IJCB
INTERNATIONAL JOURNAL OF CENTRAL BANKING
Volume 4, Number 2
June 2008

Volume 4, Number 2 June 2008

IJCB

INTERNATIONAL JOURNAL OF CENTRAL BANKING

IJCB International Journal of Central Banking

Volume 4, Number 2 June 2008

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

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 at an annual rate of $100 (USD).

Print subscription orders may be placed online at www.ijcb.org, by phone (+49 69 1344 7623), via fax (+49 69 1344 8553), 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
DG Research
European Central Bank
Postfach 16 03 19
D-60066 Frankfurt
Germany

Phone: +49 69 1344 7623
Fax: +49 69 1344 8553
E-mail: editor@ijcb.org

ISSN: 1815-4654

Inflation Forecasts and the New Keynesian Phillips Curve*

Sophocles N. Brissimis[a,b] and Nicholas S. Magginas[c]
[a]Economic Research Department, Bank of Greece
[b]Department of Economics, University of Piraeus
[c]Strategy and Economic Research Division, National Bank of Greece

We examine the ability of the New Keynesian Phillips curve to explain U.S. inflation dynamics when inflation forecasts (from the Federal Reserve's Greenbook and the Survey of Professional Forecasters) are used as a proxy for inflation expectations. The New Keynesian Phillips curve is estimated against the alternative of the hybrid Phillips curve, which allows for a backward-looking component in the price-setting behavior in the economy. The results are compared with those obtained using actual data on future inflation as conventionally employed in empirical work under the assumption of rational expectations. The empirical evidence provides, in contrast to most of the relevant literature, considerable support for the standard forward-looking New Keynesian Phillips curve when inflation expectations are measured using inflation forecasts that are observable in real time. In this case, lagged-inflation terms become insignificant in the hybrid specification. The evidence in favor of the New Keynesian Phillips curve becomes even stronger when real-time data on lagged inflation are used instead of the final inflation data used in standard specifications. Our work is closely related to the work of Roberts (1997), who used survey measures of inflation expectations in an empirical inflation model and found evidence that it is less-than-perfectly rational expectations and not the underlying structure of the economy that account for the presence of lagged inflation in empirical estimates of the New Keynesian model.

JEL Codes: C13, C52, E31, E37, E50, E52.

*We are grateful to John Taylor (editor), an anonymous referee, and Heather Gibson for very helpful comments. The views expressed in this paper are of

1. Introduction

Inflation's short-run dynamics and cyclical interaction with real economic variables is a central issue in macroeconomics and especially in monetary policy analysis. In this respect, important advances have been made during the last two decades in the theoretical modeling of inflation dynamics. Much of the modern analysis of inflation is based on what Roberts (1995) termed the "New Keynesian Phillips curve," a model of price setting based on nominal rigidities (Taylor 1980; Calvo 1983), which implies that current inflation is determined by next period's expected inflation and by real marginal cost as the driving variable. This model is widely used in the analysis of monetary policy, leading Bennett McCallum (1997) to call it "the closest thing there is to a standard specification."

Despite the increasing attention that the New Keynesian Phillips curve has attracted in recent years, there have been conflicting results regarding its empirical validity (Roberts 2005). A large empirical literature has focused on estimating this model, both as a single equation and in the context of a general equilibrium model. Fuhrer and Moore (1995) have argued that the standard New Keynesian model with sticky prices and rational expectations does not fit U.S. postwar data, while Fuhrer (1997a) and Roberts (1997) have shown that modifying the model so as to include lags of inflation not implied by the standard model with rational expectations allows it to fit the data satisfactorily. The work of Chadha, Masson, and Meredith (1992) and Roberts (1998) also provided mixed evidence about the ability of the New Keynesian Phillips curve to fit the data adequately. Recent contributions by Galí and Gertler (1999), Galí, Gertler, and López-Salido (2001), and Sbordone (2001, 2002) have offered evidence in favor of the New Keynesian Phillips curve for the United States and the euro area, while Rudd and Whelan (2005a, 2005b) argue that traditional backward-looking price-setting rules appear to be preferable to the forward-looking alternatives in describing inflation behavior. The ambiguity over the ability of the

the authors and do not necessarily reflect those of their respective institutions. Corresponding author: Brissimis: Economic Research Department, Bank of Greece, 21 E. Venizelos Avenue, 10250 Athens, Greece; E-mail: sbrissimis@bankofgreece.gr; Tel: +30 210 3202388.

New Keynesian Phillips curve to square with the facts appears to arise inter alia from a central implication of the model. Although the price level is sticky in this model, the inflation rate, by contrast, is perfectly flexible—a situation that is at odds with the empirical evidence (Mankiw and Reis 2002). Thus, the model has trouble explaining why shocks to monetary policy have a delayed and gradual effect on inflation (Mankiw 2001). Other sources of difficulty concern the characteristics of the proper measure of the driving variable (i.e., real unit labor cost or output gap) as well as the assumption about the measure of expected inflation used.

A central point in the debate is whether a modification of the model can account for the persistence of inflation detected in the data. A common view is that this is possible insofar as a backward-looking component is allowed for: however, this poses problems from a modeling standpoint, as this component is introduced as an ad hoc feature of the model (Cogley and Sbordone 2005). Thus, the baseline theory underlying the New Keynesian Phillips curve is extended to allow for a subset of firms that set prices according to a backward-looking rule of thumb, creating the so-called hybrid Phillips curve (see Galí and Gertler 1999). There are also other ways to reconcile the empirical evidence with Phillips-curve theory appealing to expectations-related factors. Roberts (1997) shows that models derived under less-than-perfectly rational expectations and models based on alternative microfoundations implying inflation stickiness are in several cases observationally equivalent. The empirical results of Roberts (1997)—obtained by using survey measures of inflation expectations—suggest that it is imperfectly rational expectations and not the underlying structure of the economy that account for the presence of lagged inflation in empirical estimates of the New Keynesian model.

In this paper we reconsider estimates of the New Keynesian Phillips curve in light of recent advances in inflation modeling and use inflation forecasts, which are observable in real time, as a proxy for inflation expectations. Building on the work of Roberts (1997), we estimate the New Keynesian Phillips curve for the United States using inflation forecasts from the Federal Open Market Committee's (FOMC's) Greenbook and the Survey of Professional Forecasters (SPF) and compare the results with the estimates obtained on the basis of actual data conventionally used in empirical work under the assumption of rational expectations. We also evaluate the baseline theory underlying the New Keynesian Phillips curve against the alternative of a hybrid Phillips curve that allows for a subset of firms to set prices according to a backward-looking rule of thumb. Doing so allows us to directly estimate the degree of departure from a pure forward-looking model needed for the Phillips-curve relationship to track the observed inflation persistence. Moreover, given the changes in the definition of the GDP deflator over the sample period, a real-time data set on this deflator (obtained from the Federal Reserve Bank of Philadelphia) is used to derive the lagged-inflation variable in the hybrid specification, with a view to ensuring consistency with the Greenbook and SPF inflation forecasts.

We estimate the alternative Phillips-curve specifications on quarterly U.S. data spanning the period from 1968:Q4 to 2000:Q4 (1968:Q4 to 2006:Q4 in a specification including inflation forecasts obtained from the Survey of Professional Forecasters). The beginning of the sample corresponds to the earlier quarter for which both SPF and Greenbook inflation-forecast data is available, whereas 2000:Q4 is the latest quarter for which Greenbook data is available. All the estimations are made by using the generalized method of moments (GMM). Several results stand out and appear to be quite robust in these estimations. Using the Greenbook and SPF inflation forecasts as proxies for private-sector inflation expectations, we find—in contrast to the findings of Fuhrer (1997a) and Rudd and Whelan (2005a, 2005b, 2006)—that expected inflation becomes the main determinant of current inflation. Overall, the empirical relevance of the hybrid specifications appears to depend largely on the assumption of rational expectations (i.e., the use of actual data on future inflation). Indeed, the lagged-inflation terms in the hybrid specification become insignificant when we approximate inflation expectations with inflation forecasts, which may deviate from full rationality, whereas significant and plausible estimates for the effect of expected inflation and the real unit labor cost are obtained.

The paper proceeds as follows. Section 2 reviews the basic theory underlying the New Keynesian Phillips curve as well as the hybrid Phillips curve and discusses the existing empirical literature. Section 3 presents estimates of different specifications of the Phillips curve using, in turn, actual data and inflation forecasts in estimation, and shows that the forecast-based specifications do a reasonably good job of describing the data. Section 4 puts the results under the perspective of the relevant theoretical literature. Some concluding remarks and tentative implications are provided in section 5.

2. Modeling Inflation Dynamics

2.1 The New Keynesian Phillips Curve

A large part of the literature has used what today is called the New Keynesian Phillips curve, in which the inflation rate is a function of the expected future inflation rate and a measure of real marginal cost, typically the output gap or real unit labor cost (Lindé 2005). The New Keynesian Phillips curve can be derived from microeconomic foundations; see, e.g., Roberts (1995), Woodford (1996), and Rotemberg and Woodford (1997). In particular, as shown in Roberts (1995), the forward-looking dynamics that underlie the New Keynesian Phillips curve emerge from optimal firm responses to obstacles to adjusting prices of the type introduced by Rotemberg (1982) and Calvo (1983).

The New Keynesian Phillips curve, as advocated by Galí and Gertler (1999), is based on a model of price setting[1,2] by monopolistically competitive firms and is given by

$$\pi_t = \frac{(1-\theta)(1-\theta\beta)}{\theta - \theta\eta\mu} s_t + \beta E_t \pi_{t+1}, \tag{1}$$

where s is excess demand, θ is the probability that firms will keep their price unchanged (or proportional to trend inflation), μ is the firm's demand elasticity, and η is the elasticity of marginal cost with respect to output.

Several authors, such as Fuhrer and Moore (1995) and Estrella and Fuhrer (2002), argue that the pure forward-looking New Keynesian Phillips curve has implications that are inconsistent with

[1] This price-adjustment rule is in the spirit of Taylor's (1980) staggered-contracts model.

[2] Similar reduced-form Phillips-curve equations can be obtained using the quadratic adjustment-cost model of Rotemberg (1982).

the data, because of the "jump dynamics" in inflation adjustment that would imply a costless disinflation, which is counterfactual.

Thus, largely empirical reasons provided motivation for the introduction of the hybrid Phillips curve.[3] Fuhrer (1997b) and Roberts (1998) have shown that modifying the model so as to include lags of inflation not implied by the standard model with rational expectations allows it to fit the data satisfactorily. In this vein, Galí and Gertler (1999), with a view to capturing inflation inertia, extend the basic Calvo model to allow a proportion ω of firms to use a backward-looking rule of thumb. The net result is the following hybrid Phillips curve that nests equation (1):

$$\pi_t = \lambda \left(\frac{1}{(1-\eta\mu)} \right) s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1}, \tag{2}$$

where

$$\begin{aligned} \lambda &= \left(\frac{(1-\omega)(1-\theta)(1-\theta\beta)}{\theta} \right) \phi^{-1}, \\ \gamma_f &= \beta\theta\phi^{-1}, \\ \gamma_b &= \omega\phi^{-1}, \\ \phi &= \theta + \omega[1-\theta(1-\beta)]. \end{aligned}$$

While the story may be plausible, it is not derived from an explicit optimization problem, in contrast to the New Keynesian Phillips-curve formulation.

Oddly enough, however, even the hybrid Phillips curve has met with rather limited success in providing a stable and consistent description of inflation behavior. In particular, the relation has proved inadequate to describe inflation dynamics at the

[3]There is also some theoretical work supporting the hybrid-Phillips-curve specification. Brayton et al. (1997) extend the quadratic adjustment-cost model to allow for higher-order adjustment costs, leading to the appearance of lagged inflation in the reduced form of the price-adjustment equation. Smets and Wouters (2003, 2005) and Christiano, Eichenbaum, and Evans (2005) show that the hybrid model can be motivated by a form of dynamic price indexing. Another possibility is that lagged inflation might reflect some form of least-squares learning on the part of private agents, as suggested by Erceg and Levin (2003), Collard and Dellas (2005), and others.

quarterly frequency. Chadha, Masson, and Meredith (1992) and Roberts (1997, 1998) obtain reasonable parameter estimates only with annual and semiannual data. With quarterly data, there are also difficulties in obtaining significant estimates of the effect of the output gap on inflation. In this case, the empirical relevance of the standard specification was improved through the substitution of real unit labor cost for the output gap as the driving variable in the model (Galí and Gertler 1999).[4]

2.2 Near-Rational Expectations and the Phillips-Curve Specification

There are also other ways to reconcile the empirical evidence with Phillips-curve theory. It may be possible to invoke expectations-related factors, such as central bank imperfect credibility or bounded rationality (Roberts 1997, 1998, 2005), to explain sluggish inflation dynamics. In this respect, inflation lags may be thought of as capturing inflation expectations that deviate from full rationality.

Roberts (1997) shows that inflation models derived under imperfectly rational expectations and models based on alternative microfoundations (such as the sticky-inflation model of Fuhrer and Moore 1995) are in several cases observationally equivalent. He argues that real-time measures of expectations, such as those obtained from inflation surveys, "can be used to distinguish between the structural and expectational sources of lagged inflation." To this end, he derives an empirical model that nests the sticky-inflation model as well as the sticky-price model, under the assumption that the inflation surveys are good proxies for inflation expectations in the latter

[4]Recent studies by Galí and Gertler (1999), Galí, Gertler, and López-Salido (2001, 2005), and Sbordone (2001, 2005) have argued that the New Keynesian Phillips curve (as well as the hybrid Phillips curve) is empirically valid, provided that real unit labor cost rather than detrended output is used as the variable driving inflation. Galí, Gertler, and López-Salido (2001) conclude that real unit labor cost is not closely related to the output gap and that monetary policy models need therefore to take into account labor market rigidities. One interpretation provided by the authors is that the results imply that the relationship between real unit labor cost and the output gap is weak. If the labor market is not competitive, labor frictions must be taken into account. Incorporating labor market imperfections is then necessary to model the response of inflation to a monetary policy shock.

model. On the basis of this empirical model and by using survey measures of inflation (obtained from the Michigan and Livingston surveys), he formulates a direct test of the sticky-inflation hypothesis by examining a variety of specifications and finds that, in most cases, the results appear to favor the hypothesis of sticky prices under less-than-perfectly rational expectations over the hypothesis that inflation is inherently sticky. These results would appear to imply that it is imperfectly rational expectations and not the underlying structure of the economy that account for the presence of lagged inflation in empirical estimates of the New Keynesian model. The findings based on survey data would be reinforced if direct estimation of the New Keynesian Phillips curve on the basis of alternative measures of expectations—such as the Greenbook and SPF forecasts, which are generally considered unbiased and broadly efficient measures of expected inflation—yielded similar results.

Thus, the present paper investigates the ability of these two other measures of inflation expectations—i.e., the forecasts included in the FOMC's Greenbook and the SPF forecasts—to account for the actual inflation-expectation formation process in the economy and provides evidence supporting the pure New Keynesian Phillips curve.[5] To the extent that such forecasts provide good proxies for private-sector expectations, they allow us to disregard issues related to the detailed specification of the actual expectation formation process. Thus, we are able to focus exclusively on the question of whether the New Keynesian Phillips curve is correctly specified and describes properly inflation dynamics once expectations are approximated by forecasts observable in real time.

[5]To our knowledge there are a few papers that attempt to estimate the New Keynesian Phillips curve for the United States by using survey measures of expectations. None of them, however, consider the Greenbook as a source of real-time inflation expectations. Roberts (1995, 1997) estimated the Phillips curve using the Livingston and Michigan survey data, arguing that this specification can describe inflation dynamics at a semiannual or annual frequency, although in his model lagged inflation remains significant. Adam and Padula (2003) used the inflation forecasts obtained from the Survey of Professional Forecasters in a quarterly model, whereas Paloviita and Mayes (2005) estimated the Phillips curve for a panel of euro-area countries by using forecast data obtained from the OECD *Economic Outlook*. The lagged-inflation term also remains an important part of the description of inflation dynamics in both papers.

Both the Greenbook and the SPF forecasts appear to incorporate efficiently a large amount of information from all sectors of the economy as well as forecasters' judgmental adjustments, making them ideal as proxies of the private sector's inflation expectations. The SPF forecasts, especially, are considered to have an edge as a summary of the private sector's inflation expectations (Carroll 2003). Moreover, these forecasts are considered to be free of several problems that usually plague other survey forecast data (e.g., excessive gradualism, inefficiency in information processing, etc.).

3. Phillips-Curve Estimation on the Basis of Inflation Forecasts

In this section we present estimates of the New Keynesian Phillips curve as well as of "hybrid" variants of the model, including both forward-looking and backward-looking components, along the lines of Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2001, 2005). We use quarterly U.S. data from 1968:Q4 through 2000:Q4 (2006:Q4 in a specification including the SPF inflation forecasts), where the beginning of the sample is determined by the earlier available inflation-forecast data, and the end of the sample period is determined by the availability of the Greenbook and/or SPF inflation forecasts. Inflation is measured as the annualized quarterly change of the GDP/GNP deflator, which—to ensure consistency with the Greenbook and SPF forecasts—corresponds to the GNP deflator from 1968 to end-1991, the GDP deflator from 1992 to end-1995, and the GDP price index since 1996. The Greenbook and SPF inflation forecasts correspond to the annualized quarterly change in the same deflator one quarter ahead. Given that both measures of inflation forecasts are constructed in real time and are not subject to any revisions—in contrast to the deflator variable, which reflects final/revised data—we also estimate the Phillips curve by using real-time data on the GDP/GNP deflator as provided by the Federal Reserve Bank of Philadelphia to obtain the lagged-inflation variable. Finally, real unit labor cost concerns the nonfarm business sector and is obtained from the National Bureau of Labor Statistics. This is the measure used by Galí and Gertler (1999) and Sbordone (2002).

3.1 Observable Measures of Expectations and Instrumental-Variables Estimation

Given that measurement errors may affect the left-hand-side and especially the right-hand-side variables of the Phillips curve, estimation requires the use of an instrumental-variables (IV) estimator. Measurement errors with respect to the explanatory variables could arise as a result of the use of inflation forecasts as proxies for (unobservable) inflation expectations and of the real unit labor cost as a proxy for real marginal cost. Estimates of the alternative specifications are obtained by using the two-step generalized method of moments.[6]

In implementing the instrumental-variables estimator, we replace the mathematical expectation of inflation with observable inflation forecasts under the assumption that the instruments used correspond to the agents' information set at the time expectations were formed. The use of instrumental variables also helps avoid the possibility that the error term in the equation is correlated either with the demand variable or with the difference between lagged and future inflation.

Given that the Greenbook and SPF forecasts proved to be unbiased and broadly efficient (see, among others, Romer and Romer 2000 and Swanson 2004), it is likely that the forecast errors will be orthogonal to the information available to agents at the time of the forecast. This is important for the consistency of the instrumental-variables estimator, which assumes orthogonality of forecast errors to lagged variables of the information set.

Thus, under the assumption that the forecast error of π_{t+1} is uncorrelated with information dated t and earlier, it follows from equation (2) that

$$E_t\left\{\left(\pi_t - \lambda\left(\frac{1}{(1-\eta\mu)}\right)s_t - \gamma_f E_t\pi_{t+1} - \gamma_b\pi_{t-1}\right)z_t\right\} = 0, \quad (3)$$

[6]It is well known that full-information estimation methods, such as those used by Fuhrer (1997a, 1997b) and Lindé (2005), display greater econometric efficiency when the correct specification of the model is known, but that does not seem to be the case in most monetary policy models. On the other hand, limited-information methods, such as GMM, are robust to incorrect model specification and to uncertainty about modeling assumptions (Roberts 2005).

where z_t is a vector of variables dated t and earlier (and, thus, orthogonal to the inflation surprise in period $t + 1$). The orthogonality condition given by equation (3) then forms the basis for estimating the model with GMM.

The use of future inflation as a proxy for expectations suggests that under the assumption of less-than-perfectly rational expectations, the instruments must reflect information available in real time, i.e., dated at period t and earlier and, in the case of serially correlated errors in the model, at period $t - 1$ or earlier. If we also take into account publication lags, so that agents forming their expectations in period t have information only up to period $t - 1$, then the instruments must be dated at period $t - 1$ and earlier.

Thus, our instrument set includes two lags of the real unit labor cost, the output gap, and nominal-wage growth and three lags of inflation (real-time inflation in the specifications estimated with real-time inflation data). To allow for the possibility of serially correlated errors, all instrument lags are dated at period $t-1$ and earlier, while in the GMM estimation we use the Newey-West weighting matrix, allowing for up to sixth-order serial correlation.

3.2 Estimating the New Keynesian and Hybrid Phillips Curves

Tables 1 and 2 present estimates of different—reduced-form—specifications of the New Keynesian and the hybrid Phillips curves obtained by using, alternatively, actual data on future inflation or the respective Greenbook and SPF inflation forecasts. Probabilities of the J-test for instrument exogeneity are also presented in the last column of the tables.

With respect to estimates of the New Keynesian Phillips curve (presented in table 1), it appears that the inclusion of the real-time forecast of next period's inflation makes relatively little difference to the results, compared with the estimation based on actual inflation data. The coefficient on expected inflation (the discount rate) is in line with what theory would predict as well as with the respective estimate obtained from the specification including actual future inflation. The coefficient on the marginal cost is statistically significant and has the correct sign.

Table 1. New Keynesian Phillips Curve (1968:Q4–2000:Q4)

	π^e_{t+1}	ulc_t	$\bar{R}^2$	J-test
Specification with Greenbook Forecast	0.91 (17.09)	0.04 (7.11)	0.83	0.69
Specification with SPF Forecast	0.95 (23.12)	0.06 (7.65)	0.85	0.71
Specification with Final Data on Future Inflation	0.94 (17.59)	0.03 (2.11)	0.81	0.59

Note: π^e_{t+1} denotes inflation expected by the private sector for period $t+1$, expressed in terms of the annualized rate of change in the GDP deflator; π_{t-1} is the lagged value of the annualized rate of change of the GDP deflator; and ulc_t is real unit labor cost. Numbers in parentheses are t-statistics, and the last column shows the p-values associated with a test of the model's overidentifying restrictions (Hansen's J-test).

As a next step, we evaluate the New Keynesian Phillips curve against the alternative of the hybrid Phillips curve that includes a lagged-inflation term. From table 2, it is clear that the balance of expectation formation moves strongly toward the forward-looking side when we use inflation forecasts (Greenbook or SPF) instead of actual data on future inflation. Indeed, while in the estimate based on actual data that is presented in the last line of table 2 the lagged-inflation term remains significant, explaining about 40 percent of current inflation,[7] it becomes insignificant when the Greenbook or SPF forecasts are used as proxies for expected inflation.[8]

[7]These estimates of the hybrid model including final inflation data are broadly in line with the respective estimates of Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2001, 2005), whereas the small difference in the estimated coefficient on the driving variable is largely attributed to the different sample period (1968:Q4–2000:Q4 in our paper compared with 1960:Q1–1997:Q4 in these papers) as well as to the more parsimonious instrument set used in the present paper.

[8]Adam and Padula (2003) find a statistically significant effect of lagged inflation when estimating a hybrid Phillips curve for the United States using SPF inflation-forecast data. Given the significant measurement errors affecting the explanatory variables, their results are likely to be contaminated by the use of an OLS estimator.

In these forecast-based specifications, the coefficient on expected inflation ranges from 0.84 to 0.86 and is significantly higher than the respective estimates of Galí and Gertler (1999), Galí, Gertler, and López-Salido (2001, 2005), and Sbordone (2002), whereas the coefficient on the real marginal cost is statistically significant and has the correct sign.[9] Moreover, in sharp contrast to these contributions, the estimated coefficient on the backward-looking inflation term is insignificant in all forecast-based specifications. Overall, the inclusion of the observable inflation forecasts appears to increase substantially the weight of the forward-looking variable in inflation

Table 2. Hybrid Phillips Curve (1968:Q4–2000:Q4)

	π^e_{t+1}	π_{t-1}	ulc_t	$\bar{R}^2$	J-test
Specification with Greenbook Forecast	0.84 (4.30)	0.18 (1.03)	0.04 (5.06)	0.85	0.68
Specification with SPF Forecast	0.86 (5.03)	0.21 (1.51)	0.05 (5.25)	0.86	0.59
Specification with Final Data on Future Inflation	0.61 (6.19)	0.38 (3.98)	0.01 (0.28)	0.83	0.65

Note: π^e_{t+1} denotes inflation expected by the private sector for period $t+1$, expressed in terms of the annualized rate of change in the GDP deflator; π_{t-1} is the lagged value of the annualized rate of change of the GDP deflator; and ulc_t is real unit labor cost. Numbers in parentheses are t-statistics, and the last column shows the p-values associated with a test of the model's overidentifying restrictions (Hansen's J-test).

[9]We also estimated the Phillips curve using three alternative measures of the output gap as proxies of the real marginal cost. One was obtained by detrending the GDP series through the application of the Hodrick-Prescott (HP) filter, a second was based on the Congressional Budget Office's measure of potential output, and a third was obtained by using—in view of the significant revisions in the output-gap series—a real-time measure of the output gap constructed on the basis of the Federal Reserve Bank of Philadelphia data set and a recursive estimation of potential output through the application of the HP filter on the GDP data available at each point of time. All the output-gap-based specifications are characterized by a large degree of instability across different subsamples as regards the coefficients of the forward- and backward-looking inflation components as well as a wrongly signed and/or statistically insignificant coefficient on this variable in the earlier part of the sample (i.e., from 1968:Q4 to 1979:Q2).

determination, suggesting that the significance of lagged inflation in conventional specifications reflects largely its role as a proxy for deviations of inflation expectations from full rationality. That being said, the hybrid Phillips curve including the inflation forecasts is rejected in favor of the New Keynesian Phillips curve.

It must be noted that data on the GDP deflator are often subject to substantial revisions. Given that the one-quarter-ahead inflation forecasts from both sources (the Greenbook and the SPF) are constructed before the first revision of the GDP deflator data (which is usually released with a delay of about a quarter), conditioning the pricing decisions of the private sector on final inflation data does not appear to provide a realistic description of the price-setting behavior in real time. Final/revised data on the GDP deflator do not reflect the private sector's information set in real time, nor are the data at the disposal of forecasters at the time they prepare their forecasts. Thus, the significance of inflation forecasts could reflect to a significant extent the inability of the standard hybrid-Phillips-curve specification to account for issues related to real-time availability of data. Therefore, we estimate the hybrid specification using real-time data for lagged inflation. The results presented in table 3 suggest that the use of a real-time lagged-inflation measure confirms the relevance of the New Keynesian specification, as the coefficient on expected inflation increases in both

Table 3. Hybrid Phillips Curve with Real-Time Lagged Inflation (1968:Q4–2000:Q4)

	π^e_{t+1}	π_{t-1}	ulc_t	$\bar{R}^2$	J-test
Specification with Greenbook Forecast	0.95 (4.97)	0.01 (0.57)	0.05 (6.64)	0.85	0.65
Specification with SPF Forecast	0.97 (4.77)	0.16 (0.93)	0.05 (5.85)	0.75	0.63

Note: π^e_{t+1} denotes inflation expected by the private sector for period $t+1$, expressed in terms of the annualized rate of change in the GDP deflator; π_{t-1} is the lagged value of the annualized rate of change of the GDP deflator; and ulc_t is real unit labor cost. Numbers in parentheses are t-statistics, and the last column shows the p-values associated with a test of the model's overidentifying restrictions (Hansen's J-test).

forecast-based specifications. On the other hand, the coefficient on lagged inflation is reduced further and becomes virtually zero, providing some indication that its significance, in specifications based on final data, is likely to reflect the additional—possibly forward-looking—information incorporated through subsequent data revisions rather than its role as a benchmark for rule-of-thumb price setters.

Similar results are obtained when we reestimate the SPF-based specifications on an extended sample ending in 2006:Q4 (table 4). The most notable differences relate to the declining weight on the forward-looking inflation component in the real-time specification, where the coefficient on expected inflation declines to 0.84 (compared with 0.97 in the smaller sample), and also to the declining coefficient on unit labor cost. This latter result could reflect the contamination of unit labor cost data by stock-option exercises, which obscure real marginal cost developments in the most recent part of the sample. Finally, in the last row of table 4, we reestimate

Table 4. Hybrid Phillips Curve: Specification with SPF Forecast—Extended Sample (1968:Q4–2006:Q4)

	π^e_{t+1}	π_{t-1}	ulc_t	$\bar{R}^2$	J-test
Specification with Final Data on Lagged Inflation	0.82 (3.27)	0.25 (1.62)	0.05 (5.06)	0.83	0.71
Specification with Real-Time Lagged Inflation	0.84 (4.15)	0.22 (1.39)	0.04 (5.48)	0.82	0.63
Specification with Real-Time Lagged and Contemporaneous Inflation	0.88 (4.45)	0.18 (0.83)	0.03 (3.52)	0.77	0.59

Note: π^e_{t+1} denotes inflation expected by the private sector for period $t+1$, expressed in terms of the annualized rate of change in the GDP deflator; π_{t-1} is the lagged value of the annualized rate of change of the GDP deflator; and ulc_t is real unit labor cost. Numbers in parentheses are t-statistics, and the last column shows the p-values associated with a test of the model's overidentifying restrictions (Hansen's J-test).

the hybrid model for the full sample, including real-time data on contemporaneous as well as lagged inflation, with no significant changes in the results.

3.3 Evaluating the Robustness of the Results

To evaluate the robustness of our preferred specifications based on inflation forecasts, we first include in the model additional lags of the explanatory variables and test for their significance. It turns out that both lagged measures of real unit labor cost and lagged inflation are not statistically significant, with the exception of the third lag of inflation that appears significant in the specifications based on SPF forecasts.

We further explore the stability of the Phillips curve by considering the estimates obtained from two subsamples—one corresponding to the Chairmanships of Arthur Burns and G. William Miller (1968:Q1–1979:Q2) and the other to the Chairmanships of Paul Volcker and Alan Greenspan (1979:Q3–2000:Q4). The inflation experience was quite different in the two subperiods: during the 1970s, inflation was rising and volatile; then it dropped sharply during the 1980s and was low and relatively stable during the 1990s.

Estimates for the different subperiods (presented in table 5) confirm that the forward-looking behavior remains dominant irrespective of the sample period examined.[10] Most notably, the coefficient on the inflation forecast during the Volcker-Greenspan period declines to 0.78 (from 0.86 during the pre-Volcker period) for the Greenbook-based specification, and from 1.09 to 0.75 for the SPF-based specification, suggesting that the forward-looking behavior in price setting was even more relevant during the earlier period. Overall, the evidence from subperiod estimates supports the ability of the New Keynesian Phillips curve, in which inflation expectations are approximated by the inflation forecasts, to adequately describe inflation dynamics in the United States during the 1968–2000 period.

[10]The instrument set used includes two lags of inflation, the real unit labor cost, the output gap, and nominal-wage growth.

Table 5. Subsample Estimates of the Forecast-Based Specifications of the New Keynesian Phillips Curve

	π^e_{t+1}	π_{t-1}	ulc_t	$\bar{R}^2$	J-test
1968:Q4–1979:Q2 (Pre-Volcker Period) Greenbook-Based Specification	0.86 (4.16)	0.17 (1.08)	0.05 (4.17)	0.59	0.54
1979:Q3–2000:Q4 (Volcker-Greenspan Period) Greenbook-Based Specification	0.78 (3.91)	0.29 (1.77)	0.02 (2.35)	0.67	0.69
1968:Q4–1979:Q2 (Pre-Volcker Period) SPF-Based Specification	1.09 (4.88)	−0.001 (−0.006)	0.03 (2.13)	0.57	0.55
1979:Q3–2000:Q4 (Volcker-Greenspan Period) SPF-Based Specification	0.75 (5.33)	0.27 (1.96)	0.02 (3.30)	0.89	0.67

Note: π^e_{t+1} denotes inflation expected by the private sector for period $t+1$, expressed in terms of the annualized rate of change in the GDP deflator; π_{t-1} is the lagged value of the annualized rate of change of the GDP deflator; and ulc_t is real unit labor cost. Numbers in parentheses are t-statistics, and the last column shows the p-values associated with a test of the model's overidentifying restrictions (Hansen's J-test).

4. Interpreting the Empirical Results

According to the results presented in the previous section, when allowing for deviations from rationality in inflation expectations (by using observed inflation forecasts), the New Keynesian Phillips curve appears to provide an adequate description of inflation dynamics in the United States during the 1968–2006 period. This result contrasts with a large part of the empirical literature, which generally favors Phillips-curve specifications including lags of inflation.

Under the assumption of rational expectations, inflation expectations by themselves cannot explain the persistence of the inflation process. However, relatively small deviations from the assumption of rational expectations can change significantly this result (Angeloni et al. 2006). The significance of the forward-looking inflation term provides a strong case in favor of theories modeling expectations formation through limited/asymmetric information and information-processing constraints (Mankiw and Reis 2002; Woodford 2003; Adam 2004) or bounded rationality and learning (Evans and Honkapohja 2001; Sims 2003).

Consequently, the source of the observed inflation persistence may be due not to structural parameters stemming from the characteristics of the agents' price-setting behavior or institutional constraints (such as indexation) but rather to expectations-related factors such as expectations about future monetary policy movements or the private sector's gradual learning of monetary policy's inflation target (Erceg and Levin 2003), or uncertainty about the nature of inflationary shocks and their persistence (Ehrmann and Smets 2003).

Such sources of less-than-perfectly rational behavior of agents in their price-setting decisions and the persistence it implies can spuriously be reflected as significant lag dynamics in the hybrid New Keynesian Phillips curve. Thus, the insignificance of the lagged-inflation term when inflation forecasts are included in the specification suggests that inflation inertia is likely to stem from imperfectly rational behavior in a purely forward-looking price-setting process.

The significance of the less-than-perfectly rational forward-looking component in the New Keynesian Phillips curve bears an important policy implication: as inflation dynamics are likely to reflect the combined influence of several sources of less-than-perfectly rational behavior, the inflation-expectations management is a very difficult task, with potentially significant costs in terms of output volatility even in the absence of backward-looking price setters.

5. Conclusions

This paper examined the ability of the New Keynesian and hybrid Phillips curves to explain U.S. inflation dynamics if inflation forecasts obtained from the Federal Reserve's Greenbook or the Survey of Professional Forecasters are used as proxies of inflation expectations. The empirical evidence provides considerable support for the standard forward-looking New Keynesian Phillips curve insofar as deviations from rationality as reflected in inflation forecasts are taken into account in estimation. In particular, theoretically plausible coefficient estimates of expected inflation and real unit labor cost have been obtained. Overall, the empirical relevance of the hybrid specifications used in the literature to explain the persistence of inflation detected in the data appears to depend on the standard assumption of rational expectations usually made (reflected in the use of actual data on future inflation). Thus, lagged-inflation terms in the hybrid Phillips curve, intended to capture inflation inertia, are not significant when we consider less-than-perfectly rational forecast proxies of inflation expectations.

Appendix. Data Sources

All data series are quarterly, beginning in 1968:Q4 and ending in 2006:Q4 (with the exception of Greenbook inflation forecasts, which are available from 1968:Q4 through 2000:Q4). Data on gross domestic product, the GDP deflator, and nominal-wage growth are all from the Federal Reserve System's Database (FRED). Data on the GDP deflator forecasts were taken from the FOMC's Greenbook and the Survey of Professional Forecasters data sets available at the Federal Reserve Bank of Philadelphia. Real-time data on the GDP/GNP deflator were compiled on the basis of the relevant data set available at the Federal Reserve Bank of Philadelphia. Finally, data on unit labor cost in the nonfarm business sector were taken from the U.S. Bureau of Labor Statistics.

References

Adam, K. 2004. "Optimal Monetary Policy with Imperfect Common Knowledge." CEPR Discussion Paper No. 4594.

Adam, K., and M. Padula. 2003. "Inflation Dynamics and Subjective Expectations in the United States." ECB Working Paper No. 222.

Angeloni, I., L. Aucremanne, M. Ehrmann, J. Galí, A. Levin, and F. Smets. 2006. "New Evidence on Inflation Persistence and Price

Stickiness in the Euro Area: Implications for Macro Modeling." *Journal of the European Economic Association* 4 (2–3): 562–74.

Brayton, F., A. Levin, R. Tryon, and J. C. Williams. 1997. "The Evolution of Macro Models at the Federal Reserve Board." *Carnegie-Rochester Conference Series on Public Policy* 47: 43–81.

Calvo, G. A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.

Carroll, C. D. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics* 118 (1): 269–98.

Chadha, B., P. R. Masson, and G. Meredith. 1992. "Models of Inflation and the Costs of Disinflation." *IMF Staff Papers* 39 (2): 395–431.

Christiano, L. J., M. Eichenbaum, and C. L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45.

Cogley, T., and A. M. Sbordone. 2005. "A Search for a Structural Phillips Curve." Federal Reserve Bank of New York Staff Report No. 203.

Collard, F., and H. Dellas. 2005. "The New Keynesian Model with Imperfect Information and Learning." Forthcoming in *Journal of Money, Credit, and Banking.*

Ehrmann, M., and F. Smets. 2003. "Uncertain Potential Output: Implications for Monetary Policy." *Journal of Economic Dynamics and Control* 27 (9): 1611–38.

Erceg, C. J., and A. T. Levin. 2003. "Imperfect Credibility and Inflation Persistence." *Journal of Monetary Economics* 50 (4): 915–44.

Estrella, A., and J. C. Fuhrer. 2002. "Dynamic Inconsistencies: Counterfactual Implications of a Class of Rational-Expectations Models." *American Economic Review* 92 (4): 1013–28.

Evans, G. W., and S. Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton, NJ: Princeton University Press.

Fuhrer, J. C. 1997a. "Inflation/Output Variance Trade-Offs and Optimal Monetary Policy." *Journal of Money, Credit, and Banking* 29 (2): 214–34.

———. 1997b. "The (Un)Importance of Forward-Looking Behavior in Price Specifications." *Journal of Money, Credit, and Banking* 29 (3): 338–50.

Fuhrer, J., and G. Moore. 1995. “Inflation Persistence.” *Quarterly Journal of Economics* 110 (1): 127–59.

Galí, J., and M. Gertler. 1999. “Inflation Dynamics: A Structural Econometric Analysis.” *Journal of Monetary Economics* 44 (2): 195–222.

Galí, J., M. Gertler, and J. D. López-Salido. 2001. “European Inflation Dynamics.” *European Economic Review* 45 (7): 1237–70.

———. 2005. “Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve.” *Journal of Monetary Economics* 52 (6): 1107–18.

Lindé, J. 2005. “Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach.” *Journal of Monetary Economics* 52 (6): 1135–49.

Mankiw, N. G. 2001. “The Inexorable and Mysterious Tradeoff between Inflation and Unemployment.” *Economic Journal* 111 (471): C45–61.

Mankiw, N. G., and R. Reis. 2002. “Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve.” *Quarterly Journal of Economics* 117 (4): 1295–1328.

McCallum, B. 1997. “Comment on ‘An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy.’ ” In *NBER Macroeconomics Annual 1997*, 355–59.

Paloviita, M., and D. Mayes. 2005. “The Use of Real-Time Information in Phillips-Curve Relationships for the Euro Area.” *North American Journal of Economics and Finance* 16 (3): 415–34.

Roberts, J. M. 1995. “New Keynesian Economics and the Phillips Curve.” *Journal of Money, Credit, and Banking* 27 (4): 975–84.

———. 1997. “Is Inflation Sticky?” *Journal of Monetary Economics* 39 (2): 173–96.

———. 1998. “Inflation Expectations and the Transmission of Monetary Policy.” Finance and Economics Discussion Series Paper No. 43, Board of Governors of the Federal Reserve System.

———. 2005. “How Well Does the New Keynesian Sticky-Price Model Fit the Data?” *Contributions to Macroeconomics* 5 (1): 1206–42.

Romer, C. D., and D. H. Romer. 2000. “Federal Reserve Information and the Behavior of Interest Rates.” *American Economic Review* 90 (3): 429–57.

Rotemberg, J. J. 1982. "Sticky Prices in the United States." *Journal of Political Economy* 90 (6): 1187–1211.

Rotemberg, J. J., and M. Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." In *NBER Macroeconomics Annual*, ed. B. S. Bernanke and J. J. Rotemberg, 297–346. Cambridge, MA: MIT Press.

Rudd, J., and K. Whelan. 2005a. "Does Labor's Share Drive Inflation?" *Journal of Money, Credit, and Banking* 37 (2): 297–312.

———. 2005b. "New Tests of the New-Keynesian Phillips Curve." *Journal of Monetary Economics* 52 (6): 1167–81.

———. 2006. "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?" *American Economic Review* 96 (1): 303–20.

Sbordone, A. M. 2001. "An Optimizing Model of U.S. Wage and Price Dynamics." Departmental Working Paper No. 10, Rutgers University.

———. 2002. "Prices and Unit Labor Costs: A New Test of Price Stickiness." *Journal of Monetary Economics* 49 (2): 265–92.

———. 2005. "Do Expected Future Marginal Costs Drive Inflation Dynamics?" *Journal of Monetary Economics* 52 (6): 1183–97.

Sims, C. A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.

Smets, F., and R. Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1123–75.

———. 2005. "Comparing Shocks and Frictions in US and Euro Area Business Cycles: A Bayesian DSGE Approach." *Journal of Applied Econometrics* 20 (2): 161–83.

Swanson, E. T. 2004. "Assessing the Rationality of Federal Reserve 'Greenbook' Forecasts." Unpublished Manuscript, Board of Governors of the Federal Reserve System.

Taylor, J. B. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy* 88 (1): 1–23.

Woodford, M. 1996. "Control of the Public Debt: A Requirement for Price Stability?" NBER Working Paper No. 5684.

———. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ: Princeton University Press.

Optimal Monetary Policy in an Estimated DSGE Model of the Euro Area with Cross-Country Heterogeneity*

Eric Jondeau[a] and Jean-Guillaume Sahuc[b]
[a]University of Lausanne and Swiss Finance Institute
[b]Banque de France and Audencia School of Management

This paper investigates the implications of cross-country heterogeneity within the euro area for the design of optimal monetary policy. We build an optimization-based multicountry model (MCM) describing the euro area in which differences between structural parameters across countries are allowed. Using Bayesian techniques, we estimate the MCM and its area-wide counterpart (AWM). We then question which model is the most appropriate for monetary policy purposes. Several results emerge. First, using an AWM induces relatively large and significant welfare losses. Second, this is not the use of a rule based on aggregated variables that is costly in terms of welfare, but rather the use of a suboptimal forecasting model. Third, allowing for habit on consumption has important implications for the dynamics of models, but taking into account differences in price indexation has more drastic effects on welfare losses.

JEL Codes: C51, E52, F41.

*We thank S. Adjémian, J.-P. Benassy, M. Casares, P. Fève, M. Giannoni, P. Jacquinot, M. Juillard, J.-P. Laffargue, H. Le Bihan, A. Levin, J. Matheron, F. Mihoubi, E. Ortega, L. Patureau, the editor (F. Smets), and two anonymous referees for fruitful discussions and remarks. We have also benefited from comments by participants in seminars at the European Central Bank, workshops hosted by the Banque de France, CEPREMAP, Bank of Canada, UQAM, and the T2M conference. Financial support from the National Center of Competence in Research "Financial Valuation and Risk Management" is gratefully acknowledged. The National Centers of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation. The views expressed in this paper are the responsibility of the authors and do not necessarily reflect those of the Banque de France. The usual disclaimer applies.
Corresponding author: Sahuc: Banque de France, Economic and Financial Research Division, 31 rue Croix des Petits Champs, F-75049 Paris, France. E-mail: jean-guillaume.sahuc@banque-france.fr.

1. Introduction

The Maastricht Treaty (article 105) states that the primary objective of the European Central Bank (ECB) is to maintain price stability within the European Monetary Union. Although the ECB may use a battery of economic indicators—including country-specific ones—in order to fulfill this primary objective, it makes decisions on the basis of aggregate developments, while national idiosyncrasies are left to the care of national governments. The consequences of such a constraint on the monetary policy of the euro area are obviously related to the extent and the nature of heterogeneity of countries within the area. Since the decisions have to be made on the basis of aggregate developments only, it may be argued that, since objectives are defined in terms of aggregate variables, an area-wide model (AWM) would be sufficient for capturing most characteristics of the euro-area economy. On the other hand, a multicountry model (MCM) may help capture the heterogeneity of countries and therefore offer valuable information about the state of the euro-area economy. Consequently, it would help define a more appropriate monetary policy rule. As a result, it is not clear a priori what type of forecasting model (multicountry or area-wide) should be used for implementing an optimal monetary policy.

To investigate the role devoted to country-specific information in the decision process regarding the Eurosystem, the standard approach to policy evaluation can be followed (see the contributions in Taylor 1999): the optimal policy rule is determined so as to minimize the expected value of an intertemporal loss function, under the constraint provided by a low-dimensional multicountry model of the euro area. Assuming that the monetary authority is exclusively interested in area-wide objectives, it is possible to compare the performance of two optimal reaction functions based on an MCM and an AWM, respectively. Such a comparison has already been performed in a few sets of contributions, revealing that the loss associated with the neglect of country-specific information might be large.[1] However, in these studies, the underlying macroeconomic

[1]The literature includes De Grauwe (2000), De Grauwe and Piskorki (2001), Aksoy, De Grauwe, and Dewachter (2002), Angelini et al. (2002), and Monteforte and Siviero (2003), among others.

models are not designed in an optimization-based framework. Consequently, the optimal monetary policy deduced from such models is subject to the Lucas critique, since it is based on reduced-form, rather than structural, parameters. This is a serious limitation when the welfare resulting from an optimal policy rule has to be evaluated.

The objective of this paper is to reassess and generalize the preceding results in investigating how heterogeneity of agents across euro-area countries is likely to affect the optimal monetary policy using an *optimization-based framework.* More precisely, we measure the welfare cost of using an AWM instead of an MCM to evaluate the optimal monetary policy. The basic idea is that the MCM is designed to capture the cross-country heterogeneity and thus to describe more accurately the way monetary policy affects the economy. Consequently, a welfare-maximizing central bank may be able to implement a more efficient monetary policy, even if the policy rule is assumed to be based on aggregate variables only. An obvious shortcoming of the MCM is that its estimation is much more demanding, since it requires modeling the joint dynamics of several economies as well as the international transmission mechanisms. In addition, the MCM is likely to induce a large amount of country-specific uncertainty, while an AWM may average these errors. Conversely, the estimation of an AWM is likely to induce an aggregation bias if structural parameters actually differ across countries. Such a bias has already been highlighted in the context of the Phillips curve (see Demertzis and Hughes Hallett 1998, Benigno and López-Salido 2006, or Altissimo, Mojon, and Zaffaroni 2007), but its consequences on the optimal monetary policy in a complete model have not been explored up to now.

In the class of dynamic stochastic general equilibrium (DSGE) models, several alternative models provide plausible frameworks to take heterogeneity into account. Our approach consists of designing a low-dimensional model that fits the data fairly well. The first reason for doing so is that in a simple structural model, the underlying mechanisms are easy to understand and the subsequent welfare analysis is transparent. Second, to measure the effect of cross-country heterogeneity on monetary policy, we need to estimate the structural parameters of all the individual economies. As a consequence, a low-dimensional model is required to obtain consistent estimates. Estimation is based on Bayesian techniques, which seem to be the best way to determine an empirically realistic specification of the heterogeneity.

The model we consider resorts to the "new open-economy macroeconomics" literature, initiated by Obstfeld and Rogoff (1995). In addition to significant frictions in the form of nominal rigidities, other mechanisms are introduced: (i) cross-country differences in the structural parameters are allowed, since we are primarily interested in the effect of such heterogeneity on the design of the optimal monetary policy, (ii) perfect risk sharing and a home bias in preferences are incorporated in the model to deal with exchange rate indeterminacy, and (iii) cross-country correlations between shocks are introduced to capture co-movement in the joint dynamics of national conditions.

Following the strategy described above, we first estimate two models, mimicking the way the ECB forecasts macroeconomic developments within the Eurosystem. In the first one, we adopt an open-economy framework and model the joint dynamics of the data for the major countries in the euro area (Germany, France, and Italy). In the second one, we model the dynamics of area-wide macroeconomic data. Our empirical evidence suggests that there exists some significant heterogeneity within the euro area, even among core countries. First, we obtain some significant differences between estimates of the structural parameters at the euro-area level and at the country level, suggesting an aggregation bias. But more importantly, we find that the main source of heterogeneity is the weak correlation between shocks across countries.

Then, we investigate how cross-country heterogeneity affects the design of optimal monetary policy within the euro area. We consider two alternative modeling approaches. In both of them, the central bank defines its preferences and its loss function at the area-wide level, and the reaction function is designed in terms of aggregate variables only. In the first approach the MCM, estimated using country-specific data, is used for computing the loss function, while in the second approach the AWM, estimated using aggregated data, is used. Then, we evaluate the optimal monetary policy that maximizes the aggregate welfare, both under the AWM and the MCM, and we measure the welfare cost of using the AWM (suboptimal) forecasting model. We obtain that the welfare cost is quite significant, both statistically and economically. It appears to be related mainly to nominal rigidities rather than to real rigidities.

The remainder of the paper is organized as follows. In section 2, we describe the theoretical MCM. In section 3, we present the data and the estimates of the MCM and AWM. In section 4, we determine the optimal monetary policy under the two forecasting models and evaluate the welfare implications of using the (suboptimal) AWM model. Section 5 summarizes our main findings and concludes.

2. Structure of the Stylized Multicountry Model

The euro area is modeled as the aggregate of several economies. For each country, we formulate a small-size open-economy model, which is inspired by recent theoretical models derived from the "new open-economy macroeconomics" literature.[2] The model is enriched in several dimensions to offer a comprehensive framework that encompasses other previous contributions. In terms of dynamics, first, key modifications are the explicit incorporation of habit formation in the households' preferences and partial indexation in a price-setting framework à la Calvo (1983). These assumptions provide us with microfounded "hybrid" versions of the IS and Phillips curves. Second, contrary to most recent studies on DSGE models, we do not assume that preferences and technologies are the same across countries, since we are interested in measuring the effect of heterogeneity on the optimal monetary policy of the area. In addition, domestic and foreign shocks are allowed to be imperfectly correlated. Third, we resort to the perfect-risk-sharing assumption. Although this assumption is admittedly heroic in empirical work, it avoids assuming nonrational expectations of exchange rate, which has been shown to be an alternative way of dealing with nonstationarity. Finally, households are assumed to have a taste bias toward goods produced in their home country. Since preferences differ across countries, the price of consumption bundles will differ when expressed in

[2] See, among others, Monacelli (2001), Clarida, Galí, and Gertler (2002), Smets and Wouters (2002), Benigno and Benigno (2003), Devereux and Engel (2003), P. Benigno (2004), Corsetti and Pesenti (2005), and Galí and Monacelli (2005).

a common currency. The real exchange rate thus deviates from purchasing power parity (PPP).[3] This assumption is crucial, because it allows the perfect-risk-sharing equation to determine uniquely the dynamics of the terms of trade.

In order to lighten the notations, we assume that there are two countries in the euro area, denoted H (home) and F (foreign). Since commercial links are much stronger between countries within the area than with countries outside the area, we neglect trade with the rest of the world. The population of the euro area is a continuum of agents on the interval $[0, 1]$. The population of country H belongs to $[0, n)$, while the foreign population belongs to $[n, 1]$. Therefore, n is the relative measure of the home-country size in the euro area. An agent in the home country is indexed $h \in [0, n)$, while a foreign agent is indexed $f \in [n, 1]$. Variables in the home country are denoted X_t, while foreign variables are denoted X_t^*. The home economy produces a continuum of differentiated goods indexed on the interval $[0, n)$. Foreign goods (or, equivalently, goods produced in the rest of the area) are indexed on the interval $[n, 1]$. All goods are tradable.

2.1 Households

The home economy is populated by infinitively living households consuming Dixit-Stiglitz aggregates of domestic and imported goods. A home household h owns a firm producing goods h and receives dividends from it. We assume that households in a given country have the same preferences and endowments. Although there may be idiosyncratic shocks among households, we assume that households have access to complete markets for state-contingent claims, so that there is no heterogeneity among agents in a given country. Consequently, all households in the same country behave in the same manner, and then we consider the optimization problem of a representative household. The representative household in country H maximizes the following expected sequence of present and future utility flows

[3] An earlier contribution that introduced home-country bias in preferences is due to Warnock (2000).

that depends positively on consumption (C_t) and negatively on labor (hours worked, L_t):[4]

$$\mathcal{U}_t = \mathbb{E}_t \sum_{k=0}^{\infty} \beta^k \varepsilon_{p,t+k} \left[\frac{1}{1-\sigma} (C_{t+k} - \gamma \mathcal{H}_{t+k})^{1-\sigma} - \frac{1}{1+\varphi} (L_{t+k})^{1+\varphi} \right], \tag{1}$$

subject to a series of real-period budget constraints:

$$C_t + \frac{B_{t+1}}{(1+i_t)P_t} + \frac{T_t}{P_t} = \frac{W_t L_t}{P_t} + \frac{B_t}{P_t} + \frac{\Pi_t}{P_t}, \tag{2}$$

where $\mathbb{E}_t$ denotes the expectation operator conditional on the information set at time t; β is the intertemporal discount factor, with $0 < \beta < 1$; σ is the inverse of the intertemporal elasticity of substitution of consumption; and φ is the inverse of the elasticity of labor disutility with respect to hours worked. In addition, W_t is the nominal wage income, Π_t is the dividend received from home firms, T_t represents nominal lump-sum government transfers, and i_t is the nominal interest rate. $\varepsilon_{p,t}$ denotes a country-specific preference shock that affects the intertemporal substitution of all households in the same manner in the home economy. Preferences display external habit formation as in Abel (1990). The habit stock is supposed to equal the level of aggregate consumption in the previous period ($\mathcal{H}_t = C_{t-1}$), and γ represents the habit-persistence parameter, measuring the effect of past consumption on current utility ($0 \leq \gamma < 1$). Finally, we assume complete markets for state-contingent claims. Consequently, households can transfer wealth to the next period by holding B_{t+1} units of the one-period nominal bond denominated in the domestic currency.[5]

---

[4] We abstract from money in this model since the central bank adjusts money supply to satisfy money demand with a simple feedback rule.

[5] More precisely, at date t, home households hold $B(s^{t+1}) = B_{t+1}$ units of the one-period bond denominated in home currency that pay 1 at date $t+1$ if state s_{t+1} occurs and 0 otherwise, where $s^t = (s_0, \cdots, s_t)$ denotes the story of events up to date t. Foreign households hold $B_t^*(s^{t+1}) = B_{t+1}^*$ units of such bond. The price of this bond in home currency is denoted by $\Phi(s^t, s^{t+1}) = \Phi_{t,t+1}$. The price at date t of the portfolio held by home households is thus given by $E_t[\Phi_{t,t+1} B_{t+1}]$. We define the one-period interest rate as $1 + i_t = 1/E_t[\Phi_{t,t+1}]$.

The maximization problem of the home household consists in maximizing equation (1) subject to constraint (2), yielding the optimal profile of consumption—holdings of domestic bond and labor supply. The first-order conditions imply[6]

$$\mathcal{U}_{C,t} = \varepsilon_{p,t}(C_t - \gamma\mathcal{H}_t)^{-\sigma}, \tag{3}$$

$$(1+i_t)^{-1} = \beta\mathbb{E}_t\left[\frac{\mathcal{U}_{C,t+1}}{\mathcal{U}_{C,t}}\frac{P_t}{P_{t+1}}\right], \tag{4}$$

$$\frac{\mathcal{U}_{L,t}}{\mathcal{U}_{C,t}} = \frac{W_t}{P_t}, \tag{5}$$

where $\mathcal{U}_{X,t}$ denotes the derivative of utility $\mathcal{U}$ with respect to variable X at period t. Equation (3) defines the marginal utility of consumption. Equation (4) is the usual Euler equation for intertemporal consumption flows. It establishes that the ratio of marginal utility of future and current consumption is equal to the inverse of the real interest rate. Equation (5) is the condition for the optimal consumption-leisure arbitrage, implying that the marginal rate of substitution between consumption and labor is equated to the real wage.

2.1.1 Composite Consumption Index

The aggregate consumption index for home households and the corresponding consumption index for foreign households are defined by[7]

$$C_t = \frac{(C_{H,t})^{\omega}(C_{F,t})^{1-\omega}}{\omega^{\omega}(1-\omega)^{1-\omega}} \quad \text{and} \quad C_t^* = \frac{\left(C_{H,t}^*\right)^{\omega^*}\left(C_{F,t}^*\right)^{1-\omega^*}}{(\omega^*)^{\omega^*}(1-\omega^*)^{1-\omega^*}}, \tag{6}$$

where ω and ω^* denote the share of home goods in the consumption of home and foreign households, respectively. $C_{H,t}$ (resp. $C_{F,t}$)

Note that, since bonds are state-contingent, including bonds denominated in foreign currency would be redundant. For more details, see Chari, Kehoe, and McGrattan (2002).

[6]We abstract here from the optimal intratemporal allocations between domestic and foreign goods.

[7]As shown by Corsetti and Pesenti (2005), the Cobb-Douglas consumption index is a sufficient condition for the trade to be invariably balanced.

is the subindex of consumption of imperfectly substitutable home (resp. foreign) goods, which is in turn given by the following CES aggregators:

$$C_{H,t} = \left[\left(\frac{1}{n}\right)^{1/\theta} \int_0^n C_t(h)^{\frac{\theta-1}{\theta}} \mathrm{d}h \right]^{\frac{\theta}{\theta-1}} \quad \text{and}$$

$$C_{F,t} = \left[\left(\frac{1}{1-n}\right)^{1/\theta} \int_n^1 C_t(f)^{\frac{\theta-1}{\theta}} \mathrm{d}f \right]^{\frac{\theta}{\theta-1}}, \qquad (7)$$

where $C_t(h)$ (resp. $C_t(f)$) is consumption of the generic good h (resp. f) produced in country H (resp. F). Parameter θ denotes the elasticity of substitution across goods produced within a given country. The corresponding consumption price indexes (CPIs) are given by

$$P_t = (P_{H,t})^{\omega} (P_{F,t})^{1-\omega} \quad \text{and} \quad P_t^* = \left(P_{H,t}^*\right)^{\omega^*} \left(P_{F,t}^*\right)^{1-\omega^*}.$$

Here, $P_{H,t}$ (resp. $P_{F,t}$) is the price subindex for home-produced (resp. foreign-produced) goods expressed in the home currency, defined as

$$P_{H,t} = \left[\frac{1}{n} \int_0^n P_{H,t}(h)^{1-\theta} \mathrm{d}h \right]^{\frac{1}{1-\theta}} \quad \text{and}$$

$$P_{F,t} = \left[\frac{1}{1-n} \int_n^1 P_{F,t}(f)^{1-\theta} \mathrm{d}f \right]^{\frac{1}{1-\theta}},$$

where $P_{H,t}(h)$ (resp. $P_{F,t}(f)$) is the price in units of country H of a generic good h (resp. f) produced in country H (resp. F).

We also assume that prices are set in the producer currency and that the law of one price holds. We then have $P_{H,t}(h) = P_{H,t}^*(h)S_t$ and $P_{F,t}(f) = P_{F,t}^*(f)S_t$, where S_t is the nominal exchange rate expressed as units of domestic currency needed for one unit of foreign currency.[8] Since we assume the same elasticity of substitution

[8] Although it has been investigated in a number of recent papers, we do not consider here the presence of imperfect exchange rate pass-through. A reason is that it is not likely to be an important feature across countries within the euro area. In addition, this feature is obviously irrelevant from the euro-area point of view.

among goods in a given country, we also have $P_{H,t} = P^*_{H,t}S_t$ and $P_{F,t} = P^*_{F,t}S_t$. Yet, from the definition of the CPI, we obtain that

$$P_t = P^*_t S_t \left(\frac{P_{H,t}}{P_{F,t}}\right)^{\omega-\omega^*}.$$

Therefore, if we assume that there exists a home bias in preferences ($\omega \neq \omega^*$), PPP does not necessarily hold; i.e., $P_t \neq P^*_t S_t$. We expect $\omega > \omega^*$, so that home households put a higher weight on home goods than foreign households.

2.1.2 International Risk Sharing

Under the assumption of complete markets, domestic and foreign households trade in state-contingent claims denominated in the home currency. This implies the following perfect-risk-sharing condition (Chari, Kehoe, and McGrattan 2002):

$$Q_t = \kappa \frac{\mathcal{U}^*_{C^*,t}}{\mathcal{U}_{C,t}}, \tag{8}$$

where the real exchange rate, defined as $Q_t \equiv S_t P^*_t / P_t$, is proportional to the ratio of the marginal utility of consumption between the two countries.[9] The assumption of international market completeness ensures that, in our model, the real exchange rate and consumption are stationary variables (see also G. Benigno 2004).

Since the real exchange rate deviates from PPP because of home bias in preferences, we also have

$$Q_t = \left(\frac{S_t P^*_{H,t}}{P_{H,t}}\right)^{\omega^*} \left(\frac{S_t P^*_{F,t}}{P_{F,t}}\right)^{1-\omega^*} \left(\frac{P_{F,t}}{P_{H,t}}\right)^{\omega-\omega^*} = (\mathcal{T}_t)^{\omega-\omega^*}, \tag{9}$$

where $\mathcal{T}_t$ represents the home terms of trade, i.e., the relative price between foreign and home bundles of goods as perceived by the home resident. It is defined as[10]

$$\mathcal{T}_t = \frac{P_{F,t}}{P_{H,t}} = \frac{S_t P^*_{F,t}}{P_{H,t}}. \tag{10}$$

[9] $\kappa = [S_0 P^*_0 \mathcal{U}_{C,0}]/[P_0 \mathcal{U}^*_{C^*,0}]$ is a constant that depicts the initial condition.

[10] The foreign terms of trade are simply given by $\mathcal{T}^*_t = P^*_{H,t}/P^*_{F,t} = 1/\mathcal{T}_t$, because the law of one price holds.

Using equations (3), (8), and (9), this definition implies the following:

$$(\mathcal{T}_t)^{\omega-\omega^*} = \kappa \frac{\varepsilon^*_{p,t}(C_t - \gamma C_{t-1})^{\sigma}}{\varepsilon_{p,t}\left(C^*_t - \gamma^* C^*_{t-1}\right)^{\sigma^*}}. \tag{11}$$

Note that, when there is no home bias in preferences ($\omega = \omega^*$), the perfect-risk-sharing assumption does not allow us to determine the terms of trade anymore.

Combining Euler equation (4) with the perfect-risk-sharing equation (8), we obtain the following dynamics for the real exchange rate and the terms of trade:

$$\mathbb{E}_t\left[\frac{Q_{t+1}}{Q_t}\right] = \mathbb{E}_t\left[\frac{\mathcal{U}^*_C(C^*_{t+1})\mathcal{U}_C(C_t)}{\mathcal{U}^*_C(C^*_t)\mathcal{U}_C(C_{t+1})}\frac{P^*_t P_{t+1}}{P_t P^*_{t+1}}\right] = \frac{1+i_t}{1+i^*_t} \tag{12}$$

$$\mathbb{E}_t\left[\frac{\mathcal{T}_{t+1}}{\mathcal{T}_t}\right] = \mathbb{E}_t\left[\frac{P^*_{F,t+1}P_{H,t}}{P_{H,t+1}P^*_{F,t}}\frac{1+i_t}{1+i^*_t}\right]. \tag{13}$$

Equation (12) is the uncovered interest rate parity (UIP) condition, which states that the expected change in the exchange rate is exactly compensated by the real interest rate differential. It is worth emphasizing that the UIP condition is not an additional implication in the model, but rather a redundant relation.

2.2 Firms

There is a continuum of infinitely living and monopolistically competitive firms indexed by h on the interval $[0, n)$ for the home country and by f on the interval $[n, 1]$ for the foreign country. These firms produce differentiated goods that are bundled into homogeneous home and foreign goods by a constant returns to scale of the Dixit-Stiglitz form:

$$Y_t = \left[\left(\frac{1}{n}\right)^{1/\theta}\int_0^n Y_t(h)^{\frac{\theta-1}{\theta}}\mathrm{d}h\right]^{\frac{\theta}{\theta-1}} \quad \text{and}$$

$$Y^*_t = \left[\left(\frac{1}{1-n}\right)^{1/\theta}\int_n^1 Y^*_t(f)^{\frac{\theta-1}{\theta}}\mathrm{d}f\right]^{\frac{\theta}{\theta-1}}.$$

The production technology of the representative home firm h combines labor as primary input and a country-specific productivity shock:

$$Y_t(h) = \varepsilon_{a,t} L_t(h). \tag{14}$$

Output is normalized by population size, so that it is expressed in per capita terms. We thus deduce that total home labor demand is given by

$$L_t = \int_0^n L_t(h)\mathrm{d}h = \frac{\Delta_t Y_t}{\varepsilon_{a,t}}, \tag{15}$$

where $\Delta_t = \int_0^n \frac{Y_t(h)}{Y_t}\mathrm{d}h$ represents the dispersion of production across firms in the home economy.

Since input markets are perfectly competitive and country specific, the standard static first-order condition for cost minimization implies that all domestic firms have identical real marginal cost, MC_t, given by

$$MC_t = \frac{1}{(1+\vartheta)} \frac{W_t}{P_{H,t}\varepsilon_{a,t}}, \tag{16}$$

where ϑ is a subsidy for output that offsets the effect of imperfect competition in goods markets on the steady-state level of output $(0 \leq \vartheta < 1)$.

A firm's price-setting decision is modeled through a modified version of Calvo's (1983) staggering mechanism. In addition to the baseline mechanism, we allow for the possibility that firms that do not optimally set their prices may nonetheless adjust them to keep up with the previous-period increase in the general price level (see Sbordone 2003 and Christiano, Eichenbaum, and Evans 2005 for details concerning this assumption). In each period, a firm faces a constant probability, $1 - \alpha$, of being able to reoptimize its price and chooses the new price $\tilde{P}_{H,t}(h)$ that maximizes the expected discounted sum of profits

$$\mathbb{E}_t \sum_{k=0}^{\infty} \alpha^k \Upsilon_{t,t+k} \left[\frac{\tilde{P}_{H,t}(h)\Psi^H_{t,t+k}}{P_{H,t+k}} - MC_{t+k} \right] Y_{t+k}(h), \tag{17}$$

subject to the sequence of demand equations

$$Y_{t+k}(h) = \left(\frac{\tilde{P}_{H,t}(h)\Psi^H_{t,t+k}}{P_{H,t+k}} \right)^{-\theta} Y_{t+k}, \tag{18}$$

where $\Upsilon_{t,t+k} = \beta^k \mathcal{U}_C(C_{t+k})/\mathcal{U}_C(C_t)$ is the discount factor between time t and $t+k$, and

$$\Psi^H_{t,t+k} = \begin{cases} \prod_{\nu=0}^{k-1} (\bar{\pi}_H)^{1-\xi} (\pi_{H,t+v})^{\xi} & k > 0 \\ 1 & k = 0, \end{cases} \tag{19}$$

where $\bar{\pi}_H$ is the domestic trend inflation and the coefficient $\xi \in [0,1]$ indicates the degree of indexation to past prices during the periods in which the firm is not allowed to reoptimize. $\Psi^H_{t,t+k}$ is a correcting term that accounts for the fact that if the firm h does not reoptimize its price, it updates its price according to the rule

$$P_{H,t}(h) = (\bar{\pi}_H)^{1-\xi} (\pi_{H,t-1})^{\xi} P_{H,t-1}(h). \tag{20}$$

Consequently, the first-order condition associated with the profit maximization implies that firms set their price equal to the discounted stream of expected future real marginal costs:

$$\mathbb{E}_t \sum_{k=0}^{\infty} \alpha^k \Upsilon_{t,t+k} \left[(\bar{\pi}_H)^{(1-\xi)k} \left(\frac{P_{H,t+k-1}}{P_{H,t-1}} \right)^{\xi} \frac{\tilde{P}_{H,t}(h)}{P_{H,t+k}} - \frac{\theta}{\theta - 1} MC_{t+k} \right] \times Y_{t+k}(h) = 0. \tag{21}$$

If flexible prices is assumed ($\alpha = 0$), this expression gives the optimal relative price $\tilde{P}_{H,t}(h)/P_{H,t} = \mu MC_t$, where $\mu \equiv \theta/(\theta - 1)$ is the optimal markup in a flexible-price economy. As there are no firm-specific shocks in this economy, all firms that are allowed to reoptimize their price at date t select the same optimal price $\tilde{P}_{H,t}(h) = \tilde{P}_{H,t}$, $\forall h$.

Staggered price setting under partial indexation implies the following expression for the evolution of the domestic price index:

$$P_{H,t} = \left[\alpha ((\bar{\pi}_H)^{1-\xi} (\pi_{H,t-1})^{\xi} P_{H,t-1})^{1-\theta} + (1-\alpha)(\tilde{P}_{H,t})^{1-\theta} \right]^{\frac{1}{1-\theta}}. \tag{22}$$

The price-setting problem solved by firms in the foreign country is similar and leads to an optimal rule analogous to equation (21). However, we allow foreign structural parameters (α^*, ξ^*) and country-specific shocks $(\varepsilon^*_{a,t})$ to differ from their home-country counterparts.

2.3 Market-Clearing Conditions

Demands for goods are given by the subindex of consumption (7); the allocation of demand across each of the goods produced within a given country for consumers H, F is given by

$$C_t(h) = \frac{1}{n}\left(\frac{P_{H,t}(h)}{P_{H,t}}\right)^{-\theta} C_{H,t} \quad \text{and} \quad C^*_t(h) = \frac{1}{n}\left(\frac{P^*_{H,t}(h)}{P^*_{H,t}}\right)^{-\theta} C^*_{H,t},$$

$$C_t(f) = \frac{1}{1-n}\left(\frac{P_{F,t}(f)}{P_{F,t}}\right)^{-\theta} C_{F,t} \quad \text{and}$$

$$C^*_t(f) = \frac{1}{1-n}\left(\frac{P^*_{F,t}(f)}{P^*_{F,t}}\right)^{-\theta} C^*_{F,t}.$$

The consumption aggregator (6) implies that home and foreign demands for composite home and foreign goods are given by

$$C_{H,t} = \omega\left(\frac{P_t}{P_{H,t}}\right) C_t \quad \text{and} \quad C^*_{H,t} = \omega^*\left(\frac{P^*_t}{P^*_{H,t}}\right) C^*_t,$$

$$C_{F,t} = (1-\omega)\left(\frac{P_t}{P_{F,t}}\right) C_t \quad \text{and} \quad C^*_{F,t} = (1-\omega^*)\left(\frac{P^*_t}{P^*_{F,t}}\right) C^*_t.$$

Then, goods market clearing in the home and foreign countries implies

$$\begin{aligned} Y_t(h) &= nC_t(h) + (1-n)C^*_t(h) \\ &= \left(\frac{P_{H,t}(h)}{P_{H,t}}\right)^{-\theta}\left(\frac{P_t}{P_{H,t}}\right)\left(\omega C_t + \mathcal{T}_t^{\omega-\omega^*}\frac{1-n}{n}\omega^* C^*_t\right) \end{aligned}$$

and

$$\begin{aligned}Y_t^*(f) &= nC_t(f) + (1-n)C_t^*(f)\\ &= \left(\frac{P_{F,t}(f)}{P_{F,t}}\right)^{-\theta}\left(\frac{P_t}{P_{F,t}}\right)\left(\frac{n}{1-n}(1-\omega)C_t + (1-\omega^*)\mathcal{T}_t^{\omega-\omega^*}C_t^*\right),\end{aligned}$$

so that aggregate outputs in home and foreign goods are

$$Y_t = \omega(\mathcal{T}_t)^{1-\omega}C_t + \frac{1-n}{n}\omega^*(\mathcal{T}_t)^{1-\omega^*}C_t^* \tag{23}$$

and

$$Y_t^* = (1-\omega)(\mathcal{T}_t)^{-\omega}\frac{n}{1-n}C_t + (1-\omega^*)(\mathcal{T}_t)^{-\omega^*}C_t^*. \tag{24}$$

Together with equation (10), these relations show that aggregate output only depends on home and foreign consumptions and preference shocks.

2.4 Log-Linear Equilibrium

In order to estimate the model, we log-linearize it around the steady state. We also close the model by specifying a fairly simple monetary policy rule for each country, in which the short-term nominal interest rate responds to lagged interest rate as well as to deviations of inflation to its steady-state value and deviations of domestic aggregate output to its flexible-price equilibrium (or natural) value. This specification includes an additional monetary policy shock. Notice that, since the historical policy rule has not been necessarily optimal, the parameters of the reaction function cannot be viewed as structural ones. Consequently, we adopt for the moment a widely accepted specification, in order to estimate structural parameters reflecting the behavior of private agents.[11] Finally, we need to specify the dynamics of the shocks. We assume that the preference, productivity, and monetary policy shocks follow AR(1) processes,

[11]See Dieppe, Küster, and McAdam (2005) for a comparison of several policy rules using an AWM.

with autoregressive parameters denoted ρ_p, ρ_a, and ρ_i and standard deviations denoted σ_p, σ_a, and σ_i, respectively.

The resulting system, expressed in percentage deviation around the steady state, is presented in appendix 1. Details of the construction of the flexible-price output are provided in appendix 2. The determination of the optimal monetary policy consistent with our structural model is performed in section 4.

In the case of an area with more than two countries, the broad structure of the model remains essentially unchanged. The major change is that, in an N-country model, international transmission mechanisms pass through $(N - 1)$ independent terms of trade. Consequently, since the Phillips curve depends on the terms of trade through movements in real marginal cost, inflation dynamics is affected by demand conditions in all countries. Moreover, domestic consumption is affected by the average of real interest rates prevailing in all countries of the area.

3. Estimation

We now concentrate on the two models that we will use to evaluate the optimal monetary policy rules: (i) an MCM that incorporates information on individual countries, allowing model parameters to differ from one country to another, and (ii) an AWM that implicitly assumes that the heterogeneity of behaviors and the asymmetry of shocks across countries can be neglected in the design of monetary policy.[12]

Models used for policy analysis must naturally make quantitative predictions that can match observed regularities in the data. In this way, our small DSGE models search to match, in broad terms, the relationship among consumption (or output), inflation, and interest rates. But we do not explore the details of our model's fit to the data in this research. A technical appendix (available from the authors upon request) gives all the details concerning the empirical performance of the models and shows that both the MCM and the AWM

[12] For the AWM, we resort to the closed-economy version of the model described above, estimated over aggregated data of the euro area.

are able to reproduce most dynamics of the data, although the data do not support all the restrictions imposed by the DSGE.[13]

It is clear that introducing additional mechanisms would enrich the model and substantially improve its fit. Such an extension would be crucial especially if the model was to be used for implementing policy applications (alternative policy scenarios, forecasting, etc.). In our context, we emphasize the ability of our models to provide some evidence on the consequences of heterogeneity for optimal monetary policy. As we will show in the next section, even within our simple framework, we obtain that the use of an AWM induces relatively large and significant welfare losses. The omitted heterogeneity is actually incorporated in shocks that do not play a great role in the welfare measure (only the preference shocks enter the welfare measure). Since we obtain a large gap between the two welfare measures even when not all the sources of heterogeneity are taken into account, adding additional sources would result in even higher welfare losses. In section 4.4, we discuss some alternative modeling specifications.

Estimation is performed using Bayesian techniques.[14] The dynamical systems are cast in a state-space representation for the set of observable variables. The Kalman filter is then used to evaluate the likelihood of the observed variables and form the posterior distribution of the structural parameters by combining the likelihood function with a joint density characterizing some prior beliefs. Given the specification of the model, the posterior distribution cannot be recovered analytically but can be evaluated numerically, using a Markov-Chain Monte Carlo (MCMC) sampling approach. More specifically, we rely on the Metropolis-Hastings (MH) algorithm to obtain random draws from the posterior distribution of the parameters.[15]

[13] In the technical appendix, we first compare the posterior distributions of the DSGE and VAR models (see Geweke 1999); then we compare the DSGE-based cross-covariances (or autocorrelations) and/or impulse-response functions with those obtained from a VAR model.

[14] The Bayesian strategy has been proposed by, among others, Schorfheide (2003), Smets and Wouters (2003), and Fernandez-Villaverde and Rubio-Ramirez (2004).

[15] We simulate a sample of 200,000 random draws. The first 50,000 observations are discarded to eliminate any dependence on the initial values. The mode and the Hessian of the posterior distribution evaluated at the mode are used to initialize the MH algorithm.

3.1 Data

The data are drawn from OECD's Business Sector Database for individual countries. The sample period runs from 1970:Q1 to 1998:Q4 at a quarterly frequency. We suppose that the euro area is represented by the three largest countries of the area (Germany, France, and Italy), which cover some 70 percent of the area-wide GDP. The MCM is then estimated for these three countries, while the AWM is estimated on the weighted average of series pertaining to the three countries under study.

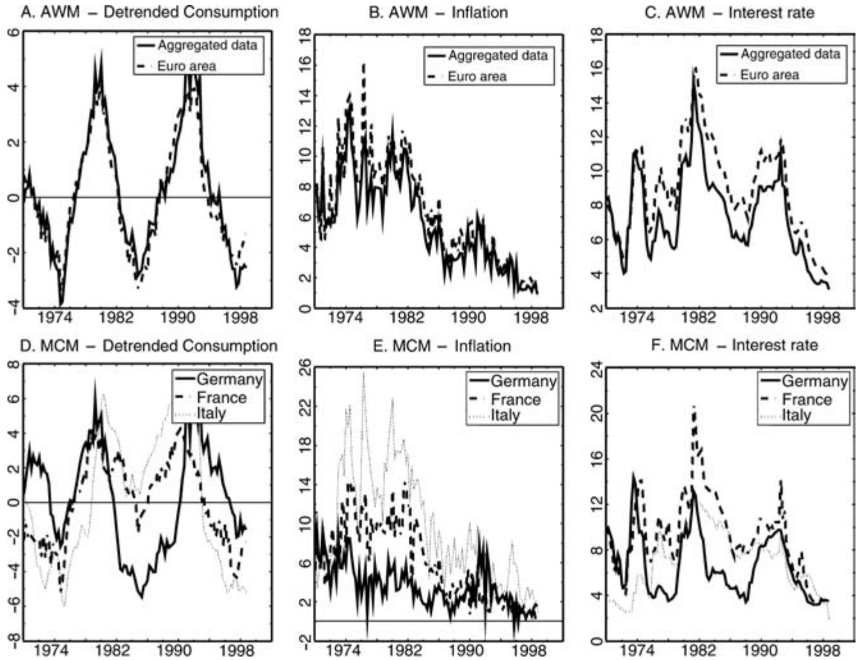
The estimation of the model is based ultimately on three key macroeconomic variables for each country: real consumption, the inflation rate, and the nominal short-term interest rate. Neither the terms of trade nor the real marginal cost are necessary for the estimation of the model, since they are defined as exact functions of the other macroeconomic variables. Consumption is defined as real consumption expenditures, linearly detrended. We measure inflation as the annualized quarterly percent change in the implicit GDP deflator. The interest rate is the three-month money-market rate. Figure 1 displays the historical path of the various series under consideration for each country or area. In the case of the euro area, we also plot the series extracted from the AWM database provided by Fagan, Henry, and Mestre (2005). We first notice that the two data sets for the euro area look very similar. We also observe a downward trend in inflation and interest rate, which mainly corresponds to the convergence process of economic conditions within the euro area. The structural model presented above is clearly not designed to capture such an empirical feature. Therefore, inflation and the nominal interest rate are detrended by the same quadratic trend in inflation.[16]

3.2 Prior Distribution and Calibrated Parameters

In the definition of the prior distributions, we followed the now standard procedure, assuming a beta distribution for parameters bounded between zero and one, a uniform distribution for standard

[16]We also examined if the estimates are modified when consumption is detrended using the regression on a quadratic time trend or a Hodrick-Prescott filter. The results were not altered in any significant way.

Figure 1. Data



deviations of shocks, and a normal distribution for the remaining parameters. The main characteristics of the prior distributions are reported in table 1. In most cases, distribution parameters are the same as in Smets and Wouters (2003) for the euro area, but we also incorporate some information drawn from Onatski and Williams (2004).[17]

While the shocks in a given country are assumed to be uncorrelated, we allow a nonzero correlation between a given shock in two countries. We thus denote δ_p, δ_a, and δ_i the correlations between domestic and foreign preference shocks, productivity shocks, and monetary policy shocks, respectively. Correlations across countries are assumed to follow a normal distribution, with a mean of 0.2 and

[17]Onatski and Williams (2004) provide an interesting investigation of some shortcomings of the standard Bayesian approach in the context of DSGE models. In particular, they put forward that parameter estimates are very sensitive to the way priors are introduced. We took advantage of some of their results.

Table 1. Prior Distribution for the Parameters

		Type	Mean	Std. Error
Consumption Habit	γ	Beta	0.700	0.100
Consumption Elast. of Subst.	σ	Normal	2.000	0.250
Labor Disutility	φ	Normal	2.000	0.250
Price Indexation	ξ	Beta	0.700	0.100
Calvo Probability	α	Beta	0.700	0.100
PR Lagged Interest Rate	ψ_i	Beta	0.700	0.100
PR Inflation	ψ_π	Normal	1.500	0.100
PR Output Gap	ψ_y	Normal	0.500	0.100
Serial Corr. Preference Shock	ρ_p	Beta	0.600	0.100
Serial Corr. Productivity Shock	ρ_a	Beta	0.600	0.100
Serial Corr. Mon. Policy Shock	ρ_i	Beta	0.600	0.100
Vol. Preference Shock	σ_p	Uniform	0.000	2.000
Vol. Productivity Shock	σ_a	Uniform	0.000	2.000
Vol. Mon. Policy Shock	σ_i	Uniform	0.000	2.000
Cross-Corr. Preference Shocks	δ_p	Normal	0.200	0.100
Cross-Corr. Productivity Shocks	δ_a	Normal	0.200	0.100
Cross-Corr. Mon. Policy Shocks	δ_i	Normal	0.200	0.100

a standard error of 0.1. We use the same priors for all countries and the euro area in turn.

Finally, we imposed dogmatic priors over the discount factor β and the elasticity of substitution across goods produced in a given country, θ. The values we use ($\beta = 0.99$ and $\theta = 10$) are conventional in the literature. The consumption/output ratio s is set equal to 0.57 for all countries. The selection of the parameters of home bias in preferences (ω) is done by using the trade flows statistics provided by the International Monetary Fund's Direction of Trade Statistics database. We first calculate the relative share of each country in the imports of the other countries. We then deduce the renormalized weights to be consistent with our assumption that the three countries under study cover the whole external trade of each other. We therefore set these parameters as follows, in order to reflect the weight of each country in the external trade of the others:

the weights of German, French, and Italian goods in the consumption of German households are $(0.8; 0.11; 0.09)$. For French and Italian households, the weights are $(0.13; 0.8; 0.07)$ and $(0.13; 0.07; 0.8)$, respectively. We checked that marginally altering these values would not change significantly our results.

3.3 Parameter Estimates

3.3.1 Results for the MCM

The joint dynamics of the whole system is estimated simultaneously for Germany, France, and Italy. Table 2 provides two sets of information regarding parameter estimates. The first set reports the posterior mode of parameters, which is obtained directly by maximizing the log of the posterior distribution with respect to parameters. The second set contains the 5th, 50th (labeled "Median"), and 95th percentiles of the posterior distribution of parameters. The posterior distribution of some parameters (namely, σ, φ, and ψ_π) is rather close to the prior distribution. This suggests that these parameters do not strongly affect the likelihood and translates to the rather large associated standard deviations.

As regards the behavior of households, our estimates of the consumption elasticity of substitution (σ) range between 1.5 and 2, while the inverse of the elasticity of labor disutility (φ) is close to 2. Although we select the same priors for all countries, we obtain significant differences for the habit-persistence parameter γ. This parameter is estimated to be medium in Germany (0.63) and France (0.69), and large in Italy (0.78). Using the usual 5 percent significance level, we reject the null hypothesis that the three parameters are equal across countries, suggesting that there is some heterogeneity in structural parameters across countries. Turning to the behavior of firms, we obtain some disparity in the parameters of price indexation ξ, which range between 0.28 for Germany and 0.43 for Italy, although the difference does not turn out to be significant. The Calvo probabilities α are estimated around 0.82, suggesting that the average duration of price contracts is less than six quarters. Reaction-function parameters display similar patterns across countries. The long-run reactions of short-term interest rate to inflation and output gap are close to 1.5 and 0.5, respectively, in the three countries. The interest-rate-smoothing parameter ψ_i is about 0.87.

Table 2. Parameter Estimates for the MCM

		Germany				
		Mode	**Std. Dev.**	**5%**	**Median**	**95%**
Consumption Habit	γ	0.630	0.050	0.553	**0.632**	0.714
Consumption Elasticity of Substitution	σ	1.542	0.232	1.162	**1.533**	1.922
Labor Disutility	φ	1.934	0.253	1.522	**1.929**	2.349
Price Indexation	ξ	0.290	0.078	0.157	**0.283**	0.406
Calvo Probability	α	0.839	0.019	0.809	**0.840**	0.869
Policy Rule: Lagged Interest Rate	ψ_i	0.871	0.020	0.841	**0.873**	0.901
Policy Rule: Inflation	ψ_π	1.507	0.100	1.340	**1.510**	1.666
Policy Rule: Output Gap	ψ_y	0.458	0.104	0.288	**0.462**	0.627
Serial Corr. Preference Shock	ρ_p	0.640	0.065	0.531	**0.643**	0.741
Serial Corr. Productivity Shock	ρ_a	0.740	0.067	0.635	**0.741**	0.854
Serial Corr. Monetary Policy Shock	ρ_i	0.506	0.067	0.395	**0.508**	0.617
Vol. Preference Shock	σ_p	0.048	0.008	0.035	**0.047**	0.061
Vol. Productivity Shock	σ_a	0.037	0.006	0.026	**0.036**	0.047
Vol. Monetary Policy Shock	σ_i	0.244	0.020	0.211	**0.243**	0.276
Cross-Correlations across Countries		**Mode**	**Std. Dev.**	**5%**	**Median**	**95%**
Preference Shock – 1/2	δ_{p12}	0.311	0.063	0.201	**0.313**	0.410
Preference Shock – 1/3	δ_{p13}	0.166	0.067	0.059	**0.168**	0.273
Preference Shock – 2/3	δ_{p23}	0.279	0.071	0.166	**0.279**	0.397
Productivity Shock – 1/2	δ_{a12}	0.194	0.067	0.077	**0.196**	0.300
Productivity Shock – 1/3	δ_{a13}	−0.032	0.076	−0.156	**−0.032**	0.096
Productivity Shock – 2/3	δ_{a23}	0.135	0.075	0.018	**0.138**	0.258
Monetary Policy Shock – 1/2	δ_{i12}	0.198	0.070	0.087	**0.200**	0.317
Monetary Policy Shock – 1/3	δ_{i13}	0.124	0.066	0.016	**0.127**	0.229
Monetary Policy Shock – 2/3	δ_{i23}	0.239	0.069	0.132	**0.237**	0.355

(*Continued*)

The standard deviations of the preference and productivity shocks are very close for the three countries, although they are smaller than the area-wide counterparts. In contrast, some large differences in the variability of the monetary policy shock are found. While the volatility is low in Germany and Italy (around 0.24 percent), it is large in France (at 0.42 percent). This result may be related to some aspects of the French monetary policy that are not

Table 2. (Continued)

France					Italy				
Mode	**Std. Dev.**	**5%**	**Median**	**95%**	**Mode**	**Std. Dev.**	**5%**	**Median**	**95%**
0.688	0.045	0.617	**0.691**	0.765	0.777	0.029	0.730	**0.777**	0.823
1.851	0.226	1.482	**1.851**	2.228	2.009	0.218	1.656	**2.009**	2.373
2.015	0.252	1.595	**2.019**	2.428	1.922	0.247	1.511	**1.919**	2.316
0.324	0.083	0.191	**0.318**	0.455	0.436	0.102	0.257	**0.428**	0.593
0.822	0.017	0.794	**0.823**	0.848	0.794	0.022	0.759	**0.795**	0.830
0.820	0.027	0.778	**0.822**	0.864	0.906	0.014	0.885	**0.908**	0.929
1.517	0.101	1.353	**1.518**	1.681	1.497	0.094	1.344	**1.500**	1.648
0.482	0.102	0.314	**0.480**	0.645	0.522	0.091	0.375	**0.522**	0.670
0.509	0.077	0.380	**0.510**	0.633	0.793	0.036	0.739	**0.795**	0.851
0.660	0.075	0.536	**0.661**	0.780	0.854	0.035	0.796	**0.855**	0.911
0.447	0.067	0.337	**0.445**	0.557	0.414	0.071	0.300	**0.412**	0.534
0.063	0.010	0.047	**0.062**	0.078	0.055	0.008	0.043	**0.054**	0.068
0.038	0.007	0.028	**0.038**	0.050	0.035	0.006	0.026	**0.035**	0.045
0.426	0.034	0.372	**0.423**	0.482	0.228	0.021	0.196	**0.226**	0.261

incorporated in the model, such as the implicit anchoring to the German monetary policy from 1983 on.

Concerning the serial correlation of shocks, the table reveals some significant differences across countries for the preference shock ($\rho_p = 0.51$ in France and 0.80 in Italy) and for the productivity shock ($\rho_a = 0.66$ in France and 0.86 in Italy). In contrast, the estimates of ρ_i are all very close to 0.45. Most cross-country correlations between shocks are significantly positive. Note, however, that shocks are far from being perfectly correlated across countries. This result is of importance, because it suggests that the asymmetry of shocks may be rather large across countries. It appears as the main source of heterogeneity within the euro area.

3.3.2 Results for the AWM

Table 3 reports the parameter estimates for the AWM model. As regards the behavior of households, our estimate of the inverse of the consumption elasticity of substitution (σ) is equal to 2.08,

Table 3. Parameter Estimates for the AWM

		Estimated ML		Posterior Distribution			Smets-Wouters	Onatski-Williams
		Mode	Std. Dev.	5%	Median	95%	Median	Median
Consumption Habit	γ	0.867	0.040	0.800	**0.871**	0.932	0.595	0.400*
Consumption Elast. of Subst.	σ	2.074	0.242	1.674	**2.078**	2.465	1.371	2.178
Labor Disutility	φ	1.972	0.227	1.600	**1.979**	2.350	2.491	3.000*
Price Indexation	ξ	0.485	0.102	0.310	**0.478**	0.646	0.472	0.323
Calvo Probability	α	0.929	0.020	0.900	**0.933**	0.956	0.905	0.930*
PR Lagged Interest Rate	ψ_i	0.855	0.026	0.814	**0.858**	0.897	0.958	0.962
PR Inflation	ψ_π	1.480	0.098	1.310	**1.480**	1.632	1.688	1.684
PR Output Gap	ψ_y	0.163	0.157	−0.032	**0.108**	0.407	0.095	0.099
Serial Corr. Preference Shock	ρ_p	0.436	0.103	0.270	**0.426**	0.610	0.842	0.876
Serial Corr. Productivity Shock	ρ_a	0.591	0.101	0.429	**0.599**	0.757	0.815	0.957
Serial Corr. Mon. Policy Shock	ρ_i	0.553	0.081	0.413	**0.551**	0.681	0.865	0.582
Vol. Preference Shock	σ_p	0.114	0.036	0.068	**0.106**	0.161	0.336	0.240
Vol. Productivity Shock	σ_a	0.127	0.063	0.048	**0.106**	0.219	0.598	0.343
Vol. Mon. Policy Shock	σ_i	0.210	0.017	0.181	**0.208**	0.237	0.081	1.000*

while the inverse of the elasticity of labor disutility (φ) is equal to 1.98. These estimates are consistent with the MCM estimates. In contrast, the habit parameter γ is as high as 0.87. This estimate significantly differs from the MCM estimate and suggests that the aggregate model is more persistent than the multicountry model.

Focusing on the behavior of firms, the parameter of price indexation is $\xi = 0.48$, while the probability that firms are not allowed to reoptimize their price is $\alpha = 0.93$. The degree of price stickiness is rather large, since the average duration of price contracts is about fifteen quarters. This figure is somewhat larger than microeconomic evidence, but it is in the range of previous macroeconomic estimates (see Smets and Wouters 2003 and Onatski and Williams 2004).

The estimate of the monetary policy rule is only indicative of how short-term interest rates reacted to macroeconomic developments over the sample period. In the absence of a common central bank over the sample, this estimate cannot be taken as reflecting plausibly the behavior of monetary authorities. The long-run response to inflation is $\psi_\pi = 1.5$, while the reaction to output gap is $\psi_y = 0.11$. Once again, these estimates contrast sharply with the MCM estimates reported above.

The estimated model is able to capture most of the persistence found in the data. This translates to low estimates of the serial correlation of shocks. Our median estimates range between 0.43 and 0.6.

To sum up, our estimates suggest that the AWM suffers from heterogeneity bias. This bias is sizable and significant for at least three key parameters that reflect the behavior of households (the habit persistence γ), firms (the Calvo probability α), and monetary authorities (the reaction to output gap ψ_y). In all cases, it points toward a higher persistence in the dynamics of the AWM. Similar results have been reported in simpler models, such as those using autoregressive processes (Granger 1980).[18]

[18]Interestingly, Pytlarczyk (2005) finds very close estimates for common parameters (for Germany and the euro area) although using a more complicated model than those presented here.

4. Optimal Monetary Policy

We now turn to the evaluation of the optimal monetary policy in the context of the euro area. Therefore, we acknowledge that there is a unique central bank within the euro area, and we keep the nominal exchange rate constant and equal to one within the area. An advantage of having developed a structural model based on optimizing behaviors is that it provides a natural objective for monetary policy—namely, the maximization of the welfare, defined as the expected utility of the representative household. Following Woodford (2003) and Giannoni and Woodford (2005), we compute the second-order Taylor-series approximation to this objective function as a quadratic function of variables and shocks. Various aspects of our model, such as inflation inertia and external habit formation, require that we derive an appropriate welfare-based stabilization objective.

Two important issues arise when considering the evaluation of welfare in the context of an open economy with habit formation. First, as discussed in Rotemberg and Woodford (1998), under the assumption that the constant subsidy for output ϑ neutralizes the distortion associated with firms' market power, it can be shown that in a closed economy the optimal monetary policy is the one that replicates the flexible-price equilibrium allocation.[19] In an open economy, as noted by Corsetti and Pesenti (2001) and Galí and Monacelli (2005), a second source of distortion comes from the fact that the transmission of monetary policy affects demand not only through the relative cost of borrowing but also through its effect on the terms of trade. This is a consequence of the imperfect substitutability between home and foreign goods, combined with sticky prices. As in Benigno and Benigno (2003), we assume that the subsidy for output exactly offsets the combined effects of market power and the terms-of-trade distortions in the steady state.

Second, in an open-economy framework, most previous studies investigated the way the optimal monetary policy may be designed for a given type of monetary arrangement between central banks.

[19] The intuition is straightforward: with the subsidy in place, there is only one distortion left in the economy—namely, sticky price. By stabilizing markups at their frictionless level, nominal rigidities cease to be binding, since firms do not feel any desire to adjust their price.

Typical extreme cases are noncooperation and full cooperation. Our evaluation of the optimal monetary policy obviously presumes full cooperation, since only one central bank is involved. More specifically, our focus is not on whether coordination may improve the global welfare, but rather on whether the fully cooperative monetary policy should be based on an aggregated model or on a multicountry model.

4.1 The Welfare Objective

4.1.1 Expression for the Welfare

DSGE models deliver a natural measure of welfare based on the representative household's utility. It is defined as the conditional expectation of the current and discounted future values of the approximated utility function. In appendix 3, we derive the welfare for the two-country model. In the closed-economy version, which corresponds to our AWM, the aggregated welfare at date 0 can be approximated by

$$\begin{aligned}\mathcal{W}_0^{AWM} \approx -\frac{\bar{\mathbb{U}}_{\bar{C}}\bar{C}}{2}\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\Big\{&(\hat{c}_t+\beta\gamma)^2+(\beta\gamma(1+\rho_p)-\sigma-1)\hat{c}_t^2\\ &+\frac{\sigma}{1-\gamma}(\hat{c}_t-\gamma\hat{c}_{t-1})^2+\left(\frac{\sigma}{1-\gamma}+\varphi\right)\left(\hat{c}_t-\hat{c}_t^n\right)^2\\ &-\frac{\sigma\gamma}{(1-\gamma)}\left(\hat{c}_t-\hat{c}_{t-1}^n\right)^2-\beta\gamma\rho_p\left(\hat{c}_t-\hat{\varepsilon}_{p,t}\right)^2\\ &+\frac{\theta\alpha}{(1-\alpha)(1-\beta\alpha)}(\hat{\pi}_t-\xi\hat{\pi}_{t-1})^2\Big\}+t.i.p., \end{aligned} \tag{25}$$

where all variables denote area-wide variables and parameters are those pertaining to the AWM. $\hat{c}_t^n$ is the natural value of aggregate consumption and *t.i.p.* regroups terms independent of the actual policy.[20] Expression (25) combines features implied by the introduction of inflation inertia and external habit formation. Interestingly, we notice that, in our estimated model, it is optimal for the central bank to put a much higher weight (about 100 times more) on the

[20]See appendix 2 for details on the log-linearized flexible-price equilibrium.

stabilization of goods price inflation than on the stabilization of the other variables. In addition, no concern about interest rate stabilization is present in this expression. This issue will be addressed in the subsequent sensitivity analysis.

The aggregated welfare in the two-country approach (taking into account the heterogeneity across countries) is defined as the weighted average of the national welfare functions:

$$\mathcal{W}_0^{MCM} = n\mathcal{W}_0 + (1-n)\mathcal{W}_0^*, \tag{26}$$

where $\mathcal{W}_0$ and $\mathcal{W}_0^*$ are detailed in appendix 3.[21]

4.1.2 Evaluation of the Optimal Policy Rule

We evaluate the optimal monetary policy by taking the unconditional expectation of expressions (25) and (26) with respect to the distribution of exogenous shocks, and under the assumption that all endogenous variables in the initial period are at their unconditional expectation of zero. This assumption ensures that the desirability of the chosen plan does not depend upon initial conditions at time 0. We thus define the unconditional expectation of the welfare as $\breve{\mathcal{W}}_0^i = (1-\beta)\mathbb{E}\mathcal{W}_0^i$, $i = MCM, AWM$.[22]

Since our aim is to compare the welfare consequences of adopting as forecasting model the (suboptimal) AWM instead of the MCM, we proceed as follows, considering two approaches in turn:

(i) In the *multicountry approach*, the central bank uses the MCM to forecast national variables. The policy rule is assumed to be defined in terms of aggregate variables, since the policy rule is designed on the basis of area-wide developments only. Its expression is given by

$$\hat{\imath}_t = F_{MCM} \times \Xi \times \hat{s}_t^{MCM},$$

[21] In the N-country case, the total welfare is given by $\mathcal{W}_0^{MCM} = \sum_{j=1}^{N} n_j \mathcal{W}_0^j$, where n_j is the weight of the country j in the euro-area GDP and $\sum_{j=1}^{N} n_j = 1$. In our evaluation, we hold the following weights: 0.4 for Germany and 0.3 for France and Italy.

[22] By maximizing unconditional welfare, we are implicitly maximizing welfare in the steady state. This welfare comparison ignores the possibility of losses in welfare on the transition path from one steady state to another (see Schmitt-Grohé and Uribe 2004).

where in a two-country setup

$$\hat{s}_t^{MCM} = \left(\hat{\varepsilon}_{p,t}, \hat{\varepsilon}_{p,t}^*, \hat{a}_t, \hat{a}_t^*, \hat{c}_t^n, \hat{c}_t^{*n}, \hat{c}_{t-1}^n, \hat{c}_{t-1}^{*n}, \hat{c}_t, \hat{c}_t^*, \hat{c}_{t-1}, \hat{c}_{t-1}^*, \hat{\pi}_{H,t}, \hat{\pi}_{F,t}^*, \hat{\imath}_{t-1}, \hat{\imath}_{t-1}^*\right)'$$

denotes the vector of state variables under the MCM, and Ξ is an aggregation matrix that defines the area-wide aggregates as functions of country variables. Then, the constrained optimal monetary policy rule (F_{MCM}) is obtained by maximizing the weighted average of national welfare functions (expression (26)), allowing cross-country heterogeneity of behaviors. It should be noted that this rule is not in general fully optimal under the MCM, since it imposes several constraints on the parameters of the rule. Indeed, domestic and foreign variables are constrained to have the same weight in the reaction function.[23] The maximal value of welfare is denoted $\breve{\mathcal{W}}_0^{MCM}$. For further use, we also define the fully optimal policy rule as F_{MCM}^{opt} and the corresponding welfare as $\breve{\mathcal{W}}_0^{opt}$.

(ii) In the *aggregated approach*, the central bank forecasts area-wide variables (using the AWM) and adopts a policy rule designed in terms of aggregate variables only, in the form

$$\hat{\imath}_t = F_{AWM} \times \hat{s}_t^{AWM},$$

where

$$\hat{s}_t^{AWM} = \left(\hat{\varepsilon}_{p,t}, \hat{a}_t, \hat{c}_t^n, \hat{c}_{t-1}^n, \hat{c}_t, \hat{c}_{t-1}, \hat{\pi}_t, \hat{\imath}_{t-1}\right)'$$

denotes the vector of state variables under the AWM. The optimal monetary policy rule is then obtained by maximizing the aggregated welfare (expression (25)), assuming homogeneity of behaviors across countries. The resulting welfare would not be directly comparable to $\breve{\mathcal{W}}_0^{MCM}$, since the two expressions are evaluated under two different sets

[23] An important consequence is that it cannot be computed using the standard approach, based on solving the Bellman equation. Rather, the constrained rule F_{MCM} is obtained by numerically maximizing the welfare among all policy rules that include aggregate variables only.

of assumptions. To obtain comparable welfare functions, we evaluate the welfare associated with the AWM policy rule (F_{AWM}) using the MCM. The welfare of the area is therefore computed as the weighted average of national welfares. This expression collapses to the aggregated welfare under full homogeneity only. The maximal value of welfare associated with the AWM policy rule but evaluated under the MCM is denoted $\breve{\mathcal{W}}_0^{AWM}$. We then deduce the cost of using the (suboptimal) aggregated approach from the comparison of $\breve{\mathcal{W}}_0^{AWM}$ and $\breve{\mathcal{W}}_0^{MCM}$.

4.2 Welfare Implications of Heterogeneity

The constrained optimal rule evaluated under the multicountry approach (F_{MCM}) is expected to induce a higher welfare than the optimal rule under the aggregated approach (F_{AWM}). The reason is that, although both rules are defined in terms of aggregate variables only, the parameters for F_{MCM} are obtained by maximizing the welfare under the "true" model. Assessing whether the central bank should be concerned about heterogeneity therefore requires that the welfare cost of using the AWM be economically significant. For this purpose, we compute two measures that provide some information on the welfare reduction due to the use of the AWM.

The first measure gives the cost of using the suboptimal forecasting model AWM as a permanent percentage shift in steady-state aggregate consumption. It is defined by scaling the welfare loss ($\breve{\mathcal{W}}_0^{AWM} - \breve{\mathcal{W}}_0^{MCM}$) by $\bar{\mathbb{U}}_{\bar{C}}\bar{C}$:

$$\delta_1 = -\frac{\breve{\mathcal{W}}_0^{AWM} - \breve{\mathcal{W}}_0^{MCM}}{\bar{\mathbb{U}}_{\bar{C}}\bar{C}}. \tag{27}$$

This measure has been previously investigated by Erceg, Henderson, and Levin (2000), Amato and Laubach (2004), Tchakarov (2004), and Benigno and López-Salido (2006).[24]

[24]Since $\breve{\mathcal{W}}_0^i = (1-\beta)\,\mathbb{E}\,(\mathcal{W}_0^i)$, expression (27) is also equivalent to

$$\delta_1 = -(1-\beta)\frac{\mathbb{E}\,(\mathcal{W}_0^{AWM}) - \mathbb{E}\,(\mathcal{W}_0^{MCM})}{\bar{\mathbb{U}}_{\bar{C}}\bar{C}},$$

where the welfare functions are evaluated under the MCM.

The second measure is the fraction of the gap (in terms of welfare) between the AWM-based rule and the fully optimal MCM-based rule that is filled by the constrained MCM-based rule. It is defined as

$$\delta_2 = \frac{\breve{\mathcal{W}}_0^{AWM} - \breve{\mathcal{W}}_0^{MCM}}{\breve{\mathcal{W}}_0^{AWM} - \breve{\mathcal{W}}_0^{opt}}. \tag{28}$$

This measure allows us to compare our evaluations with those performed by Angelini et al. (2002) and Monteforte and Siviero (2003) in the context of ad hoc loss functions.

Table 4 reports the welfare obtained for the various policy rules considered, using the median of the posterior distribution of estimated parameters. The table gives the welfare under the AWM, the constrained MCM, and the fully optimal MCM, as well as the two measures of welfare cost. We obtain that the use of the AWM to define the monetary policy rule implies a welfare reduction as compared with the use of the constrained MCM (shown in panel A, the benchmark estimate). If we measure the welfare cost as the permanent percentage shift in steady-state aggregate consumption, we obtain that a cost of using the AWM is equal to $\delta_1 = 0.0037$. This suggests that the steady-state aggregate consumption level obtained using the AWM is almost 0.37 percent lower than the steady-state aggregate consumption obtained using the constrained MCM. This evaluation of the cost of using a suboptimal forecasting model is rather large as compared with some previous welfare evaluations.[25] Note, however, that the measure δ_2 provides additional insight into the source of welfare loss in using an AWM. Indeed, it indicates that the constrained MCM-based rule makes up for 98 percent of the distance between the AWM-based rule and the fully optimal MCM-based rule. This result suggests that, consistent with previous evidence, this is not the use of a restricted policy rule based on aggregate variables that is costly, but rather the use of a suboptimal forecasting model.

[25] Benigno and López-Salido (2006) estimate the cost of monetary policies in the context of heterogeneous Phillips curves within the euro area. They obtain that the cost of using an HICP-targeting policy rule instead of the optimal monetary policy is about 0.02 percent of steady-state consumption.

Table 4. Welfare Comparison

	Value of Welfare			Measures of Welfare Cost	
Model	**AWM**	**Constrained MCM**	**Optimal MCM**	δ_1 **(in %)**	δ_2
A. Benchmark	**−1.4700**	**−1.1024**	**−1.0965**	**0.0037**	**0.9842**
B. Sensitivity					
Without Habit Formation	−2.2330	−1.9980	−1.9890	0.0024	0.9631
Without Price Indexation	−1.7370	−1.6210	−1.6192	0.0012	0.9847
Without Habit Formation or Price Indexation	−2.8200	−2.7832	−2.7827	0.0004	0.9866
C. Robustness					
With Output Data Rather than Consumption Data	−1.9520	−1.5635	−1.5598	0.0039	0.9906
With an Inflation Target Shock	−1.7420	−1.4019	−1.3888	0.0034	0.9629
With Fixed Costs in the Production Function	−2.9462	−2.3367	−2.3271	0.0061	0.9845

4.3 Sensitivity Analysis

As a first sensitivity exercise, we investigate the role of the two sources of endogenous persistence mechanisms we introduced in the model to reproduce the properties of the data—namely, external habit formation and price indexation. We measure how varying both of these assumptions affects the value of the cost of using an AWM rather than an MCM. To this end, we reestimate the AWM and the MCM under alternative assumptions, with and without habit formation and with and without price indexation (i.e., eight sets of estimates). In panel B of table 4, the second row reports the results for the two measures of welfare cost for the model without habit formation, the third row for the model without price indexation, and the fourth row without habit formation or price indexation. As may be expected, removing these friction mechanisms reduces the difference in welfare between the AWM and the MCM. Indeed, the welfare cost falls from 0.37 percent under the full specification to only 0.04 percent in the absence of habit formation and price indexation. We also notice that the welfare cost of using the AWM is more widely reduced when we assume no price indexation than when we assume no habit formation. In the former case, we obtain $\delta_1 = 0.12$ percent, while we have $\delta_1 = 0.24$ percent in the latter case. The main reason is that the price-indexation parameter (ξ) affects the welfare through the expression $(\hat{\pi}_t - \xi\hat{\pi}_{t-1})^2$, which has a weight on the aggregate welfare 100 times larger than the weights on the other variables. Therefore, the rather large welfare cost of using the AWM appears to be mainly attributable to the introduction in our model of price indexation rather than to habit formation.[26]

As a second sensitivity exercise, we investigate the role of interest rate smoothing, a feature that has been found to be necessary to reproduce the observed monetary policy rules. It is known that introducing a microfounded concern for interest rate smoothing is rather complicated, especially in the presence of habit formation (Woodford 2003). To eliminate this problem, we propose to simply include an ad hoc interest-rate-smoothing objective $\Lambda_i(\hat{\imath}_t - \hat{\imath}_{t-1})^2$ in

[26]But we find that habit formation is very important for the model dynamics. Justiniano and Preston (2004) have also found such a result.

the expressions (25) and (26) of the aggregate welfare.[27] We focus on a grid over the weight on interest rate smoothing $\Lambda_i = [0, 1]$. Figure 2 displays the value of the welfare cost for each weight. We first observe that introducing some concern for interest rate volatility in the welfare measure would not affect our main result that the use of the (suboptimal) AWM is costly as compared with a model that incorporates cross-country heterogeneity. Second, the larger the weight on interest rate smoothing, the higher the welfare cost. The welfare cost increases from $\delta_1 = 0.37$ percent when there is no interest rate smoothing to $\delta_1 = 0.86$ percent in the presence of strong interest rate smoothing ($\Lambda_i = 1$). This implies that, when we introduce an interest-rate-smoothing objective, the cost of using the AWM instead of the MCM is larger than under the baseline central bank preferences. The reason is that, under the AWM, the economy is less reactive to changes in the short-term interest rate. For instance, the consumption-habit and price-indexation parameters are larger in the AWM than in the MCM. Consequently, the central bank has to be more reactive and increase its short-term rate more severely, which in turn decreases the welfare.

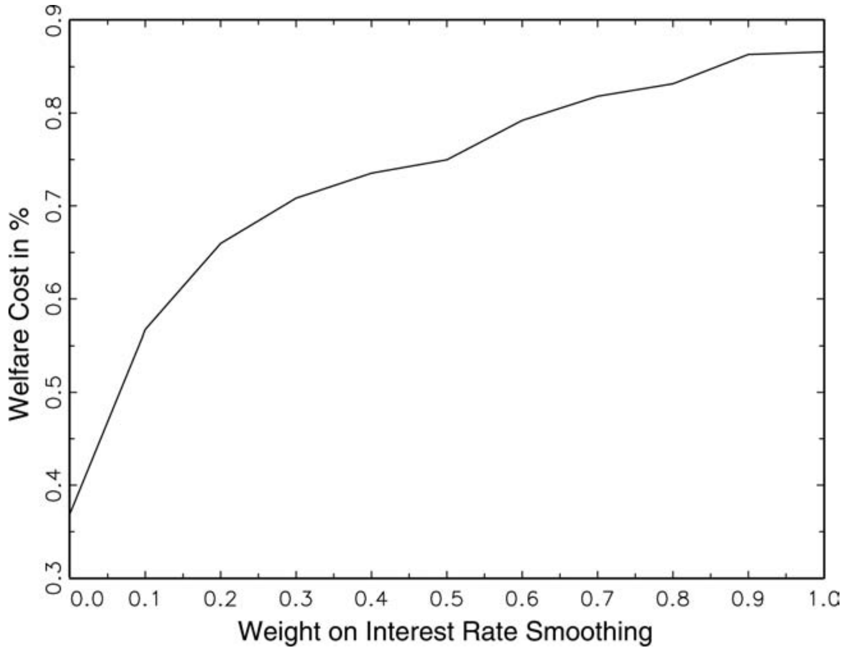
4.4 Robustness Analysis

In this section, we address some robustness issues that may be raised by the specification we adopted to model European economies.

A first issue is related to the choice of consumption as opposed to output gap in the estimation of the model. Indeed, it may be argued that the specification adopted in the paper does not make explicit reference to the trade flows across countries. In fact, the MCM can be easily accommodated to include output as an observable variable in place of consumption, using equations (23) and (24). In this case, the trade flows across countries would be more transparent, although the genuine specification of the model would not be altered. Regarding the AWM, exactly the same specification can be estimated, since consumption and output are theoretically equivalent in our closed economy. Unreported evidence shows that the

[27] We have also replaced the interest-rate-smoothing objective ($\Lambda_i(\hat{\imath}_t - \hat{\imath}_{t-1})^2$) with the variance of the interest rate ($\Lambda_i \hat{\imath}_t^2$), without finding any significant difference.

Figure 2. Welfare Losses with Interest Rate Smoothing



choice of output in place of consumption mainly results in a decrease of the overall persistence of the system.[28] The habit parameter γ decreases in all countries, consistent with the more volatile dynamics of output. The Calvo parameter and the interest-rate-smoothing parameter are also affected in the same direction, but to a lesser extent. As table 4 reveals (first row of panel C), the consequences on welfare costs are negligible.

A second issue relies on detrending. As described above, nominal interest and inflation rates in Europe have a downward trend over the estimation period. We took this trend into account by linearly detrending these series. Clearly, this might bias our estimates and the subsequent welfare analysis. A first solution would be to estimate the model in levels instead of deviations from trends. However, it would raise additional difficulties—in particular, due to the

[28] All the results presented in this section are in the technical appendix available upon request from the authors.

near nonstationarity of the original series. An alternative approach consists of introducing a time-varying inflation target.[29] We assume that this shock follows a first-order autoregressive process, as in Smets and Wouters (2003). Estimation of this model results in an increase in the parameters driving the persistence of the system. This was expected, given that inflation and interest rates are no longer detrended. However, when we turn to the welfare evaluation, we observe that the cost of using the suboptimal AWM is not significantly altered ($\delta_1 = 0.34$ percent).

As a last robustness check, we consider a model with fixed costs in the production function, as in Rotemberg and Woodford (1998). The main objective of this specification is to ensure that economic profits are equal to zero at the steady state. In the presence of fixed costs, the welfare has to be redefined accordingly, so that the welfare cost of using the AWM is very likely to be affected. Our estimates indicate that fixed costs play a significant role in the model, since the share of the fixed costs in production is significant in all countries (50 percent for the AWM; between 10 percent and 30 percent in the MCM). We still have a perfect illustration of the aggregation bias. In addition, this specification seems to generate more internal persistence since Calvo parameters and habit parameters, for instance, increase (implying a decrease in the serial correlation of shocks). In terms of welfare, we observe in the last row of table 4 (panel C) a significant increase in the welfare cost of using the AWM as a forecasting model. It is now almost twice the benchmark estimate ($\delta_1 = 0.61$ percent instead of 0.37 percent). The reason may come from two sources: (i) a modification of the welfare expression (introducing a fixed cost is not neutral) and (ii) a change in the structural parameters.

5. Conclusion

In this paper, we evaluate the cost of ignoring the cross-country heterogeneity within the euro area when implementing the optimal monetary policy. To address this issue, we develop a multicountry DSGE model, which is used to estimate the dynamics of national

[29] Such a shock appears simultaneously in the Phillips curve and in the monetary policy rule.

economies within the euro area. This model incorporates frictions required to reproduce the persistence of the actual data, including the presence of sticky-price setting and external habit formation in consumption. An additional characteristic of the model is the introduction of heterogeneous behaviors across countries that allows us to investigate the cost of using an AWM instead of an MCM.

Using Bayesian techniques, we estimate the MCM and AWM and provide evidence that the behavioral parameters in Germany, France, and Italy display some significant differences, and that shocks affecting the different economies are only very weakly correlated. Our results therefore highlight that heterogeneity can be mainly attributable to the asymmetry of shocks across countries rather than to differences in the behavioral parameters.

Since our model is suitable for the analysis of optimal monetary policy, we then compare the two models on the basis of their ability to maximize the welfare of the area-wide representative household. The welfare costs associated with the two optimal rules are then compared, allowing heterogeneity of behaviors. We find that using an AWM generates a relatively large welfare loss that corresponds to a permanent decrease in steady-state aggregate consumption by around 0.37 percent. Our results also suggest that this is not the use of a rule based on aggregate variables that is costly in terms of welfare, but rather the use of a suboptimal forecasting model. Moreover, the rather large welfare cost of using the AWM appears to be mainly attributable to the introduction in our model of price indexation rather than to habit formation. Finally, we investigate the implications of heterogeneity when an additional ad hoc interest-rate-smoothing objective is allowed. Introducing some concern for interest rate volatility in the welfare measure would not affect the previous results.

It may be argued that the cost of designing, estimating, and using an MCM is rather large, suggesting that the AWM would be less costly to implement. However, our estimate of the difference between the AWM and the MCM in terms of welfare is very sizable. In addition, it is worth emphasizing that our evaluation is based on the three largest countries of the area that may be viewed as very similar economies. It is likely that including additional economies would even widen the discrepancies between the two models. Using larger models incorporating different fiscal policies and labor market characteristic (i.e., increasing heterogeneity) should also tend toward higher welfare losses.

Appendix 1. The Log-Linearized Dynamic Equilibrium

This appendix displays the log-linearized dynamic equilibrium in the case of a two-country model. $\hat{x}_t$ denotes the log-deviation from the steady-state value $\bar{x}$, i.e., $\hat{x}_t = \log(x_t/\bar{x})$.

Model

- Home IS curve:

$$\hat{c}_t = \frac{\gamma}{1+\gamma}\hat{c}_{t-1} + \frac{1}{1+\gamma}\mathbb{E}_t\hat{c}_{t+1} - \frac{(1-\gamma)}{(1+\gamma)\sigma}(\hat{\imath}_t - \mathbb{E}_t\hat{\pi}_{H,t+1}) + \frac{(1-\gamma)(1-\omega)}{(1+\gamma)\sigma}\mathbb{E}_t\Delta\hat{\tau}_{t+1} + \frac{(1-\rho_p)(1-\gamma)}{(1+\gamma)\sigma}\hat{\varepsilon}_{p,t} \quad (29)$$

- Home Phillips curve:

$$\hat{\pi}_{H,t} = \frac{\xi}{1+\beta\xi}\hat{\pi}_{H,t-1} + \frac{\beta}{1+\beta\xi}\mathbb{E}_t\hat{\pi}_{H,t+1} + \frac{(1-\beta\alpha)(1-\alpha)}{(1+\beta\xi)\alpha}\widehat{mc}_t \quad (30)$$

- Home marginal cost:

$$\widehat{mc}_t = \left(\frac{\sigma}{1-\gamma} + \varphi\omega s\right)\hat{c}_t - \frac{\gamma\sigma}{1-\gamma}\hat{c}_{t-1} + \varphi(1-\omega s)\hat{c}_t^* - (1+\varphi)\hat{\varepsilon}_{a,t} + [(1-\omega)(1+\varphi\omega s) + \varphi(1-\omega^*)(1-\omega s)]\hat{\tau}_t \quad (31)$$

- Home aggregate output:

$$\hat{y}_t = [(1-\omega)\omega s + (1-\omega^*)(1-\omega s)]\hat{\tau}_t + \omega s\hat{c}_t + (1-\omega s)\hat{c}_t^* \quad (32)$$

- Home preference shock:

$$\hat{\varepsilon}_{p,t} = \rho_p\hat{\varepsilon}_{p,t-1} + \eta_{p,t} \quad (33)$$

- Home productivity shock:

$$\hat{\varepsilon}_{a,t} = \rho_a\hat{\varepsilon}_{a,t-1} + \eta_{a,t} \quad (34)$$

- Foreign IS curve:

$$\hat{c}_t^* = \frac{\gamma^*}{1+\gamma^*}\hat{c}_{t-1}^* + \frac{1}{1+\gamma^*}\mathbb{E}_t\hat{c}_{t+1}^* - \frac{(1-\gamma^*)}{(1+\gamma^*)\sigma^*}\left(\hat{\imath}_t^* - \mathbb{E}_t\hat{\pi}_{F,t+1}^*\right) - \frac{(1-\gamma^*)\omega^*}{(1+\gamma^*)\sigma^*}\mathbb{E}_t\Delta\hat{\tau}_{t+1} + \frac{\left(1-\rho_p^*\right)(1-\gamma^*)}{(1+\gamma^*)\sigma^*}\hat{\varepsilon}_{p,t}^* \tag{35}$$

- Foreign Phillips curve:

$$\hat{\pi}_{F,t}^* = \frac{\xi^*}{1+\beta\xi^*}\hat{\pi}_{F,t-1}^* + \frac{\beta}{1+\beta\xi^*}\mathbb{E}_t\hat{\pi}_{F,t+1}^* + \frac{(1-\beta\alpha^*)(1-\alpha^*)}{(1+\beta\xi^*)\alpha^*}\widehat{mc}_t^* \tag{36}$$

- Foreign marginal cost:

$$\widehat{mc}_t^* = \left(\frac{\sigma^*}{1-\gamma^*} + \varphi^*(1-\omega^*)s^*\right)\hat{c}_t^* - \frac{\gamma^*\sigma^*}{1-\gamma^*}\hat{c}_{t-1}^* + \varphi^*[1-(1-\omega^*)s^*]\hat{c}_t - (1+\varphi^*)\hat{\varepsilon}_{a,t}^* - [\omega[1+\varphi^*(1-(1-\omega^*)s^*)] + \omega^*\varphi^*(1-\omega^*)s^*]\hat{\tau}_t \tag{37}$$

- Foreign aggregate output:

$$\hat{y}_t^* = [1-(1-\omega^*)s^*]\hat{c}_t + (1-\omega^*)s^*\hat{c}_t^* - (\omega[1-(1-\omega^*)s^*] + \omega^*(1-\omega^*)s^*)\hat{\tau}_t \tag{38}$$

- Foreign preference shock:

$$\hat{\varepsilon}_{p,t}^* = \rho_p^*\hat{\varepsilon}_{p,t-1}^* + \eta_{p,t}^* \tag{39}$$

- Foreign productivity shock:

$$\hat{\varepsilon}_{a,t}^* = \rho_a^*\hat{\varepsilon}_{a,t-1}^* + \eta_{a,t}^* \tag{40}$$

- Terms of trade:

$$\hat{\tau}_t = \frac{1}{\omega-\omega^*}\left[\frac{\sigma}{1-\gamma}\hat{c}_t - \frac{\gamma\sigma}{1-\gamma}\hat{c}_{t-1} - \frac{\sigma^*}{1-\gamma^*}\hat{c}_t^* + \frac{\gamma^*\sigma^*}{1-\gamma^*}\hat{c}_{t-1}^* + \hat{\varepsilon}_{p,t}^* - \hat{\varepsilon}_{p,t}\right] \tag{41}$$

Notice that $s = \bar{C}/\bar{Y}$ and $s^* = \bar{C}^*/\bar{Y}^*$.

Taylor-Type Rules

- Home monetary policy rule:

$$\hat{\imath}_t = \psi_i \hat{\imath}_{t-1} + (1 - \psi_i)\left[\psi_\pi \hat{\pi}_{H,t} + \psi_y\left(\hat{y}_t - \hat{y}_t^n\right)\right] + \hat{\varepsilon}_{i,t} \tag{42}$$

- Foreign monetary policy rule:

$$\hat{\imath}_t^* = \psi_i^* \hat{\imath}_{t-1}^* + \left(1 - \psi_i^*\right)\left[\psi_\pi^* \hat{\pi}_{F,t}^* + \psi_y^*\left(\hat{y}_t^* - \hat{y}_t^{*n}\right)\right] + \hat{\varepsilon}_{i,t}^* \tag{43}$$

- Home monetary policy shock:

$$\hat{\varepsilon}_{i,t} = \rho_i \hat{\varepsilon}_{i,t-1} + \eta_{i,t} \tag{44}$$

- Foreign monetary policy shock:

$$\hat{\varepsilon}_{i,t}^* = \rho_i^* \hat{\varepsilon}_{i,t-1}^* + \eta_{i,t}^* \tag{45}$$

Appendix 2. The Log-Linearized Flexible-Price Output

The so-called natural output is obtained as the level of output that would prevail under flexible prices in the absence of cost-push shocks. In this case, the optimal pricing decision for the firm h—i.e., the price that would maximize profits at each period—is given by

$$P_{H,t}(h) = \frac{\mu}{(1+\vartheta)} \frac{W_t}{\varepsilon_{a,t}},$$

where $\mu \equiv \theta/(\theta - 1)$ is the optimal markup and ϑ is the subsidy for output that offsets the effect on imperfect competition in goods markets on the steady-state level of output. Using the demand for good h, $Y_t(h) = \left(\frac{P_{H,t}(h)}{P_{H,t}}\right)^{-\theta} Y_t$, we note that the relative supply of good h must in turn satisfy

$$\left(\frac{Y_t(h)}{Y_t}\right)^{-1/\theta} = \frac{\mu}{(1+\vartheta)} \frac{W_t}{P_{H,t}\varepsilon_{a,t}}.$$

Note also that, in steady state,

$$\frac{\bar{\mathcal{U}}_{\bar{L}}}{\bar{\mathcal{U}}_{\bar{C}}} = \frac{(1+\vartheta)}{\mu} = 1.$$

Because all wages are the same in the case of flexible wages, we have $W_t(h) = W_t$ and $L_t(h) = L_t$ for all h. This implies that all sellers supply a quantity Y_t^n satisfying

$$1 = \frac{\mathcal{U}_{L^n,t}}{\mathcal{U}_{C^n,t}} \frac{P_t}{P_{H,t}} \frac{1}{\varepsilon_{a,t}} = \frac{\left(L_t^n\right)^{\varphi}}{\left(C_t^n - \gamma C_t^n\right)^{-\sigma}} \frac{\left(\mathcal{T}_t^n\right)^{1-\omega}}{\varepsilon_{a,t}}$$
$$= \frac{\left(Y_t^n/\varepsilon_{a,t}\right)^{\varphi}}{\left(C_t^n - \gamma C_{t-1}^n\right)^{-\sigma}} \frac{\left(\mathcal{T}_t^n\right)^{1-\omega}}{\varepsilon_{a,t}}.$$

Log-linearizing this expression yields

$$\hat{y}_t^n = -\frac{\sigma}{(1-\gamma)\varphi}\hat{c}_t^n + \frac{\sigma\gamma}{(1-\gamma)\varphi}\hat{c}_{t-1}^n - \frac{(1-\omega)}{\varphi}\hat{\tau}_t^n + \frac{(1+\varphi)}{\varphi}\hat{\varepsilon}_{a,t}.$$

By using the terms-of-trade expression

$$\hat{\tau}_t^n = \frac{1}{\omega - \omega^*}\left[\frac{\sigma}{1-\gamma}\hat{c}_t^n - \frac{\gamma\sigma}{1-\gamma}\hat{c}_{t-1}^n - \frac{\sigma^*}{1-\gamma^*}\hat{c}_t^{*n} + \frac{\gamma^*\sigma^*}{1-\gamma^*}\hat{c}_{t-1}^{*n} + \hat{\varepsilon}_{p,t}^* - \hat{\varepsilon}_{p,t}\right]$$

with the definition of the aggregate output

$$\hat{y}_t^n = \omega s\hat{c}_t^n + (1-\omega s)\hat{c}_t^{n*} + [(1-\omega)\omega s + (1-\omega^*)(1-\omega s)]\hat{\tau}_t^n,$$

we obtain

$$\left(\frac{\sigma}{1-\gamma} + \varphi\omega s + \frac{\sigma\Psi}{(1-\gamma)}\right)\hat{c}_t^n = \frac{\gamma\sigma}{1-\gamma}(1+\Psi)\hat{c}_{t-1}^n - \left(\varphi(1-\omega s) - \frac{\sigma^*\Psi}{(1-\gamma^*)}\right)\hat{c}_t^{*n} - \frac{\gamma^*\sigma^*\Psi}{(1-\gamma^*)}\hat{c}_{t-1}^{*n} - \Psi\left(\hat{\varepsilon}_{p,t}^* - \hat{\varepsilon}_{p,t}\right) + (1+\varphi)\hat{\varepsilon}_{a,t} \tag{46}$$

and

$$\hat{y}_t^n = \left(\omega s + \frac{\sigma\tilde{\Psi}}{(1-\gamma)}\right)\hat{c}_t^n - \left(\frac{\gamma\sigma\tilde{\Psi}}{(1-\gamma)}\right)\hat{c}_{t-1}^n + \left(1 - \omega s - \frac{\sigma^*\tilde{\Psi}}{(1-\gamma^*)}\right)\hat{c}_t^{*n} + \left(\frac{\gamma^*\sigma^*\tilde{\Psi}}{(1-\gamma^*)}\right)\hat{c}_{t-1}^{*n} + \tilde{\Psi}\left(\hat{\varepsilon}_{p,t}^* - \hat{\varepsilon}_{p,t}\right), \tag{47}$$

where $\Psi = [(1-\omega)(1+\varphi\omega s)+\varphi(1-\omega^*)(1-\omega s)]/(\omega-\omega^*)$ and $\tilde{\Psi} = [(1-\omega)\omega s+(1-\omega^*)(1-\omega s)]/(\omega-\omega^*)$.

The same calculations for the foreign country yield

$$\begin{aligned}\left(\frac{\sigma^*}{1-\gamma^*}+\varphi^*(1-\omega^*)s^*+\frac{\sigma^*\Psi^*}{(1-\gamma^*)}\right)\hat{c}_t^{*n} &= \frac{\gamma^*\sigma^*}{1-\gamma^*}(1+\Psi^*)\hat{c}_{t-1}^{*n}\\ &- \left(\varphi^*[1-(1-\omega^*)s^*]-\frac{\sigma\Psi^*}{(1-\gamma)}\right)\hat{c}_t^n-\frac{\gamma\sigma\Psi^*}{(1-\gamma)}\hat{c}_{t-1}^n\\ &+ \Psi^*\left(\hat{\varepsilon}_{p,t}^*-\hat{\varepsilon}_{p,t}\right)+(1+\varphi^*)\hat{\varepsilon}_{a,t}^* \end{aligned} \tag{48}$$

and

$$\begin{aligned}\hat{y}_t^{*n} = &\left[1-(1-\omega^*)s^*-\frac{\sigma\tilde{\Psi}^*}{(1-\gamma)}\right]\hat{c}_t^n+\frac{\gamma\sigma\tilde{\Psi}^*}{(1-\gamma)}\hat{c}_{t-1}^n\\ &+\left((1-\omega^*)s^*+\frac{\sigma^*\tilde{\Psi}^*}{(1-\gamma^*)}\right)\hat{c}_t^{*n}-\frac{\gamma^*\sigma^*\tilde{\Psi}^*}{(1-\gamma^*)}\hat{c}_{t-1}^{*n}\\ &-\tilde{\Psi}^*\left(\hat{\varepsilon}_{p,t}^*-\hat{\varepsilon}_{p,t}\right), \end{aligned} \tag{49}$$

where $\Psi^* = [\omega[1+\varphi^*(1-(1-\omega^*)s^*)]+\omega^*\varphi^*(1-\omega^*)s^*]/(\omega-\omega^*)$ and $\tilde{\Psi}^* = (\omega[1-(1-\omega^*)s^*]+\omega^*(1-\omega^*)s^*)/(\omega-\omega^*)$.

Appendix 3. Approximation of the Welfare Criterion

The second-order approximation of the home representative household's utility is derived in this section, using methods discussed in more detail in Woodford (2003). The average utility flow of the representative household at date t is given by

$$\mathbb{W}_t = \mathbb{U}(C_t,\mathcal{H}_t,\varepsilon_{p,t})-\frac{1}{n}\int_0^n \mathbb{V}(L_t(h),\varepsilon_{p,t})\mathrm{d}h, \tag{50}$$

where

$$\mathbb{U}(C_t,\mathcal{H}_t,\varepsilon_{p,t}) = \frac{\varepsilon_{p,t}}{1-\sigma}(C_t-\gamma\mathcal{H}_t)^{1-\sigma} \quad \text{and}$$
$$\mathbb{V}(L_t(h),\varepsilon_{p,t}) = \frac{\varepsilon_{p,t}}{1+\varphi}(L_t(h))^{1+\varphi}.$$

Taylor Expansion of the Utility Function

The second-order Taylor expansion of $\mathbb{U}(C_t, \mathcal{H}_t, \varepsilon_{p,t})$ around the steady state $\bar{\mathbb{U}} = \mathbb{U}(\bar{C}, \bar{\mathcal{H}}, \bar{\varepsilon}_p)$ yields

$$\begin{aligned}\mathbb{U}(C_t, \mathcal{H}_t, \varepsilon_{p,t}) \approx \bar{\mathbb{U}} + \bar{\mathbb{U}}_{\bar{C}}\tilde{C}_t + \bar{\mathbb{U}}_{\bar{\mathcal{H}}}\tilde{\mathcal{H}}_t + \bar{\mathbb{U}}_{\bar{\varepsilon}_p}\tilde{\varepsilon}_{p,t} + \frac{1}{2}\bar{\mathbb{U}}_{\bar{C}\bar{C}}\tilde{C}_t^2 \\ + \frac{1}{2}\bar{\mathbb{U}}_{\bar{\mathcal{H}}\bar{\mathcal{H}}}\tilde{\mathcal{H}}_t^2 + \frac{1}{2}\bar{\mathbb{U}}_{\bar{\varepsilon}_p\bar{\varepsilon}_p}(\tilde{\varepsilon}_{p,t})^2 + \bar{\mathbb{U}}_{\bar{C}\bar{\mathcal{H}}}\tilde{C}_t\tilde{\mathcal{H}}_t \\ + \bar{\mathbb{U}}_{\bar{C}\bar{\varepsilon}_p}\tilde{C}_t\tilde{\varepsilon}_{p,t} + \bar{\mathbb{U}}_{\bar{\mathcal{H}}\bar{\varepsilon}_p}\tilde{\mathcal{H}}_t\tilde{\varepsilon}_{p,t} + \mathcal{O}(\|\zeta\|^3), \qquad (51)\end{aligned}$$

where $\tilde{X}_t = X_t - \bar{X}$, $\mathcal{O}(\|\zeta\|^3)$ denotes the order of residual and $\|\zeta\|$ is a bound on the amplitude of exogenous disturbances.

Applying a second-order Taylor expansion ($\tilde{X}_t/\bar{X} = \hat{x}_t + \frac{1}{2}\hat{x}_t^2 + \mathcal{O}(\|\zeta\|^3)$, where $\hat{x}_t = \ln X_t - \ln \bar{X}$), we obtain

$$\begin{aligned}\mathbb{U}(C_t, \mathcal{H}_t, \varepsilon_{p,t}) \approx \bar{\mathbb{U}} + \bar{\mathbb{U}}_{\bar{C}}\bar{C}\left(\hat{c}_t + \frac{1}{2}\hat{c}_t^2\right) + \bar{\mathbb{U}}_{\bar{\mathcal{H}}}\bar{\mathcal{H}}\left(\hat{h}_t + \frac{1}{2}\hat{h}_t^2\right) \\ + \bar{\mathbb{U}}_{\bar{\varepsilon}_p}\bar{\varepsilon}_p\left(\hat{\varepsilon}_{p,t} + \frac{1}{2}\hat{\varepsilon}_{p,t}^2\right) + \frac{1}{2}\bar{\mathbb{U}}_{\bar{C}\bar{C}}\bar{C}^2\hat{c}_t^2 \\ + \frac{1}{2}\bar{\mathbb{U}}_{\bar{\mathcal{H}}\bar{\mathcal{H}}}\bar{\mathcal{H}}^2\hat{h}_t^2 + \frac{1}{2}\bar{\mathbb{U}}_{\bar{\varepsilon}_p\bar{\varepsilon}_p}\bar{\varepsilon}_p^2\hat{\varepsilon}_{p,t}^2 + \bar{\mathbb{U}}_{\bar{C}\bar{\mathcal{H}}}\bar{C}\bar{\mathcal{H}}(\hat{c}_t\hat{h}_t) \\ + \bar{\mathbb{U}}_{\bar{C}\bar{\varepsilon}_p}\bar{C}\bar{\varepsilon}_p(\hat{c}_t\hat{\varepsilon}_{p,t}) + \bar{\mathbb{U}}_{\bar{\mathcal{H}}\bar{\varepsilon}_p}\bar{\mathcal{H}}\bar{\varepsilon}_p(\hat{h}_t\hat{\varepsilon}_{p,t}) + \mathcal{O}(\|\zeta\|^3) \\ (52)\end{aligned}$$

with

$$\begin{aligned}&\bar{\mathbb{U}}_{\bar{C}} = \bar{\varepsilon}_p(\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma}, \\ &\bar{\mathbb{U}}_{\bar{C}\bar{C}} = -\sigma\bar{\varepsilon}_p(\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma-1} = \tfrac{-\sigma}{(\bar{C}-\gamma\bar{\mathcal{H}})}\bar{\mathbb{U}}_{\bar{C}}, \\ &\bar{\mathbb{U}}_{\bar{\mathcal{H}}} = -\gamma\bar{\varepsilon}_p(\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma} = -\gamma\bar{\mathbb{U}}_{\bar{C}}, \\ &\bar{\mathbb{U}}_{\bar{\mathcal{H}}\bar{\mathcal{H}}} = -\gamma^2\sigma\bar{\varepsilon}_p(\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma-1} = \tfrac{-\gamma^2\sigma}{(\bar{C}-\gamma\bar{\mathcal{H}})}\bar{\mathbb{U}}_{\bar{C}}, \\ &\bar{\mathbb{U}}_{\bar{C}\bar{\mathcal{H}}} = \sigma\gamma\bar{\varepsilon}_p(\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma-1} = \tfrac{\sigma\gamma}{(\bar{C}-\gamma\bar{\mathcal{H}})}\bar{\mathbb{U}}_{\bar{C}}, \\ &\bar{\mathbb{U}}_{\bar{\varepsilon}_p} = \tfrac{1}{1-\sigma}(\bar{C} - \gamma\bar{\mathcal{H}})^{1-\sigma} = \tfrac{(\bar{C}-\gamma\bar{\mathcal{H}})}{(1-\sigma)\bar{\varepsilon}_p}\bar{\mathbb{U}}_{\bar{C}}, \\ &\bar{\mathbb{U}}_{\bar{\varepsilon}_p\bar{\varepsilon}_p} = 0, \\ &\bar{\mathbb{U}}_{\bar{C}\bar{\varepsilon}_p} = (\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma} = \tfrac{\bar{\mathbb{U}}_{\bar{C}}}{\bar{\varepsilon}_p}, \\ &\bar{\mathbb{U}}_{\bar{\mathcal{H}}\bar{\varepsilon}_p} = -\gamma(\bar{C} - \gamma\bar{\mathcal{H}})^{-\sigma} = \tfrac{-\gamma}{\bar{\varepsilon}_p}\bar{\mathbb{U}}_{\bar{C}}.\end{aligned}$$

Replacing $\mathcal{H}_t$ with C_{t-1}, the utility of consumption simplifies to

$$\begin{aligned}\mathbb{U}(C_t, C_{t-1}, \varepsilon_{p,t}) \approx \bar{\mathbb{U}}_{\bar{C}}\bar{C}\Bigg\{&(\hat{c}_t - \gamma\hat{c}_{t-1}) + \frac{1}{2}\left(\hat{c}_t^2 - \gamma\hat{c}_{t-1}^2\right)\\ &-\frac{\sigma}{2(1-\gamma)}(\hat{c}_t - \gamma\hat{c}_{t-1})^2 + \hat{c}_t\hat{\varepsilon}_{p,t} - \gamma\hat{c}_{t-1}\hat{\varepsilon}_{p,t}\Bigg\}\\ &+ t.i.p. + \mathcal{O}(\|\zeta\|^3), \qquad (53)\end{aligned}$$

where $t.i.p.$ denotes terms independent of the actual policy, such as constant terms involving only exogenous variables.

Taylor Expansion of the Disutility of Work

The second-order Taylor expansion for $\mathbb{V}(L_t(h), \varepsilon_{p,t})$ around the steady state $\bar{\mathbb{V}} = \mathbb{V}(\bar{L}, \bar{\varepsilon}_p)$ is

$$\begin{aligned}\mathbb{V}(L_t(h),\varepsilon_{p,t}) \approx \bar{\mathbb{V}} &+ \bar{\mathbb{V}}_{\bar{L}}\bar{L}\left(\hat{l}_t(h) + \frac{1}{2}\hat{l}_t^2(h)\right) + \bar{\mathbb{V}}_{\bar{\varepsilon}_p}\bar{\varepsilon}_p\left(\hat{\varepsilon}_{p,t} + \frac{1}{2}(\hat{\varepsilon}_{p,t})^2\right)\\ &+ \frac{1}{2}\bar{\mathbb{V}}_{\bar{L}\bar{L}}\bar{L}^2\hat{l}_t^2(h) + \frac{1}{2}\bar{\mathbb{V}}_{\bar{\varepsilon}_p\bar{\varepsilon}_p}\bar{\varepsilon}_p^2\hat{\varepsilon}_{p,t}^2 + \bar{\mathbb{V}}_{\bar{L}\bar{\varepsilon}_p}\bar{L}\bar{\varepsilon}_p(\hat{l}_t(h)\hat{\varepsilon}_{p,t})\\ &+ \mathcal{O}(\|\zeta\|^3) \qquad (54)\end{aligned}$$

with

$$\begin{aligned}&\bar{\mathbb{V}}_{\bar{L}} = \bar{\varepsilon}_p\bar{L}^{\varphi},\\ &\bar{\mathbb{V}}_{\bar{\varepsilon}_p} = \tfrac{1}{1+\varphi}(\bar{L})^{1+\varphi} = \tfrac{\bar{L}}{(1+\varphi)\bar{\varepsilon}_p}\bar{\mathbb{V}}_{\bar{L}},\\ &\bar{\mathbb{V}}_{\bar{L}\bar{L}} = \varphi\bar{\varepsilon}_p\bar{L}^{\varphi-1} = \tfrac{\varphi}{\bar{L}}\bar{\mathbb{V}}_{\bar{L}},\\ &\bar{\mathbb{V}}_{\bar{\varepsilon}_p\bar{\varepsilon}_p} = 0,\\ &\bar{\mathbb{V}}_{\bar{L}\bar{\varepsilon}_p} = \bar{L}^{\varphi} = \tfrac{\bar{\mathbb{V}}_{\bar{L}}}{\bar{\varepsilon}_p}.\end{aligned}$$

The disutility of work becomes

$$\begin{aligned}\mathbb{V}(L_t(h), \varepsilon_{p,t}) \approx \bar{\mathbb{V}}_{\bar{L}}\bar{L}\left\{\hat{l}_t(h) + \frac{1+\varphi}{2}\hat{l}_t^2(h) + \hat{l}_t(h)\hat{\varepsilon}_{p,t}\right\}\\ + t.i.p. + \mathcal{O}(\|\zeta\|^3). \qquad (55)\end{aligned}$$

Individual Labor to Composite Labor

Now define the composite labor index:

$$L_t = \int_0^n L_t(h)\mathrm{d}h = \int_0^n \frac{Y_t(h)}{\varepsilon_{a,t}}\mathrm{d}h = \frac{Y_t}{\varepsilon_{a,t}} \int_0^n \left(\frac{\tilde{P}_t(h)}{P_t}\right)^{-\theta} \mathrm{d}h.$$

Taking a second-order Taylor expansion of the logarithm of this equation yields

$$\hat{l}_t = \hat{y}_t - \hat{\varepsilon}_{a,t} + \hat{u}_t, \tag{56}$$

where $\hat{u}_t = \ln \int_0^n \left(\frac{\tilde{P}_t(h)}{P_t}\right)^{-\theta} \mathrm{d}h$ is of second order. As shown by Woodford (2003, chap. 6), one has

$$\hat{u}_t = \frac{\theta\alpha}{2(1-\alpha)(1-\beta\alpha)}(\hat{\pi}_{H,t} - \xi\hat{\pi}_{H,t-1})^2 + \mathcal{O}(\|\zeta\|^3). \tag{57}$$

Welfare Expressions

We first integrate equation (55) over h and replace $\int_0^n L_t(h)\mathrm{d}h$ and $\hat{u}_t$ with their respective expressions. We then take the present discounted sum of equations (53) and (55) and subtract the second expression from the first one to obtain

$$\begin{aligned}\sum_{t=0}^{\infty}\beta^t \mathbb{W}_t = \bar{\mathbb{U}}_{\bar{C}}\bar{C}\sum_{t=0}^{\infty}\beta^t \Bigg\{ &(\hat{c}_t - \gamma\hat{c}_{t-1}) + \frac{1}{2}\left(\hat{c}_t^2 - \gamma\hat{c}_{t-1}^2\right) \\ &- \frac{\sigma}{2(1-\gamma)}(\hat{c}_t - \gamma\hat{c}_{t-1})^2 + \hat{c}_t\hat{\varepsilon}_{p,t} - \gamma\hat{c}_{t-1}\hat{\varepsilon}_{p,t} - s^{-1}\hat{y}_t \\ &- \frac{1+\varphi}{2s}(\hat{y}_t - \hat{\varepsilon}_{a,t})^2 - s^{-1}\hat{y}_t\hat{\varepsilon}_{p,t} - \frac{\theta\alpha}{2(1-\alpha)(1-\beta\alpha)s} \\ &\times (\hat{\pi}_{H,t} - \xi\hat{\pi}_{H,t-1})^2 \Bigg\} + t.i.p. + \mathcal{O}(\|\zeta\|^3). \end{aligned} \tag{58}$$

Recall that $\bar{\mathbb{V}}_{\bar{L}} = \bar{\mathbb{U}}_{\bar{C}}$ and $s = \bar{C}/\bar{Y}$. Given that

$$\sum_{t=0}^{\infty}\beta^t x_{t-1} = x_{-1} + \beta\sum_{t=0}^{\infty}\beta^t x_t = \beta\sum_{t=0}^{\infty}\beta^t x_t + t.i.p.$$

and using the fact that

$$(1+\varphi)\hat{\varepsilon}_{a,t} = A_1\hat{c}_t^n + A_2\hat{c}_{t-1}^n + A_3\hat{c}_t^{*n} + A_4\hat{c}_{t-1}^{*n} + A_5\hat{\varepsilon}_{p,t} + A_6\hat{\varepsilon}_{p,t}^*,$$

where parameters A_j $(j = 1, \cdots, 6)$ find their counterparts in equation (46), it yields

$$\begin{aligned}
\mathcal{W}_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \mathbb{W}_t = -\bar{\mathbb{U}}_{\bar{C}}\bar{C}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Big\{ & -(1-\beta\gamma)\hat{c}_t + s^{-1}\hat{y}_t \\
& + \frac{\sigma}{2(1-\gamma)}(\hat{c}_t - \gamma\hat{c}_{t-1})^2 - \frac{(1-\beta\gamma)}{2}\hat{c}_t^2 + \frac{1+\varphi}{2s}\hat{y}_t^2 \\
& + (\gamma\beta\rho_p - 1)\hat{c}_t\hat{\varepsilon}_{p,t} \\
& - s^{-1}\big(A_1\hat{c}_t^n + A_2\hat{c}_{t-1}^n + A_3\hat{c}_t^{*n} + A_4\hat{c}_{t-1}^{*n} + A_5\hat{\varepsilon}_{p,t} + A_6\hat{\varepsilon}_{p,t}^*\big)\hat{y}_t \\
& + s^{-1}\hat{y}_t\hat{\varepsilon}_{p,t} + \frac{\theta\alpha}{2(1-\alpha)(1-\beta\alpha)s}(\hat{\pi}_{H,t} - \xi\hat{\pi}_{H,t-1})^2 \Big\} \\
& + t.i.p. + \mathcal{O}(\|\zeta\|^3). \qquad (59)
\end{aligned}$$

Finally, replacing the cross-product $x_{1,t}x_{2,t}$ with $(x_{1,t}^2 + x_{2,t}^2 - (x_{1,t} - x_{2,t})^2)/2$, we can rewrite the home welfare criterion as

$$\begin{aligned}
\mathcal{W}_0 = -\frac{\bar{\mathbb{U}}_{\bar{C}}\bar{C}}{2}\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Big\{ & (\hat{c}_t - \Psi_c)^2 + \frac{(1+\hat{y}_t)^2}{s} \\
& + (\beta\gamma(1+\rho_p) - 3)\hat{c}_t^2 \\
& + \frac{1+\varphi - A_1 - A_2 - A_3 - A_4 - A_5 - A_6}{s}\hat{y}_t^2 \\
& + \frac{\sigma}{(1-\gamma)}(\hat{c}_t - \gamma\hat{c}_{t-1})^2 - (\beta\gamma\rho_p - 1)(\hat{c}_t - \hat{\varepsilon}_{p,t})^2 \\
& + \frac{A_1}{s}\big(\hat{y}_t - \hat{c}_t^n\big)^2 + \frac{A_2}{s}\big(\hat{y}_t - \hat{c}_{t-1}^n\big)^2 + \frac{A_3}{s}\big(\hat{y}_t - \hat{c}_t^{*n}\big)^2 \\
& + \frac{A_4}{s}\big(\hat{y}_t - \hat{c}_{t-1}^{*n}\big)^2 - \frac{1-A_5}{s}(\hat{y}_t - \hat{\varepsilon}_{p,t})^2 + \frac{A_6}{s}\big(\hat{y}_t - \hat{\varepsilon}_{p,t}^*\big)^2 \\
& + \frac{\theta\alpha}{(1-\alpha)(1-\beta\alpha)s}(\hat{\pi}_{H,t} - \xi\hat{\pi}_{H,t-1})^2 \Big\} + t.i.p. + \mathcal{O}(\|\eta\|^3),
\end{aligned}$$

where $\Psi_c = (1-\beta\gamma)$.

The calculations for the welfare of the foreign representative household yield

$$
\begin{aligned}
\mathcal{W}_0^* = -\frac{\bar{\mathbb{U}}_{\bar{C}^*}^* \bar{C}^*}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ & \left(\hat{c}_t^* - \Psi_c^*\right)^2 + \frac{\left(1+\hat{y}_t^*\right)^2}{s^*} \\
& + \left(\beta\gamma^*\left(1+\rho_p^*\right) - 3\right)\hat{c}_t^{*2} \\
& + \frac{1+\varphi^* - A_1^* - A_2^* - A_3^* - A_4^* - A_5^* - A_6^*}{s^*}\hat{y}_t^{*2} \\
& + \frac{\sigma^*\gamma^*}{(1-\gamma^*)}\left(\hat{c}_t^* - \gamma^*\hat{c}_{t-1}^*\right)^2 - \left(\beta\gamma^*\rho_p^* - 1\right)\left(\hat{c}_t^* - \hat{\varepsilon}_{p,t}^*\right)^2 \\
& + \frac{A_1^*}{s^*}\left(\hat{y}_t^* - \hat{c}_t^n\right)^2 + \frac{A_2^*}{s^*}\left(\hat{y}_t^* - \hat{c}_{t-1}^n\right)^2 + \frac{A_3^*}{s^*}\left(\hat{y}_t^* - \hat{c}_t^{*n}\right)^2 \\
& + \frac{A_4^*}{s^*}\left(\hat{y}_t^* - \hat{c}_{t-1}^{*n}\right)^2 + \frac{A_5^*}{s^*}\left(\hat{y}_t^* - \hat{\varepsilon}_{p,t}\right)^2 - \frac{1-A_6^*}{s^*}\left(\hat{y}_t^* - \hat{\varepsilon}_{p,t}^*\right)^2 \\
& + \frac{\theta\alpha^*}{(1-\alpha^*)(1-\beta\alpha^*)s^*}\left(\hat{\pi}_{F,t}^* - \xi^*\hat{\pi}_{F,t-1}^*\right)^2 \Bigg\} + t.i.p. + \mathcal{O}(\|\eta\|^3),
\end{aligned}
$$

where $\Psi_c^* = (1-\beta\gamma^*)$.

References

Abel, A. 1990. "Asset Prices under Habit Formation and Catching Up with the Joneses." *American Economic Review* 80 (2): 38–42.

Aksoy, Y., P. De Grauwe, and H. Dewachter. 2002. "Do Asymmetries Matter for European Monetary Policy?" *European Economic Review* 46 (3): 443–69.

Altissimo, F., B. Mojon, and P. Zaffaroni. 2007. "Fast Micro and Slow Macro: Can Aggregation Explain the Persistence of Inflation?" ECB Working Paper No. 729.

Amato, J. D., and T. S. Laubach. 2004. "Implications of Habit Formation for Optimal Monetary Policy." *Journal of Monetary Economics* 51 (2): 305–25.

Angelini, P., P. Del Giovane, S. Siviero, and D. Terlizzese. 2002. "Monetary Policy Rules for the Euro Area: What Role for National Information?" Working Paper No. 457, Banca d'Italia.

Benigno, G. 2004. "Real Exchange Rate Persistence and Monetary Policy Rules." *Journal of Monetary Economics* 51 (3): 473–502.

Benigno, G., and P. Benigno. 2003. "Price Stability in Open Economies." *Review of Economic Studies* 70 (4): 743–64.

Benigno, P. 2004. "Optimal Monetary Policy in a Currency Area." *Journal of International Economics* 63 (2): 293–320.

Benigno, P., and J. D. López-Salido. 2006. "Inflation Persistence and Optimal Monetary Policy in the Euro Area." *Journal of Money, Credit, and Banking* 38 (3): 587–614.

Calvo, G. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan. 2002. "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?" *Review of Economic Studies* 69 (3): 533–63.

Christiano, L. J., M. Eichenbaum, and C. L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45.

Clarida, R., J. Galí, and M. Gertler. 2002. "A Simple Framework for International Monetary Policy Analysis." *Journal of Monetary Economics* 49 (5): 879–904.

Corsetti, G., and P. Pesenti. 2001. "Welfare and Macroeconomic Interdependence." *Quarterly Journal of Economics* 116 (2): 421–46.

———. 2005. "International Dimensions of Optimal Monetary Policy." *Journal of Monetary Economics* 52 (2): 281–305.

De Grauwe, P. 2000. "Monetary Policies in the Presence of Asymmetries." *Journal of Common Market Studies* 38 (4): 593–612.

De Grauwe, P., and T. Piskorski. 2001. "Union-Wide Aggregates versus National Data Based Monetary Policies: Does It Matter for the Eurosystem?" CEPR Discussion Paper No. 3036.

Demertzis, M., and A. Hughes Hallett. 1998. "Asymmetric Transmission Mechanisms and the Rise in European Unemployment: A Case of Structural Differences or of Policy Failures?" *Journal of Economic Dynamics and Control* 22 (6): 869–86.

Devereux, M. B., and C. Engel. 2003. "Monetary Policy in the Open Economy Revisited: Price Setting and Exchange-Rate Flexibility." *Review of Economic Studies* 70 (4): 765–83.

Dieppe, A., K. Küster, and P. McAdam. 2005. "Optimal Monetary Policy Rules for the Euro Area: An Analysis Using the Area Wide Model." *Journal of Common Market Studies* 43 (3): 507–37.

Erceg, C. J., D. W. Henderson, and A. T. Levin. 2000. "Optimal Monetary Policy with Staggered Wage and Price Contracts." *Journal of Monetary Economics* 46 (2): 281–313.

Fagan, G., J. Henry, and R. Mestre. 2005. "An Area-Wide Model for the Euro Area." *Economic Modelling* 22 (1): 39–59.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez. 2004. "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach." *Journal of Econometrics* 123 (1): 153–87.

Galí, J., and T. Monacelli. 2005. "Monetary Policy and Exchange Rate Volatility in a Small Open Economy." *Review of Economic Studies* 72 (3): 707–34.

Geweke, J. 1999. "Computational Experiments and Reality." Manuscript, University of Minnesota and Federal Reserve Bank of Minneapolis.

Giannoni, M. P., and M. Woodford. 2005. "Optimal Inflation-Targeting Rules." In *The Inflation Targeting Debate*, ed. B. S. Bernanke and M. Woodford, 93–162. Chicago: University of Chicago Press.

Granger, C. W. J. 1980. "Long Memory Relationships and the Aggregation of Dynamic Models." *Journal of Econometrics* 14 (2): 227–38.

Justiniano, A., and B. Preston. 2004. "Small Open Economy DSGE Models: Specification, Estimation and Model Fit." Working Paper. (Now circulated as "Monetary Policy and Uncertainty in an Empirical Small Open Economy Model." 2006.)

Monacelli, T. 2001. "New International Monetary Arrangements and the Exchange Rate." *International Journal of Finance and Economics* 6 (4): 389–400.

Monteforte, L., and S. Siviero. 2003. "Aggregate vs. Disaggregate Euro-Area Macro-Modelling." Manuscript, Banca d'Italia.

Obstfeld, M., and K. Rogoff. 1995. "Exchange Rate Dynamic Redux." *Journal of Political Economy* 103 (3): 624–60.

Onatski, A., and N. Williams. 2004. "Empirical and Policy Performance of a Forward-Looking Monetary Model." Manuscript, Columbia University.

Pytlarczyk, E. 2005. "An Estimated DSGE Model for the German Economy within the Euro Area." Deutsche Bundesbank Working Paper No. 33/2005.

Rotemberg, J. J., and M. Woodford. 1998. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: Expanded Version." NBER Technical Working Paper No. 233.

Sbordone, A. M. 2003. "A Limited Information Approach to the Simultaneous Estimation of Wage and Price Dynamics." Manuscript, Rutgers University.

Schmitt-Grohé, S., and M. Uribe. 2004. "Optimal Operational Monetary Policy in the Christiano-Eichenbaum-Evans Model of the U.S. Business Cycle." Manuscript, Duke University.

Schorfheide, F. 2003. "Labor-Supply Shifts and Economic Fluctuations." *Journal of Monetary Economics* 50 (8): 1751–68.

Smets, F., and R. Wouters. 2002. "Openness, Imperfect Exchange Rate Pass-Through and Monetary Policy." *Journal of Monetary Economy* 49 (5): 947–81.

———. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1 (5): 1123–75.

Taylor, J. B. 1999. *Monetary Policy Rules.* Chicago: University of Chicago Press.

Tchakarov, I. 2004. "The Gains from International Monetary Cooperation Revisited." International Monetary Fund Working Paper No. 04/1.

Warnock, F. E. 2000. "Idiosyncratic Tastes in a Two-Country Optimizing Model: Implications of a Standard Assumption." International Finance Discussion Paper No. 631, Board of Governors of the Federal Reserve System.

Woodford, M. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ: Princeton University Press.

Optimal and Simple Monetary Policy Rules with Zero Floor on the Nominal Interest Rate*

Anton Nakov
Banco de España

Recent treatments of the issue of a zero floor on nominal interest rates have been subject to some important methodological limitations. These include the assumption of perfect foresight or the introduction of the zero lower bound as an initial condition or a constraint on the variance of the interest rate, rather than an occasionally binding non-negativity constraint. This paper addresses these issues, offering a global solution to a standard dynamic stochastic sticky-price model with an explicit occasionally binding non-negativity constraint on the nominal interest rate. It turns out that the dynamics and sometimes the unconditional means of the nominal rate, inflation, and the output gap are strongly affected by uncertainty in the presence of the zero lower bound. Commitment to the optimal rule reduces unconditional welfare losses to around one-tenth of those achievable under discretionary policy, while constant price-level targeting delivers losses that are only 60 percent larger than those under the optimal rule. Even though the unconditional performance of simple instrument rules is almost unaffected by the presence of the zero lower bound, conditional on a strong deflationary shock, simple instrument rules perform substantially worse than the optimal policy.

JEL Codes: E31, E32, E37, E47, E52.

*I would like to thank Kosuke Aoki, Fabio Canova, Wouter den Haan, Jordi Galí, Albert Marcet, Ramon Marimon, Bruce Preston, Morten Ravn, Michael Reiter, Thijs van Rens, John Taylor, and an anonymous referee for helpful comments and suggestions. I am grateful also for comments to seminar participants at Universitat Pompeu Fabra, ECB, Banco de España, Central European University, Warwick University, Mannheim University, New Economic School, Rutgers

1. Introduction

An economy is said to be in a "liquidity trap" when the monetary authority cannot achieve a lower nominal interest rate in order to stimulate output. Such a situation can arise when the nominal interest rate has reached its zero lower bound (ZLB), below which nobody would be willing to lend, if money can be stored at no cost for a nominally riskless zero rate of return.

The possibility of a liquidity trap was first suggested by Keynes (1936) with reference to the Great Depression of the 1930s. At that time he compared the effectiveness of monetary policy in such a situation to trying to "push on a string." After WWII and especially during the high-inflation period of the 1970s, interest in the topic receded, and the liquidity trap was relegated to a hypothetical textbook example. As Krugman (1998) noticed, of the few modern papers that dealt with it, most concluded that "the liquidity trap can't happen, it didn't happen, and it won't happen again."

With the benefit of hindsight, however, it did happen, and to no less than Japan. Figure 1 illustrates this, showing the evolution of output, inflation, and the short-term nominal interest rate following the collapse of the Japanese real estate bubble of the late 1980s. The figure exhibits a persistent downward trend in all three variables and, in particular, the emergence of deflation since 1998 coupled with a zero nominal interest rate since 1999.

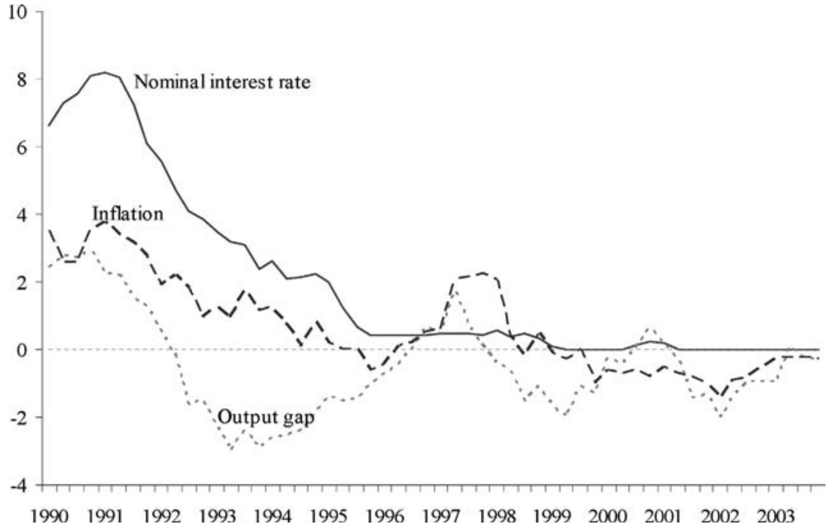
Motivated by the recent experience of Japan, the aim of the present paper is to contribute a quantitative analysis of the ZLB issue in a standard sticky-price model under alternative monetary policy regimes. On the one hand, the paper characterizes optimal monetary policy in the case of discretion and commitment.[1] On the other hand, it studies the performance of several simple monetary policy rules, modified to comply with the zero floor, relative to the optimal commitment policy. The analysis is carried out within

---

University, HEC Montreal, EBRD, Koc University, and Universidad de Navarra. Financial support from the Spanish Ministry of Foreign Affairs (Becas MAE) and the ECB are gratefully acknowledged.

[1]The part of the paper on optimal policy is similar to independent work by Adam and Billi (2006, 2007). The added value is to quantify and compare the performance of optimal commitment policy with that of a number of suboptimal rules in the same stochastic sticky-price setup.

Figure 1. Japan's Fall into a Liquidity Trap



a stochastic general equilibrium model with monopolistic competition and Calvo (1983) staggered price setting, under a standard calibration to the postwar U.S. economy.

The main findings are as follows: the optimal discretionary policy with zero floor involves a deflationary bias, which may be significant for certain parameter values and which implies that any quantitative analyses of discretionary biases of monetary policy that ignore the zero lower bound may be misleading. In addition, optimal discretionary policy implies much more aggressive cutting of the interest rate when the risk of deflation is high, compared with the corresponding policy without zero floor. Such a policy helps mitigate the depressing effect of private-sector expectations on current output and prices when the probability of falling into a liquidity trap is high.[2]

[2]An early version of this paper comparing the performance of optimal discretionary policy with three simple Taylor rules was circulated in 2004; optimal commitment policy and more simple rules were added in a version circulated in 2005. Optimal discretionary policy was studied independently by Adam and Billi (2004b), and optimal commitment policy by Adam and Billi (2004a).

In contrast, optimal commitment policy involves less preemptive lowering of the interest rate in anticipation of a liquidity trap, but it entails a promise for sustained monetary policy easing following an exit from a trap. This type of commitment enables the central bank to achieve higher expected (and actual) inflation and lower real rates in periods when the zero floor on nominal rates is binding.[3] As a result, under the baseline calibration, the expected welfare loss under commitment is only around one-tenth of the loss under optimal discretionary policy. This implies that the cost of discretion may be much higher than normally considered when abstracting from the zero-lower-bound issue.

The average welfare losses under simple instrument rules are eight to twenty times larger than those under the optimal rule. However, the bulk of these losses stem from the intrinsic suboptimality of simple instrument rules and not from the zero floor per se. This is related to the fact that under these rules the zero floor is hit very rarely—less than 1 percent of the time—compared with optimal commitment policy, which visits the liquidity trap one-third of the time. On the other hand, conditional on a large deflationary shock, the relative performance of simple instrument rules deteriorates substantially vis-à-vis the optimal commitment policy.

Issues of deflation and the liquidity trap have received considerable attention recently, especially after the experience of Japan.[4] In an influential article, Krugman (1998) argued that the liquidity trap boils down to a credibility problem in which private agents expect any monetary expansion to be reverted once the economy has recovered. As a solution, he suggested that the central bank should commit to a policy of high future inflation over an extended horizon.

More recently, Jung, Teranishi, and Watanabe (2005) have explored the effect of the zero lower bound in a standard sticky-price model with Calvo price setting under the assumption of perfect foresight. Consistent with Krugman (1998), they conclude that optimal

[3]This basic intuition was suggested already by Krugman (1998), based on a simpler model.

[4]A partial list of relevant studies includes Krugman (1998), Wolman (1998), McCallum (2000), Reifschneider and Williams (2000), Eggertsson and Woodford (2003), Klaeffling and Lopez-Perez (2003), Coenen, Orphanides, and Wieland (2004), Jung, Teranishi, and Watanabe (2005), Kato and Nishiyama (2005), and Adam and Billi (2006, 2007).

commitment policy entails a promise of a zero nominal interest for some time after the economy has recovered. Eggertsson and Woodford (2003) study optimal commitment policy with zero lower bound in a similar model in which the natural rate of interest is allowed to take two different values. In particular, it is assumed to become negative initially and then to jump to its "normal" positive level with a fixed probability in each period. These authors also conclude that the central bank should create inflationary expectations for the future. Importantly, they derive a moving price-level targeting rule that delivers the optimal commitment policy in this model.

One shortcoming of much of the modern literature on monetary policy rules is that it largely ignores the ZLB issue or at best uses rough approximations to address the problem. For instance, Rotemberg and Woodford (1997) introduce nominal rate targeting as an additional central bank objective, which ensures that the resulting path of the nominal rate does not violate the zero lower bound too often. In a similar vein, Schmitt-Grohe and Uribe (2004) exclude from their analysis instrument rules that result in a nominal rate with an average that is less than twice its standard deviation. In both cases, therefore, one might argue that for sufficiently large shocks that happen with a probability as high as 5 percent, the derived monetary policy rules are inconsistent with the zero lower bound.

On the other hand, of the few papers that do introduce an explicit non-negativity constraint on nominal interest rates, most simplify the stochastics of the model—e.g., by assuming perfect foresight (Jung, Teranishi, and Watanabe 2005) or a two-state low/high economy (Wolman 1998; Eggertsson and Woodford 2003). Even then, the zero lower bound is effectively imposed as an initial ("low") condition and not as an occasionally binding constraint.[5] While this assumption may provide a reasonable first pass at a quantitative analysis, it may be misleading to the extent that it ignores the occasionally binding nature of the zero interest rate floor.

Other studies (e.g., Coenen, Orphanides, and Wieland 2004) lay out a stochastic model but knowingly apply inappropriate solution techniques that rely on the assumption of certainty equivalence. It

[5]Namely, the zero floor binds for the first several periods, but once the economy transits to the "high" state, the ZLB never binds thereafter.

is well known that this assumption is violated in the presence of a nonlinear constraint such as the zero floor, but nevertheless these researchers have imposed it for reasons of tractability (admittedly, they work with a larger model than the one studied here). Yet forcing certainty equivalence in this case amounts to assuming that agents ignore the risk of the economy falling into a liquidity trap when making their optimal decisions.

The present study contributes to the above literature by solving numerically a stochastic general equilibrium model with monopolistic competition and sticky prices with an explicit occasionally binding zero lower bound, using an appropriate global solution technique that does not rely on certainty equivalence. It extends the analysis of Jung, Teranishi, and Watanabe (2005) to the stochastic case with an AR(1) process for the natural rate of interest.

After a brief outline of the basic framework adopted in the analysis (section 2), the paper characterizes and contrasts the optimal discretionary and optimal commitment policies (sections 3 and 4). It then analyzes the performance of a range of simple instrument and targeting rules (sections 5 and 6) consistent with the zero floor.[6] Sections 4–6 include a comparison of the conditional performance of all rules in a simulated liquidity trap, while section 7 presents their average performance, including a ranking according to unconditional expected welfare. Section 8 studies the sensitivity of the findings to various parameters of the model, as well as the implications of endogenous inflation persistence for the ZLB issue, and the last section concludes.

2. Baseline Model

While in principle the zero-lower-bound phenomenon can be studied in a model with flexible prices, it is with sticky prices that the liquidity trap becomes a real problem. The basic framework adopted in this study is a stochastic general equilibrium model with monopolistic competition and staggered price setting à la Calvo (1983) as

[6]These include truncated Taylor-type rules reacting to contemporaneous, expected, or past inflation, output gap, or price level, and with or without "interest rate smoothing"; truncated first-difference rules; the "augmented Taylor rule" by Reifschneider and Williams (2000); and flexible price-level targeting.

in Galí (2003) and Woodford (2003). In its simplest log-linearized version,[7] the model consists of three building blocks, describing the behavior of households, firms, and the monetary authority.

The first block, known as the "IS curve," summarizes the household's optimal consumption decision,

$$x_t = E_t x_{t+1} - \sigma\big(i_t - E_t \pi_{t+1} - r_t^n\big). \tag{1}$$

It relates the "output gap" x_t (i.e., the deviation of output from its flexible-price equilibrium) positively to the expected future output gap and negatively to the gap between the ex ante real interest rate, $i_t - E_t\pi_{t+1}$, and the "natural" (i.e., flexible-price equilibrium) real rate, r_t^n (which is observed by all agents at time t). Consumption smoothing accounts for the positive dependence of current output demand on expected future output demand, while intertemporal substitution implies the negative effect of the ex ante real interest rate. The interest rate elasticity of output, σ, corresponds to the inverse of the coefficient of relative risk aversion in the consumers' utility function.

The second building block of the model is a "Phillips curve"-type equation, which derives from the optimal price-setting decision of monopolistically competitive firms under the assumption of staggered price setting à la Calvo (1983),

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t, \tag{2}$$

where β is the time discount factor and κ, the "slope" of the Phillips curve, is related inversely to the degree of price stickiness.[8] Since firms are unable to adjust prices optimally every period, whenever

[7]It is important to note that, like in the studies cited in the introduction, the objective here is a modest one, in that the only source of nonlinearity in the model stems from the ZLB. Solving the fully nonlinear sticky-price model with Calvo (1983) contracts can be a worthwile enterprise; however, it increases the dimensionality of the computational problem by the number of states and co-states that one should keep track of (e.g., the measure of price dispersion and, in the case of optimal policy, the Lagrange multipliers associated with all forward-looking constraints).

[8]In the underlying sticky-price model, the slope κ is given by $[\theta(1+\varphi\varepsilon)]^{-1}(1-\theta)(1-\beta\theta)(\sigma^{-1}+\varphi)$, where θ is the fraction of firms that keep prices unchanged in each period, φ is the (inverse) wage elasticity of labor supply, and ε is the elasticity of substitution among differentiated goods.

they have the opportunity to do so, they choose to price goods as a markup over a weighted average of current and expected future marginal costs. Under appropriate assumptions on technology and preferences, marginal costs are proportional to the output gap, resulting in the above Phillips curve. Here this relation is assumed to hold exactly, ignoring the so-called cost-push shock, which sometimes is appended to generate a short-term trade-off between inflation and output-gap stabilization.

The final building block models the behavior of the monetary authority. The model assumes a "cashless-limit" economy in which the instrument controlled by the central bank is the nominal interest rate. One possibility is to assume a benevolent monetary policymaker seeking to maximize the welfare of households. In that case, as shown in Woodford (2003), the problem can be cast in terms of a central bank that aims to minimize (under discretion or commitment) the expected discounted sum of losses from output gaps and inflation, subject to the optimal behavior of households (1) and firms (2), and the zero nominal interest rate floor:

$$\underset{i_t,\pi_t,x_t}{\mathrm{Min}} \; E_0 \sum_{t=0}^{\infty} \beta^t \left(\pi_t^2 + \lambda x_t^2\right) \tag{3}$$

$$\text{s.t. } (1), (2)$$

$$i_t \geq 0, \tag{4}$$

where λ is the relative weight of the output gap in the central bank's loss function.[9]

An alternative way of modeling monetary policy is to assume that the central bank follows some sort of simple decision rule that relates the policy instrument, implicitly or explicitly, to other variables in

[9] Arguably, Woodford's (2003) approximation to the utility of the representative consumer is accurate to second order only in the vicinity of the steady state with zero inflation. To the extent that the shock inducing a zero interest rate pushes the economy far away from that steady state, the approximation error could in principle be large. In that case, the welfare evaluation in section 7 can be interpreted as a relative ranking of alternative policies based on an ad hoc loss criterion, under the assumption that the central bank targets zero inflation. Studying the welfare implication of different rules in the fully nonlinear model lies outside the scope of this paper.

the model. An example of such a rule, consistent with the zero floor, is a truncated Taylor rule,

$$i_t = \max[0, r^* + \pi^* + \phi_\pi(\pi_t - \pi^*) + \phi_x x_t], \tag{5}$$

where r^* is an equilibrium real rate, π^* is an inflation target, and ϕ_π and ϕ_x are response coefficients for inflation and the output gap.

To close the model, one needs to specify the behavior of the natural real rate. In the fuller model, the latter is a composite of a variety of real shocks, including shocks to preferences, government spending, and technology. Following Woodford (2003), here I assume that the natural real rate follows an exogenous mean-reverting process,

$$\widehat{r}_t^n = \rho \widehat{r}_{t-1}^n + \epsilon_t, \tag{6}$$

where $\widehat{r}_t^n \equiv r_t^n - r^*$ is the deviation of the natural real rate from its mean, r^*; ϵ_t are i.i.d. $\mathrm{N}(0, \sigma_\epsilon^2)$ real shocks; and $0 \leq \rho < 1$ is a persistence parameter.

The equilibrium conditions of the model therefore include the constraints (1), (2), and either a set of first-order optimality conditions (in the case of optimal policy) or a simple rule like (5). In either case the resulting system of equations cannot be solved with standard solution methods relying on local approximation because of the non-negativity constraint on the nominal rate. Hence I solve them with a global solution technique known as "collocation." The rational-expectations equilibrium with occasionally binding constraint is solved by way of parameterizing expectations (Christiano and Fischer 2000) and is implemented with the MATLAB routines developed by Miranda and Fackler (2002). The appendix outlines the simulation algorithm, while the following sections report the results.

2.1 Baseline Calibration

The model's parameters are chosen to be consistent with the "standard" Woodford (2003) calibration to the U.S. economy, which in turn is based on Rotemberg and Woodford (1997) (table 1). Thus, the slope of the Phillips curve (0.024), the weight of the output gap in the central bank loss function (0.003), the time discount factor (0.993), and the mean (3 percent per annum) and standard deviation (3.72 percent) of the natural real rate are all taken directly from

Table 1. Baseline Calibration (Quarterly Unless Otherwise Stated)

Structural Parameters		
Discount Factor	β	0.993
Real Interest Rate Elasticity of Output	σ	0.250
Slope of the Phillips Curve	κ	0.024
Weight of the Output Gap in Loss Function	λ	0.003
Natural Real-Rate Parameters		
Mean (% per Annum)	r^*	3%
Standard Deviation (Annual)	$\sigma(r^n)$	3.72%
Persistence (Quarterly)	ρ	0.65
Simple Instrument Rule Coefficients		
Inflation Target (% per Annum)	π^*	0%
Coefficient on Inflation	ϕ_π	1.5
Coefficient on Output Gap	ϕ_x	0.5
Interest-Rate-Smoothing Coefficient	ϕ_i	0

Woodford (2003). The persistence (0.65) of the natural real rate is assumed to be between the one used by Woodford (2003) (0.35) and that estimated by Adam and Billi (2006) (0.8) using a more recent sample period.[10] The real interest rate elasticity of aggregate demand (0.25)[11] is lower than the elasticity assumed by Eggertsson and Woodford (2003) (0.5), but as these authors point out, if anything, a lower degree of interest sensitivity of aggregate expenditure biases the results toward a more modest output contraction as a result of a binding zero floor.[12] In the simulations with simple rules, the baseline target inflation rate (0 percent) is consistent with the implicit zero target for inflation in the central bank's loss function.

[10]These parameters for the shock process imply that the natural real interest rate is negative about 15 percent of the time on an annual basis. This is slightly more often than with the standard Woodford (2003) calibration (10 percent).

[11]This corresponds to a constant relative risk aversion of 4 in the underlying model.

[12]With the Woodford (2003) value of this parameter (6.25), the model predicts unrealistically large output shortfalls when the zero floor binds—e.g., an output gap around –30 percent for values of the natural real rate around –3 percent.

The baseline reaction coefficients on inflation (1.5), the output gap (0.5), and the lagged nominal interest rate (0) are standard in the literature on Taylor (1993)-type rules. Section 8 studies the sensitivity of the results to various parameter changes.

3. Discretionary vs. Commitment Policy

Since the seminal work of Kydland and Prescott (1977) and Barro and Gordon (1983), the literature has focused on two (arguably extreme) ways of dealing with problems in which agents' expectations of future policy actions affect their current behavior. One is assuming full discretion, meaning that policymakers are unable to make any promises about their own (or their successors') future actions. The alternative is to suppose that policymakers have free access to a perfect commitment technology, which guarantees that they will never default on any of their past promises. While these two polar settings provide important insights into a wide variety of macroeconomic problems, their predictions sometimes differ considerably.

This turns out to be so in the context of the zero-lower-bound issue. In particular, this section shows that if the central bank cannot make any credible promises about the future course of monetary policy, then the zero lower bound is invariably associated with deflation. On the other hand, if the central bank is able to commit to the optimal state-contingent policy, then hitting the zero lower bound need not be associated with a falling price level, and may even result in slightly positive inflation. Intuitively, by committing to *future* inflation (once the zero lower bound ceases to bind), the central bank is able to reduce the real interest rate and stimulate demand at times when output is unusually low and the interest rate is constrained by the zero floor. At the same time, forward-looking price-setting behavior implies that some of the expected future inflation is built into current pricing decisions, which may result in slightly positive inflation even while the zero floor is binding. This way of affecting behavior is just unavailable to a discretionary policymaker; therefore, in the discretion case the private sector correctly anticipates that the zero lower bound will prevent the central bank from offsetting fully the effects of large-enough negative shocks on inflation.

3.1 Optimal Discretionary Policy

Abstracting from the zero floor, the solution to the discretionary optimization problem is well known (Clarida, Galí, and Gertler 1999).[13] Under discretion, the central bank cannot manipulate the beliefs of the private sector, and it takes expectations as given. The private sector is aware that the central bank is free to reoptimize its plan in each period; therefore, in a rational-expectations equilibrium, the central bank should have no incentives to change its plans in an unexpected way. In the baseline model with no endogenous state variables, the discretionary policy problem reduces to a sequence of static optimization problems in which the central bank minimizes current-period losses by choosing the current inflation, output gap, and nominal interest rate as a function only of the exogenous natural real rate, r_t^n.

The solution without zero bound then is straightforward: inflation and the output gap are fully stabilized at their (zero) targets in every period and state of the world, while the nominal interest rate moves one-for-one with the natural real rate. This is depicted by the dashed lines in figure 2. With this policy, the central bank is able to achieve the globally minimal welfare loss of zero at all times.

With the zero floor, the basic problem of discretionary optimization (without endogenous state variables) can still be cast as a sequence of static problems. The period-t Lagrangian is given by

$$\frac{1}{2}\left(\pi_t^2 + \lambda x_t^2\right) + \phi_{1t}[x_t - f_{1t} + \sigma(i_t - f_{2t})] \\ + \phi_{2t}[\pi_t - \kappa x_t - \beta f_{2t}] + \phi_{3t} i_t, \qquad (7)$$

where ϕ_{1t} is the Lagrange multiplier associated with the IS curve (1), ϕ_{2t} with the Phillips curve (2), and ϕ_{3t} with the zero constraint (4). The functions $f_{1t} = E_t(x_{t+1})$ and $f_{2t} = E_t(\pi_{t+1})$ are the private-sector expectations that the central bank takes as given. Noticing that $\phi_{3t} = -\sigma\phi_{1t}$, the Kuhn-Tucker conditions for this problem can be written as

$$\pi_t + \phi_{2t} = 0 \qquad (8)$$

$$\lambda x_t + \phi_{1t} - \kappa\phi_{2t} = 0 \qquad (9)$$

[13] In this section, attention is restricted to Markov-perfect equilibria only.

$$i_t \phi_{1t} = 0 \tag{10}$$

$$i_t \geq 0 \tag{11}$$

$$\phi_{1t} \geq 0. \tag{12}$$

Substituting (8) and (9) into (10), and combining the result with (1), (2), and (4), a Markov-perfect rational-expectations equilibrium should satisfy

$$x_t - E_t x_{t+1} + \sigma\left(i_t - E_t \pi_{t+1} - r_t^n\right) = 0 \tag{13}$$

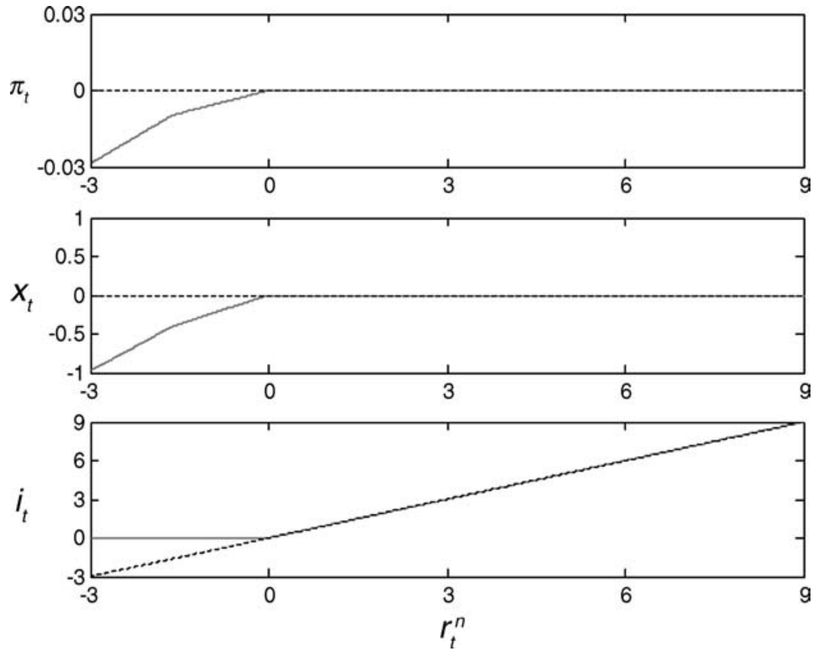
$$\pi_t - \kappa x_t - \beta E_t \pi_{t+1} = 0 \tag{14}$$

$$i_t(\lambda x_t + \kappa \pi_t) = 0 \tag{15}$$

$$i_t \geq 0 \tag{16}$$

$$\lambda x_t + \kappa \pi_t \leq 0. \tag{17}$$

Figure 2. Optimal Discretionary Policy with Perfect Foresight



Notice that (15) implies that the typical "targeting rule" involving inflation and the output gap is satisfied whenever the zero floor on the nominal interest rate is not binding,

$$\lambda x_t + \kappa \pi_t = 0, \tag{18}$$

$$\text{if } i_t > 0. \tag{19}$$

However, when the zero floor is binding, from (13) the dynamics are governed by

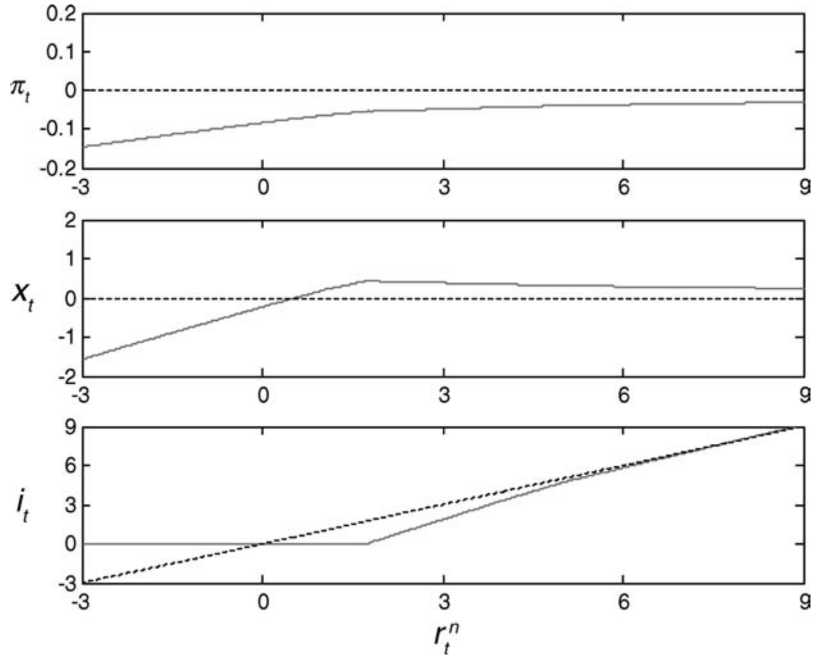
$$x_t + \sigma p_t - \sigma r_t^n = E_t x_{t+1} + \sigma E_t p_{t+1} \tag{20}$$

$$\text{if } i_t = 0, \tag{21}$$

where p_t is the (log) price level. Notice that it is no longer possible to set inflation and the output gap to zero at all times, for such a policy would require a negative nominal rate when the natural real rate falls below zero. Moreover, (20) implies that if the natural real rate falls so that the zero floor becomes binding, then since next period's output gap and price level are independent of today's actions, for expectations to be rational, the sum of the current output gap and price level must fall. The latter is true for any process for the natural real rate that allows it to take negative values.

An interesting special case, which replicates the findings of Jung, Teranishi, and Watanabe (2005), is the case of perfect foresight. By perfect foresight it is meant here that the natural real rate jumps initially to some (possibly negative) value, after which it follows a deterministic path (consistent with an AR(1) process) back to its steady state. In this case, the policy functions are represented by the solid lines in figure 2. As anticipated in the previous paragraph, at negative values of the natural real rate, both the output gap and inflation are below target. On the other hand, at positive levels of the natural real rate, prices and output can be stabilized fully in the case of discretionary optimization with perfect foresight. The reason for this is simple: once the natural real rate is above zero, deterministic reversion to steady state ensures that it will never be negative in the future. This means that it can always be tracked one-for-one by the nominal rate (as in the case

Figure 3. Optimal Discretionary Policy in the Stochastic Case



without zero floor), which is sufficient to fully stabilize prices and output.

One of the contributions of this paper is to extend the analysis in Jung, Teranishi, and Watanabe (2005) to the more general case in which the natural real rate follows a stochastic AR(1) process. Figure 3 plots the optimal discretionary policy in the stochastic environment. Clearly, optimal discretionary policy differs in several important ways, both from the optimal discretionary policy unconstrained by the zero floor and from the constrained perfect-foresight solution.

First of all, given the zero floor, it is in general no longer optimal to set either inflation or the output gap to zero in any period. In fact, in the solution with zero floor, inflation falls short of the target at *any* level of the natural real rate. This gives rise to a "deflationary bias" of optimal discretionary policy—in other words, an *average* rate of inflation below the target. Sensitivity analysis shows that for some plausible parameter values, the deflationary bias becomes quantitatively significant.[14] This implies that any quantitative analysis of discretionary biases in monetary models that does not take into account the zero lower bound can be misleading.

Secondly, as in the case of perfect foresight, at negative levels of the natural real rate, both inflation and the output gap fall short of their respective targets. However, the deviations from target are larger in the stochastic case—up to 1.5 percentage points for the output gap and up to 15 basis points for inflation at a natural real rate of –3 percent. As we will see in the following section, the fall of inflation under discretionary optimization is in contrast with the case of commitment, when prices are much better stabilized and may even slightly *increase* while the nominal interest rate is at its zero lower bound.

Third, above a positive threshold for the natural real rate, the optimal output gap becomes positive, peaking at around +0.5 percent.

Finally, at positive levels of the natural real rate, the optimal nominal interest rate policy with zero floor is both *more expansionary* (i.e., prescribing a lower nominal rate) and *more aggressive* (i.e., steeper) compared with the optimal discretionary policy without zero floor.[15] As a result, the nominal rate hits the zero floor at levels of the natural real rate as high as 1.8 percent (and is constant at zero for lower levels of the natural real rate).

These results hinge on two factors: (i) the nonlinearity induced by the zero floor and (ii) the stochastic nature of the natural real rate. The combined effect is an asymmetry in the ability of the central bank to respond to positive versus negative shocks when the natural real rate is close to zero. Namely, while the central bank can fully offset any positive shocks to the natural real rate because nothing prevents it from raising the nominal rate by as much as is necessary, it cannot fully offset large-enough negative shocks. The most it can do in this case is to reduce the nominal rate to zero, which is still higher than the rate consistent with zero output gap

[14] For example, the deflationary bias becomes half a percentage point with $\rho = 0.8$ and $r^* = 2$ percent.

[15] This is also true when the optimal nominal interest rate policy is compared with the optimal discretionary policy with zero floor and perfect foresight.

and inflation. Taking private-sector expectations as given, the latter implies a higher than desired *real* interest rate, which depresses output and prices through the IS and Phillips curves.

At the same time, when the natural real rate is close to zero, private-sector expectations reflect the asymmetry in the central bank's problem: a positive shock in the following period is expected to be neutralized, while an equally probable negative one is expected to take the economy into a liquidity trap. This gives rise to a "deflationary bias" in expectations, which in a forward-looking economy has an immediate impact on the current evolution of output and prices. Absent an endogenous state, the current evolution of the economy is all that matters today, and so it is rational for the central bank to partially offset the depressing effect of expectations on today's outcome by more aggressively lowering the nominal rate when the risk of deflation is high.

At sufficiently high levels of the natural real rate, the probability for the zero floor to become binding converges to zero. In that case, optimal discretionary policy approaches the unconstrained one—namely, zero output gap and inflation and a nominal rate equal to the natural real rate. However, around the deterministic steady state, the differences between the two policies—with and without zero floor—remain significant.

Since in the baseline model the discretionary optimization problem is equivalent to a sequence of static problems, optimal discretionary policy is independent of history. This means that it is only the current risk of falling into a liquidity trap that matters for current policy, regardless of whether the economy is approaching a liquidity trap or has just exited one. This is in sharp contrast with the optimal policy under commitment, which involves a particular type of history dependence, as will become clear in the following section.

3.2 Optimal Commitment Policy

In the absence of the zero lower bound, the equilibrium outcome under optimal discretion is globally optimal, and therefore it is observationally equivalent to the outcome under optimal commitment policy. The central bank manages to stabilize fully inflation and the output gap while adjusting the nominal rate one-for-one with the natural real rate.

However, this observational equivalence no longer holds in the presence of a zero interest rate floor. While full stabilization under either regime is not possible, important gains can be obtained from the ability to commit to future policy. In particular, by committing to deliver inflation in the future, the central bank can affect private-sector expectations about inflation, and thus the real rate, even when the nominal interest rate is constrained by the zero floor. This channel of monetary policy is simply unavailable to a discretionary policymaker.

Using the same Lagrange method as before, but this time taking into account the dependence of expectations on policy choices, it is straightforward to obtain the equilibrium conditions that govern the optimal commitment solution:

$$x_t - E_t x_{t+1} + \sigma\left(i_t - E_t \pi_{t+1} - r_t^n\right) = 0 \tag{22}$$

$$\pi_t - \kappa x_t - \beta E_t \pi_{t+1} = 0 \tag{23}$$

$$\pi_t - \phi_{1t-1}\sigma/\beta + \phi_{2t} - \phi_{2t-1} = 0 \tag{24}$$

$$\lambda x_t + \phi_{1t} - \phi_{1t-1}/\beta - \kappa\phi_{2t} = 0 \tag{25}$$

$$i_t \phi_{1t} = 0 \tag{26}$$

$$i_t \geq 0 \tag{27}$$

$$\phi_{1t} \geq 0. \tag{28}$$

From conditions (24) and (25), it is clear that the Lagrange multipliers inherited from the past period will have an effect on current policy. They in turn will depend on the history of endogenous variables and in particular on whether the zero floor was binding in the past. In this sense, the Lagrange multipliers summarize the effect of commitment, which (in contrast to optimal discretionary policy), involves a particular type of history dependence.

Figures 4–6 plot the optimal policies in the case of commitment. The figures illustrate specifically the dependence of policy on ϕ_{1t-1}, the Lagrange multiplier associated with the zero floor, while holding ϕ_{2t-1} fixed. When the nominal interest rate is constrained by the zero floor, ϕ_1 becomes positive, implying that the central bank commits to a lower nominal rate, higher inflation, and higher output gap in the following period, conditional on the value of the natural real rate.

Figure 4. Optimal Commitment Policy (Inflation)

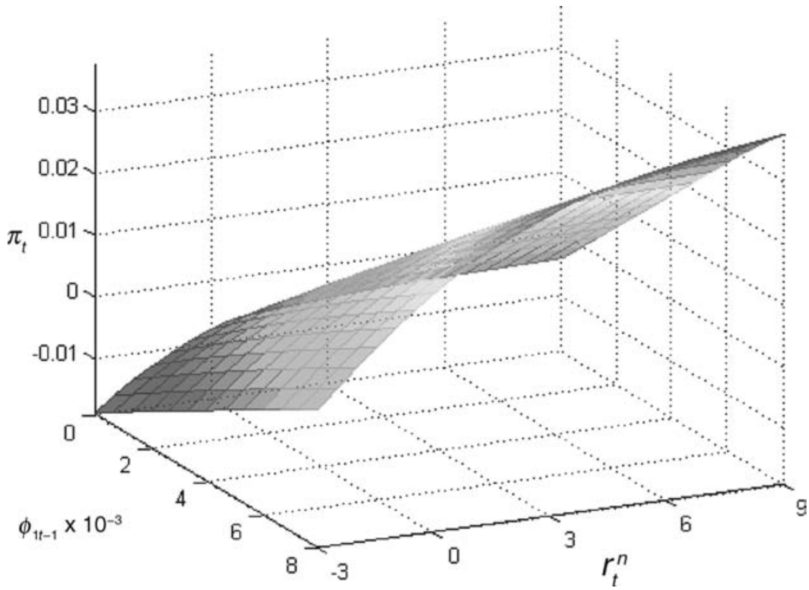


Figure 5. Optimal Commitment Policy (Output Gap)

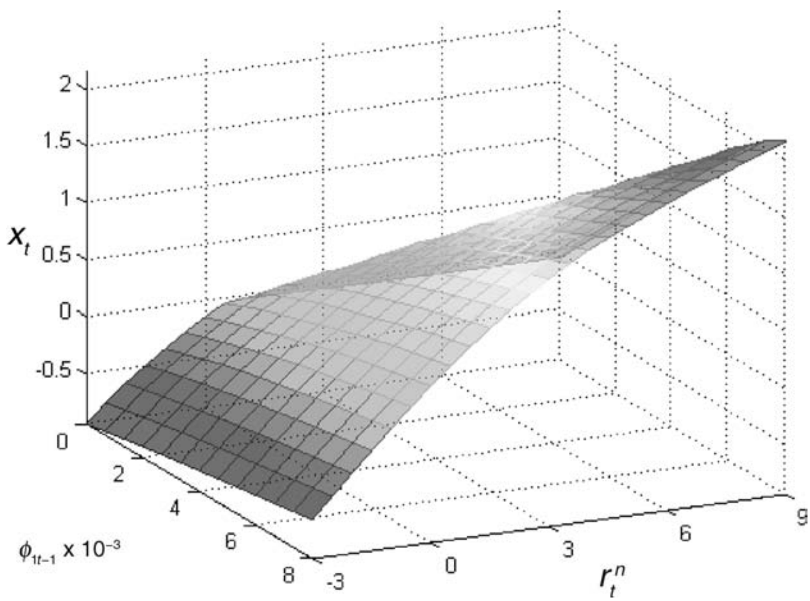
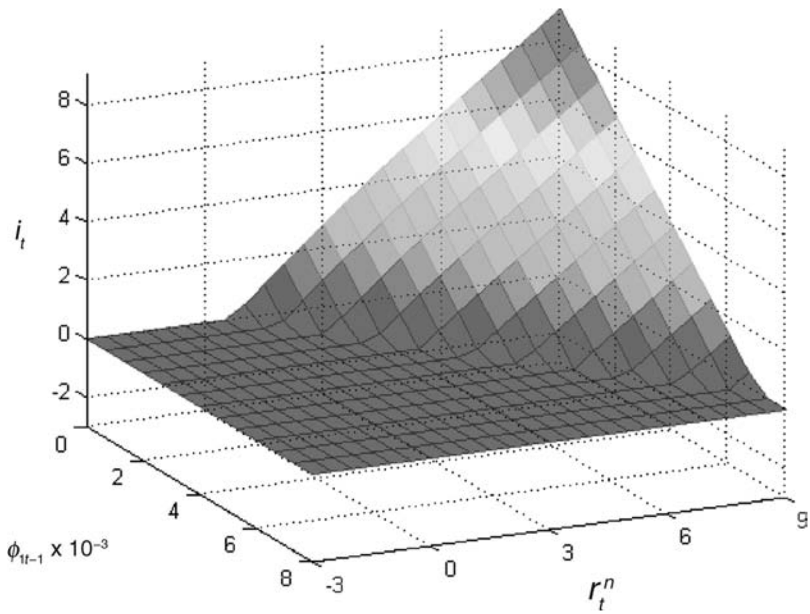


Figure 6. Optimal Commitment Policy (Nominal Interest Rate)



Since the commitment is assumed to be credible, it enables the central bank to achieve higher expected inflation and a lower real rate in periods when the nominal rate is constrained by the zero floor. The lower real rate reinforces expectations for higher future output and thus further stimulates current output demand through the IS curve. This, together with higher expected inflation, stimulates current prices through the expectational Phillips curve. Commitment therefore provides an additional channel of monetary policy, which works through expectations and through the ex ante real rate, and which is unavailable to a discretionary monetary policymaker.

A standard way to illustrate the differences between optimal discretionary and commitment policies is to compare the dynamic evolution of endogenous variables under each regime in response to a single shock to the exogenous natural real rate. Figures 7 and 8 plot the impulse responses to a small and a large negative shock to the natural real rate, respectively. In figure 7, notice that in the case

Figure 7. Impulse Responses to a Small Shock: Commitment vs. Discretion

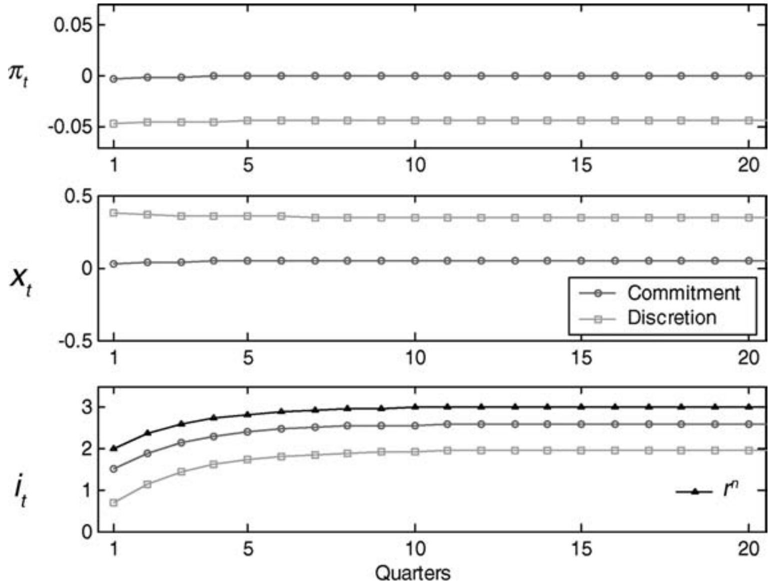
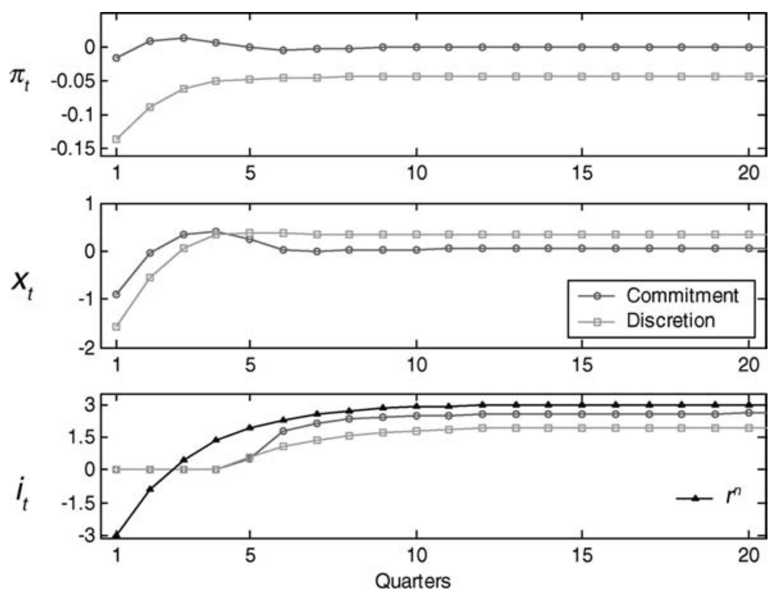


Figure 8. Impulse Responses to a Large Shock: Commitment vs. Discretion



of a small shock to the natural real rate from its steady state of 3 percent down to 2 percent, inflation and the output gap under optimal commitment policy (lines with circles) remain almost fully stabilized. In contrast, under discretionary optimization (lines with squares), inflation stays slightly below target and the output gap remains about half a percentage point above target, consistent with equation (18), as the economy converges back to its steady state. The nominal interest rate under discretion is about 1 percent lower than the rate under commitment throughout the simulation, yet it remains strictly positive at all times.

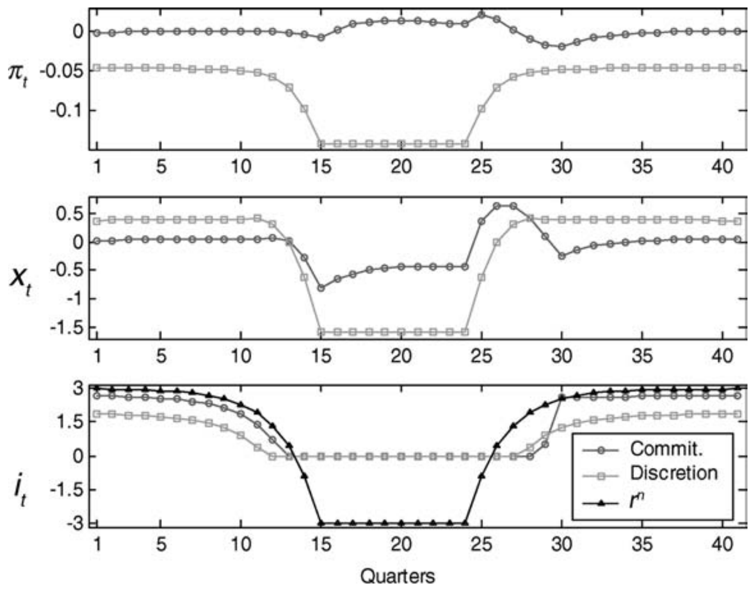
The picture changes substantially in the case of a large negative shock to the natural interest rate to –3 percent (see figure 8). Notably, under both commitment and discretion, the nominal interest rate hits the zero lower bound and remains there until two quarters after the natural interest rate has returned to positive.[16] Under discretionary optimization, both inflation and the output gap fall on impact, consistent with equation (20), after which they converge toward their steady state. The initial shortfall is significant, especially for the output gap, amounting to about 1.5 percent. In contrast, under the optimal commitment rule, the initial output loss and deflation are much milder, owing to the ability of the central bank to commit to a positive output gap and inflation once the natural real rate has returned to positive.

An alternative way to compare optimal discretionary and commitment policies in the stochastic environment is to juxtapose the dynamic paths that they prescribe for endogenous variables under a chosen evolution for the stochastic natural real rate.[17] The experiment is shown in figure 9, which plots a simulated "liquidity trap" under the two regimes. The line with triangles in the bottom panel is the assumed evolution of the natural real rate. It slips down from +3 percent (its deterministic steady state) to −3 percent over a period of fifteen quarters, then remains at −3

[16]The fact that the zero-interest-rate policy terminates in the same quarter under commitment and under discretion is a coincidence in this experiment. The relative duration of a zero-interest-rate policy under commitment versus discretion depends on the parameters of the shock process as well as the particular realization of the shock.

[17]In the model, agents observe only the current state; i.e., the future evolution of the natural real rate is unknown to them in this experiment.

Figure 9. Optimal Paths in a Liquidity Trap—Commitment vs. Discretion



percent for ten quarters before recovering gradually (consistent with the assumed AR(1) process) to +3 percent in another fifteen quarters.

The top and middle panels of figure 9 show the responses of inflation and the output gap under each of the two regimes. Not surprisingly, under the optimal commitment regime, both inflation and the output gap are closer to target than under the optimal discretionary policy. In particular, under optimal discretion, inflation is always *below* the target as it falls to -0.15 percent, shadowing the drop in the natural real rate. Compared to that, under optimal commitment, prices are almost fully stabilized, and in fact they even slightly *increase* while the natural real rate is negative!

In turn, under optimal discretion the output gap is initially around $+0.4$ percent, but then it declines sharply to -1.6 percent with the decline in the natural real rate. In contrast, under optimal commitment, output is initially at its potential level and the largest negative output gap is only half the size of the one under optimal discretion.

Supporting these paths of inflation and the output gap are corresponding paths for the nominal interest rate. Under discretionary optimization, the nominal rate starts at around 2 percent and declines at an increasing rate until it hits zero two quarters before the natural real rate has turned negative. It is then kept at zero while the natural real rate is negative, and only two quarters after the latter has returned to positive territory does the nominal interest rate start rising again. Nominal rate increases following the liquidity trap mirror the decreases while approaching the trap, so that the tightening is more aggressive in the beginning and then gradually diminishes as the nominal rate approaches its steady state.

In contrast, the nominal rate under optimal commitment begins closer to 3 percent, then declines to zero one quarter before the natural real rate turns negative. After that, it is kept at its zero floor until three quarters after the recovery of the natural real rate to positive levels, which is one quarter longer compared with optimal discretionary policy. Interestingly, once the central bank starts increasing the nominal rate, it raises it very quickly; the nominal rate climbs nearly 3 percentage points in just two quarters. This is equivalent to six consecutive monthly increases by 50 basis points each. The reason is that once the central bank has validated the inflationary expectations (which help mitigate deflation during the liquidity trap), there is no more incentive to keep inflation above target when the natural interest rate has returned to normal.

Under discretion, the paths of inflation, output, and the nominal rate are symmetric with respect to the midpoint of the simulation period because optimal discretionary policy is independent of history. Therefore, inflation and the output gap inherit the dynamics of the natural real rate, the only state variable on which they depend. This is in contrast with the asymmetric paths of the endogenous variables under commitment, reflecting the optimal history dependence of policy under this regime. In particular, the fact that under commitment the central bank can promise higher output gap and inflation in the wake of a liquidity trap is precisely what allows it to engage in less preemptive easing of policy in anticipation of the trap and at the same time deliver a superior inflation and output-gap performance compared with the optimal policy under discretion.

4. Suboptimal Rules with Zero Floor

4.1 Targeting Rules

In the absence of the zero floor, targeting rules take the form

$$\alpha_\pi E_t \pi_{t+j} + \alpha_x E_t x_{t+k} + \alpha_i E_t i_{t+l} = \tau, \tag{29}$$

where α_π, α_x, and α_i are weights assigned to the different objectives; j, k, and l are forecasting horizons; and τ is the target. These are sometimes called *flexible* inflation-targeting rules to distinguish them from *strict* inflation targeting of the form $E_t\pi_{t+j} = \tau$.[18] When j, k, or $l > 0$, the rules are called inflation *forecast* targeting to distinguish them from rules targeting contemporaneous variables.

As demonstrated by (20) in section 3, in general, such rules are not consistent with equilibrium in the presence of the zero floor, for they would require negative nominal interest rates at times. A natural way to modify targeting rules so that they comply with the zero floor is to write them as a complementarity condition,

$$i_t(\alpha_\pi E_t \pi_{t+j} + \alpha_x E_t x_{t+k} + \alpha_i E_t i_{t+l} - \tau) = 0 \tag{30}$$

$$i_t \geq 0, \tag{31}$$

which requires that either the target τ is met or the nominal interest rate must be at its zero floor. In this sense, a rule like (30)–(31) can be labeled "flexible inflation targeting with a zero-interest-rate floor."

In fact, section 3 showed that the optimal policy under discretion takes this form with $\alpha_i = 0$, $\alpha_\pi = \kappa$, $\alpha_x = \lambda$, $j = k = 0$, and $\tau = 0$—namely,

$$i_t(\lambda x_t + \kappa\pi_t) = 0 \tag{32}$$

$$i_t \geq 0. \tag{33}$$

In the absence of the zero floor, it is well known that optimal commitment policy can be formulated as optimal *speed-limit*

[18] Notice that the zero lower bound implies that *strict* inflation targeting is simply not feasible: from the New Keynesian Phillips curve, $\pi_t = C$ implies $x_t = C(1-\beta)$ at all times, and the IS equation is not satisfied for large-enough negative shocks to r_t^n.

targeting,

$$\Delta x_t + \frac{\kappa}{\lambda}\pi_t = 0, \tag{34}$$

where $\Delta x_t = \Delta y_t - \Delta y_t^{flex}$ is the growth rate of output relative to the growth rate of flexible-price output (the *speed limit*). In contrast to discretionary optimization, however, the optimal commitment rule with zero floor cannot be written in the form (30)–(31). This is because, with zero floor, the optimal target involves a particular type of history dependence, as shown by Eggertsson and Woodford (2003).[19] In particular, manipulating the first-order conditions of the optimal commitment problem, one can arrive at the following speed-limit targeting rule with zero floor:

$$i_t \left[\Delta x_t + \frac{\kappa}{\lambda}\pi_t - \frac{1}{\lambda}\left(\frac{\kappa\sigma + \beta}{\beta}\phi_{1t-1} - \phi_{1t} + \frac{1}{\beta}\Delta\phi_{1t-1}\right)\right] = 0 \tag{35}$$

$$i_t \geq 0. \tag{36}$$

Since $\kappa\sigma$ is small and β is close to one, and for plausible values of ϕ_{1t} consistent with the assumed stochastic process for the natural real rate,[20] the above rule is approximately the same as

$$i_t \left[\Delta y_t + \frac{\kappa}{\lambda}\pi_t - \tau_t\right] = 0 \tag{37}$$

$$i_t \geq 0, \tag{38}$$

where $\tau_t \approx \Delta y_t^{flex} + \lambda^{-1}\Delta^2\phi_{1t}$ is a history-dependent target (speed limit). In normal circumstances when $\phi_{1t} = \phi_{1t-1} = \phi_{1t-2} = 0$, the target is equal to the growth rate of flexible-price output, as in the problem without zero bound; however, if the economy falls into a liquidity trap, the speed limit is adjusted in each period by the speed of change of the penalty (the Lagrange multiplier) associated with the non-negativity constraint. The faster the economy is plunging into the trap, therefore, the higher is the speed-limit target that the

[19] These authors derive the optimal commitment policy in the form of a moving *price-level* targeting rule. Alternatively, it can be formulated as a moving *speed-limit* targeting rule as demonstrated here.

[20] ϕ_{1t} is two orders of magnitude smaller than the natural real rate.

central bank promises to achieve contingent on the interest rate's return to positive territory.

While the above rule is optimal in this framework, it is perhaps not very practical. Its dependence on the unobservable Lagrange multipliers makes it very hard, if not impossible, to implement or communicate to the public. Moreover, as pointed out by Eggertsson and Woodford (2003), credibility might suffer if all that the private sector observes is a central bank that persistently undershoots its target yet keeps raising it for the following period. To overcome some of these drawbacks, Eggertsson and Woodford (2003) propose a simpler constant price-level targeting rule, of the form

$$i_t \left[x_t + \frac{\kappa}{\lambda} p_t \right] = 0$$
$$i_t \geq 0, \tag{39}$$

where p_t is the log price level.[21]

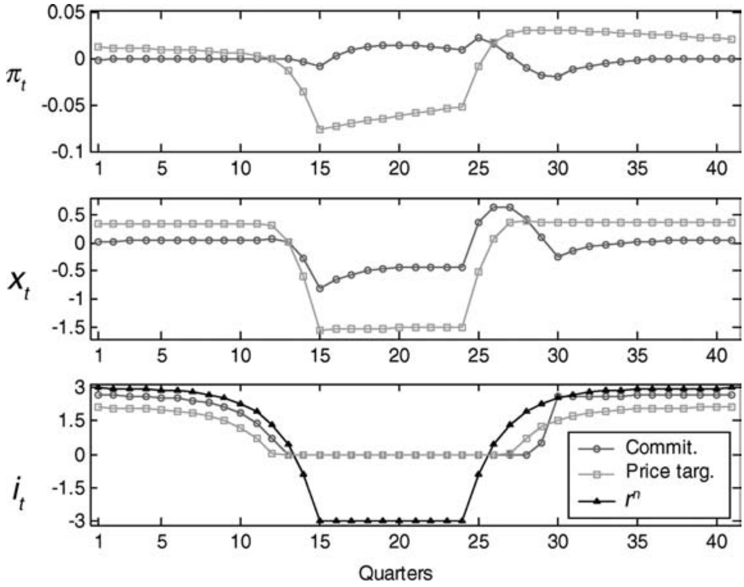
The idea is that committing to a price-level target implies that any undershooting of the target resulting from the zero floor is going to be undone in the future by positive inflation. This raises private-sector expectations and eases deflationary pressures when the economy is in a liquidity trap. Figure 10 demonstrates the performance of this simpler rule in a simulated liquidity trap. Notice that while the evolution of the nominal rate and the output gap is similar to that under the optimal discretionary rule, the path of inflation is much closer to the target. Since the weight of inflation in the central bank's loss function is much larger than that of the output gap, the fact that inflation is better stabilized accounts for the superior performance of this rule in terms of welfare.

4.2 Simple Instrument Rules

The practical difficulties with communicating and implementing rules like (35) or even (39) have led many researchers to focus on simple instrument rules of the type proposed by Taylor (1993). These rules have the advantage of postulating a relatively straightforward

[21] Notice that the weight on the price level is optimal within the class of constant price-level targeting rules. In particular, it is related to $\kappa/\lambda = \varepsilon$, the degree of monopolistic competition among intermediate goods producers.

Figure 10. Dynamic Paths under Constant Price-Level Targeting



relationship between the nominal interest rate and a limited set of variables in the economy. While the advantage of these rules lies in their simplicity, at the same time—absent the zero floor—some of them have been shown to perform close enough to the optimal rules in terms of the underlying policy objectives (Galí 2003). Hence, it has been argued that some of the better simple instrument rules may serve as a useful benchmark for policy, while facilitating communication and transparency.

In most of the existing literature, however, simple instrument rules are specified as linear functions of the endogenous variables. This is, in general, inconsistent with the existence of a zero floor because for large-enough negative shocks (e.g., to prices), linear rules would imply a negative value for the nominal interest rate. For instance, a simple instrument rule reacting only to past period's inflation,

$$i_t = r^* + \pi^* + \phi_\pi(\pi_{t-1} - \pi^*), \tag{40}$$

where r^* is the equilibrium real rate, π^* is the target inflation rate, and ϕ_π is an inflation response coefficient, can clearly imply negative values for the nominal rate.

In the context of liquidity trap analysis, a natural way to modify simple instrument rules is to truncate them at zero with the max$(\cdot)$ operator. For example, the truncated counterpart of the above Taylor rule can be written as

$$i_t = \max[0, r^* + \pi^* + \phi_\pi(\pi_{t-1} - \pi^*)]. \tag{41}$$

In what follows, I consider several types of truncated instrument rules, including the following:

- Truncated Taylor rules (TTRs) that react to past, contemporaneous, or expected future values of the output gap and inflation ($j \in \{-1, 0, 1\}$),

$$i_t^{TTR} = \max[0, r^* + \pi^* + \phi_\pi(E_t\pi_{t+j} - \pi^*) + \phi_x(E_t x_{t+j})] \tag{42}$$

- TTRs with partial adjustment or "interest rate smoothing" (TTRSs),

$$i_t^{TTRS} = \max\left\{0, \phi_i i_{t-1} + (1 - \phi_i) i_t^{TTR}\right\} \tag{43}$$

- TTRs that react to the price *level* instead of inflation (TTRPs),

$$i_t^{TTRP} = \max[0, r^* + \phi_\pi(p_t - p^*) + \phi_x x_t] \tag{44}$$

  where p_t is the log price level and p^* is a constant price-level target; and
- Truncated "first-difference" rules (TFDRs) that specify the *change* in the interest rate as a function of the output gap and inflation,

$$i_t^{TFDR} = \max[0, i_{t-1} + \phi_\pi(\pi_t - \pi^*) + \phi_x x_t]. \tag{45}$$

  This formulation ensures that if the nominal interest rate ever hits zero, it will be held there as long as inflation and the output gap are negative, thus extending the duration of a zero-interest-rate policy relative to a truncated Taylor rule.

- The "augmented Taylor rule" (ATR) of Reifschneider and Williams (2000),

$$i_t^{ATR} = \max\left[0, i_t^{TR} - \alpha Z_t\right] \tag{46}$$

$$i_t^{TR} = r^* + \pi^* + \phi_\pi(\pi_t - \pi^*) + \phi_x x_t \tag{47}$$

$$Z_t = Z_{t-1} + \left(i_t^{ATR} - i_t^{TR}\right). \tag{48}$$

This last rule keeps track of the amount by which the interest rate was higher than an unconstrained Taylor rule due to a binding zero lower bound, and allows for a compensating lower nominal interest rate once the natural real rate has returned to positive levels. Reifschneider and Williams (2000) simulate a stochastic economy with this policy (under the assumption of certainty equivalence) and show that it improves performance substantially compared with the standard Taylor rule. The augmented Taylor rule is interesting also because it is thought to have influenced the conduct of monetary policy in the United States during the 2003–05 episode when announcements by Federal Reserve Chairman Alan Greenspan suggested a "considerable period" of low interest rates, followed by a "measured pace" of interest rate increases.[22]

As before, I illustrate the performance of each family of simple instrument rules by simulating a liquidity trap and plotting the implied paths of endogenous variables under each regime. In addition, I contrast the performance of optimal commitment policy to the augmented rule of Reifschneider and Williams (2000), assuming that the Federal Reserve followed their rule in the period since 2001:Q3. The more rigorous evaluation of welfare of alternative policies is reserved for the following section.

Given the model's simplicity, the focus here is not on finding the optimal values of the parameters within each class of rules but rather on evaluating the performance of alternative monetary policy regimes. To do that I use values of the parameters commonly estimated and widely used in simulations in the literature. I make sure that the parameters satisfy a sufficient condition for local uniqueness of equilibrium. Namely, the parameters are required to observe the

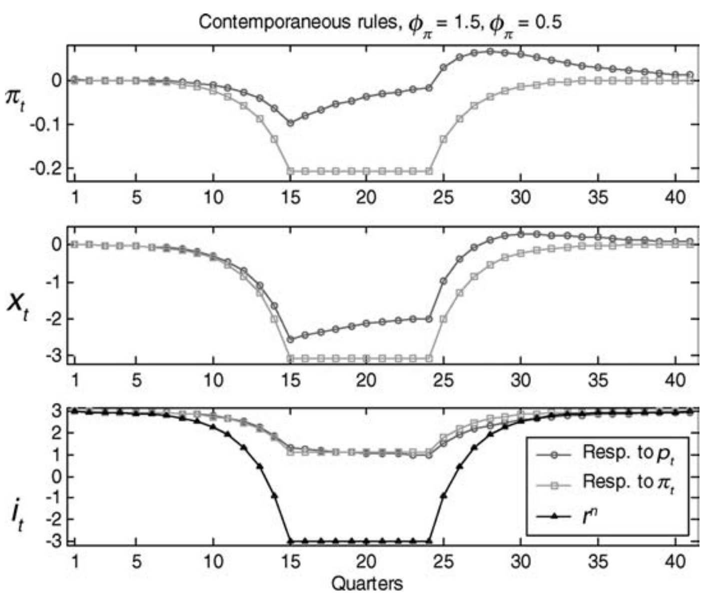
[22] I thank the editor John Taylor for pointing this out to me and suggesting the additional exercise with the Reifschneider and Williams (2000) rule.

so-called Taylor principle, according to which the nominal interest rate must be adjusted more than one-to-one with changes in the rate of inflation, implying $\phi_\pi > 1$. I further restrict $\phi_x \geq 0$ and $0 \leq \phi_i \leq 0.8$.

Figure 11 plots the dynamic paths of inflation, the output gap, and the nominal interest rate that result under regimes TTR and TTRP, conditional on the same path for the natural real rate as before. Both the truncated Taylor rule (TTR, lines with squares) and the truncated rule responding to the price level (TTRP, lines with circles) react contemporaneously with coefficients $\phi_\pi = 1.5$ and $\phi_x = 0.5$, and $\pi^* = 0$.

Several features of these plots are worth noticing. First of all, and not surprisingly, under the truncated Taylor rule, inflation, the output gap, and the nominal rate inherit the behavior of the natural real rate. Perhaps less expected, though, while both inflation and especially the output gap deviate further from their targets compared with the optimal rules in figure 9, the nominal interest rate

Figure 11. Truncated Taylor Rules Responding to the Price Level or to Inflation

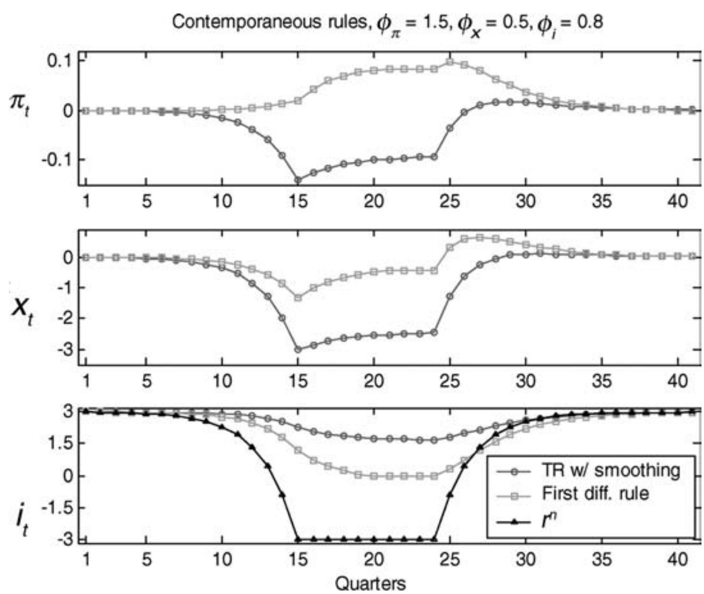


always stays above 1 percent, even when the natural real rate falls as low as −3 percent! This suggests that—contrary to popular belief—an equilibrium real rate of 3 percent may provide a sufficient buffer from the zero floor even with a truncated Taylor rule targeting *zero* inflation.

Secondly, figure 11 demonstrates that in principle the central bank can do even better than a TTR by reacting to the price level rather than to the rate of inflation. The reason for this is clear—by committing to react to the price level, the central bank promises to undo any past disinflation by higher inflation in the future. As a result, when the economy is hit by a negative real-rate shock, current inflation falls by less because expected future inflation increases.

Figure 12 plots the dynamic paths of endogenous variables under regimes TTRS and TFDR, again with $\phi_\pi = 1.5$, $\phi_x = 0.5$, and $\pi^* = 0$. The TTRS (lines with circles) is a partial adjustment version of the TTR, with smoothing coefficient $\phi_i = 0.8$. The TFDR

Figure 12. Truncated Taylor Rule with Smoothing vs. Truncated First-Difference Rule

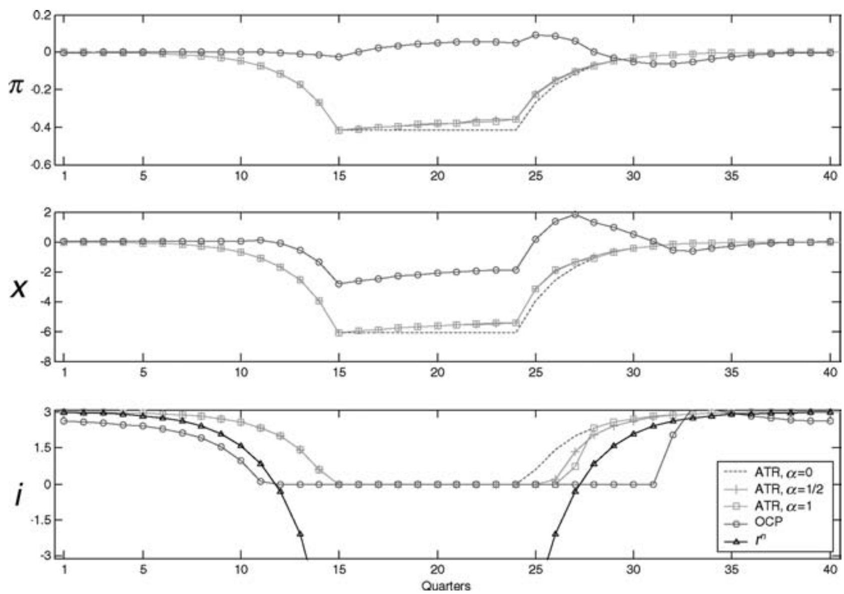


(lines with squares) is a truncated first-difference rule that implies more persistent deviations of the nominal interest rate from its steady-state level.

The figure suggests that interest rate smoothing (TTRS) may improve somewhat on the truncated Taylor rule (TTR) and may do a bit worse than the rule reacting to the price level (TTRP). However, it implies the least instrument volatility. On the other hand, the truncated first-difference rule (TFDR) seems to be doing an even better job at stabilization in a liquidity trap. Notice that under this rule, the nominal interest rate deviates most from its steady state, hitting zero for five quarters. Interestingly, the paths for inflation and the output gap under the TFDR resemble, at least qualitatively, those under the optimal commitment policy, suggesting that this rule may be approximating the optimal history dependence of policy.

Finally, figure 13 contrasts the liquidity trap performance of the "augmented Taylor rule" (ATR) of Reifschneider and Williams (2000) to that of the optimal commitment policy. With $\alpha = 0$, the

Figure 13. Dynamic Paths: Augmented Taylor Rule vs. Optimal Commitment Policy



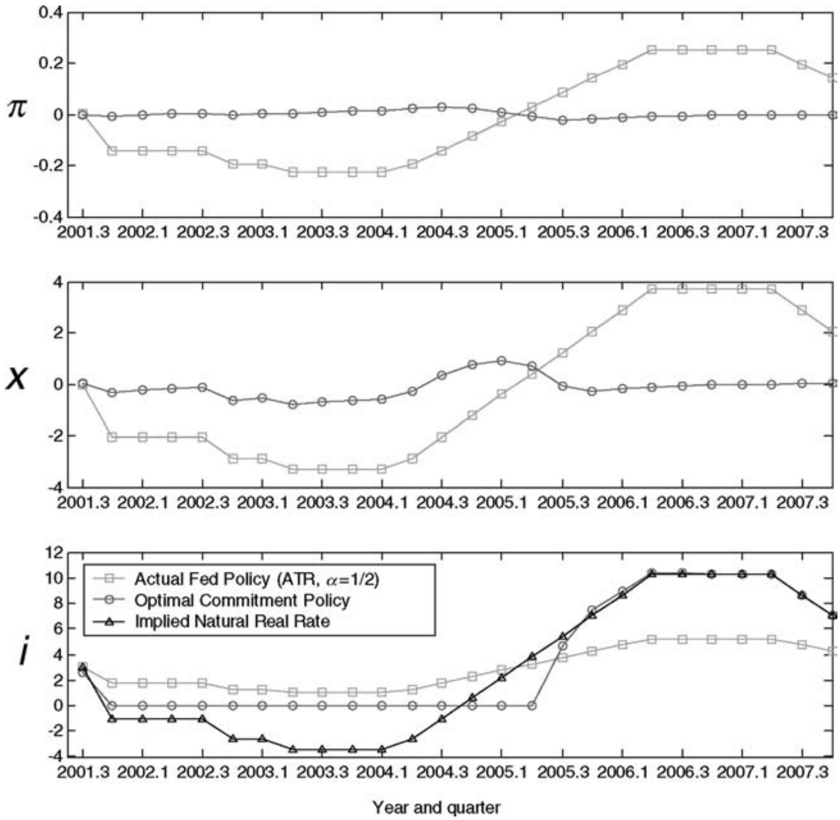
ATR is the same as the standard Taylor rule. Since we have seen in figure 11 that with response coefficients $\phi_\pi = 1.5$ and $\phi_x = 0.5$ the nominal interest rate under the Taylor rule remains positive even as the natural real rate falls to -3 percent, in this particular exercise we assume that the natural real rate falls much more (to –9 percent) so as to allow the mechanism of the Reifschneider and Williams (2000) rule to kick in. The figure shows that, in that case, with $\alpha = 1/2$, the augmented rule implies a zero nominal interest rate for one additional quarter and a lower nominal interest rate compared to the TTR during five quarters. Notice, however, that since under the ATR the zero-interest-rate policy is terminated later, and eventually the interest rate must converge to the standard Taylor rule, the pace of interest rate increases is *faster* than that of the standard TTR. This is even more pronounced with $\alpha = 1$, in which case the interest rate is kept at zero two additional quarters but then is raised very rapidly and converges to the TTR in just two quarters.

The top and middle panels of the figure show, not surprisingly, that the augmented rules with $\alpha = 1/2$ (lines with crosses) or $\alpha = 1$ (lines with squares) achieve better stabilization outcomes than the standard TTR (dashed lines). Interestingly, though, the paths of inflation and the output gap almost overlap with $\alpha = 1/2$ or $\alpha = 1$, suggesting that—within this class of rules and provided that α is positive—the particular time profile of the extra easing of policy is not so important.

What seems to make a big difference in a liquidity trap situation, however, is the total amount of easing following the recovery of the natural real rate. This can be seen by contrasting the inflation and output-gap performance of the ATR with that of the optimal commitment rule (lines with circles). Under the optimal commitment policy, the nominal interest rate is kept at zero for as much as seven quarters more than the standard Taylor rule, and five quarters more than the ATR with $\alpha = 1$. After that, as already noted in section 3.2, the interest rate is raised very rapidly to $+3$ percent in just two quarters, much faster than the TTR and even than the ATR with $\alpha = 1$.

Figure 14 illustrates this point in the context of the recent U.S. experience. The line with squares plots the end-of-quarter actual federal funds rate from 2001:Q3 (right after September 11) to 2008:Q1. The line with triangles is the implied path of the natural real interest

Figure 14. The Recent U.S. Episode: Actual Federal Reserve Policy, Implied Natural Real Rate, and Optimal Commitment Policy



rate, assuming that actual Federal Reserve policy followed the augmented Taylor rule of Reifschneider and Williams (2000). And the line with circles is the optimal commitment policy, given the imputed path of the exogenous natural real rate. The contrast between the two policies is quite clear: through the lens of the standard three-equation monetary policy model, the "considerable period" of low interest rates ended "too soon" (the federal funds rate was kept at 1 percent during four quarters), while the subsequent "measured pace" of interest rate increases was much "too slow."

In particular, according to our model, a policymaker following the optimal commitment policy would have set the nominal interest rate to zero for fifteen quarters (from 2001:Q4 through 2005:Q2), followed by an aggressive closing of the gap between the actual and the natural rate of interest in a single quarter. This policy would have essentially stabilized prices and would have resulted in only a modest and short-lived output boom (output above the natural level) between 2004:Q3 and 2005:Q2. In comparison, under the augmented Taylor rule, inflation and the output gap both were much lower than the target between 2001:Q4 and 2004:Q4, and then much higher than the target between 2005:Q3 and 2007:Q2. This stark contrast between the performance of the two rules may be interpreted as a caveat to the advisability of "measured pace" of interest rate increases following a liquidity trap. What optimal policy seems to dictate instead is the creation of expectations (and subsequent delivery) of a *zero* nominal interest rate during a *prolonged* period, followed by a *rapid* catch-up with a more normal policy stance once the economy has recovered and a zero interest rate is no longer needed.

As a final qualification, it is important to keep in mind that the simulations in figures 9–13 are conditional on one particular path for the natural real rate. It is, of course, possible that a suboptimal rule that appears to perform well while the economy is in a liquidity trap turns out to perform badly on average. In the following section, I undertake the ranking of alternative rules according to an unconditional expected welfare criterion, which takes into account the stochastic nature of the economy, time discounting, and the relative cost of inflation vis-à-vis output-gap fluctuations.

5. Welfare Ranking of Alternative Rules

A natural criterion for the evaluation of alternative monetary policy regimes is the central bank's loss function. Woodford (2003) shows that under appropriate assumptions the latter can be derived as a second-order approximation to the utility of the representative consumer in the underlying sticky-price model.[23] Rather than

[23]See footnote 9.

normalizing the weight of inflation to one, I normalize the loss function so that utility losses arising from deviations from the flexible-price equilibrium can be interpreted as a fraction of steady-state consumption,

$$WL = \frac{\overline{U} - U}{U_c C} = \frac{1}{2} E_0 \sum_{t=0}^{\infty} \beta^t \left[\varepsilon(1 + \varphi\varepsilon)\zeta^{-1}\pi_t^2 + (\sigma^{-1} + \varphi)x_t^2 \right] \quad (49)$$

$$= \frac{1}{2}\varepsilon(1 + \varphi\varepsilon)\zeta^{-1} E_0 \sum_{t=0}^{\infty} \beta^t L_t, \quad (50)$$

where $\zeta = \theta^{-1}(1-\theta)(1-\beta\theta)$; θ is the fraction of firms that keep prices unchanged in each period; φ is the (inverse) elasticity of labor supply; and ε is the elasticity of substitution among differentiated goods. Notice that $(\sigma^{-1} + \varphi)[\varepsilon(1 + \varphi\varepsilon)]^{-1}\zeta = \kappa/\varepsilon = \lambda$ implies the last equality in the above expression, where L_t is the central bank's period loss function, which is being minimized in (3).

I rank alternative rules on the basis of the *unconditional* expected welfare. To compute it, I simulate 2,000 paths for the endogenous variables over 1,000 quarters and then compute the average loss per period across all simulations. For the initial distribution of the state variables, I run the simulation for 200 quarters prior to the evaluation of welfare. Table 2 ranks all rules according to their welfare score. It also reports the volatility of inflation, the output gap, and the nominal interest rate under each rule, as well as the frequency of hitting the zero floor.

Table 2. Properties of Optimal and Simple Rules with Zero Floor

	OCP	**PLT**	**TFDR**	**ODP**	**TTRP**	**TTRS**	**ATR**	**TTR**
std(π) $\times 10^2$	1.04	3.47	4.59	3.85	7.23	9.12	12.8	12.9
std(x)	0.45	0.69	1.04	0.71	1.61	1.91	1.89	1.90
std(i)	3.21	3.20	1.36	3.27	1.06	0.56	1.14	1.14
Loss $\times 10^5$	6.97	10.9	52.3	54.2	62.9	103	146	147
Loss/OCP	1	1.56	7.50	7.77	9.01	14.8	20.95	21.09
Pr($i = 0$)%	32.6	32.0	1.29	36.8	0.24	0.00	0.48	0.44

One thing to keep in mind in evaluating the welfare losses is that in the benchmark model with nominal price rigidity as the only distortion and a shock to the natural real rate as the only source of fluctuations, absolute welfare losses are quite small—typically less than 1/100 of a percent of steady-state consumption for any sensible monetary policy regime.[24] Therefore, the focus here is on evaluating rules on the basis of their welfare performance relative to that under the optimal commitment rule with zero floor.

In particular, in terms of unconditional expected welfare, the optimal discretionary policy (ODP) delivers losses that are nearly eight times larger than the ones achievable under the optimal commitment policy (OCP). Recall that abstracting from the zero floor and in the absence of shocks other than to the natural real rate, the outcome under discretionary optimization is the same as under the optimal commitment rule. Hence, the cost of discretion is substantially understated in analyses that ignore the existence of the zero lower bound on nominal interest rates. Moreover, *conditional* on the economy's fall into a liquidity trap, the cost of discretion is even higher.

Interestingly, the frequency of hitting the zero floor is quite high—around one-third of the time—under the optimal commitment policy, as well as under the optimal discretionary policy. This result is sensitive to the assumption that the central bank targets zero inflation in the long run. If instead the central bank targeted a rate of inflation of 2 percent, the frequency of hitting the zero floor would decrease to around 12 percent of the time. The latter is still much higher than what has been observed in the United States (or even in Japan) and suggests either that policy has not been conducted optimally (note that the frequency is much lower under the simple instrument rules) or that there may be other unmodeled costs associated with low or volatile interest rates, unrelated to the ability of the central bank to achieve its inflation and output-gap targets. Indeed, in the model presented, hitting the zero lower bound is desirable because commitment to a zero-interest-rate policy is precisely what enables the central bank to achieve inflation and output-gap paths closer to the targets.

[24]To be sure, output gaps in a liquidity trap are considerable; however, the output gap is attributed negligible weight in the central bank loss function of the benchmark model.

Table 2 further confirms Eggertsson and Woodford's (2003) intuition about the desirable properties of an (optimal) constant price-level targeting rule (PLT)—here losses are only 56 percent greater than those under the optimal commitment rule. It also involves hitting the zero floor around one-third of the time.

In contrast, losses under the truncated first-difference rule (TFDR) are 7.5 times as large as those under the optimal commitment rule. Interestingly, however, the TFDR narrowly outperforms optimal discretionary policy. Even though the implied volatility of inflation and the output gap is slightly higher under this rule, it does a better job than ODP at keeping inflation and the output gap closer to target on average. This is possibly related to the highly inertial nature of this rule. An additional advantage—albeit one that is not reflected in the benchmark welfare criterion—is that instrument volatility is less than half of that under any of the optimal policies. This is why the zero floor is hit only around 1.3 percent of the time under this rule.

Similarly, losses under the truncated Taylor rule reacting to the price level (TTRP) are nine times larger than under OCP but only slightly worse than optimal discretionary policy. Moreover, instrument volatility under this rule is smaller than under the TFDR, which is why it involves hitting the zero floor even more rarely—only one quarter every 100 years on average.

Not surprisingly, the rule with the least instrument volatility among the studied simple rules—less than one-fifth of that under OCP—is the truncated Taylor rule with smoothing (TTRS). As a consequence, under this rule the nominal interest rate virtually never hits the zero lower bound. However, welfare losses are almost fifteen times larger than under OCP.

Finally, under the simplest truncated Taylor rule (TTR) without smoothing, the zero lower bound is hit only two quarters every 100 years, while welfare losses are around twenty times larger than those under OCP. Nevertheless, even under this simplest rule, losses are very small in absolute terms.

The fact that the zero lower bound is hit so rarely under the standard truncated Taylor rule (as well as under the other considered simple instrument rules) explains why the expected welfare gains of following the augmented Taylor rule (ATR) are negligible in our setup: the zero floor binds so rarely that the mechanism of additional easing embedded in the augmented rule is triggered only once every 100 years or so. It also suggests that the zero constraint plays a minor role for unconditional expected welfare under many sensible simple instrument rules. Indeed, computing their welfare score without the zero floor (by removing the maximum operator), reveals that close to 99 percent of the welfare losses associated with the five simple instrument rules stem from their intrinsic suboptimality rather than from the zero floor per se. Put differently, if one reckons that the stabilization properties of a standard Taylor rule are satisfactory in an environment in which nominal rates can be negative, then adding the zero lower bound to it leaves unconditionally expected welfare virtually unaffected. Nevertheless, as was illustrated in the previous section, *conditional* on a sufficiently negative evolution of the natural real rate, the losses associated with most of the studied simple instrument rules are substantially higher relative to the optimal commitment policy.

6. Sensitivity Analysis

In this section, I analyze the sensitivity of the main findings with respect to the parameters of the shock process, the strength of reaction and the timing of variables in truncated Taylor-type rules, and an extension of the model with endogenous inflation persistence.

6.1 Parameters of the Natural Real-Rate Process

6.1.1 Larger Variance

Table 3 reports the effects of an increase of the standard deviation of r^n to 4.5 percent (a 20 percent increase), while keeping the

Table 3. Properties of Selected Rules with Higher std(r^n)

	OCP	ODP	TTR
std(r^n)	4.46	4.46	4.46
std(π)	$\times$1.52	$\times$1.50	$\times$1.20
std(x)	$\times$1.46	$\times$1.45	$\times$1.20
std(i)	$\times$1.14	$\times$1.14	$\times$1.19
Loss	$\times$2.12	$\times$2.60	$\times$1.42
Pr($i = 0$)%	$\times$1.23	$\times$1.19	$\times$3.24

persistence constant, under three alternative regimes—optimal commitment policy, discretionary optimization, and a truncated Taylor rule.

Under OCP, the zero floor is hit around 23 percent more often, while welfare losses more than double. Figure 15 shows that the higher volatility implies that both the preemptive easing of policy and the commitment to future loosening are somewhat stronger. In turn, figure 16 shows that under ODP, preemptive easing is much stronger and the deflation bias is larger; table 3 shows that welfare losses increase by a factor of 2.6. Finally, under the TTR, the zero floor is hit three times more often, while welfare losses are up by 40 percent.

6.1.2 Stronger Persistence of Shocks

Table 4 and figures 17–19 show the effect of an increase in the persistence of shocks to the natural real rate to 0.8, while keeping the variance of r^n unchanged.

Figure 15. Sensitivity of OCP to $\sigma(r^n)$

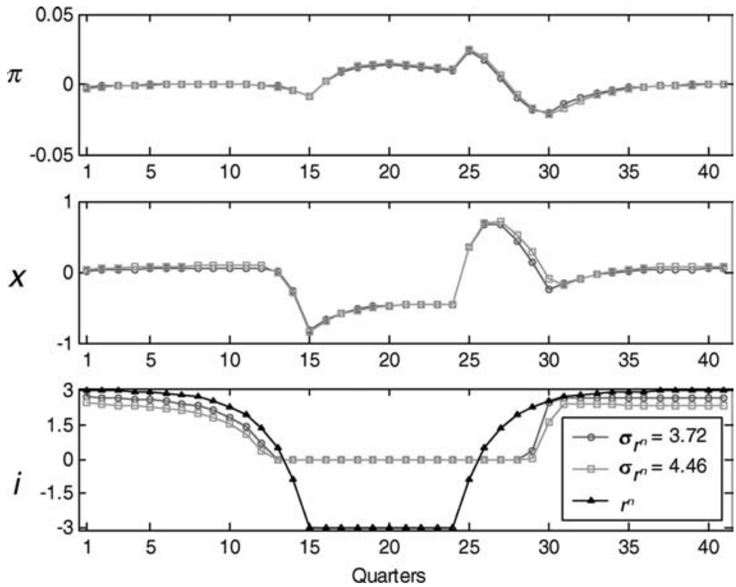
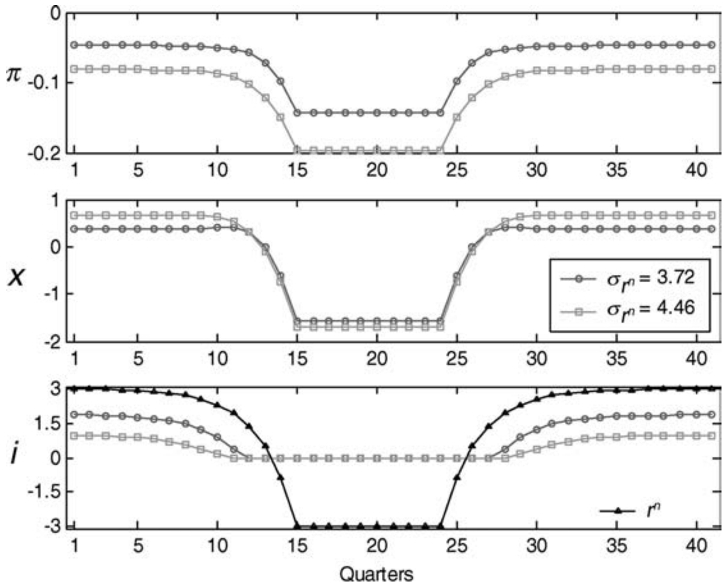


Figure 16. Sensitivity of ODP to $\sigma(r^n)$



Under OCP (figure 17), preemptive easing is a bit stronger, while future monetary loosening is much more prolonged. As a result of the stronger persistence, welfare losses under OCP more than double. Under ODP (figure 18), preemptive easing is much stronger, the deflation bias is substantially larger, and welfare losses increase by a factor of 5.5. And under the TTR (figure 19), deviations of inflation

Table 4. Properties of Selected Rules with More Persistent $\hat{r}^n$

	OCP	**ODP**	**TTR**
$\rho(r^n)$	0.80	0.80	0.80
std(π)	×2.14	×2.94	×2.44
std(x)	×1.55	×1.76	×1.42
std(i)	×1.02	×1.04	×1.52
Loss	×2.69	×5.47	×4.39
Pr($i = 0$)%	×1.09	×1.21	×11.20

Figure 17. Sensitivity of OCP to ρ

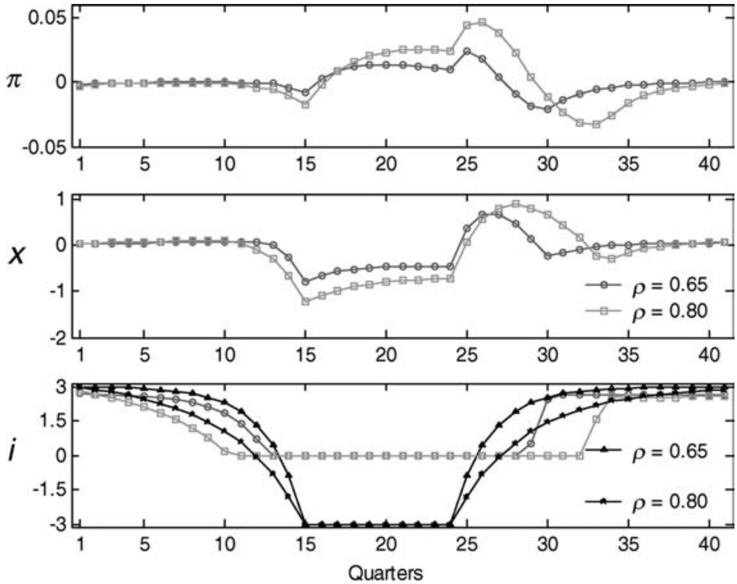


Figure 18. Sensitivity of ODP to ρ

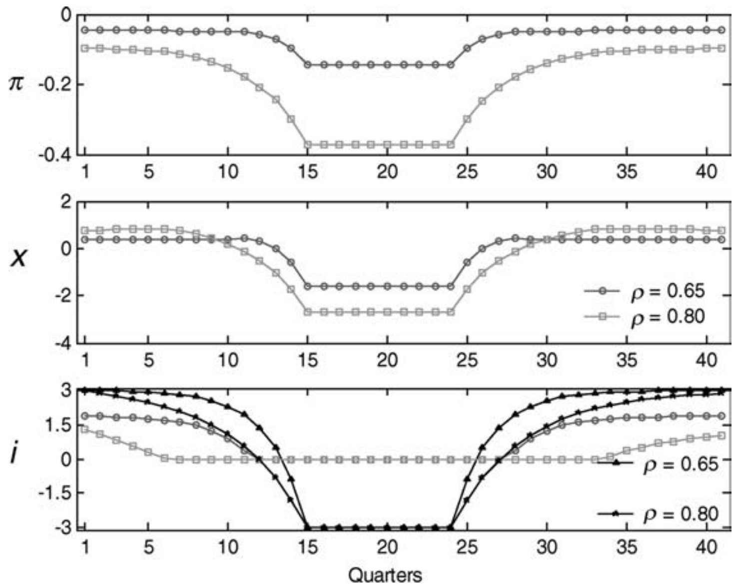
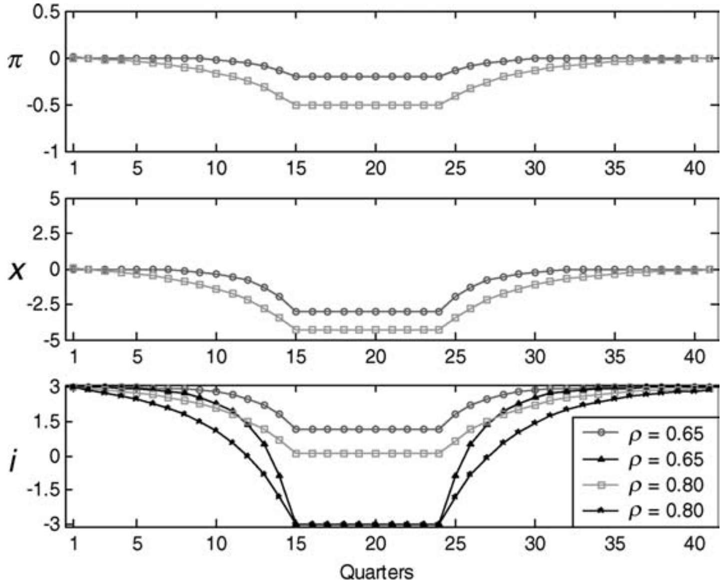


Figure 19. Sensitivity of TTR to ρ



and the output gap from target become larger and more persistent, the frequency of hitting the zero floor increases by a factor of 11, and welfare losses more than quadruple.

6.1.3 Lower Mean

The effects of a lower steady state of the natural real rate at 2 percent—keeping the variance and persistence of r^n constant—are illustrated in figures 20 and 21 and summarized in table 5.[25]

Under OCP, preemptive easing is a bit stronger, while future monetary policy loosening is much more prolonged; losses more than double. Interestingly, under ODP, preemptive easing is so strong that the nominal rate is zero more than half of the time. The deflation bias is larger, and losses increase by a factor of 4.5. And under the

[25]Notice that for simple rules such as the TTR, it is the sum $r^* + \pi^*$ that provides a "buffer" against the zero lower bound. Therefore, up to a constant shift in the rate of inflation, varying r^* is equivalent to testing for sensitivity with respect to π^*.

Figure 20. Sensitivity of OCP to r^*

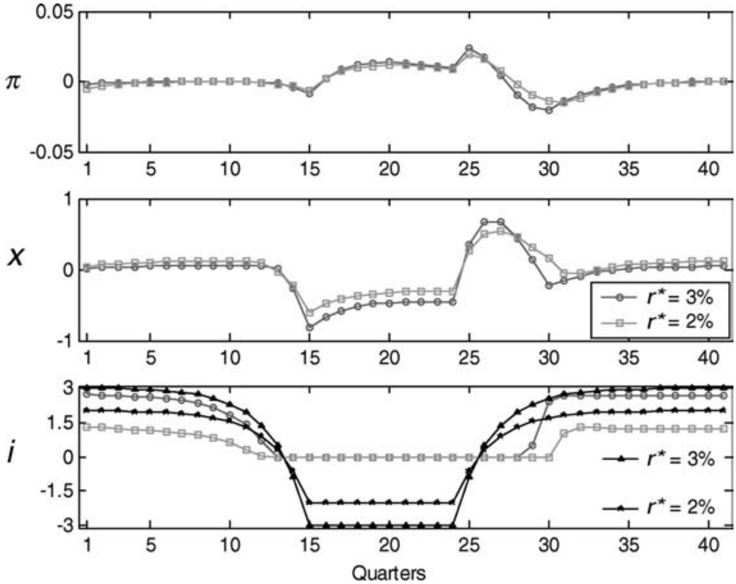


Figure 21. Sensitivity of ODP to r^*

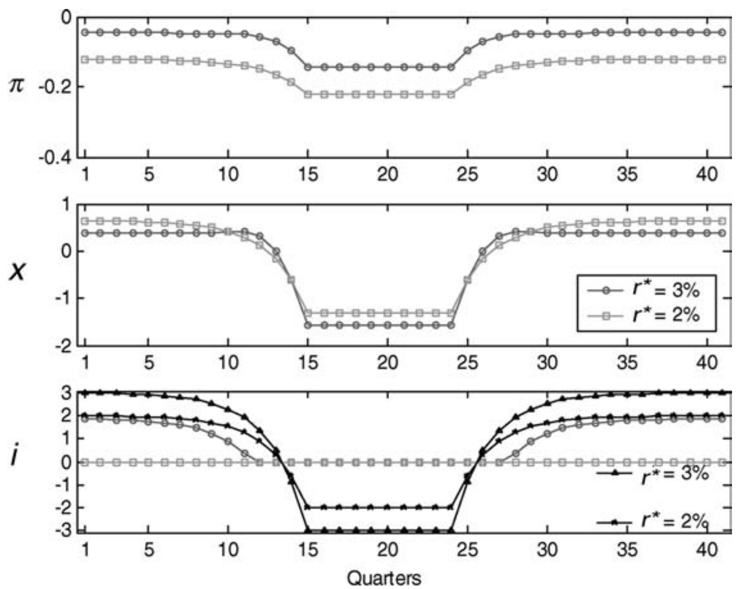


Table 5. Properties of Selected Rules with Lower r^*

	OCP	ODP	TTR
r^*	2%	2%	2%
std(π)	×1.74	×1.79	×1.00
std(x)	×1.55	×1.62	×1.00
std(i)	×0.89	×0.87	×0.98
Loss	×2.55	×4.47	×1.48
Pr($i = 0$)%	×1.51	×1.59	×9.27

TTR, the zero floor is hit nine times more often, while losses increase by 50 percent.

6.2 Instrument Rule Specification

6.2.1 The Strength of Response

Table 6 reports the dependence of welfare losses on the size of response coefficients in the truncated Taylor rule. It turns out that losses can be reduced substantially by having the interest rate react more aggressively to deviations of inflation (and to some extent the output gap) from target. For instance, losses are halved with $\phi_\pi = 10$ and $\phi_x = 1$ relative to the benchmark case $\phi_\pi = 1.5$ and $\phi_x = 0.5$. And they are reduced further to one-fifth with $\phi_\pi = 100$.

Table 6. Relative Losses under TTRs with Different Response Coefficients

	ϕ_π								
ϕ_x	**1.01**	**1.5**	**2**	**2.5**	**3**	**5**	**10**	**50**	**100**
0	×1.89	1.80	1.71	1.65	1.57	1.34	0.95	0.27	0.18
0.5	×1.03	1	0.97	0.94	0.90	0.80	0.59	0.24	0.18
0.75	×0.81	0.79	0.76	0.74	0.72	0.65	0.55	0.23	0.17
1	×0.65	0.63	0.62	0.60	0.58	0.53	0.50	0.22	0.17
1.5	×3.16	1.07	0.74	0.62	0.57	0.48	0.39	0.21	0.17
2	n.a.	2.05	1.13	0.77	0.64	0.45	0.41	0.24	0.23
3	n.a.	n.a.	4.44	2.64	1.92	0.83	0.45	0.29	0.27

In the absence of cost-push shocks, there is no policy trade-off, and in the model without zero lower bound, the central bank can approximate arbitrarily well the first-best outcome by threatening to adjust the interest rate by an infinite amount ($\phi_\pi \rightarrow \infty, \phi_x \rightarrow \infty$) in response to deviations of inflation and the output gap from target. In the case with zero lower bound, however, such a threat is constrained by the zero floor and implies that the nominal interest rate will be zero much of the time in response to infinitesimal target shortfalls. In practice, with response coefficients above $\phi_\pi = 100$ and $\phi_x = 3$, our numerical algorithm fails to converge. Nevertheless, table 6 shows that within the feasible range, welfare losses decline in a monotone fashion as the inflation reaction coefficient is increased. On the other hand, there is a nonlinearity when varying the output-gap coefficient. Namely, losses are higher with a small output-gap coefficient, then decline as the coefficient is raised; but as the output-gap response coefficient is increased further, welfare losses start rising again.

6.2.2 Forward-Looking, Contemporaneous, or Backward-Looking Reaction

For given response coefficients of a truncated Taylor rule, welfare losses turn out to be smallest under a backward-looking rule and highest under a forward-looking specification. While losses are still small in absolute value, with $\phi_\pi = 1.5$ and $\phi_x = 0.5$, they are up by 25 percent under the forward-looking rule and are around 15 percent lower under the backward-looking rule, relative to the contemporaneous one. The frequency of hitting the zero floor is similarly higher under a forward-looking specification and lower under a backward-looking one. The reason for the dominance of the backward-looking rule can be that under it the interest rate tends to be kept lower following periods of deflation, in a way that resembles the optimal history dependence under commitment. On the other hand, under forward-looking rules, the effective response to a given shock to the natural real rate is lower, given the assumed autoregressive nature of the natural real rate.

6.3 Endogenous Inflation Persistence and the Zero Floor

Wolman (1998), among others, has argued that stickiness of inflation is crucial in generating costs of deflation associated with the zero floor. To follow up on this hypothesis, I extend the present framework by incorporating endogenous inflation persistence.[26] Lagged dependence of inflation may result if firms that do not reoptimize prices index them to past inflation. In this case, the (log-linearized) inflation dynamics can be represented with the following modified Phillips curve (Christiano, Eichenbaum, and Evans 2001; Woodford 2003):

$$\widehat{\pi}_t = \beta E_t \widehat{\pi}_{t+1} + \kappa x_t, \tag{51}$$

where $\widehat{\pi}_t = \pi_t - \gamma\pi_{t-1}$ is a quasi-difference of inflation and γ measures the degree of price indexation.

An important thing to keep in mind is that in principle the welfare-relevant loss function is endogenous to the structure of the model. Hence, strictly speaking, one cannot compare welfare in the two environments—with and without inflation persistence—using the same loss criterion. On the other hand, Woodford (2003) shows that in the case of indexation to past inflation, the welfare-relevant loss function takes the same form as (3), except that inflation is replaced by its quasi-difference $\widehat{\pi}_t$. This implies that inflation persistence (as measured by γ) does not affect welfare under an optimal targeting rule that takes into account the existing degree of economy-wide indexation. However, since micro evidence on price changes rejects the presence of such indexation, I use the same criteria as in the baseline model to evaluate the performance of rules in an environment with endogenous inflation persistence.

Table 7 reports the properties of selected regimes relative to the baseline environment without endogenous inflation persistence. Under the optimal constant price-level targeting rule, an increase in the persistence of inflation to 0.8 results in doubling of inflation volatility and almost tripling of the baseline loss measure. Similarly, inflation volatility more than doubles and welfare losses nearly quadruple under the baseline truncated Taylor rule when the stickiness of inflation increases to 0.8. Interestingly, the properties of optimal discretionary policy are found to depend in a nonlinear way on the degree of inflation persistence. Namely, while an increase in inflation persistence to 0.5 raises inflation volatility by 80 percent

[26]In this case, lagged inflation becomes a state variable and the first-order conditions are adjusted accordingly.

Table 7. Performance of Selected Rules with Endogenous Inflation Persistence Relative to the Baseline Environment without Inflation Persistence

	PLT	**PLT**	**ODP**	**ODP**	**TTR**	**TTR**
γ	0.5	0.8	0.5	0.8	0.5	0.8
std(π)	×1.46	×2.21	×1.80	×1.02	×1.54	×2.35
std(x)	×1.01	×1.02	×1.12	×0.81	×0.98	×0.95
std(i)	×1.00	×1.00	×1.01	×1.02	×1.06	×1.10
Loss	×1.50	×2.77	×3.74	×2.66	×1.83	×3.70
Pr($i = 0$)%	×1.00	×1.00	×1.04	×1.15	×1.46	×2.00

and nearly quadruples losses, a further increase of inflation persistence to 0.8 leads to relatively smaller losses. The reason is that high inflation persistence serves as an additional channel of policy, making it possible for the central bank to "steer away" from an approaching liquidity trap by choosing higher current inflation.

7. Conclusions

Recent treatments of the zero-lower-bound issue have suffered from some important limitations. These include assuming perfect foresight or forcing certainty equivalence, or treating the zero floor as an initial condition rather than an occasionally binding non-negativity constraint. This paper addresses these issues, providing a global solution to a standard stochastic sticky-price model with an explicit occasionally binding ZLB on the nominal interest rate. As it turns out, the dynamics (and in some cases the unconditional means) of the nominal rate, inflation, and the output gap are strongly affected by uncertainty in the presence of the zero interest rate floor.

In particular, optimal discretionary policy involves a deflationary bias and interest rates are cut more aggressively when the risk of deflation is high, implying that they are kept lower both before and after a liquidity trap. The extent of such lowering of rates is found to increase in the variance and persistence of shocks to the natural real rate, and to decrease in its unconditional mean. Moreover, the preemptive lowering of rates is even more important under discretionary policy in the presence of endogenous inflation persistence. Compared with that, under optimal commitment policy, the need for preemptive lowering of interest rates is limited since the central bank can commit to a period of looser monetary policy conditional on the economy's recovery from a possible liquidity trap.

Imposing the zero lower bound correctly in the stochastic model allows us to evaluate quantitatively the performance of a variety of monetary policy regimes. Thus, commitment to the optimal rule reduces welfare losses to one-tenth of those achievable under discretionary policy. Constant price-level targeting delivers losses that are only 60 percent greater than those under the optimal commitment policy. In contrast, under a truncated Taylor rule, losses are twenty times greater than under the optimal commitment policy. Another interesting finding is that the unconditional welfare losses associated with simple instrument rules are almost unaffected by the zero lower bound per se and instead derive from the suboptimal responses to shocks characteristic of simple rules. This is related to the fact that under simple instrument rules, the zero lower bound is hit very rarely, while optimal commitment policy involves a zero nominal interest rate around one-third of the time.

In fact, in an extension of the model with money, optimal policy might be expected to visit the liquidity trap even more often. Hitting the zero lower bound in that case would be good because it eliminates the opportunity cost of holding cash balances. An interesting question to address in that setup would be how the optimal mean of the nominal interest rate is affected by the existence of the zero lower bound. Solving the fully nonlinear problem would be another useful extension; however, it increases the dimensionality of the computational problem. A limitation of the solution technique employed here is that it is practical only for models with a limited number of states.

Appendix. Numerical Algorithm

This section illustrates the algorithm used to solve the problem in the case of discretionary optimization. The cases with commitment and with simple rules are solved in a similar way. I apply the routines for rational-expectations models included in the CompEcon toolkit of Miranda and Fackler (2002). These solve for the optimal response x as a function of the state s, when equilibrium responses are governed by an arbitrage-complementarity condition of the form

$$f[s_t, x_t, E_t h(s_{t+1}, x_{t+1})] = \phi_t, \tag{52}$$

where s follows the state transition function

$$s_{t+1} = g(s_t, x_t, \varepsilon_{t+1}) \tag{53}$$

and x_t and ϕ_t satisfy the complementarity conditions

$$a(s_t) \leqslant x_t \leqslant b(s_t), \quad x_{jt} > a_j(s_t) \Rightarrow \phi_{jt} \leqslant 0, \quad x_{jt} < b_j(s_t) \Rightarrow \phi_{jt} \geqslant 0, \tag{54}$$

where ϕ_t is a vector whose j^{th} element, ϕ_{jt}, measures the marginal loss from activity j. In equilibrium, ϕ_{jt} must be nonpositive (non-negative) if x_{jt} is greater (less) than its lower (upper) bound; otherwise, agents can gain by reducing (increasing) activity j. If x_{jt} is neither at its upper nor at its lower bound, ϕ_{jt} must be zero to preclude arbitrage possibilities.

In the context of the monetary policy model under discretion, f_{jt} is the derivative of the complementarity condition (15) with respect to the nominal interest rate, and ϕ_{jt} is the Lagrange multiplier ϕ_{1t} associated with the non-negativity constraint on the nominal interest rate:

$$-(\lambda x_t + \kappa \pi_t) = \phi_{1t}. \tag{55}$$

Since there is no upper bound on the interest rate, $b(s_t) = +\infty$, and $x_t < b(s_t)$ always holds so that ϕ_{1t} is non-negative. This, together with $a(s_t) = 0$, implies that in the case of discretionary optimization, the above complementarity problem reduces to

$$i_t \geqslant 0, \quad \phi_{1t} \geqslant 0, \quad i_t > 0 \Rightarrow \phi_{1t} = 0, \tag{56}$$

which also can be written as

$$i_t \geqslant 0, \quad \phi_{1t} \geqslant 0, \quad i_t \phi_{1t} = 0. \tag{57}$$

An approximate solution is obtained with the method of collocation, which in this case consists of approximating the expectation functions $E_t x_{t+1}$ and $E_t \pi_{t+1}$ by linear combinations of known basis functions, θ_j, whose coefficients, c_j, are determined by requiring the approximants to satisfy the equilibrium equations exactly at n collocation nodes:

$$h[s, x(s)] \approx \sum_{j=1}^{n} c_j \theta_j(s). \tag{58}$$

The coefficients are determined by the following algorithm. For a given value of the coefficient vector c, the equilibrium responses x_i are computed at the n collocation nodes s_i by solving the complementarity problem (which is transformed into a standard root-finding problem). Then, given the equilibrium responses x_i at the collocation nodes s_i, the coefficient vector c is updated solving the n-dimensional linear system

$$\sum_{j=1}^{n} c_j \theta_j(s_i) = h(s_i, x_i). \tag{59}$$

This iterative procedure is repeated until the distance between successive values of c becomes sufficiently small (Miranda and Fackler 2002).

To approximate the expectation functions, $E_t x_{t+1}$ and $E_t \pi_{t+1}$, one needs to discretize the shock to r^n. Here the normal shock to the natural rate of interest is discretized using a K-node Gaussian quadrature scheme:

$$Eh[s, x(s)] \approx \sum_{k\ =\ 1}^{K} \sum_{j=1}^{n} \omega_k c_j \theta_j [g(s_i, x, \varepsilon_k)], \tag{60}$$

where ε_k and ω_k are Gaussian quadrature nodes and weights chosen so that the discrete distribution approximates the continuous univariate normal distribution $N(0, \sigma_\epsilon^2)$.

In the discretionary optimization problem I use linear splines on a uniform grid of 2,000 points for values of the natural rate of interest between -10 percent and $+10$ percent, so that each point on the grid corresponds to 1 basis point. In this problem, linear splines work better than Chebychev polynomials or cubic splines because the response function has a kink in the place where the zero bound becomes binding.

There are two types of approximation errors. On the one hand are the deviations from the equilibrium first-order conditions. In this case the "arbitrage benefits" are negligible for each of the three equilibrium equations (13), (14), and (15). Specifically, they are of the order of 10^{-16} for the IS and the Phillips curves, and 10^{-19} for the complementarity condition. On the other hand are the residuals from the approximation of the expectation functions. Except for a couple of residuals of the order of 10^{-4}, concentrated mostly in the place where the zero constraint becomes binding, the rest of the residuals are of the order of 10^{-8}. Given the measurement units, a residual of 10^{-4} corresponds to 0.001 percent of annual inflation or output-gap error, which, provided that the residuals of this size are just a few, is a satisfactory level of accuracy for the problem at hand. In principle, the expectations residuals can be reduced further by concentrating more evaluation points in the neighborhood of the kink and by using more quadrature nodes, albeit at the cost of computing time.

In the case of commitment, the problem is first cast in the form specified by (52), (53), and (54) by substituting out ϕ_{2t} from (24) into (25) and the resulting expression for ϕ_{1t} into (26). In addition, the state transition vector is augmented by the two "co-state" variables ϕ_{1t} and ϕ_{2t}, which are cast in recursive form using (24) and (25):

$$\phi_{1t} = \phi_{1t-1}(1 + \kappa\sigma)/\beta + \kappa\phi_{2t-1} - \lambda x_t - \kappa\pi_t \tag{61}$$

$$\phi_{2t} = \phi_{1t-1}\sigma/\beta + \phi_{2t-1} - \pi_t. \tag{62}$$

With simple rules, the system is in the required form, and the only necessary adjustments are to the state transition vector in those cases when past endogenous variables enter the rule.

References

Adam, K., and R. M. Billi. 2004a. "Optimal Monetary Policy under Commitment with a Zero Bound on Nonimal Interest Rates." ECB Working Paper No. 377.

———. 2004b. "Optimal Monetary Policy under Discretion with a Zero Bound on Nominal Interest Rates." ECB Working Paper No. 380.

———. 2006. "Optimal Monetary Policy under Commitment with a Zero Bound on Nominal Interest Rates." *Journal of Money, Credit, and Banking* 38 (7): 1877–1905.

———. 2007. "Optimal Monetary Policy under Discretion with a Zero Bound on Nominal Interest Rates." Forthcoming in *Journal of Monetary Economics.*

Barro, R. J., and D. B. Gordon. 1983. "Rules, Discretion and Reputation in a Model of Monetary Policy." *Journal of Monetary Economics* 12 (1): 101–21.

Calvo, G. A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12 (3): 383–98.

Christiano, L. J., M. Eichenbaum, and C. L. Evans. 2001. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." NBER Working Paper No. 8403.

Christiano, L. J., and J. D. M. Fischer. 2000. "Algorithms for Solving Dynamic Models with Occasionally Binding Constraints." *Journal of Economic Dynamics and Control* 24 (8): 1179–1232.

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., A. Orphanides, and V. Wieland. 2004. "Price Stability and Monetary Policy Effectiveness when Nominal Interest Rates are Bounded at Zero." CEPR Discussion Paper No. 3892.

Eggertsson, G. B., and M. Woodford. 2003. "The Zero Bound on Interest Rates and Optimal Monetary Policy." *Brookings Papers on Economic Activity* 1.

Galí, J. 2003. "New Perspectives on Monetary Policy, Inflation, and the Business Cycle." In *Advances in Economic Theory*, Vol. III, ed. M. Dewatripont, L. Hansen, and S. Turnovsky, 151–97. Cambridge University Press.

Jung, T., Y. Teranishi, and T. Watanabe. 2005. "Optimal Monetary Policy at the Zero-Interest-Rate Bound." *Journal of Money, Credit, and Banking* 37 (5): 813–35.

Kato, R., and S.-I. Nishiyama. 2005. "Optimal Monetary Policy when Interest Rates are Bounded at Zero." *Journal of Economic Dynamics and Control* 29 (1–2): 97–133.

Keynes, J. M. 1936. *The General Theory of Employment, Interest, and Money.* London: MacMillan.

Klaeffling, M., and V. Lopez-Perez. 2003. "Inflation Targets and the Liquidity Trap." ECB Working Paper No. 272.

Krugman, P. 1998. "It's Baaaack: Japan's Slump and the Return of the Liquidity Trap." *Brooking Papers on Economic Activity* 2: 137–87.

Kydland, F. E., and E. C. Prescott. 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *The Journal of Political Economy* 85 (3): 473–92.

McCallum, B. T. 2000. "Theoretical Analysis Regarding a Zero Lower Bound on Nominal Interest Rates." *Journal of Money, Credit, and Banking* 32 (4): 870–904.

Miranda, M. J., and P. L. Fackler. 2002. *Applied Computational Economics and Finance*. Cambridge, MA: MIT Press.

Reifschneider, D., and J. C. Williams. 2000. "Three Lessons for Monetary Policy in a Low-Inflation Era." Federal Reserve Bank of Boston Conference Proceedings, 936–78.

Rotemberg, J. J., and M. Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." In *NBER Macroeconomics Annual 1997*, ed. B. S. Bernanke and J. J. Rotemberg, 297–346.

Schmitt-Grohe, S., and M. Uribe. 2004. "Optimal Operational Monetary Policy in the Christiano-Eichenbaum-Evans Model of the U.S. Business Cycle." Mimeo.

Taylor, J. B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39: 195–214.

Wolman, A. L. 1998. "Staggered Price Setting and the Zero Bound on Nominal Interest Rates." *Economic Quarterly* (Federal Reserve Bank of Richmond) 84 (4): 1–24.

Woodford, M. 2003. *Interest and Prices*. Princeton, NJ: Princeton University Press.

Specification and Calibration Errors in Measures of Portfolio Credit Risk: The Case of the ASRF Model*

Nikola Tarashev and Haibin Zhu
Bank for International Settlements

This paper focuses on the asymptotic single-risk-factor (ASRF) model in order to analyze the impact of specification and calibration errors on popular measures of portfolio credit risk. Violations of key assumptions of this model are found to be virtually inconsequential, especially for large, well-diversified portfolios. By contrast, flaws in the calibrated interdependence of credit risk across exposures, caused by plausible small-sample estimation errors or rule-of-thumb values of asset return correlations, can lead to significant inaccuracies in measures of portfolio credit risk. Similar inaccuracies arise under standard assumptions regarding the tails of the distribution of asset returns.

JEL Codes: G21, G28, G13, C15.

*We are indebted to Michael Gordy and Kostas Tsatsaronis for their encouragement and useful comments on an earlier version of this paper. We thank Klaus Düllmann, Wenying Jiangli, and participants at a BIS workshop, the 3rd International Conference on Credit and Operational Risk at HEC Montréal, and the 17th Annual Derivatives Securities and Risk Management Conference at the FDIC for insightful suggestions. We are also grateful to Marcus Jellinghaus for his valuable help with the data. The views expressed in this paper are our own and do not necessarily represent those of the Bank for International Settlements. Author contact: Nikola Tarashev: Research and Policy Analysis, Bank for International Settlements, Basel, Switzerland; Tel.: 41-61-280-9213; Fax: 41-61-280-9100; E-mail: nikola.tarashev@bis.org. Haibin Zhu: Research and Policy Analysis, Bank for International Settlements, Basel, Switzerland; Tel.: 41-61-280-9164; Fax: 41-61-280-9100; E-mail: haibin.zhu@bis.org.

1. Introduction

Assessments of portfolio credit risk have attracted much attention in recent years. One reason is that participants in the increasingly popular market for structured finance products rely heavily on estimates of the interdependence of credit risk across various exposures.[1] Such estimates are also of principal interest to financial supervisors who, in enforcing new standards in the banking and insurance industries, have to ensure that regulatory capital is closely aligned with credit risk.

This paper investigates the well-known asymptotic single-risk-factor (ASRF) model of portfolio credit risk, which underpins the internal-ratings-based (IRB) approach of the Basel II framework (Basel Committee on Banking Supervision 2005). The popularity of this model stems from its implication that the contribution of each exposure to the credit value-at-risk (VaR) of the portfolio—defined as the maximum default loss that can be incurred with a given probability over a given horizon—is independent of the characteristics of the other exposures. This implication has been derived rigorously in Gordy (2003) and is known as *portfolio invariance* of marginal credit VaR contributions. The implication has been interpreted as alleviating the data requirements and computational burden on users of the model. Indeed, portfolio invariance implies that credit VaR can be calculated solely on the basis of exposure-specific parameters, including individual probability of default (PD), loss given default (LGD), and dependence on the common factor.

Its popularity notwithstanding, the "portfolio invariance" implication of the ASRF model hinges on two strong assumptions that have been criticized as sources of *specification errors*. Namely, the model assumes that the systematic component of credit risk is governed by a single common factor and that the portfolio is so finely grained that all idiosyncratic risks are diversified away. Violations of the "single-factor" and "perfect-granularity" assumptions would translate directly into erroneous assessments of portfolio credit risk.

[1] Examples of structured finance products are collateralized debt obligations (CDOs), nth-to-default credit default swaps (CDSs), and CDS indices.

Moreover, putting specification issues aside, a practical implementation of the ASRF model may be quite challenging. An important reason is that the portfolio invariance implication does *not* allow a user of the model to consider any particular exposure in isolation. Namely, estimates of exposure-specific dependence on the common factor, which are required for a calibration of the ASRF model, hinge on information about the correlation structure in the portfolio or about the common factor itself.[2] When such portfolio- or market-wide information is imperfect, the user will implement a *flawed calibration* of the model, which will be another source of errors in measured portfolio credit risk.

A contribution of this paper is to develop a *unified* method for quantifying the importance of model specification and calibration errors in assessments of portfolio credit risk. In order to implement this method, we rely on a large data set that comprises Moody's KMV estimates of PDs and pairwise asset return correlations for nearly 11,000 nonfinancial corporations worldwide. We use these estimates as the actual credit-risk parameters of hypothetical portfolios that are designed to match the industrial-sector concentration of typical portfolios of U.S. wholesale banks. For each hypothetical portfolio, we derive the "true" probability distribution of default losses and then condense this distribution into unexpected losses, which are defined as a credit VaR net of expected losses. This summary statistic is equivalent to a "target" capital measure necessary to cover default losses with a desired probability.[3] Target capital can be compared directly to a "shortcut" capital measure, which relies on the ASRF model and incorporates a rule-of-thumb calibration of the interdependence of credit risk across exposures.

[2]The IRB capital formula of Basel II abstracts from this calibration issue by postulating that firm-specific dependence on the single common factor is determined fully by the level of the corresponding PD.

[3]In this paper, we use the terms "assessment of portfolio credit risk" and "capital measure" interchangeably. Importantly, our capital measures do not correspond to "regulatory capital," which reflects considerations of bank supervisors, or to "economic capital," which reflects additional strategic and business objectives of financial firms.

We decompose the difference between the target and shortcut capital measures into four non-overlapping and exhaustive components. Two of these components, which we attribute to a "multifactor" effect and a "granularity" effect, relate to misspecification of the ASRF model. The other two components, which we derive after transforming the correlation structure to be consistent with the ASRF model, relate to errors in the calibration of the interdependence of credit risk across exposures.[4] Specifically, the calibration errors we consider arise either from an overall bias in the measured correlations of firms' asset returns—which gives rise to what we dub a "correlation level" effect—or from noise in the measured dispersion of these correlations across pairs of firms—a "correlation dispersion" effect.

Another contribution of this paper is that it provides two additional perspectives on flaws in the calibration of the ASRF model. First, we calculate deviations from a desired capital buffer that arise *not* from the adoption of rule-of-thumb parameter values but from plausible small-sample errors in asset return correlation *estimates.* Second, motivated by the analysis in Gordy (2000) and Frey and McNeil (2003), we examine the importance of errors in the calibration of tail dependence among asset returns. The impact of such errors on capital measures is similar to but materializes independently of the impact of errors in estimated asset return correlations.

Our conclusion is that errors in the practical implementation, as opposed to the specification, of the ASRF model are the main sources of potential miscalculations of credit risk in large portfolios. Specifically, the misspecification-driven multifactor effect leads a user of the model to underpredict target capital buffers by only 1 percent. This is because a single-factor approximation, if chosen optimally, fits well the correlation structure of asset returns in our data. In addition, the granularity effect results in a 5 percent underprediction of the target level. In comparison, assessments of portfolio credit risk are considerably more sensitive to possible miscalibrations of the single-factor model. The correlation dispersion effect, for example, leads to capital measures that are 12 percent higher than the target measure. In turn, the correlation level effect causes a roughly 8 percent

[4] In order to sharpen the analysis, we do not analyze the implications of errors in the estimates of PDs and LGDs.

overprediction (underprediction) of the target measure for each percentage point of positive (negative) error in the average correlation coefficient. Furthermore, plausible small-sample errors in correlation estimates—arising when users of the model have five to ten years of monthly asset returns data—translate into capital measures that may deviate from the target level by 30 to 45 percent. Finally, data on asset returns are at odds with the conventional multinormality assumption. Specifically, this assumption implies tail-of-distribution dependence among asset returns that is too low and translates into an underestimation of the target capital by 22 to 86 percent.

Among the four effects on capital measures, only the granularity effect is sensitive to the number of exposures in the portfolio. When this number decreases, the portfolio maintains a larger portion of idiosyncratic risks. Thus, we are not surprised to find that the granularity effect leads to a 19 percent underestimation of target capital in typical small portfolios.

To the best of our knowledge, this is the first paper that develops a unified framework for analyzing a wide range of errors in assessments of portfolio credit risk. Most of the related literature has focused exclusively on misspecifications of the ASRF model and has proposed partial corrections that do not impair the model's tractability. Empirical analyses of violations of the perfect granularity assumption include Martin and Wilde (2002), Vasicek (2002), Emmer and Tasche (2003), and Gordy and Lütkebohmert (2007). For their part, Pykhtin (2004), Düllmann (2006), Düllmann and Masschelein (2006), and Garcia Cespedes et al. (2006) have analyzed implications of the common-factor assumption under different degrees of portfolio concentration in industrial sectors. In addition, Heitfield, Burton, and Chomsisengphet (2006) and Düllman, Scheicher, and Schmieder (2006) have examined both granularity and sector concentration issues in the context of U.S. and European bank portfolios, respectively.[5] A small branch of the related literature, which includes Loeffler (2003) and Morinaga and Shiina (2005), has considered only calibration issues and has derived that noise in model parameters can have a significant impact on assessments of portfolio credit risk.

[5] The recent working paper by the Basel Committee on Bank Supervision (2006) provides an extensive review of these articles.

The remainder of this paper is organized as follows. Section 2 outlines the ASRF model and the empirical methodology applied to it. Section 3 describes the data and section 4 reports the empirical results. Finally, section 5 concludes.

2. Methodology

In this section, we first outline the ASRF model. Then, we discuss how violations of its key assumptions or flawed calibration of its parameters can affect assessments of portfolio credit risk. Finally, we develop an empirical methodology for quantifying and comparing alternative sources of error in such assessments.

2.1 The ASRF Model[6]

The ASRF model of portfolio credit risk—introduced by Vasicek (1991)—postulates that an obligor defaults when its assets fall below some threshold. In addition, the model assumes that asset values are driven by a single common factor:

$$V_{iT} = \rho_i \cdot M_T + \sqrt{1 - \rho_i^2} \cdot Z_{iT}, \tag{1}$$

where V_{iT} is the value of assets of obligor i at time T; M_T and Z_{iT} denote the common and idiosyncratic factors, respectively; and $\rho_i \in [-1, 1]$ is the obligor-specific loading on the common factor. The common and idiosyncratic factors are independent of each other and scaled to random variables with mean 0 and variance 1.[7] Thus, the asset return correlation between borrowers i and j is given by $\rho_i \rho_j$.

The ASRF model delivers a closed-form approximation to the probability distribution of default losses on a portfolio of N exposures. The accuracy of the approximation increases when the number of exposures grows, $N \to \infty$, and the largest exposure weight

[6]This section provides an intuitive discussion of the ASRF model. For a rigorous and detailed study of this model, see Gordy (2003). In addition, Frey and McNeil (2003) examine the calibration of the ASRF model under the so-called Bernoulli mixture representation.

[7]The ASRF model can accommodate distributions with infinite second moments. Nonetheless, we abstract from this generalization in order to streamline the analysis.

shrinks, $\sup_i(w_i) \to 0$. In these limits, in which the portfolio is perfectly granular, the probability distribution of default losses can be derived as follows. First, let the indicator $\mathcal{I}_{iT}$ equal 1 if obligor i is in default at time T and 0 otherwise. Conditional on the value of the common factor, the expectation of this indicator equals

$$\begin{aligned} E(\mathcal{I}_{iT}|M_T) &= \Pr(V_{iT} < \mathcal{F}^{-1}(PD_{iT})|M_T) \\ &= \Pr(\rho_i \cdot M_T + \sqrt{1-\rho_i^2} \cdot Z_{iT} < \mathcal{F}^{-1}(PD_{iT})|M_T) \\ &= \mathcal{H}\left(\frac{\mathcal{F}^{-1}(PD_{iT}) - \rho_i M_T}{\sqrt{1-\rho_i^2}}\right), \end{aligned}$$

where PD_{iT} is the unconditional probability that obligor i is in default at time T; the cumulative distribution function (CDF) of Z_{iT} is denoted by $\mathcal{H}(\cdot)$; and the CDF of V_{iT} is $\mathcal{F}(\cdot)$, implying that the default threshold equals $\mathcal{F}^{-1}(PD_{iT})$.

Second, under perfect granularity, the Law of Large Numbers implies that the conditional total loss on the portfolio, $TL|M$, is deterministic for any value of the common factor M:

$$\begin{aligned} TL|M &= \sum_i w_i \cdot E(LGD_i) \cdot E(\mathcal{I}_i|M) \\ &= \sum_i w_i \cdot E(LGD_i) \cdot \mathcal{H}\left(\frac{\mathcal{F}^{-1}(PD_i) - \rho_i M}{\sqrt{1-\rho_i^2}}\right), \end{aligned} \tag{2}$$

where time subscripts have been suppressed. In addition, the loss given default of obligor i, LGD_i, is assumed to be independent of both the common and idiosyncratic factors.[8]

Finally, the conditional total loss $TL|M$ is a decreasing function of the common factor M and, consequently, the unconditional distribution of TL can be derived directly on the basis of equation (2) and the CDF of the common factor, $\mathcal{G}(\cdot)$. Denoting by

[8]The ASRF model does allow for interdependence between asset returns and the LGD random variable. Such interdependence leads to another dimension in the study of portfolio credit risk, which is explored by Kupiec (2008). We abstract from this additional dimension in order to focus on the correlation of default events.

$TL_{1-\alpha}$ the $(1-\alpha)^{th}$ percentile in the distribution of total losses, i.e., $\Pr(TL < TL_{1-a}) = 1-\alpha$, it follows that

$$\begin{aligned} TL_{1-\alpha} &= \sum_i w_i \cdot E(LGD_i) \cdot \mathcal{H}\left(\frac{\mathcal{F}^{-1}(PD_i) - \rho_i \mathcal{G}^{-1}(\alpha)}{\sqrt{1-\rho_i^2}}\right) \\ &= TL|M_\alpha, \end{aligned} \tag{3}$$

where $M_\alpha \equiv G^{-1}(\alpha)$ is the α^{th} percentile in the distribution of the common factor. The magnitude $TL_{1-\alpha}$ is also known as the credit VaR at the $(1-\alpha)$ confidence level.

The capital buffer that covers unexpected (i.e., total minus expected) losses on the entire portfolio with probability $(1-\alpha)$ equals[9]

$$\begin{aligned} \kappa &= TL_{1-\alpha} - \sum_i w_i \cdot E(LGD_i) \cdot PD_i \\ &= \sum_i w_i \cdot E(LGD_i) \cdot \left[\mathcal{H}\left(\frac{\mathcal{F}^{-1}(PD_i) - \rho_i \mathcal{G}^{-1}(\alpha)}{\sqrt{1-\rho_i^2}}\right) - PD_i\right] \\ &\equiv \sum_i w_i \cdot \kappa_i. \end{aligned} \tag{4}$$

As implied by this equation, the capital buffer for the portfolio can be set on the basis of exposure-specific parameters, which comprise the exposure's weight in the portfolio, as well as its LGD, PD, and loading on the common factor. The flip side of this implication

[9]In line with the IRB approach of Basel II, this paper effectively incorporates a one-period model, in which a default occurs only at the end of the horizon (in our case, T). In contrast, Gupta et al. (2005) consider obligors that default if their assets are below the default threshold at any one of *multiple* dates over the horizon. We do not consider such a multiperiod variant of the model because one of its key inputs is the term structure of asset return correlations, which cannot be estimated on the basis of our data set (see section 3). Importantly, abstracting from "multiperiod" considerations is unlikely to influence our conclusions. Gupta et al. (2005) find that switching from a one-period to a multiperiod setup lowers a 99.93 percent credit VaR by roughly 5 percent. Although not trivial, this magnitude is several times smaller than the magnitude of errors we identify and analyze below.

is that the portion of the capital buffer attributed to any particular exposure is independent of the rest of the portfolio and, thus, is *portfolio invariant.*[10]

In practice, an implementation of the ASRF model requires that one specify the distribution of the common and idiosyncratic factors of asset returns. It is standard to assume normal distributions and rewrite equation (4) as

$$\kappa = \sum_i w_i \cdot E(LGD_i) \cdot \left[\Phi \left(\frac{\Phi^{-1}(PD_i) - \rho_i \Phi^{-1}(\alpha)}{\sqrt{1-\rho_i^2}} \right) - PD_i \right], \quad (5)$$

where $\Phi(\cdot)$ is the CDF of a standard normal variable. Equation (5) underpins the regulatory capital formula in the IRB approach of Basel II, in which $E(LGD_i) = 45$ percent and $\alpha = 0.1$ percent.

2.2 Impact of Model Misspecification

The portfolio invariance implication of the ASRF model hinges on two key assumptions—i.e., that the portfolio is of perfect granularity and that there is a single common factor. In light of this, we examine at a conceptual level how violations of either of these assumptions—which give rise to "granularity" and "multifactor" effects—affect capital measures. For the illustrative examples in this section, we use the ASRF formula in equation (5) and set $\alpha = 0.1$ percent.

2.2.1 Granularity Effect

The granularity effect arises empirically either because of a limited number of exposures or because of exposure concentration in a small number of borrowers. In either of these cases, idiosyncratic risk is not fully diversified away. Therefore, the existence of a granularity effect implies that capital measures based on the ASRF model would be insufficient to cover unexpected losses.

The top-left panel in figure 1 provides an illustrative example of the granularity effect. This example focuses on a homogeneous portfolio, which gives rise to a desired capital level that decreases with

[10] Given portfolio invariance, the contribution of each exposure to portfolio credit risk is simply $w_i \kappa_i / \kappa$. Ordovas and Thompson (2003) develop a procedure for deriving this contribution in a more general setting.

the number of exposures (solid line).[11] The figure also plots the capital measure obtained under the ASRF model and, thus, incorporates the assumption that there is an infinite number of exposures (dotted line). The difference between the dotted and solid lines equals the magnitude of the granularity effect. As expected, the granularity effect is always negative and decreases when the number of exposures increases.

Gordy and Lütkebohmert (2007) derive a closed-form "granularity adjustment," which approximates (the negative of) the granularity effect. When the portfolio is homogeneous, the approximation is linear in the reciprocal of the number of exposures, which is largely in line with the properties of the granularity effect plotted in figure 1.[12]

2.2.2 Multifactor Effect

The impact of various macroeconomic and industry-specific conditions on portfolio credit risk may be best accounted for by generalizing equation (1) to incorporate *multiple* (potentially unobservable) common factors. Multiple common factors affect the likelihood of default clustering (i.e., the likelihood of a large number of defaults occurring over a given horizon), which influences the tails of the probability distribution of credit losses.[13] In line with our empirical results (reported in section 4 below), we treat a fattening of these tails (reflected in greater kurtosis) as implying unambiguously a higher level of the desired capital buffer.[14]

Importantly, the existence of multiple common factors of credit risk would violate the single-factor assumption of the ASRF model, leading to what we call a multifactor effect in capital measures. Depending on the characteristics of the credit portfolio, the multifactor effect could be either negative—i.e., implying that the ASRF

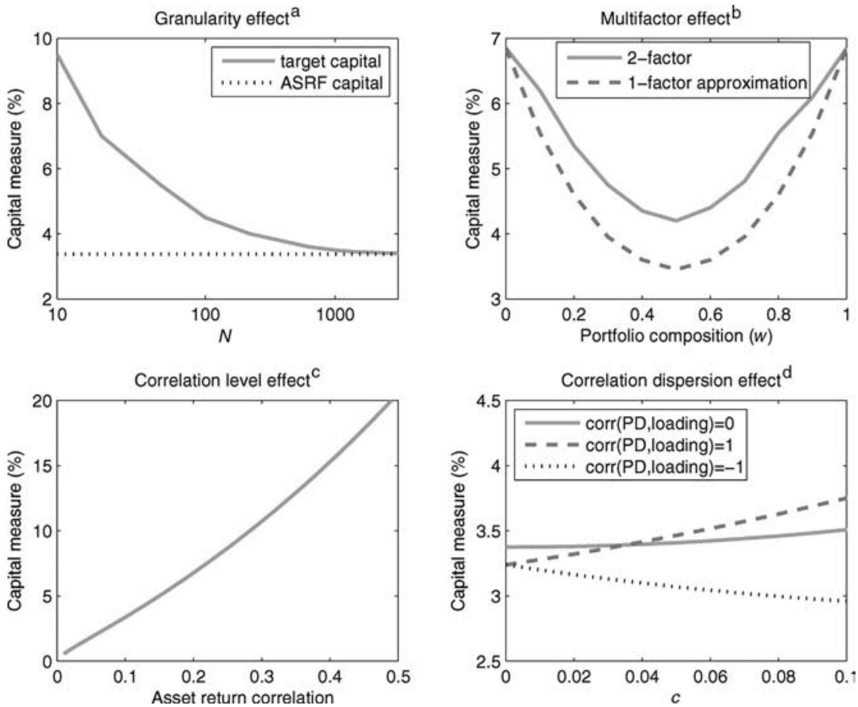
[11] The calculation of the desired capital level uses a Gaussian copula (see appendix 1).

[12] Further comparison between the granularity effect and the granularity adjustment of Gordy and Lütkebohmert (2007) is reported in section 4.1.

[13] These losses were denoted by TL in section 2.1.

[14] Note that, depending on the confidence level of the targeted credit VaR, a fattening of the tails of the probability distribution of credit losses may either raise or lower the desired capital buffer. Fatter tails of the loss distribution translate into a higher desired level of the capital buffer only if the value of α in equation (4) is sufficiently close to zero.

Figure 1. Four Sources of Error in Capital Measures



Note: Capital measures, in percent and per unit of aggregate exposure, are shown on the vertical axes. For each panel (unless noted otherwise), PD = 1 percent and LGD = 50 percent are the same across exposures.
[a]The solid line plots target capital for a portfolio in which all pairwise asset return correlations equal 10 percent. The number of exposures in the portfolio (N) varies across the horizontal axis. The dotted line plots the corresponding capital estimate when $N \to \infty$. [b]The portfolio consists of 1,000 exposures that are divided into two groups, with w denoting the weight of the first group. Within each group, the asset return correlations equals 20 percent for all exposure pairs. Intergroup correlations are zero. The solid line plots the target capital level, which incorporates the two common factors in the simulated data. The dashed line plots the capital calculated under a one-common-factor approximation of the correlation structure (see appendix 3). [c]Capital measures are shown for different levels of homogeneous pairwise asset return correlations. [d]The solid line plots capital measures under the assumption that PD = 1 percent, there is a single common factor, and the loadings on this factor are distributed uniformly in the cross-section between $\sqrt{0.1} - c$ and $\sqrt{0.1} + c$. For the other two lines, PDs are distributed uniformly in the cross-section between 0.5 percent and 1.5 percent and have a positive (dashed line) or negative (dotted line) linear relationship with the common-factor loadings.

model underestimates the desired capital—or positive. To illustrate the two possibilities, we generalize equation (1) to

$$V_i = \rho_{1,i} \cdot M_1 + \rho_{2,i} \cdot M_2 + \sqrt{1 - \rho_{1,i}^2 - \rho_{2,i}^2} \cdot Z_i, \tag{6}$$

where M_1, M_2, and Z are mutually independent standard normal variables.

In our first example, we consider a portfolio in which all exposures have equal weights, have the same PD, and are divided into two groups according to their dependence on the common factors. For exposures in the first group, $0 < \rho_{1,i} = \rho < 1$ and $\rho_{2,i} = 0$, while $\rho_{1,j} = 0$ and $0 < \rho_{2,j} = \rho < 1$ for exposures in the second group. Thus, the common factors are group specific and underpin positive and homogeneous within-group pairwise correlations and zero across-group correlations. The solid line in the top-right panel of figure 1 plots the desired capital measure for such a portfolio as a function of the relative weight of exposures in group 1.[15] This measure is lowest when the portfolio is most diversified between the two groups of exposures and, thus, the probability of large default losses is minimized. In addition, the dashed line in the panel plots an alternative capital measure, which is based on the ASRF model and, thus, incorporates a single-factor structure of the asset return correlations. We choose this structure optimally in order to provide the ASRF model with as much information about the true correlations as possible, subject to the single-factor constraint (see appendix 3). As it turns out, the single-factor approximation matches extremely well the true *average* asset return correlation but approximates only roughly the dispersion of correlation coefficients in the cross-section of exposures.

[15] The desired capital buffer is calculated on the basis of Monte Carlo simulations (see appendix 2) for a portfolio consisting of 1,000 exposures.

The difference between the dashed and solid lines equals the multifactor effect. This effect is *negative* because the single-factor assumption of the ASRF model ignores the fact that the common factors are two independent sources of default clustering, which leads to an underestimation of the desired capital. The underestimation is largest when the two groups enter the portfolio with equal weights, in which case the role of multiple factors is greatest.

It is possible, however, to construct another example, in which imposing an erroneous single-factor structure on portfolio credit risk distorts the interaction between asset return correlations and individual PDs in a way that leads to a *positive* multifactor effect. Consider a portfolio comprising two groups of exposures, with the exposures in the first group being individually riskier but less correlated among themselves than the exposures in the second group. In terms of equation (6), this can be formalized by postulating that firms with high PDs feature $0 < \rho_{1,i} = \rho < 1$ and $\rho_{2,i} = 0$, whereas firms with low PDs feature $0 < \rho_{1,j} = \rho_{2,j} = \rho < 1$. A single-factor approximation to this correlation structure would match the average correlation coefficient but would also imply too high a correlation among riskier exposures. Raising the probability of default clustering, this would lead to a capital buffer that is larger than desired.

2.3 Impact of Calibration Errors

Errors in the calibration of the ASRF model will affect assessments of portfolio credit risk even if this model is well specified. In this paper, we focus on errors in the calibration of the interdependence of credit risk across exposures, which can be driven by noise in the adopted values of asset return correlations or by a flawed assumption regarding the distribution of asset returns. When analyzing the consequences of such errors, we maintain our earlier practice and treat fattening of the tails of the loss distribution as implying unambiguously a higher 99.9 percent credit VaR and, thus, a higher desired level of the capital buffer. In this way, we sharpen the conceptual analysis and keep it in line with our empirical findings.

We study two general types of errors in calibrated asset return correlations: errors in the average correlation coefficient and errors in the dispersion of correlation coefficients across exposure pairs. It is important to keep in mind how these calibration-driven errors relate to errors induced by the two misspecification effects that we examined in sections 2.2.1 and 2.2.2. First, given that they occur within the constraints of the ASRF model, calibration errors are independent of the granularity effect. Second, given the properties of the one-factor approximation that underlies the extraction of the multifactor effect (see section 2.2.2), this effect is effectively separated from errors in the average correlation. By contrast, errors in the calibrated dispersion of asset return correlations could arise either as a result of ignoring the importance of multiple factors (recall section 2.2.2) or as a result of noise in the estimated factor loadings when there is a single common factor. In this section, we are concerned with the second case, as it is consistent with a correct specification of the ASRF model and refers only to calibration errors.

The two types of errors in calibrated asset return correlations have various potential sources. One possibility is that a user of the ASRF model relies entirely on rule-of-thumb values, which may simply be correlation estimates for a popular credit index. Such estimates will lead to a discrepancy between desired and calculated capital to the extent that the popular index is not representative of the user's own portfolio. Alternatively, a user of the model may have access only to short time series of data on the assets of the obligors in its portfolio, which would lead to small-sample estimation errors in asset return correlations. Indeed, this second source of error is likely to be important in practice because (i) asset value estimates are typically available at low (i.e., monthly or quarterly) frequencies and (ii) supervisory texts require that financial institutions possess only five years of relevant data.[16]

A positive error in the average level of asset return correlations leads to a capital measure that is higher than the desired one (figure 1, bottom-left panel). This result reflects the intuition that inflating asset return correlations increases the likelihood of default clustering, which fattens the tails of the loss distribution. In the remainder

[16]Data limitations are likely to be important irrespective of how a user of the model estimates asset return correlations. Such estimates may rely on balance sheet information and stock market data. Alternatively, as derived in Tarashev and Zhu (2006), asset return correlations can be extracted from the CDS market.

of this paper, the impact of errors in the *average* correlation on capital measures is dubbed the "correlation level" effect.

In turn, the effect of noise in the estimated dispersion of correlation coefficients can be seen in the following example. Suppose that all firms in one portfolio have homogeneous PDs and exhibit homogeneous pairwise asset return correlations. Suppose further that a second portfolio is characterized by the same PDs and average asset return correlation but includes a group of firms that are more likely to default together. The second portfolio, in which pairwise correlations exhibit dispersion, is more likely to experience several simultaneous defaults and, thus, has a loss distribution with fatter tails. Consequently, of the two portfolios, the second one requires higher capital in order to attain solvency with the same probability. This is portrayed by the upward slope of the solid line in the bottom-right panel of figure 1 and is a particular instance of what we dub the "correlation dispersion" effect, which arises in the context of a single common factor.

This result can be strengthened (dashed line in the same panel) but also weakened or even reversed if PDs vary across firms. To see why, consider the previous example but suppose that the strongly correlated firms in the second portfolio are the ones that have the lowest individual PDs. In other words, the firms that are likely to generate multiple defaults are less likely to default. As a result, greater dispersion of asset return correlations may lower the probability of default clustering in the second portfolio to an extent that depresses the desired capital level below that for the first portfolio. This is illustrated by the negative slope of the dotted line in the bottom-right panel of figure 1.

Even if asset return correlations were known with certainty, a flawed calibration of the overall distribution of asset returns would still drive errors in the calibrated interdependence of credit risk across exposures. Although the ASRF model imposes quite weak restrictions on asset return distributions, it is common practice to adopt distributions whose main advantages stem not from realistic features but from operational convenience. In particular, the consensus in the literature is that asset returns have fatter tails than those imposed by the conventional normality assumption. To the extent that the fatness of the tails reflects the distribution of the common factor, the probability of default clustering and, thus, the desired capital level would be higher than those derived under normality (Hull and White 2004; Tarashev and Zhu 2006). We study this issue by considering Student-t distributions for both the common and idiosyncratic factors of asset returns.

2.4 Evaluating Various Sources of Error

An important contribution of this paper is to present a unified empirical method for quantifying the impact of several sources of error in model-based assessment of portfolio credit risk. In particular, we focus on the difference between target capital measures and shortcut ones, the latter of which are based on the ASRF model and probable erroneous calibration of its parameters. We dissect this difference into four non-overlapping and exhaustive components, attributing them to the multifactor, granularity, correlation level, and correlation dispersion effects. In order to probe further the likely magnitude of the last two effects, we derive plausible small-sample errors that could affect direct estimates of asset return correlations. Finally, we also examine the implications of erroneous assumptions regarding the distribution of asset returns.

2.4.1 Baseline Method

The baseline empirical method consists of two general steps. In the first step, we construct a hypothetical portfolio that is either "large"—consisting of 1,000 equal exposures—or "small"—consisting of 200 equal exposures.[17] The sectoral composition of such a portfolio is constrained to be in line with the typical loan portfolio of large wholesale banks in the United States.[18] Given the constraint, the portfolio is sampled at random from the entire population of firms in our data set. Since each simulated portfolio is subject to sampling noise, we examine 3,000 different draws for both large and small portfolios.

[17] The distinction between what we dub large and small portfolios does not reflect the size of the aggregate exposure but rather different degrees of diversification across individual exposures (see section 3.2 for further detail).

[18] Such a portfolio does not incorporate consumer loans and, thus, may not be representative of all aspects of credit risk.

For a portfolio constructed in the first step, the second step calculates five alternative capital measures, which differ in the underlying assumptions regarding the interdependence of credit risk across exposures. Each of these alternatives employs the same set of PD values and assumes that asset returns are normally distributed. In addition, each alternative incorporates random LGDs that are independent of all random variables driving defaults. The LGD distributions are symmetric, triangular, peak at 50 percent, have a continuous support on the interval $[0\%, 100\%]$, and are identical and independent across exposures.[19]

Each measure differs from a previous one owing to a single assumption:

1. The *target capital* measure incorporates data on asset return correlations, which are treated as representing the "truth." Using these correlations, we conduct Monte Carlo simulations to construct the "true" probability distribution of default losses at the one-year horizon. The implied 99.9 percent credit VaR minus expected losses equals target capital (see appendix 2 for further detail).[20]
2. The second capital measure differs from target capital only owing to a restriction on the number of common factors governing asset returns. In particular, we adopt a correlation matrix that fits the original one as closely as possible under the constraint that correlation coefficients should be consistent with the presence of a single common factor (see appendix 3). The fitted single-factor correlation matrix is used to derive the one-year probability distribution of joint defaults

[19] The LGD specification warrants an explanation. The independence between the incidence of defaults and LGDs implies that, in the absence of simulation noise, only the mean of the LGD distribution enters (as a multiplicative factor) capital measures. However, the entire LGD distribution affects measures obtained from Monte Carlo simulations. Importantly, assuming a continuous distribution for LGDs smoothes the derived probability distribution of joint defaults, which improves the robustness of simulation-based capital measures.

[20] In order to reduce simulation errors to levels that do not affect our conclusions, we estimate each portfolio loss distribution on the basis of 500,000 draws. Alternatively, it is possible to circumvent the use of Monte Carlo simulations by approximating loss distributions analytically, on the basis of the so-called saddle-point method of Browne, Martin, and Thompson (2001).

on the basis of the so-called Gaussian copula method (see appendix 1). This distribution is then mapped into a probability distribution of default losses and, finally, into a capital measure.

3. The third capital measure differs from the second one only in that it assumes that all idiosyncratic risk is diversified away. This assumption allows us to use the fitted single-factor correlation matrix, which underpins measure 2 in the ASRF formula (equation (5)).
4. The fourth capital measure differs from the third one only in that it is based on the assumption that loading coefficients on the single common factor are the same across exposures. The resulting common correlation coefficient, which is set equal to the average of the pairwise correlations underpinning measures 2 and 3, is used as an input to the ASRF formula (equation (5)).
5. Finally, the *shortcut* capital measure differs from the fourth one only in that it incorporates alternative, rule-of-thumb, values for the common correlation coefficient.

The three intermediate measures lead to a straightforward dissection of the difference between target and shortcut capital.[21] Specifically, the difference between measures 5 and 1 is the sum of the following four components: (i) the difference between measures 2 and 1, which equals the multifactor effect; (ii) the difference between measures 3 and 2, which equals the granularity effect; (iii) the difference between measures 4 and 3, which equals the correlation dispersion effect; and (iv) the difference between measures 5 and 4, which equals the correlation level effect.

The specific ordering and choice of the three intermediate capital measures is a result of the following reasoning. As far as *measure 2* is concerned, it is determined by the necessity to extract the multifactor effect first. The reason for this is twofold. First, deriving a capital measure that assumes an infinite number of exposures but allows for

[21] Importantly, the method also applies to alternative definitions of *target* and *shortcut* capital, so long as the *true* correlation structure and shortcut correlation estimates chosen by the user are well defined.

multiple factors (i.e., extracting the granularity effect before the multifactor effect) is subject to approximation errors (see, e.g., Pykhtin 2004). Second, it is possible to isolate calibration errors (via measures 4 and 5) only after the extraction of the multifactor effect (via measure 2) has modified the original correlation matrix so that it is consistent with the ASRF model. Likewise, an application of this model for the extraction of calibration errors requires the assumption of infinite granularity, which explains why measure 3 is calculated before measures 4 and 5.[22] Finally, modifying measure 4 by preserving the cross-sectional distribution of the single-factor correlation coefficients but changing their average level would reverse the order in which the correlation level and dispersion effects are extracted. An important problem with this procedure is that it would allow only for an imperfect estimate of the correlation level effect because this estimate would be influenced by changes in the *structure* of the single-factor correlation matrix.

2.4.2 Two Extensions

In an attempt to delve further into the impact of *plausible* calibration errors on capital measures, we conduct two additional exercises. Each exercise focuses on a specific type of error in the calibrated interdependence of defaults and incorporates the assumption that the true PDs are identical across exposures. This assumption insulates capital measures from the impact of interaction between heterogeneous PDs and errors in the calibrated interdependence of defaults.

In the first exercise, we derive the extent to which plausible limitations on the size of available data can affect assessments of portfolio credit risk by affecting the estimates of asset return correlations. Specifically, we draw time series of asset returns from a joint distribution in which all pairwise correlations equal the correlation underpinning measure 4. Then we use the sample correlation matrix of the simulated series, a typical value for the probability of default,

[22]Despite this observation, we have also experimented with imposing the infinite granularity assumption only after we calculate the correlation dispersion and level effects (i.e., deriving measure 3 after measures 4 and 5). This modifies the meaning of these two effects but alters negligibly their magnitudes, as well as the magnitude of the granularity effect.

and the ASRF formula in equation (5) in order to derive a capital measure.[23] The difference between this measure and the desired capital, which employs the *exact* correlation structure, is driven by small-sample noise in the estimates of asset return correlations.

In our second exercise, we examine how measure 4 would change if the common and idiosyncratic factors of asset returns are in fact driven by Student-t distributions. The results of this exercise reveal how flawed calibration of the tail-of-distribution dependence among asset returns affects capital calculations. In order to carry out the exercise, we use the general ASRF formula in equation (4) and make two technical adjustments to the empirical setup. The first adjustment rescales the common and idiosyncratic factors to random variables with a unit variance.[24] The second adjustment addresses the fact that the generalized CDF of asset returns, $\mathcal{F}(\cdot)$, does not exist in closed form. In concrete terms, we calculate the default threshold $\mathcal{F}^{-1}(PD)$ on the basis of 10 million Monte Carlo simulations.

3. Data Description

This section describes the two major blocks of data that we rely on: (i) credit-risk parameter estimates provided by Moody's KMV and (ii) the sectoral distribution of exposures in typical portfolios of U.S. wholesale banks.

3.1 Credit-Risk Parameters

Our sample includes the universe of firms covered in July 2006 by both the expected default frequency (EDF$^{\text{TM}}$) model and the global correlation (GCorr$^{\text{TM}}$) model of Moody's KMV. These two models deliver, respectively, estimates of one-year physical PDs and physical asset return correlation coefficients for publicly traded companies. We abstract from financial firms and work with 10,891 companies.

The sample covers firms with diverse characteristics. Specifically, 5,709 of the firms are headquartered in the United States; 4,383 in Western Europe; and the remaining 799 in the rest of the world. Further, the distribution of the 10,891 firms across

[23] This measure abstracts from the granularity and multifactor effects.

[24] A Student-t variable with $r > 2$ degrees of freedom has a variance of $\frac{r}{r-2}$.

industrial sectors is reported in the last column in table 1, with the largest share of firms (10.4 percent) coming from the business service sector. Importantly, only 1,434 (or 13.2 percent) of the firms have a rating from either S&P or Moody's, which matches the stylized fact that the majority of bank exposures are unrated.

There are several reasons why EDFs and GCorr correlations are natural data for our exercise. First, the two measures are derived within mutually consistent frameworks, which build on the model of Merton (1974) and are in the spirit of the ASRF model (see Das and Ishii 2001, Crosbie and Bohn 2003, and Crosbie 2005 for detail). Second, in line with their role in this paper, Moody's KMV EDFs have been widely used as proxies for actual default probabilities (see, e.g., Berndt et al. 2005 and Longstaff, Mithal, and Neis 2005). Third, the GCorr correlations are underpinned by a multifactor structure, which is crucial for our study of the multifactor effect. In particular, this model incorporates 120 common factors that comprise 2 global economic factors, 5 regional economic factors, 7 sector factors, 61 industry-specific factors, and 45 country-specific factors.

Table 2 and figure 2 report summary statistics of the Moody's KMV one-year PD and asset return correlation estimates. The cross-sectional distribution of PDs has a long right tail and, thus, its median (0.39 percent) is much lower than its mean (2.67 percent). In addition, the favorable credit conditions in July 2006 resulted in 1,217 firms (i.e., about 11.2 percent of the total) having the lowest EDF score (0.02 percent) allowed by the Moody's KMV empirical methodology. At the same time, the upper bound on the Moody's KMV PD estimates (20 percent) is attained by 643 firms. For their part, GCorr correlations are limited between 0 and 65 percent. Clustered mainly between 5 percent and 25 percent, these correlations average 9.24 percent.[25]

[25]The GCorr correlation estimates are quite in line with correlation estimates reported in other studies. For instance, Lopez (2004) documents an average asset correlation of 12.5 percent for a large number of U.S. firms and Düllman, Scheicher, and Schmieder (2006) estimate a median asset return correlation of 10.1 percent for European firms.

Table 1. Sectoral Composition of Simulated Portfolios

	Large Portfolio		Small Portfolio		
Sector	**Number of Names**	**Exposure Weight (%)**	**Number of Names**	**Exposure Weight (%)**	**Number of Firms in the Sample**
Aerospace and Defense	31	3.1			105
Agriculture	9	0.9			56
Air Transportation	4	0.4			83
Apparel, Footwear, and Textiles	17	1.7			357
Automotive	51	5.1	19	9.5	198
Broadcast Media	43	4.3	16	8.0	191
Business Services	23	2.3			1,132
Chemicals	42	4.2	15	7.5	940
Computer Equipment	10	1.0			746
Construction	43	4.3	16	8.0	277
Electric, Gas, and Sanitary	107	10.7	39	19.5	335
Electronics and Electrical	18	1.8			693
Entertainment and Leisure	33	3.3			294
Fabricated Metals	17	1.7			146
Food, Beverages, and Tobacco	63	6.3	23	11.5	490
General Retail	31	3.1			133
Glass and Stone	6	0.6			149
Health Care	38	3.8	14	7.0	178
Legal and Other Services	16	1.6			452
Lodging	22	2.2			70
Machinery and Equipment	36	3.6	13	6.5	645

(continued)

Table 1. (Continued)

Sector	Large Portfolio		Small Portfolio		
	Number of Names	**Exposure Weight (%)**	**Number of Names**	**Exposure Weight (%)**	**Number of Firms in the Sample**
Medical Equipment	10	1.0			334
Mining	6	0.6			486
Miscellaneous Manufacturing	18	1.8			130
Nondefense Trans. and Parts	2	0.2			53
Oil and Gas Exploration	55	5.5	20	10.0	100
Oil Refining and Delivery	24	2.4			458
Other Trans. Services	22	2.2			108
Paper and Forestry	23	2.3			172
Personal Services	7	0.7			31
Primary Metals	11	1.1			188
Printing and Publishing	28	2.8			186
Repair Services and Rental	13	1.3			37
Restaurants	9	0.9			141
Rubber and Plastics	18	1.8			120
Semiconductors	2	0.2			177
Telecommunications	69	6.9	25	12.5	177
Trucking and Warehousing	5	0.5			68
Water Transportation	5	0.5			104
Wood, Furniture, and Fixtures	13	1.3			151
Total	1,000	100	200	100	10,891
Sector Concentration Index	0.0432		0.1135		
Name Concentration Index	0.0010		0.0050		

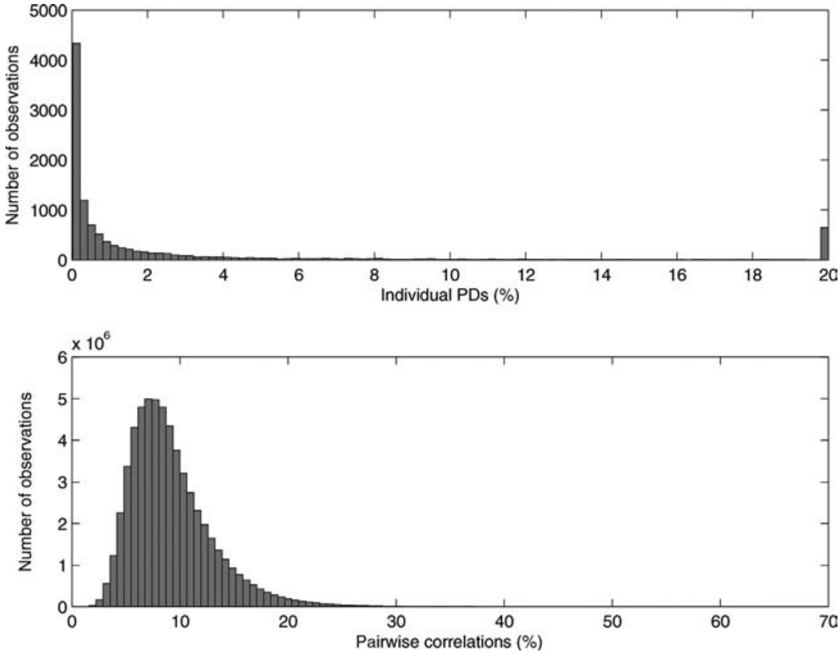
Note: A concentration index equals the sum of squared weights, set either at the firm or sector level.

Table 2. Summary Statistics (in percent)

	Mean	Std. Dev.	Skewness	Median	Minimum	Maximum
One-Year PDs	2.67	5.28	2.49	0.39	0.02	20.00
Pairwise Correlations	9.24	3.86	1.87	8.45	0.29	65.00

Note: The sample includes 10,891 nonfinancial firms.

Figure 2. Distribution of Individual PDs and Pairwise Correlations



Note: The parameter estimates relate to 10,891 nonfinancial firms in July 2006.
Data Source: Moody's KMV.

3.2 Characteristics of Hypothetical Portfolios

The portfolios we simulate match the sectoral distribution of the typical portfolio of U.S. wholesale banks. Specifically, to construct a large portfolio (1,000 exposures), we apply the forty nonfinancial sector weights reported by Heitfield, Burton, and Chomsisengphet (2006) (see table 1). For a small portfolio (200 exposures), we rescale the ten largest sectoral weights so that they sum to unity, and we set all other weights to zero. Within each sector, we draw firms at random.[26] All firms in a portfolio receive equal weights (up to a rounding error) and, thus, there is a one-to-one correspondence between the number of firms in a sector and that sector's weight in the portfolio.

The sector and name concentration indices, reported at the bottom of table 1, provide justification for our design of large and small portfolios. Calculated as the sum of squared sectoral (name) weights, the sector (name) concentration index of the large portfolios studied in this paper equals 0.0432 (0.001). This belongs to the range of concentration indices $[0.03, 0.045]$ ($[0.000, 0.003]$) reported in Heitfield, Burton, and Chomsisengphet (2006) for large portfolios of U.S. banks. For small portfolios, the analogous indices and ranges are 0.1135 (0.005) and $[0.035, 0.213]$ ($[0.001, 0.008]$), respectively.

4. Empirical Results

We implement the empirical methodology described in section 2.4 in order to quantify the impact of various sources of error in ASRF-based assessments of portfolio credit risk. Before reporting our findings, it is useful to highlight several aspects of the methodology.

First, as far as calibration of the model is concerned, the analysis in this paper focuses exclusively on errors in the values of parameters that relate to the interdependence of credit risk across exposures. Considering the impact of noise in PD and LGD estimates would make it extremely difficult to isolate the correlation level and dispersion effects we focus on. This is because noise in PDs and LGDs would interact with noise in correlation inputs in a highly nonlinear fashion.

[26] Within each industry sector, we draw randomly *with* replacement. If the same firm is drawn twice, the corresponding pairwise correlation is set equal to the average correlation for the sector. Drawing randomly *without* replacement does not affect materially the results.

Second, we make the stylized assumption that portfolios consist of equally weighted exposures. Considering disparate exposure sizes would require considering an additional dimension of portfolio characteristics, as it will no longer be the case that the granularity of a larger portfolio is necessarily finer. In addition, lower granularity that results from higher concentration in a small number of borrowers would also have a bearing on the number and importance of common factors affecting the portfolio and on the overall correlation of risk across exposures. This would make it impossible to isolate the granularity effect from the other three effects we consider.

Finally, our analysis treats the correlation matrix provided by Moody's KMV as revealing the "true" correlation of asset returns. Of course, this matrix is itself an estimate that is subject to errors. Nevertheless, the Moody's KMV correlation matrix provides a reasonable benchmark to work from. In addition, we have verified that results regarding the *relative* importance of alternative sources of error depend only marginally on the accuracy of the GCorr estimates, even though the *absolute* impact of alternative sources of error does change with the benchmark correlation level.

4.1 Various Errors in Shortcut Capital Measures

To study various sources of error in assessments of portfolio credit risk, we calculate the five capital measures listed in section 2.4 for 3,000 large and as many small hypothetical portfolios. Even though they have the same sectoral composition by construction, the simulated portfolios differ from each other with respect to the individual constituent exposures and, thus, with respect to the underlying risk parameters (see table 3).

Table 4 reports summary statistics of the target and shortcut capital measures (i.e., the two extremes described in section 2.4). For large portfolios, the target capital averages 3.31 percent (per unit of aggregate exposure) across the 3,000 simulated portfolios. The corresponding shortcut level (based on a rule-of-thumb asset return correlation of 12 percent) is 81 basis points higher.

Decomposing the difference between target and shortcut capital for large portfolios reveals that errors caused by model misspecification play a minor role. In qualitative terms, the multifactor effect

Table 3. Characteristics of Simulated Loan Portfolios (in percent)

A. Large Portfolios (1,000 Firms)					
	Mean	**Std. Dev.**	**Median**	**Minimum**	**Maximum**
Average PD	2.42	0.19	2.42	1.79	3.12
Std. Dev. of Individual PDs	5.16	0.26	5.16	4.25	6.14
Median PD	0.26	0.03	0.26	0.18	0.36
Average Correlation	9.78	0.22	9.77	9.14	10.73
Std. Dev. of Loadings	9.33	0.31	9.32	8.33	10.47
Corr. (PD, Loadings)	−20.00	2.04	−20.10	−26.70	−12.80
B. Small Portfolios (200 Firms)					
	Mean	**Std. Dev.**	**Median**	**Minimum**	**Maximum**
Average PD	2.28	0.36	2.26	1.24	3.68
Std. Dev. of Individual PDs	5.05	0.53	5.06	3.01	6.89
Median PD	0.24	0.05	0.23	0.11	0.55
Average Correlation	10.49	0.44	10.48	8.99	12.00
Std. Dev. of Loadings	10.54	0.70	10.55	7.80	12.79
Corr. (PD, Loadings)	−19.80	4.59	−20.20	−31.80	−1.20

Note: The results are based on 3,000 simulated portfolios and are obtained in two steps. First, portfolio-specific characteristics (specified by row headings) are calculated for each simulated portfolio. Second, summary statistics (specified by column headings) are calculated for each of the characteristics obtained in the first step. “Loadings” are estimated under a one-common-factor approximation of the correlation structure and refer to the firm-specific loadings of asset returns on the single common factor.

Table 4. Capital Measures and Four Sources of Errors (in percent)

	Mean	Std. Dev.	Median	95% Interval	50% Interval
A. Large Portfolios (1,000 Firms)					
Target Capital[a]	**3.31**	**0.18**	**3.30**	**[2.97, 3.67]**	**[3.19, 3.42]**
Deviation from Target Due To:[b]					
Multifactor Effect[c]	−0.03	0.02	−0.03	[−0.08, 0.01]	[−0.05, −0.02]
Granularity Effect[d]	−0.16	0.004	−0.15	[−0.16, −0.15]	[−0.16, −0.15]
Correlation Dispersion Effect[e]	0.39	0.05	0.39	[0.30, 0.48]	[0.36, 0.42]
Correlation Level Effect[f]	0.61	0.06	0.62	[0.49, 0.73]	[0.57, 0.66]
Shortcut Capital (corr = 12%)	**4.12**	**0.20**	**4.12**	**[3.75, 4.51]**	**[3.99, 4.25]**
Correlation Level Effect if:					
corr = 6%	−1.06	0.08	−1.06	[−1.23, −0.92]	[−1.11, −1.01]
corr = 18%	2.24	0.10	2.24	[2.05, 2.43]	[2.17, 2.30]
corr = 24%	3.86	0.14	3.86	[3.59, 4.14]	[3.77, 3.95]
B. Small Portfolios (200 Firms)					
	Mean	**Std. Dev.**	**Median**	**95% Interval**	**50% Interval**
Target Capital[a]	**3.86**	**0.32**	**3.85**	**[3.25, 4.55]**	**[3.64, 4.07]**
Deviation from Target Due To:[b]					
Multifactor Effect[c]	−0.04	0.04	−0.04	[−0.12, 0.04]	[−0.07, −0.01]
Granularity Effect[d]	−0.73	0.03	−0.73	[−0.79, −0.67]	[−0.75, −0.71]
Correlation Dispersion Effect[e]	0.42	0.12	0.42	[0.19, 0.65]	[0.34, 0.50]
Correlation Level Effect[f]	0.40	0.13	0.40	[0.16, 0.67]	[0.32, 0.48]
Shortcut Capital (corr = 12%)	**3.91**	**0.38**	**3.90**	**[3.17, 4.70]**	**[3.64, 4.16]**

(continued)

Table 4. (Continued)

B. Small Portfolios (200 Firms)					
	Mean	**Std. Dev.**	**Median**	**95% Interval**	**50% Interval**
Correlation Level Effect if:					
corr = 6%	−1.19	0.13	−1.19	[−1.46, −0.95]	[−1.28, −1.10]
corr = 18%	1.95	0.21	1.94	[1.57, 2.38]	[1.81, 2.08]
corr = 24%	3.50	0.30	3.49	[2.94, 4.12]	[3.30, 3.70]

Note: Summary statistics are for the simulated portfolios underpinning table 3. The column entitled "95% Interval" reports the 2.5th and 97.5th percentiles of the statistics specified in the particular row heading. The column entitled "50% Interval" reports the corresponding 25th and 75th percentiles. In all calculations, LGD_i has a symmetric triangular distribution between 0 and 1.

[a]Target capital is based on Moody's KMV estimates of PDs and asset return correlations and a Monte Carlo procedure for calculating the probability distribution of default losses.

[b]Four sources of deviation from the target capital level are shown; a negative sign implies underestimation. The sum of the target capital level and the four deviations equals the shortcut capital level. Each deviation is based on the assumptions underlying previous deviations plus one additional assumption.

[c]For the multifactor effect, the correlation matrix underpinning the target capital level is approximated under the assumption that there is a single common factor.

[d]For the granularity effect, there is the additional assumption that the number of firms is infinite.

[e]For the correlation dispersion effect, the additional assumption is that the loadings on the single common factor are the same across exposures.

[f]For the correlation level effect, the additional assumption imposes a different, shortcut level on the constant pairwise correlation.

can be of either sign (fourth column in table 4) but is more likely to be negative (fourth and fifth columns in table 4). In light of the discussion in section 2.2.2, imposing a single-factor framework is more likely to lead to too low a capital buffer because such a framework ignores the existence of multiple sources of default clustering. In quantitative terms, however, the multifactor effect entails an average discrepancy that amounts to less than 1 percent of the average target capital level.[27] This is because the single-factor approximation fits closely the raw correlation matrix. Indeed, our single-factor approximation matches almost perfectly the level of average correlations (with a maximum discrepancy across simulated portfolios of less than 4 basis points) and explains on average 76 percent of the variability of pairwise correlations in the cross-section of exposures.[28,29]

Similarly, the granularity effect is with the expected negative sign but, for large portfolios, leads to a small deviation from target capital. With an average of −16 basis points (or roughly 5 percent of target capital), this deviation is nonetheless significantly higher than that induced by the multifactor effect. Not surprisingly, the granularity effect we calculate is approximated extremely well by (the negative of) the closed-form granularity adjustment of Gordy and Lütkebohmert (2007), which averages −17 basis points for large portfolios in our sample. In addition, the correlation between the granularity effect and the granularity adjustment across large simulated portfolios is 66 percent.

By contrast, erroneous calibration of the ASRF model leads to much greater deviations from target. For large portfolios, the correlation dispersion effect raises the capital measure by 39 basis points,

[27]A similar result is obtained by Düllmann and Masschelein (2007), who rely on Pykhtin (2004) to approximate the multifactor effect in loan portfolios of German banks.

[28]The goodness-of-fit measure for the one-factor approximation is described in appendix 3. Across the 3,000 simulations of large portfolios, this measure ranges between 67 and 85 percent. For small portfolios, the range is between 63 and 86 percent.

[29]Principal-component analysis confirms this result. Specifically, the portion of the total variance of asset returns explained by the first principal component is at least ten times larger than the portion explained by the second principal component.

which amounts to roughly 12 percent of the target level.[30] The sign of this effect reflects the fact that, in our data, exposures with higher PDs tend to be less correlated with the rest of the portfolio (see the last row in each panel of table 3).[31] The shortcut capital measure ignores this regularity and, in line with the intuition provided in section 2.3, overshoots the target.

The correlation level effect has a similarly important implication. Specifically, this effect reveals that raising the average correlation coefficient from 9.78 percent (the one observed in the data) to a rule-of-thumb value of 12 percent leads to an 18 percent overestimation of the target capital level. The sign of the deviation is not surprising in light of the discussion in section 2.3. Importantly, the shortcut measure drops (rises) by roughly 8 percent with each percentage point decrease (increase) in the homogeneous correlation coefficient. Thus, using a rule-of-thumb correlation of 6 percent leads to a 32 percent *underestimation* of the target level.[32]

Turning to small portfolios, the decomposition results are qualitatively the same, with the notable exception of the granularity effect. In these portfolios, a much smaller portion of the idiosyncratic risk is diversified away and the granularity effect equals −73 basis points, which implies a 19 percent underestimation of the target capital. This underestimation is approximated well by the Gordy and Lütkebohmert (2007) granularity adjustment, (the negative of) which averages −86 basis points for small portfolios and exhibits an 89 percent correlation with the corresponding granularity effect.

4.2 Regression Analysis of Calibration Errors

Given the dominant role of correlation level and dispersion effects as determinants of model-based assessments of portfolio credit risk, we investigate the sources of these two effects via a regression analysis.

[30] On the basis of a hypothetical portfolio of U.S. firms, Hanson, Pesaran, and Schuermann (2007) also demonstrate the importance of accounting for cross-sectional heterogeneity in credit-risk parameters.

[31] The negative relationship between PDs and correlations (or loading coefficients in a single-factor setting) is likely to be a general phenomenon. See, e.g., Lopez (2004), Arora, Bohn, and Korablev (2005), and Dev (2006), who find that global factors often play bigger roles for firms of better credit quality.

[32] The rule-of-thumb asset return correlations reported in the literature range between 5 and 25 percent.

The regressions—run on the cross-section of simulated portfolios—are simple linear models of *calibration-driven capital discrepancy*, which is defined as shortcut capital (based on a correlation of 12 percent) minus target capital net of the multifactor and granularity effects.

We consider two blocks of explanatory variables. The first block comprises the average level and the dispersion of the asset return correlation coefficients underlying each simulated portfolio.[33] These variables are natural drivers of the correlation level and dispersion effects and would explain the two effects completely if assessments of portfolio credit risk did not depend on the interaction of asset return correlations with PDs. In order to account for such interaction, we include a second block of explanatory variables, which comprises average PDs and the cross-sectional correlations between PDs and single-factor loading coefficients. One would recall that the PDs underlying target capital are identical to those underlying the shortcut measure. Thus, the regression coefficient of the first variable in the second block reflects how a general rise in single-name credit risk *interacts* with the different average correlations and correlation structures behind the two capital measures. In turn, the coefficient of the last explanatory variable captures the component of the correlation dispersion effect that is driven by a systematic relationship between individual firms' riskiness and their dependence on the common factor.

The regression results, reported in table 5, reveal that the correlation level and dispersion variables have strong explanatory power. Depending on the portfolio size, these variables explain one-third or more of the variation in calibration-driven capital discrepancies across simulated portfolios and enter the regressions with statistically significant coefficients of the expected signs. First, given that the correlation underpinning the shortcut measure stays constant across simulated portfolios, the positive impact of a higher average correlation on target capital translates into a negative impact on capital discrepancy. Second, the correlation dispersion variable enters with a positive coefficient because—given that the empirical relationship between PDs and asset return correlations is negative

[33]The correlation dispersion variable is calculated as the standard deviation of the common-factor loading coefficients, which are obtained from the single-factor approximation of the correlation matrix.

Table 5. Explaining Calibration-Driven Capital Discrepancies

	Regression 1			Regression 2		
	Large Portfolio	**Small Portfolio**	**Pooled Sample**	**Large Portfolio**	**Small Portfolio**	**Pooled Sample**
Constant	3.01	3.00	3.29	2.58	2.40	2.46
	(65.6)	(53.0)	(111.6)	(129.6)	(84.0)	(159.8)
Average Corr.	−0.24	−0.23	−0.24	−0.28	−0.26	−0.27
	(34.8)	(31.5)	(48.8)	(95.6)	(74.4)	(114.5)
Std. Dev. of Loading Coefficients	0.037	0.022	0.010	0.031	0.029	0.028
	(7.5)	(4.8)	(3.1)	(14.6)	(13.9)	(19.9)
Average PD				0.21	0.19	0.19
				(86.2)	(61.6)	(94.2)
Corr. (PD, Loading Coefficient)				−0.018	−0.019	−0.019
				(80.5)	(79.5)	(113.9)
Adjusted R^2	0.39	0.33	0.56	0.89	0.86	0.91

Note: t-statistics are shown in parentheses. The regression is based on 3,000 simulations for each portfolio size. The dependent variable equals shortcut capital (based on asset return correlation of 12 percent) minus target capital, net of the granularity and multifactor effects (see table 4). Loading coefficients are estimated under a one-common-factor approximation of the correlation structure of asset returns. The last regressor equals the Kendall rank correlation between PDs and loading coefficients.

and that the shortcut capital measure abstracts from correlation dispersion—shortcut capital overpredicts the target level by more when correlation dispersion is greater. In order to visualize this phenomenon, refer back to the dotted line in the bottom-right panel of figure 1, which represents the case of a negative correlation between PDs and common-factor loadings. In this plot, shortcut capital appears at zero correlation dispersion ($c = 0$), and a rise in correlation dispersion (i.e., a rise in c) translates into a downward movement of target capital along the dotted line.

The second part of the analysis reveals that the main driver of the correlation level and dispersion effects is the *interaction* between correlation coefficients and PDs in assessments of portfolio credit risk. In particular, adding the second block of explanatory variables to the regression raises the goodness-of-fit measures (adjusted R^2) by 50 percentage points (to 89 percent) for large portfolios and by 53 percentage points (to 86 percent) for small portfolios. In addition, the positive statistically significant coefficient of average PDs indicates that, although an increase in this variable raises both the target and shortcut capital measures, the effect is stronger under the higher (homogeneous) asset return correlation, underpinning the latter measure. Finally, the statistically significant coefficient of the correlation between PDs and common-factor loadings is with the expected negative sign. This is because target capital tends to increase in the correlation between PDs and loadings on the single factor—as illustrated by an upward movement *across* the lines in the bottom-right panel of figure 1—whereas the shortcut measure abstracts from this correlation.

Importantly, the regression results are extremely robust across portfolio sizes. The robustness can be seen in that the values of the goodness-of-fit measures, as well as the coefficient estimates and t-statistics, obtained in the context of large portfolios match almost exactly their small-portfolio counterparts. In a further test of the robustness of the regression results, we pool observations across the two portfolio sizes and observe that all estimates change only marginally, leaving the message of the regression analysis intact.[34]

[34]Background checks reveal that the residuals of the full regressions can be attributed to a large extent to interactions among PDs and asset return correlations that are nonlinear and difficult to pin down.

4.3 Estimation Errors

The above results show that capital measures based on shortcut input estimates can deviate substantially from the target level. In practice, shortcut measures are likely to be adopted by less-sophisticated users of the ASRF model who face severe constraints in terms of data and analytical capacity. By contrast, larger and more-sophisticated users are likely to construct their own estimates of asset return correlations on the basis of in-house data. This section demonstrates that, for realistic sizes of such data, small-sample estimation errors in the correlation parameters are likely to lead to large flaws in assessments of portfolio credit risk.

In order to quantify plausible estimation errors, we consider a portfolio whose "true" credit-risk parameters match those of the "typical" portfolio in our data set. For this portfolio, we impose the simplifying assumption of homogeneous PDs (1 percent), LGDs (50 percent), and pairwise asset return correlations (9.78 percent) and consider different numbers of underlying exposures (see table 6). Abstracting from issues related to granularity and multiple factors, this assumption allows us to use the ASRF model, which implies that the desired capital buffer, dubbed "benchmark," equals 3.31 percent for each portfolio size. Referring to table 4, this is recognized as the average value of the target capital buffer examined in section 4.1 above.[35]

Then, we place ourselves in the shoes of a model user who does not know the exact asset return correlations but estimates them from available data.[36] Specifically, we endow the user with 60, 120, or 300 months of asset returns data—drawn from the true underlying distribution—and calculate the sample correlation matrix. In order to quantify a plausible range of errors in the estimate of the correlation matrix, we repeat this exercise 1,000 times. As reported in panels A and B of table 6, the sample correlations contain estimation errors that are likely to be substantial even for 300 months (or twenty-five years) of data.

[35]Given that we abstract from model misspecification in this subsection, the benchmark capital measure is conceptually equivalent to what we earlier called target capital.

[36]In order to focus on issues in the estimation of the interdependence of credit risk across exposures, we assume that the user knows the true PD and LGD.

Table 6. Impact of Estimation Errors (in percent)

	$N = 100$	$N = 200$	$N = 500$	$N = 1{,}000$
A. Sample Average of Pairwise Correlations				
$T = 60$	9.72 [6.5, 13.3]	9.67 [6.4, 13.3]	9.64 [6.6, 13.0]	9.63 [6.9, 12.7]
$T = 120$	9.77 [7.4, 12.4]	9.77 [7.6, 12.1]	9.76 [7.6, 12.1]	9.72 [7.7, 12.0]
$T = 300$	9.75 [8.3, 11.3]	9.79 [8.3, 11.3]	9.74 [8.4, 11.2]	9.77 [8.4, 11.25]
B. Sample Standard Deviation of Loading Coefficients				
	$N = 100$	$N = 200$	$N = 500$	$N = 1{,}000$
$T = 60$	12.11 [10.3, 14.1]	11.84 [10.5, 13.2]	11.65 [10.8, 12.5]	11.58 [10.8, 12.2]
$T = 120$	8.50 [7.4, 9.8]	8.34 [7.5, 9.1]	8.17 [7.6, 8.8]	8.13 [7.7, 8.6]
$T = 300$	5.35 [4.6, 6.2]	5.24 [4.7, 5.7]	5.16 [4.8, 5.5]	5.12 [4.9, 5.4]
C. Estimated Capital, Based on One-Factor Loading Structure				
	$N = 100$	$N = 200$	$N = 500$	$N = 1{,}000$
$T = 60$	3.87 [2.8, 5.1]	3.83 [2.8, 5.0]	3.80 [2.8, 4.9]	3.79 [2.9, 4.8]
$T = 120$	3.59 [2.9, 4.4]	3.58 [2.9, 4.4]	3.56 [2.9, 4.3]	3.54 [2.9, 4.3]
$T = 300$	3.41 [3.0, 3.9]	3.42 [3.0, 3.9]	3.40 [3.0, 3.9]	3.41 [3.0, 3.9]
Benchmark	3.31	3.31	3.31	3.31

Note: Results are based on 1,000 simulations of the asset returns of N firms over T months, with the true pairwise correlation, PD, and LGD fixed at 9.78 percent, 1 percent, and 50 percent, respectively. Each cell contains the mean and the 95 percent confidence interval of the average of sample estimates of pairwise correlations (panel A), the standard deviation of the loading coefficients in a one-factor approximateion (panel B), and the implied capital measure (panel C). The "Benchmark" row in panel C refers to the level of the capital measure obtained on the basis of the true risk parameters.

Panel C of this table reveals how estimation errors in correlation coefficients translate into deviation from the desired benchmark capital buffer. First, these deviations are affected little by the number of exposures in the portfolio. Second, at standard confidence levels, the deviations decrease in the size of the available time series of asset returns but remain substantial even if this size is assumed to be unrealistically large. For example, if a portfolio comprises 1,000 exposures and a user has 120 months of data, estimated capital buffers can deviate from the benchmark level by as much as 30 percent with a 95 percent probability. For longer time series, covering 300 months, the 95 percent confidence interval does become much tighter but remains consistent with errors as high as 18 percent of the benchmark capital level. Third, estimated capital buffers exhibit a positive bias relative to the benchmark level; i.e., their average level is invariably higher than 3.31 percent. This is because the true correlation structure is assumed to be homogeneous, while small-sample errors introduce dispersion in estimated correlation coefficients. By the intuition presented in section 2.3, this dispersion raises the implied capital buffer in the presence of homogeneous PDs.

4.4 Alternative Asset Return Distributions

There is general consensus in the literature that the distribution of asset returns have tails that are fatter than the tails of the convenient normal distribution. Importantly, an erroneous normality assumption tends to bias capital buffers downward to the extent that the empirical distribution of asset returns is driven by fat tails in the distribution of the common factor.[37] Such a distribution of the common factor implies great tail dependence among asset returns, which leads to a large probability of default clustering.

In order to quantify the impact of alternative asset return distributions on capital measures, we consider a homogeneous portfolio in which all PDs equal 1 percent, all LGDs equal 50 percent, and all asset return correlations equal 9.78 percent (the same as in section 4.3). Given these risk parameters, we follow the literature on

[37] For existing theoretical and empirical analysis of the treatment of tail dependencies by credit-risk models, see, e.g., Gordy (2000), Lucas et al. (2002), and Frey and McNeil (2003).

the pricing of portfolio credit risk (see Hull and White 2004; Kalemanova, Schmid, and Werner 2007) and consider the case in which both the common and idiosyncratic factors of asset returns have the same Student-t distribution. Experimenting with different distributional specifications, we do see that fatter tails of the distribution of the common factor (i.e., fewer degrees of freedom) translate into larger deviations from a capital buffer derived under the normality assumption (table 7, left panel).

In order to examine which distributional specification is supported by the data, we rely on time series of asset returns estimated by Moody's KMV. For each of the 10,891 firms in our sample, we use the available fifty-nine months of estimated returns (from September 2001 through July 2006) to calculate the sample kurtosis, which is the standard measure of tail fatness. The mean and median of this statistic across firms equal 7.96 and 5.28, respectively. Then, on the basis of 10,000 Monte Carlo simulations, we derive the distributions of the estimators of these mean and median when (i) the data size matches the size of the Moody's KMV data on asset return estimates and (ii) both the common and idiosyncratic factors of asset returns follow the same Student-t distribution. As revealed by the confidence intervals for these estimators, the Moody's KMV data support only the "double-t" specification with 3 degrees of freedom for each factor (table 7, center and right panels).

If the asset returns are indeed driven by such distributions—which would be in line with findings in Kalemanova, Schmid, and Werner (2007)—then a normality assumption will lead to tremendous underpredictions of the desired capital buffer. Specifically, we find that a double-t specification with 3 degrees of freedom implies a capital buffer of 6.17 percent per unit of aggregate exposure. This buffer is 86 percent higher than the capital buffer calculated under a normality assumption.

Even though this result may be undermined by probable errors in the Moody's KMV estimates of asset returns, alternative double-t specifications studied in the related literature also lead to capital measures that are significantly higher than those implied by a normal distribution. Indeed, given that the available time series of Moody's KMV asset return estimates are short, plausible *systemic* errors in these estimates across firms might affect substantially the cross-sectional mean and median of the sample kurtosis. This casts

Table 7. Alternative Distributional Assumptions

Degrees of Freedom[a]	Capital Measure[b] (in percent)	Mean Kurtosis[c]		Median Kurtosis[c]	
		Mean	95% Interval	Mean	95% Interval
(3, 3)	6.17	7.83	[7.30, 8.27]	5.54	[5.11, 6.54]
(4, 4)	4.59	5.72	[5.47, 5.92]	4.43	[4.25, 4.59]
(5, 5)	4.03	4.79	[4.64, 4.93]	3.96	[3.85, 4.06]
(6, 6)	3.81	4.31	[4.19, 4.40]	3.70	[3.63, 3.78]
(7, 7)	3.67	4.02	[3.93, 4.10]	3.55	[3.49, 3.60]
(8, 8)	3.59	3.83	[3.76, 3.89]	3.44	[3.39, 3.49]
(9, 9)	3.53	3.70	[3.64, 3.75]	3.36	[3.32, 3.40]
(10, 10)	3.49	3.60	[3.55, 3.65]	3.30	[3.26, 3.34]
(∞, ∞)	3.31	3.00	[2.99, 3.01]	2.89	[2.88, 2.90]

[a]This panel shows the degrees of freedom of the Student-t distribution of asset returns' common and idiosyncratic factors, respectively. The last row refers to a Gaussian specification.

[b]Obtained by applying the general ASRF formula (equation (4)) to a homogeneous portfolio, in which PD = 1 percent and LGD = 50 percent for each exposure and asset return correlations equal 9.78 percent for each pair of exposures. The default boundary for such exposures is calculated on the basis of 10 million simulations.

[c]This panel shows the means and the 95 percent confidence intervals (based on 10,000 Monte Carlo simulations) of the estimators of the cross-sectional mean and median of the kurtosis of asset returns when (i) the marginal Student-t distributions of the common and idiosyncratic factors have the degrees of freedom specified in the row heading and asset return correlations equal 9.78 percent, and (ii) the available data comprise time series of fifty-nine returns for 10,891 firms.

doubt on the validity of the double-t specification with 3 degrees of freedom and prompts us to consider alternative specifications, with 4 degrees of freedom (which is recommended by Hull and White 2004) and 5 degrees of freedom (which is reportedly a market standard). As revealed in table 7, these alternatives imply capital measures that are, respectively, 39 and 22 percent higher than the measure incorporating normal distributions.

5. Concluding Remarks

In this paper, we have quantified the relative importance of alternative sources of error in portfolio credit-risk measures based on the popular ASRF model. Our data have revealed that violations of key modeling assumptions—namely, that granularity is perfect and that there is a single common risk factor—are likely to have a limited impact on such measures, especially for large, well-diversified portfolios. By contrast, erroneous calibration of the ASRF model—driven by flaws in popular rule-of-thumb values of asset return correlation, plausible small-sample estimation errors, or a wrong assumption regarding the distribution of asset returns—has the potential to affect substantially measures of portfolio credit risk.

Given these results and the fact that we have abstracted from several additional sources of error in portfolio credit-risk measures, the task of risk managers and supervisors appears challenging. In particular, we have assumed that PDs and LGDs are free of estimation noise. However, such noise is likely to be sizable in practice and to interact with noise in correlation estimates in generating errors in measured portfolio credit risk. In addition, there may be time variation in credit-risk parameters that relate to the likelihood and severity of default losses as well as to the correlation of the occurrence of such losses across exposures. Such time variation, which could be due either to cyclical developments or to structural changes in credit markets, would impair the useful content of the available data and, thus, would make it even more difficult to measure portfolio credit risk.

That said, the ASRF model is by no means the only way to measure portfolio credit risk (see Gordy 2000). It is, thus, important to study the degree to which specification and calibration errors affect implications of alternative models as well. Such a generalization of the analysis in this paper would be valuable to risk managers and supervisors.

Appendix 1. Gaussian Copula

The Gaussian copula is an efficient algorithm for measuring portfolio credit risk when a portfolio consists of a finite number of exposures, the correlation matrix is driven by a common-factor loading structure, and underlying distributions are normal. The efficiency of the algorithm stems from the fact that, conditional on the realization of the common factor(s), default occurrences are independent across exposures. This allows for a closed-form solution for the conditional probability of joint defaults. The corresponding unconditional probability is then derived by integrating over the probability distribution of the common factor(s). For further detail, see Gibson (2004).

Appendix 2. Monte Carlo Simulations

Monte Carlo simulations deliver the target capital level. This method can be applied to any portfolio comprising N equally weighted exposures, provided that the exposure-specific probabilities of default, PD_i; the distribution of LGD_i; and the correlation matrix of asset returns, R; are known.

Given that LGD_i is assumed to be independent of the factors underlying PD_i, and that the distribution of LGD_i is identical across exposures, the simulation of portfolio credit losses can be divided into two parts. The first part calculates the probability distribution of joint defaults. Given this distribution, the second part incorporates the LGD distribution to derive the probability distribution of portfolio losses.

Specifically, drawing on section 2.1, we estimate the probability of joint defaults as follows:

1. Using the vector $\{PD_i\}_{i=1}^{N}$ and the assumption that asset returns are distributed as standard normal variables, we obtain an $N \times 1$ vector of default thresholds.
2. We draw an $N \times 1$ vector from N standard normal variables whose correlation matrix is R. The number of entries in this

vector that are smaller than the corresponding default threshold is the number of simulated defaults for the particular draw.

3. We repeat the previous step 500,000 times to derive the probability distribution of the number of defaults, $\Pr(nd = k)$, where nd refers to the number of defaults and $k = 0, 1, \cdots, N$.

Then, we estimate the probability distribution of portfolio credit losses as follows:

1. For a given number of defaults, k, we draw LGDs for the defaulted exposures 1,000 times and calculate the conditional loss distribution, $\Pr(TL|nd = k)$.
2. We conduct the above exercise for each $k = 1, \ldots, N$, and then calculate the unconditional probability distribution of portfolio credit losses. Specifically, $\Pr(TL) = \sum_k \Pr(TL|nd = k) \cdot \Pr(nd = k)$.

Finally, we set the capital measure to equal $TL_{1-\alpha} - \sum_{i=1}^{N} E(PD_i) \cdot E(LGD_i)$.

Appendix 3. Fitting a Single-Factor Correlation Structure

A single-factor approximation of an empirical correlation matrix is obtained as follows. Denote the empirical correlation matrix by Σ and its elements σ_{ij}, for $i, j \in \{1, \cdots, N\}$. The single-factor loading structure $\rho \equiv [\rho_1, \ldots \rho_N]$ that minimizes the discrepancies between the elements of Σ and their fitted counterparts are given by

$$\min_{\rho} \sum_{i=1}^{N-1} \sum_{j>i} \left(\sigma_{ij} - \rho_i \rho_j\right)^2.$$

Andersen, Sidenius, and Basu (2003) propose an efficient algorithm to solve this minimization problem. The fitted correlation matrix $\hat{\Sigma}$ has elements $\rho_i \rho_j$.

We also construct a measure that reflects the "explanatory power" of the single-factor approximation:

$$\text{Goodness-of-Fit Measure } \equiv 1 - \frac{var(\epsilon)}{var(\sigma)},$$

where σ is a vector of all pairwise correlation coefficients σ_{ij} $(i, j = 1, \cdots, N, i < j)$ and ϵ is a vector of the errors $\sigma_{ij} - \rho_i\rho_j$ $(i, j = 1, \cdots, N, i < j)$. This measure reflects the degree to which the cross-sectional variation in pairwise correlations can be explained by common-factor loadings in a single-factor framework.

References

Andersen, L., J. Sidenius, and S. Basu. 2003. "All Your Hedges in One Basket." *Risk* 16 (November): 67–72.

Arora, N., J. Bohn, and I. Korablev. 2005. "Power and Level Validation of the EDF^{TM} Credit Measure in the U.S. Market." Moody's KMV White Paper.

Basel Committee on Banking Supervision. 2005. "An Explanatory Note on the Basel II IRB Risk Weight Functions." BCBS Technical Note, Bank for International Settlements.

———. 2006. "Studies on Credit Risk Concentration: An Overview of the Issues and a Synopsis of the Results from the Research Task Force Project." BCBS Research Task Force Working Paper, Bank for International Settlements.

Berndt, A., R. Douglas, D. Duffie, M. Ferguson, and D. Schranz. 2005. "Measuring Default Risk Premia from Default Swap Rates and EDFs." BIS Working Paper No. 173.

Browne, C., R. Martin, and K. Thompson. 2001. "Taking to the Saddle." *Risk* 14 (June): 91–94.

Crosbie, P. 2005. "Global Correlation Factor Structure: Modeling Methodology." Moody's KMV Documents.

Crosbie, P., and J. Bohn. 2003. "Modeling Default Risk." Moody's KMV White Paper.

Das, A., and S. Ishii. 2001. "Methods for Calculating Asset Correlations: A Technical Note." KMV Documents.

Dev, A. 2006. "The Correlation Debate." *Risk* 19 (October).

Düllmann, K. 2006. "Measuring Business Sector Concentration by an Infection Model." Deutsche Bundesbank Discussion Paper Series 2, No. 2006/03.

Düllmann, K., and N. Masschelein. 2006. "Sector Concentration in Loan Portfolios and Economic Capital." Deutsche Bundesbank Discussion Paper Series 2, No. 2006/09.

———. 2007. "A Tractable Model to Measure Sector Concentration Risk in Credit Portfolios." *Journal of Financial Services Research* 32 (1): 55–79.

Düllmann, K., M. Scheicher, and C. Schmieder. 2006. "Asset Correlations and Credit Portfolio Risk: An Empirical Analysis." Working Paper.

Emmer, S., and D. Tasche. 2003. "Calculating Credit Risk Capital Charges with the One-Factor Model." *Journal of Risk* 7 (2): 85–101.

Frey, R., and A. J. McNeil. 2003. "Dependent Defaults in Models of Portfolio Credit Risk." *Journal of Risk* 6 (1): 59–92.

Garcia Cespedes, J. C., J. Antonio de Juan Herrero, A. Kreinin, and D. Rosen. 2006. "A Simple Multifactor 'Factor Adjustment' for the Treatment of Credit Capital Diversification." *Journal of Credit Risk* 2 (3): 57–85.

Gibson, M. S. 2004. "Understanding the Risk of Synthetic CDOs." Finance and Economics Discussion Series, No. 2004-36, Board of Governors of the Federal Reserve System.

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.

Gordy, M., and E. Lütkebohmert. 2007. "Granularity Adjustment for Basel II." Deutsche Bundesbank Discussion Paper No. 01/2007.

Gupta, S., A. McLeod, P. Teklos, and K. Thompson. 2005. "Time for Multi-Period Capital Models." *Risk* 18 (October): 74–78.

Hanson, S. G., M. H. Pesaran, and T. Schuermann. 2007. "Firm Heterogeneity and Credit Risk Diversification." Forthcoming in *Journal of Empirical Finance.*

Heitfield, E., S. Burton, and S. Chomsisengphet. 2006. "Systematic and Idiosyncratic Risk in Syndicated Loan Portfolios." *Journal of Credit Risk* 2 (3): 3–31.

Hull, J., and A. White. 2004. "Valuation of a CDO and an n-th to Default CDS without Monte Carlo Simulation." *Journal of Derivatives* 12 (Winter): 8–23.

Kalemanova, A., B. Schmid, and R. Werner. 2007. "The Normal Inverse Gaussian Distribution for Synthetic CDO Pricing." *Journal of Derivatives* 14 (Spring).

Kupiec, P. 2008. "A Generalized Single Common Factor Model of Portfolio Credit Risk." *Journal of Derivatives* 15 (Spring): 25–40.

Loeffler, G. 2003. "The Effects of Estimation Error on Measures of Portfolio Credit Risk." *Journal of Banking and Finance* 27 (8): 1427–53.

Longstaff, F. A., S. Mithal, and E. Neis. 2005. "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market." *Journal of Finance* 60 (5): 2213–53.

Lopez, J. A. 2004. "The Empirical Relationship between Average Asset Correlation, Firm Probability of Default and Asset Size." *Journal of Financial Intermediation* 13 (2): 265–83.

Lucas, A., P. Klaassen, P. Spreij, and S. Straetmans. 2002. "Extreme Tails for Linear Portfolio Credit Risk Models." In *Risk Measurement and Systemic Risk: Proceedings of the 3rd Joint Central Bank Research Conference.* Basel, Switzerland: Bank for International Settlements.

Martin, R., and T. Wilde. 2002. "Unsystematic Credit Risk." *Risk* 15 (November): 123–28.

Merton, R. C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–70.

Morinaga, S., and Y. Shiina. 2005. "Underestimation of Sector Concentration Risk by Mis-assignment of Borrowers." Working Paper.

Ordovas, R., and K. Thompson. 2003. "Credit Ensembles." *Risk* 16 (April): 67–72.

Pykhtin, M. 2004. "Multi-factor Adjustment." *Risk* 17 (March): 85–90.

Tarashev, N., and H. Zhu. 2006. "The Pricing of Portfolio Credit Risk." Bank for International Settlements Working Paper No. 214.

Vasicek, O. 1991. "Limiting Loan Loss Probability Distribution." KMV Working Paper.

———. 2002. "Loan Portfolio Value." *Risk* 15 (December): 160–62.

The Impact of Central Bank Announcements on Asset Prices in Real Time*

Carlo Rosa[a,b] and Giovanni Verga[c]
[a]Institute for the World Economy
[b]Centre for Economic Performance
[c]Department of Economics, University of Parma

This paper examines the effect of European Central Bank (ECB) communication on the price discovery process in the Euribor futures market using a *new* tick-by-tick data set. First, we show that two pieces of news systematically hit financial markets on Governing Council meeting days: the ECB policy rate decision and the explanation of its monetary policy stance. Second, we find that the *unexpected* component of ECB explanations has a significant and sizable impact on futures prices. Third, we investigate how communication interacts with learning by the public about the credibility of the central bank: financial market participants needed around three years, from 1999 through 2001, to learn how to interpret and believe ECB announcements. Finally, our results suggest that the Euribor futures market is efficient.

JEL Codes: E52, E58, G14.

*We would like to thank seminar participants at the III Banca d'Italia/CEPR Conference on Money, Banking and Finance, Royal Economic Society annual meeting 2007, Kiel Institute for the World Economy, London School of Economics, LUISS, and University of Ancona. We are grateful for comments to the discussant Filippo Altissimo, Gianluca Benigno, Willem Buiter, Kai Carstensen, Michael Ehrmann, Christian Julliard, Thomas Lux, Jochen Mankart, Andrew Patton, Riccardo Puglisi, Dennis Snower, Rick Van Der Ploeg, Pietro Veronesi, and Alwyn Young. We also thank the *International Journal of Central Banking* editor Hyun Song Shin for detailed comments on an earlier draft of the paper that led to significant improvements. This paper is based on a chapter of Carlo Rosa's doctoral dissertation at the London School of Economics. The first author is extremely grateful to Gianluca Benigno for his guidance, continuous stimulus, and support throughout his graduate studies, and to the

"Financial markets evidence indeed indicates that we as a central bank have managed to be understood by market professionals. . . . One of the main goals of a responsible central bank now is to have a reliable communication guiding expectations in a rapidly changing environment."

(Trichet 2005)

"A month ago Jean-Claude Trichet gave what markets see as his standard nod and wink: the European Central Bank (ECB), said its president, would continue to exercise 'vigilance' against inflationary pressures. Stand by, in other words, for another increase in interest rates at the bank's next rate-setting meeting on October 5. ECB-watchers were therefore well prepared when rates duly rose, by a quarter of a percentage point, to 3.25%."

(*The Economist*, October 2006)

1. Introduction

Do European Central Bank (ECB) qualitative announcements affect market expectations about the future path of the ECB's monetary policy? How long does it take for these messages to be promptly incorporated into asset prices? Is it possible to assess the effect of central bank statements without relying on the exogeneity assumption of monetary policy shocks? What is the degree of efficiency of the Euribor futures market? To address these questions, this paper uses a novel data set to present extensive evidence regarding the impact of central bank announcements on asset prices.

This work contributes to the rapidly expanding literature on central bank communication. Since central banking is increasingly becoming the art of managing expectations, communication has developed into a key monetary policy instrument. The value added of this study consists in examining the real-time effects of central bank announcements on financial markets. We show that market

Ph.D. examiners Charles Goodhart and Paul Mizen for their perceptive comments. All remaining errors are ours. Author contact: Rosa: Institute for the World Economy, Düsternbrooker Weg 120, 24105 Kiel, Germany *or* Centre for Economic Performance, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom; E-mail: carlo.rosa@ifw-kiel.de; Web site: http://carlorosa1.googlepages.com. Verga: Department of Economics, University of Parma, Via Kennedy 6, 43100 Parma, Italy; E-mail: giovanni.verga@unipr.it.

participants respond to *two different pieces of news* rather than just one piece of news, as is commonly analyzed in the monetary economics literature. Therefore, in order to properly describe the central bank conduct of monetary policy, we find that two dimensions are needed: both central bank actions and central bank words. Put differently, whereas the workhorse model used so far in the literature (Kuttner 2001) has been based only on monetary policy shocks—i.e., a single factor—our results suggest that nowadays central banks are also able to affect asset prices through their bias statements, a second policy instrument.

We apply our empirical methodology to investigate the effectiveness of the ECB's communication policy because of its unique institutional characteristic of first announcing its policy rate decisions (i.e., the minimum bid rate for the main refinancing operations of the Eurosystem) and then, after about forty-five minutes, explaining its monetary policy stance. By using a *new* intraday data set, we are able to neatly investigate the effects on asset prices of the latter announcement separately from the policy decision. We can thus circumvent the endogeneity and omitted-variables problems (i.e., interest rate changes and monetary policy shocks can be influenced by each other and by other common variables) that affect most of the previous work.

Our main findings can be summarized as follows. First, by using number-of-transactions and volumes data, we show qualitatively that financial markets immediately react to the two pieces of news that systematically reach them on Governing Council meeting days—i.e., the policy decision announcement communicated at 12:45 (throughout the paper, London time is always used) and the information released at the ECB President's monthly press conference starting at 13:30.

Second, by using a new tick-by-tick data set, we show quantitatively that changes in market interest rates can be explained by *unexpected* ECB announcements—i.e., the difference between what the ECB announces and what the market expects the ECB to announce. In particular, we find that the news shock is not only statistically and economically significant, but also quantitatively important. This evidence suggests that ECB words and deeds have been consistent with each other; otherwise, market participants would not have reacted to central bank announcements.

Third, we qualify the degree of efficiency of the Euribor futures market: futures prices incorporate the news stemming from the ECB actions very quickly, in less than five minutes, and the news stemming from the ECB President's speech in around one hour.

Fourth, by estimating a state-space model, we investigate how communication interacts with learning by the public about the credibility of the central bank.[1] We find that the importance of the ECB press conference has increased over time: financial market participants needed around three years, from 1999 through 2001, to learn how to interpret and believe central bank announcements.

The rest of the paper is organized as follows. In the next section, we discuss the measurement of the tone of ECB announcements. In section 3, we describe the three-month Euribor futures data. And, in order to illustrate the advantages of using high-frequency data, we analyze the futures price dynamics in specific announcement days. Finally, we provide some qualitative analysis based on number of transactions and volumes of contracts exchanged during a trading day. In section 4, we estimate the effect of the news shock on futures rates using intraday tick-by-tick data. In section 5, we perform some important robustness checks and sensitivity analysis. In section 6, we analyze the implications of our findings for central bank communication and monetary policy. In section 7, we summarize and conclude by suggesting some important issues for future research.

2. Measuring the Tone of ECB Announcements

Since its inception, the ECB has paid considerable attention to its announcement policy, and especially to its choices of medium, form, and content. In order to communicate with the public effectively and address the informational needs of various target groups such

[1] We have in mind Blinder's (1998, 64) definition of credibility: "Matching deeds to words. . . . Credibility means that your pronouncements are believed—even though you are bound by no rule and may even have a short-run incentive to renege. In the real world, such credibility is not normally created by incentive-compatible compensation schemes nor by rigid precommitment. Rather, it is painstakingly built up by a history of matching deeds to words. A central bank that consistently does what it says will acquire credibility by this definition almost regardless of the institutional structure."

as politicians, academics, the press, and participants in the financial markets, the ECB uses various instruments. These include the *Monthly Bulletin*, the President's monthly press conference (and its Questions and Answers session), the Testimony to the Committee on Monetary Affairs of the European Parliament (which currently takes place four times per year), and frequent speeches by its President and/or members of the Governing Council.

In its *Monthly Bulletin* of November 2002 (page 64), the ECB noted that "the monthly press conferences held by the President and the Vice-President and the *Monthly Bulletin* are two of the most important communication channels adopted by the ECB." In particular, in its *Monthly Bulletin* of January 2006 (page 57), the ECB confirmed that its President's press conference "provides a detailed explanation of the economic outlook for the euro area and the risks to price stability. This communication is aimed at improving the public's understanding of the current decision and the possible future course of policy interest rates."

In this study we restrict the analysis to the text of the introductory statements to the monthly ECB press conferences. We consider this analysis as the first stage of a broader research agenda, which will eventually analyze all ECB channels of communication. The introductory statement to the monthly press conference represents a natural candidate to initiate this research agenda, since it is simple and systematic in terms of its frequency and structure.

In order to make the European monetary authority's statements suitable for statistical computation, we assign a number to each monthly announcement. This number is intended to summarize the ECB's overall monetary policy stance as communicated by its Governing Council.[2] In his press conferences, the ECB President employs

[2]The seminal paper of Romer and Romer (1989) pioneered this so-called narrative approach. In particular, they examined the records of Federal Reserve policy deliberations in order to identify exogenous (according to their claim) monetary policy shocks. More generally, the classification of statements is often referred to as content analysis (see Krippendorf 2004 and Weber 2004): it consists of a set of techniques to extract the content of a message. A similar methodology for the classification of monetary policy statements is also applied by, among others, Ehrmann and Fratzscher (2007a), Gerlach (2007), Guthrie and Wright (2000), and Jansen and De Haan (2005).

a very standardized form of language, and its main conclusions consist of a limited number of keywords or strings.[3] Hence, it is possible to represent explicitly our mapping between words and numbers (therefore, an ordered scale) through the construction of a glossary that is reported in table 1.[4] It should be emphasized that over the years, the ECB has made considerable effort to systematically explain to the public the "meaning" that it attaches to most of the code words reported in table 1. For instance, on 6 March 2003, during the Questions and Answers session, President Duisenberg highlighted that if the ECB "use[s] the word 'appropriate' we expect it [the level of the policy rate] to remain valid for a considerable period of time." In another occasion, on 6 July 2006, President Trichet commented that "it is up to you to draw the appropriate conclusion from the words that are in the introductory statement and those that I use in this session. Until now, it seems to me, these have been quite well understood."

The wording indicator, *Index*, is converted into a variable on a three-value scale from -1 to $+1$. The value of zero suggests that the current level of the repo rate is appropriate to maintain price stability over the medium term. The value -1 characterizes an easing period—it is possible that the repo rate will be cut in the near future—whereas the value $+1$ assigned to the press conference statement suggests the ECB desires tighter monetary conditions. In the appendix, we report our assigned value of risk, *Index*, for each ECB monetary policy announcement, along with a few examples of introductory statements and our coding.

Since words are not precise quantitative data, the ranking of statements according to their assessment of ECB future monetary policy moves (tightening, neutral, or easing) is necessarily influenced

[3] There is no difference between the tone of Duisenberg's and Trichet's speeches: the President, in fact, simply reads a statement prepared by the whole Governing Council.

[4] Even though we have done the coding of each statement by reading the full press conference, including its Questions and Answers section, the synthetic assessment of the ECB perceived risk to price stability seems the most important. The keywords reported in the glossary only serve to provide a *parsimonious* and *transparent* background of the coding.

Table 1. Glossary of ECB's Official Statements and Their Ranking

ECB's Main Statements: The Most Important Keywords	***Index***
Strong vigilance/vigilant [with regard to upside risks to price stability] It is imperative that upward pressure be contained Monitor closely/carefully all upward risks/pressures Risks to price stability are upward/upside Upward pressure remains	+1
Appropriate Favorable Compatible Consistent In line Balanced Absence of significant pressures either upward or downward	0
Appropriate/favorable, but there are/remain some [downside] risks Some of the downward risks had materialized Downside risks are relevant/still cause for concern	−1
Note: Based on Rosa and Verga (2006).	

by personal judgment. Although we acknowledge that our assessment is subjective, given the ECB's success in explaining its meaning, the overall tone of its announcements is usually unambiguous. Moreover, since the main goal of this article is to propose a methodology to assess the informational value of the ECB press conference and disentangle the effects of qualitative announcements compared with monetary policy decisions, it is important to note that *the econometric results* that we present in this work are qualitatively very similar and remain highly statistically significant even if we use *other people's wording indicators* of the ECB monetary policy stance, such as those of Musard-Gies (2006).

3. Euribor Futures Market Data

3.1 Description

On 1 January 1999, the euro became Europe's main currency. Since then, new financial markets have been set up, including the Euribor,[5] the Eonia (Euro OverNight Index Average), and the euro-denominated short-term interest rate derivatives market. In particular, the three-month Euribor futures contracts are cash-settled short-term interest rate financial instruments with the Euribor rate for a three-month euro deposit of a face value of €1,000,000 as the underlying asset.

The Euribor futures contract that we consider in this study is traded at the Euronext Liffe (London International Financial Futures and Options Exchange) from 7 to 18. Futures prices are quoted on a daily basis, and the contracted interest rate equals 100 less the futures price. Each contract moves in fixed increments (or discrete units/ticks) of 0.005, which corresponds to a value of €12.5. The last trading day of each futures contract is two trading days prior to the third Wednesday of the delivery month, while the delivery date is the first business day after the last trading day. At a given point in time, twenty-five contracts are usually being actively traded. The standard delivery months are March, June, September, and December, known as quarterly expiries. There are also serial expiry contracts that expire in the nearest following six calendar months and that do not correspond to the quarterly sequence. Typically, serial expiry contracts exhibit lower liquidity.

[5] The Euribor (Euro Interbank Offered Rate) is a daily reference rate based on the interest rates at which banks offer to lend unsecured funds to other banks in the euro wholesale (or "interbank") money market. The Euribor is determined (fixed) by the European Banking Federation (EBF) at about 10:00 each day and is a filtered average of interbank deposit rates offered by a large panel of designated contributor banks (currently more than fifty), for maturities ranging from one week to one year. Euribor rates can be downloaded at www.euribor.org.

It is possible to build two different types of futures price time series: by position and by contract. Position time series are constructed by merging price data of different futures contracts. At a given point in time, the first position is defined as the contract that expires next in the quarterly sequence. The second position is represented by the second contract to expire in the same quarterly sequence. The third, fourth, etc., positions are constructed similarly. On the other hand, as the name suggests, the contract time series starts on the opening date of the contract and stops when the futures contract expires.

In this paper, we restrict our attention to the first-position three-month Euribor futures contracts (basically the three-month-forward three-month-ahead—implicit—Euribor rate) for two reasons. First, we do not need to adjust futures prices for a different number of months left to expiration and thus we avoid unnecessary complications. For example, Piccinato et al. (1999) find that the intraday statistical properties of futures prices are a function of the time remaining before expiry (i.e., seasonality that depends on the "time-to-maturity" effect). Second, studying futures by position can be justified on the basis of how the futures market works. In fact, in order to stay in the market, traders holding close-to-expiry contracts need to roll their position forward into the next expiry futures contract. By doing so, they are constructing a time series by position that extends beyond the expiry of each contract.

Nowadays, the first-position contract displays very high liquidity. For instance, during the last quarter of 2005, the average daily volume (i.e., number of exchanged contracts) was approximately 125,400 futures contracts, an increase of 50 percent compared with the same period in 2004 (83,842 futures contracts) and 68 percent compared with two years earlier (74,317 in the fourth quarter of 2003).

The data used in this study is provided by the Institute for Financial Markets (www.theifm.org). The data set contains several pieces of information, such as transaction-by-transaction price (around 2,500,000 transaction ticks), time-of-trade execution to the nearest second (both January 1999–June 2006), and volumes (July 2003–June 2006). We have trade data (transaction prices) in our database. However, we do not have bid-ask quotes.

3.2 Specific Announcement Days

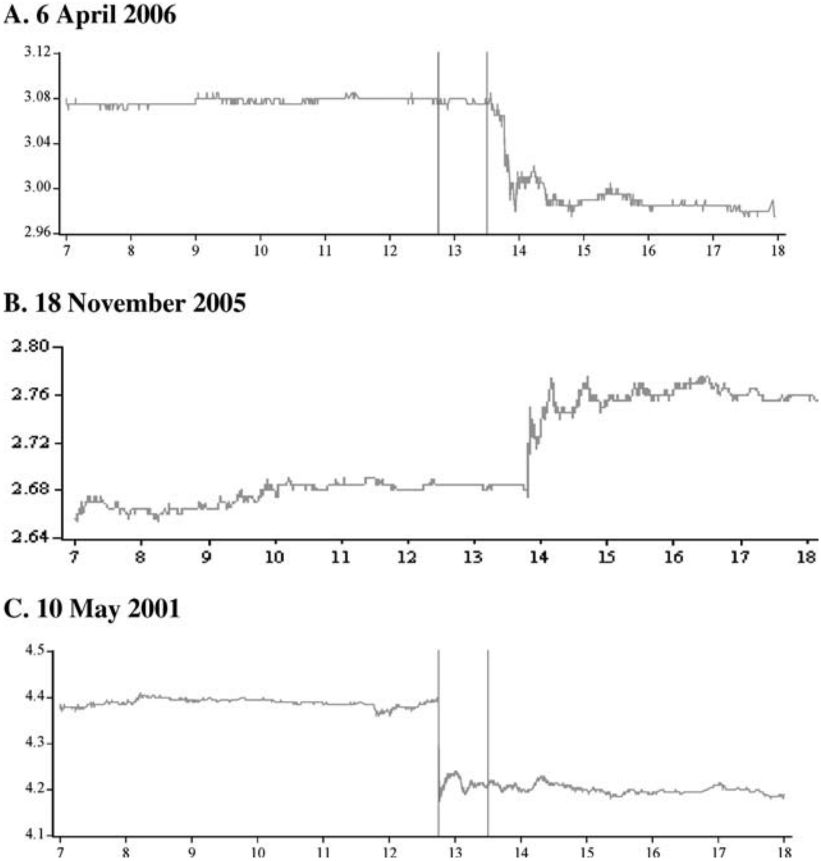
The ECB conduct of monetary policy is characterized by the unique institutional feature that on the same day and at two different points in time, the ECB Governing Council announces its monetary policy decision and explains its monetary policy stance. At 12:45, the ECB communicates the new level of its policy rate through a press release. Forty-five minutes later, at 13:30, the monthly press conference starts and the ECB President explains to the public the monetary policy decision taken and also the Governing Council's view of recent economic developments. The speech is very important, especially for traders, because it conveys strong hints about the future path of ECB monetary policy.

The advantages of using high-frequency data are best illustrated in figure 1, which reports the tick-by-tick three-month Euribor futures price movements on three specific days.

- On 6 April 2006 (Governing Council meeting day), financial market participants fully anticipated the ECB policy rate decision. Indeed, there was no sharp market reaction at around 12:45. However, the futures rate fell sharply at around 14:00. Everything happened in twenty minutes: it went from a rate of 3.085 at 13:33 to 2.98 at 13:56. One explanation could be a more dovish (than expected) speech given by ECB President Trichet; recall that the press conference starts at 13:30. One of his answers to journalists' questions (reported below) was extremely clear about ECB future monetary policy moves, and it may shed light on the immediate response of the three-month futures price movements. In his monthly introductory statement, the President did not mention explicitly the keyword "vigilant," which seems to indicate a strong risk for policy rate spikes in the near future (cf. glossary in table 1).

  > **Question:** Mr. Trichet, the markets were expecting you to say vigilance in order to prepare them or prepare for an interest rate rise in May. *You did not say vigilance, was that deliberate?* And second, did the Council discuss raising rates today?
  > **Trichet:** As we do in all our meetings which concentrate on monetary policy, we discussed the issue of rates. We

Figure 1. Three-Month Euribor Futures Tick-by-Tick Rate



Note: The horizontal axis represents the time of the day, while the vertical axis represents the futures rate that equals 100 less the futures price. The vertical lines are placed at 12:45 and 13:30.

discussed it at length. It is our responsibility to be as clear and transparent as possible with market participants, investors and savers. *I would say that the current suggestions regarding the high probability of an increase of rates in our next meeting do not correspond to the*

> *present sentiment of the Governing Council. I would also add that the sentiment that I see from time to time in some remarks or market literature concerning the perception that we do not increase rates when we are out of Frankfurt is equally not at all the sentiment of the Governing Council.* I trust that, for the sake of clarity, transparency and simplicity, it was perhaps useful to make these two remarks. And *it is true, vigilance is not mentioned in the introductory remarks, as you very wisely remarked.* [Emphasis added]

This example illustrates two important points. First, the ECB is able to move asset prices using words alone, without any need for contemporaneously implementing policy deeds. Second, the immediate response of the futures price is consistent with the efficiency of the Euribor futures market.

- At the beginning of November 2005, the ECB left its policy rate on hold. However, Trichet said the Bank remained highly vigilant on inflation and stood ready to raise interest rates. He added, "We stand ready to move any time when it is required by our mandate and by the situation ... we are very clear that we clearly could move any time." On Friday afternoon, 18 November, at around 14:00 at the European Banking Congress in Frankfurt, Trichet told the press that "after two years and a half of maintaining rates at a historical low, I consider that the Governing Council is ready to take a decision to move interest rates from the present level in order to take into account the level of risk." Panel B of figure 1 clearly shows that traders immediately placed bets that the ECB would increase the policy rate in December: the three-month futures rate jumped up steeply. The message of this last example is that, as long as it is not fully anticipated, ECB communication is able to move asset prices on any day, not only during Governing Council scheduled meeting days.
- Finally, panel C reports a case where the news is represented by ECB monetary policy actions rather than by announcements on its overall monetary policy stance. Note that futures prices adjust immediately: it took less than sixty seconds to completely price in the monetary policy shock. Indeed, it can

be shown econometrically (see separate technical appendix, available at www.ijcb.org) that monetary policy shocks are completely incorporated in futures rates in less than five minutes.[6] Therefore, this finding implies that we can assess how central bank qualitative announcements affect yields, safely disregarding the surprise in the target release. Moreover, it suggests that financial markets seem to understand numbers better than words.

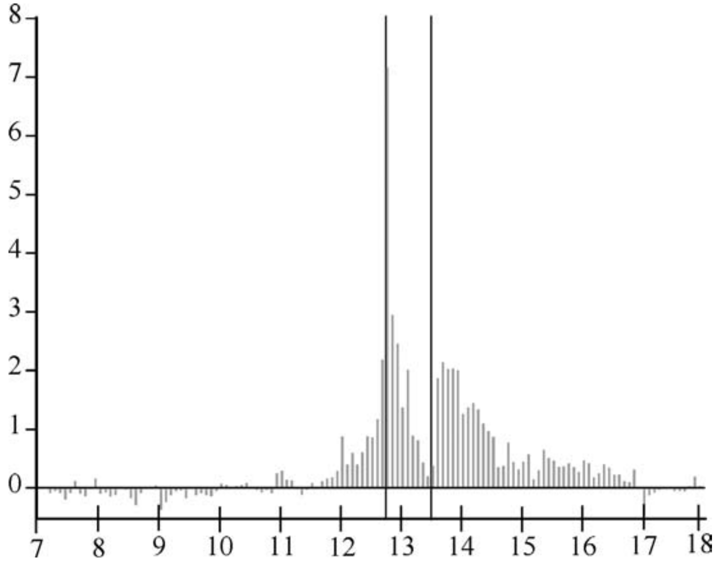
3.3 Qualitative Analysis of Volumes and Number of Transactions

In this subsection, in order to better understand the relationship between the price discovery process and ECB communication, we present some qualitative results on average number of transactions, a proxy for the market activity, and volumes by distinguishing between Governing Council meeting days and all other Thursdays. Note that since Governing Council meetings take place on Thursday and in order to explicitly take into account day-of-the-week effects, we compare market activity on Governing Council meeting days with market activity on all other Thursdays, rather than with activity on all other trading days.

Figure 2 shows that the five-minute average number of transactions is substantially higher on Governing Council meeting days (full sample January 2000–June 2006). A value larger than zero indicates that monetary policy decisions and communication induce a larger number of transactions than could be considered "normal" had the announcements not been made. For instance, a value of one indicates that the five-minute average number of transactions in that time window has been 100 percent higher during Governing

[6] We approximate monetary policy shocks by the difference between the new repo rate communicated at 12:45 and the one-month Euribor rate quoted at 10:00. By doing so, we are implicitly assuming that the risk premia have stayed constant during our sample period. An alternative measure of monetary policy shocks, which is free of both the risk-premium issue and market noise, is provided by survey data. However, as shown by Andersson (2007, appendix B), these two proxies are very similar, with a correlation coefficient of 0.75 for the ECB target surprise and 0.80 for the Federal Reserve target surprise. Moreover—and this is a crucial aspect for our exercise—expectations derived from the financial markets are real time, i.e., based on the latest available information.

Figure 2. Ratio of Average Number of Transactions per Quarter of an Hour (from January 2000)



Note: Plot of the ratio between the average number of transactions on Governing Council meeting days and all other Thursdays. Two vertical lines indicate 12:45 and 13:30 London time.

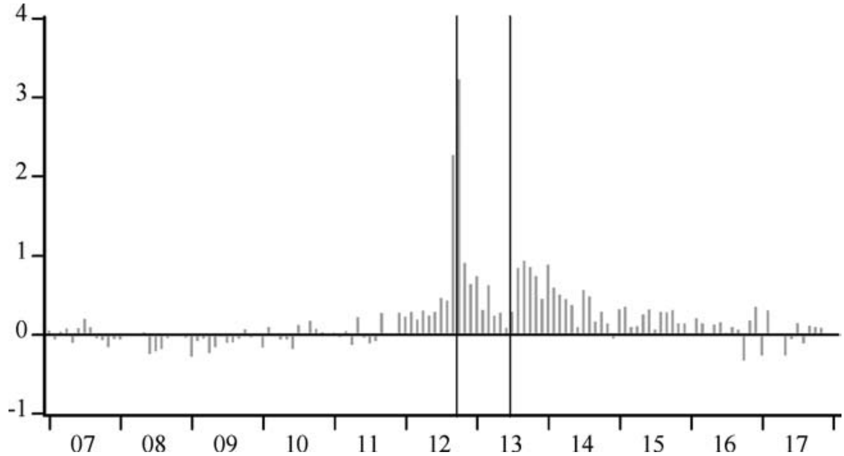
Council meeting days compared with all other Thursdays. Of course, the greater the news content of an announcement, the stronger the financial market activity should be.

It is eye-catching that there are two peaks: the first one corresponds closely to the new repo-rate announcements (12:45), while the second one takes place at the start of the ECB President's press conference (13:30). Market expectations seem to be quite heterogeneous at the time of the surprise, but then they start to converge. It is interesting to note that the convergence is much faster for monetary policy shocks than for news shocks. Apparently, quantitative announcements are easier to interpret than qualitative ones.

Figure 3 plots the ratio between futures price volatility on Governing Council meeting days with respect to all Thursdays.[7] Again,

[7]We use the absolute deviation of the (five-minute window) futures prices because it better captures the autocorrelation and the seasonality of the data

Figure 3. Five-Minute Futures Price Volatility (from January 2000)



Note: Plot of the ratio between futures price volatility (measured as the five-minute absolute deviation) on Governing Council meeting days with respect to all other Thursdays.

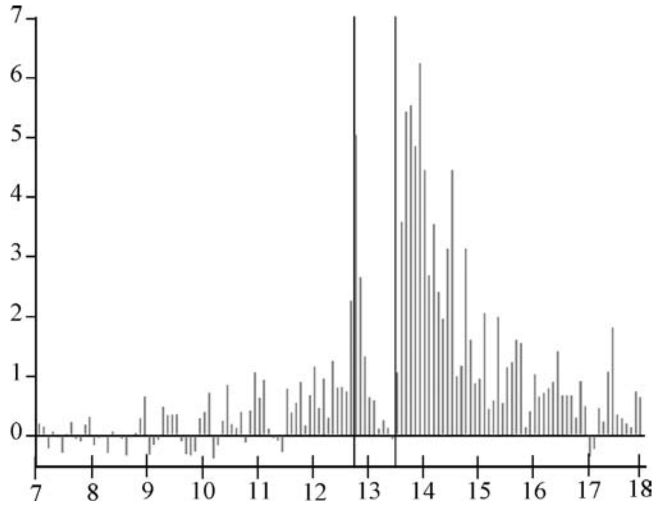
two pieces of news seem to systematically hit the market. Moreover, it is clear that futures prices incorporate the first one, the monetary policy shock, very quickly.

These findings are completely corroborated if we use average volumes rather than either average number of transactions or asset price volatility. In figure 4, we plot the ratio between average volumes on Governing Council meeting days and all other Thursdays for the sample period July 2003–June 2006 (recall that we do not have volume data before July 2003). Contrary to figures 2 and 3, financial market participants' expectations of future monetary policy actions now seem to be less heterogeneous than their expectations about ECB announcements.

This result is also confirmed by plotting the ratio between the average number of transactions on Governing Council meeting days

(Piccinato et al. 1999) compared with the more usual standard-deviation definition. For completeness, the latter measure of the volatility is also computed. There are no significant differences between the two definitions.

Figure 4. Ratio of Average Volume (from July 2003)

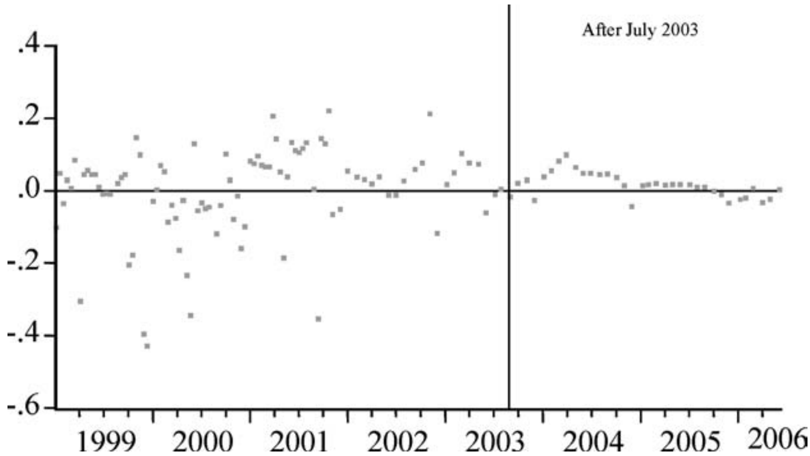


Note: Plot of the ratio between the average volume on Governing Council meeting days and all other Thursdays. Two vertical lines indicate 12:45 and 13:30 London time.

and all other Thursdays, restricting the sample period to July 2003–June 2006 (figure available in the working paper version of this article, Rosa and Verga 2006). A possible explanation for this phenomenon is that ECB monetary policy actions have recently become more predictable than they were at the beginning of the ECB's life. Indeed, figure 5 plots the monetary policy shock for the whole sample (January 1999–June 2006) and shows that its absolute value is much smaller nowadays. Recall also that from June 2003 to December 2005, the ECB did not move its policy rate.[8]

[8]Prior to November 2001, the ECB Governing Council could change the policy rate twice a month. For this reason, prior to November 2001, when we calculate the change in the one-month Euribor rate, we get not only the monetary policy surprise at the current meeting but also the change in market expectations for the next intramonth meeting. Figure 5 would remain qualitatively very similar, as would our conclusions, if we used the change in the one-week Euribor rates as a proxy of the monetary policy shock. Note that, unfortunately, two- and three-week Euribor rates data are not available prior to October 2001.

Figure 5. Monetary Policy Shocks



Note: The monetary policy shock is defined as the difference between the new repo rate communicated at 12:45 and the one-month Euribor rate quoted at 10:00. We add to it the mean equilibrium (liquidity and risk) spread between the repo and the one-month Euribor rate, in the specific case 0.11. The vertical line indicates July 2003, which corresponds to the starting point of the data on volumes.

Hence, on Governing Council meeting days, two pieces of news systematically hit financial markets: the ECB policy rate decision (standard in the literature) and the explanation of its monetary policy stance. In order to describe central bank monetary policy, we need two dimensions: both the current policy rate and its future path. We conclude that, at least qualitatively, financial markets seem to pay attention to both news items.

4. Tick-by-Tick Data: Econometric Results

In this section, we estimate the impact of unexpected central bank announcements on the short end of the term structure, using a *new* tick-by-tick data set from the Euribor futures market.

Since we are interested in investigating and measuring only the innovations in expectations caused by the ECB President's press conference, we restrict our econometric analysis to Governing Council meeting days. In other words, we apply a standard event-study approach (see, among others, Campbell, Lo, and MacKinlay 1997, chap. 4, and MacKinlay 1997).

Our goal is to assess quantitatively whether financial markets react to ECB communication and, more specifically, to examine the informational value of the press conference beyond that contained in the monetary policy decision. In this respect, it is crucially important to realize that the news does not consist of the ECB announcement itself but rather of its unexpected component, i.e., the difference between what the ECB declares and what the market expects the ECB to declare. Therefore, to verify empirically the effectiveness of ECB words, we need to proceed in two steps. First, we have to pin down what the market expects the ECB to declare. Second, we investigate the sensitivity of asset prices to the news shock.

We first posit and then verify empirically that the market tries to predict the new ECB announcement, $Index_t^{NEW}$, through the following regression:

$$Index_t^{NEW} = \alpha + \gamma_1 Index_t^{OLD} + \gamma_2\big(f_{t-h} - R_t^{NEW}\big) + \varepsilon_t, \qquad (1)$$

where α is a constant term and γs are regression coefficients. f_{t-h} stands for the Euribor futures rate quoted immediately before the press conference takes place, i.e., at 13:25. R_t^{NEW} stands for the new repo-rate level communicated at time 12:45. $Index_t^{OLD}$ is the wording indicator for the previous month's press conference. ε_t stands for a zero-mean noise term uncorrelated with the regressors.

In words, we assume that *Index* follows an AR(1) process: the economic environment usually does not change too much in the course of one month, and thus the ECB monetary policy stance and its statement also cannot be completely revised. However, in order to construct market expectations about ECB declarations using the very latest (indeed, real-time) market participants' information, we include as an explanatory variable the slope of the term structure immediately before the ECB President's press conference takes place, approximated by $f_{t-h} - R_t^{NEW}$. The rationale for including this term is as follows. If the futures rate (net of the risk premium already captured by α[9]) is higher than the new level of the repo rate, then the market expects the ECB to increase its policy rate in the near future. Hence, other things being equal, it expects a greater value of *Index* to be announced. In other words, if the short end of the term structure is upward sloping, a hawkish declaration is likely. Vice versa, if the short end of the term structure is downward sloping, then the market expects the ECB to cut its policy rate in the near future. Therefore, it expects a dovish announcement.

Since the wording indicator variable, *Index*, takes only discrete values (i.e., integers from -1 to $+1$), ordered-probit regression is the most appropriate estimator. Table 2 reports the estimated regressor coefficients of equation (1), together with its limit points δs, for the period January 2000–June 2006.[10] Interestingly, both coefficients γ_1 and γ_2 have the expected positive sign and are highly statistically significant.

Furthermore, the independent variables explain fairly well the announced tone of the ECB President's declaration (the goodness of fit measured by the pseudo-R^2 is around 0.57).

We construct market participants' expectations about the ECB announcement as follows:

$$E_{t-h}\left[Index_t^{NEW}\right] = \sum_{i=-1}^{+1} \Pr\left(Index_t^{NEW} = i\right) \cdot i, \tag{2}$$

where $E_{t-h}[.]$ stands for the expectation operator conditional on the time $t-h$ information set, which is immediately before the ECB's

[9] Note that the futures contract is different from a repo contract stipulated with the central bank. In fact, the futures contract refers (more or less) to a three-month-ahead three-month-forward rate, while the repo contract refers to an immediate one-month-forward rate. For this reason, a more sophisticated approach would consider a risk premium that varies over time. In this case, business-cycle indicators, such as the default spread (i.e., a return increase from high-grade to low-grade bonds, from bonds to stocks, and from large to small stocks) and term spread (i.e., premium for maturity risks from long-term to short-term securities), may track risk-premium dynamics.

[10] Our data set starts in January 1999. However, on the one hand, we consider the year 1999 as a period when financial market participants were learning to better interpret ECB announcements. On the other hand, the practitioners whom we consulted suggested that the Euribor futures market was not very liquid in the beginning. This fact is confirmed by volume and number-of-tick data in 1999 compared with the following years.

Table 2. Auxiliary Regression to Measure the Expected ECB Announcement using Ordered Probit

$f_{t-h} - R_t^{NEW}$	1.871^{**}
	(0.742)
$Index_t^{OLD}$	1.967^{***}
	(0.354)
δ_1	-1.553^{***}
	(0.348)
δ_2	1.180^{***}
	(0.333)
Log-Likelihood	−30.030
Pseudo-R^2	0.574
Observations	70

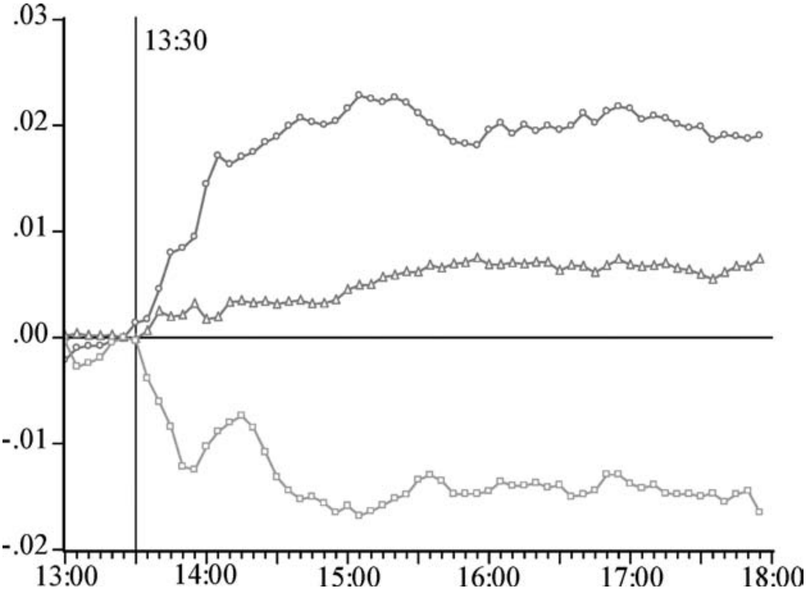
Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ML – Ordered Probit (Quadratic hill climbing). ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

press statement is released. $\Pr(Index_t^{NEW} = i)$ is computed analytically by the ordered-probit model (see Ruud 2000 or the working paper version of this article, Rosa and Verga 2006, for more details). Then we define the measure of unexpected central bank announcement, the *news shock*, as follows:

$$NS_t = Index_t^{NEW} - E_{t-h}\left[Index_t^{NEW}\right]. \tag{3}$$

Figure 6 shows the futures rate response to unexpected hawkish, neutral, and dovish announcements made by the ECB President during the monthly press conference. A statement is defined as hawkish when the news surprise belongs to the top 20th percentile of the news shocks observed in our sample period (i.e., 0.142). Vice versa, a statement is defined as dovish when the news surprise belongs to the bottom 20th percentile of the news shocks observed in our sample period (i.e., −0.228). In the remaining cases, the central bank statement is classified as neutral. Overall, the price response is consistent with the tone of the news. Indeed, futures rates increase after an unexpected hawkish announcement, decrease after a dovish one, and are basically unaffected by neutral declarations.

Figure 6. Futures Rate Reactions to Central Bank Announcements



Note: The chart plots average futures rate changes following unexpected hawkish (line with circles), neutral (line with triangles), and dovish (line with squares) ECB announcements on Governing Council meeting days. A statement is defined as hawkish when the news surprise, the difference between $Index_t^{NEW}$ and $E_{t-h}[Index_t^{NEW}]$, belongs to the top 20th percentile of the news shocks observed in our sample period (i.e., 0.142). A statement is defined as dovish when the news surprise belongs to the bottom 20th percentile of the news shocks observed in our sample period (i.e., −0.228). In the remaining cases, the central bank statement is classified as neutral. The horizontal axis is time of the day, and the vertical axis is futures rates (basis points/100).

We now test econometrically the effectiveness of ECB communication by estimating the following regression:

$$f_{t+h} - f_{t-h} = \alpha + \beta NS_t + \varepsilon_t, \tag{4}$$

where NS_t stands for the news surprise defined in equation (3), and f_{t-h} and f_{t+h} stand for, respectively, the futures rate quoted

Table 3. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t+h=$ 14:45 $t-h=$ 13:25	$t+h=$ 15:45 $t-h=$ 13:25	$t+h=$ 16:45 $t-h=$ 13:25
Constant	0.003	0.005	0.006
	(0.003)	(0.004)	(0.004)
NS_t	0.031**	0.033**	0.038***
	(0.012)	(0.013)	(0.014)
R^2	0.152	0.130	0.156
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

immediately before and immediately after the ECB announcement. The rest of the notation is the same as above.[11]

Table 3 reports the estimations for three different time windows—specifically, for $t + h =$ 14:45, 15:45, and 16:45—and for $t - h =$ 13:25 (all averaged for a five-minute interval).

It turns out that the coefficient on the news shock, β, is always positive (as expected) and statistically significant at least at the 5 percent level: the news shock can systematically explain the futures

[11] Since every Thursday at 13:30 there is the release of a U.S. jobless claims figure, as a robustness check we estimate equation (4) by also controlling for its surprise component, defined as the difference between the actual release and market expectations measured through the mean response of a Bloomberg survey among market participants. However, this surprise component is never significant, and for this reason it has been dropped. We kindly thank Michael Ehrmann for sharing with us the surprise component of the U.S. jobless claim figures.

As a further robustness check (see separate technical appendix) we reestimate equation (4) by also controlling for the surprise in the target release on the right-hand side, approximated by the difference between the new repo rate communicated at 12:45 and the one-month Euribor rate quoted at 10:00. All the econometric results hold both qualitatively and quantitatively. Hence, the dependent variable is not simply picking up a delayed effect of the policy rate release.

price change around the time of the ECB President's announcements. Table 3 suggests that the ECB can influence the money-market interest rates to some extent by simply using words, rather than deeds such as a change in its policy rate. For example, when the ECB President declares "it is imperative to contain upward pressure to price stability," while the market is expecting a value of *Index* of zero, the futures rate increases on average by about 4 basis points.[12] This finding shows that the ECB unexpected announcements have a significant and sizable impact on futures prices. To gain some idea of the importance of the news shock, note that the standard deviation of the daily percentage price change for the three-month futures is 0.026 on Thursdays with no Governing Council meeting. Thus, a one standard deviation of the news shock, corresponding to 0.365, leads to a price change of about 50 percent of the normal daily volatility of price changes. The size of the response of the futures price induced by the news shock is comparable to the reaction due to the monetary policy shock and generally much greater than the reaction of the bond market to macroeconomic news (see, e.g., Balduzzi, Elton, and Green 2001). Is the response of futures prices to the news shock "appropriate," or is it an overreaction? Unfortunately, in the absence of a fully developed structural asset pricing model, we cannot provide a definitive answer to this question. Preliminary evidence based both on visual inspection of figure 6 and more formal regression analysis indicates that the adjustment of futures rates to the news shock is gradual, thus suggesting that there may not be overreaction. Moreover, the additional volatility of futures prices on Governing Council meeting days compared with all other Thursdays (between 16:00 and 13:30) has the same order of magnitude of the reaction of futures rates to a one standard deviation of the news shock. Hence, market participants seem to give the "appropriate" weight to the new public information released by the central bank.

So far, we have shown that asset prices react to ECB communication, but how long does this reaction take? In other words, what

[12]Because of attenuation bias due to measurement error in the explanatory variable (Johnston and DiNardo 1996, 154) of equation (2) (also due to the artificial discreteness of our wording indicator), this number should be interpreted as a lower bound on the ECB's ability to move asset prices by simply making announcements.

Table 4. Time Needed to Incorporate the News Shock (Futures Rates), Dependent Variable $f_{17} - f_t$

	t = 14:15	t = 14:30	t = 14:45
Constant	0.002	0.003	0.003*
	(0.002)	(0.002)	(0.002)
NS_t	0.020**	0.012*	0.008
	(0.009)	(0.007)	(0.005)
R^2	0.120	0.054	0.030
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

is the degree of efficiency (i.e., the speed with which prices incorporate information) of the Euribor futures market? We answer this question, and hence we measure the speed of futures price response to ECB announcements, by estimating the following regression for different f_{t^*} (see Balduzzi, Elton, and Green 2001):

$$f_{17:00} - f_{t^*} = \alpha + \beta NS_t + \varepsilon_t, \tag{5}$$

where f_{t^*} stands for the (five-minute average) futures rate at time t^*.

Table 4 clearly indicates that as time unfolds, futures rates incorporate more and more the news shock. At around 14:45, unexpected announcements are no longer significant. Hence, about one hour after the end of the press conference, futures prices have completely incorporated the news originating from the ECB President's monthly press conference.

The seemingly quick adjustment in asset prices is consistent with the efficiency of the Euribor futures market. Also, Bernoth and von Hagen (2004) documented the efficiency of the Euribor futures market. However, they use daily data and thus are not able to quantify the degree of efficiency. Moreover, on a typical Governing Council meeting day, a large, potentially uncountable number of news items hit the financial market. This makes our estimates much more efficient, resulting in smaller standard errors in the coefficients of the news shock. Finally, and most importantly, Bernoth and von Hagen (2004) only analyze the response of futures rates to monetary policy shocks, while we separately identify the effect of the two systematic events that take place on Governing Council meeting days: the ECB policy rate announcement and, especially, the ECB press conference.

Throughout the paper, we use a five-minute average quotation rather than specific ticks, since the initial reaction of bond prices to the "unexpected" ECB announcement may be larger (overshooting) or smaller (undershooting) than its "true" effect (cf. Faust, Swanson, and Wright 2004). Asset prices should incorporate news instantaneously but actually do not. This procedure may introduce a possible bias in our estimations (cf. Blume and Stambaugh 1983). Nevertheless, we think that this bias is not important since we consider a very liquid market. Ideally, we want to give more importance to a quotation price that corresponds to a high traded volume. However, we cannot construct average futures prices weighted by volumes, since volume data are available only from July 2003.

As a further robustness check, we also rerun the previous regressions (see separate technical appendix) using equally spaced data instead of averaged tick-by-tick data. We construct these artificial data by linear interpolation of the transaction prices immediately before and after the relevant point in time. Then we obtain futures rate returns as the first difference of the new prices (see Andersen et al. 2003, 593). Our empirical findings discussed in this section are qualitatively very similar if we use equally spaced data.

5. Robustness Checks: Generated-Regressor Issue and State-Space Model

5.1 Generated-Regressor Issue

So far, the econometric estimations have been carried out in two steps. First, we determine market expectations about ECB announcements immediately before the start of the press conference. Then, we use the news shock to explain the futures price discovery process. In other words, in the second step we employ generated regressors (cf. Oxley and McAleer 1993).

This fact may give rise to underestimated standard errors and hence to spurious significant regressor coefficients. In order to solve this issue and to check the statistical validity of our conclusions, we reestimate the same baseline regression of the previous subsection all in one step. More formally, we estimate the following equation by OLS:

$$f_{t+h} - f_{t-h} = c_1 + c_2 \cdot \Big(Index_t^{NEW} \\ - c_3 \cdot \big(f_{t-h} - R_t^{NEW}\big) - c_4 \cdot Index_t^{OLD}\Big) + \varepsilon_t,$$

where cs are regressor coefficients, and the rest of the notation is the same as before.

The econometric results continue to hold both qualitatively and quantitatively (see separate technical appendix).

In order to account for the generated-regressor problem when computing coefficient estimates' standard errors, we also check the robustness of our conclusions by using a bootstrap approach to statistical inference (see, e.g., Efron and Tibshirani 1993). More specifically, we apply a sampling-with-replacement raw residuals bootstrap scheme with 1,000 repetitions. The empirical results (see separate technical appendix) are qualitatively very similar to those obtained in the previous section when White's (1980) robust standard errors are used. In particular, the 99 percent confidence bands of the coefficient of the news shock in equation (4) never include negative numbers. This fact confirms that the ECB is indeed able to move asset prices significantly in the desired direction.

5.2 State-Space Model

An implicit assumption of all the econometric models specified so far is that the regressor coefficient of the news shock remains constant over time. This implies that we have completely ruled out by assumption a learning period.

In this section, we specify and estimate a state-space model that explicitly allows us to incorporate unobservable variables, known as state variables, into the observable model. In other words, we relax the above assumption and allow the regressor coefficient of the news shock to vary over time. Specifically, we specify the following linear state-space representation:

$$
\begin{aligned}
f_{t+h} - f_{t-h} = \alpha + \beta_t \bigl(& Index_t^{NEW} - \gamma_1 Index_t^{OLD} \\
& - \gamma_2 \cdot \bigl(f_{t-h} - R_t^{NEW} \bigr) \bigr) + \varepsilon_t \\
\beta_t = \beta_{t-1} + u_t, &
\end{aligned}
$$

where t stands for a Governing Council meeting day. ε_t and u_t are random variables assumed to be serially independent and independently normally distributed. The rest of the notation is the same as before.

For simplicity, we assume that the unobserved state variable β_t moves over time as a first-order autoregression—specifically, a random-walk, stochastic process.

To solve for the model's parameters, we use the Kalman filter, which is a recursive algorithm for sequentially updating the one-step-ahead estimate of the state mean and variance (i.e., $E_{t-1}[\beta_t]$ and $Var_{t-1}[_{\beta_t}]$). In order to implement the Kalman filter, we maximize the sample log-likelihood function using numeric derivatives and standard iterative techniques (Marquardt optimization algorithm) and taking into account that ε_t and u_t are normally distributed.

Table 5 reports the estimation results. The regression coefficients γs are statistically significant and economically meaningful, i.e., with the expected positive sign.

The bottom part (row 4 of the table) displays the final one-step-ahead forecast value of the state variable, $E_{T-1}[\beta_T]$, where T stands for the final sample date, and its root mean-squared error (MSE) value. It is statistically significant and has a magnitude of 0.030.

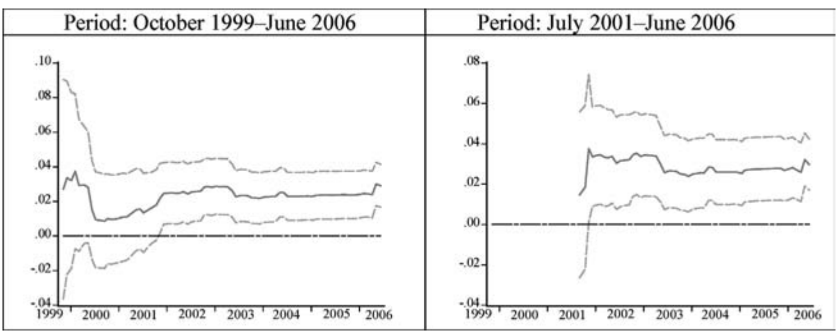
Figure 7 displays the entire path of the one-step-ahead forecast of the state variable together with its confidence bands. This chart suggests that the coefficient of the news shock has varied over time. In particular, financial market participants needed around three years to believe, and thus react to, ECB announcements. It is interesting

Table 5. State-Space Model (Futures Rates)

α	0.006* (0.003)
γ_1	0.775*** (0.260)
γ_2	1.112** (0.460)
$\beta_{T/T-1}$	0.030*** (0.006)
Log-Likelihood	154.590
Observations	73

Note: Monthly observations on days of ECB Governing Council meetings, October 1999–June 2006. The dependent variable is the five-minute average change in futures prices between 13:25 and 14:45. The econometric method is maximum likelihood (Marquardt optimization algorithm). ML standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Figure 7. One-Step-Ahead State Variable Prediction



Note: The chart displays the entire path of the state variable (solid line) together with its confidence bands of two standard deviations (dashed lines). Two sample periods have been considered: May 1999–June 2006 (left panel) and July 2001–June 2006 (right panel). Note that the results are mutually consistent. The signal variable is the five-minute average futures rate quoted at 14:45. To facilitate the readability of the right panel, we drop the first observation (centered at zero) that features very large confidence bands.

to see that after this learning period, the coefficient of the news shock has stayed relatively constant over time.

5.3 Using Alternative Wording Indicators

In order to investigate the effects of central bank qualitative announcements on asset prices, a key step is represented by the application of the narrative approach to categorize the hawkishness of the ECB rhetoric. In this section, as a robustness check we consider alternative classification schemes of the tone of the announcement. In particular, we use both a wording indicator based on a more finely graded scale ranging from −2 (strong inclination to lower rates) to +2 (strong inclination to increase rates), coded as *Index2*, and other people's wording indicators of the ECB monetary policy stance, such as those used by Musard-Gies (2006). Then, we employ these new *Indexes* to construct market participants' expectations about the ECB declaration, given by equation (2), and to construct the related news shock, given by equation (3). Finally, as we did before, we investigate the impact of the surprise component of central bank announcements on asset prices.

By reestimating the baseline regressions of section 4, we find that the econometric results remain qualitatively very similar and highly statistically significant despite the fact that we employ a different wording indicator.[13] However, we do find that a more finely graded scale produces more precise point estimates (see tables 6 and 7) compared with a scale that ranges from −1 to +1—i.e., one that only distinguishes between easing, neutral, and tightening. This evidence suggests that although a three-value classification of the announcement is arguably less controversial, it represents a measure of the tone of the statement that is too coarse and that may neglect important slight nuances. For instance, in the latter case, all the qualifications of the keyword "vigilance"—such as strong, extreme, moderate, etc. (see Jansen and de Haan 2007a for an exhaustive list)—are not differentiated with a substantial information loss.

[13]In the interest of space, we report in a separate technical appendix the econometric analysis using Musard-Gies' (2006) wording indicators.

Table 6. Auxiliary Regression to Measure the Expected ECB Announcement using Ordered Probit

$f_{t-h} - R_t^{NEW}$	2.448***
	(0.691)
$Index2_t^{OLD}$	1.496***
	(0.237)
δ_1	−2.896***
	(0.422)
δ_2	−1.648***
	(0.347)
δ_3	1.012***
	(0.288)
δ_4	3.254***
	(0.482)
Log-Likelihood	−47.688
Pseudo-R^2	0.538
Observations	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ML – Ordered Probit (Quadratic hill climbing). ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Table 7. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t+h=$ **14:45** $t-h=$ **13:30**	$t+h=$ **15:45** $t-h=$ **13:30**	$t+h=$ **16:45** $t-h=$ **13:30**
Constant	0.003	0.005	0.005
	(0.003)	(0.004)	(0.004)
NS_t	0.023***	0.028***	0.029***
	(0.006)	(0.007)	(0.007)
R^2	0.188	0.199	0.195
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

6. Central Bank Communication and Monetary Policymaking

Central bank communication and its effects on financial markets have recently received increasing attention in the monetary economics literature both theoretically (Woodford 2005) and empirically (Ehrmann and Fratzscher 2007a, 2007c; Gerlach 2007; Kohn and Sack 2003; Jansen and de Haan 2005, 2006, 2007a, 2007c; Rosa 2007a, 2007b, 2007c; Rosa and Verga 2007).

The workhorse model used so far in the literature (Kuttner 2001) to describe the effects of monetary policy on asset prices has been based only on monetary policy shocks, i.e., a single factor. However, nowadays central banks have adopted a more transparent conduct of monetary policy up to pre-announcing their future policy moves. Hence, it turns out that central banks mostly affect asset prices through their bias statements (a second policy instrument) by influencing financial market expectations of their future policy actions, rather than by unexpected deeds, i.e., monetary policy shocks. The former effect is not only significant but also has a sizable impact on futures prices.

In addition, Bomfim (2003) and Gurkaynak, Sack, and Swanson (2005) find that at least two factors are required in order to capture adequately the effects of U.S. monetary policy on asset prices. They interpret the first one as the current target federal funds rate and the second one as the future path of policy, which is closely associated with Federal Open Market Committee announcements. We solve a related empirical exercise. However, there remain important differences. On the one hand, the methodology is different. We first identify the surprise component of the ECB press conference. Then, we use it to explain the change in the futures rate. While Gürkaynak, Sack, and Swanson (2005) assume that the second factor of a factorial analysis on the futures price changes with maturity less than a year corresponds to central bank announcements. Then they use both factors to explain other asset price movements. By doing so, they implicitly assume that the two factors are at least weak exogenous with respect to bond and stock prices, while we do not make any exogeneity assumption. Put differently, *first* we measure explicitly the news shock and *then* explain its effects, while they do not interpret central bank statements simply because the surprise is posited equal to the second factor, rather than derived from first principles. On the other hand, we analyze the ECB, while they focus on the U.S. Federal Reserve. This is extremely important because we are able not only to separately and sequentially identify both the monetary policy and the news shock but also to separately investigate their effects. We also test the degree of efficiency of the Euribor futures market.

Rosa (2007b) applies the methodology developed in this paper to investigate the impact of the unexpected component of the Federal Reserve's monetary policy decisions and balance-of-risk statements on the full spectrum of the U.S. yield curve. Then he does a comparative study exercise between the effectiveness of the ECB and Federal Reserve communication on their domestic interest rates, and study the cross-effects—namely, the Federal Reserve's ability to move European interest rates and the corresponding ECB's capacity to move U.S. rates. In this paper, we also examine the effect of ECB communication on the price discovery process for the European money-market rates. However, we use *high-frequency intraday data* rather than daily data. As we mentioned above, this is a crucial improvement for the identification of the effects of the news shocks because it allows us to fully exploit the unique institutional feature of ECB monetary policy conduct, i.e., the fact that on the same Governing Council meeting day, the ECB announces its policy decision and explains its monetary policy stance at two different points in time. Since the monetary policy shock is immediately incorporated into asset prices, we can separately analyze the effects of the news shock, and thus we do not need to make any exogeneity assumption of the monetary policy shock that is typically encountered in the literature. Second, we can investigate how communication interacts with learning by the public about the credibility of the central bank. Third, by considering a narrower time window—i.e., the futures price change immediately before and after the press conference takes place—we not only avoid potential bias due to omitted variables, but we also obtain more efficient point estimates of the regressor coefficients. Finally, we provide a thorough qualitative description of the market reaction (volume, number of transactions, volatility, etc.) to qualitative information, and this is interesting from a market microstructure standpoint.

Brand, Buncic, and Turunen (2006) investigate the impact of ECB monetary policy decisions and communication on the yield curve by using high-frequency data. Their methodology is based on Gürkaynak, Sack, and Swanson (2005), and thus the news shock is assumed rather than derived from first principles. Moreover, we use futures tick-by-tick data from LIFFE (and we complement our analysis by studying volumes and number-of-transactions data), while they use real-time quotes of deposit and swap rates from Reuters observed at five-minute intervals. Also, Ehrmann and Fratzscher (2007b) analyze the information content of the ECB press conference. However, the present work goes one step further by explaining the change in the level of futures rates caused by the ECB's unexpected announcement, rather than looking only at changes in second moments such as absolute returns. In other words, we document the ability of the ECB to move rates in the *desired direction*, instead of simply introducing noise, when its announcement differs from what the market expected.

There is an open question that this paper brings to the fore: if the words of the ECB President can be easily and unambiguously quantified in the way we suggest, then why is this piece of information not presented in a precise numerical form, analogous to the ECB policy rate decision?

Monetary policymakers are interested in permanently moving futures rates using their statements rather than the precise estimates of the timing and impact of news, i.e., its initial reaction. Table 8 addresses this question and provides the futures price change for three weeks, fifteen trading days, following the ECB President's press conference. Unexpected hawkish or dovish announcements are defined as in section 4, specifically as in figure 6.

It is interesting to see that the initial response to ECB statements is part of a larger, long-term reaction. However, the evidence indicates that there is an asymmetric long-term response. On the one hand, the futures price change becomes increasingly negative and increasingly significant during the month after a dovish announcement. This statistical pattern is uncovered despite standard errors increasing with the measurement interval; this holds true even if the multiday tests lack power against the alternatives that the price reacts permanently to the tone of the central bank declaration over

Table 8. Futures Price Response over Longer Horizons

Days	Hawkish	Dovish
1	0.022**	–0.014*
2	0.022*	–0.017
3	0.019*	–0.023*
4	0.021*	–0.017
5	0.016	–0.022
6	0.030	–0.031*
7	0.040	–0.046**
8	0.043*	–0.052***
9	0.045*	–0.061***
10	0.041*	–0.067***
11	0.036	–0.070***
12	0.027	–0.076***
13	0.008	–0.081***
14	0.006	–0.084***
15	–0.013	–0.098***

Note: We compute futures price changes as the difference between the five-minute average futures rate (between 17:00 and 17:05, London time) on trading day t after the press conference and the thirty-minute average futures rate (between 13:25 and 13:30, London time) taken on Governing Council meeting days, which is immediately before the press conference takes place. A statement is defined as hawkish when the news surprise, the difference between $Index_t^{NEW}$ and $E_{t-h}\left[Index_t^{NEW}\right]$, belongs to the top 20th percentile of the news shocks observed in our sample period (i.e., 0.142). A statement is defined as dovish when the news surprise belongs to the bottom 20th percentile of the news shocks observed in our sample period (i.e., –0.228). In the remaining cases, the central bank statement is classified as neutral. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

the following month.[14] On the other hand, futures prices increase, but only in a marginally significant way, by about 3–4 basis points after a hawkish announcement and maintain the new level for about two weeks.

[14]Technically, future prices follow a unit-root process; thus, the variance of futures price changes between date t and $t+m$ is proportional to m.

Overall, we provide strong and economically relevant evidence that central bank communication impacts futures prices permanently, and not only transitorily. Hence, central bankers' announcements seem to be a very powerful tool to systematically drive market expectations and, eventually, the actual evolution of the real economy.

7. Conclusions

Event-study analysis is now an important part of finance—especially corporate finance, where it is used to highlight empirical regularities in the response of stock prices to investment decisions, financing decisions, and changes in corporate control. In this paper, we apply its methods in order to investigate the reaction of asset prices to unexpected central bank qualitative announcements.

Given the unique institutional features of ECB monetary policy conduct, we think that high-frequency intraday is the proper frequency for our event-study analysis. Since the news shock and monetary policy shock hit the financial market at two different points in time, not only can we distinguish one surprise from the other, but we can also investigate their effects independently.

The interpretation of central bankers' statements and actions is of considerable importance to monetary policymakers, financial market participants, and—more generally—the overall public. In this paper, we analyze the relationship between central bank words and deeds and changes in asset prices. More specifically, we examine the effect of European Central Bank communication on the price discovery process in the Euribor futures market using a *new* tick-by-tick data set.

First, we find that the number of transactions and the number of exchanged futures contracts (volume) data confirm that two news items systematically hit financial markets on Governing Council meeting days: the ECB policy rate decision and the explanation of its monetary policy stance.

Second, we show that when the tone of the press conference is different from what the market expects, the futures rate experiences a statistically and economically significant quick (less than an hour) reaction. Put differently, we show that communication is an important tool in the process of conducting monetary policy stance.

Third, our results establish the degree of efficiency of the Euribor futures market.

Finally, by estimating a state-space model, we find that the importance of the ECB press conference has increased over time, especially since the first years of the ECB's life. This evidence suggests that financial market participants needed around three years to learn how to interpret (and also fully believe) the central bank announcements.

The fact that the ECB is able to move asset prices by simply using words seems to indicate that financial markets believe that the European Central Bank does what it says it will do. In other words, even if it is a relatively young central bank, the ECB has already acquired some reputation for telling the truth. Hence, the ECB has already built up some credibility capital. However, credibility is a matter of degree, and this paper does not answer the question of how credible the ECB is.

There are, of course, several other important issues not considered here that require further study.

To interpret an event study, we need to assess quantitatively our ability to detect the presence of an abnormal asset price change. In other words, we also need to evaluate the power of the test, i.e., the probability of rejecting a false null hypothesis (i.e., ECB unexpected announcements have no impact on the behavior of asset prices). In this paper, we make specific assumptions about the distribution of abnormal price changes. Hence we use parametric estimation methods. Alternatively, nonparametric methods (such as either the sign or the rank test), which are free of specific distributional assumptions, are available and can be used.

As a first step, we restricted our sample to Governing Council meeting days. It would be interesting to extend our analysis to include all ECB President speeches. We would thus be able to break down news shocks further into two separate factors: path (change in the near-term path of policy expectations) and time (changes in the expected timing of policy speeches). Moreover, we could also disentangle and separate news about the future path of monetary policy from news about the future economic outlook—i.e., the evolution of macroeconomic or monetary variables, such as output, price indexes, exchange rates, M3 growth, etc.

We test market efficiency in real time. We look at the effects of the ECB Presidents' announcements on Euribor futures rates using a new high-frequency data set. We explain price changes, but we do not statistically investigate the informational content of the number of observations and volumes (number of exchanged contracts) within a specific time interval (Demos and Goodhart 1996).

We apply standard event-study econometric methods, but at the same time we overlook market microstructure issues, such as non-synchronous trading effects (transactions usually take place at time intervals of irregular length and thus transaction data are sampled at irregular random intervals) and price discreteness (prices are always quoted in discrete units). We believe that the three-month futures market institutional structure can be safely ignored for our purpose of assessing the response of asset prices to ECB unexpected announcements. However, it is possible that our results could be biased (cf. Campbell, Lo, and MacKinlay 1997, chap. 3). The computation of further diagnostic tests could be particularly fruitful to gauge the robustness of our preliminary findings.

Appendix

Coding of ECB President Press Conferences

Table 9 reports the assigned value of risk, *Index*, for each ECB monetary policy announcement from January 2000 through June 2006.

Examples of Introductory Statements and Their Coding

It should be noted that the excerpts reported in this appendix only provide some references to the classification but do not completely exhaust the information we use to pin down the ECB future policy inclination (see Rosa and Verga 2007 for further details). Emphasis has been added.

Date: 2 March 2000
The Governing Council also concluded that **the balance of risks to price stability in the medium term remains on the upside. These upside risks will need to be monitored and assessed**

Table 9. ECB President Announcements about Future Monetary Policy Moves

Date	*Index*	*Index2*	Date	*Index*	*Index2*
05/01/2000	1	2	03/04/2003	0	0
03/02/2000	1	2	08/05/2003	−1	−1
02/03/2000	1	2	05/06/2003	−1	−1
13/04/2000	1	2	10/07/2003	−1	−1
11/05/2000	1	2	NA		
08/06/2000	1	2	04/09/2003	0	0
06/07/2000	1	2	02/10/2003	0	0
NA			06/11/2003	0	0
14/09/2000	1	2	04/12/2003	0	0
05/10/2000	1	2	08/01/2004	0	0
02/11/2000	1	2	05/02/2004	0	0
14/12/2000	1	2	04/03/2004	0	0
NA			01/04/2004	0	0
01/02/2001	1	1	06/05/2004	0	0
01/03/2001	1	1	03/06/2004	1	1
11/04/2001	1	1	01/07/2004	1	1
10/05/2001	0	0	NA		
07/06/2001	0	0	02/09/2004	1	1
05/07/2001	0	0	07/10/2004	1	1
NA			04/11/2004	1	1
30/08/2001	−1	−1	02/12/2004	1	1
11/10/2001	−1	−1	13/01/2005	1	1
08/11/2001	0	0	03/02/2005	1	1
06/12/2001	0	0	03/03/2005	1	1
03/01/2002	0	0	07/04/2005	1	1
07/02/2002	0	0	04/05/2005	1	1
07/03/2002	0	0	02/06/2005	1	1
04/04/2002	0	0	07/07/2005	1	1
02/05/2002	0	0	NA		
06/06/2002	1	1	01/09/2005	1	1
04/07/2002	0	0	06/10/2005	1	2
NA			03/11/2005	1	2
12/09/2002	−1	−2	01/12/2005	1	1
10/10/2002	−1	−1	12/01/2006	1	2
07/11/2002	−1	−2	02/02/2006	1	2
05/12/2002	−1	−2	02/03/2006	1	2
09/01/2003	−1	−1	06/04/2006	0	1
06/02/2003	−1	−2	04/05/2006	1	2
06/03/2003	−1	−1	08/06/2006	1	2

Note: January 2000–June 2006. We report the monetary policy intentions indicators communicated by the ECB on Governing Council meeting days. *Index* ranges from −1 to +1, while *Index2* ranges from −2 to +2. Note that we have considered only the first press conference of each month. NA indicates that the press conference did not take place.

continuously in order to ensure that timely action can be taken, if and when required. (...)

The Governing Council concluded that **vigilance is required** and pointed to several factors.
Coding: +1

Date: 14 September 2000
The annual rate of increase in the Harmonised Index of Consumer Prices (HICP) was 2.4% in July 2000. Recent consumer price developments in the euro area have been very much influenced by the strong rise in oil prices and the depreciation of the exchange rate of the euro.

While monetary policy cannot address short-term developments in prices, **it is imperative for monetary policy that medium-term upward pressure on prices be contained**. The risk that the current pressure on the HICP might spill over onto costs and prices determined in the domestic economy must be taken seriously. This holds true in particular in the context of the favourable prospects for economic growth.
Coding: +1

Date: 7 March 2002
As usual, at today's meeting we examined recent monetary, financial and economic developments. The Governing Council concluded that the information which had become available in recent weeks confirmed that the current level of key ECB interest rates remains **appropriate** for the maintenance of price stability over the medium term. Against this background, the Governing Council decided to leave the key ECB interest rates unchanged.
Coding: 0

Date: 4 March 2004
Overall, the Governing Council confirmed its previous assessment of a **favourable outlook for price stability** in the euro area over the medium term. Against this background, we concluded that the **current stance of monetary policy remains appropriate**. The key ECB interest rates have therefore been left unchanged at their low levels. Our monetary policy stance provides support to the economic recovery in the euro area. (...)

To sum up, the economic analysis continues to indicate that the main scenario for price developments in the coming years **is in line with price stability**. Cross-checking with the monetary analysis does not alter this picture for the time being.
Coding: 0

Date: 7 November 2002
We have reviewed monetary, financial and economic developments and updated our assessment in the light of the information available. In view of the high uncertainty on future growth, and its implication for medium-term inflationary developments, the Governing Council **has discussed extensively the arguments for and against a cut in the key ECB interest rates**. The view has prevailed to keep interest rates unchanged. However, the Governing Council **will monitor closely the downside risks** to economic growth in the euro area.
Coding: −1

Date: 12 September 2002
Our conclusion is that risks to price stability appear rather **balanced**. Against this background, the current level of key ECB interest rates is **appropriate**.

(...) Nevertheless, **risks to the economic outlook**, both inside and outside the euro area, **need to be monitored closely**.
Coding: −1

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2003. "Modelling and Forecasting Realized Volatility. *Econometrica* 71 (2): 579–625.

Andersson, M. 2007. "Using Intraday Data to Gauge Financial Market Responses to Fed and ECB Monetary Policy Decisions." ECB Working Paper No. 726.

Balduzzi, P., E. J. Elton, and T. C. Green. 2001. "Economic News and Bond Prices: Evidence from the U.S. Treasury Market." *Journal of Financial and Quantitative Analysis* 36 (4): 523–43.

Bernoth, K., and J. von Hagen. 2004. "The Euribor Futures Market: Efficiency and the Impact of ECB Policy Announcements." *International Finance* 7 (1): 1–24.

Blinder, A. S. 1998. *Central Banking in Theory and Practice*. Cambridge, MA: MIT Press.

Blume, M. E., and R. F. Stambaugh. 1983. "Biases in Computed Returns: An Application to the Size Effect." *Journal of Financial Economics* 12 (3): 387–404.

Bomfim, A. N. 2003. "Monetary Policy and the Yield Curve." Finance and Economics Discussion Series No. 2003-15, Board of Governors of the Federal Reserve System.

Brand, C., D. Buncic, and J. Turunen. 2006. "The Impact of ECB Monetary Policy Decisions and Communication on the Yield Curve." ECB Working Paper No. 657.

Campbell, J. Y., A. W. Lo, and A. C. Mackinlay. 1997. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Demos, A. A., and C. A. E. Goodhart. 1996. "The Interaction between the Frequency of Market Quotations, Spread and Volatility in the Foreign Exchange Market." *Applied Economics* 28 (3): 377–86.

Efron, B., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall.

Ehrmann, M., and M. Fratzscher. 2007a. "Communication and Decision-Making by Central Bank Committee Members: Different Strategies, Same Effectiveness?" *Journal of Money, Credit, and Banking* 39 (2–3): 509–41.

———. 2007b. "Explaining Monetary Policy in Press Conferences." ECB Working Paper No. 767.

———. 2007c. "The Timing of Central Bank Communication." *European Journal of Political Economy* 23 (1): 124–45.

Faust, J., E. T. Swanson, and J. H. Wright. 2004. "Identifying VARs Based on High Frequency Futures Data." *Journal of Monetary Economics* 51 (6): 1107–31.

Gerlach, S. 2007. "Interest Rate Setting by the ECB 1999–2006: Words and Deeds." *International Journal of Central Banking* 3 (3): 1–45.

Gürkaynak, R. S., B. Sack, and E. T. 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.

Guthrie, G., and J. Wright. 2000. "Open Mouth Operations." *Journal of Monetary Economics* 46: 489–516.

Jansen, D.-J., and J. de Haan. 2005. "Talking Heads: The Effects of ECB Statements on the Euro-Dollar Exchange Rate." *Journal of International Money and Finance* 24 (2): 343–61.

———. 2006. "Look Who's Talking: ECB Communication during the First Years of EMU." *International Journal of Finance and Economics* 11 (3): 219–28.

———. 2007a. "The Importance of Being Vigilant: Has ECB Communication Influenced Euro Area Inflation Expectations?" DNB Working Paper No. 148.

———. 2007b. "Were Verbal Efforts to Support the Euro Effective? A High-Frequency Analysis of ECB Statements." *European Journal of Political Economy* 23 (1): 245–59.

Johnston, J., and J. DiNardo. 1996. *Econometric Methods*. McGraw-Hill/Irwin.

Kohn, D. L., and B. P. Sack. 2003. "Central Bank Talk: Does It Matter and Why?" Finance and Economics Discussion Series No. 2003-55, Board of Governors of the Federal Reserve System.

Krippendorf, K. 2004. *Content Analysis: An Introduction to Its Methodology*. 2nd ed. Thousand Oaks, CA: Sage.

Kuttner, K. 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Markets." *Journal of Monetary Economics* 47 (3): 523–44.

MacKinlay, A. C. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature* 35 (1): 13–39.

Musard-Gies, M. 2006. Do European Central Bank's Statements Steer Interest Rates in the Euro Zone? *The Manchester School* 74 (s1): 116–39.

Oxley, L., and M. McAleer. 1993. "Econometric Issues in Macroeconomic Models with Generated Regressors." *Journal of Economic Surveys* 7 (1): 1–40.

Piccinato, B., G. Ballocchi, M. M. Dacorogna, and R. Gencay. 1999. "Intraday Statistical Properties of Eurofutures." *Derivatives Quarterly* 6 (2): 28–44.

Romer, C. D., and D. H. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." In *NBER Macroeconomics Annual 1989*, ed. O. J. Blanchard and S. Fischer, 121–70.

Rosa, C. 2007a. "Providing Content to ECB Announcements." Forthcoming in *Rivista Internazionale di Scienze Sociali.*

———. 2007b. "Talking Less and Moving the Market More: Is This the Recipe for Monetary Policy Effectiveness? Evidence from the ECB and the Fed." CEP Discussion Paper No. 855.

———. 2007c. "What Is the ECB Reaction Function? A Dynamic Probit Analysis." Unpublished Paper.

Rosa, C., and G. Verga. 2006. "The Impact of Central Bank Announcements on Asset Prices in Real Time: Testing the Efficiency of the Euribor Futures Market." CEP Discussion Paper No. 764.

———. 2007. "On the Consistency and Effectiveness of Central Bank Communication: Evidence from the ECB." *European Journal of Political Economy* 23 (1): 146–75.

Ruud, P. A. 2000. *An Introduction to Classical Econometric Theory.* Oxford University Press.

The Economist. 2006. "Feeling Fitter." Finance and Economics Section, 7 October, vol. 381, issue 8498.

Trichet, J.-C. 2005. "Communication, Transparency and the ECB's Monetary Policy." Keynote speech at the New Year's reception of the International Club of Frankfurt Economic Journalists, Frankfurt am Main, Germany, 24 January. Available at http://www.ecb.int/press/key/date/2005/html/sp050125.en.html.

Weber, R. P. 2004. *Basic Content Analysis.* 2nd ed. Newbury Park, CA: Sage.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.

Woodford, M. 2005. "Central-Bank Communication and Policy Effectiveness." NBER Working Paper No. 11898.

Technical Appendix to "The Impact of Central Bank Announcements on Asset Prices in Real Time"

Generated-Regressor Issue

Direct Method

We estimate the following regression by OLS:

$$f_{t+h} - f_{t-h} = c_1 + c_2 \cdot \big(Index_t^{NEW} - c_3 \cdot \big(f_{t-h} - R_t^{NEW}\big) - c_4 \cdot Index_t^{OLD}\big) + \varepsilon_t, \quad \text{(A1)}$$

where cs are regressor coefficients, and the rest of the notation is the same as in the paper.

Table A1. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t+h=$ 14:45 $t-h=$ 13:25	$t+h=$ 15:45 $t-h=$ 13:25	$t+h=$ 16:45 $t-h=$ 13:25
c_1	0.006	0.008*	0.008*
	(0.004)	(0.004)	(0.005)
c_2	0.029**	0.029**	0.035***
	(0.011)	(0.013)	(0.013)
c_3	1.169**	1.060*	1.019**
	(0.537)	(0.610)	(0.494)
c_4	0.778***	0.805***	0.708***
	(0.145)	(0.187)	(0.172)
R^2	0.224	0.161	0.178
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Empirical Confidence Bands

In order to account for the generated-regressor problem when computing coefficient estimates' standard errors, we also check the robustness of our conclusions by using a bootstrap approach to statistical inference (see, e.g., Efron and Tibshirani 1993). More specifically, we apply a sampling-with-replacement raw residuals bootstrap scheme with 1,000 repetitions. The empirical results are qualitatively very similar to those reported in the paper (tables 2, 3, and 4).

Table A2. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t+h=$ **14:45** $t-h=$ **13:25**	$t+h=$ **15:45** $t-h=$ **13:25**	$t+h=$ **16:45** $t-h=$ **13:25**
Constant	0.003	0.005	0.006
NS_t	0.031***	0.033***	0.038***
R^2	0.152	0.130	0.156
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. ***, **, and * indicate empirical significance at the 99, 95, and 90 percent level, respectively. The bootstrap scheme is based on 1,000 repetitions.

Econometric Analysis Using Andersen et al. (2003) Equally Spaced Returns

As a further robustness check, we also rerun the previous regressions (estimations reported below) using equally spaced data instead of averaged tick-by-tick data. We construct these artificial data by linear interpolation of the transaction prices immediately before and after the relevant point in time. Then we obtain futures rate returns as the first difference of the new prices (see Andersen et al. 2003, 593). Note that, overall, the empirical findings discussed in this section are qualitatively very similar if we use equally spaced data.

Table A3. Auxiliary Regression to Measure the Expected ECB Announcement Using Ordered Probit

$f_{t-h} - R_t^{NEW}$	1.864**
	(0.742)
$Index_t^{OLD}$	1.967***
	(0.354)
δ_1	−1.552***
	(0.348)
δ_2	1.180***
	(0.333)
Log-Likelihood	−30.052
Pseudo-R^2	0.574
Observations	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ML – Ordered Probit (Quadratic hill climbing). ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Table A4. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t + h =$ 14:45 $t - h =$ 13:30	$t + h =$ 15:45 $t - h =$ 13:30	$t + h =$ 16:45 $t - h =$ 13:30
Constant	0.003	0.006	0.005
	(0.003)	(0.004)	(0.004)
NS_t	0.033***	0.032**	0.040***
	(0.011)	(0.013)	(0.014)
R^2	0.170	0.125	0.169
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Table A5. Time Needed to Incorporate the News Shock (Futures Rates), Dependent Variable $f_{17} - f_{t^*}$

	$t^* = 14{:}15$	$t^* = 14{:}30$	$t^* = 14{:}45$	$t^* = 15{:}00$
Constant	0.004*	0.003	0.004*	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
NS_t	0.018*	0.016**	0.007	0.007
	(0.009)	(0.007)	(0.005)	(0.005)
R^2	0.097	0.105	0.023	0.026
Observations	70	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

The Effects of Monetary Policy Shocks

In this section we analyze the effects of monetary policy shocks to futures rates. In particular:

- We explain the change in futures rates around the new policy rate release by the monetary policy shock.
- We show econometrically that monetary policy shocks are completely incorporated in futures rates in less than five minutes.
- Potentially some of the movements that are taking place in the market during the press conference may still be the consequence of the policy rate announcement. If this is the case, our methodology does suffer from an omitted-variables problem. In other words, because the news surprise might be correlated with the interest rate announcement, then our estimates might contain some residuals effect from the interest rate announcement. This will bias the estimated coefficients of the news shock upward. Given the above finding, we can assess how words affect yields, safely disregarding the surprise in the target release. However, to take explicitly into account this issue, as a further robustness check, we reestimate the main regression presented in the paper (see table 3 in the article), controlling also for the monetary policy shock.

Table A6. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t + h = 12{:}45$ $t - h = 12{:}40$	$t + h = 13{:}00$ $t - h = 12{:}40$
Constant	0.009***	0.006***
	(0.002)	(0.002)
MPS_t	0.069***	0.043***
	(0.017)	(0.016)
R^2	0.312	0.124
Observations	99	99

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively. Note that we have not included two observations (8 June 2000 and 10 May 2001) because they induce two outlier residuals. However, the results continue to hold even if we include both observations.

Table A7. Time Needed to Incorporate the News Shock (Futures Rates), Dependent Variable $f_{13:25} - f_{t^*}$

	$t^* = 12{:}40$	$t^* = 12{:}45$	$t^* = 12{:}55$
Constant	0.006**	−0.003	−0.001
	(0.002)	(0.002)	(0.002)
MPS_t	0.039**	−0.030	−0.007
	(0.019)	(0.024)	(0.021)
R^2	0.100	0.062	0.007
Observations	99	99	99

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Table A8. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t+h=$ **14:45** $t-h=$ **13:25**	$t+h=$ **15:45** $t-h=$ **13:25**	$t+h=$ **16:45** $t-h=$ **13:25**
Constant	0.007	0.010	0.010
	(0.007)	(0.009)	(0.008)
NS_t	0.030**	0.032**	0.037***
	(0.012)	(0.014)	(0.014)
MPS_t	0.040	0.044	0.039
	(0.051)	(0.071)	(0.065)
R^2	0.163	0.140	0.163
Observations	70	70	70

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–June 2006. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively. There are seventy observations because for the period January 2000–October 2001 we have only considered the first Governing Council meeting of every month.

Econometric Analysis Using Musard-Gies' (2006) Wording Indicator, IndexMG

As a robustness check, in this section we reestimate the main regressions reported in the article using the wording indicator variable, *IndexMG*, proposed by Musard-Gies (2006).

IndexMG takes on a four-value scale from −1 to +2. The value of zero suggests that the current level of the policy rate is appropriate to maintain price stability over the medium term. A negative value characterizes an easing period: it is possible that the policy rate will be cut in the near future. On the other hand, a positive value characterizes a potential future monetary policy tightening.

IndexMG spans a shorter sample, from January 1999 through October 2004. In the econometric analysis reported below, we consider the set of Governing Council meetings where also our *Index* is available (cf. table 9).

Table A9. Auxiliary Regression to Measure the Expected ECB Announcement Using Ordered Probit

$f_{t-h} - R_t^{NEW}$	4.693***
	(1.531)
$IndexMG_t^{OLD}$	3.068***
	(0.711)
δ_1	−2.370***
	(0.687)
δ_2	1.787***
	(0.625)
δ_3	7.068***
	(1.862)
Log-Likelihood	−19.366
Pseudo-R^2	0.717
Observations	51

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–October 2004. The econometric method is ML – Ordered Probit (Quadratic hill climbing). ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

Table A10. Explanation of Innovation in Expectations (Futures Rates), Dependent Variable $f_{t+h} - f_{t-h}$

	$t + h = 14{:}45$ $t - h = 13{:}30$	$t + h = 15{:}45$ $t - h = 13{:}30$	$t + h = 16{:}45$ $t - h = 13{:}30$
Constant	0.008**	0.010**	0.011**
	(0.004)	(0.004)	(0.004)
NS_t	0.025**	0.028**	0.029**
	(0.011)	(0.012)	(0.014)
R^2	0.115	0.097	0.095
Observations	51	51	51

Note: Monthly observations on days of ECB Governing Council meetings, January 2000–October 2004. The econometric method is ordinary least squares. Heteroskedasticity-consistent standard errors are in parentheses. ***, **, and * denote significance at the 10, 5, and 1 percent level, respectively.

The Danger of Inflating Expectations of Macroeconomic Stability: Heuristic Switching in an Overlapping-Generations Monetary Model*

Alex Brazier, Richard Harrison, Mervyn King, and Tony Yates
Bank of England

We use a monetary overlapping-generations model to discuss the cause and durability of the marked fall in the volatility of inflation in recent decades. In our model, agents have to forecast inflation, and they do so using two "heuristics." One is based on lagged inflation, the other on an inflation target announced by the central bank. Agents switch between those heuristics based on an imperfect assessment of how each has performed in the past. The way the economy propagates productivity shocks into inflation depends on the proportion of agents using each heuristic. Movements in these proportions generate fluctuations in small-sample measures of economic volatility. We use this simple model of heuristic switching to contrast the performance of monetary policy rules. We find that, relative to the rule that would be optimal under rational expectations, a rule that responds to both productivity shocks and inflation expectations better stabilizes the economy but does not prevent agents from switching between heuristics. Finally, we study the impact of introducing an explicit inflation target, which can be used by agents as a simple heuristic, into an economy that did not previously have one. Depending on the heuristics agents have access to before the introduction of the target, this can result in reduced inflation volatility.

JEL Codes: E32, E37, E52.

1. Introduction

The United Kingdom has experienced a period in which the volatility of both real and nominal variables has fallen. From 1995 to 2005, the standard deviation of output growth was less than one-third of its value from 1975 to 1985; the standard deviation of inflation was less than one-tenth of its value from 1975 to 1985. Inflation persistence has also fallen dramatically. Similar developments are apparent in other advanced economies such as the United States and the euro area. These changes have found various names: the "Great Stability," the "Great Moderation," or the "NICE" (non-inflationary consistently expansionary) decade.[1] Policymakers face a challenge in judging how to react to these changes because their causes, and therefore their durability, are uncertain, as Velde's (2004) lucid survey of the research so far makes clear.

There are two types of explanation for these changes. The first is that the reduction in volatility is due to better monetary and fiscal policy. The second is that it reflects either smaller shocks or changes in the way those shocks are propagated into output and inflation volatility. Thus far, econometric studies have tended to attribute most of the improvement to what Velde described as policymakers having a "good hand" rather than engaging in "good play": witness the line of work including Stock and Watson (2002), Sims and Zha (2004), Cogley and Sargent (2005), and many others. But Bernanke (2004) suggested that what is counted as good luck in such studies includes the effect of better monetary policy in anchoring inflation expectations.

Our paper presents a model in which the link between fluctuations in the time-series properties of inflation and expectations formation is explicit. We work with a monetary overlapping-generations model, in which we assume agents form expectations by choosing amongst simple rules of thumb, or "heuristics." Agents work when

necessarily reflect those of the Monetary Policy Committee. We are grateful for insightful discussions with Charles Bean, James Proudman, Tim Taylor, Gertjan Vlieghe, and Fabrizio Zampolli. We also thank Bill Branch, Buz Brock, Cars Hommes, Hyun Shin, and Frank Smets for comments. All errors and omissions are our own responsibility. This paper was prepared for an ECB colloquium celebrating the career of Otmar Issing, Frankfurt, March 16–17, 2006.

[1] See, for example, Bernanke (2004) and King (2003).

they are young and sell their output to the old in exchange for money, which is the only store of value available to them. They consume that money when they are old. Young agents seek to minimize the disutility from working when young and maximize the utility they will gain from consuming when old. In doing this, they face the problem of forecasting the future purchasing power of the money balances they accumulate when young: they need to forecast the change in the price level. Uncertainty about future inflation is generated by our assumption that the productivity of young agents is subject to shocks.

We contrast the rational-expectations equilibrium with that which emerges when agents use a finite set of heuristics to make their forecasts of inflation. They choose between the heuristics on the basis of their performance in forecasting inflation in the recent past. We assume they observe that performance with some noise, but the better the true past performance of a heuristic, the greater chance there is that an agent uses it to make the next period's forecast. These heuristics—as Gigerenzer, Todd, and the ABC Research Group (1999) and others have noted—are both fast to compute and frugal in their information requirements. Advocates of the heuristics approach argue that model-consistent expectations are attractive devices for those who work with model economies, but it may not be rational for agents to have acquired them, given the informational and computational costs of doing so. In our model, agents choose between two heuristics: one that sets forecast inflation equal to the steady-state value, which we term loosely an "inflation-target" heuristic; and one in which forecast inflation is set to the latest realization of inflation, which we term the "lagged-inflation" heuristic.

Our model is closed by a process for nominal money growth, which characterizes central bank behavior. We use two such processes to study the dynamics of inflation: one in which the central bank follows the rule that would be optimal in the event that expectations were rational, and another that assumes the central bank attempts to take account of heuristic behavior.

Our strategy is to use a model of heuristics to explain the Great Stability. We are therefore exploring an idea put forward by Branch and Evans (2007). And in combining a monetary overlapping-generations model with heuristics, we are borrowing from Brock and de Fontnouvelle (2000), who did this in their quest to see whether heuristic behavior could sustain equilibria in which paper money is valued.

When agents switch between inflation-forecasting heuristics, the time-series properties of inflation change over time. On average, the majority of agents use the inflation-target heuristic. But there are times when everyone does, and times when no one does. The way the economy propagates productivity shocks into inflation depends on the proportion of agents using each heuristic. Because this proportion fluctuates, so does the way shocks are propagated into inflation. And the changes in heuristic use generates higher inflation volatility than in a rational-expectations version of the model. Moreover, there are greater fluctuations in the volatility of inflation and in the persistence of inflation. This model, for either of the money processes we use, exhibits pronounced episodes of high inflation volatility, followed by low inflation volatility and persistence. When agents use the inflation-target heuristic, inflation tends to be less variable and less persistent than when more agents use the lagged-inflation heuristic.

We contrast the money-growth process that would be optimal under rational expectations with one that attempts to take account of heuristics. We do so with the usual caveats that must accompany welfare analysis in overlapping-generations models. Our welfare criterion is the unconditional expectation of the sum of the welfare of the old and young in any time period. This is equivalent to maximizing the average level of welfare over all generations.

Under rational expectations, the optimal policy is for money growth to respond to the level of productivity. Such a rule eliminates both the volatility of labor supply, which is costly to the young, and the volatility of consumption, which is costly to the old. The success of monetary policy under rational expectations can be attributed to its leverage over expectations. By committing to future policy actions, monetary policy has extra leverage over current labor supply and inflation.

That leverage is not available when agents use heuristics, so we investigate how policy might adapt in those circumstances. The model under heuristics is highly nonlinear. There is no analytical expression for optimal policy available, so we confine ourselves to a search for a rule that responds linearly to two important state variables in the model: productivity and expected inflation. The best rule—according to our welfare criterion—in this class of rule increases money growth when productivity is high, and by more than under rational expectations, and it reduces money growth when inflation expectations rise. The welfare benefits from shifting away from the rational-expectations policy are greater during periods when agents are using the backward-looking heuristic. Despite a monetary policy that attempts to take account of heuristics, heuristic switching still occurs and so there are still fluctuations in inflation volatility and inflation persistence. At the same time, this model generates fluctuations in the estimated disturbances to linear autoregressive equations for inflation, echoing the findings of econometricians using macroeconomic time series.

The message from the paper to this point is that very stable macroeconomic outturns should not be taken for granted. But we go on to explore the notion that the widespread adoption of explicit inflation objectives by central banks can be modeled as the provision of a heuristic to which agents did not previously have access. When we introduce an inflation-target heuristic to agents, we find that at least some adopt it immediately and that subsequently the volatility of inflation is lower, despite the heuristic switching that ensues. We illustrate how the impact of the introduction of the inflation target depends on the performance of the heuristic with which agents start out.

2. The Model

Our model is an overlapping-generations model with money. It is deliberately stylized and was chosen as the simplest possible model in which agents must forecast future inflation.

Agents live for two periods. They work when young and consume when old. Young agents minimize the disutility from work (L) when young and maximize the expected utility from consumption (C) when old.[2] Their output is produced with a linear technology, denoted $A_t L_t$, where A_t is productivity, known at time t when young agents determine their labor supply. Their output is sold at price P_t. Young agents accumulate nominal money balances (M_t) equal to the

[2]Note that, for simplicity, we assume that there is no discounting of future consumption.

value of their output. Their consumption when old is determined by the real value of those same money balances $\left(\frac{M_t}{P_{t+1}}\right)$. We denote expectations formed by agents using the operator E_t. In some cases that will refer to rational expectations and in others it will refer to a heuristic. At each stage we will make clear how agents are forming their expectations.

Formally, young agents solve the following problem:

$$\max_{L_t} E_t \left[-\frac{L_t^{1+\eta}}{1+\eta} + \frac{C_{t+1}^{1-\alpha}}{1-\alpha} \right] \eta > 0,\ 0 < \alpha < 1 \tag{1}$$

subject to

$$M_t = A_t L_t P_t \tag{2}$$

$$C_{t+1} = \frac{M_t}{P_{t+1}}. \tag{3}$$

The problem that old agents solve is degenerate. They maximize utility by spending all their real balances on consumption goods. The young accumulate money from the old and from the government. The government's budget constraint implies that the nominal money stock evolves according to

$$M_t = M_{t-1} + P_t D_t, \tag{4}$$

where $D_t > 0$ is output purchased from the private sector in exchange for money. We assume that government purchases are used for purposes that do not yield private utility.[3] The instrument of monetary policy is the growth rate of the nominal money stock, G:

$$M_t = M_{t-1} + G_t M_{t-1} = (1 + G_t) M_{t-1}$$

so that, since $P_t D_t = G_t M_{t-1}$, the nominal value of government purchases equals the increase in the nominal money supply: there is no distinction between fiscal and monetary policy in this model.

[3]We could, analogously, assume that government purchases are redistributed back to agents and that these redistributions enter utility in a way that was additively separable from other components. Our marginal condition for labor supply would be identical in this model, although consumption and mean levels of welfare would not be. Dropping the simplification used here would not affect the impact of heuristic switching on the dynamics of macroeconomic outcomes.

The young consumer's problem can now be written as

$$\max_{L_t} E_t \left[-\frac{L_t^{1+\eta}}{1+\eta} + \frac{1}{1-\alpha} \left(A_t L_t \frac{P_t}{P_{t+1}} \right)^{1-\alpha} \right].$$

Denoting inflation as $\Pi_{t+1} = \frac{P_{t+1}}{P_t}$, the first-order condition for labor supply is given by

$$L_t^{\eta+\alpha} = E_t \big(A_t \Pi_{t+1}^{-1} \big)^{1-\alpha}.$$

This equation makes it clear that young agents have to make forecasts. If expected inflation tomorrow is high, agents expect the value of any money balances they accumulate by working when young to be eroded when they are old. Their demand for money balances will be lower.

Uncertainty about the future price level is introduced by a simple, stochastic process for productivity (A_t):

$$A_t = A_{t-1}^{\rho} Z_t, \tag{5}$$

where $\ln Z_t$ is normally distributed.

For ease of exposition, we proceed by taking a first-order approximation around the nonstochastic steady state. Using lowercase letters to denote log-deviations from the steady state, the (log-linearized) first-order condition for labor supply is

$$l_t = \frac{1-\alpha}{\eta+\alpha} a_t - \frac{1-\alpha}{\eta+\alpha} E_t \pi_{t+1}. \tag{6}$$

We use m_t to denote the log-deviation of *real* money balances, $\frac{M_t}{P_t}$, from the steady state. The *real* money demand condition is

$$m_t = a_t + l_t$$
$$m_t = \frac{1+\eta}{\eta+\alpha} a_t - \frac{1-\alpha}{\eta+\alpha} E_t \pi_{t+1}. \tag{7}$$

The linearized version of the government budget constraint is given by equation (8) below, where we denote the steady-state inflation rate as Π and use g_t to denote the absolute (note, not log) deviation of the growth rate of nominal money from its steady-state level:[4]

$$m_t = \Pi^{-1} g_t + m_{t-1} - \pi_t. \tag{8}$$

We linearize around a positive steady-state inflation rate ($\Pi > 1$) to ensure that the frequency of negative government spending levels D implied by money growth g is negligible: we do not regard such outcomes as economically meaningful.

Linearizing the productivity process gives

$$a_t = \rho a_{t-1} + \zeta_t, \zeta_t \sim N(0, \sigma^2), \tag{9}$$

where ζ_t is the log-deviation of the disturbance Z_t from its steady-state value, 1.

To summarize the model: to maximize their expected utility, young agents must forecast inflation. Uncertainty about future inflation is introduced by fluctuations in the demand for real money balances arising from shocks to productivity. If those movements are not matched by equal movements in the nominal money stock, inflation will fluctuate. In the next section we calculate the monetary policy that maximizes welfare when agents form rational expectations of inflation.

3. Rational Expectations and Optimal Policy

The model is described by equations (6), (7), (8), and (9) together with an equation for money growth, g_t. We assume that monetary policy is characterized by the design of a rule for money growth to which the policymaker commits. The rule is designed to maximize a particular measure of welfare. It is designed before any realization of productivity is observed, so although money growth can respond to realizations of productivity, the policy rule itself is invariant to changes in productivity.

Our welfare measure is the sum of the utility of the young and old agents:

$$W_t \equiv -\frac{L_t^{1+\eta}}{1+\eta} + \frac{C_t^{1-\alpha}}{1-\alpha}.$$

[4]The coefficient on g results from the fact that $(1+g) = \Pi$ in the steady state.

This differs from the utility function of a young agent (equation (1)) because it adds the utility of today's old to the disutility of work experienced by today's young. We assume that policy is designed to maximize the unconditional expectation of welfare. This maximizes the average level of welfare across all generations and across all possible realizations of productivity.[5]

We assume that monetary policy maximizes welfare, taking the steady-state level of money growth as given. In this model, there would be welfare improvements from lowering the mean level of money growth and the associated government purchases (which do not yield private utility). We abstract from that component of policy to focus on the stabilization role of monetary policy. Hence, the curvature of the welfare function means that, by stabilizing the economy, we maximize the average level of welfare. Note that, conditional on a level of productivity that is known and different from the steady-state level of productivity, agents will not prefer steady-state levels of labor supply and future consumption. But, before the value of productivity is revealed, they will prefer stable over variable labor supply and consumption because of the curvature in utility. Welfare is maximized when labor supply and consumption do not deviate from their steady-state levels.

Our welfare function is

$$E[W_t - W] = -\frac{\eta}{2}E\left[l_t^2\right] - \frac{\alpha}{2}E\left[c_t^2\right], \tag{10}$$

which we derive as the second-order Taylor approximation to the welfare measure. Policy maximizes the unconditional expectation of a weighted sum of the variances of young agents' labor supply and old agents' consumption. Note that the linear terms that are anticipated in a second-order Taylor expansion drop out: the unconditional expectations of linear terms in log-deviations from the steady state are zero.

Under rational expectations, we now demonstrate that monetary policy can stabilize labor supply and consumption completely

[5] Our procedure is similar to the practice of maximizing "period utility" in the monetary policy design literature that uses representative agent models.

by committing to a rule for money growth that feeds back from the model's driving variable, productivity:

$$g_t = \chi a_t. \tag{11}$$

It is straightforward to show that, for an arbitrary value of χ, the rational-expectations solutions for real money balances and inflation are given by

$$m_t = a_t + l_t = \frac{1+\eta-(1-\alpha)\rho\frac{\chi}{\Pi}}{1+\eta-(1-\alpha)\rho} a_t \tag{12}$$

and

$$\begin{aligned} \pi_t = & \left[\frac{\chi}{\Pi}\rho + \frac{1+\eta-(1-\alpha)\rho\frac{\chi}{\Pi}}{1+\eta-(1-\alpha)\rho}(1-\rho)\right] a_{t-1} \\ & + \left[\frac{\chi}{\Pi} - \frac{1+\eta-(1-\alpha)\rho\frac{\chi}{\Pi}}{1+\eta-(1-\alpha)\rho}\right] \zeta_t. \end{aligned} \tag{13}$$

Policy can completely stabilize employment when $m_t = a_t$. From equation (12), this is the case when $\chi = \Pi$. Under this rule there are no welfare costs to young agents from macroeconomic volatility. But what happens to the volatility of inflation and (hence) the utility of old agents? We know that the consumption of the old generation is determined by their accumulated money balances adjusted for subsequent inflation:

$$c_t = m_{t-1} - \pi_t$$

and, when $\chi = \Pi$, the equilibrium inflation equation (13) can be simplified to

$$\pi_t = a_{t-1}.$$

We already know that real money balances equal productivity because labor supply is stabilized:

$$m_t = a_t \Rightarrow m_{t-1} = a_{t-1}$$

so that

$$c_t = m_{t-1} - \pi_t = 0.$$

A policy rule in the form of equation (11), setting $\chi = \Pi$, eliminates all of the welfare costs of macroeconomic instability. Such a rule generates movements in inflation in the next period that are equal to the realization of productivity in the current period. This strategy means that the real value when old of any money balances accumulated when young is unaffected by realizations of productivity. Anticipating this, the young have no incentive to change their labor supply in response to changes in productivity. With labor supply constant and the impact of productivity on real money balances offset by inflation, the consumption of the old is constant. The key to the success of monetary policy in stabilizing both labor supply and consumption is its leverage over not only the current money stock but also overanticipated future inflation. Indeed, it is clear from (6) that monetary policy can stabilize labor supply in the face of productivity disturbances *only* through its leverage over inflation expectations.

To reemphasize, note that complete stabilization of consumption and employment is optimal because of the curvature of agents' utility (a feature preserved by our quadratic approximation). Note too that monetary policy does not prevent agents from responding to productivity shocks; it simply creates conditions that mean it is optimal for agents not to.

4. Modeling the Choice of Heuristic

So far we have assumed model-consistent expectations to provide a benchmark against which to compare subsequent departures from that assumption. Many have argued that in reality agents would find it too costly or would not have the means to collect the information and carry out the computations required for a rational-expectations equilibrium to be achieved. The route we choose is to adopt a model in which agents may have heterogeneous expectations and in which those expectations are based on simple heuristics.

4.1 The Heuristic Choice Literature

The literature on heuristics is itself now very large and ably surveyed by one of its recent leaders in Hommes (2005). He charts the history of this strand of thought from the suggestion by Keynes (1936) that fluctuations in sentiment would influence the macroeconomy, through Simon (1957), who explained that agents were "boundedly rational" in the face of costs of collecting information and computing the outcomes of their decisions. Another landmark is the emergence of experimental evidence that agents use simple heuristics to make decisions, culminating in Kahneman's (2003) Nobel lecture. This led to a large research program exploring why it may have proven beneficial for nature to endow us with such heuristics—a topic that occupies, for example, Gigerenzer, Todd, and the ABC Research Group (1999). We use a model in which agents choose between a finite set of heuristics based on noisy observations of past forecast performance. The papers from which we draw most inspiration in this respect are Brock and Hommes (1997), Brock and de Fontnouvelle (2000), and Branch and Evans (2006, 2007), who in turn ground their decision-making model in the discrete-decision, multinomial logit models set out in Manski and McFadden (1981).[6]

We are not the first to combine a monetary overlapping-generations model with a model of heuristic expectations formation. Brock and de Fontnouvelle (2000) do just this. But their concern is very different. Early students of rational-expectations, monetary overlapping-generations models noted that these models generated equilibria in which money had value and equilibria in which it did not. This was a source of discomfort since paper money in reality is pervasive, and yet there was no guide as to which of the model's equilibria should or would be selected. Brock and de Fontnouvelle (2000) is an effort to see whether heuristic behavior can lead to monetary equilibria: they find that it can.

4.2 Heuristic Choice in Our Model

Our agents select from two heuristics described by

$$E_{1,t}\pi_{t+1} = \pi_{t-1}$$
$$E_{2,t}\pi_{t+1} = 0.$$

[6]See also de Grauwe and Grimaldi (2006). They show how exchange rate dynamics and fluctuations in the performance of fundamentals models of the exchange rate are affected by heuristic switching, embedding the Brock and Hommes approach, using the same model of predictor choice that we employ.

The first predictor ($E_{1,t}\pi_{t+1}$) sets expected inflation equal to the latest observed outturn. We term this the "lagged-inflation" predictor. This predictor is based on lagged inflation (π_{t-1}) and not current inflation (π_t), which will itself depend on agents' expectations and will not be realized at the time agents are forming their expectations. The second predictor ($E_{2,t}\pi_{t+1}$) sets expected inflation equal to the target (since π represents the deviation of inflation from target, we have $E_{2,t}\pi_{t+1} = 0$). This we term the "inflation-target" predictor.[7] This particular set of predictors includes plausible models for agents to use to forecast, but is itself arbitrary. For most of our analysis, exactly what is in this set of predictors is not important. What is important is that there are different predictors and that switching amongst them will generate changes in the way the model propagates shocks: this requires that the heuristics in the set are not too similar. Later in the paper, we interpret the inflation-target predictor as one that can be added to the set of available predictors if the central bank declares an explicit inflation objective. At that point it will be crucial to consider predictor sets that initially exclude, and later include, the inflation-target predictor, so our predictor set must be taken more literally.

One of the difficulties of working with a model of nonrational expectations is that there are so many to choose from. So there is an inevitable arbitrariness about our choice of heuristics. But we do not view this as too much of a drawback, since the points we make will be qualitative ones. Our choice of heuristics is therefore guided by simplicity and plausibility. The lagged-inflation heuristic is simple and appeals to much of the empirical literature on inflation expectations (often termed "naive expectations"). The inflation-target heuristic is designed to capture the potential effect of inflation targets as anchors for expectations, so here the heuristic is effectively chosen for us.

Agents in our model differ from those embedded within adaptive learning models. In those models, the tools that agents use to forecast encompass the true model. In variants where agents have access to the entire history of data, they may eventually learn the true coefficients. Our agents' models are both misspecified, and agents

[7] Diron and Mojon (2005) document how using the central bank's stated target as a forecast rule of thumb can perform well relative to alternative models.

have a fixed window for evaluating their predictors that prevents the apparent performance of these predictors converging over time.

We follow our predecessors in this literature and assume that the heuristics are selected according to their recent forecast performance. Specifically, we define the objective function as

$$F_{i,t} = -\frac{1}{H}\sum_{j=1}^{H}[\pi_{t-j} - E_{i,t-j-1}\pi_{t-j}]^2 \tag{14}$$

for $i = 1, 2$. The term on the right-hand side is the "mean squared error" of the heuristic, calculated over the previous H periods. This captures the ability of the heuristic to match the behavior of inflation in the recent past. The objective can be thought of as some form of "utility function": agents prefer heuristics with higher F scores.[8]

The proportion of agents choosing each predictor, $n_{i,t}$, is determined by the following function:

$$n_{i,t} = \frac{\exp(\theta F_{i,t})}{\sum_{j=1}^{2}\exp(\theta F_{j,t})}, \tag{15}$$

where the parameter $\theta > 0$ is referred to in previous work as the "intensity of choice." Brock and de Fontnouvelle (2000) note that in this model θ can be related to the amount of noise in observing the forecast error function F.[9] The larger is θ, the more accurately agents observe the past forecast performance of the heuristics, and the more the portion of agents using each heuristic responds to forecast performance. The limit of $\theta = \infty$ represents the case in which all agents observe perfectly—and hence choose—the best heuristic in each period. As θ approaches zero, we approach a situation in which the noise in observing predictor performance is so large that predictor choice is entirely nonsystematic. To emphasize, with a finite

[8]The thought experiment that agents are conducting here is flawed, and it highlights the difference between their behavior and that under rational expectations: the performance of a heuristic in forecasting actually depends on how many agents use it for forecasting. Agents neglect this fact when they compute F from recent observations on π.

[9]The authors steer the reader to the unabridged (1996) version of this paper, University of Wisconsin Working Paper No. 9624, for a complete account of this interpretation (and others) of the model.

θ, the presence of measurement error means that agents will not always pick the best-performing heuristic. But the probability that they will pick a particular heuristic will increase with its past forecasting performance. The share of the population using each of the two heuristics will equal the probability that any individual picks that heuristic.

Aggregating across young agents, we have the following:

$$E_t \pi_{t+1} = n_{1,t} \pi_{t-1}.$$

Thus the real-money-demand relation under heuristics is given by

$$m_t = \frac{1+\eta}{\eta+\alpha} a_t - \frac{1-\alpha}{\eta+\alpha} n_{1,t} \pi_{t-1}. \tag{16}$$

5. Model Properties under Rational Expectations and a Single Heuristic

We simulate the model comprising the equation for n_1; the portion using the lagged-inflation heuristic, (15); and the linearized equations for real money demand, the government budget constraint, and the productivity and money processes (equations (16), (8), (9), and (11), respectively).

We use the following parameter values: $\eta = 0.2$; $\alpha = 0.41$; $\Pi = 1.02$; $\theta = 100,000$; $\rho = 0.925$; $\sigma^2 = 0.000075$; and $H = 50$. Critically assessing the suitability of these parameters is difficult, given the highly stylized structure of the model. We emphasize simply that we are using this model in the hope that it can say something interesting about the dynamics of an economy over business-cycle frequencies and be of interest to monetary policymakers who have to design a policy to stabilize the economy over such time periods.

Nevertheless, some discussion of our chosen parameters is warranted. Our choices for η and α imply that the elasticity of real money demand to expected inflation (equal to $\frac{1-\alpha}{\eta+\alpha}$) is close to unity, which means that real money balances are relatively responsive to expected inflation. Marcet and Nicolini (2003) use parameter values that imply that real money demand is rather less responsive to changes in expected inflation (their parameters would imply a slope

$\frac{1-\alpha}{\eta+\alpha}$ of around 0.15), but simulations under this type of parameterization are qualitatively similar to those we present here.

Our choice of Π implies that the steady-state inflation rate is 2 percent per period, which matches the rate chosen by some central banks if we interpret a period as one year. This choice bounds our choice for the variance of the productivity disturbance. This—together with the design of the process for monetary policy, g—will govern the frequency with which the implied level of government spending is negative, which we want to keep to a minimum. The degree of persistence in the shocks affects the chance of lagged inflation proving to be a good forecaster of future inflation, and therefore of agents using it as a heuristic. The variance of productivity implied by our assumed values for σ^2 and ρ is of a similar order of magnitude to cyclical output variations.[10]

The ability of the model to generate switches in heuristic use is also determined by the evaluation horizon H and the intensity of choice θ (which we prefer to interpret as the accuracy with which heuristic performance is observed). The shorter the evaluation horizon, the larger the fluctuations in observed forecast performance. The greater the intensity of choice, the larger the response of heuristic choice to movements in forecast performance. The important thing for the story in this paper is that some economically significant degree of heuristic switching occurs.

Table 1 records some time-series properties of three versions of our overlapping-generations model. We report variances as an index for which 100 equals the rational-expectations case. In each case the model is solved under the money process that is optimal under rational expectations. The first column reports the rational-expectations version of the model discussed in section 3. The variance and autocorrelation of inflation are calculated from the equivalent moments of the forcing process, productivity. For the other cases, statistics are computed from 1,000 Monte Carlo replications of 20,000 periods each. We summarize this Monte Carlo experiment by reporting the mean, 5th, and 95th percentiles respectively, in each cell. "Lagged inflation" refers to a

[10] The standard deviation of log-productivity (a) is given by $\sqrt{\sigma^2/(1-\rho^2)} \approx 0.023$. The variance of residuals from a regression of annual UK (log) GDP on a time trend is around 0.03.

Table 1. Time-Series Properties of the Rational-Expectations and Single-Heuristic Models

	Rational Expectations	Lagged Inflation	Inflation Target
var(II)	100	1,020 (918, 1110)	129 (127, 130)
var(var(II))	100	36,600 (25800, 50600)	110 (107, 112)
ρ(II)	0.925 (0.920, 0.929)	0.541 (0.536, 0.546)	0.711 (0.697, 0.725)

Note: Variances relative to rational-expectations case (=100). Numbers in parentheses are 5th, 95th percentiles.

model in which agents are restricted to the heuristic that inflation tomorrow is equal to inflation yesterday. "Inflation target" refers to a model in which they are restricted to the inflation-target heuristic.

These results serve as a benchmark against which we compare our model when agents switch between the two heuristics. They also provide some intuition about what happens to the time-series properties of variables as the number of agents using each heuristic switches between the extremes implied by these first simulations. The first row of table 1 shows the variance of inflation, which is about ten times larger when all agents use the lagged-inflation heuristic compared with the rational-expectations benchmark. The second row shows the variance of the variance of inflation. This is computed by first forming a time series of a rolling fifty-period variance of inflation and then calculating the variance of that. We are interested in this statistic because it connects with our concern to examine the durability of the "Great Moderation" seen in the variance of inflation in developed economies recently. When all agents use the lagged-inflation heuristic, this measure is about 360 times larger than in the rational-expectations case. The final row shows the coefficient from a first-order autoregression of inflation. This illustrates how the estimated time-series behavior of inflation depends on the method with which agents are forecasting inflation. The results for the "inflation-target" model are similar to those for "rational expectations."

6. Model Properties under Heuristic Switching

In this section we report the results from simulating the model when agents switch between the two heuristics depending on their past forecasting performance. As a benchmark, we continue to assume that money growth follows the process that would be optimal if agents formed rational expectations. The summary statistics are shown in table 2. Here, we perform 1,000 replications of 200,000 periods each, computing our statistics based on the final 20,000 periods. We use longer simulations to purge the effect of our initial conditions for the heuristics (we assume that all agents start out using the lagged-inflation heuristic). Experimentation showed that 200,000 periods was long enough for the estimates of the statistics of interest to converge.[11] We continue to normalize all variances to equal 100 in the rational-expectations case.

Overall, the variance of inflation in this heuristic-switching economy is higher than when all agents were forced to use the inflation-target heuristic but lower than in the economy where all agents used

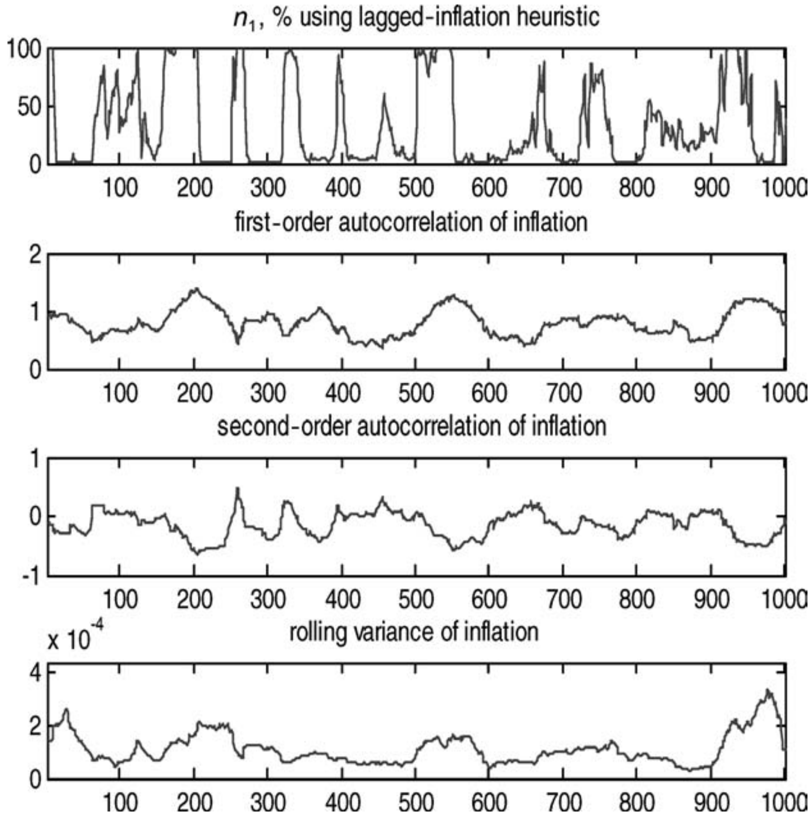
Table 2. Time-Series Properties of Heuristic-Switching Model

var(II)	165
	(163, 166)
var(var(II))	316
	(289, 348)
ρ(II)	0.707
	(0.698, 0.715)

Note: Variances relative to rational-expectations case (=100). Numbers in parentheses are 5th, 95th percentiles.

[11] Repeating the experiment using twice as many periods (400,000) gives statistics that are essentially identical to those reported here. For example, the estimates of inflation persistence reported in the paper are within 0.001 for the estimates using twice as many periods.

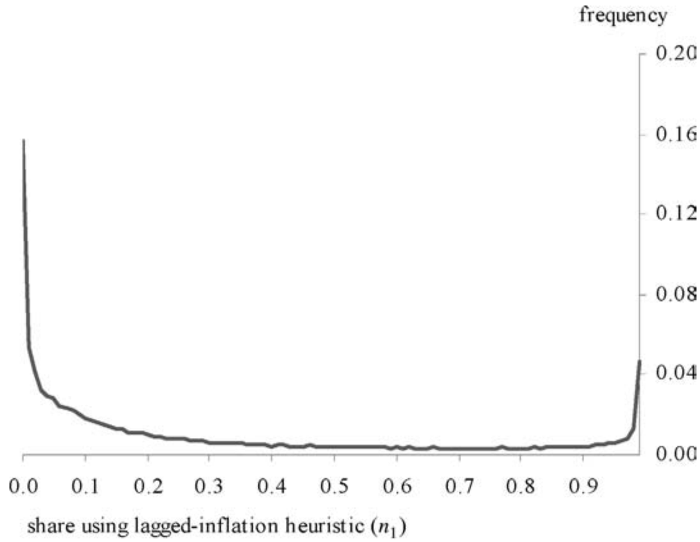
Figure 1. Heuristic Switching under Rational-Expectations Policy (First- and Second-Order Autocorrelations Are Regression Coefficients on First and Second Lags of Inflation)



the lagged-inflation heuristic. The same is true of fluctuations in the small-sample variance of inflation.

In figure 1, we plot a 1,000-period extract from one 20,000-period simulation to illustrate the dynamics of this heuristic-switching economy. The top panel of the figure shows how the proportion of agents using the lagged-inflation heuristic, n_1, fluctuates. It sometimes reaches the upper bound of 100 percent but is generally close to zero. On average, the proportion of agents using the lagged-inflation heuristic is about 30 percent. Switching between the two

Figure 2. Share (n_1) of Agents Using Lagged-Inflation Heuristic under Rational-Expectations Policy



heuristics is an important determinant of the time-series behavior of variables.[12]

Figure 2 is an alternative—histogram—representation of these movements in n_1. It shows that the distribution of n_1 is bimodal. If the intensity of choice (θ) was infinite, then we would expect the observations to be either $n_1 = 0$ or $n_1 = 1$ as agents are able to perfectly observe the best performing predictor. But since θ is finite (though large), there are some observations between these extremes.

Though the model spends most of the time in a region where the majority of agents are using the inflation-target heuristic, there are episodes where almost all are using the lagged-inflation heuristic. These results reflect the fact that agents in our model use a finite sample of recent data to evaluate predictor performance: in

[12]Indeed, when plotting inflation alongside the series for productivity (a_t), it is difficult to discern by eye how the productivity shocks are transmitted into inflation outcomes. The reason is simply that heuristic switching changes the coefficients in the model equations—that is, the mapping from exogenous shocks to endogenous variables.

the jargon of the learning literature, they assess forecast performance using "constant gain." If instead we allowed agents in the model to learn with "decreasing gain" (that is, using the entire history of the data), the model would generate a histogram centered around a single, interior value of n_1. This is because our model exhibits what has been called "negative feedback" from heuristic use to heuristic performance. These aspects of macroeconomic models with predictor choice are discussed in Branch and Evans (2007), who suggest that this negative feedback effect may be relatively uncommon in macroeconomic models. Instead, they construct a simple model with "positive feedback," characterized by multiple equilibria, some of which are unstable. At such equilibria, disturbances that, for example, increase the proportion of agents using a given predictor improve the relative performance of that predictor, further increasing the proportion, and so on.

Positive feedback and multiple equilibria can be generated in our model under suitable parameterizations for the productivity process and the conduct of monetary policy. For example, we found that the monetary reaction function

$$\Pi^{-1} g_t = -m_{t-1} + 0.5a_t - 0.25n_{1,t}\pi_{t-1}$$

was able to generate these properties when we set $\rho = 0.6$.[13] But under policy that is optimal when agents form rational expectations—and, indeed, under the policy that attempts to take account of heuristic switching that we derive below—we have negative feedback between heuristic use and performance.

The bottom three panels of figure 1 illustrate how heuristic switching generates small-sample fluctuations in the time-series properties of inflation. The panels labeled "first-" and "second-order autocorrelation" report rolling coefficients from a regression of inflation on two lags of itself. The bottom panel plots the variance of inflation. These moments are calculated over a horizon of fifty periods. When the proportion of agents using the lagged-inflation heuristic is high for a sustained period, so is the variance of inflation; at these times the coefficient of the first lag of inflation in an

[13]The coefficient on the lag of real money balances is suggested by the form of the reaction function used by Branch and Evans (2007).

autoregression of inflation is high, and the coefficient on the second lag is low. We gain some insight into these fluctuations by fixing n_1 and writing the reduced form for inflation:

$$\pi_t = \Pi^{-1} g_t - \frac{1+\eta}{\eta+\alpha} a_t + \frac{1+\eta}{\eta+\alpha} a_{t-1} + \frac{1-\alpha}{\eta+\alpha} n_1 \pi_{t-1} - \frac{1-\alpha}{\eta+\alpha} n_1 \pi_{t-2}.$$

As we see in the simulations, also in this reduced-form equation for inflation we notice that the higher is n_1, the higher is the coefficient on π_{t-1} and the lower is the corresponding coefficient on π_{t-2}.

These fluctuations in the autocorrelation function for inflation echo the debates about what has caused the fluctuations in inflation persistence, documented by, amongst others, Benati (2004) and Levin and Piger (2004). That debate has generated two broad answers: (i) that changes in inflation persistence have come about because of structural change or (ii) that they reflect changes in monetary policymaking and the introduction of inflation targeting. Our model generates changes in small-sample moments of inflation that reflect neither, but instead are the result of heuristic switching.

7. Monetary Policy under Heuristic Switching

So far we have worked with the money-growth process that would be optimal under rational expectations. We now consider if the central bank can improve on this process in light of its knowledge about expectations formation. There are two motivations. From a positive standpoint, we can check that the heuristic-switching explanation for the appearance (and possible disappearance) of low inflation volatility is robust to cases in which the central bank follows a more sensible policy. From a normative standpoint, we can highlight the cost of the central bank incorrectly assuming that expectations are rational.

In section 3, we showed that, under rational expectations, a rule for money growth that responded to productivity could stabilize labor input and consumption. It did so through its impact on anticipated future money growth and inflation. When agents use heuristics, commitment to a policy rule no longer delivers any direct leverage on expected future inflation. Policy only affects expectations indirectly through past inflation. The lack of direct leverage over expectations means that, unlike the rational-expectations case, policy cannot offset all the welfare losses arising from productivity shocks. It needs to adapt to the use of heuristics.

Additional complications arise in attempting a study of the welfare consequences of policy under heuristics. Heuristic switching makes the model nonlinear, even when the individual decision rules are linearized.[14] This nonlinearity causes two problems.

The first problem is that we cannot derive an optimal monetary policy analytically, even when we use the quadratic approximation to welfare explained above. So we have to resort to numerical methods. We define a class of candidate monetary policy processes and then simulate the model under each rule within that class, compute welfare, and look for the rule that scores the highest. The particular nonlinear nature of our model means that we have to simulate for millions of periods to get reliable estimates of our welfare function. So we must confine our search across alternative policy rules to make the exercise manageable. We will work with the following class of rules for money growth:

$$g_t = \chi_1 a_t + \chi_2 E_t \pi_{t+1}.$$

This process allows the policymaker to respond to productivity and to data on expected inflation. We assume that policymakers receive data on expected inflation but do not attempt directly to internalize the interaction between policy, endogenous inflation outcomes, and n_1. (Indirectly, policymakers will choose the combinations of χ_1 and χ_2 that generate the most beneficial paths for n_1, the proportion using the lagged-inflation heuristic.) We search for the values of χ_1 and χ_2 that deliver the best welfare for our agents, defined by our criterion in equation (10).

The second problem caused by the nonlinearity of the model is that alternative policy rules will generate small differences in the mean rates of inflation. These will cause the average levels of utility to differ according to the policy rule, as the government budget constraint means that higher average inflation implies higher average government spending and higher resource destruction. The differences in means will not affect the welfare criterion we have chosen,

[14]The fraction (n_1) of agents using the lagged-inflation heuristic affects the coefficients of the decision rules. And n_1 itself varies over time, in response to the behavior of the economy.

which is defined on variances. So it must be stressed that our search can rank policy rules only according to their stabilization properties, and not their effect on means.[15]

We focus on rules that respond to productivity and inflation expectations for two reasons. First, this class of rules allows us to nest the optimal policy under rational expectations, which responds to the only state variable in that model, productivity. Second, it also allows the policymaker to respond to another state variable in the heuristic-switching model, expected inflation. And that happens to echo the concerns of policymakers in reality.[16]

We can get some intuition for why a rule like this is likely to work by considering an extreme case that the policymaker will face: one in which all agents use the inflation-target heuristic. When everyone is using the inflation-target heuristic ($n_1 = 0$), the labor supply function (6) collapses to

$$l_t = \frac{1-\alpha}{\eta+\alpha} a_t.$$

Fluctuations in labor supply are inevitable. The average expected welfare of young agents is lower than when agents have rational expectations, and policy responds optimally. Under heuristics, monetary policy is powerless to influence this. But monetary policy can help old agents. The consumption of old agents at date t is

$$c_t = m_{t-1} - \pi_t$$

and the evolution of real money balances is given by

$$m_t = \Pi^{-1} g_t + m_{t-1} - \pi_t$$

[15]These small differences in mean inflation will also have a small effect on the performance of the inflation-target heuristic under the alternative policy rules. The higher the mean inflation rate, the worse the (zero) inflation-target heuristic performs, and the smaller the portion of agents who use it.

[16]Expectations-based rules have been argued to have benefits in other contexts. For example, Evans and Honkapohja (2003) have recommended them as devices for implementing monetary policy to ensure that the rational-expectations equilibrium is stable under least-squares learning.

so that the policymaker can fully stabilize c_t by committing to the policy rule:

$$\begin{aligned} g_t &= \Pi m_t \\ &= \Pi \frac{1+\eta}{\eta+\alpha} a_t, \end{aligned}$$

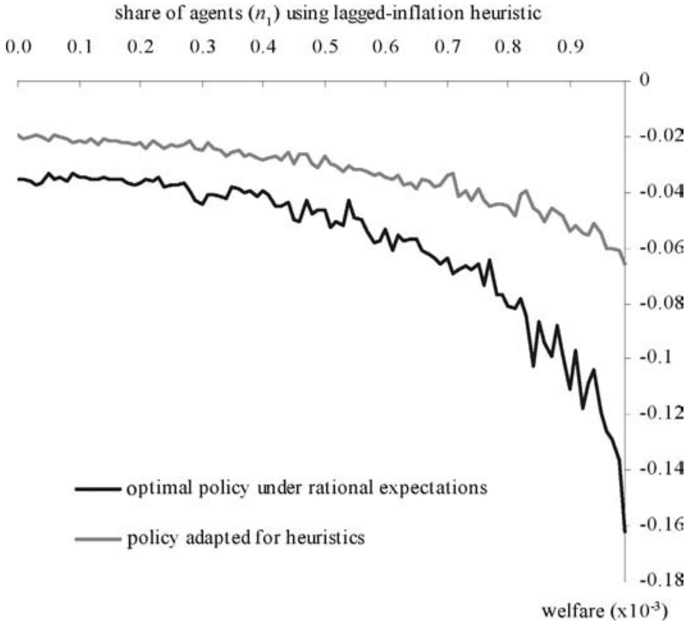
which, since $\frac{1+\eta}{\eta+\alpha} > 1$, implies a stronger response to productivity shocks than under rational expectations.

In the general case, where some agents use the lagged-inflation heuristic, labor supply and the demand for real money balances depend on inflation expectations, which in turn depend on lagged inflation. In that case, even in the absence of a current productivity shock, labor supply and output can fluctuate. Without any policy action, inflation will move to bring the real value of money balances into line with output. These fluctuations are costly, so monetary policy might do better by responding to inflation expectations as well as to productivity. Of course, one thing this discussion reveals is that the ideal response to productivity and inflation expectations should itself depend on n_1. However, to make the analysis more tractable, we confine our search to rules that involve constant, independent values of χ_1 and χ_2.[17]

The best rule in our grid search is one with values of $\chi_1 = 2$ and $\chi_2 = -1.75$. This policy shares a feature with the optimal policy under rational expectations in that money growth is expanded when productivity is unusually high. A positive shock to productivity reduces the price level; a positive money-growth response by policy therefore acts to offset that. The policy response under heuristics is to respond more aggressively (recall that under rational expectations, χ equals Π, the steady-state rate of inflation, which is 1.02). We believe that this response allows the policy to perform well when few agents believe the inflation target: as described above, in this setting, an aggressive response to productivity can help to stabilize the consumption of old agents. The heuristics policy also suggests

[17]Using this shortcut naturally raises the issue of whether it would be appropriate to build a model of heuristic policy design on the part of the central bank to go with the heuristic expectations formation on the part of agents in the model. We leave that issue for future research.

Figure 3. Welfare Generated by Alternative Policy Rules as the Share of Agents Using the Lagged-Inflation Heuristic Varies



that money growth should fall when expected inflation rises. When expected inflation rises, labor supply and demand for real balances fall. Monetary policy can stabilize inflation by contracting the money supply.

The rule considered here generates higher welfare than arbitrary persistent processes for money growth, fixed money growth, and the policy that would be optimal under rational expectations (derived in section 3). The welfare surface appeared well behaved in the space used for the grid search. Figure 3 shows how welfare differs under the two policy rules at different values of n_1, the portion using the lagged-inflation heuristic. We arrange the simulated periods according to their associated value of n_1 and calculate average welfare at each value of n_1.

As we can see, when the central bank tries to take account of heuristics, it delivers higher welfare than the rational-expectations

Table 3. Time-Series Properties of Heuristic-Switching Model

	Policy Process	
	Rational Expectations	**Heuristics**
var(II)	165	163
	(163, 166)	(158, 168)
var(var(II))	316	170
	(289, 348)	(145, 199)
ρ(II)	0.707	0.657
	(0.698, 0.715)	(0.653, 0.661)

Note: Variances relative to rational-expectations case (=100). Numbers in parentheses are 5th, 95th percentiles.

policy at all values of n_1. The welfare improvement achieved by the heuristics-adapted policy is greater for larger values of n_1: the more agents are using the lagged-inflation heuristic, the greater the benefit of following the policy adapted for heuristics, or, put another way, the greater the cost of policymakers mistakenly following the policy that would be appropriate under rational expectations.[18]

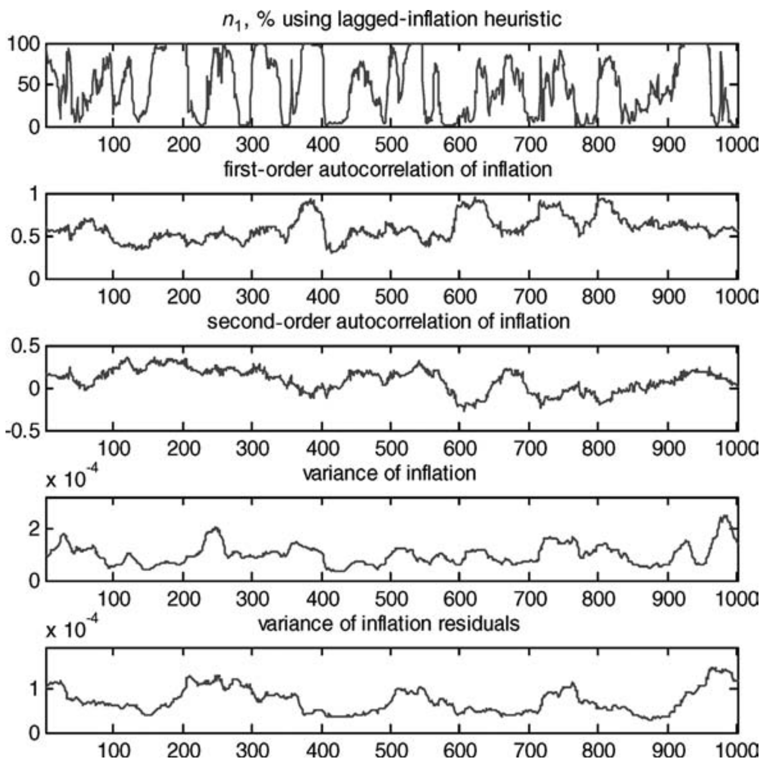
Table 3 shows summary statistics that compare two versions of the heuristic-switching model. In one, monetary policy follows the process that would be optimal under rational expectations. In the other, monetary policy is adapted for heuristic switching. As before, we report moments from 1,000 replications of simulations of 200,000 periods each, with statistics computed using the final 20,000 periods. We continue to report moments of inflation—persistence aside—as an index where 100 is the value for the model under rational expectations and the associated optimal policy.

[18]We have calculated that the minimum value for these costs, when few or no agents are using the lagged-inflation heuristic, is still more than ten times the welfare cost of mistakenly pursuing the heuristics policy when agents actually have rational expectations. This is an indication that if policymakers were unsure how agents arrived at their forecasts, a safe policy would be to assume that agents did not have rational expectations. This contrasts somewhat with Gaspar, Smets, and Vestin (2006), who found that the optimal rational-expectations policy does quite a good job of replicating the optimal policy in a model where agents form expectations using adaptive learning.

When agents switch between heuristics, the variance of inflation under a policy rule that takes switching into account is roughly the same as the variance of inflation under the rational-expectations policy. Under the policy that adapts to heuristics, the volatility of the small-sample estimates of the variance of inflation is less than one-fourth that under the rational-expectations policy. But note that it is still more than four times the figure we observe for the model under rational expectations. Note too that inflation is a little less persistent under the policy adapted to heuristics.

Figure 4 plots an extract from one 20,000-period simulation of the model with policy adapted to heuristics. Notice that the fluctuations

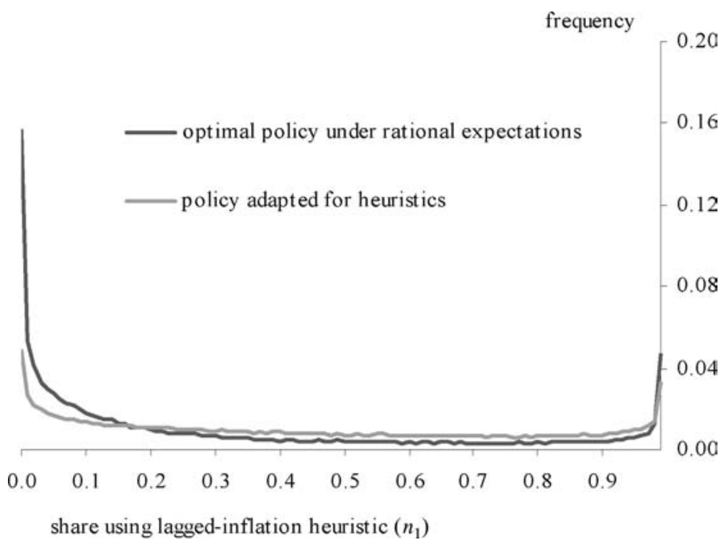
Figure 4. Heuristic Switching under a Heuristics Policy (First- and Second-Order Autocorrelations Are Regression Coefficients on First and Second Lags of Inflation)



in n_1, the proportion of agents using the lagged-inflation heuristic, are, to the eye, as pronounced as those under the policy that would be optimal under rational expectations. Figure 5 compares the histograms for n_1 that are generated in the heuristic-switching economy both when policy follows the rational-expectations optimal rule and when it adapts to the use of heuristics. Relative to the rational-expectations optimal policy, the heuristics-adapted policy reduces the probability mass at both extremes of n_1 and increases it slightly at interior values.

Under both policies heuristic switching generates small-sample fluctuations in the time-series properties of inflation. We can see this from the volatility in the coefficients on lagged inflation in an autoregression for inflation. The bottom panel of figure 4 plots the variance of the residuals from a rolling fifty-period regression for inflation on its own lags. This variance is clearly moving over time and tends to be high when the variance of inflation is high, and vice versa. We plot this time series to link our analysis to the econometric studies that report that large fractions of recent declines

Figure 5. Share of Agents Using Lagged-Inflation Heuristic under Alternative Policy Rules



in macroeconomic volatility are due to "good luck."[19] Here, very loosely, when the number using the inflation-target heuristic is low, the variance of inflation is low, and the variance of the shocks in a simple autoregression is low. In the language of the applied literature on the Great Stability, the econometrician estimates there to have been a period of good luck, when the true variance of the disturbances to our model economy is unchanging.

8. Model Properties after the Introduction of an Inflation-Target Heuristic

Thus far, we have investigated whether switching amongst heuristics can generate fluctuations in small-sample estimates of the volatility of inflation that are consistent with the marked reduction in volatility seen in recent decades. And our contention is that it can. These fluctuations occur regardless of whether monetary policy adopts a different rule. So far we have considered the set of heuristics as something beyond the control of policymakers. In this section, we assume that the monetary policy framework can influence the set of heuristics from which agents choose, and we consider what happens when agents are given access to an inflation-target heuristic that was not previously available to them. We suggest that this may be a way of formalizing what happened when many central banks adopted numerical objectives for inflation. This exercise is related to one conducted by Orphanides and Williams (2005). They interpret the introduction of a numerical objective for the central bank as equivalent to giving agents knowledge of the constant in the inflation process: this knowledge improves agents' estimates of the dynamics of that process.

Table 4 presents simulations of the introduction of an inflation-target heuristic into four different models. The four models correspond to the table columns and comprise two different initial lagged-inflation heuristics, derived under two different processes for monetary policy. Under the columns headed "Rational Expectations" we have results that use our baseline process for money that would be optimal under rational expectations. Within this we use

[19] See, for example, Stock and Watson (2002) and Cogley and Sargent (2005).

Table 4. Impact of Introducing the Inflation-Target Heuristic

Policy Process:	**Rational Expectations**		**Persistent**	
Heuristic:	**Lagged Inflation**	**Best AR**	**Lagged Inflation**	**Best AR**
Before Target				
var(II)	100	12.7	160	17.1
var(var(II))	100	0	0.108	0
ρ(II)	0.541	0.705	0.521	0.636
After Target				
mean(n_1)	30.3	42.2	17.7	72.3
n_1 impact	0	42.3	0	66.4
var(II)	16.3	12.7	18	16.4
var(var(II))	0.912	0.310	0.996	0.397
ρ(II)	0.707	0.708	0.593	0.626

Note: Variances relative to rational-expectations/lagged-inflation case in top-left quadrant.

two heuristics. The first, "Lagged Inflation," is our familiar lagged-inflation heuristic. The column headed "Best AR" refers to a model in which expectations of inflation are determined by the projection of inflation tomorrow on inflation yesterday implied by the model itself. Specifically, we assume that agents set $E_t(\pi_{t+1}) = \rho_h \pi_{t-1}$. We determine $E_t(\pi_{t+1})$ by the following process. First, agents collect all data to time $t-1$ and run a regression $\pi_s = \rho_h^{ols} \pi_{s-2}$ for $s = \{1, \ldots, t-1\}$. Second, agents use ρ_h^{ols} to form $E_t(\pi_{t+1})$. Third, another data point for time t is generated. Agents add this to their data set and return to the second step. The value of ρ_h used to compute numbers under the "Best AR" column in table 4 is the number to which this iterative process converges.[20]

The two columns under "Persistent" repeat this analysis, but using a persistent process for money growth where the persistence

[20]The point to which this iteration converges might be referred to as a restricted-perceptions equilibrium. Subject to the restricted perceptions of the inflation process that agents have, their projections are optimal.

and variance are set equal to the values chosen for the productivity process (and with no correlation between the two). Results are, as before, derived from 1,000 simulations of length 200,000 periods. In each replication, we introduce the inflation target after 40,000 periods. We simulate the model for a further 160,000 periods: aside from where we are interested in the impact effect of the introduction of the inflation-target heuristic, we compute statistics based on the final 20,000 periods of the simulation. We weight the bulk of our simulation time toward the period when we have two heuristics, because we need longer simulations to get accurate estimates of the statistics for the two-heuristic model.[21] (In table 4, figures are reported relative to the rational-expectations/lagged-inflation case, which is indexed to 100 and appears in the top-left quadrant.)

We report several details. First, in the top rows, we give statistics for the economy before the introduction of the inflation-target heuristic. These are the variance of inflation (row labeled "var(Π)"); the variance of short-sample estimates of that variance ("var(var(Π))"); and the persistence of inflation ("$\rho(\Pi)$"). For the second half of the simulation, after the introduction of the inflation target, we report these same statistics, but with two additions. First, we report the average value of n_1 in the five periods immediately following the introduction of the target and label this row "n_1 impact." Second, we report the mean of n_1 over the life of the rest of the simulation (labeled "mean(n_1)"). In this table, we normalize variances and the variance of variances relative to those computed for the top left-hand case in this table—the case where agents have a single, simple lagged-inflation heuristic, and policy is conducted according to the rule that would be optimal under rational expectations.

The basic message is that the immediate impact effect of the introduction of the inflation-target heuristic is maximal when, prior to that, agents use only the lagged-inflation heuristic. In both the "lagged-inflation" simulations, n_1, the number using the lagged-inflation heuristic, drops to zero in the period immediately following introduction of the inflation target (albeit rising again thereafter). This is shown by the zeros recorded in the row labeled "n_1 impact."

[21] The single-heuristic models are linear models, and we know (from checking the appropriate analytics) that the relevant statistics are estimated accurately with short simulations.

It turns out that in our model, if we exogenously impose that $n_1 = 1$, it greatly worsens the forecast performance of that heuristic, which is why when agents are free to choose between two heuristics, they jump to using the inflation target for a while.

This begs the question of why agents were content to use only the lagged-inflation heuristic prior to the introduction of the target. It is beyond the scope of this paper to model the complete process that specifies the evolution of the set of heuristics that agents use. But for comparison, we have the simulations where agents start out life using a heuristic based on an optimal projection of inflation tomorrow on inflation yesterday (the "best AR" simulations). With the use of such a projection, one which performs better than the simple lagged-inflation heuristic, the effect of the new target heuristic is more muted: this is true under both our "rational-expectations" and "persistent" monetary policy processes.

Similarly, we see that when agents are constrained to use the simple lagged-inflation heuristic, the introduction of the inflation target has its largest effect on the time-series properties of inflation, reducing the variability of inflation and the fluctuations in small-sample estimates of this variability.[22]

To summarize, the ability of the model to provide a dramatic reduction in inflation volatility and for that reduction to be durable depends on the sophistication of agents' forecasting methods before the introduction of the inflation target.

9. Conclusions

In the past decade, both inflation and output growth seem to have become more stable in advanced economies. This coincided with the convergence of inflation expectations on inflation targets. We have illustrated how an economy populated by agents who choose amongst heuristics for forecasting inflation can generate fluctuations in the variance of inflation. There are periods in which agents use the

[22]We repeated the experiment many times and found that the main determinant of the impact effect was the assumption about the heuristic that agents used before the introduction of the inflation target. This was more important than, for example, the recent history of productivity shocks in the periods preceding the target introduction.

inflation-target heuristic, and there are periods when many agents choose to use a heuristic based on lagged inflation. In the former, a given shock will generate less variability in inflation. But a sequence of shocks that reduces the ability of the inflation-target heuristic to match inflation in the past can lead agents to switch to the lagged-inflation heuristic.

We asked how monetary policy might adapt to agents' use of heuristics. Under rational expectations, a rule for money growth that responded to productivity could stabilize completely labor supply and consumption. It did so through its leverage over expectations. When agents use heuristics, monetary policy has no direct leverage over inflation expectations, which are determined entirely by the past behavior of inflation. Relative to the policy that would be optimal under rational expectations, a money-growth rule that reacts to both productivity and inflation expectations can better stabilize the economy. Even under such a policy, agents switch back and forth between heuristics, and the time-series properties of inflation tend to fluctuate.

Our final exercise was to simulate the introduction of an inflation-target heuristic. When we did this, there was some evidence that the introduction of this heuristic improves macroeconomic outcomes by reducing the volatility of inflation. By how much, and to what extent, agents use the new heuristic depends on the performance of the heuristics they had before. These results suggest that some of the improvements seen in the United Kingdom and elsewhere could be locked in, at least if the inflation-targeting regime can be thought of as having made available the simple heuristic that "inflation will equal the target."

References

Benati, L. 2004. "UK Monetary Regimes and Macroeconomic Stylised Facts." Bank of England Working Paper No. 290.

Bernanke, B. S. 2004. "The Great Moderation." Remarks delivered at the meetings of the Eastern Economic Association, February 20, Washington, DC.

Branch, W. A., and G. W. Evans. 2006. "Intrinsic Heterogeneity in Expectations Formation." *Journal of Economic Theory* 127 (1): 264–95.

———. 2007. "Model Uncertainty and Endogenous Volatility." *Review of Economic Dynamics* 10 (2): 207–37.

Brock, W. A., and P. de Fontnouvelle. 2000. "Expectational Diversity in Monetary Economies." *Journal of Economic Dynamics and Control* 24 (5–7): 725–59.

Brock, W. A., and C. H. Hommes. 1997. "A Rational Route to Randomness." *Econometrica* 65 (5): 1059–95.

Cogley, T., and T. J. Sargent. 2005. "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US." *Review of Economic Dynamics* 8 (2): 262–302.

De Grauwe, P., and M. Grimaldi. 2006. "Exchange Rate Puzzles: A Tale of Switching Attractors." *European Economic Review* 50 (1): 1–33.

Diron, M., and B. Mojon. 2005. "Forecasting the Central Bank's Inflation Objective Is a Good Rule of Thumb." ECB Working Paper No. 564.

Evans, G. W., and S. Honkapohja. 2003. "Adaptive Learning and Monetary Policy Design." *Journal of Money, Credit, and Banking* 35 (6, part 2): 1045–72.

Gaspar, V., F. Smets, and D. Vestin. 2006. "Adaptive Learning, Persistence, and Optimal Monetary Policy." *Journal of the European Economic Association* 4 (2–3): 376–85.

Gigerenzer, G., P. M. Todd, and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press.

Hommes, C. 2005. "Heterogeneous Agent Models in Economics and Finance." In *Handbook of Computational Economics*, Vol. 2, ed. K. Judd and L. Tesfatsion. Elsevier.

Kahneman, D. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93 (5): 1449–75.

Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*. New York: MacMillan.

King, M. 2003. "The Governor's Speech at the East Midlands Development Agency/Bank of England Dinner." *Bank of England Quarterly Bulletin* (Winter).

Levin, A. T., and J. Piger. 2004. "Is Inflation Persistence Intrinsic in Industrial Economies?" ECB Working Paper No. 334.

Manski, C. F., and D. L. McFadden, eds. 1981. *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, MA: MIT Press.

Marcet, A., and J. P. Nicolini. 2003. "Recurrent Hyperinflations and Learning." *American Economic Review* 93 (5): 1476–98.

Orphanides, A., and J. C. Williams. 2005. "Imperfect Knowledge, Inflation Expectations, and Monetary Policy." In *The Inflation-Targeting Debate*, ed. B. S. Bernanke and M. Woodford, 201–34. Chicago: University of Chicago Press.

Simon, H. A. 1957. *Models of Man.* New York: John Wiley & Sons.

Sims, C. A., and T. Zha. 2004. "Were there Regime Switches in US Monetary Policy?" Federal Reserve Bank of Atlanta Working Paper No. 14.

Stock, J. H., and M. W. Watson. 2002. "Has the Business Cycle Changed, and Why?" In *NBER Macroeconomics Annual 2002*, ed. M. Gertler and K. Rogoff. Cambridge, MA: MIT Press.

Velde, F. R. 2004. "Poor Hand or Poor Play? The Rise and Fall of Inflation in the U.S." *Economic Perspectives* (Federal Reserve Bank of Chicago) 28 (1): 34–51.