears at every point in time. An important limitation of this approach is its inability to capture *credit rationing*, which may occur during episodes of severe restriction of loan supply and is typically related to specific factors.[18] Exploiting the BLS information on the factors behind the supply restriction, we can account for the presence of credit rationing in our estimates by using a variant of the "quantitative approach" by Fair and Jaffee (1972). According to that approach, the structural representation of a disequilibrium model consists of a demand equation, a supply equation, and an equation determining the level of excess demand. In this paper we test whether specific BLS supply indicators provide information on the existence of excess demand or supply, therefore capturing market disequilibrium phenomena. From a practical perspective, this approach boils down to testing whether any of the supply factors can be significantly added to equation (2). The complete description and derivation of the equations for this approach are provided in the appendix.

4. The Empirical Results

Table 3 reports the results of the econometric estimation. First, we discuss the equilibrium model (equations (1)–(2)). Columns (a)–(a') show the estimates of the supply and demand equations, including only the BLS indicators. The coefficient for the loan spread in the demand equation is highly significant and negative, suggesting a downward-sloping demand curve. The estimated elasticity is high: a 100 basis point increase in the markup is associated with a reduction of more than 2 percentage points in the quarter-on-quarter growth rate of loans. In the supply equation, the coefficient of loan growth

[18] Credit-rationing episodes occur when, at the prevailing interest rate, the demand for credit exceeds the supply and lenders will not supply additional credit even if the borrowers are ready to pay higher margins. Seminal work by Bernanke and Lown (1991) points to the role of banks' capital constraints; Stiglitz and Weiss (1981) focus on the role of asymmetric information. Additional causes of credit rationing relate to funding or liquidity constraints and risk aversion.

Table 3. Structural Equations for Loans to Enterprises

	Equilibrium Model				Disequilibrium Model	
	(a)	**(a')**	**(b)**	**(b')**	**(c)**	**(c')**
	Supply Curve	**Demand Curve**	**Supply Curve**	**Demand Curve**	**Supply Curve**	**Demand Curve**
	Δspread (t)	**Δloan (t)**	**Δspread (t)**	**Δloan (t)**	**Δspread (t)**	**Δloan (t)**
Endogenous Variables:						
Δloan (t)	0.067*		0.082*		0.081**	
Δspread (t)		−2.189***		−2.468***		−2.286***
Predetermined Variables:						
Δloan (t–2)		0.191***		0.129**		0.152**
Δspread (t–1)	−0.404***		−0.426***		−0.424***	
Δspread (t–2)	−0.204**		−0.252***		−0.252***	
Exogenous Variables:						
BLS Demand, Increase (t)		1.140***		1.067***		0.984**
BLS Demand, Decrease (t)		−0.964**		−0.538		−0.288
BLS Supply, Capital Position, Tightening (t)	−0.018		0.008			−2.001***
BLS Supply, Funding Conditions, Tightening (t)	0.472***		0.407***		0.407***	
BLS Supply, Funding Conditions, Tightening (t–1)	0.323**		0.290*		0.292*	
BLS Supply, Risk Perception, Tightening Considerably (t)	0.597**		0.487***		0.487***	
BLS Supply, Risk Perception, Tightening Somewhat (t) * Crisis Dummy	0.167**		0.124		0.124	
Crisis Dummy	0.101		0.030		0.029	

(*continued*)

Table 3. (Continued)

	Equilibrium Model				Disequilibrium Model	
	(a)	**(a')**	**(b)**	**(b')**	**(c)**	**(c')**
	Supply Curve	**Demand Curve**	**Supply Curve**	**Demand Curve**	**Supply Curve**	**Demand Curve**
	Δspread (t)	**Δloan (t)**	**Δspread (t)**	**Δloan (t)**	**Δspread (t)**	**Δloan (t)**
Other Control Variables:						
ΔNominal GDP (t)			−0.094**	0.427**	−0.095**	0.379**
ΔFinancing Needs (t)			−0.009	0.032	−0.009	0.031
Bank Size (t)			−0.047	−0.336	−0.047	−0.359
Bank Share of Core Loans (t)			−0.015	0.087*	−0.011	0.086*
Bank Loan Quality (t)			0.047*	−0.338***	0.046**	−0.403***
Bank Funding Gap (t)			−0.002	0.049	−0.001	0.054
Bank Capital-to-Asset Ratio (t)			0.021	−0.105	0.017	−0.007
ΔMarginal Cost of Funding (t)			0.081**	0.256	0.081**	0.330*
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Seasonal Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Estimation Technique	2S-GMM	2S-GMM	2S-GMM	2S-GMM	2S-GMM	2S-GMM
Number of Observations (N)	247	247	245	245	245	245
Number of Regressors (K)	12	7	20	15	19	16
Number of Endogenous Regressors (K1)	1	1	1	1	1	1
Number of Instruments (L)	14	14	22	22	22	22
Number of Excluded Instruments (L1)	3	8	3	8	4	7
R-squared	0.162	0.136	0.195	0.196	0.169	0.251
Identification Diagnostics:						
Under-identification Test:						
Kleibergen-Paap rk LM Statistic	20.88	20.78	14.99	17.03	21.83	17.05
p-value	0.00	0.01	0.00	0.03	0.00	0.02
Weak-Identification Test:						
F-statistic of Excluded Instruments	7.48	9.56	4.51	6.12	5.12	6.98
Over-identification Test for All Variables:						
Hansen J Statistic	4.10	16.97	3.24	12.46	3.28	2.30
p-value	0.13	0.02	0.20	0.08	0.35	0.89

(*continued*)

Table 3. (Continued)

	Equilibrium Model				Disequilibrium Model	
	(a)	(a')	(b)	(b')	(c)	(c')
	Supply Curve	Demand Curve	Supply Curve	Demand Curve	Supply Curve	Demand Curve
	Δspread (t)	Δloan (t)	Δspread (t)	Δloan (t)	Δspread (t)	Δloan (t)
C-Statistics (p-value) for Each Single Variable:						
BLS Demand, Increase (t)	0.67		0.82		0.82	
BLS Demand, Decrease (t)	0.11		0.13		0.15	
BLS Supply, Capital Position, Tightening (t)		0.10		0.01	0.27	
BLS Supply, Funding Conditions, Tightening (t)		0.73		0.40		0.40
BLS Supply, Funding Conditions, Tightening (t–1)		0.30		0.74		0.74
BLS Supply, Risk Perception, Tightening Considerably (t)		0.95		0.91		0.91
BLS Supply, Risk Perception, Tightening Somewhat (t) * Crisis Dummy		0.88		0.47		0.47

Notes: The dependent variables, "Δspread" and "Δloan," are, respectively, the quarterly change in bank markup (computed as the difference between the average rate on new loans and the EONIA rate) and the quarter-on-quarter growth rate in loan quantity. "BLS Supply, Capital Position, Tightening," "BLS Supply, Funding Conditions, Tightening," and "BLS Supply, Risk Perception, Tightening" (also specifying whether the bank reported that this factor contributed considerably or somewhat to a tightening) are dummy variables taking the value of 1 if a bank reported that this factor contributed to a tightening in credit supply conditions. "BLS Demand, Decrease" and "BLS Demand, Increase" are dummy variables taking the value of 1 if the bank reported, respectively, decrease/increase in demand. *, **, and *** denote significance, respectively, at 10 percent, 5 percent, and 1 percent. The "Under-identification Test" is a Lagrange-Multiplier test of the null hypothesis that the equation is under-identified (i.e., the matrix of reduced-form coefficients on the $L1$ excluded instruments has rank $K1-1$), while the alternative hypothesis is that the equation is identified (i.e., the matrix has rank exactly equal to $K1$). Under the null hypothesis, the test statistic is distributed as chi-squared in $(L1-K1+1)$ degrees of freedom. The "Over-identification Test" is based on the Sargan-Hansen statistic: under the null hypothesis that the exclusion restrictions are valid; the test statistic is distributed as chi-squared in the $(L-K)$ number of tested over-identifying restrictions. The C-statistic (or "difference-in-Sargan" statistic) allows a test of the exogeneity of one single instrument. Under the null hypothesis that both the smaller set of instruments and the additional, suspect instruments are valid, the C-statistic is distributed as chi-squared in the number of instruments tested.

is positive but very small and only marginally significant, suggesting that the credit market supply curve is almost flat.

As to the other variables in the supply equation, none of the "easing" dummies is significant (and they are therefore not reported in the table). This is not unexpected, as it presumably reflects the marked asymmetry of the BLS supply indicators, which almost never report an easing. "Tightening" replies are associated with significant effects on the loan spread. Specifically, banks' funding conditions have a significant effect both on impact and with a one-quarter lag: the estimates indicate that if all banks in the sample reported a tightening related to this factor, the spread would be around 80 basis points higher than if no bank signaled such an effect. A tightening related to a worsening of banks' risk perception is also found to have a significant effect on loan rates: the impact effect is estimated at 60 basis points when banks reported that this factor contributed "considerably" to the tightening, and at 17 basis points during the crisis when it was reported to have contributed "somewhat."[19] Given the interest rate elasticity of credit demand, the resulting effects on quantities would amount to a reduction of 1.8 percentage points in the quarter-on-quarter lending growth for funding conditions and of between 0.4 and 1.3 points for risk perception. A tightening connected with the banks' capital position is not statistically significant. In the loan demand equation, the dummies of the BLS demand indicator are significant and have the expected signs in the case of both "increase" and "decrease." The coefficients suggest that the relationship is roughly linear: both an increase and a decrease in demand are related to a corresponding change in credit growth by about 1 percentage point.

Columns (b)–(b') report the results of the regression including bank-level and macroeconomic covariates. The results are similar to the previous regression, with minor differences in the magnitude of the coefficients. The elasticity of demand increases somewhat in absolute value while that of the supply curve remains broadly unchanged. In both equations, GDP and bank-level loan quality are significant and have the expected signs; higher marginal cost of funding is associated with a rise in the cost of credit. The inclusion of these controls crowds out the significance of the BLS indicator for

[19] In the latter case, it is significant only when interacted with the crisis dummy.

demand decrease in the demand equation and of the moderate risk perception indicator in the supply equation. The overall impact on the loan spread of a tightening due to funding conditions is around 70 basis points (after one quarter); that of a considerable tightening due to risk perception is 50 basis points. The corresponding effects on loan growth are 1.6 and 1.1 percentage points.

Overall, the fact that the results for the BLS indicators are very similar between the models reported in columns (a)–(a') and (b)–(b') provides further evidence that endogeneity is not a significant concern. An additional formal check of the quality of the identification strategy is the Sargan-Hansen test, which tests the null hypothesis that the instruments are correctly excluded from the structural equation to be identified. In the regressions just discussed (both with and without controls), the test suggests that the supply equation is correctly identified. For the demand equation, however, the null hypothesis cannot be accepted at a 10 percent confidence level. The "difference-in-Sargan" C-statistic—testing for the exclusion of each instrument separately—suggests that this failure is related to the restriction that the dummy for the tightening of banks' capital position is excluded from the quantity equation. The exclusion of all the other instruments, instead, is accepted with a generally high confidence level. This suggests that the BLS indicator of a bank's capital position could enter the demand equation, thus carrying information on possible credit rationing.

Columns (c)–(c') of table 2 show the results for the disequilibrium model, i.e., the system of equations (7)–(8) described in the appendix, in which the existence of excess loan demand is related to the BLS supply factor involving capital position.[20] The direct effect of the BLS capital position indicator on loan demand is negative and highly significant; a tightening of this factor is associated with a decline of 2 percentage points in the quarter-on-quarter growth rate of loans. The coefficients of the remaining variables are almost identical to the previous specification, including the estimated slopes of the demand and supply curves and the coefficients associated with the BLS demand and supply indicators.

[20]In unreported regressions, we tried to use the other BLS supply factors as rationing indicators; consistently with the difference-in-Sargan test mentioned above, none were significant.

The diagnostic tests appear to support our identification scheme. In particular, the overall Sargan-Hansen test p-value increases substantially for both the supply and demand equations, and the null hypothesis is now comfortably accepted also for the demand equation. All individual instruments—including the BLS capital position in the demand equation—also comfortably pass the C-test for single instrument exclusion. Overall, these results suggest that the credit supply curve is almost flat in normal times and becomes temporarily price inelastic when banks report a worsening in their capital position, consistent with the presence of credit rationing during a financial crisis.

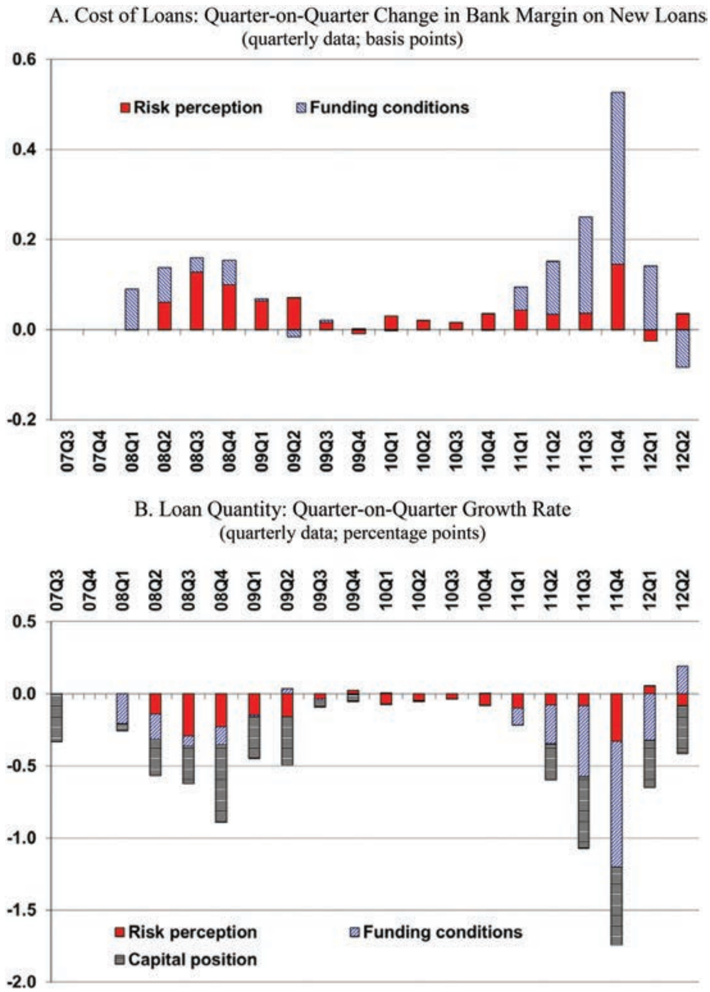
5. Was the Sovereign Debt Crisis Different from the Global Crisis?

In the previous sections we have used the BLS demand and supply indicators to identify and estimate structural demand and supply curves for business lending. Now we use these estimates to quantify the contribution of supply factors to changes in the cost and the volume of loans during the two phases of the financial crisis. To this end we perform a counterfactual exercise setting the supply indicators constant throughout the crisis at their pre-crisis (2007:Q2) levels.[21] Based on the estimated parameters, we can calculate the values of the interest rate spread and the loan growth rate in this counterfactual scenario and compare them with the fitted values based on actual supply indicators, thus obtaining an estimate of each factor's contribution.

We perform the counterfactual analysis using the specification presented in columns (c)–(c') of table 3. Figure 3 (panels A and B) shows the quarterly contribution of each factor to the change in the cost and the growth rate of loans to firms. Table 4 provides a more compact illustration of the effects of supply factors, in terms of their quantification and sources, by reporting the estimates of the cumulative effects in the two phases of the financial crisis.

[21] In the exercise, we assume that the BLS demand indicators and the control variables equal their realized values.

Figure 3. Counterfactual Exercise: Estimated Contribution of Supply Factors to the Cost and Growth Rate of Loans to Enterprises (based on columns (c)–(c') of table 3)



The results indicate that supply factors—as measured by the BLS indicators—had a substantial effect on both the cost and the availability of credit throughout the crisis. The magnitude of the effects was stronger on average during the sovereign debt crisis than during the global financial crisis.

Table 4. Counterfactual Exercise: Cumulative Contribution of Supply Factors to the Cost and Growth Rate of Loans to Enterprises (based on columns (c)–(c') of table 3)

	Global Crisis 2007:Q3–2010:Q1	**Sovereign Debt Crisis 2010:Q2–2012:Q2**	**Whole Period 2007:Q3–2012:Q2**
	Effect on the Cost of Credit (Basis Points)		
BLS Supply Factors:	25	82	107
BLS Funding Conditions	46	34	80
BLS Perception of Risk	71	116	187
Total Supply Indicators			
	Effect on the Stock of Loans (Percent)		
BLS Supply Factors:			
BLS Capital Position	−2.2	−2.0	−4.1
BLS Funding Conditions	−0.6	−1.9	−2.4
BLS Perception of Risk	−1.1	−0.8	−1.8
Total Supply Indicators	−3.8	−4.6	−8.4

As regards the cost of credit, the tightening of supply conditions is estimated to have determined a quarterly rise of 53 basis points at the peak of the sovereign debt crisis (2011:Q4), compared with about 15 basis points at the peak of the global crisis (2008:Q3 and Q4). The cumulative effect from the beginning of the crisis through 2012:Q2 is estimated at around 190 basis points, of which about one-third came during the global crisis and two-thirds during the sovereign debt crisis.

The two phases of the crisis were characterized by differing relative importance of the various supply factors. During the global crisis, risk perception played a predominant role in affecting the cost of lending, while the impact of funding conditions was smaller. By contrast, during the sovereign debt crisis, the factors relating to difficulties in access to funding became much more important, determining on average around 70 percent of the rise in interest rates due to all supply factors. The effects on the growth rate of loans via the elasticity of demand to cost differed correspondingly. The impact of supply factors was greatest in the last quarter of 2011, when it is estimated to have reduced the quarter-on-quarter growth rate of loans by almost 2 percentage points, compared with almost 1 point in 2008:Q4. The contributions of credit-rationing effects related to the banks' capital position were similar in the two phases of the crisis.[22]

At the end of the sample period (2012:Q2), supply factors are estimated to have determined a cumulative reduction of about 8 percent in the stock of loans. The part of this negative effect that can be ascribed to credit rationing is estimated to be about 4 percent, equally distributed in the two phases of the crisis.

6. Robustness Analysis

6.1 Assessing the Endogeneity of the BLS Supply Indicators

As discussed above, the use of the BLS demand and supply indicators to identify demand and supply may raise endogeneity concerns.

[22]It is worthwhile noting that these differences may reflect not only the relative importance of the various factors in the two periods but also changes in the way lending standards affect credit markets over time.

Adding to the evidence provided by the correlation analysis (in section 3) and that discussed in section 4, we further address this issue here by replicating the approach proposed by BCDZ (2014). First, we estimate logit regressions for each BLS supply factor, including, besides the standard control variables, the BLS demand indicators as explanatory variables. Results (reported in table 5) show that changes in the BLS demand conditions have marginal information content only for the supply indicator that reflects banks' funding conditions; they have no predictive power for the supply indicators related to the banks' assessment of their capital position or their perception of risk.

Using the estimated parameters of the logit regressions, we then derive the "adjusted" indicators of BLS supply factors.[23] Figure 4 shows these indicators together with the corresponding "unadjusted" measures. The dynamics of the two types of indicator are very similar. As expected, during the sovereign debt crisis, the BLS "adjusted" indicator for bank funding conditions is somewhat lower than the corresponding "adjusted" measure.

Next, we reestimate the model (with rationing), replacing all the BLS supply indicators with the corresponding "adjusted" measures. The coefficients for the supply and demand curves are reported in table 6 and are very similar to those obtained with the benchmark specification. In particular, the semi-elasticity of loan demand and the estimated effect of credit rationing, as captured by banks' assessments of their capital position, remain highly significant and of the same magnitude. We also replicate the counterfactual exercise described in section 5 using these estimated coefficients and the "adjusted" BLS supply indicators. The main results remain broadly unchanged.[24]

[23]In particular, these are the differences between the bank-specific reported outcome and the bank-specific predicted probabilities of a tightening at each point in time. As BCDZ (2014) observe, asymptotically these residuals share the zero-mean and orthogonality properties of the residuals from an OLS regression.

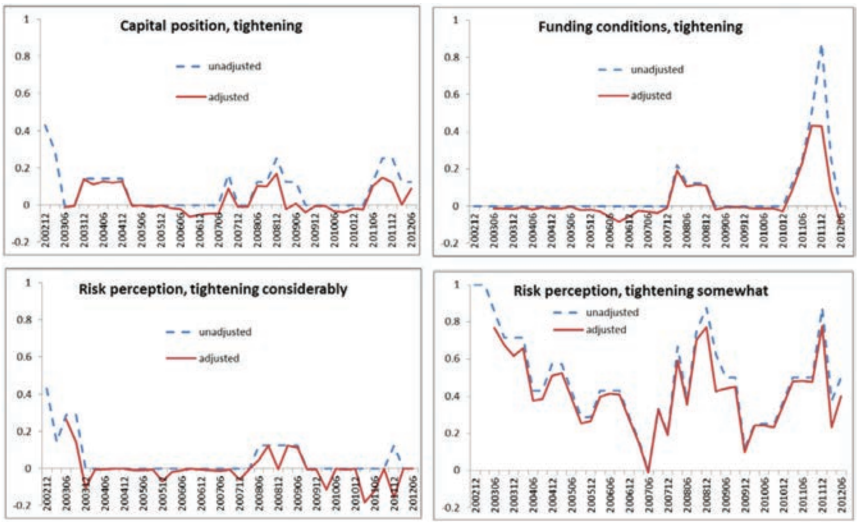
[24]Specifically, the cumulative effect of the supply factors on the cost and the stock of loans, from the beginning of the crisis through 2012:Q2, reduces somewhat to, respectively, 160 basis points and –6.8 percent. The results also confirm the predominant role of banks' funding conditions during the sovereign debt crisis, as well as the presence of a similar impact of credit-rationing effects in the two phases of the crisis.

Table 5. Logit Regressions for BLS Supply Factors

	Dependent Variable			
	BLS Supply, Capital Position, Tightening	**BLS Supply, Funding Conditions, Tightening**	**BLS Supply, Risk Perception, Tightening Considerably**	**BLS Supply, Risk Perception, Tightening Somewhat**
Lagged Dependent Variable	1.502**	1.458**	3.901**	1.881***
BLS Demand, Decrease (t)	1.237	1.658*	−0.331	−0.596
BLS Demand, Increase (t)	1.371	2.551**	−20.713	0.783
ΔNominal GDP (t)	−0.474	−0.019	0.740	−0.444**
ΔFinancing Needs (t)	−0.085	−0.005	0.128	0.021
Bank Size (t)	12.920**	4.721	9.303*	−0.986
Bank Loan Quality (t)	−0.683**	0.181	−0.846	0.063
Bank Share of Core Loans (t)	0.126	0.040	−0.412	−0.004
Bank Funding Gap (t)	0.081	0.137*	0.370*	0.000
Bank Capital-to-Asset Ratio (t)	−0.364	0.053	2.474	−0.101
ΔMarginal Cost of Funding (t)	0.342	0.476*	4.693*	0.447**
Bank-Specific Fixed Effects	Yes	Yes	Yes	Yes

Notes: "BLS Supply, Capital Position, Tightening," "BLS Supply, Funding Conditions, Tightening," and "BLS Supply, Risk Perception, Tightening" (also specifying whether the bank reported that this factor contributed considerably or somewhat to a tightening) are dummy variables taking the value of 1 if a bank reported that this factor contributed to a tightening in credit supply conditions. "BLS Demand, Decrease" and "BLS Demand, Increase" are dummy variables taking the value of 1 if the bank reported, respectively, decrease/increase in demand. *, **, and *** denote significance, respectively, at 10 percent, 5 percent, and 1 percent.

Figure 4. "Adjusted" vs. "Unadjusted" BLS Supply Indicators (quarterly data)



Source: Authors' calculation based on Bank of Italy data.
Notes: Each "unadjusted" BLS supply indicator is computed as the share of banks reporting that the indicated factor contributed to a tightening in credit supply conditions. Each "adjusted" indicator is computed as the difference between the corresponding "unadjusted" indicator and the predicted probability of a tightening related to the indicated factor, which is obtained by the logit regressions described in section 6.1.

All in all, the results reported here confirm that the correlation between the BLS supply and demand indicators is not a critical issue in our empirical investigation.

6.2 Using Alternative Measures for the Cost of Credit

As a second robustness check, we reestimate our structural model under three alternative definitions of the cost of credit, replacing the difference between the loan rate and the EONIA, which we used in our baseline regressions. First, practical considerations suggest that loan demand may simply depend on the cost of credit, rather than on the margin over some market rate; therefore, we rerun our regressions using simply the level of the loan rate. Second, since the use of an overnight rate could result in attributing an improper term

Table 6. Structural Equations for Loans to Enterprises: Removing Correlation between BLS Supply Factors and BLS Demand Indicators

	(a)	(a')
	Supply Curve	**Demand Curve**
	Δspread (t)	**Δloan (t)**
Endogenous Variables:		
Δloan (t)	0.092**	
Δspread (t)		−2.025**
Predetermined Variables:		
Δloan (t–2)		0.152**
Δspread (t–1)	−0.413***	
Δspread (t–2)	−0.268***	
Exogenous Variables:		
BLS Demand, Increase (t)		0.834**
BLS Demand, Decrease (t)		−0.451
BLS Supply, Capital Position, Tightening (t), Adjusted		−1.922***
BLS Supply, Funding Conditions, Tightening (t), Adjusted	0.430***	
BLS Supply, Funding Conditions, Tightening (t–1), Adjusted	0.432***	
BLS Supply, Risk Perception, Tightening Considerably (t), Adjusted	0.346**	
BLS Supply, Risk Perception, Tightening Somewhat (t), Adjusted * Crisis Dummy	0.096	
Crisis Dummy	0.062	
Other Control Variables:		
ΔNominal GDP (t)	−0.112*	0.438**
ΔFinancing Needs (t)	−0.011	0.041
Bank Size (t)	−0.016	−0.573
Bank Share of Core Loans (t)	−0.017*	0.080
Bank Loan Quality (t)	0.053**	−0.396***
Bank Funding Gap (t)	−0.003	0.054
Bank Capital-to-Asset Ratio (t)	0.029	−0.039
ΔMarginal Cost of Funding (t)	0.111***	0.267
Fixed Effects	Yes	Yes
Seasonal Dummies	Yes	Yes
Estimation Technique	2S-GMM	2S-GMM
Number of Observations (N)	245	245
Number of Regressors (K)	19	16
Number of Engodenous Regressors (K1)	1	1
Number of Instruments (L)	22	22
Number of Excluded Instruments (L1)	4	7
R-squared	0.095	0.264

(*continued*)

Table 6. (Continued)

	(a)	(a')
	Supply Curve	**Demand Curve**
	Δspread (t)	**Δloan (t)**
Identification Diagnostics:		
Under-identification Test:		
Kleibergen-Paap rk LM Statistic	19.41	17.31
p-value	0.00	0.02
Weak-Identification Test:		
F-statistic of Excluded Instruments	4.59	6.38
Over-identification Test for All Variables:		
Hansen J Statistic	1.99	1.84
p-value	0.57	0.93

Notes: "BLS Supply Funding Conditions, Tightening" and "BLS Supply, Risk Perception, Tightening" (also specifying whether the bank reported that this factor contributed considerably or somewhat to a tightening) are residuals from the logit regressions reported in table 3. For the definition of all other variables and the statistical tests, see the notes of table 3.

premium to our measure of loan cost, we replicated the analysis by computing the margin with respect to the three-month EURIBOR (to which 95 percent of business loans are indexed). One drawback of EURIBOR (which is quoted for unsecured transactions) is that it was not an appropriate measure of the monetary policy stance during the global crisis, when it abnormally jumped because of the increase in risk aversion in the interbank market (see Angelini, Nobili, and Picillo 2011). For this reason, we considered as a third measure the spread vis-à-vis the three-month EUREPO, which is based on secured interbank transactions and was basically unaffected by tensions in the interbank market during the crisis.

Table 7 reports the estimated coefficients for the alternative measures of the cost of credit. The results are very similar to those for the benchmark system.

6.3 Including the Sovereign Spread

During the sovereign debt crisis, analysts and policymakers paid considerable attention to the sovereign yield spreads between the euro-area countries hit by the tensions and Germany, which was often regarded as a sort of "sufficient statistic" to measure the

Table 7. Estimated Coefficients in Loan Demand and Supply Curves: Using Alternative Measures for the Cost of Credit

Semi-elasticity of Enterprises' Credit Demand to the Cost of Credit	
Memo: Spread between the Loan Rate and the EONIA	*−2.286***
Loan Rate	−2.914***
Spread between the Loan Rate and the Three-Month EURIBOR	−1.819*
Spread between the Loan Rate and the Three-Month EUREPO	−1.967***
Semi-elasticity of Credit Supply to a Change in Loans to Enterprises	
Memo: Spread between the Loan Rate and the EONIA	*0.081***
Loan Rate	0.066*
Spread between the Loan Rate and the Three-Month EURIBOR	0.012
Spread between the Loan Rate and the Three-Month EUREPO	0.072*

Notes: In the credit demand equation, the dependent variable is the quarter-on-quarter growth rate of the loan quantity. The loan rate is the average rate on new loans to enterprises. For each measure of the cost of credit, the quarterly change is considered in the regressions. *, **, and *** denote significance, respectively, at 10 percent, 5 percent, and 1 percent.

severity of the strains. Thus, we deemed it useful to investigate whether including the ten-year sovereign spread affects the estimated coefficients of the credit demand and supply curves and added the sovereign spread as a control variable. Results are reported in table 8.

The estimated semi-elasticity of credit demand remains unchanged at about 2 percent. The slope of the loan supply curve is also similar to that obtained with the benchmark specification. The coefficient of the sovereign spread is positive and highly significant: a 100 basis point increase in the spread is associated with a pass-through of around 60 basis points after one quarter, a magnitude similar to those found in other studies based on aggregate data (Neri 2013; Zoli 2013; Albertazzi et al. 2014).[25] The inclusion

[25] We also checked whether the effects of changes in the sovereign spread differed depending on whether they stem from changes in the yield on Italian or on German government bonds, by separately including the BTP and the bund yields in the regressions in the place of the sovereign spread. The results show that a rise in the BTP yield has a stronger effect on the cost of loans to enterprises than a reduction in the bund yield. The pass-through after one quarter is around 75 basis points in the former case and 50 basis points in the latter.

Table 8. Structural Equations for Loans to Enterprises: Including the Sovereign Spread

	(a)	(a')
	Supply Curve	**Demand Curve**
	Δspread (t)	**Δloan (t)**
Endogenous Variables:		
Δloan (t)	0.061*	
Δspread (t)		−2.020**
Predetermined Variables:		
Δloan (t–2)		0.143**
Δspread (t–1)	−0.462***	
Δspread (t–2)	−0.177**	
Exogenous Variables:		
BLS Demand, Increase (t)		0.904**
BLS Demand, Decrease (t)		−0.253
BLS Supply, Capital Position, Tightening (t)		−2.254***
BLS Supply, Funding Conditions, Tightening (t)	0.089	
BLS Supply, Funding Conditions, Tightening (t–1)	0.109	
BLS Supply, Risk Perception, Tightening Considerably (t)	0.376***	
BLS Supply, Risk Perception, Tightening Somewhat (t) * Crisis Dummy	0.083	
Crisis Dummy	0.030	
Other Control Variables:		
ΔNominal GDP (t)	−0.044	0.381**
ΔFinancing Needs (t)	−0.006	0.021
Bank Size (t)	−0.071	−0.449
Bank Share of Core Loans (t)	−0.013	0.089*
Bank Loan Quality (t)	0.032	−0.442***
Bank Funding Gap (t)	−0.005	0.050
Bank Capital-to-Asset Ratio (t)	0.010	−0.025
ΔMarginal Cost of Funding (t)	0.033	0.211
ΔSovereign Spread (t)	0.201***	0.682
ΔSovereign Spread (t–1)	0.416***	−0.194
Fixed Effects	Yes	Yes
Seasonal Dummies	Yes	Yes
Estimation Technique	2S-GMM	2S-GMM
Number of Observations (N)	245	245
Number of Regressors (K)	21	18
Number of Engodenous Regressors (K1)	1	1
Number of Instruments (L)	24	24
Number of Excluded Instruments (L1)	4	7
R-squared	0.327	0.274

(*continued*)

Table 8. (Continued)

	(a)	(a')
	Supply Curve	**Demand Curve**
	Δspread (t)	**Δloan (t)**
Identification Diagnostics:		
Under-identification Test:		
Kleibergen-Paap rk LM Statistic	22.48	24.64
p-value	0.00	0.00
Weak-Identification Test:		
F-statistic of Excluded Instruments	5.20	5.95
Over-identification Test for All Variables:		
Hansen J Statistic	4.94	4.44
p-value	0.17	0.62

Notes: "ΔSovereign Spread" is the quarterly change in the difference between the yield on the ten-year Italian government bond and the corresponding German one. For the definition of all other variables and the statistical tests, see the notes of table 3.

of the sovereign spread wipes out the significance of the BLS funding conditions indicator and the bank-specific marginal cost and lowers the coefficient of the BLS risk-perception indicators. These results suggest that during the sovereign debt crisis the correlation between banks' funding difficulties and credit developments largely reflected the common shock related to the sovereign debt markets. This common effect dominated the idiosyncratic components, potentially captured by the individual banks' cost of funding and their survey answers.

7. Conclusions

In this paper we used firm-level responses of Italian banks to the Eurosystem's quarterly Bank Lending Survey to estimate structural relationships in the credit market. Our aim was to assess the effect of credit supply tightening on the volume and cost of bank lending, distinguishing between restrictions connected to borrowers' riskiness and those due to worsening in banks' balance sheet conditions. We evaluated whether such effects differed during the sovereign debt crisis compared with the global financial crisis.

We find that the impact of supply shocks on loan rates and quantities is significantly larger when the restriction reflects banks' balance sheet constraints as compared with borrowers' riskiness. The data also indicate that credit rationing was present during the most acute phases of financial tension and was related to banks' capital constraints. In addition, a counterfactual exercise suggests that the effects of the supply restriction on both the cost and the availability of credit were, on average, stronger during the sovereign debt crisis than during the global crisis. And whereas throughout the global crisis supply effects on the cost of credit were mostly related to the banks' risk perception, during the sovereign crisis funding conditions became predominant. Credit-rationing effects related to the banks' capital position were similar in the two phases of the crisis.

A caveat common to all studies based on survey data is that the quality of the results depends on the truthfulness of the respondents' answers. On the one hand, banks may be inclined to report tighter credit standards than those actually applied because they fear that the information could be exploited for supervisory purposes. On the other hand, during the crisis public criticism and political pressure may have induced banks to portray their policies as less restrictive than they actually were. In the latter case, our estimates of the effects of supply restriction on credit conditions could be considered as a lower bound.

Our results suggest two policy considerations. First, well-capitalized banks are less likely to generate strong procyclical changes in credit supply conditions through rationing. In this regards, the recent structural reforms in the banking sector should help to stabilize the credit cycle. Second, the differences we found as regards the effects of supply restriction corroborate the mix of policy measures adopted by the European Central Bank's Governing Council to counteract the effects of the crisis on lending, which attenuated the negative spiral between tightening credit conditions and the deterioration of the real economy. An interesting development for future research would be to extend our model to include bank-level data on recourse to the longer-term refinancing operations or sales of financial assets to central banks in the context of the Extended Asset Purchase Programme. This would contribute to the growing empirical literature on the effects of unconventional monetary policy on credit conditions and the real economy.

Appendix. Description of the Credit-Rationing Model

The system of equations (1) and (2) describes a framework in which the changes in the interest rate always ensure that the quantity of credit supplied equals the amount demanded at each point in time, so that the market clears. An important limitation of this approach is its inability to capture credit rationing, namely episodes of excess demand over supply.[26] To make the model tractable and suitable for empirical analysis, we follow the "quantitative approach" by Fair and Jaffee (1972), according to which the structural representation of a disequilibrium model consists in a demand equation, a supply equation, and a "short-side" rule assuming that the traded quantity is the lower between supply and demand. The system of equations is, therefore, modified as follows:

$$\Delta spread_{it} = a_{1i} + \sum_{f,c} \beta_1^{f,c}(L) BLS\_S_{i,t}^{f,c} + \theta_1 \cdot \Delta loans_{it}^S + \gamma_1 X_{it}^1 + \mu_{it}^S \tag{3}$$

$$\Delta loans_{it}^D = a_{2i} + \sum_{c} \beta_2^{c}(L) BLS\_D_{i,t}^{c} + \theta_2 \cdot \Delta spread_{it} + \gamma_2 X_{it}^2 + \mu_{it}^D \tag{4}$$

$$\Delta loans_{it} = min(\Delta loans_{it}^S, \Delta loans_{it}^D). \tag{5}$$

To close the model, the seminal approach also included additional equations where excess demand and excess supply are related to positive and negative changes in the price level. This approach may be questionable in an empirical analysis of the credit market insofar as the interest rate is an endogenous variable. In this paper we instead assume that the information on the existence of excess

[26] Credit-rationing episodes occur when, at the prevailing interest rate, the demand for credit exceeds the supply and lenders will not supply additional credit even if the borrowers are ready to pay higher margins. The possible causes include banks' balance sheet constraints, risk aversion, and asymmetric information. As for the latter, banks may not want to raise lending rates above a certain level in order to avoid financing riskier borrowers (adverse selection) or to discourage firms from taking on additional risk (moral hazard), as in the seminal work by Stiglitz and Weiss (1981).

demand or supply is provided by the BLS supply indicators rather than changes in the loan interest rate, as follows:

$$\Delta loans_{it}^{D} - \Delta loans_{it}^{S} = \begin{cases} \sum_f \sigma_1^f(L) BLS\_S_{i,t}^{f,tightening} & \text{if } \Delta loans_{it}^{D} - \Delta loans_{it}^{S} > 0 \\ -\sum_f \sigma_2^f(L) BLS\_S_{i,t}^{f,easing} & \text{if } \Delta loans_{it}^{D} - \Delta loans_{it}^{S} \leq 0. \end{cases} \tag{6}$$

Equation (6) relates excess demand and excess supply, respectively, to the banks reporting a tightening and an easing of the various BLS supply factors. This assumption is more in line with the literature of credit rationing, which typically occurs in the form of non-price allocation of credit. Following the discussion in Fair and Jaffee (1972), the system of equations (3)–(6) can be expressed in closed form as follows:

$$\Delta spread_{it} = a_{1i} + \sum_{f,c} \beta_1^{f,c}(L) BLS\_S_{i,t}^{f,c} + \theta_1 \Delta loans_{it} + \gamma_1 X_{it}^1 - \sum_f \sigma_2^{f,easing}(L) BLS\_S_{i,t}^{f,easing} + \mu_{it}^S \tag{7}$$

$$\Delta loans_{it} = a_{2i} + \theta_2 \Delta spread_{it} + \sum_c \beta_2^c(L) BLS\_D_{i,t}^c + \gamma_2 X_{it}^2 - \sum_f \sigma_1^{f,tightening}(L) BLS\_S_{i,t}^{f,tightening} + \mu_{it}^S. \tag{8}$$

We allow all the BLS supply factors to potentially capture a disequilibrium in the credit market and test which ones actually do so by looking at the statistical significance of the coefficients $\sigma_1(L)$ and $\sigma_2(L)$. In practice, since banks very rarely reported that some factors had contributed to an easing of credit standards (see table 1), it is impossible to estimate the coefficients $\sigma_2(L)$. Therefore, our analysis is accordingly confined to assess which (if any) of the coefficients $\sigma_1^{f,tightening}(L)$ is significantly different from zero. Model (7)–(8) is estimated with the same GMM estimator as model (1)–(2).[27]

[27] Another approach to estimating disequilibrium models relies on maximum-likelihood methods (Amemiya 1974; Maddala and Nelson 1974), which have been

References

Affinito, M. 2011. "Do Interbank Customer Relationships Exist? And How Did They Function in the Crisis? Learning from Italy." Economic Working Paper No. 826, Bank of Italy.

Albertazzi, U., T. Ropele, G. Sene, and F. M. Signoretti. 2014. "The Impact of the Sovereign Debt Crisis on the Activity of Italian Banks." *Journal of Banking and Finance* 46: 387–402.

Allain, L., and N. Oulidi. 2009. "Credit Market in Morocco: A Disequilibrium Approach." IMF Working Paper No. 09/53.

Amemiya, T. 1974. "A Note on the Fair and Jaffee Model." *Econometrica* 42 (4): 759–62.

Angelini, P., A. Nobili, and C. Picillo. 2011. "The Interbank Market after August 2007: What Has Changed, and Why?" *Journal of Money, Credit and Banking* 43 (5): 923–58.

Barajas, A., and S. Steiner. 2002. "Credit Stagnation in Latin America." IMF Working Paper No. 02/53.

Bassett, W. F., M. B. Chosak, J. C. Driscoll, and E. Zakrajšek. 2014. "Changes in Bank Lending Standards and the Macroeconomy." *Journal of Monetary Economics* 62: 23–40.

Bayoumi, T., and O. Melander. 2008. "Credit Matters: Empirical Evidence on U.S. Macro-Financial Linkages." IMF Working Paper No. 08/169.

Berg, J., A. van Rixtel, A. Ferrando, G. de Bondt, and S. Scopel. 2005. "The Bank Lending Survey for the Euro Area." ECB Occasional Paper No. 23.

Bernanke, B. S., and M. Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9 (4): 27–48.

Bernanke, B. S., and C. S. Lown. 1991. "The Credit Crunch." *Brookings Papers on Economic Activity* 22 (2): 205–48.

---

used extensively in studies of credit-rationing episodes during financial crises (e.g., Kim 1999; Barajas and Steiner 2002; Allain and Oulidi 2009). However, maximum likelihood requires stronger distributional assumptions than GMM; maximizing the likelihood function in a disequilibrium model may be difficult (Goldfeld and Quandt 1975), and the complexity of the likelihood function in the presence of autocorrelation is high (Laffont and Monfort 1977).

Bofondi, M., L. Carpinelli, and E. Sette. 2013. "Credit Supply during a Sovereign Crisis." Economic Working Paper No. 909, Bank of Italy.

Ciccarelli, M., A. Maddaloni, and J. L. Peydró. 2015. "Trusting the Bankers: A New Look at the Credit Channel of Monetary Policy." *Review of Economic Dynamics* 18 (4): 979–1002.

Cunningham, T. J. 2006. "The Predictive Power of the Senior Loan Officer Survey: Do Lending Officers Know Anything Special?" Working Paper No. 24, Federal Reserve Bank of Atlanta.

de Bondt, G., A. Maddaloni, J. L. Peydró, and S. Scopel. 2010. "The Euro Area Bank Lending Survey Matters: Empirical Evidence for Credit and Output Growth." ECB Working Paper No. 1160.

Degryse, H., M. Kim, and S. Ongena. 2009. *Microeconometrics of Banking: Methods, Applications, and Results*, 1st Edition. Oxford University Press.

Del Giovane, P., G. Eramo, and A. Nobili. 2011. "Disentangling Demand and Supply in Credit Developments: A Survey-Based Analysis for Italy." *Journal of Banking and Finance* 35 (10): 2719–32.

Fair, R. C., and D. M. Jaffee. 1972. "Methods of Estimation for Markets in Disequilibrium." *Econometrica* 40 (3): 497–514.

Freixas, X., and J. C. Rochet. 2008. *Microeconomics of Banking*, 2nd Edition. MIT Press.

Goldfeld, S. M., and R. Quandt. 1975. "Estimation in a Disequilibrium Model and the Value of Information." *Journal of Econometrics* 3 (4, November): 325–48.

Hayashi, F. 2000. *Econometrics.* Princeton, NJ: Princeton University Press.

Hempell, H. S. 2004. "New Bank Lending Survey of the Eurosystem: Interpretation and Use of First Results for Germany." *Journal of Business Cycle Measurement and Analysis* 1 (3): 387–408.

Hempell, H. S., and C. Kok Sorensen. 2010. "The Impact of Supply Constraints on Bank Lending in the Euro Area — Crisis Induced Crunching?" ECB Working Paper No. 1262.

Kim, H. 1999. "Was the Credit Channel a Key Monetary Transmission Mechanism Following the Recent Financial Crisis in the Republic of Korea?" Policy Research Working Paper No. 2103, World Bank.

Kleibergen, F., and R. Paap. 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 133 (1): 97–126.

Laffont, J., and A. Monfort. 1977. "Econometrie des modeles d'equilibre avec rationnement." No. A157 (February), Ecole Polytechnique Laboratoire d'Econometrie.

Lown, C. S., and D. P. Morgan. 2006. "The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey." *Journal of Money, Credit and Banking* 38 (6): 1575–97.

Lown, C., D. P. Morgan, and S. Rohatgi. 2000. "Listening to Loan Officers: The Impact of Commercial Credit Standards on Lending and Output." *Economic Policy Review* (Federal Reserve Bank of New York) 6 (2): 1–16.

Maddala, G. S., and F. D. Nelson. 1974. "Maximum Likelihood Methods for Models of Markets in Disequilibrium." *Econometrica* 42 (6): 1013–30.

Neri, S. 2013. "The Impact of the Sovereign Debt Crisis on Bank Lending Rates in the Euro Area." Occasional Paper No. 170, Bank of Italy.

Peek, J. and E. S. Rosengren. 2005. "Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan." *American Economic Review* 95 (4): 1144–66.

Quandt, R. E. 1978. "Tests of the Equilibrium vs. Disequilibrium Hypotheses." *International Economic Review* 19 (2): 435–52.

Stiglitz, J., and A. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (3): 393–410.

Swiston, A. 2008. "A U.S. Financial Conditions Index: Putting Credit Where Credit Is Due." IMF Working Paper No. 08/161.

Zoli, E. 2013. "Italian Sovereign Spreads: Their Determinants and Pass-through to Bank Funding Costs and Lending Conditions." IMF Working Paper No. 13/84.

Joint Validation of Credit Rating PDs under Default Correlation*

Ricardo Schechtman
Central Bank of Brazil

This study investigates new proposals of statistical tests for validating the PDs (probabilities of default) of credit rating models (CRMs). The proposed tests recognize the existence of default correlation, deal jointly with the default behavior of all the ratings, and, in contrast to previous literature, control the error of validating incorrect CRMs. Power-sensitivity analysis and strategies for power improvement are discussed for the calibration tests, whereas a non-typical goal is proposed for the tests of discriminatory power, leading to results of power dominance. Finally, Monte Carlo simulations investigate the finite sample bias for varying scenarios of parameters.

JEL Codes: C12, G21, G28.

1. Introduction

This paper studies issues of validation for credit rating models (CRMs). In this article, CRMs are defined as a set of risk buckets (ratings) to which borrowers are assigned and which indicate the likelihood of default (usually through a measure of probability of default, PD) over a fixed time horizon (usually one year). Examples include rating models of external credit agencies such as Moody's and Standard & Poor's and banks' internal credit rating models.

*The author would like to thank Axel Munk, Dirk Tasche, Getulio Borges da Silveira, and Kostas Tsatsaronis for helpful conversations along the project as well as seminar participants at the Bank for International Settlements (BIS), the 17th FDIC Annual Derivatives Securities and Risk Management Conference, the C.R.E.D.I.T Conference, and Central Bank of Brazil research workshops. The author also acknowledges the hospitality at the BIS during his fellowship there, when most of the project was conducted. Views expressed in this paper do not necessarily reflect those of the Central Bank of Brazil or the BIS. Author contact: Research Department, Central Bank of Brazil, Av. Presidente Vargas 730, Rio de Janeiro, Brazil, 20071-900. E-mail: ricardo.schechtman@bcb.gov.br.

CRMs are key tools to credit risk management and have had their relevance increased, as the Basel II Accord (Basel Committee on Banking Supervision 2006a) allows the PDs of the internal ratings to function as inputs in the computation of banks' regulatory levels of capital.[1] Its goal was not only to make regulatory capital more risk sensitive, and therefore to diminish the problems of regulatory arbitrage, but also to strengthen stability in financial systems through better assessment of borrowers' credit quality. However, the great challenge for Basel II, in terms of implementation, lies still in the validation of CRMs, particularly the validation of bank-estimated rating PDs.[2] Besides satisfying regulatory demands, PD validation is also crucial for banks not to be left in competitive disadvantage towards their peers. However, the recent financial crisis has also promoted doubts about the efficacy with which banks and rating agencies had been validating their CRMs. Regulators are currently examining whether to place limits on the use of models to prevent banks from attempting to understate the riskiness of their portfolios[3] (Watt 2013).

In fact, validation of CRMs has been considered a difficult job due to two main factors. Firstly, the typically long credit time horizon of one year or so results in few observations available for *backtesting*. This means, for instance, that the bank/supervisor will, in most practical situations, have to judge the CRM based solely on five to ten observations available at the database. Secondly, as borrowers are usually sensitive to a common set of factors in the economy (e.g., industry, geographical region), variation of macro conditions over the forecasting time horizon induces correlation among defaults. Both these factors contribute to decreasing the power of quantitative methods of validation. This paper does not aim at a prescription to surpass the aforementioned unavoidable difficulties but instead at discussing the trade-offs and limitations involved in the validation task from a statistical perspective.

---

[1]The higher the PD, the higher is the regulatory capital.

[2]According to BCBS (2005b), validation is above all a bank task, whereas the supervisor's role should be to certificate this validation.

[3]That would represent a further enhancement of the Basel Accords, after the recent Basel III Accord (BCBS 2011). Basel III didn't bring any major modifications on the use of CRMs for regulatory purposes.

The judgment of the performance of a CRM is generally a twofold issue. It involves the aspects of calibration and discriminatory power. Calibration is the ability to forecast accurately the ex post *(long-run)* default rate of each rating (e.g., through an ex ante estimated PD). Discriminatory power is the ability to ex ante discriminate, based on the rating, between defaulting borrowers and non-defaulting borrowers.

As BCBS (2006a) is explicit about the demand for banks' internal models to possess good calibration, testing calibration is the starting point of this paper.[4] According to BCBS (2005b), quantitative techniques for testing calibration are still in the early stages of development. BCBS (2005b) reviews some simple tests, namely, the binomial test, the Hosmer-Lemeshow test, a normal test, and the traffic lights approach (Blochwitz et al. 2004). These techniques all have the disadvantage of being univariate (i.e., designed to test a single rating PD per time) and/or making the unrealistic assumption of cross-sectional default independency. An approach that tests each rating PD per time may translate into a joint procedure with rather higher error rates than those of the employed univariate test (e.g., Hochberg and Tamhane 1987). Similarly, a false assumption of default independency generally produces substantially higher error rates and can also lead to similar probabilities of rejecting correctly and incorrectly specified CRMs (e.g., BCBS 2005b, Bluemke 2013).

More recent proposed approaches have similar or new limitations. Balthazar (2004) proposes using the same Basel II model of capital requirement, which recognizes default dependency, for PD validation, but restricts the analysis to the univariate case. Miu and Ozdemir (2008) explore deeper the idea of Balthazar (2004), but still their analysis remains restricted to the univariate case. Blochlinger (2012) proposes a multivariate method that also recognizes default correlation but, on the other hand, is inconsistent with the functional form of the Basel II model (see the discussion in Gordy 2000). Inconsistency with the Basel II model cannot ensure that the validated PDs would be appropriate inputs to the Basel II capital requirement formula. Bluemke (2013) addresses the multivariate case in a Basel II-like model, but his approach does not provide a closed formula for the critical region or the power of the test. Therefore, apart from

[4]According to BCBS (2006a), PDs should resemble long-run average default rates for all ratings.

simulations, the author cannot discuss trade-offs and strategies for power improvement involved in the proposed validation test.[5] Additionally, a crucial concern common to all approaches in the literature is that they do not control for the error of accepting a miscalibrated CRM. Instead, they control for the error of rejecting correct CRMs, which, from a prudential viewpoint, is of secondary importance.

This paper reverses the roles of the hypothesis used throughout the CRM validation literature, in order to control for the error of validating incorrectly specified CRMs. Furthermore, this paper presents an asymptotic analytical framework to jointly test several PDs under the assumption of default correlation. The approach generalizes the Basel II model in a similar fashion to Demey et al. (2004) but with a new configuration oriented towards validation purposes. The results include a new simple one-sided CRM calibration test and the discussion of the relative roles played by the distinct elements that affect the power of the proposed test (e.g., differences between consecutive PD ratings, indifference regions of validation, asset correlations, ratings driving the power). Under the new formulation of the hypothesis, power is the probability of accepting a correctly specified CRM and, therefore, should achieve minimum levels for the test to be practical. Strategies for power improvement are also analyzed. The paper also discusses, to a considerable extent, the greater particular difficulties and conceptual problems related to two-sided CRM calibration testing.

Good discriminatory power is also a desirable property of CRMs, as it allows rating-based yes/no decisions (e.g., credit granting) to be made with less error and therefore less cost by the bank (see Blochlinger and Leippold 2006, for instance). BCBS (2005b) comprehensively reviews some well-established techniques for examining discriminatory power, including the area under the receiver operating characteristic (ROC) curve (Engelmann, Hayden, and Tasche 2003), the accuracy ratio, and the Kolgomorov-Smirnov statistic.

Although the use of the above-mentioned techniques of discriminatory power is widespread in banking industry, two constraining points should be noted. First, the pursuit of perfect discrimination is inconsistent with the pursuit of perfect calibration in realistic

[5] Additionally, Miu and Ozdemir (2008) and Blochlinger (2012) examine only very briefly power considerations.

CRMs. The reason is that to increase discrimination, one would be interested in having, over the long run, the ex post rating distributions of the default and non-default groups of borrowers as separate as possible, and this involves having default rates as low as possible for good-quality ratings (in particular, lower than the PDs of these ratings) and as high as possible for bad-quality ratings (in particular, higher than the PDs of these ratings). See appendix 1 for a graphical example. Second, although scarcely remarked in the literature (e.g., Blochlinger 2012), usual measures of discriminatory power are a function of the cross-sectional dependency between borrowers. This fact potentially represents an undesired property of traditional measures to the extent that the level and structure of default correlation is mainly a portfolio characteristic rather than a property intrinsic to the performance of CRMs.[6] Using the same framework employed in calibration testing, this paper proposes and discusses tests of "rating" discriminatory power that (i) can be seen as a necessary requisite to perfect calibration and (ii) are not a function of the default dependency structure. Power of these tests is also discussed, including results of power dominance between distinct proposed tests.

This text is organized as follows. Section 2 develops a default rate asymptotic probabilistic model (DRAPM) upon which validation will be discussed. The model leads to a unified theoretical framework for checking calibration and discriminatory power. Section 3 discusses briefly the formulation of the testing problem for CRM validation. The discussion of calibration testing, both one-sided and two-sided, is contained in section 4. Theoretical aspects of discriminatory power testing are investigated in section 5. Section 6 contains a Monte Carlo analysis of the finite sample properties of DRAPM and their consequences for calibration testing. Section 7 concludes.

2. The Default Rate Asymptotic Probabilistic Model (DRAPM)

The model of this section provides a default rate probability distribution upon which statistical testing is possible. It is based on

[6]It is not solely a portfolio characteristic because default correlation among the ratings potentially depends on the design of the CRM too.

an extension of the Basel II underlying model of capital requirement. In fact, this paper generalizes the idea first proposed by Balthazar (2004), of using the Basel II model for validation, to a multi-rating setting.[7] The applied extension is close to Demey et al. (2004)[8] and refers to including an additional systemic factor for each rating. While in Basel II the reliance on a single factor is crucial to the derivation of portfolio-invariant capital requirements (cf. Gordy 2003), for validation purposes a richer structure is necessary to allow for non-singular variance matrix among the ratings, as becomes clearer ahead in this section.

The formulation of DRAPM starts with a decomposition of z_{in}, the normalized return on assets of a borrower n with rating i. Close in spirit to the Basel II model, z_{in} is expressed as

$$z_{in} = \rho_B^{1/2} x + (\rho_W - \rho_B)^{1/2} x_i + (1 - \rho_W)^{1/2} \varepsilon_{in}, \tag{1}$$

for each rating $i = 1 \ldots I$ and each borrower $n = 1 \ldots N$, where x, x_i, ε_{ij} $(i = 1 \ldots I,\ j = 1 \ldots N)$ are independent and standard normal distributed.

Above, x represents a common systemic factor affecting the asset return of all borrowers, x_i a systemic factor affecting solely the asset return of borrowers with rating i, and ε_{in} an idiosyncratic shock. The parameters ρ_B and ρ_W lie in the interval [0 1]. Note that $\mathrm{Cov}(z_{in}, z_{jm})$ is equal to ρ_W if $i = j$ and to ρ_B otherwise, so that ρ_W represents the "within-rating" asset correlation and ρ_B the "between-rating" asset correlation. The Basel II model (and the validation approaches that are based on it such as Balthazar 2004) is a particular case of DRAPM when $\rho_W = \rho_B$. In other words, there is no systemic factor associated with rating i in the latter.

The model description continues with the statement that a borrower n with rating i defaults at the end of the forecasting time horizon if $z_{in} < \Phi^{-1}(PD_i)$ at that time, where Φ denotes the standard

[7] This idea is also adopted in Miu and Ozdemir (2008) and Bluemke (2013), among others. The reader is referred to BCBS (2005a) for a detailed presentation of the Basel II underlying model.

[8] The purpose of Demey et al. (2004) is to estimate correlations, while the focus here is on developing a minimal non-degenerate multivariate structure useful for testing.

normal cumulative distribution function.[9] Consequently, the conditional probability of default $PD_i(\mathbf{x})$, where $\mathbf{x} = (x, x_1, \ldots, x_i)'$ denotes the vector of systemic factors, can be expressed by

$$PD_i(\mathbf{x}) \equiv \mathrm{Prob}(z_{in} < \Phi^{-1}(PD_i)|\mathbf{x}) = \Phi((\Phi^{-1}(PD_i) - \rho_B^{1/2}x - (\rho_W - \rho_B)^{1/2}x_i)/(1-\rho_W)^{1/2}). \quad (2)$$

Let DR_{iN} denote the default rate computed using a sample of N borrowers with rating i at the start of the forecasting horizon. It is easy to see, as in Gordy (2003), that

$$\Phi^{-1}(\mathbf{DR_N}) - \Phi^{-1}(\mathbf{PD}(\mathbf{x})) \rightarrow 0 \text{ a.s. when } N \rightarrow \infty, \quad (3)$$

where $\mathbf{DR_N} = (DR_{1N}, DR_{2N}, \ldots, DR_{IN})'$ and $\mathbf{PD}(\mathbf{x}) = (PD_1(\mathbf{x}), PD_2(\mathbf{x}), \ldots, PD_i(\mathbf{x}))'$.

More concretely, the limiting default rate joint distribution is

$$\Phi^{-1}(\mathbf{DR}) \approx N(\boldsymbol{\mu}, \textstyle\sum), \quad (4)$$

where $\mu_i = \Phi^{-1}(PD_i)/(1-\rho_W)^{1/2}$, $\sum_{ij} = \rho_W/(1-\rho_W)$ if $i = j$, and $\sum_{ij} = \rho_B/(1-\rho_W)$ otherwise.

This asymptotic distribution is a multi-rating extension of the univariate limiting distribution presented in Gordy (2003) and also analyzed in Vasicek (2002). That is the distribution upon which all the tests of this paper will be derived. A limiting normal distribution is mathematically convenient to the derivation of likelihood ratio multivariate tests. The cost to be paid is that the approach is asymptotic, so that the discussions and results of this paper are not suitable for CRMs with a small number of borrowers per rating, such as, for example, rating models for large corporate exposures. Even for moderate numbers of borrowers, section 6 reveals that the departure from the asymptotic limit can be substantial, significantly

[9]Note that the probability of this event is therefore, by construction, PD_i. Without generalization loss, PD_i is assumed to increase in i. The characterization of default as the event defined by the fall of the (normalized return of) assets below a certain threshold is motivated by the structural approach to credit risk modeling developed by Black and Scholes (1973) and Merton (1974). Other implications of structural models for default probabilities can be found, for example, in Leland (2004).

altering the theoretical size and power of the tests. Application of the tests of the next sections should then be extremely careful.

Some comments on the choice of the form of $\sum$ are warranted.[10] To the extent that borrowers of each rating present similar distributions of economic and geographic sectors of activity, which ultimately govern borrowers' asset correlations, ρ_B is likely to be close to ρ_W, as this situation resembles the single systemic factor case. Nevertheless, it is reasonable to assume $0 < \rho_B < \rho_W$, in opposition to $\rho_B = \rho_W$, the implicit assumption of previous Basel II-like approaches, in order to leave open the possibility of some degree of association between rating PDs and borrowers' sectors of activity. As a result, borrowers in the same rating are considered to behave more dependently than borrowers in different ratings, because the profile of borrowers' sectors of activity is likely more homogeneous within than between ratings. Indeed, more realistic modeling is likely to require a higher number of asset correlation parameters and a portfolio-dependent approach; therefore the choice of just a pair of correlation parameters is regarded here as a practical compromise for general testing purposes. Furthermore, notice that $\rho_B \neq \rho_W$ is crucial to guarantee a non-singular matrix $\sum$ and, therefore, a non-degenerate asymptotic distribution. A singular matrix in the context of equation (4) would mean that the default rates of all ratings are asymptotically transformations of one another, which is unrealistic for joint validation purposes.

This paper further assumes that the correlation parameters ρ_W and ρ_B are known. The typically small number of years that banks have at their disposal suggests that the inclusion of correlation estimation in the testing procedure is not feasible, as it would diminish considerably the power of the tests. Instead, this paper relies on the Basel II Accord to extract some information on correlations.[11] By matching the variances of the non-idiosyncratic parts of the asset returns in the Basel II and DRAPM models, ρ_W can be seen as the

[10] Note that the structure of $\sum$ defines DRAPM more concretely than the chosen decomposition of the normalized asset return, because the decomposition is not unique given $\sum$.

[11] An important distinction to the Basel II model or Balthazar (2004), however, is that this paper does not make correlations dependent on the rating. In fact, the empirical literature on asset correlation estimation contains ambiguous results on this sensitivity.

asset correlation parameter present in the Basel II formula and in studies that make use of a single systemic factor. For corporate borrowers, for example, the Basel II Accord chooses $\rho_W \in [0.12\ 0.24]$. On the other hand, as the configuration present in equation (1) is new, there is no available information on ρ_B. Sensitivity analysis of the power of the tests on the choices of both ρ_W and ρ_B parameters is carried out in section 4. It should be noted, however, that the supervisory authority may have a larger set of information to estimate correlations and/or may even desire to set their values publicly for testing purposes.

Finally, serial independency is assumed for the annual default rate time series. Therefore, the (Φ^{-1}-transformed) average annual default rate, used as the test statistic for the tests of the next sections, has the normal distribution above, with $\sum/\mathrm{Y}$ in place of $\sum$, where Y is the number of years available to *backtest*. According to BCBS (2005b), serial independency is less inadmissible than cross-sectional independency. Furthermore, Blochinger (2012) argues that if the anticipated parts of the systemic factors are already factored into the allocations of borrowers to rating PDs, the resulting rating default rates are indeed serially independent.

3. The Formulation of the Testing Problem

Any configuration of a statistical test should start with the definitions of the null hypothesis $\mathrm{H_o}$ and the alternative one $\mathrm{H_1}$. In testing a CRM, a crucial decision refers to where the hypothesis "the rating model is correctly specified" should be placed.[12] If the bank/supervisor only wishes to abandon this hypothesis if data strongly suggests it is false, then the "correctly specified" hypothesis should be placed under $\mathrm{H_0}$, as in Balthazar (2004), BCBS (2005b), Miu and Ozdemir (2008), and Blochlinger (2012), among others. But if the bank/supervisor wants to know if the data provided enough evidence confirming the CRM is correctly specified, then this hypothesis should be placed in $\mathrm{H_1}$ and its opposite in $\mathrm{H_o}$. The reason is that the result of a statistical test is reliable knowledge only when the null hypothesis is rejected, usually at a low significance

[12] For this general discussion, one can think of "correctly specified" as meaning either correct calibration or good discriminatory power.

level. The latter option is pursued throughout this paper. Thus the probability of accepting an incorrect CRM will be the error to be controlled for at the significance level α.

To be precise, Bluemke (2013) also tries to control for the error of accepting an incorrect CRM but, in placing this hypothesis under the null, he is led to try to limit the type II error, which is generally not liable to uniform limitation. Indeed, Bluemke (2013) restricts the error II limitation to a single point of the alternative H_1. By reversing the role of the hypotheses found in previous validation approaches, this paper seems to be the first to uniformly control for the error of accepting an incorrect CRM.

Placing the "correctly specified" hypothesis under H_1 has immediate consequences. For a statistical test to make sense, H_0 usually needs to be defined by a closed set and H_1, therefore, by an open set.[13] This implies that the statement that "the CRM is correctly specified" needs to be translated into some statement about the parameters' PD_is lying in an *open* set—in particular, there shouldn't be equalities defining H_1 and the inequalities need to be strict. It is, for example, statistically inappropriate to try to conclude that the PD_is are equal to the bank-postulated values. In cases like that, the solution is to enlarge the desired conclusion by means of the concept of an indifference region. The configuration of the indifference region should convey the idea that the bank/regulator is satisfied with the eventual conclusion that the true **PD** vector lies there. In the previous case, the indifference region could be formed, for example, by open intervals around the postulated PD_is. The next sections make use of the concept to a great extent. At this point it is desirable only to remark that the feature of an indifference region shouldn't be seen as a disadvantage of the approach of this paper. Rather, it reflects better the fact that not necessarily all the borrowers in the same rating i have exactly the same theoretical PD_i and that it is, therefore, more realistic to see the ratings as defined by PD intervals.[14]

[13] H_0 and $H_0 \cup H_1$ need to be closed sets in order to guarantee that the maximum of the likelihood function is attained.

[14] However, in the context of Basel II, ratings need not be related to PD intervals but merely to single PD values. In light of this study's approach, this represents a gap in information needed for validation.

4. Calibration Testing

This section distinguishes between one-sided and two-sided tests for calibration. One-sided tests (which are only concerned about $\mathrm{PD_i}$s being sufficiently high) are useful to the supervisory authority by allowing to conclude that Basel II capital requirements derived by the approved PD estimates are sufficiently conservative in light of the banks' realized default rates. From a broader view, however, not only is excess of regulatory capital undesirable by banks, but also BCBS (2006b) states that the PD estimates should ideally be consistent with the banks' managerial activities such as credit granting and credit pricing.[15] To accomplish these goals, PD estimates must, without adding distortions, reflect the likelihood of default of every rating, something to be verified more effectively by two-sided tests (which are concerned about $\mathrm{PD_i}$s being within certain ranges). Unfortunately, the difficulties present in two-sided calibration testing are greater than in one-sided testing, as indicated ahead in this section. The analysis of one-sided calibration testing starts the section.

4.1 One-Sided Calibration Testing

Based on the arguments of the previous section about the proper roles of $\mathrm{H_o}$ and $\mathrm{H_1}$, the formulation of a one-sided calibration test is proposed below. Note that the desired conclusion, configured as an intersection of strict inequalities, is placed in $\mathrm{H_1}$.

$$\begin{aligned}\mathrm{H_o} &: PD_i \geq u_i \text{ for some i} = 1 \ldots \mathrm{I}\\ \mathrm{H_1} &: PD_i < u_i \text{ for every i} = 1 \ldots \mathrm{I},\end{aligned}$$

where $PD_i \equiv \Phi^{-1}(\mathrm{PD_i})$ and $u_i \equiv \Phi^{-1}(\mathrm{u_i})$. (This convention of representing Φ^{-1}-transformed figures in italic is followed throughout the rest of the text.)[16]

[15]More specifically, if the PDs used as inputs to the regulatory capital differ from the PDs used in managerial activities, at least some consistency must be verified between the two sets of values for validation purposes; cf. BCBS (2006b).

[16]As Φ^{-1} is strictly increasing, statements about italic figures imply equivalent statements about non-italic figures.

Here u_i is a fixed known number that defines an indifference acceptable region for PD_i. Its value should ideally be slightly larger than the value postulated for PD_i so that the latter is within the indifference region. Besides, u_i should preferably be smaller than the value postulated for PD_{i+1} so that at least the rejection of H_0 could conclude that $PD_i <$ postulated PD_{i+1}.[17] That is also an advantage of this paper's approach, since the monotonicity of PDs between individual rating grades is not always ensured in the methods proposed in the literature (e.g., Bluemke 2013).

According to DRAPM and based on the results of Sasabuchi (1980) and Berger (1989), which investigate the problem of testing homogeneous linear inequalities concerning normal means, a size-α critical region can be derived for the test.[18]

Reject H_0 (i.e., validate the CRM) if

$$\overline{DR}_i \le u_i/(1-\rho_W)^{1/2} - z_\alpha(\rho_W/(Y(1-\rho_W)))^{1/2}$$
$$\text{for every } i = 1 \ldots I, \tag{5}$$

where $\overline{DR}_i = \frac{\sum_{y=1}^{Y} \Phi^{-1}(DR_{iy})}{Y}$ is the (transformed) average annual default rate of rating i, and $z_\alpha = \Phi(1-\alpha)$ is the $1-\alpha$ percentile of the standard normal distribution.[19]

This test is a particular case of a min test, a general procedure that calls for the rejection of a union of individual hypotheses if each one of them is rejected at level α. In general, the size of a min test will be much smaller than α, but the results of Sasabuchi (1980) and Berger (1989) guarantee that the size is exactly α for the previous one-sided calibration test.[20] This means that the CRM is validated at size α if each PD_i is validated as such.

A min test has several good properties. First, it is uniformly more powerful (UMP) among monotone tests (Laska and Meisner

[17]As banks have the capital incentive to postulate lower PDs, one could argue that $PD_i <$ postulated PD_{i+1} also leads to $PD_i <$ true PD_{i+1}. Specific configurations of u_i are discussed later in the section.

[18]Size of a test is the maximum probability of rejecting H_0 when it is true.

[19]This definition of $\overline{DR}_i$ is used throughout the paper.

[20]More formally, this is the description of a union-intersection test, of which the min test is a particular case when all the individual critical regions are intervals not limited on the same side.

1989), which gives a solid theoretical foundation for the procedure since monotonicity is generally a desired property.[21] Second, as the transformed default rate variables are asymptotically normal in DRAPM, the min test is also asymptotically the likelihood-ratio test (LRT). Finally, the achievement of size α is robust to violation of the assumption of normal copula for the transformed default rates (Wang, Hwang, and Dasgupta 1999) so that, for size purposes, the requirement of *joint* normality for the systemic factors can be relaxed.

From a practical point of view, it should be noted that the decision to validate or not validate the CRM does not depend on the parameter ρ_B, which is useful for applications since ρ_B is not present in Basel II framework and so there is not much knowledge about its reasonable values. However, the power of the test—i.e., the probability of validating the CRM when it is correctly specified—does depend on ρ_B. The power is given by the following expression.

$$\begin{aligned} \text{Power} = \Phi_I(&-z_\alpha + (u_1 - PD_1)/(\rho_W/Y)^{1/2}, \ldots, \\ &- z_\alpha + (u_i - PD_i)/(\rho_W/Y)^{1/2}, \ldots, \\ &- z_\alpha + (u_I - PD_I)/(\rho_W/Y)^{1/2}; \rho_B/\rho_W), \end{aligned} \tag{6}$$

where $\Phi_I(\ldots; \rho_B/\rho_W)$ is the cumulative distribution function of an I^{th}-variate normal of mean 0, variances equal to 1, and covariances equal to ρ_B/ρ_W.

Berger (1989) remarks that if the ratio ρ_B/ρ_W is small, then the power of this test can be quite low for the PD_is only slightly smaller than u_is and/or a large number of ratings I. This is intuitive, as a low ratio ρ_B/ρ_W indicates that ex post information about one rating does not contain much information about other ratings and so is less helpful to conclude for validation. On the other hand, as previously noted in section 2, DRAPM is more realistic when ρ_B/ρ_W is close to 1 so that the referred theoretical problem becomes less relevant in the practical case.

More generally, it is easy to see that the power increases when PD_is decrease, u_is increase, Y increases, I decreases, ρ_B increases, or

[21] In the context of this paper, a test is monotone if the fact that average annual default rates are in the critical region implies that smaller average default rates are still in the critical region. Monotonicity is further discussed later in the paper.

Table 1. $u_i \times PD_i$

PD_i(%)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
u_i (%)	2	4	6	8	9	11	12	14	15	17	18	20	21	22	24	25	26	28	29	30

Note: PD_i validated if $H_0 : PD_i \geq u_i$ rejected in favor of $H_1 : PD_i < u_i$, considering the base scenario: $Y = 5$, $\rho_W = 0.15$, $\alpha = 15\%$, and $\beta = 80\%$, where Y = number of years, ρ_W = asset correlation, α = size of the test, and β = power at the true $PD \in H_1$.

ρ_W decreases.[22] In fact, it is worth examining the trade-off between the configuration of the indifference region in the form of the u_is and the attained power. If high precision is demanded (u_is close to postulated PD_is), then power must be sacrificed; if high power is demanded (u_is far from postulated PD_is), then precision must be sacrificed. Some numerical examples are analyzed below in order to provide further insights on this trade-off.

The case I = 1 represents an upper bound to the power expression above. In this case, for a desired power of β when the probability of default is exactly equal to the postulated PD, it is true that

$$u - PD = (z_\alpha - z_\beta) \times (\rho_W/Y)^{1/2}. \tag{7}$$

In a base-case scenario given by $Y = 5$, $\rho_W = 0.15$, $\alpha = 15\%$, and $\beta = 80\%$, the right-hand side of the previous equation is approximately equal to 0.32. This scenario is considered here sufficiently conservative, with a realistic balance between targets of power and size. In this case, it holds that

$$u_i = \Phi(0.32 + \Phi^{-1}(PD_i)). \tag{8}$$

Table 1 displays pairs of values of u_i and PD_i that conform to the equality above.

As, in a multi-rating context, any reasonable choice of u_i must satisfy $u_i \leq PD_{i+1}$, table 1 illustrates, for the numbers of the base-case scenario, an approximate lower bound for PD_{i+1} in terms of

[22] Obviously, the power also increases when the level α increases.

Table 2. PDs (%) Chosen According to u_i Specification and CRM Design

	PD_is Follow Arithmetic Progression		PD_is Follow Geometric Progression	
	$u_i = PD_{i+1}$	$u_i = (PD_{i+1} + PD_i)/2$	$u_i = PD_{i+1}$	$u_i = (PD_{i+1} + PD_i)/2$
I = 3	1.22, 11.82, 22.42	6.52, 17.12, 27.72	1.22, 3.66, 11	1.83, 5.5, 16.5
I = 4	2, 9.5, 17, 24.5	5.75, 13.25, 20.75, 28.25	2, 4, 8, 16	2.66, 5.33, 10.66, 21.33

Notes: CRMs validated if $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ rejected in favor of $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$. Distinct CRMs have the same u_0 and u_1 for each I = 3,4, where I is the number of ratings.

PD_i.[23] More generally, table 1 provides examples of whole rating scales that conform to the restriction $PD_{i+1} \geq u_i$, e.g., $PD_1 = 1\%$, $PD_2 = 2\%$, $PD_3 = 4\%$, $PD_4 = 8\%$, $PD_5 = 14\%$, $PD_6 = 22\%$, $PD_7 = 36\%$ (note the shaded cells). Note that such conforming rating scales must possess increasing PD differences between consecutive ratings (i.e., $PD_{i+1} - PD_i$ increasing in i), a characteristic found indeed in the design of many real-world CRMs. Therefore, DRAPM suggests a validation argument in favor of that design choice. Notice that this feature of increasing PD differences is directly related to the non-linearity of Φ, which in turn is a consequence of the asymmetry and kurtosis of the distribution of the untransformed default rate.

To further investigate the feature of increasing PD differences and choices of $\mathbf{u} = (u_1, u_2, \ldots, u_I)'$ in the one-sided calibration test, the cases I = 3 and I = 4 are explicitly analyzed in the sequence. For each I, four CRMs are considered, with their PD_is depicted in table 2. CRMs of table 2 can have PD_is following either an arithmetic progression or a geometric progression. Besides, two strategies of configuration of the indifference region are considered: a liberal one with $u_i = PD_{i+1}$ and a more precise one with $u_i = (PD_{i+1} + PD_i)/2$. In order to allow for a fair comparison of power among distinct CRMs, PD_is figures of table 2 are chosen with the purpose that the resulting sets of ratings of each CRM cover equal ranges in the

[23] This is approximate because the computation was based on I = 1. (In fact, the true attained power in a multi-rating setup is smaller.) Also, the discussion of this paragraph assumes that true **PD** = postulated **PD**.

Table 3. Power Comparison among CRM Designs and u_i Choices, I = 3

$\rho_B/\rho_W = 0.8, \alpha = 0.15$

	PD_is Follow Arithmetic Progression		**PD_is Follow Geometric Progression**	
	$u_i = PD_{i+1}$	$u_i = (PD_{i+1}+PD_i)/2$	$u_i = PD_{i+1}$	$u_i = (PD_{i+1}+PD_i)/2$
$\rho_W = 0.12$, Y = 10	0.97	0.60	1.00	0.95
In-between	0.86	0.43	0.98	0.81
$\rho_W = 0.18$, Y = 5	0.72	0.34	0.91	0.67

Notes: Power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, computed at the postulated PDs of table 2 (case I = 3). ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, in-between scenario characterized by $(\rho_W/Y) = 0.15^2$, Y = number of years, α = size of test.

PD scale. More specifically, this goal is interpreted here as all CRMs having equal u_0 and u_i.[24]

The power figures of the one-sided calibration test at the postulated **PDs** are shown in tables 3 and 4, according to values set to parameters ρ_W and Y. The values of these parameters are chosen considering three feasible scenarios: a favorable one characterized by ten years of data and a low within-rating correlation of 0.12, an unfavorable one characterized by the minimum number of five years prescribed by Basel II (cf. Basel 2006a) and a high ρ_W at 0.18, and an in-between scenario.[25]

Tables 3 and 4 show that CRMs with the feature of increasing $(PD_{i+1} - PD_i)$ usually achieve significantly higher levels of power than CRMs with equally spaced PD_is, confirming the intuition derived from table 1. The tables also reveal that, even when

[24] u_0 corresponds to the fictitious PD_o. In table 2, PD_o can be easily figured out from the constructional logic of the PD_i progression. For the construction of the CRMs of table 2, $u_0 = 1.22\%$ and $u_3 = 33\%$ for I = 3, and $u_0 = 2\%$ and $u_4 = 32\%$ for I = 4. Furthermore, the ratio of the PD_i geometric progression is set equal to 3 for I = 3 and 2 for I = 4.

[25] As ρ_B/ρ_W is fixed in tables 3 and 4, what matters for the power calculation is just the ratio (ρ_W/Y). Therefore, the in-between scenario can be thought of as characterized by adjusting both Y and ρ_W or just one of them. In tables 3 and 4 it is given by $(\rho_W/Y)^{1/2} = 0.15$.

Table 4. Power Comparison among CRM Designs and u_i Choices, I = 4

$\rho_B/\rho_W = 0.8, \alpha = 0.15$

	PD$_i$s Follow Arithmetic Progression		**PD$_i$s Follow Geometric Progression**	
	$\mathbf{u_i = PD_{i+1}}$	$\mathbf{u_i = (PD_{i+1}+ PD_i)/2}$	$\mathbf{u_i = PD_{i+1}}$	$\mathbf{u_i = (PD_{i+1}+ PD_i)/2}$
$\rho_W = 0.12$, Y = 10	0.82	0.39	0.95	0.68
In-between	0.62	0.28	0.81	0.48
$\rho_W = 0.18$, Y = 5	0.49	0.22	0.65	0.37

Notes: Power of the test $H_0 : PD_i \geq u_i$ for some i = 1...I against $H_1 : PD_i < u_i$ for every i = 1...I, computed at the postulated PDs of table 2 (case I = 4). ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, in-between scenario characterized by $(\rho_W/Y) = 0.15^{1/2}$, Y = number of years, α = size of test.

solely focusing on the former, more demanding requirements for u_i (cf. $u_i = (PD_{i+1}+ PD_i)/2$) may produce overly conservative tests, with, for example, power on the level of only 37 percent. Therefore liberal strategies for u_i (cf. $u_i = PD_{i+1}$) seem to be necessary for realistic validation attempts, and attention is focused on these strategies in the remainder of this section. Further from the tables, the power is found to be very sensitive to the within-rating correlation ρ_W and to the number of years Y. It can increase more than 80 percent from the worst to the best scenario (cf. last column of table 4).

While in previous tables the between-rating correlation parameter ρ_B is held fixed, tables 5 and 6 examine its effect, along a set of feasible values, on the power of the test. Power is computed at the postulated **PDs** of CRMs of table 2 with $u_i = PD_{i+1}$, I = 4 and for the in-between scenario of parameters of ρ_W and Y. The tables show just a minor effect of ρ_B, regardless of the size of the test and the CRM design. Therefore, narrowing down the uncertainty in the value of ρ_B value is not of great importance if just approximate levels of power are desired at postulated **PDs**. The elements that indeed drive the power of the test are unveiled in the next analysis.

Tables 7 and 8 provide insights on the relative role played by the different ratings on the power. Power is computed at postulated **PDs** for a sequence of four embedded CRMs, starting with the CRM with equally spaced PDs of the second line of table 7 (the CRM with

Table 5. Effect of ρ_B when PD_is Follow Arithmetic Progression

$u_i = PD_{i+1}$, $(\rho_W/Y)^{1/2} = 0.15$, $I = 4$

	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
$\rho_B/\rho_W = 0.6$	0.32	0.47	0.58
$\rho_B/\rho_W = 0.7$	0.35	0.50	0.60
$\rho_B/\rho_W = 0.8$	0.38	0.52	0.62
$\rho_B/\rho_W = 0.9$	0.41	0.55	0.65

Notes: Power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, computed at the postulated PDs of table 2 (case $I = 4$, $u_i = PD_{i+1}$, PD_is follow arithmetic progression). ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, Y = number of years, α = size of the test.

Table 6. Effect of ρ_B when PD_is Follow Geometric Progression

$u_i = PD_{i+1}$, $(\rho_W/Y)^{1/2} = 0.15$, $I = 4$

	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
$\rho_B/\rho_W = 0.6$	0.54	0.69	0.78
$\rho_B/\rho_W = 0.7$	0.56	0.71	0.79
$\rho_B/\rho_W = 0.8$	0.60	0.73	0.81
$\rho_B/\rho_W = 0.9$	0.62	0.74	0.82

Notes: Power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, computed at the postulated PDs of table 2 (case $I = 4$, $u_i = PD_{i+1}$, PD_is follow geometric progression). ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, Y = number of years, α = size of the test.

increasing PD differences of the second line of table 8). Each next CRM in table 7 (table 8) is built from its antecedent by dropping the less risky (riskiest) rating. Power is computed for the in-between scenario and $u_i = PD_{i+1}$. The tables reveal that as the number of ratings diminishes, the power increases just to a minor extent, provided the riskiest (less risky) ratings are always kept in the CRM. Thus it can be said that in table 7 (table 8) the highest (lowest) PD_is drive the power of the test. This is partly intuitive because the highest (lowest) PD_is correspond to the smallest differences $(u_i - PD_i)$ in

Table 7. Influence of Distinct PD_is on Power

PD_is follow arithmetic progression;
$\rho_B/\rho_W = 0.6$; $(\rho_W/Y)^{1/2} = 0.15$; $u_i = PD_{i+1}$

PD_is	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
2%, 9.5%, 17%, 24.5%	0.32	0.47	0.58
9.5%, 17%, 24.5%	0.32	0.47	0.58
17%, 24.5%	0.34	0.49	0.59
24.5%	0.44	0.58	0.68

Notes: Power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, for $I = 4 \ldots 1$, computed at the PDs of the first column. ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, Y = number of years, α = size of the test.

Table 8. Influence of Distinct PD_is on Power

PD_is follow geometric progression;
$\rho_B/\rho_W = 0.6$; $(\rho_W/Y)^{1/2} = 0.15$; $u_i = PD_{i+1}$

PD_is	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
2%, 4%, 8%, 16%	0.54	0.69	0.78
2%, 4%, 8%	0.54	0.69	0.78
2%, 4%	0.56	0.71	0.79
2%	0.65	0.77	0.84

Notes: Power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, for $I = 4 \ldots 1$, computed at the PDs of the first column. ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, Y = number of years, α = size of the test.

the CRMs of table 7 (table 8) and because distinct PD_is contribute to the power differently just to the degree their differences $(u_i - PD_i)$ vary.[26] The surprising part of the result refers to the degree of relative low importance of the dropped PD_is: the variation of power between I = 1 and I = 4 can be merely around 10 percent. This

[26]It is easy to see that for the CRMs with equally spaced PD_is, $(u_i - PD_i)$ is trivially constant in i but the Φ^{-1}-transformed difference $(u_i - PD_i)$ decreases in i. For the CRMs with increasing $(PD_{i+1} - PD_{i,})$, $(u_i - PD_i)$ trivially increases in i and the Φ^{-1}-transformed difference $(u_i - PD_i)$ increases in i too.

latter observation should be seen as a consequence of the functional form of DRAPM, particularly the choice of the normal copula for the (transformed) default rates and the form of Σ.[27]

A message embedded in the previous tables is that in some quite feasible cases (e.g., Y = 5 years available at the database, $\rho_W = 0.18$ reflecting the portfolio default volatility, $\alpha < 15\%$ desired), the one-sided calibration test can have substantially low power (e.g., lower than 50 percent at the postulated **PD**). Another related problem refers to the test not being similar on the boundary between the hypotheses and therefore biased (if $I > 1$).[28] To cope with these *deficiencies*, the statistical literature contains some proposals of non-monotone uniformly more powerful tests for the same problem, such as in Liu and Berger (1995) and McDermott and Wang (2002). The new tests are constructed by carefully enlarging the rejection region in order to preserve the size α. The enlargement trivially implies power dominance. The new tests have two main disadvantages though. First, from a supervisory standpoint, non-monotone rejection regions are harder to defend on an intuitive basis because they imply that a bank could pass from a state of validated CRM to a state of non-validated CRM if default rates for some of the ratings *decrease.* Second, from a theoretical point of view, Perlman and Wu (1999) note that the new tests do not dominate the original test in the decision-theoretic sense because the probability of validation under H_0 (i.e., when the CRM is incorrect) is also higher for them. The authors conclude that UMP tests should not be pursued at any cost, particularly at the cost of intuition. This is the view adopted in this study, so the new tests are not explored further in this paper.

Yet, one may try to include some prior knowledge in the formulation of the one-sided calibration test as a strategy for power improvement. Notice, first, that the size α of the test is attained when all but

[27]Bluemke (2013) also shows situations in which the power of validation tests is driven by a single rating, but he does not address the influence of the CRM design in determining what this rating is.

[28]A test is α-similar on a set A if the probability of rejection is equal to α everywhere there. A test is unbiased at level α if the probability of rejection is smaller than α everywhere in H_0 and greater than α everywhere in H_1. Every unbiased test at level α with a continuous power function is α-similar in the boundary between H_0 and H_1 (Gourieroux and Monfort 1995).

one of the PD_is go to 0 while the remaining one is set fixed at u_i.[29] This is probably a very unrealistic scenario against which the bank or the supervisor would like to be protected. The bank/supervisor may alternatively remove by assumption this unrealistic case from the space of **PD** possibilities and rather consider that part of the information to be tested is true. Notably, it can be assumed that the postulated PD_{i-1}, not 0, represents a lower bound for PD_i, for every rating i. A natural modification of the test consists then on replacing z_α with a smaller constant $c > 0$ to adjust to the removed unrealistic **PD** scenarios,[30] with resulting enlargement of the critical region and achievement of a more powerful test. (Recall the definition of the critical region in (5).) Hence, c is defined by the requirement that the size of the modified test (with c instead of z_α) in the reduced **PD** space is α. Similarly to Sasabuchi (1980), the determination of c needs the examination of only the **PD** vectors with all but one of their coordinates' PD_is equal to their lower bounds (the postulated PD_{i-1}s), and the remaining one, say PD_j, set at u_j, for j varying in 1...I. More formally,

$$\begin{aligned}&\mathrm{Max}_{1\leq j\leq I}(\Phi_I(-c + (u_1 - PD_0)/(\rho_W/Y)^{1/2}, \ldots, -c, \ldots,\\ &\quad - c + (u_I - PD_{I-1})/(\rho_W/Y)^{1/2}; \rho_B/\rho_W) = \alpha^{31},\end{aligned} \tag{9}$$

from which the value of c can be derived.

However, produced results indicate the previous modification approach is of limited efficacy to power improvement. More specifically, computed results indicate that the power increase is relevant only in the region of small (probably unrealistic) ratio ρ_B/ρ_W or for ambitious choices of u_i (i.e., close to PD_i). In the latter case, the increase is not sufficient, however, to the achievement of reasonable levels of power because the original levels are already too low (cf. table 1, for example). Those results are consistent with the intuition derived from the analysis of tables 7 and 8.

[29] Note $PD_i \rightarrow 0 \Rightarrow PD_i \rightarrow -\infty$. The limiting **PD** vector is in H_0 and, therefore, should not be validated. It has a probability of validation equal to α.

[30] As the coordinates of the input to the power function cannot go to infinity as before, $-c > -z_\alpha$ for the size to be achieved.

[31] PD_0 is here just a lower bound to PD_1. It could be $-\infty$ or defined subjectively based on accumulated practical experience. Note that the new critical region will now depend on ρ_B and that the calculation of c needs some computational effort.

On the other hand, one may also try to derive the LRT based on the restricted **PD** parameter space:

$$\begin{aligned}
\mathrm{H_o} : &\ PD_{\mathrm{i}} \geq u_{\mathrm{i}} \text{ for some i} = 1 \ldots \mathrm{I} \text{ and} \\
&\ PD_{\mathrm{i}} \geq \text{postulated } PD_{\mathrm{i}-1} \text{ for every i} = 1 \ldots \mathrm{I} \\
\mathrm{H_1} : &\ PD_{\mathrm{i}} < u_{\mathrm{i}} \text{ for every i} = 1 \ldots \mathrm{I} \text{ and} \\
&\ PD_{\mathrm{i}} \geq \text{postulated } PD_{\mathrm{i}-1} \text{ for every i} = 1 \ldots \mathrm{I}.^{32}
\end{aligned}$$

The LRT will differ from the modification approach with respect to the information contained in the observed default rates. The LRT will have very small observed average default rates, providing lower relative evidence in favor of $\mathrm{H_1}$ because, by assumption, they cannot be explained by very small PDs.[33] Accordingly, the null distribution of the likelihood-ratio (LR) statistic doesn't need to put mass on those unrealistic **PD** scenarios. Unfortunately, to the best of the author's knowledge, the derivation of the LRT critical region for such a problem is lacking in the statistical literature. Its complexity arises from the facts that, in contrast to the original one-sided calibration test, $\mathrm{H_0}$ and $\mathrm{H_1}$ do not share the same boundary in $\Re^{\mathrm{I}}$ and that the boundary indeed shared is a limited set. Thus, it is reasonable to conjecture that the null distribution of the LR statistic will be fairly complicated. And similarly to the previous strategy, if $u_i >>$ postulated PD_{i-1} for most ratings, the increase in power is likely to negligible again.[34]

4.2 Two-Sided Calibration Testing

The section now comments on two-sided calibration testing, mostly from a theoretical perspective. Similarly to the one-sided version, the hypotheses of a two-sided test can be stated as follows:

---

[32] $\mathrm{H_1}$ need not be defined only by strict inequalities here since the union $\mathrm{H_0}$ U $\mathrm{H_1}$ does not span the full $\Re^{\mathrm{I}}$ space.

[33] Very small observed average default rates in the sense that $\Phi^{-1}(\mathrm{DR_i})/(1-\rho_{\mathrm{W}})^{1/2} < \Phi^{-1}(\text{postulated } \mathrm{PD_{i-1}})$.

[34] It is important to remark that if I is large, strategies of power improvement will generally have more chances of *relative* success, although they depart from lower original levels of power.

$$\mathrm{H_o} : PD_i \geq u_i \text{ or } PD_i \leq l_i \text{ for some } \mathrm{i} = 1 \ldots \mathrm{I}$$
$$\mathrm{H_1} : l_i < PD_i < u_i \text{ for every } \mathrm{i} = 1 \ldots \mathrm{I}.$$

Now the acceptable indifference region is defined by two parameters $\mathrm{u_i}$ and $\mathrm{l_i}$ for each rating i, with ideally $\mathrm{l_i} \geq$ postulated $\mathrm{PD_{i-1}}$ and $\mathrm{u_i} \leq$ postulated $\mathrm{PD_{i+1}}$. Under that formulation, the test belongs to the class of multivariate equivalence tests, which are tests designed to show similarity rather than difference and are widely employed in the pharmaceutical industry (under the denomination of bio-equivalent tests) to demonstrate that drugs are equivalent. Berger and Hsu (1996) comprehensively review the recent development of equivalence tests in the univariate case ($\mathrm{I} = 1$). The standard procedure to test univariate equivalence is the TOST test (two one-sided tests—called this because the procedure is equivalent to performing two size-α one-sided tests and concluding equivalence only if both reject). Wang, Hwang, and Dasgupta (1999) discuss the extension of TOST to the multivariate case, making use of the intersection-union method.[35] When applied to the DRAPM distribution, that extension results in the following critical region for the two-sided calibration test.[36]

Reject $\mathrm{H_o}$ (i.e., validate the CRM) if

$$\begin{aligned} & l_i/(1-\rho_{\mathrm{W}})^{1/2} + \mathrm{z}_\alpha(\rho_{\mathrm{W}}/(\mathrm{Y}(1-\rho_{\mathrm{W}})))^{1/2} \\ & \quad \leq \overline{DR}_i \leq u_i/(1-\rho_{\mathrm{W}})^{1/2} - -\mathrm{z}_\alpha(\rho_{\mathrm{W}}/(\mathrm{Y}(1-\rho_{\mathrm{W}})))^{1/2} \end{aligned} \tag{10}$$

for every $i = 1 \ldots \mathrm{I}$.

As the maximum power of the test occurs in the middle point of the cube $[l_i u_i]^{\mathrm{I}}$, it is reasonable to make the cube symmetric around the postulated **PD** (in other words, to make $u_i - PD_i = PD_i - l_i$ for every i), so that the highest probability of validating the CRM occurs exactly at the postulated **PD**. Additional configurations of the indifference region may include, as in the one-sided test, choosing $u_i = PD_{i+1}$ or $l_i = PD_{i-1}$ (but not both).

[35] Wang, Hwang, and Dasgupta (1999) also show that TOST is basically an LR test.

[36] The standard TOST is formulated assuming unknown variance, while the proposed two-sided calibration test of this paper assumes known variance. Therefore the reference to the term TOST encompasses here some freedom of notation.

Similarly to the one-sided test, the two-sided version has problems of lack of power and bias.[37] In this respect, the statistical literature contains some proposals for improving TOST (Berger and Hsu 1996; Brown, Hwang, and Munk 1998), which are again subject to criticism from an intuitive point of view by Perlman and Wu (1999).[38] Furthermore, an additional drawback of the two-sided test, in contrast to the original TOST, is its excess of conservatism because the test is only level α, while its size may be much smaller.[39] That observation indicates the magnified difficulty in performing two-sided calibration testing.

Yet, two additional approaches to testing multivariate equivalence deserve comments. The first one is developed by Brown, Casella, and Hwang (1995). Applied to the problem of $\mathbf{PD}$ calibration testing, it consists of accepting an alternative hypothesis H_1 (i.e., validating the CRM) if the Brown confidence set for the $\mathbf{PD}$ vector is entirely contained in H_1. The approach would allow the bank or the supervisor to separate the execution of the test from the task of defining an indifference region because H_1 configuration could be discussed at a later stage, after the knowledge of the *form* of the set. In particular, the confidence set can be seen as the smallest indifference region that still permits validation of the calibration. Brown, Casella, and Hwang (1995) propose an optimal confidence set in the sense that if the true $\mathbf{PD}$ vector is equal to the postulated one, then the expected volume of that set is minimal, which means that, in average terms, maximal precision is achieved when calibration is *exactly* right. The cost of this optimality is larger set volumes for $\mathbf{PDs}$ different from the postulated one. Munk and Pfluger (1999) show in simulation exercises that the power of Brown's procedure can be substantially lower than those of more standard tests, like the TOST, for a wide range of $\mathbf{PDs}$ close to the postulated one. Therefore, in light of the view of this paper that ratings could more

[37] If $I > 1$, the test is not similar on the boundary between the hypotheses and is therefore biased.

[38] However, in the case of calibration testing with known variance, the bias is not as pronounced as in the standard TOST with unknown variance, due to the impossibility of making the variance go to 0 as in Berger and Hsu (1996).

[39] It can be shown that the degree of conservatism depends on ρ_B. The reason for the discrepancy with the standard TOST relates again to the impossibility of making the variance go to 0 as in Berger and Hsu (1996).

realistically be seen as PD intervals, the benefit of the optimality at a single point is doubtful at a minimum. Consequently, Brown's approach is regarded here as of more theoretical than practical value to calibration testing.[40]

The second different approach to testing multivariate equivalence is developed by Munk and Pfluger (1999). So far, this paper has just considered rectangular sets in the H_1 statements of the calibration tests. The goal has been to show that the true **PD** lies in a rectangle or in a quadrant of the space $\Re^I$. The referred authors analyze instead the use of ellipsoidal alternatives for the multivariate equivalence problem, which, for purposes of calibration testing, can be exemplified as follows:

$$H_o : \boldsymbol{e}^t \mathbf{D} \boldsymbol{e} \geq \Delta$$

$$H_1 : \boldsymbol{e}^t \mathbf{D} \boldsymbol{e} < \Delta,$$

where $\boldsymbol{e} = \boldsymbol{PD}$ – postulated $\boldsymbol{PD}$, $\mathbf{D}$ is a positive definite matrix, which conceives a notion of distance in $\Re^I$, and Δ denotes a fixed tolerance bound. $\mathbf{D}$ and Δ define an indifference region for **PD**.

Munk and Pfluger (1999) advocate this formulation to allow the notion of equivalence to be interpreted as a combined measure of several parameters (e.g., a combination of the PD_is, $i = 1 \ldots I$). As a consequence, this implies that very good *marginal* equivalence (e.g., the true PD_1 is very close to the postulated PD_1) should allow larger indifference regions for the other parameters (e.g., the other PD_is). Conceptually, though, this point is hard to justify in the validation of CRMs unless miscalibration were necessarily derived from a systematic erroneous estimation of all the PD_is. Nevertheless, the view of this paper is that miscalibration could be rather rating specific. Furthermore, note that the rectangular alternatives already permit a lot of flexibility in allowing different indifference interval lengths for different ratings. Consequently, for purposes of calibration testing, ellipsoidal alternatives are regarded here more as a practical complication.[41]

[40] Other confidence set approaches to calibration testing are also possible. Some of them are, however, dominated by the multivariate TOST (Munk and Pfluger 1999).

[41] However, for purposes of power improvement, it still might be useful to investigate ellipsoidal alternatives inscribed or approximating rectangular alternatives. This investigation is not addressed in this paper.

5. Tests of Rating Discriminatory Power

One of the most traditional measures of discriminatory power is the area under the ROC curve (AUROC).[42] Let n and m be two distinct random borrowers with probabilities of default $\mathrm{PD_n}$ and $\mathrm{PD_m}$, respectively. Following Bamber (1975), AUROC is defined as

$$\begin{aligned}\mathrm{AUROC} = {} & \mathrm{Prob}(\mathrm{PD_n} > \mathrm{PD_m} |\ \text{n defaults and m doesn't}) \\ & + {}^1\!/_2 \cdot \mathrm{Prob}(\mathrm{PD_n} = \mathrm{PD_m} |\ \text{n defaults and m doesn't}).\end{aligned} \tag{11}$$

High values of AUROC (close to 1) are typically interpreted as evidence of good CRM discriminatory performance. However, the definition of AUROC as the probability of an event concerning the realizations of two (random) borrowers makes it a function not only of the **PD** vector but also of the default correlation structure.[43] To the extent that the CRM should not be held accountable for the effect of default dependency between borrowers, the AUROC measure of discrimination becomes distorted.[44] Blochlinger (2012) shows this distortion by means of a numerical example. The proposition below shows formally the dependency of AUROC on the asset correlation parameters.

PROPOSITION. Consider an extension of DRAPM in which (ρ_{ij}) is the matrix of asset correlations between borrowers of ratings i and j, i,j = 1...I. Let P(i,j) denote the probability of two random borrowers having ratings i and j and P(i) the probability of one random borrower having rating i. Then

$$AUROC = \frac{\sum_{i>j} \Phi_2\left(PD_i, -PD_j, -\rho_{ij}\right)\mathrm{P(i,j)} + \frac{1}{2}\sum_i \Phi_2\left(PD_i, -PD_i, -\rho_{ii}\right)\mathrm{P(i)}}{\sum_{i,j} \Phi_2\left(PD_i, -PD_j, -\rho_{ij}\right)\mathrm{P(i,j)}} \tag{12}$$

---

[42]ROC = receiver operating characteristic curve (cf. Bamber 1975). $0 \leq \mathrm{AUROC} \leq 1$.

[43]It is a function of the distribution of borrowers across the ratings too.

[44]Note that, in contrast, the definition of good calibration is always *purely* linked to the good quality of the **PD** vector, although the way to *empirically* conclude that will typically depend on the default correlation values, as shown in section 4.

Proof. See appendix 2.

Blochlinger (2012) proposes a measure of discriminatory power that is not a function of default dependency. However, in contrast to AUROC, it is a function of the portfolio-wide true (unknown) PD and, besides, his asymptotic testing results are based on a multiplicative form for the conditional PDs not obeyed by the Basel II model or DRAPM (equation (2) is not of multiplicative form). This section describes alternatives for tests of *rating* discriminatory power built upon the DRAPM distribution. The qualifying term *rating* is added purposefully to the traditional expression "discriminatory power" to emphasize that the property desired to be concluded/measured here is different from that embedded in traditional measures of discriminatory power. Rather than verifying that the ex post rating distributions of the default and non-default groups of borrowers are as separate as possible, the proposed tests of *rating* discriminatory power aim at showing that $\mathrm{PD_i}$ is a strictly increasing function of i. In other words, the discriminatory power should be present *at the rating level* or, more concretely, low-quality ratings should have larger $\mathrm{PD_i}$s. Note that this is a less stringent requirement than correct two-sided calibration and the alternative hypothesis here will, therefore, strictly contain the $\mathrm{H_1}$ of the two-sided calibration test.[45] In this sense, the fulfillment of good rating discriminatory power is consistent with the pursuit of correct calibration. Furthermore, as the proposed tests are based on hypotheses involving solely the **PD** vector, they are not functions of default correlations; consequently, they address the two pitfalls of traditional measures of discriminatory power that were discussed in the introduction. Finally, showing PD monotonicity along the rating dimension is also useful to corroborate the assumptions of some methods of PD inference on low default portfolios (e.g., Pluto and Tasche 2005).

This section distinguishes between a test of *general* rating discriminatory power and a test of *focal* rating discriminatory power. The former addresses a situation where the bank or supervisor is

[45]Provided $\mathrm{u_i} < \mathrm{l_{i+1}}$ for $\mathrm{i} = 1 \ldots \mathrm{I} - 1$, as expected in practical applications. On its turn, Blochlinger (2012) investigates a less stringent requirement than correct two-sided calibration but a more demanding one than PD monotonicity, namely that PD ratios between ratings equal specific constants.

uncertain about the increasing PD behavior along the whole rating scale, whereas the latter focuses on a pair of consecutive ratings. The formulation of the general test is proposed below.

$$\mathrm{H_o} : PD_i \geq PD_{i+1} \text{ for some i} = 1 \ldots \mathrm{I} - 1$$
$$\mathrm{H_1} : PD_i < PD_{i+1} \text{ for every i} = 1 \ldots \mathrm{I} - 1.$$

By viewing $PD_{i+1} - PD_{\mathrm{i}}$ as the unknown parameter to be estimated (up to a constant) by $DR_{i+1} - DR_i$ for every rating i, the previous test involves testing strict homogeneous inequalities about normal means. (The key observable variables are now default rate differences between consecutive ratings, rather than the default rates themselves, as in the one-sided calibration test.) So, similarly to the one-sided calibration test, a size-α likelihood-ratio critical region can be derived.

Reject $\mathrm{H_0}$ (i.e., validate the CRM) if

$$\overline{DR}_{i+1} - \overline{DR}_i > \mathrm{z}_\alpha (2(\rho_\mathrm{W} - \rho_\mathrm{B})/(\mathrm{Y}(1 - \rho_\mathrm{W})))^{1/2} \quad \text{for every i} = 1 \ldots \mathrm{I} - 1. \tag{13}$$

It is worth noting above that, differently from the calibration tests, there is no need for the configuration of an indifference region, as the desired $\mathrm{H_1}$ conclusion is already defined by strict inequalities. On the other hand, now the critical region and—therefore, the decision itself to validate the CRM—depends on the unknown parameter ρ_B. The Basel II case ($\rho_\mathrm{B} = \rho_\mathrm{W}$) represents the extreme liberal situation where just an observed increasing behavior of the average annual default rates along the rating dimension is sufficient to validate the CRM (regardless of the confidence level α), whereas the case $\rho_\mathrm{B} = 0$ places the strongest requirement in the incremental increase of the default rate averages along the rating scale.[46] In practical situations, the bank or the supervisor may want to determine the highest value of ρ_B such that the general test still validates the CRM and then check how this value conforms to its beliefs about reality.

When theoretically compared with the power of the one-sided calibration test, the power of the general test is notably affected by

[46] This is again intuitive, as low values of ρ_B mean that ex post information about one rating does not contain much information about other ratings.

a trade-off of three factors.[47] First, the fact that now the underlying normal variables are likely to have smaller variances ($\mathrm{Var}(DR_{\mathrm{i}+1} - DR_i) = 2(\rho_{\mathrm{W}} - \rho_{\mathrm{B}})/(1 - \rho_{\mathrm{W}}) < \mathrm{Var}(DR_i) = \rho_{\mathrm{W}}/(1 - \rho_{\mathrm{W}})$, provided $\rho_{\mathrm{B}}/\rho_{\mathrm{W}} > 1/2$) contributes to an increase in power. On the other hand, the now not positive underlying correlations ($\mathrm{Corr}(DR_{i+1} - DR_i, DR_j - DR_{j-1}) = -1/2$ if i = j and 0 otherwise, compared with $\mathrm{Corr}(DR_i,DR_j) = \rho_{\mathrm{B}}/\rho_{\mathrm{W}} > 0$ for $\mathrm{i} \neq j$) contributes to a decrease in power.[48] Finally, the presence of I – 1 statements in H_1, instead of I, implies a slight increase in power too. In general, the resulting dominating force is to be determined by the particular choices of ρ_{B}, ρ_{W}, and I. However, computed results indicate that discrimination test power will usually be larger than calibration power for CRM designs including both arithmetic and geometric progressions for the $\mathrm{PD_i}$s and reasonable specifications for the testing parameters.[49] Finally, as with calibration testing, similar comments on possible strategies for power improvement and their limitations apply here as well.

It is also worthwhile to discuss the situation where the bank or the supervisor is satisfied by the "general level" of rating discrimination except for a particular pair of consecutive ratings. Suppose the bank/supervisor wants to find evidence that two consecutive ratings (say ratings 1 and 2, without loss of generality) indeed distinguish the borrowers in terms of their creditworthiness. From a supervisory standpoint, a suspicion of regulatory arbitrage may, for instance, motivate the concern.[50] To examine this issue, this section formulates a test of focal rating discriminatory power, whose hypotheses are stated as follows.[51]

---

[47] Similarly to the calibration case, the power expression can be easily derived.

[48] Therefore, not necessarily validating rating discriminatory power is easier than validating (one-sided) calibration.

[49] Also, computed results in line with previous calibration findings indicate that CRMs whose $\mathrm{PD_i}$s follow geometric progression will generally achieve higher levels of power than when $\mathrm{PD_i}$s follow arithmetic progression and their power is basically driven by the first pairs of consecutive ratings, in the high-credit-quality part of the scale.

[50] Suspicion of regulatory arbitrage may derive from a situation where large credit risk exposures are apparently rated with slightly better ratings so that the resulting capital charge of Basel II is diminished.

[51] The discussion of this section is easily generalized to the situation where more than one pair of consecutive ratings are to have their rating discriminatory power verified.

$$\mathrm{H_o}: PD_1 = PD_2 \leq PD_3 \leq \ldots \leq PD_I$$
$$\mathrm{H_1}: PD_1 < PD_2 \leq PD_3 \leq \ldots \leq PD_I.$$

From a mathematical point of view, the development of the likelihood-ratio test for such a problem is more complex than the majority of the tests considered so far in this paper, because now the union of the null and the alternative hypotheses do not span the full $\Re^{\mathrm{I}}$, nor do the hypotheses share a common boundary. But, in contrast to the section 4 one-sided calibration LRT under **PD** restriction, now both $\mathrm{H_0}$ and $\mathrm{H_1}$ are convex cones. This implies that the null distribution of the LR will depend on the structure of the cone $\mathrm{C} = \mathrm{H_o}\ \mathrm{U}\ \mathrm{H_1}$, whether obtuse or acute with respect to norm induced by $\sum^{-1}$.[52] In the first case, the LR statistic follows a $\chi 2$ *bar* distribution under $\mathrm{H_0}$ (Menendez, Rueda, and Salvador 1992b).[53] In the second case, the distribution of the LR statistic is intractable, but the test is dominated in power by a *reduced* test comprised of testing just the *different parts* of the hypotheses $\mathrm{H_o}$ and $\mathrm{H_1}$ (Menéndez and Salvador 1991; Menéndez, Rueda, and Salvador 1992a). It can be shown that the structure of $\sum$ adopted in this paper makes the cone C acute, so that the second case is the relevant one.[54] The reduced dominating test takes the form below.

$$\mathrm{H_o}: PD_1 = PD_2$$
$$\mathrm{H_1}: PD_1 < PD_2.$$

The test above is just a particular case of the general rating discriminatory power test with I = 2. Accordingly, its rejection rule is given as follows.

[52]See Martín and Salvador (1988) and Menéndez, Rueda, and Salvador (1992b) for the definitions of those cone types. The norm induced by $\sum^{-1}$ is defined as $\|x\|_{\Sigma^{-1}} = x^T \mathbf{\Sigma}^{-1} x$.

[53]Although $\chi 2$ *bar* distributions are common in the theory of order-restricted inference (Robertson, Wright, and Dykstra 1988), application of the focal test in this circumstance is not very practical, as the determination of both the LRT statistic and the p-values are computationally intensive.

[54]This is true because $\mathbf{a_i'\Sigma a_j} \leq 0$, $\mathrm{i} \neq \mathrm{j}$, where the $\mathbf{a_i}$'s ($\mathbf{a}_\mathrm{i} = (0, \ldots, -1, 1, \ldots, 0)$') generate the linear restrictions defining the cone C. More specifically, it is true that $\mathbf{a_i'\Sigma a_j} = (\rho_\mathrm{B} - \rho_\mathrm{W})/(1 - \rho_\mathrm{W})$ if $|\mathrm{i} - \mathrm{j}| = 1$ or 0 if $|\mathrm{i} - \mathrm{j}| \geq 2$. See the mentioned references for further details. Whether more general but still realistic variance structures $\mathbf{\Sigma}$ might lead to a different conclusion is an interesting question not addressed in this paper.

Reject H_0 (i.e., validate the CRM) if

$$\overline{DR}_2 - \overline{DR}_1 > z_\alpha(2(\rho_W - \rho_B)/(Y(1 - \rho_W)))^{1/2}. \tag{14}$$

The dominance of the focal test by a reduced test is a surprising result and was long considered an anomaly of the LR principle (e.g., Warrack and Robertson 1984). In the context of CRMs, this means that in order to judge the discriminatory performance of a particular pair of consecutive ratings, the bank or the supervisor would be in a better position if it simply disregards the prior knowledge of the performance of the other ratings. But how can less information be better? Only most recently Perlman and Wu (1999) showed that indeed the overall picture was not so much in favor of the "dominating" test, arguing that the latter presents controversial properties. For example, it rejects **PDs** *closer* to H_0 than to H_1.[55] Nevertheless, the practitioner does not have another choice besides using the power dominating test, because, as just observed, the null distribution of the LRT statistic for the focal test is unknown. Keeping that in mind, the analysis of this section provides the theoretical foundation to an easy-to-implement procedure that focuses solely on the supposedly problematic pair of ratings. More interestingly, however, a generalization of the results discussed in this section suggests a uniform procedure to check rating discriminatory power: select the ratings whose discriminatory capacity are at stake and apply the general test to them.

6. Finite-Sample Properties

All the tests discussed in this paper are based on the asymptotic distribution of DRAPM, which assumes an infinite number of borrowers for each rating. This section analyzes the implications to the performance of the one-sided calibration test of a finite but still large number of borrowers (N = 100 is chosen as the base case).[56] Due

[55]Perlman and Wu (1999) conclude once again that UMP size-α tests should not be pursued at any cost.

[56]The analysis is restricted to the one-sided calibration test not only because it is the main focus of this paper but also because the finite sample properties of discriminatory tests are more complex to analyze when distributions of default rate *differences* are involved. Also, as perceived later in the section, the issues of most concern related to the finite-sample properties of the two-sided calibration test derive from the analysis of the one-sided case.

to the strong reliance of the test on the asymptotic normality of the marginal distributions of DRAPM, it is important to verify how the real marginals compare to the asymptotic ones. The focus on a particular marginal allows then, for the sake of clarity, to direct the attention initially to the case I = 1. This section conducts Monte Carlo simulations of DRAPM, at the stage in which idiosyncratic risk is not yet diversified away, for N = 100 and Y = 5, unless stated otherwise. Based on a large set of simulated average annual default rates and for I = 1, the effective significance level is computed as a function of the nominal significance level α, for varying scenarios of the parameters true PD and ρ_W. In general, 200,000 simulations are run for each scenario.

$$\text{Effective confidence level} = \hat{P}rob\left(\frac{\sqrt{1-\rho_W}\,\overline{DR}_n - PD}{\sqrt{\rho_W/Y}} < -z_\alpha\right), \tag{15}$$

where the probability is estimated by the empirical frequency of the event and $\overline{DR}_n$ denotes a particular simulation result.

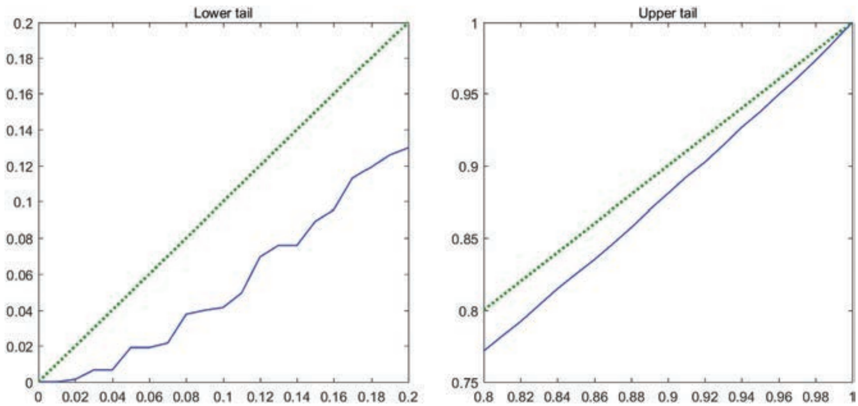
The effective level measures the real size of the asymptotic size-α one-sided test. Alternatively, since it is expressed in the form of a probability of rejection, the effective level can also be seen as the real power at the postulated PD, when the asymptotic power is equal to α, of an asymptotic size-δ one-sided test, with $\delta < \alpha$.[57] From both interpretations, the occurrence of effective levels lower than nominal levels means that the test is more conservative, with a smaller probability of validation in general than what is suggested by the analysis of section 4 based on DRAPM. Effective levels higher than nominal levels indicates the opposite: a finite-sample liberal *bias*.

A first general important finding derived from the performed simulations for the case I = 1 is that the convergence of the lower tails of the average (transformed) default rate distributions to their normal asymptotic limits is slower and less smooth than in the case of the upper tails, for realistic PD values. The situation is illustrated by the pair of graphs (figure 1) calculated based on the scenario PD = 3%, $\rho_W = 0.20$, N = 100, and Y = 5. The solid line represents

[57] More specifically, it is easy to see that $\delta = \Phi(-z_\alpha - (u - PD)/(\rho_W/Y)^{1/2})$.

Figure 1. Lower and Upper Tails

PD = 3%, $\rho_W = 0.20$, N = 100, Y = 5



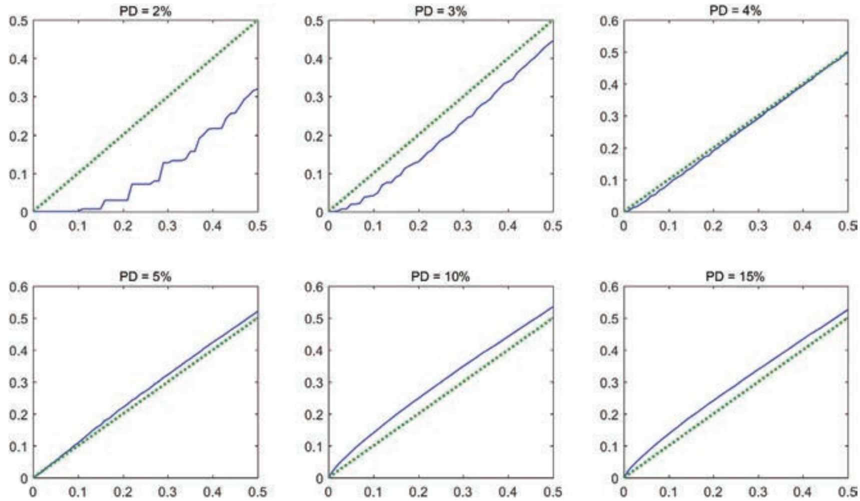
Notes: Solid line: Effective confidence level against the nominal size α of the asymptotic one-sided test H_0 : PD $\geq$ u against H_1 : PD < u. Dotted straight line is the identity function to ease comparison. PD = true probability of default, ρ_W = asset correlation, N = number of borrowers, Y = number of years.

the effective confidence level for each nominal level depicted at the x-axes, while the dotted straight line is the identity function merely denoting the nominal level to facilitate comparison. Note that the effective level is much farther from the nominal value in the lower tail of the distribution (depicted in the right-hand graph) than in the upper tail (depicted in the left-hand graph). In particular, if the one-sided calibration test is employed at the nominal level of 10 percent, the test will be much more conservative in reality, as the effective size will be approximately only 4 percent.

Indeed, the fact that the lower tail is less well behaved is strongly relevant to this paper's one-sided calibration test. Under the approach of placing the undesired conclusion in H_0 (e.g., PD $\geq$ u), rejection of the null, or equivalently validation, is obtained if average default rates are small, so that the one-sided test is based in fact on the lower tail of the distribution. On the contrary, the upper tail would be the relevant part of the distribution had the approach of placing the "CRM correctly specified" hypothesis in H_0 been adopted, as in the rest of the literature. Since convergence of the upper tail is better behaved, the finite-sample departure from

Figure 2. Effect of PD

$\rho_W = 0.20$, N = 100, Y = 5



Notes: Solid line: Effective confidence level against the nominal size α of the asymptotic one-sided test H_0 : PD $\geq$ u against H_1 : PD $<$ u. Dotted straight line is the identity function to ease comparison. PD = true probability of default, ρ_W = asset correlation, N = number of borrowers, Y = number of years.

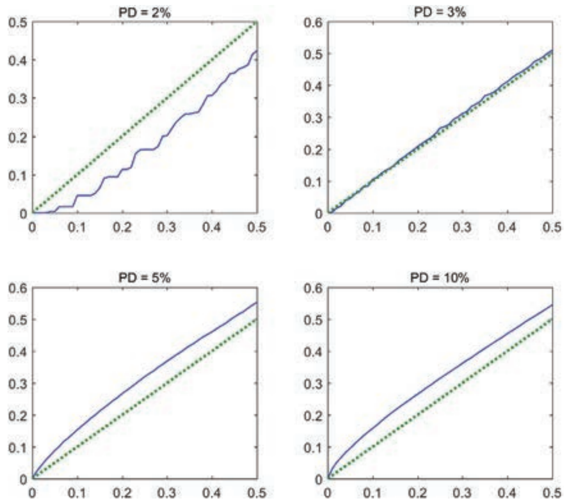
the normal limit would be smaller in this case. In the view of this paper, this would be, however, a misleading property of the latter approach because the worse relative behavior of the lower tail would not be revealed.

The main numerical findings regarding the finite sample power performance of the one-sided calibration test are described in the sequence, based on the analysis of the simulated lower tails.[58] The investigation starts with the effect of the true PD, when I = 1, on the effective confidence level. Figures 2 and 3 reveal that, in the region of 0% < PD < 10% and $0.15 < \rho_W < 0.20$, as PD increases, the test evolves from having a conservative bias (true power smaller than the

[58]Miu and Ozdemir (2008) also investigate finite-sample properties of similar validation tests, but their results are not comparable to those of this paper since they adopt H_0 : CRM correctly specified, they assume serial dependency of default rates, and they investigate different ranges of parameter values for PD, ρ_W, and Y.

Figure 3. Effect of PD

$\rho_W = 0.15$, N = 100, Y = 5



Notes: Solid line: Effective confidence level against the nominal size α of the asymptotic one-sided test H_0 : PD $\geq$ u against H_1 : PD $<$ u. Dotted straight line is the identity function to ease comparison. PD = true probability of default, ρ_W = asset correlation, N = number of borrowers, Y = number of years.

asymptotic one) to having a liberal bias (true power larger than the asymptotic one). At PD = 4% for $\rho_W = 0.20$ or at PD = 3% for $\rho_W = 0.15$, the finite sample bias is approximately null as the test matches its theoretical limiting values. On the other hand, in the region of 10% < PD < 15% and $0.15 < \rho_W < 0.20$, as PD increases, the solid line comes back a bit closer to the dotted straight one, i.e., the test diminishes its liberal bias (but not sufficiently so as to turn conservative).

As the asymptotic one-sided test based on DRAPM already suffers from problems of lack of power, this section suggests, as a possible general recommendation, consideration of the real (unmodified) applications of the test solely in the cases where the finite-sample analysis indicates a non-conservative bias. Indeed, if instead an additional layer of conservatism is added to the already conservative asymptotic test, the resulting procedure may hardly validate at all. The restriction to the finite-sample liberal cases, when I = 1, suggests against, according to figures 2 and 3, attempts of validation of low PDs (e.g., PD $\leq 3\%$).

For the case $I > 1$, it is easy to see that the effective level that measures the real size of the asymptotic size-α one-sided test is given by the maximum of the effective levels computed for each different rating assuming $I = 1$.[59] Therefore, the effective confidence level takes graphically, for each nominal level α, the form of the maximum of the solid lines corresponding to the different rating PDs that constitute the CRM. As a result, the effective level may not be reduced in the presence of low PD ratings that introduce conservative bias on a marginal basis. Nevertheless, CRMs with low PD ratings will still be particularly hard to validate due to the lower true power derived from the corresponding real marginals. To better understand the multivariate case, this section computes the true effective power at the postulated $\mathrm{PD_i}$s of the CRMs of table 2. Effective power is estimated making use of the simulated average annual default rates, in a similar fashion to equation (15) but with $-z_\alpha + (u - PD)/(\rho_W/Y)^{1/2}$ replacing $-z_\alpha$ and extending the formula to the multivariate case.[60]

The differences true powers at the postulated $\mathrm{PD_i}$s minus the asymptotic ones (shown in tables 3 and 4) are presented in tables 9 and 10.[61] Results indicate that the fall in power in the finite-sample setting is considerably more material in the CRM design where $\mathrm{PD_i}$s follow a geometric progression (unless the favorable scenario of low ρ_W and high Y is considered). This result is consistent with the finding of section 4 that the lowest $\mathrm{PD_i}$s drive the power of the multivariate one-sided test in this CRM design. Because the lowest PDs are exactly the ones that suffer most from a conservative finite-sample bias when $I = 1$, as revealed in figures 2 and 3, this bias extends comprehensibly to the multivariate case. Consequently, when failing

[59]For $I > 1$, the interpretation in terms of real power at the postulated PDs is also possible but a less direct one. For $I > 1$, the effective confidence level can be seen as the maximum real power among I asymptotic size-δ_i one-sided tests, when the asymptotic power is equal to α, with $\delta_i < \alpha$ and the power being computed for test i at the postulated PD of rating i (while other ratings have PDs = 0). δ_i has a similar expression to the case $I = 1$.

[60]Compare with equation (6), which determines the asymptotic power.

[61]The same (ρ_W, Y) scenarios of tables 3 and 4 are considered here. The in-between scenario is specified as $\rho_W = 0.1575$ and $Y = 7$ for the computation of the effective power. The assumption about ρ_B/ρ_W is also kept here.

Table 9. Difference between Effective and Asymptotic Power for Several CRM Designs and u_i Choices, N = 100, I = 3

$\rho_B/\rho_W = 0.8, \alpha = 0.15$

	PD$_i$s Follow Arithmetic Progression		**PD$_i$s Follow Geometric Progression**	
	$u_i = PD_{i+1}$	**$u_i = (PD_{i+1} + PD_i)/2$**	**$u_i = PD_{i+1}$**	**$u_i = (PD_{i+1} + PD_i)/2$**
$\rho_W = 0.12$, Y = 10	–0.01	–0.01	0.00	–0.03
In-between	–0.01	0.00	–0.05	–0.09
$\rho_W = 0.18$, Y = 5	–0.01	–0.01	–0.11	–0.13

Notes: Effective and asymptotic power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, computed at the postulated PDs of table 2 (case I = 3). ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, in-between scenario characterized by $(\rho_W/Y) = 0.15^2$, Y = number of years, α = asymptotic size of test. Effective power computed based on 200,000 Monte Carlo simulations of DRAPM.

Table 10. Difference between Effective and Asymptotic Power for Several CRM Designs and u_i Choices, N = 100, I = 4

$\rho_B/\rho_W = 0.8, \alpha = 0.15$

	PD$_i$s Follow Arithmetic Progression		**PD$_i$s Follow Geometric Progression**	
	$u_i = PD_{i+1}$	**$u_i = (PD_{i+1} + PD_i)/2$**	**$u_i = PD_{i+1}$**	**$u_i = (PD_{i+1} + PD_i)/2$**
$\rho_W = 0.12$, Y = 10	–0.02	–0.01	–0.03	–0.03
In-between	–0.02	–0.01	–0.08	–0.06
$\rho_W = 0.18$, Y = 5	–0.01	–0.01	–0.11	–0.08

Notes: Effective and asymptotic power of the test $H_0 : PD_i \geq u_i$ for some $i = 1 \ldots I$ against $H_1 : PD_i < u_i$ for every $i = 1 \ldots I$, computed at the postulated PDs of table 2 (case I = 4). ρ_W = within-rating asset correlation, ρ_B = between-rating asset correlation, in-between scenario characterized by $(\rho_W/Y) = 0.15^2$, Y = number of years, α = asymptotic size of test. Effective power computed based on 200,000 Monte Carlo simulations of DRAPM.

to validate CRMs with increasing PD differences because of large default rates of some low postulated PD_is, a possible practical advice is to apply the test only to the remainder of the postulated **PD** vector (e.g., ratings 3 to 7 in the example related to table 1). Alternatively, if suspicion of PD undercalibration is particularly placed on the low postulated PD_is, a higher nominal level α could be applied just to them.

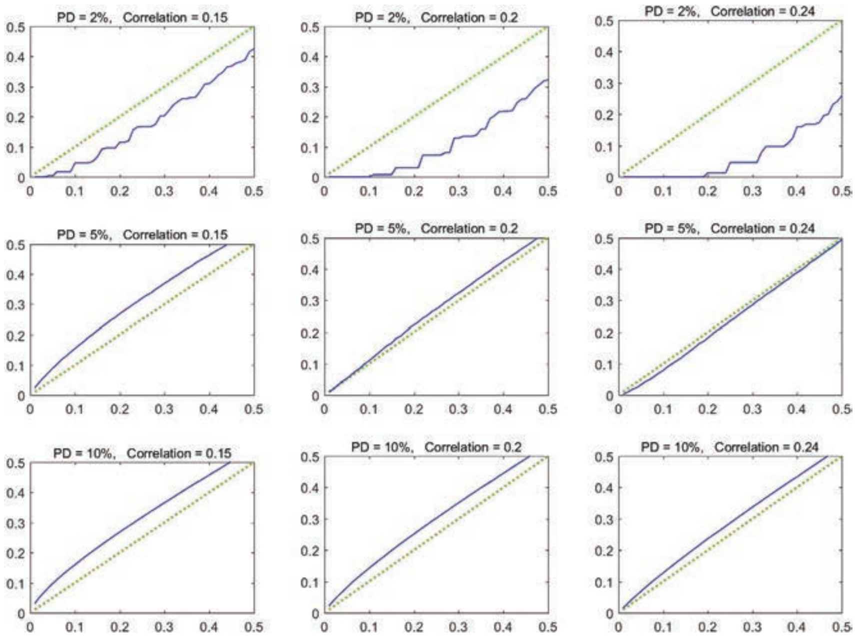
In the CRM design where PD_is follow an arithmetic progression, the highest PD_is drive the power of the test (see section 4), but they suffer instead from a small liberal finite-sample bias. Consequently, with equally spaced PD_is, the effective power of the multivariate test is not greatly affected, as noticed in tables 9 and 10. This is particularly useful since the asymptotic power in this CRM design is already low compared with CRMs with increasing PD differences (as discussed in section 4). Tables 9 and 10 finally reveal that the choice of the indifference region is generally of secondary importance to the magnitude of the finite-sample bias of the test in comparison with the choice of the rating PD_is, unless the favorable (ρ_W,Y) scenario is considered.

The influence of the within-rating asset correlation ρ_W under the base case of $N = 100$ is analyzed in figure 4. Considering initially the case $I = 1$, one notes that when ρ_W increases, the test evolves towards a more conservative bias (or towards a smaller liberal one), for every PD. Note that this represents a second channel, now through the finite-sample properties, by which ρ_W diminishes the power of the test. For the case $I > 1$, note first that the move towards a conservative bias is greater the lower the PD in figure 4. Because the lowest rating PD_is tend to drive the power of the test when CRMs possess increasing PD differences, it is expected that the finite-sample power-reducing effect of ρ_W will be stronger in precisely that type of CRM design. This is indeed the result found in tables 9 and 10 when one moves in the direction from $\rho_W = 0.12$ to $\rho_W = 0.18$ and PD_is follow a geometric progression.[62] Consequently,

[62]More precisely, in tables 9 and 10, Y also changes across the parameter scenarios. However, holding Y constant and just increasing ρ_W produces the same sort of result.

Figure 4. Effect of ρ_W

PD = 2%, 5%, or 10% depending on the row, Y = 5, N = 100



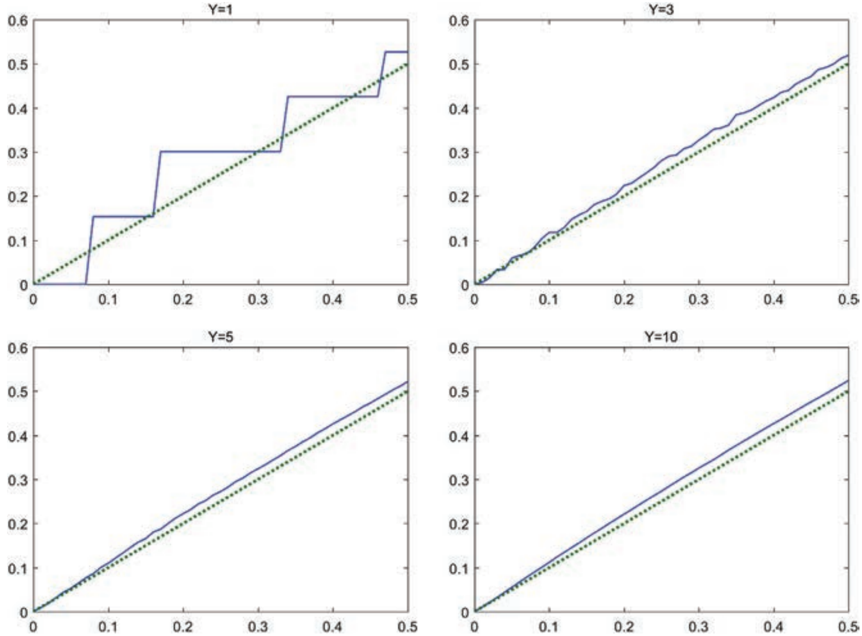
Notes: Solid line: Effective confidence level against the nominal size α of the asymptotic one-sided test H_0 : PD $\geq$ u against H_1 : PD $<$ u. Dotted straight line is the identity function to ease comparison. PD = true probability of default, ρ_W = asset correlation, N = number of borrowers, Y = number of years.

when a high value of ρ_W is an important consideration in the validation of CRMs with increasing PD differences, it may be advised to investigate separately the appropriateness of the calibration of postulated PD_is, in a similar fashion to what was suggested previously in this the section.

The influence of the number of years Y under the base case of N = 100 is analyzed in figure 5, considering again initially the case I = 1. The effect of an increase in the number of years, in the region of one to ten years, is to smooth considerably the distribution lower tail. Results not shown indicate that as N increases beyond 100,

Figure 5. Effect of Y

PD = 5%, $\rho_W = 0.20$, N = 100



Notes: Solid line: Effective confidence level against the nominal size α of the asymptotic one-sided test H_0 : PD $\geq$ u against H_1 : PD $<$ u. Dotted straight line is the identity function to ease comparison. PD = true probability of default, ρ_W = asset correlation, N = number of borrowers, Y = number of years.

the solid and dotted lines come closer at every figure, as expected. Other produced results also indicate that, for the same Y, the lower-tail discontinuity is greater the lower the PDs. Consequently, based on the same reasoning underlying the previous analysis on the effect of ρ_W, the lower-tail discontinuity propagates more strongly to the multivariate case again when CRMs possess increasing PD differences.

Finally, it is important to observe that even if the one-sided test could be totally based on the simulated distributions of this section, there would still be some extreme cases where validation is virtually impossible at traditional low confidence levels. When I = 1 and Y = 1 (cf. figure 5) or true PD = 1%, for example, the lower tail of the distribution is quite discrete and presents significant probability of zero defaults. As a result, the effective confidence level jumps several times and assumes only a small finite number of values in the lower tail. When Y = 1 (and I = 1), the first non-zero effective level is already approximately 15 percent; after that, the next value is approximately 30 percent. Therefore, validation at the 5 percent or 10 percent significance level is not possible. Hence, the Basel II prescription of a minimum of five years of data is important not only to increase the asymptotic power of the test, according to section 4, but also to remove the quite problematic finite-sample behavior of the lower tail.

7. Conclusion

This study contributes to the CRM validation literature by introducing new ways to statistically address the validation of credit rating PDs. Firstly, it proposes new formulations for H_0 and H_1 in order to control the error of accepting an incorrect CRM. Secondly, it provides an integrated analytical treatment of all ratings at once, in a way that recognizes the effect of default correlation. Finally, it provides a unified framework for testing calibration and rating discriminatory power. All these aspects are interlinked with the development of a probabilistic asymptotic normal model for the average default rate vector that recognizes default correlation. Important empirical and practical consequences stem from these proposals, as outlined in the following paragraphs.

On calibration testing, the relative roles played by the distinct elements that affect the power are unveiled for the one-sided version. The feature of increasing PD differences between consecutive ratings, found in many real-world CRMs, and, particularly, the choice of liberal indifference regions are shown to be important to the achievement of reasonable levels of power. On the other hand, the correlation between the ratings, whose calibration is not present in Basel II, possesses only a minor effect on power. Also, appropriately restricting the set of PDs to be tested may do a job almost as good as the original test in terms of power, which may offer support, in many practical circumstances, to reduce joint validation of credit rating PDs to individual validation of a few rating PDs. Another important general message of the analysis is that the power of the one-sided calibration test is unavoidably and substantially low in some cases. Regarding this issue, strategies of power improvement are discussed, suggesting limited efficacy or inappropriateness. Additionally, the paper discusses the conceptual problems of applying modern ideas in multivariate equivalence to two-sided calibration testing.

As far as discrimination is concerned, a new goal of rating discriminatory power is established for CRMs. In contrast to traditional measures of discrimination, the new aimed property is less stringent than the requirement of perfect calibration and is not dependent on default correlation. Results of uniform power dominance provide a theoretical foundation for restricting the investigation of the desired property just to the pairs of consecutive ratings whose discriminatory capacity are at stake and, therefore, lead to an easy-to-implement procedure.

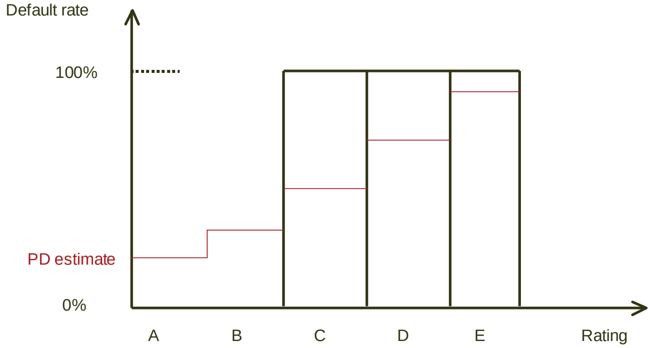
Understanding the implications of DRAPM to validation also includes an analysis of its properties when dealing with a finite-sample of borrowers. As a matter of fact, DRAPM has the disadvantage of being an asymptotic model whose finite-sample properties may introduce a significant additional layer of test conservatism besides the asymptotic one. Monte Carlo simulations show that this will likely be the case for small PDs (e.g., PD $\leq 3\%$) or a small number of years (e.g., $\mathrm{Y} < 5$) in the one-sided calibration test. A possible recommendation in the former case may be to apply the test just to the remaining ratings or to investigate the low postulated PDs at a higher nominal confidence level. On the other hand, when a liberal finite-sample bias is present, it may counterbalance the nominal conservatism, although some caution should always be exercised in the analysis. A general more robust procedure, however, would ideally try to incorporate the remaining non-systemic part of the credit risk into the validation process. Future research is warranted on this aspect.

Above all, the bank or the regulator should not demand much from statistical testing of CRMs. Even under the simplifying assumptions of DRAPM, the power of the tests of this paper, as well as other tests discussed in the literature, is negatively affected by the unavoidable presence of default correlation and by the small length of default rate time series available in banks' databases. Possibly due to this reason, BCBS (2005b) perceives validation as comprising not only quantitative but also somewhat qualitative tools. It is likely, for example, that the investigation of the continuous internal use of **PDs**/ratings by the bank may uncover further evidence, although subjective, supporting or not supporting the CRM validation. Nonetheless, this paper supports the view that the possibility of reliance on qualitative aspects opened by the Basel Committee should not dampen the incentives to extract as much quantitative feedback as possible from statistical testing, including a quantitative sense of its limitations.

Appendix 1

Figure 6 should be interpreted as a result over the long run and displays a rating model with perfect discrimination but not perfect calibration. The bars' heights represent the magnitude of the ex post default rate for each rating. All borrowers classified as C to E defaulted, whereas all borrowers classified as A to B survived. If this is the regular behavior of this CRM, knowing beforehand the rating of the obligor allows one to predict default or no default with certainty (perfect discriminatory power). The thin stepped line indicates the ex ante PD estimate for each rating. Ratings A and B had a 0 percent default rate, thus lower than the ex ante prediction. Ratings C to E had a 100 percent default rate, thus higher than the ex ante prediction. The CRM is therefore not correctly calibrated. Obviously, this example represents an extreme case (because realistic CRMs do not have perfect discriminatory power), but it is useful to illustrate that although both characteristics are desirable, they may well be inconsistent as they are aimed at their best.

Figure 6. Perfect Discrimination but Imperfect Calibration



Notes: The bold bar height of each rating represents the ex post long-run default rate of that rating, whereas the thin stepped line represents the ex ante PD estimates of the ratings.

Appendix 2

Proof of Proposition

The first parcel of the AUROC definition can be expressed as follows.

$$\begin{aligned}
&\mathrm{Prob}(\mathrm{PD_n} > \mathrm{PD_m} | \text{n defaults and m doesn't}) \\
&= \frac{\mathrm{Prob}(\text{n defaults and m doesn't}, \mathrm{PD_n} > \mathrm{PD_m})}{\mathrm{Prob}(\text{n defaults and m doesn't})} \\
&= \frac{\sum_{i,j=1}^{I} \mathrm{Prob}(\text{n defaults and m doesn't}, \mathrm{PD_n} > \mathrm{PD_m} | \text{n has rating i and m has rating j}) \mathrm{P(i,j)}}{\sum_{i,j=1}^{I} \mathrm{Prob}(\text{n defaults and m doesn't} | \text{n has rating i and m has rating j}) \mathrm{P(i,j)}} \\
&= \frac{\sum_{i,j=1,\ i>j}^{I} \mathrm{Prob}(\text{n defaults and m doesn't} | \text{n has rating i and m has rating j}) \mathrm{P(i,j)}}{\sum_{i,j=1}^{I} \mathrm{Prob}(\text{n defaults and m doesn't} | \text{n has rating i and m has rating j}) \mathrm{P(i,j)}} \\
&= \frac{\sum_{i,j=1,\ i>j}^{I} \Phi_2\left(\Phi^{-1}(\mathrm{PD_i}), -\Phi^{-1}(\mathrm{PD_j}), -\rho_{ij}\right) \mathrm{P(i,j)}}{\sum_{i,j=1}^{I} \Phi_2\left(\Phi^{-1}(\mathrm{PD_i}), -\Phi^{-1}(\mathrm{PD_j}), -\rho_{ij}\right) \mathrm{P(i,j)}}.
\end{aligned}$$

where the last equality derives from the expression for a joint probability of default and non-default implicit in a DRAPM-style model (e.g., Gordy 2000). Similarly, the second parcel of the AUROC definition can be expressed as

$$1/2\ \mathrm{Prob}(\mathrm{PD_n} = \mathrm{PD_m} | \mathrm{n\ defaults\ and\ m\ doesn't})$$
$$= \frac{\sum_{i=1}^{I} \Phi_2\left(\Phi^{-1}(\mathrm{PD_i}), -\Phi^{-1}(\mathrm{PD_i}), -\rho_{ii}\right)\mathrm{P(i)}}{2\sum_{i,j=1}^{I} \Phi_2\left(\Phi^{-1}(\mathrm{PD_i}), -\Phi^{-1}(\mathrm{PD_j}), -\rho_{\mathrm{ij}}\right)\mathrm{P(i,j)}}.$$

and the proposition is proved, observing the convention $PD_i \equiv \Phi^{-1}(\mathrm{PD_i})$.

References

Balthazar, L. 2004. "PD Estimates for Basel II." *Risk* (April): 84–85.

Bamber, D. 1975. "The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Operating Characteristic Graph." *Journal of Mathematical Psychology* 12 (4): 387–415.

Basel Committee on Banking Supervision. 2005a. "An Explanatory Note on the Basel II IRB Risk Weight Functions." Bank for International Settlements.

———. 2005b. "Studies on the Validation of Internal Rating Systems." Bank for International Settlements.

———. 2006a. "Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework — Comprehensive Version." Bank for International Settlements.

———. 2006b. "The IRB Use Test: Background and Implementation." Bank for International Settlements.

———. 2011. "Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems — Revised Version." Bank for International Settlements.

Berger, R. L. 1989. "Uniformly More Powerful Tests for Hypotheses Concerning Linear Inequalities and Normal Means." *Journal of the American Statistical Association* 84 (405): 192–99.

Berger, R. L., and J. C. Hsu. 1996. "Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets." *Statistical Science* 11 (4): 283–319.

Black, F., and M. Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81 (3): 637–54.

Blochlinger, A. 2012. "Validation of Default Probabilities." *Journal of Financial and Quantitative Analysis* 47 (5): 1089–1123.

Blochlinger, A., and M. Leippold. 2006. "Economic Benefit of Powerful Credit Scoring." *Journal of Banking and Finance* 30 (3): 851–73.

Blochwitz, S., S. Hohl, D. Tasche, and C. When. 2004. "Validating Default Probabilities on Short Time Series." Working Paper.

Bluemke, O. 2013. "Probability of Default Validation: Introducing the Likelihood-Ratio Test and Power Considerations." *Journal of Risk Model Validation* 7 (2): 29–59.

Brown, L. D., G. Casella, and G. Hwang. 1995. "Optimal Confidence Sets, Bioequivalence and the Limaçon of Pascal." *Journal of the American Statistical Association* 90 (431): 880–89.

Brown, L. D., G. Hwang, and A. Munk. 1998. "An Unbiased Test for the Bioequivalence Problem." *Annals of Statistics* 25 (6): 2345–67.

Demey, P., J. F. Jouanin, C. Roget, and T. Roncalli. 2004. "Maximum Likelihood Estimate of Default Correlations." *Risk* (November): 104–8.

Engelmann, B., E. Hayden, and D. Tasche. 2003. "Testing Rating Accuracy." *Risk* (January).

Gordy, M. B. 2000. "A Comparative Anatomy of Credit Risk Models." *Journal of Banking and Finance* 24 (1–2): 119–49.

———. 2003. "A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules." *Journal of Financial Intermediation* 12 (3): 199–232.

Gourieroux, C., and A. Monfort. 1995. *Statistics and Econometric Models*, Vol. 2 (Themes in Modern Econometrics). Cambridge University Press.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple Comparison Procedures.* John Wiley & Sons, Inc.

Laska, E. M., and M. J. Meisner. 1989. "Testing Whether an Identified Treatment Is Best." *Biometrics* 45 (4): 1139–51.

Leland, H. E. 2004. "Predictions of Default Probabilities in Structural Models of Debt." *Journal of Investment Management* 2 (2): 5–20.

Liu, H., and R. L. Berger. 1995. "Uniformly More Powerful, One-Sided Tests for Hypotheses about Linear Inequalities." *Annals of Statistics* 23 (1): 55–72.

Martín, M., and B. Salvador. 1988. "Validity of the 'Pool-Adjacent-Violator' Algorithm." *Statistics and Probability Letters* 6 (3): 143–45.

McDermott, M. P., and Y. Wang. 2002. "Construction of Uniformly More Powerful Tests for Hypotheses about Linear Inequalities." *Journal of Statistical Planning and Inference* 107 (1–2): 207–17.

Menéndez, J. A., C. Rueda, and B. Salvador. 1992a. "Dominance of Likelihood Ratio Tests under Cone Constraints." *Annals of Statistics* 20 (4): 2087–99.

———. 1992b. "Testing Non-Oblique Hypotheses." *Communications in Statistics — Theory and Method* 21 (2): 471–84.

Menéndez, J. A., and B. Salvador. 1991. "Anomalies of the Likelihood Ratio Test for Testing Restricted Hypotheses." *Annals of Statistics* 19 (2): 889–98.

Merton, R. C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–70.

Miu, P., and B. Ozdemir. 2008. "Estimating and Validating Long-Run Probability of Default with Respect to Basel II Requirements." *Journal of Risk Model Validation* 2 (2): 3–41.

Munk, A., and R. Pfluger. 1999. "1-α Equivariant Confidence Rules for Convex Alternatives are $\alpha/2$-Level Tests — With Applications to the Multivariate Assessment of Bioequivalence." *Journal of the American Statistical Association* 94 (448): 1311–19.

Perlman, M. D., and L. Wu. 1999. "The Emperor's New Tests." *Statistical Science* 14 (4): 355–69.

Pluto, K., and D. Tasche. 2005. "Thinking Positively." *Risk* (August): 72–78.

Robertson, T., F. T. Wright, and R. L. Dykstra. 1988. *Order Restricted Statistical Inference.* John Wiley & Sons.

Sasabuchi, S. 1980. "A Test of a Multivariate Normal Mean with Composite Hypotheses Determined by Linear Inequalities." *Biometrika* 67 (2): 429–39.

Shapiro, A. 1988. "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis." *International Statistical Review* 56 (1): 49–62.

Vasicek, O. 2002. "Loan Portfolio Value." *Risk* (December).

Wang, W., J. T. G. Hwang, and A. Dasgupta. 1999. "Statistical Tests for Multivariate Bioequivalence." *Biometrika* 86 (2): 395–402.

Warrack, G., and T. Robertson. 1984. "A Likelihood Ratio Test Regarding Two Nested but Oblique Order-Restricted Hypotheses." *Journal of the American Statistical Association* 79 (388): 881–86.

Watt, M. 2013. "Mending the RWA Machine." *Risk* 26 (1, January): 17–20.

Currency Wars, Coordination, and Capital Controls*

Olivier Blanchard
Petersen Institute for International Economics

The strong monetary policy actions undertaken by advanced economies' central banks have led to complaints of "currency wars" by some emerging market economies, and to widespread demands for more macroeconomic policy coordination. This paper revisits these issues. It concludes that, while advanced economies' monetary policies indeed have had substantial spillover effects on emerging market economies, there was and still is little room for coordination. It then argues that restrictions on capital flows were and are a more natural instrument for advancing the objectives of both macro and financial stability.

JEL Codes: F3, F36, F42.

1. Introduction

In September 2010, Guido Mantega, then minister of finance of Brazil, declared, "We are in the midst of an international currency war, a general weakening of currency. This threatens us because it takes away our competitiveness" (Wheatley and Garnham 2010). His complaint was relayed and amplified by others, notably by Raghuram Rajan, governor of the Central Bank of India. In April 2014, for example, Rajan said, "The disregard for spillovers could

*Prepared for the Asian Monetary Policy Forum, Singapore, May 2016. Thanks to Vivek Arora, Mathieu Bussiere, Tam Bayoumi, Ricardo Caballero, Steve Cecchetti, Bill Cline, Xavier Gabaix, Taka Ito, Olivier Jeanne, Sebnem Kalemli-Ozcan, Gian Maria Milesi Ferretti, David Vines, and two referees for discussions and comments, and to Julien Acalin for outstanding research assistance.
Author e-mail: oblanchard@piie.com.

put the global economy on a dangerous path of unconventional monetary tit for tat. To ensure stable and sustainable economic growth, world leaders must re-examine the international rules of the monetary game, with advanced and emerging economies alike adopting more mutually beneficial monetary policies."

Complaints by emerging market economies about advanced economies' monetary policies, together with calls for coordination, have been a staple of the last seven years. The purpose of this paper is to examine the validity of these complaints and the scope for coordination. It reaches two main conclusions: (i) The scope for coordination was and is limited. (ii) Restrictions on capital flows were and are the more natural instrument to achieve a better outcome.

The paper is organized as follows. Section 2 briefly reviews the cross-border effects of advanced economies' monetary policies on emerging economies, through goods markets, foreign exchange markets, and financial markets. Section 3 examines the scope for coordination and concludes that it was and still is rather limited. It argues that, given the limits on fiscal policy, restrictions on capital flows were and still are the appropriate macroeconomic instrument to achieve better outcomes, both in advanced economies and in emerging economies. Section 4 returns to the effects of capital flows on the financial systems in emerging economies, and argues for a second role for restrictions on capital flows, not only as a macroeconomic tool but also as a financial stability tool.

2. Cross-Border Effects

Expansionary monetary policy in advanced economies (AEs in what follows), conventional or unconventional, has affected emerging market economies (EMs in what follows) through three channels: increased exports, exchange rate appreciation, and the effects of capital flows on their financial system. The first two are fairly well understood, the third much less.[1]

[1]For a set of studies of the various cross-border effects, see the "Selected Issues" part of the 2011 International Monetary Fund United States Spillover Report.

2.1 Expansionary AE Monetary Policy Leads to a Higher Demand for EM Exports

This channel is straightforward: Lower interest rates lead to higher AE output, thus to higher AE imports, including higher imports from EMs.

It is useful for later to get a sense of potential magnitudes: For most EMs, exports to AEs represent between 5 percent and 10 percent of their GDP.[2] For example, Chinese exports to the AEs are equal to 10 percent of Chinese GDP; Brazilian and Indian exports are equal to 5 percent of their respective GDP.[3] Using these numbers suggests small effects of higher output in AEs: A 1 percent increase in AE output leads to an increase of 0.1 percent in Chinese output and less than half that in the other two countries.

The relevant numbers are, however, higher. First, for any EM, higher AE output leads not only to a direct increase in exports to AEs but also to an indirect effect through higher induced output in other EM countries. Second, the elasticity of AE imports to GDP is higher than unity, reflecting the share of investment in imports and the higher cyclicality of investment. Recent estimates suggest an elasticity between 1.5 and 2.0.[4] Third, multipliers are likely to increase the effect of exports on output. Overall, this suggests that an increase in U.S. output of 1 percent may lead, through higher imports (at a given exchange rate), to an increase in output in China around 0.2 percent and to a smaller number for most other emerging markets.

The other number we need is the semi-elasticity of AE demand to the real interest rate. Here again, uncertainty is substantial, but a typical number is that a sustained 100 bp decrease in the real policy rate leads, over time, to an increase in aggregate demand of 1 percent of GDP (this is roughly the semi-elasticity implicit in the FRB/US model used by the Federal Reserve).

Putting things together, and with the usual caveats, this suggests that a 1 percent sustained decrease in the AE real policy rate—or

[2]Data are from http://wits.worldbank.org/.

[3]Given the relevance of supply chains, and the fact that higher exports mechanically imply higher imports, the numbers somewhat overstate the relevant numbers.

[4]See, for example, Boz, Bussiere, and Marsilli (2015).

the equivalent of a 1 percent decrease in the policy rate in the case quantitative easing (QE) is used to decrease long rates instead—leads to an increase in EM GDP ranging from 0.1 percent to 0.2 percent, with the size of the effect depending on the ratio of exports to AEs to GDP.

This heterogeneity in the size of the effects of AE output on EMs is actually amplified through another related channel, namely the effect of AE output on commodity prices. An increase in AE output increases the demand for commodities and therefore increases their price. This implies further heterogeneity in the effects of AE output on EMs. Net commodity exporters benefit more, and commodity importers benefit less and possibly not at all from an increase in U.S. output.

2.2 Expansionary AE Monetary Policy Leads to EM Exchange Rate Appreciation

This effect has been in evidence since the beginning of the crisis, although monetary policy has been only one of the factors moving exchange rates. The acute phase of the crisis was dominated by an increase in market risk aversion and by repatriations of funds by AE banks, leading to large capital outflows and depreciations of EM currencies despite a sharp decrease in AE policy rates. Thereafter, low interest rates in advanced economies led to a return of capital flows to EMs. Adjustments in policies, current or anticipated, have led to large exchange rate movements, among them the "taper tantrum" of 2013 when the Federal Reserve indicated that it would slow down its purchases of bonds, leading to increases in long rates and large depreciations in a number of EMs.

EM policymakers have complained about the "unconventional" character of monetary policy in this context, but there is no reason to think that, with respect to exchange rate movements, unconventional monetary expansion works very differently from conventional monetary policy: To the extent that unconventional policy decreases spreads on domestic bonds, whatever their type or maturity, it makes them less attractive and leads to depreciation.

Depreciation in turn leads to an increase in net exports. The argument has been made that exchange rate changes no longer improve the trade balance. The evidence suggests, however, that they still do. A recent International Monetary Fund (IMF) study concludes that the Marshall-Lerner condition (appropriately modified to account for incomplete pass-through) still holds: A real depreciation of 10 percent leads, on average, to an increase in real net exports over time of 1.5 percent of GDP, with a fairly wide range from 0.5 percent to 3.0 percent of GDP, reflecting in part the variation in export shares across AEs and EMs.[5]

Again, it is useful for later to do a back-of-the-envelope computation. Assuming that uncovered interest parity (UIP) holds at least as an approximation, assuming that AE real interest rates are expected to be lower than EM interest rates by 1 percent for, say, three years, this implies an initial EM real appreciation of 3 percent. Putting this together with the previous numbers, and with all the proper caveats, the exchange rate channel suggests an average decrease in EM real net exports of 0.45 percent of GDP, with a range going from 0.15 percent to 0.9 percent of GDP, taking place over a number of years.

2.3 Expansionary AE Monetary Policy Affects EMs' Financial Systems

Perhaps the loudest complaints about AE monetary policies have been those aimed at gross inflows, at the so-called "tsunamis of liquidity"[6] triggered by AE monetary policies, and their perceived adverse effects on EMs' financial stability.

The image of tsunamis of liquidity rushing into EM financial systems is a very powerful one. It is, however, also a very misleading one. A decrease in the AE policy rate indeed leads AE investors to increase their demand for EM assets. Thus, at a given exchange rate, it indeed leads to an increase in gross inflows to EMs. In the absence of foreign exchange (FX) intervention, and on the assumption that net exports only adjust over time, these gross inflows must, however, be matched by equal gross outflows in order for the foreign exchange market to clear. Put another way, whatever "tsunami" of inflows is triggered by AE monetary policy must be matched

[5] See IMF (2015, ch. 3). See also the study by Bussiere, Gaulier, and Steingrass (2016), which reaches similar conclusions.

[6] I believe the expression was first used by Dilma Rousseff in 2012.

by an equal tsunami of outflows: "net tsunamis" must be equal to zero. This is achieved through the decrease in the AE exchange rate—equivalently, the appreciation of the EM currency.

This does not mean, however, that EM policymakers are wrong when they think that AE monetary policy affects their financial system. Empirical work, in particular by Hèlène Rey (for example, Miranda-Agrippino and Rey 2015) suggests that U.S. monetary policy indeed has important and complex effects on other countries' financial systems. Why might this be? It is fair to say that, despite a great deal of recent and ongoing research, we do not yet have a good sense of the specific channels and of their relative importance. For this reason, I shall leave the effect of AE monetary policy on EM financial stability out of the model in the next section. I shall, however, return to the issue in section 3, review what we know and do not know, and discuss potential implications.[7]

3. The Scope for Coordination

Do these cross-border effects, these spillovers, imply a role for coordination, as the Rajan quote in the introduction suggests? The first step in exploring the answer is to define coordination more precisely, and here I want to take exception with some of the existing rhetoric:

- Coordination is not about more communication. Surely, in the current environment, a better understanding of each other's macroeconomic policies can only help. Thus, G-7 or G-20 meetings and discussions are clearly desirable. This is, however, too unambitious a definition of coordination.
- Coordination is not about asking some countries to modify their policies to help others at their own expense. This is too ambitious a definition of coordination, and unlikely to ever happen. The argument that countries play repeated games, and thus may be willing to sacrifice in the short run in order

[7]As a result of the difference between what we know about the first two channels and the third, the paper is a bit schizophrenic. The first two sections build on an old literature, with a few new twists. The third section is highly speculative. A more ambitious paper would integrate the three channels in one model, but we are/I am not there yet.

to have others do the same in the future if and when needed, is unlikely to convince policymakers.

- Coordination is not about asking policymakers to take into account “spillbacks,” i.e., the effects of their policies on their country through their effects on other countries.[8] This may be the case if, for example, AE policies lead to major difficulties in EMs, which lead in turn to doubts about financial claims on EMs, which, finally, lead to financial problems for AE banks. Typically, these spillbacks are small, and, in any case, policymakers should take them into account on their own. This does not qualify as coordination.
- Coordination is not about asking policymakers to follow policies that they feel they cannot or simply do not want to adopt. I feel that this is part of what the “G-20 map” process, which is the G-20 version of coordination, does.[9] It suggests to countries that they should do more structural reforms, and appropriately modify monetary and fiscal policies. This may be the right advice, but if it is correct, countries should do much of it on their own, whether or not other countries do what is asked of them.

I shall instead take coordination to mean a set of changes in policies that makes all countries better off. More formally, I shall ask whether the decentralized equilibrium, which I shall take to be the Nash equilibrium, is efficient, or whether it can be improved upon.[10,11]

---

[8] See, for example, Caruana (2015).

[9] See https://www.imf.org/external/np/exr/facts/g20map.htm for a description of the G-20 map process, and the 2012 Umbrella Report for G-20 Mutual Assessment Process (http://www.imf.org/external/np/g20/pdf/062012.pdf) for more details.

[10] This is the standard academic definition and the one used, for example, by Jeff Frankel in the paper presented at the 2015 Asian Monetary Policy Forum (Frankel 2016). His paper, titled “International Coordination,” touches on many of the same points I do.

[11] I leave aside the international provision of public goods, such as the provision of liquidity by the IMF or by central banks, the harmonization of financial regulations, etc. These are obviously important but are a very different form of coordination.

With this definition, the general answer is simple and well known: If countries have as many non-distorting instruments as they have targets, then the Nash equilibrium is efficient, and there is no room for coordination to improve outcomes for all countries. The reason is obvious: Whatever other countries do, countries have sufficiently many instruments to achieve the targets they want.

A general discussion of whether countries have as many instruments as targets can get very abstract and sterile. One can think of targets as being the output gap, inflation, the exchange rate, and financial stability, and instruments as being monetary policy, fiscal policy, macroprudential policy, FX intervention, and capital controls. Simple counting of instruments and targets is unlikely to resolve the issue: Some of the policy instruments are likely to create distortions, so that they enter both as targets (minimizing the distortion) and as instruments. If all instruments are distortionary, for example, then it follows that there will always be more targets than instruments and there will always be room for coordination to improve the outcome. But if the distortions are small, the gains from coordination may be limited. It is more useful to work through a simple formal model and show what this implies.

3.1 A Two-Country Mundell-Fleming Model

For my purposes, let me start with a simple and old-fashioned two-country Mundell-Fleming model. The model is old fashioned in two ways: First, it is static and not derived from microfoundations.[12] Given the logic behind the conclusions, I am confident that they would hold in a more microfounded and more general model. Second, it leaves out the third channel discussed earlier, the effects of AE monetary policy on EM financial stability. The reason is that I feel we/I do not know how to best extend the model to capture these effects. Thus, I leave this extension to an informal discussion in the next section.

The model has two (blocks of) countries, a domestic economy (as a stand-in for advanced economies) and a foreign economy (as

[12]For a treatment of the scope for coordination in a microfounded model, see Obstfeld and Rogoff (2002).

a stand-in for emerging market economies). Foreign variables are denoted by an asterisk.

Domestic output is given by

$$Y = A + NX$$
$$A = G - cR + X$$
$$NX = a(Y^* - Y) - bE.$$

Domestic output, Y, is equal to the sum of absorption, A, and net exports, NX. Absorption depends on fiscal policy, summarized by G, on the monetary policy rate, R, and on a shock to domestic demand, X. Net exports depend positively on foreign output, Y^*, negatively on domestic output, Y, and negatively on the real exchange rate, E.

Symmetrically, foreign output is given by

$$Y^* = A^* - NX$$
$$A^* = G^* - cR^* + X^*$$
$$NX = a(Y^* - Y) - bE.$$

Finally, following UIP, the exchange rate depends on the difference between the domestic and the foreign policy rates. Under the UIP interpretation, the coefficient d measures the expected persistence of the interest differential:

$$E = d(R - R^*).$$

A decrease in the domestic policy rate over the foreign policy rate leads to a depreciation of the domestic currency—equivalently to an appreciation of the foreign currency.

Absent shocks, G, G^*, X, X^* are normalized to zero. This normalization implies that equilibrium output in the absence of shocks, which I take to be potential output, is equal to zero. So are net exports, interest rates, and the exchange rate.

Each country cares about internal balance, i.e., the deviation of output from potential, and external balance, the deviation of net exports from zero.

$$\Omega = \min Y^2 + \alpha \, NX^2$$
$$\Omega^* = \min Y^{*2} + \beta \, NX^2$$

To start with, assume that each country can use both fiscal and monetary policies. As there are two targets and two non-distorting instruments in each country, the theorem applies: The Nash equilibrium is efficient, and there is no room for coordination. Suppose we capture what has happened during the crisis by assuming that, starting from steady state in both countries—so, given the normalization, all variables are equal to zero—the domestic economy is hit by an adverse demand shock, so $X < 0$. Then, the Nash equilibrium is trivially characterized: The domestic economy uses fiscal policy, $G = -X$, to offset the shock, and the foreign economy does not need to change either G^* or R^*.[13]

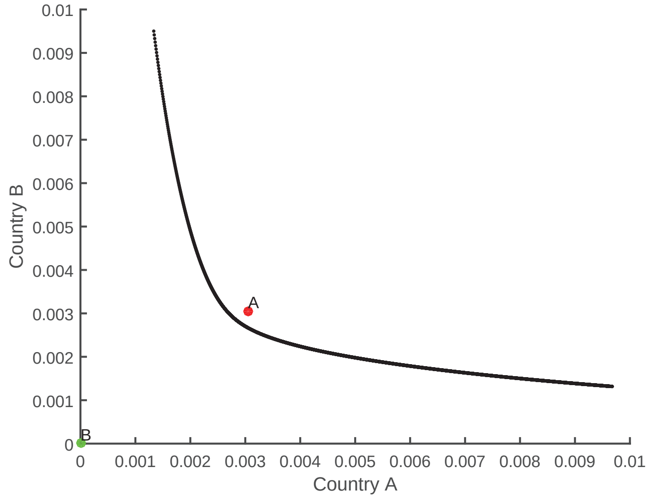
One may worry about the fact that, in the model and clearly counterfactually, the two countries completely offset the shock and return to the pre-shock equilibrium. This is not essential. The shock may be (and indeed was) a more complex one, affecting for example the supply side, so that the countries want to return to a different equilibrium after the shock. And the model is easily extended to limit the ability of policy to offset the shocks. If, for example, decisions about fiscal and monetary policies are made before X is fully revealed, the economies will be affected by the shock, but the efficiency of the Nash equilibrium will remain. Coordination cannot improve the outcome.

3.2 Coordination when Fiscal Policy Cannot Be Used

Why does the above result feel too strong? Probably because the potential role attributed to fiscal policy is too optimistic. Policymakers may/do care about the fiscal balance, in which case, formally, there are now three targets and only two instruments. Related, and more relevant at this point, given the large increase in debt associated with the crisis, are the perceived limits on the current use of fiscal policy. Indeed, a recurring theme of policy discussions has

[13] Actually, the equilibrium set of policies is not unique. One can verify that any equilibrium where R and R^* move together, implying no change in the exchange rate, and G and G^* adjust so as to maintain demand constant in each country is efficient. But this is a curiosity.

Figure 1. AE and EM Welfare under Nash and Coordination



been the extreme reliance on monetary policy due to the perceived limits on the use of fiscal policy.[14]

What happens if we assume that fiscal policy cannot be used, so that $G = G^* = 0$?[15] In this case, each country has two targets and only one instrument. The Nash equilibrium is inefficient, and there is a set of policies that improve welfare in both countries.

The set of utilities that can be achieved through coordination is obtained by maximizing a weighted average of the two countries' welfare functions, $\Omega + \lambda\Omega^*$ for different values of λ. Figure 1 plots the Nash equilibrium, A, and the utility frontier for a given set of parameters (the qualitative feature of the figure does not depend on

[14]Many economists, including me, have questioned whether fiscal policy is really unavailable. They have argued that, even at the currently high debt levels, there may be room for fiscal expansion. I leave this debate aside here. All I need for the argument made here is that there are some perceived limits on the use of fiscal expansion.

[15]Equivalently, we could assume that fiscal policy can be used, but that it creates distortions, with these distortions entering the objective function. This would lead to a more limited role for fiscal policy, and the essence of the results below would go through.

the specific set of parameters.) All the points to the southwest of A yield higher welfare for both countries.[16]

The improvement in welfare is small, and this conclusion is consistent with the literature, from Oudiz and Sachs (1984) to a recent paper by Taylor (2013).[17] The next question is what form coordination should actually take. Should coordination lead AEs to adopt a more or a less aggressive monetary policy?

The answer turns out to depend on the sign of $(ac - bd)$. This expression has a simple interpretation. The first term, ac, reflects the strength of the first channel (higher AE output, leading to a stronger demand for EM exports) above, with c measuring the effect of the policy rate on demand and a measuring the share of imports. The second term, bd, reflects the strength of the second channel (EM appreciation, leading to a decrease in demand for EM exports), with d measuring the effect of the policy rate on the exchange rate and b measuring the effect of the exchange rate on net exports.

When the first channel dominates the second, the net effect of a decrease in the domestic policy rate is to increase foreign net exports and foreign output. The coordination equilibria (I use "equilibria" as there is a (small) range of equilibria that dominate the Nash equilibrium, namely all the points to the southwest of A) are associated with a stronger response of the domestic policy rate and a weaker response of the foreign policy rate than under Nash. When the second channel dominates the first, however, the coordination equilibria

[16]Given that we are minimizing a loss function, the closer to the origin, the better.

[17]Given the simplicity of the model and the lack of a serious calibration, this conclusion cannot be given too much weight. But the result is, in fact, quite robust, and is related to the discussion that comes below: A change in the AE interest rate has two opposite effects on EM countries, higher demand for exports and an exchange rate appreciation. These largely cancel out, with the implication that the net effect on EMs of the AE policy, and thus the scope for coordination, is limited. The same reasoning applies to the model presented by Taylor. His model is a two-country Mundell-Fleming model, with a specification of demand close to this paper and a supply side characterized by staggered nominal wage setting. Each country has two targets, the standard deviations of output and of inflation, and one instrument, the policy rate. Given that there are fewer instruments than targets, there is room for coordination to improve the outcome. The effects of coordination are small, however, because the two effects of AE monetary policy on EMs, namely higher exports and an exchange rate appreciation, largely cancel out, both for output and for inflation.

Table 1. Policy Rates under Nash and Coordination

a	b	R (Nash)	R^* (Nash)	λ	R (Coord)	R^* (Coord)
0.4	0.2	–.868	–.131	1	–.882	–.117
0.2	0.4	–.767	–.230	1	–.759	–.241

are associated with a weaker response of the domestic policy rate and a stronger response of the foreign rate. When the two channels cancel, the coordinated equilibrium is the same as the Nash equilibrium: In other words, coordination does not help.

Table 1 shows the outcomes for two sets of parameters. The shock is taken to be a decrease in domestic demand, X, by 1, while X^* is unchanged. The parameters α, β, c, and d are the same in both cases and are equal, respectively, to 0.5, 0.5, 1.0, and 1.0. The two lines differ in the values of a and b (and thus the implied value of $ac - bd$, which is positive in the first case and negative in the second).

The coordinated equilibria that dominate the Nash equilibrium all have very similar interest rates, so we can just look at one of them. The table reports the Nash equilibrium domestic and foreign interest rates, and those associated with one of the dominating coordinated equilibria, the equilibrium associated with $\lambda = 1$. In the first case, the first channel dominates, and coordination yields a stronger response of the domestic rate, –88.2 bps compared with –86.8 bps. In the second case, the second channel dominates, and coordination yields a weaker response, –75.9 bps compared with –76.7 bps.[18]

These results point to the practical problem in achieving coordination in this context, namely whether we know which way the inequality goes. My reading of the history of the last seven years is that it is one of major disagreements about the strength of the two effects and, by implication, disagreements about what coordination should achieve.

To go back to the quotes at the beginning, both Guido Mantega and Raghu Rajan emphasized the second channel, the effect

[18]The differences between the rates under Nash and coordination are small, but again, the calibration is too crude for this aspect to be given too much weight.

of AE monetary policy on the exchange rate. To quote Rajan again, "Rather the mandates of systemically influential central banks should be expanded to account for spillovers, forcing policymakers to avoid unconventional measures with substantial adverse effects on other economies, particularly if the domestic benefits are questionable." In terms of our model, Rajan had in mind a small effect of the policy rate on domestic demand, a small value for c. In the limit where c tends to zero, this is indeed a zero-sum game between the two countries, and coordination should lead to smaller policy rate cuts—thus, the use of the term "currency wars."[19]

Advanced economy policymakers, on the other hand, have typically emphasized the first channel. Strong AE growth, they have argued, is essential for the world in general and for EMs in particular. In terms of the model, they have emphasized the importance of a, the effect of AE output on AE imports. In his 2015 Mundell-Fleming lecture, which deals very much with the same topics as this paper, Ben Bernanke argued, "US growth during the recent recovery has certainly not been driven by exports, and, as I will explain, the 'expenditure-augmenting' effects of US monetary policies (adding to global aggregate demand) tend to offset the 'expenditure-switching' effects (adding to demand in one country at the expense of others)" (Bernanke 2016).

Who is right? The back-of-the-envelope computations given in section 1 suggest that it is hard to assess which way the inequality works. Indeed, different econometric models give different results. Taylor (2013) gives the results of simulations from two large multi-country models, one based on the TMCM model built by Taylor himself (1993) and the other by Carabenciov et al. (2013). The first simulation focuses on the effects of U.S. monetary policy on Japan and finds a small positive impact of a U.S. monetary expansion on

[19] John Taylor (2013) has suggested an alternative interpretation of the source of EM complaints. He has argued that the problem came from suboptimal policies in AE countries, namely too-low interest rates, leading to larger adverse effects on EM economies. Within the logic of the model presented here, as well as in the logic of the model he uses, this is not convincing. Because the net effects of a change in the AE interest rate on EMs are small, the "wrong" interest rate in AEs is unlikely to have a major impact on EMs' output, trade balance, or inflation outcomes. Whether effects through other channels, such as effects on EM financial systems, can strengthen the argument is discussed in the next section.

Table 2. Effects of an AE Monetary Expansion on Output in AEs and EMs

Year	1	2	3	4	5	6
Advanced Economies	1.00	1.60	1.38	0.94	0.61	0.39
Emerging Economies	0.17	0.39	0.39	0.33	0.28	0.22

Japan's output. The second simulation also finds a small positive effect of a U.S. monetary expansion on Japan, but a negative effect on both Latin America's and emerging Asia's output. The IMF modeling team was kind enough to run another simulation of that model for this paper, and the results are given in table 2. The experiment is an AE monetary expansion in response to a decrease in domestic demand in AEs, and the table shows the effects of the monetary expansion on both AE and EM output, from year 1 to 6 (the numbers show the difference between output with and without the monetary expansion). The numbers show a net positive effect of the AE monetary response on EM output.

While such a simulation is much more sophisticated than the simple computations in section 1, it still comes with many caveats. In particular, it comes with likely large differences across EMs. EM countries with strong trade links to AEs, such as China, may indeed be better off and be in favor of more AE expansion. EM countries with weaker links to AEs, such as Brazil or India, may be worse off and want less AE expansion; this may explain why Brazil and India may have been among the most vocal critics of AE policy.

In short, given the diverging views and the lack of solid evidence, coordination means something different for AE and EM policymakers, so it is unlikely to happen.

3.3 A Deus ex Machina? Capital Controls

If, because of limits on fiscal policy, the Nash equilibrium is inefficient and the room for coordination is limited, can policymakers improve on the Nash outcome? The short answer is yes, if they are willing and able to use an additional instrument: restrictions on capital flows, i.e., capital controls.

The logic for why capital controls are useful in this context is straightforward. Advanced economies suffer from a lack of domestic demand. As we saw earlier, if they could freely use fiscal policy, they could just offset the decrease in domestic demand through a fiscal expansion. This would return both countries to the pre-shock equilibrium levels of output and exchange rate. If fiscal policy is not available, they must use monetary policy. Monetary policy, however, not only increases domestic demand but also affects the exchange rate through interest differentials. Capital controls can, at least within the logic of the model, eliminate the effect of the interest differential on the exchange rate.

This argument can be formalized as follows. Extend the equation for the exchange rate to

$$E = d(R - (R^* - x)),$$

where x may be interpreted as a tax per unit on foreign inflows (such as has been used in Chile, or more recently in Brazil). Assume, as above, that fiscal policy cannot be used; that AEs can use monetary policy, R; and that EMs can use monetary policy R^* and the tax x. Assume again that the shock is a decrease in X by 1.

Then the Nash equilibrium takes a simple form. AEs decrease the policy rate R by $1/c$. EMs increase x by $1/c$, leaving the exchange rate unchanged. AE output and net exports return to their pre-shock level (zero, by normalization). In terms of figure 1, the two countries achieve the point at the origin, a large improvement relative to the Nash or the coordinated equilibrium absent controls. Not only do EMs protect themselves, but AEs also benefit from being able to use monetary policy without having to worry about the exchange rate.

In short, (varying) capital controls are the logical macroeconomic instrument to use when fiscal policy is not available. It reduces the problems associated with an increased reliance on monetary accommodation. Such an endorsement of capital controls comes with many caveats. Before returning to them, I turn to the case for capital controls as a financial instrument.

4. Monetary Policy, Capital Controls, and FX Intervention

In the previous section, I left aside the third channel, i.e., the potential effects of AE monetary policy on gross inflows into EMs and on the EM financial system. But, as I discussed earlier, many of the EM complaints have been aimed precisely at those gross inflows and their perceived adverse effects on financial stability.

How does AE monetary policy affect gross flows to EMs and the EM financial system? Despite a lot of recent work, the answers are less clear than one would like, on both theoretical and empirical grounds.

4.1 Gross Flows and AE Monetary Policy: Theoretical Considerations

Let me first dispose quickly of the simplest but fallacious version of the "tsunami" argument, namely that monetary policy "unleashes large flows into EMs." Write down the equilibrium condition in the foreign exchange market as

$$FI = FO + FX - NX,$$

where FI denotes gross inflows, FO denotes gross outflows, FX denotes foreign exchange intervention, and NX is the current account surplus. In the very short run (say from a few minutes to a few months), the current account balance does not move very much. So, in the absence of foreign exchange intervention, to a close approximation the following equality $FI = FO$ must hold. Gross inflows must be matched by gross outflows. Put another way, foreign exchange market equilibrium implies that "tsunami" inflows must be matched by equal outflows.[20]

Even if gross inflows are offset by gross outflows, this does not imply that their effects on EM financial systems cancel each other. If both go up, for example, it may be that the effects of larger inflows are quite different from the effects of the larger outflows. To explore this, one must first look into what happens to gross inflows (and by implication, gross outflows) in response to a monetary expansion in AEs. This requires one to specify the determinants of the gross inflows and outflows.

[20] This paragraph may be seen as fighting a straw man. In my time at the IMF, however, I found that this vision of gross flows as waves of cash finding their way into EM banks, the EM stock market, etc., without taking into account the necessary countervailing outflows, was quite widespread among policymakers.

Assume that gross inflows into EMs and gross outflows from EMs are given by

$$FI = \alpha + \beta(d(R^* - R - z) + E)$$
$$FO = \alpha^* - \beta^*(d(R^* - R - \gamma z) + E).$$

Both inflows and outflows are now assumed to be less than fully elastic with respect to expected returns. Both α and α^*, and β and β^* are allowed to differ, reflecting potentially different preferences and types of AE and EM investors.[21]

The variable z shifts inflows and outflows; it can be thought of as reflecting a risk premium, reflecting the convolution of perceptions of risk and risk aversion; its effect may be different for AE and EM investors, and this is captured by the presence of coefficient γ. For example, "risk off" may lead AE investors to become more risk averse, while having less of an effect on EM investors, in which case $\gamma < 1$.

Note that as β and β^* go to infinity, and z goes to zero, the equilibrium tends to the uncovered interest parity condition $E = d(R - R^*)$.[22]

Assume, as we did above, that we are looking at the short run so we can ignore movements in the current account, so equilibrium in the foreign exchange market is simply given by

$$FI = FO + FX.$$

Suppose now that the AE central bank decreases its policy rate R by $\Delta R < 0$, that the EM central bank does not adjust its policy

[21] This assumes that it is the flows that respond to expected return differentials. A more appealing, but more complicated, assumption is that desired stocks respond to expected return differentials, with the stock adjustment taking place over time. Such a specification would, however, lead to similar conclusions as those reached in the text, with one caveat: To the extent that the wealth of domestic and foreign investors was affected differentially by the exchange rate, these wealth effects would affect inflows and outflows, and potentially lead to a change in gross flows in response to AE monetary policy. I ignore these effects here.

[22] Note also that just replacing the UIP condition in the previous section with these equations would not change the conclusions reached there about the role and the limits of coordination.

rate, so $\Delta R^* = 0$, and does not intervene, so $FX = 0$. Solving for the equilibrium gives

$$\Delta E = d\Delta R \quad \text{and} \quad \Delta FI = \Delta FO = 0.$$

In words, the exchange rate adjusts so as to keep expected relative returns the same, just as under the UIP condition, and the decrease in the exchange rate leads to unchanged gross inflows (and outflows). This is true despite less than fully elastic flows, different preferences of AE and EM investors, and possibly different risk premia.[23]

How can the result of unchanged gross flows be overturned?

Looking beyond the short run, the current account responds over time to the appreciation. Starting from the current account balance, the current account turns into deficit. Going back to the equilibrium condition, this implies a capital account surplus. Gross inflows increase, gross outflows decrease. Net inflows increase, corresponding to the deterioration of the current balance. This, however, takes time and the size of the net inflows may be small relative to the initial shift in gross flows (at a given exchange rate).

Keeping the focus on the short run, I can think of two ways to overturn the result:

(i) Demands for domestic and foreign investors differ in more fundamental ways than introduced here. I do not, however, have a sense of what plausible deviations to introduce.[24]

(ii) Monetary policy works partly through its effects on the risk premium. Suppose, for example, that lower AE rates decrease the risk premium z by Δz. Then

$$\Delta E = d\,\frac{\beta + \beta^*\gamma}{\beta + \beta^*}\,\Delta z$$

$$\Delta FI = \Delta FO = d\,\frac{\beta^*(\gamma - 1)}{\beta + \beta^*}\,\Delta z.$$

[23] This remains true even if R^* adjusts. The adjustment has an effect on the exchange rate, not on the gross flows.

[24] Following on the caveat in a previous footnote, a stock-flow specification, allowing for wealth effects due to the change in the exchange rate, could lead to a change in equilibrium gross flows. While I have not explored its empirical relevance, I suspect the effect would be small.

> If γ is less than one—that is, if EM investors are less sensitive to z than AE investors—then the exchange rate appreciation is more limited, and gross inflows and outflows increase. Thus, if a decrease in the policy rate is associated with a decrease in the risk premium, and if $\gamma < 1$, then a monetary expansion is associated with higher gross flows.

This line of explanation suggests a complex relation between monetary policy—conventional or unconventional—and gross flows, depending on co-movements between the risk premium and monetary policy. For example, QE1 may have reassured AE investors that U.S. markets would be less dysfunctional, leading to a return of AE investors to the United States and a decrease in gross flows to EMs. In contrast, QE2 may have had little effect on perceived risk, and led AE investors to increase gross flows to EMs. The taper tantrum may have led to a decrease in gross flows to EMs, not so much by tightening future U.S. monetary conditions but rather by increasing uncertainty about the course of future U.S. monetary policy.

4.2 Gross Flows and AE Monetary Policy: Empirical Evidence

Despite a large number of empirical studies, the evidence on the effects of AE monetary policy on gross flows is also unclear. The empirical difficulties are many, from the usual difficulty of identifying monetary policy shocks, compounded since the crisis by the zero lower bound and the lack of movement in the policy rate, to the use of unconventional instruments, to the issue of separating out expected and unexpected monetary policy actions, to quality or coverage issues with the flow data.

A number of studies have found an effect of monetary policy on specific gross flows. Bruno and Shin (2015), for example, using a VAR methodology over the pre-crisis period (1995:Q4 to 2007:Q4) find an effect of the federal funds rate on cross-border bank-to-bank flows; the effect is, however, barely significant. Fratzscher, Lo Duca, and Straub (2013), using daily data on portfolio equity and bond flows, find significant effects of different monetary policy announcements and actions since the beginning of the crisis.[25] Their results, however, point to the different effects of apparently largely similar monetary measures. For example, they find that QE1 announcements decreased bond flows to EMs, while QE2 announcements increased them. In terms of the equations above, this indeed suggests that, in each case, monetary policy worked partly through its effects on the risk premium, and that different announcements had different effects on that premium.

These studies cannot settle, however, the issue of whether total gross inflows increase with AE monetary expansions: The increase in the inflows the researchers have identified may be offset by a decrease in other inflows. Studies of total inflows, or of the set of inflows adding up to total inflows, yield mixed conclusions. A representative and careful paper, by Cerutti, Claessens, and Puy (2015), using quarterly flows over 2001:Q2 to 2013:Q2, suggests two main conclusions. The most significant observable variable in explaining gross flows into EMs is the VIX index: An increase in the VIX leads to a decrease in gross inflows to EMs. The coefficients on the monetary policy variables, namely the expected change in the policy rate and the slope of the yield curve, typically have the expected sign but are rarely significant. Together, these two variables explain only a small part of overall variations in capital flows.

Thus, on both theoretical and empirical grounds, the relation of monetary policy to gross inflows into EMs is less clear than is often believed by policymakers and even by researchers.[26]

4.3 Gross Inflows and EM Financial Systems: Other Channels?

Leaving aside the effects if any on the volume of gross flows, how may AE monetary policy affect the EM financial systems? One can think of two channels.

---

[25] See also Koepke (2015).

[26] This suggests that statements like "the empirical literature has long established that US interest rates are an important driver of international portfolio flows, with lower rates 'pushing' capital to emerging markets" (Koepke 2015) are too strong. To be clear, the issue is not whether they affect exchange rates—they do—but whether they lead to large increases in gross flows—which is less settled.

The first channel, which the Asian crisis put in evidence, is through the effect of the exchange rate itself on the financial system. To the extent that financial institutions, the government, firms, or households have foreign-currency-denominated claims and liabilities, the appreciation triggered by AE monetary policy will affect their balance sheets. Even if financial institutions are largely hedged, unhedged positions by the others will affect the value of their claims and affect financial stability. The effects on financial stability are likely to vary in magnitude, and even in sign, across countries, depending on the structure of foreign-currency-denominated claims. In general, given that most EM countries still borrow largely in foreign currency, the effect of an appreciation triggered by AE monetary policy should be favorable (so it does not explain the complaints of EM policymakers to AE monetary accommodation). The exact structure of claims and liabilities will, however, matter.

The second channel is through changes in the composition of gross inflows and outflows triggered by AE monetary policy. If, for example, foreign investors increase their holdings of sovereign bonds and domestic investors decrease theirs, then the effects on the financial system are likely to be limited. If instead, inflows take the form of additional funds to domestic banks, and outflows come from a decrease in holdings of sovereign bonds, then this is likely to lead to an increase in domestic credit supply. Depending on its nature and intensity, this increase may be desirable or instead lead to an unhealthy credit boom.[27]

It is clear, for example, that, at the beginning of the crisis, the repatriation of funds by AE banks had such a composition effect. The decrease in funding to EM banks by AE banks was not compensated by an increase in funding of EM banks by EM investors, leading to a tightening of credit. The issue at hand is, however, about the effects of monetary policy per se. Just as for the effect of AE monetary policy on overall gross flows, the evidence on the composition of the flows triggered by AE monetary policy is not clear. In Cerutti, Claessens, and Puy (2015), for example, there is no clear difference between the estimated effects of monetary policy variables on bank, portfolio debt, and portfolio equity flows.

[27]See Blanchard et al. (2016).

Thus, overall, it is difficult to conclude that AE monetary policy has had major, predictable effects on EM financial systems. Nevertheless, it is clearly a potentially important dimension that EM policymakers must monitor. This takes me back to the issue of capital controls, now in the context of financial stability.

4.4 Capital Controls versus FX Intervention

While the use of capital controls has been limited, many countries have relied on FX intervention to limit the movements in exchange rate caused by AE monetary policy. From the macroeconomic point of view of the previous section, i.e., leaving implications for gross inflows aside, controls and FX intervention are largely substitutes. Under the assumption that the elasticity of flows to return differentials is finite—a necessary condition for FX intervention to have an effect—both can limit the effects of lower AE interest rates on the exchange rate and achieve the same macroeconomic outcome. If, however, we take into account the channel discussed in this section, the two have very different implications. Capital controls, by assumption, can limit gross inflows. FX intervention, by limiting the exchange rate adjustment, increases gross inflows. This can be seen straightforwardly from above. If, in response to a decrease in the AE policy rate, FX intervention keeps the exchange rate unchanged, gross flows increase by

$$\Delta FI = -bd\Delta R > 0.$$

Thus, if the purpose is to limit the effects of AE monetary policy on the EM financial system, capital controls clearly dominate FX intervention.

5. Conclusions

I have looked at the interactions between AE and EM macro policies since the beginning of the crisis, interactions characterized by complaints of “currency wars” and demands for more coordination. I have offered three main sets of conclusions.

In AEs, limits on fiscal policy have led since the beginning of the crisis to an overreliance on monetary policy. This potentially opens the scope for coordination. Whether coordination would entail an increase or a decrease in interest rates in AEs is, however, difficult to assess, with AEs and EMs disagreeing about the sign. This has made and still makes coordination de facto impossible to achieve.

If there are limits on the use of fiscal policy, leading to the over-reliance on monetary policy and undesirable effects on the exchange rate, the natural instrument in this context is the use of capital controls by EMs. It allows AEs to use monetary policy to increase domestic demand, while shielding EMs from the undesirable exchange rate effects. In the context of limits on fiscal policy, controls are a natural macroeconomic instrument.

Despite some progress, how AE monetary policy affects EM financial systems remains largely unsettled, both theoretically and empirically. To the extent that AE monetary policy leads to gross inflows into EMs, to the extent that these gross flows affect the EM financial systems, and to the extent that EMs want to avoid these effects, capital controls rather than FX intervention are the right instrument.

These conclusions come with the usual and strong caveats. Technical and political issues associated with the use of capital controls as contingent instruments are still relevant. This is not an unconditional endorsement of controls, but an exploration and a starting point to a discussion.

References

Bernanke, B. S. 2016. "Federal Reserve Policy in an International Context." Forthcoming in *IMF Economic Review*. Mundell-Fleming lecture.

Blanchard, O., J. Ostry, A. Ghosh, and M. Chamon. 2016. "Capital Flows: Expansionary or Contractionary?" *American Economic Review: Papers and Proceedings* 106 (5): 565–69.

Boz, E., M. Bussiere, and C. Marsilli. 2015. "Recent Slowdown in Global Trade: Cyclical or Structural?" In *The Global Trade Slowdown: A New Normal?* ed. B. Hoekman, 55–70 (chapter 3). VoxEU.org E-book. CEPR Press.

Bruno, V., and H. S. Shin. 2015. "Capital Flows and the Risk-Taking Channel of Monetary Policy." *Journal of Monetary Economics* 71: 119–32.

Bussiere, M., G. Gaulier, and W. Steingress. 2016. "Global Trade Flows: Revisiting the Exchange Rate Elasticity." Manuscript, Banque de France (September).

Carabenciov, I., C. Freedman, R. Garcia-Saltos, D. Laxton, O. Kamenik, and P. Manchev. 2013. "GPM6: The Global Projection Model with Six Regions." IMF Working Paper No. 13/87.

Caruana, J. 2015. "The International Monetary and Financial System: Eliminating the Blind Spot." Panel remarks at the IMF conference "Rethinking Macro Policy III: Progress or Confusion?" Washington, DC, April 16.

Cerutti, E., S. Claessens, and D. Puy. 2015. "Push Factors and Capital Flows to Emerging Markets: Why Knowing Your Lender Matters More than Fundamentals." IMF Working Paper No. 15/127.

Frankel, J. 2016. "International Coordination." NBER Working Paper No. 21878.

Fratzscher, M., M. Lo Duca, and R. Straub. 2013. "On the International Spillovers of U.S. Quantitative Easing." ECB Working Paper No. 1557 (June).

International Monetary Fund. 2011. "United States Spillover Report." July 22.

———. 2015. "Exchange Rates and Trade Flows: Disconnected?" In *World Economic Outlook: Adjusting to Lower Commodity Prices* (October). Washington, DC: International Monetary Fund.

Koepke, R. 2015. "Fed Policy Expectations and Portfolio Flows to Emerging Markets." IIF Working Paper (June).

Miranda-Agrippino, S., and H. Rey. 2015. "World Asset Markets and the Global Financial Cycle." NBER Working Paper No. 21722 (October).

Obstfeld, M., and K. Rogoff. 2002. "Global Implications of Self-Oriented National Monetary Rules." *Quarterly Journal of Economics* 117 (2): 503–35.

Oudiz, G., and J. Sachs. 1984. "Macroeconomic Policy Coordination among the Industrial Economies." *Brookings Papers on Economic Activity* (1): 1–75.

Rajan, R. 2014. "Containing Competitive Monetary Easing." Project Syndicate, April 28.

Taylor, J. B. 1993. *Macroeconomic Policy in a World Economy: From Econometric Design to Practical Operation.* New York: W. W. Norton.

———. 2013. “International Monetary Policy Coordination: Past, Present, and Future.” BIS Working Paper No. 437 (December).

Wheatley, J., and P. Garnham. 2010. “Brazil in ‘Currency War’ Alert.” *Financial Times*, September 27.