Volume 11, Number 1 January 2015

IJCB

INTERNATIONAL JOURNAL OF CENTRAL BANKING

IJCB
INTERNATIONAL JOURNAL OF CENTRAL BANKING
Volume 11, Number 1
January 2015

IJCB International Journal of Central Banking

Volume 11, Number 1 January 2015

*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. Orders may be placed by phone (001 415 974 2035), via fax (001 415 974 2168), or by e-mail (editor@ijcb.org).

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

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

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

ISSN: 1815-4654

Imbalances and Rebalancing in an Estimated Structural Model for Spain*

Jan in 't Veld,[a] Andrea Pagano,[b] Rafal Raciborski,[a] Marco Ratto,[b] and Werner Roeger[a]
[a]European Commission, Directorate-General Economic and Financial Affairs
[b]European Commission, Joint Research Centre (JRC)

This paper uses an estimated DSGE model to analyze the factors behind the buildup of imbalances in the Spanish economy. Shock decompositions suggest that external imbalances have been able to build up mainly due to the reduction in real interest rates and easier access to credit following the elimination of the exchange rate risk premium. A rebalancing process in recent years with a sharp contraction in domestic demand has moved the trade balance back into surplus. The main driving factors were the collapse of the housing bubble and tightening of credit conditions.

JEL Codes: C54, E21, E62, F32.

1. Introduction

The existence of large internal and external imbalances in many European countries has become a major concern in the European Union. While the buildup of imbalances and growing current account deficits were recognized as one of the main challenges for adjustment in EMU early on (European Commission 2006, 2008), the financial and sovereign debt crises have exposed the serious vulnerabilities in some countries. In response, the European Union has created the

*This paper has benefited from useful comments from many colleagues at the European Commission. The views expressed in this paper are those of the authors and should not be attributed to the European Commission. Corresponding author: Marco Ratto, European Commission, Joint Research Centre (JRC), TP361, 21027 Ispra (VA), Italy. Author e-mails: jan.intveld@ec.europa.eu; andrea.pagano@jrc.ec.europa.eu; rafal.raciborski@ec.europa.eu; marco.ratto@jrc.ec.europa.eu; werner.roeger@ec.europa.eu.

macroeconomic imbalances procedure (MIP), which in its preventive arm tries to identify at an early stage the existence of internal and external imbalances relative to specific thresholds via a scoreboard. If imbalances are deemed to be harmful, a member state can be placed into an excessive imbalance position and corrective action can be demanded.[1] But this requires a thorough analysis of imbalances and underlines the importance of having a coherent framework to do this.

Current account deficits are not necessarily macroeconomic imbalances, but may reflect an appropriate allocation of savings, given different investment opportunities across countries. External borrowing allows countries to smooth consumption over time, and capital flows from core euro-area countries to the periphery could be a normal feature of savings looking for the highest returns. But distortions due to incorrect expectations of returns and mispricing of risks could lead to a misallocation of resources and a buildup of imbalances and vulnerabilities. As rising and persistent deficits translate into an accumulation of net foreign liabilities, the risk of being tipped into an external crisis becomes non-trivial and accelerates with further exposure. Catao and Milesi-Ferretti (2013) find that once economies' net foreign liabilities rise above 50 percent of GDP, the risk of crisis—defined as either an outright external default or the disbursement of a large multi-lateral financial support package—accelerates with further net liability exposure. Their model predicts Spain as a high-risk case, with signals flashing red since 2008.

The main contribution of our paper is to show how an estimated structural model can be used to examine the driving factors behind the buildup of imbalances in a country. We take the case of Spain and estimate our model over the period 1995–2012, including both the boom years and the subsequent bust, and use this to identify and measure the main shocks driving the Spanish economy. Large imbalances have built up in the Spanish economy since its accession to the euro area. With Spanish membership of EMU came the elimination of currency risk, and a sharp fall in interest rates, which spurred cross-border borrowing that fed a long-lasting housing and

[1] See http://ec.europa.eu/economy_finance/economic_governance/macroeconomic_imbalance_procedure/index_en.htm.

credit boom and large current account deficits and escalating external debt. A correction started in 2007, accelerated by the financial crisis, and the Spanish economy has since gone through a sharp adjustment, with unemployment soaring. But while there has been a significant improvement in the trade balance, Spain's net foreign liabilities remain above 90 percent of GDP, far exceeding euro-area averages and raising concerns about long-run sustainability. For the scoreboard used in the MIP, the "threshold" for the net international investment position is set at 35 percent of GDP, and for Spain this indicator would have been flashing since 2002. Given the stock nature of this indicator, it is unlikely this alert will stop flashing anytime soon.

Our model has some features which make it especially suitable to analyze the Spanish economy, namely residential investment and credit constraints as introduced by Kiyotaki and Moore (1997). Given the prominent role of residential investment and innovations in mortgage lending, we model housing investment explicitly and allow for collateral constraints; see, e.g., also Iacoviello (2005), Monacelli (2007), and Iacoviello and Neri (2010). This helps us in quantifying the extent to which financial innovations have contributed to the boom but also sheds light on the effects of a possible credit crunch in mortgage lending. For a historical decomposition, we use the fitted shocks of the model for a shock accounting exercise to decompose growth rates, domestic demand, and trade balance to GDP ratios to quantify the relative contributions of the different shocks. This allows us to identify the main driving factors behind the buildup of external imbalances and find policies which are most likely to be successful in rebalancing the economy. Possible causes are a loosening of credit growth; bubbles in asset markets, which may have driven up domestic demand; or loss of competitiveness because of insufficient adjustment of wages to productivity growth. By taking the model to the data, one can pinpoint certain developments in labor, housing, and credit markets.[2]

[2]Several papers have used estimated DSGE models to provide an additive decomposition of the data in terms of the estimated shocks. For the United States, Christiano, Motto, and Rostagno (2008) compares the monetary policy response to the 2001 recession in the United States and the euro area. In 't Veld et al. (2011) uses an estimated model to evaluate competing explanations about the boom-bust cycle in the United States.

Our shock decompositions suggest that one of the main factors behind the buildup of imbalances was low real interest rates and easier access to credit, linked to the inflow of cheap capital due to the disappearance of the risk premium and monetary policy set at the euro-area level. This conclusion is broadly in line with the findings in other studies. In a study on adjustment in a monetary union, the European Commission (2006) emphasized the role of the decline in the risk premium in explaining current account deficits in adjustment in a monetary union. Andres et al. (2010) also highlight the role of lower interest rates in Spain. Using an estimated structural model for Spain and comparing this with a version in which Spain is able to set its own monetary policy, they show that an independent monetary authority would have hypothetically pursued different output-inflation trade-offs for most of the sample period, and would have cooled down inflationary pressures at the cost of slightly slower economic growth (ibid., p. 93). Burriel, Fernández-Villaverde, and Rubio-Ramírez (2010) likewise refer to the adoption of the euro and the associated historically low real interest rates as a contributing factor to the long period of continuous real GDP growth since the mid-1990s. Jaumotte and Sodsriwiboon (2010) suggest other southern euro-area countries might also have benefited from a "euro bonus" in running larger deficits than sustainable. The euro helped these countries to maintain higher investment levels by improving their access to international savings, but investment took place in less productive non-tradable sectors, such as construction, and current account deficits now exceed their long-run fundamental "norms."

Our analysis indicates that the correction in the trade balance has been mainly driven by the collapse in the housing bubble and a tightening in credit conditions. The contraction in domestic demand is in line with historical evidence which suggests that current account adjustment in deficit countries has more typically relied on expenditure reduction than on expenditure switching (Lane and Milesi-Ferretti 2011). However, our results also show that the demand contraction has been accompanied by deflation and a decline in unit labor costs which has brought about some depreciation of the real effective exchange rate.

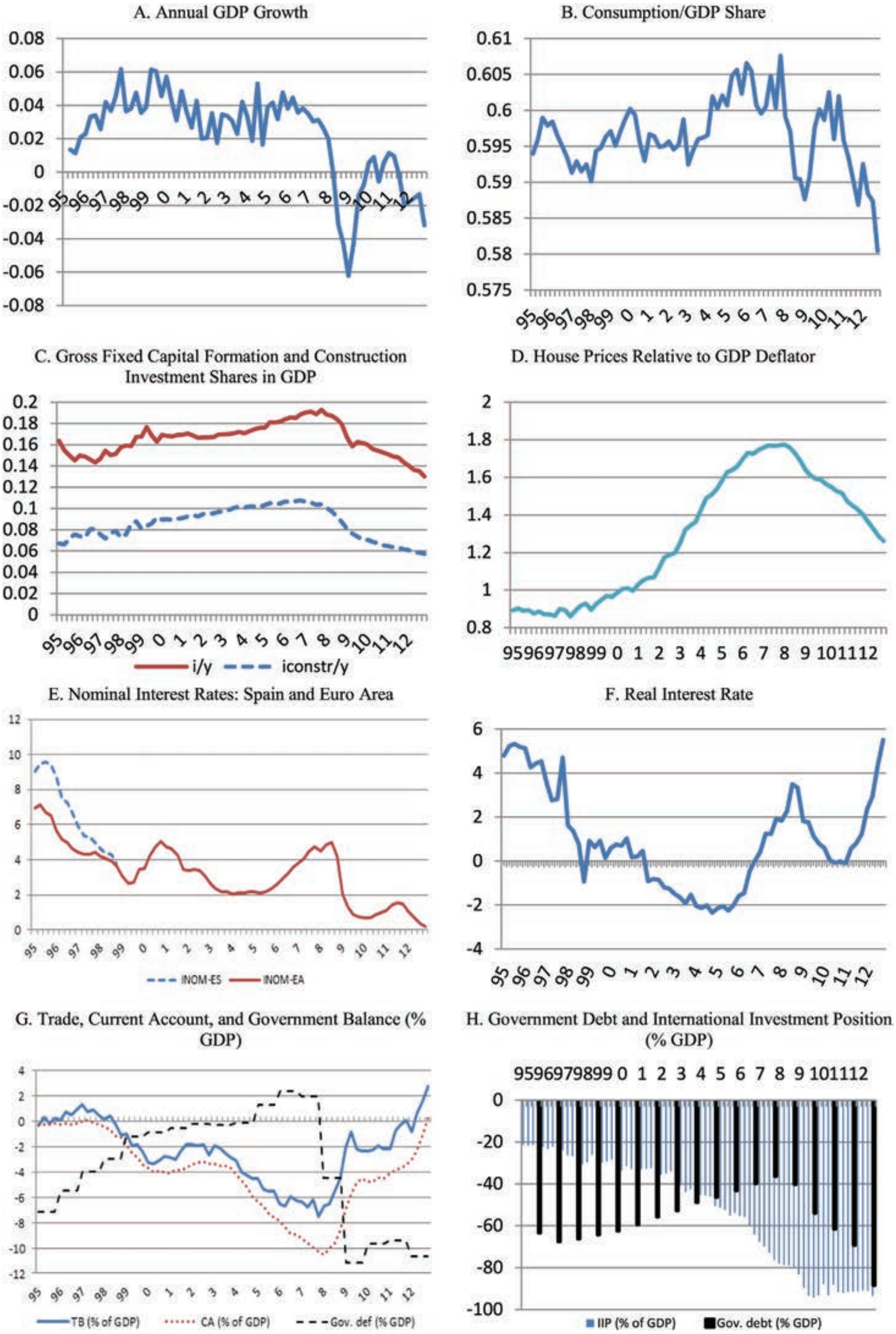
The remainder of the paper is structured as follows. The following section describes some of the main stylized facts of the Spanish economy since the adoption of the euro. Section 3 describes in detail the theoretical model; section 4 discusses the estimation and historical evolution of some of the main shocks. Section 5 discusses the shock decomposition of the main variables of interest to determine the underlying factors behind the buildup of imbalances. Section 6 concludes.

2. Some Stylized Facts of the Spanish Economy

For most of the decade up to 2007, GDP growth in Spain exceeded average growth in the euro area. But the financial and economic crisis hit Spain severely, with quarter-to-quarter annualized growth rates falling to almost −6 percent in the first quarter of 2009, a slower recovery in 2010–11 compared with the rest of the euro area, and a second-dip recession in 2012. While over the boom years all domestic demand components showed a trend increase in their share in GDP, this trend increase was most pronounced in investment, and in particular in residential construction (see figure 1). The Spanish consumption-to-GDP ratio remained relatively stable around 60 percent, about 4 percentage points higher than the euro-area average, but it had fallen to 58 percent by the end of 2012. On the other hand, investment-to-GDP and construction-investment-to-GDP ratios showed a strong positive trend from 1995 up to the last quarters of 2007 and then both dropped sharply in the crisis. This strong performance of investment (especially in the construction sector) relative to output in the 2000s, coupled with a substantial increase in house prices in this period, is suggestive of a housing bubble that developed in Spain prior to the crisis. It is noticeable that house prices in real terms have shown some correction since their peak in 2007, but were by 2012 only back at their 2003 levels.

Accession to EMU led to a gradual elimination of the risk premium on Spanish interest rates with a convergence of the policy rate towards the euro-area average by 1999. The reduction in nominal interest rates also led to a sharp reduction in the real interest rate and even negative rates between 2001 and 2006, but real rates rose sharply in 2012. The boom in domestic demand was accompanied by a strong deterioration in external balances. From the second half of the 1990s, the Spanish economy moved from small trade surpluses to a very large trade deficit of more than 7 percent by

Figure 1. Spain: 1995:Q1–2012:Q4



2007 and even larger current account deficits. The persistent trade deficits accumulated into an ever-increasing net foreign indebtedness, rising from around 20 percent of GDP in the late 1990s to more than 90 percent of GDP by 2009. Since the crisis, the Spanish trade deficit has shrunk considerably, to close to balance by the end of 2011, but its net international investment position remains around −90 percent of GDP. Spain's public finances were, up to the crisis, in better shape than the euro-area average. The Spanish government balance improved considerably between the mid-1990s and mid-2000s, even recording a surplus between 2005 and 2007.[3] Much of this improvement was not due to permanent factors but to increases in tax revenues associated with changes in the composition of GDP—in particular, transitory asset boom revenues (Martinez-Mongay, Maza Lasierra, and Yaniz Igal 2007). The crisis led to a sharp reversal of this trend, with the deficit peaking at 11 percent of GDP in 2009 and remaining persistently high since. This deterioration in the fiscal position reversed the trend decline in gross debt, which had fallen to 36 percent of GDP in 2007, but which increased rapidly to close to 90 percent of GDP by the end of 2012.

3. Model

We consider an open economy, which produces goods that are imperfect substitutes for goods produced in the rest of the world.[4] Households engage in international financial markets, and there is near-perfect international capital mobility. There are three production sectors, a final goods production sector as well as an investment-goods-producing sector and a construction sector. We distinguish between Ricardian households, which have full access to financial markets, and credit-constrained households facing a collateral constraint on their borrowing. The economy is part of a monetary union and faces an exogenous interest rate. There is a fiscal authority, which follows rules-based stabilization policies. Behavioral and

[3] In fact, the experience of Spain of fiscal and current account deficits moving in opposite directions appears to contradict the twin-deficit hypothesis.

[4] The model is an extension of the QUEST model estimated on euro-area data (Ratto, Roeger, and in 't Veld 2009) with a housing sector and collateral-constrained households. A similar version has been estimated on U.S. data (in 't Veld et al. 2011).

technological relationships can be subject to autocorrelated shocks denoted by U_t^k, where k stands for the type of shock. The logarithm of U_t^k will generally be autocorrelated with autocorrelation coefficient ρ^k and innovation ε_t^k.[5]

3.1 Firms

3.1.1 Final Goods Producers

Firms operating in the final goods production sector are indexed by j. Each firm produces a variety of the domestic good which is an imperfect substitute for varieties produced by other firms. Because of imperfect substitutability, firms are monopolistically competitive in the goods market and face a demand function for goods. Domestic final goods producers sell goods and services to domestic and foreign households, investment and construction firms, and governments. Output is produced with a Cobb-Douglas production function using capital $K_t^{j,P}$ and production workers L_t^j as inputs:

$$Y_t^j = (UCAP_t^j K_t^j)^{1-\alpha} L_t^{j\alpha} U_t^{Y\alpha}, \quad \text{with } L_t^j = \left[\int_0^1 L_t^{i,j\,\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{\theta-1}}. \tag{1}$$

Total employment of the firm L_t^j is itself a constant elasticity of substitution (CES) aggregate of labor supplied by individual households i. The parameter $\theta > 1$ determines the degree of substitutability among different types of labor. Firms also decide about the degree of capacity utilization ($UCAP_t^j$). There is an economy-wide technology shock U_t^Y following a random-walk process plus drift. The objective of the firm is to maximize profits Pr,

$$\begin{aligned} \mathrm{Pr}_t^j = {} & p_t^j Y_t^j - w_t L_t^j - i_t^K p_t^{K,j} K_t^j - (adj^P(P_t^j) + adj^L(L_t^j) \\ & + adj^{UCAP}(UCAP_t^j)), \end{aligned} \tag{2}$$

where i^K denotes the rental rate of capital. Firms also face technological and regulatory constraints which restrict their price-setting,

[5]Lowercase letters denote logarithms, i.e., $z_t = log(Z_t)$. Lowercase letters are also used for ratios and rates. In particular, we define $p_t^j = P_t^j / P_t^Y$ as the relative price of good j with regard to the GDP deflator.

employment, and capacity-utilization decisions. Price-setting rigidities can be the result of the internal organization of the firm or specific customer-firm relationships associated with certain market structures. Costs of adjusting labor have a strong job-specific component (e.g., training costs), but higher employment adjustment costs may also arise in heavily regulated labor markets with search frictions. Costs associated with the utilization of capital can result from higher maintenance costs associated with a more intensive use of a piece of capital equipment. The following convex functional forms are chosen:

$$\begin{aligned} adj^L(L_t^j) &= w_t\left(L_t^j u_t^L + \frac{\gamma_L}{2}\Delta L_t^{j2}\right) \\ adj^P(P_t^j) &= \frac{\gamma_P}{2}\left(\frac{P_t^j}{P_{t-1}^j} - 1\right)^2 Y_t \\ adj^{UCAP}(ucap_t^j) &= p_t^I K_t\left(\gamma_{ucap,1}(ucap_t^j - 1) + \frac{\gamma_{ucap,2}}{2}(ucap_t^j - 1)^2\right). \end{aligned} \tag{3}$$

The firm determines labor input, capital services, and prices optimally in each period given the technological and administrative constraints as well as demand conditions.

3.1.2 Residential Construction

Monopolistically competitive firms h in the residential construction sector use new land (J_t^{Land}) sold by (Ricardian) households and final goods (J_t^{Constr}) to produce new houses using a CES technology

$$J_t^H = \left(s_L^{\frac{1}{\sigma_L}} J_t^{Land\frac{(\sigma_L-1)}{\sigma_L}} + (1-s_L)^{\frac{1}{\sigma_L}} J_t^{Constr\frac{(\sigma_L-1)}{\sigma_L}}\right)^{\frac{\sigma_L}{\sigma_L-1}} \tag{4}$$

subject to a quadratic adjustment cost constraint

$$adj^{PH}(P_t^H) = \frac{\gamma_P}{2}\left(\frac{P_t^H}{P_{t-1}^H} - 1\right)^2 Y_t. \tag{5}$$

New and existing houses are perfect substitutes. Thus households can make capital gains or suffer capital losses, depending on house price fluctuations.

3.1.3 Investment Goods Producers

There is a perfectly competitive investment goods production sector which combines domestic and foreign final goods, using the same CES aggregators as households and governments do to produce investment goods for the domestic economy. Denote the CES aggregate of domestic and foreign inputs used by the investment goods sector by J_t^{inp}; then real output of the investment goods sector is produced by the following linear production function:

$$J_t = J_t^{inp} U_t^{PI}, \tag{6}$$

where U_t^{PI} is a technology shock to the investment goods production technology which itself follows a random walk,[6]

$$u_t^{PI} = u_{t-1}^{PI} + \varepsilon_t^{UPI}. \tag{7}$$

3.2 Households

The household sector consists of a continuum of households $h \in [0,1]$. A fraction s^r of all households are Ricardian and indexed by r, and s^c households are credit constrained and indexed by c. The period utility function is identical for each household type and specified as a nested CES aggregate of consumption (C_t^h) and housing services (H_t^h) and separable in leisure $(1 - L_t^h)$. We also allow for habit persistence in consumption. Thus temporal utility for consumption is given by

$$\begin{aligned} &U(C_t^h, H_t^h, 1 - L_t^h) \\ &\quad = \log\left\{\left[s_C^{\frac{1}{\sigma^H}}\left(C_t^h - hC_{t-1}^h\right)^{\frac{\sigma^H - 1}{\sigma^H}} + s_H^{\frac{1}{\sigma^H}} H_t^{h\frac{\sigma^H - 1}{\sigma^H}}\right]^{\frac{\sigma^H}{\sigma^H - 1}}\right\} \\ &\qquad + \exp(u_t^L)\vartheta\left(1 - L_t^h\right)^{1-\kappa}. \end{aligned} \tag{8}$$

Both types of households supply differentiated labor services to unions which maximize a joint utility function for each type of

[6]This shock is introduced to capture a divergent trend in relative investment prices.

labor i. It is assumed that types of labor are distributed equally over the two household types. Nominal rigidity in wage setting is introduced by assuming that the household faces adjustment costs for changing wages. These adjustment costs are borne by the household.

3.2.1 Ricardian Households

Ricardian households have full access to financial markets. They hold domestic government bonds ($B_t^{G,r}$) and bonds issued by other domestic and foreign households ($B_t^r, B_t^{F,r}$), real capitals (K_t^r) used in the final goods production sector, as well as the stock of land ($Land_t$) which is still available for building new houses. In addition, they hold a stock of deposits (D) with a financial intermediary who provides loans to credit-constrained households. The household receives income from labor, financial assets, rental income from lending capital to firms, selling land to the residential construction sector, plus profit income from firms owned by the household (final goods Pr_t^j, residential construction Pr_t^H, and financial intermediaries Pr_t^B). We assume that all domestic firms are owned by Ricardian households. Income from labor is taxed at rate t^w, consumption at rate t^c. In addition, households pay lump-sum taxes T^{LS}. We assume that income from financial wealth is subject to different types of risk. Domestic bonds and interest income from deposits yield risk-free real return equal to r_t. Domestic and foreign bonds are subject to (stochastic) risk premia linked to net foreign indebtedness. An equity premium on real assets arises because of uncertainty about the future value of real assets. Furthermore, the discount factor β^r is subject to random shocks.

The Lagrangian of this maximization problem is given by

$$MaxV_0^r = \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^{r^t} U(C_t^r, 1 - L_t^r, H_t^r) - \mathrm{E}_0 \sum_{t=0}^{\infty} \lambda_t^r \beta^{r^t}$$
$$\times \begin{pmatrix} (1 + t_t^c) p_t^C C_t^r + p_t^I I_t^r + p_t^H (1 + t_t^c) I_t^{H,r} + (B_t^{G,r} + B_t^r + D_t) \\ \quad + rer_t B_t^{F,r} - (1 + r_{t-1})(B_{t-1}^{G,r} + B_{t-1}^r + D_{t-1}) \\ \quad - (1 + r_{t-1}^F) rer_t B_{t-1}^{F,r} - ((1 - t^K) i_{t-1}^K + t^K \delta^K) p_{t-1}^I K_{t-1}^r \\ \quad - (1 - t_t^W) w_t L_t^r + adj^W(W_t) - p_t^L J_t^{Land} - \sum_{j=1}^{\mathrm{Pr}_t^j} - \mathrm{Pr}_t^H \\ \quad - \mathrm{Pr}_t^B + T_t^{LS,r} \end{pmatrix}$$

$$- \mathrm{E}_0 \sum_{t=0}^{\infty} q_t^{r,K} \beta^{r^t} (K_t^r - J_t^r - (1-\delta^K) K_{t-1}^r)$$
$$- \mathrm{E}_0 \sum_{t=0}^{\infty} q_t^{r,H} \beta^{r^t} \left(H_t^r - J_t^{H,r} - (1-\delta^H) H_{t-1}^{H,r} \right)$$
$$- \mathrm{E}_0 \sum_{t=0}^{\infty} q_t^{r,L} \beta^{r^t} \left(Land_t + J_t^{Land} - (1+g_t^L) Land_{t-1} \right). \tag{9}$$

The investment decisions with regard to physical capital and housing are subject to convex adjustment costs; therefore, we make a distinction between real investment expenditure ($I_t^r, I_t^{H,r}$) and physical investment ($J_t^r, J_t^{H,r}$). Investment expenditure of households including adjustment costs is given by

$$I_t^r = J_t^r \left(1 + \frac{\gamma_K}{2} \left(\frac{J_t^r}{K_t^r} \right) \right) + \frac{\gamma_I}{2} (\Delta J_t^r)^2 \tag{10a}$$
$$I_t^{H,r} = J_t^{H,r} \left(1 + \frac{\gamma_H}{2} \left(\frac{J_t^{H,r}}{H_t^r} \right) \right) + \frac{\gamma_{I^H}}{2} (\Delta J_t^{H,r})^2. \tag{10b}$$

The budget constraint is written in real terms, with all prices expressed relative to the GDP deflator (P). Investment is a composite of domestic and foreign goods.

We follow Bernanke and Gertler (1999) and assume that residential and non-residential investment decisions are subject to fundamental and non-fundamental shocks. From the first-order conditions of the household investment problem, we obtain the shadow price of non-residential and residential capital:

$$q_t^{r,K} = \beta E_t \left[q_{t+1}^{r,K} (1-\sigma^K) + \left((1-t^K) i_t^K + t^K \sigma^K \right) p_t^I \right] \tag{11a}$$
$$q_t^{r,H} = U_{H,t}^r + \beta E_t \left(q_{t+1}^{r,H} (1-\sigma^H) \right), \tag{11b}$$

which is equal to the present discounted value of fundamental shocks, namely the after-tax rental rate of capital and the marginal utility of housing services, respectively. As in Bernanke and Gertler (1999), we assume that in addition there are non-fundamental shocks x_t^i with $i \in (K, H)$, which follow a "near-rational bubble" process $x_{t+1}^i = \frac{a^i}{\beta} x_t^i + e_t^i$ with $a^i < \beta$.

We then define the modified shadow price $Q_t^{r,i} = q_t^{r,i} + x_t^i$, which follows the process

$$(1 - z_t^K)Q_t^{r,K} = \beta E_t\Big(Q_{t+1}^{r,K}(1 - \delta^K) + \big((1 - t^K)i_t^K + t^K\delta^K\big)\, p_t^I\Big) \tag{12a}$$

$$(1 - z_t^H)Q_t^{r,H} = U_{H,t}^r + \beta E_t\Big(Q_{t+1}^{r,H}(1 - \delta^H)\Big), \tag{12b}$$

where $z_t^i = (1 - a^i)\frac{x_t^i}{\beta}$.

Like Bernanke and Gertler, we use the term "bubble" loosely to denote temporary deviations of asset prices from fundamentals due to waves of optimism and excessive risk taking in periods of rising asset prices and waves of pessimism or panics in periods of increased uncertainty.

In the context of the current crisis, alternative explanations could be given for a sudden fall in asset prices. For example, an increase in z_t^i could capture what Gorton (2010) calls a "panic," to describe the uncertainty about the value of certain asset classes which have forced banks to deliver and dump assets, leading to falling asset values. A rising z_t^i could also capture what Hall (2010) refers to as "principal agent frictions," which he models by introducing an exogenous wedge shock between safe assets (government bonds) and risky assets (equity and houses) in order to empirically match rising spreads between safe and risky assets.

The interest rate that households face when making consumption and investment decisions depends on the aggregate level of foreign indebtedness (defined as $(-B_t^{F,r})/(p_tY_t)$):

$$i_t^h = i_t + rprem\left(\frac{(-B_t^F)}{P_tY_t}\right). \tag{13}$$

This specification corresponds to the debt-elastic interest rate premium in the comparison of methods studied by Schmitt-Grohe and Uribe (2003) in closing small open-economy models. The major reason for this specification is that it induces stationarity. However, we also regard the interest elasticity with regard to foreign debt as an important behavioral parameter describing the risk tolerance of foreign creditors. The parameter *rprem*, together with the rate of time preference of Ricardian households, determines the steady-state debt level of the economy.

3.2.2 Credit-Constrained Households

Credit-constrained households differ from Ricardian households in two respects. First, they have a higher rate of time preference ($\beta^c < \beta^r$) and, second, they face a collateral constraint on their borrowing. They borrow B_t^c exclusively from domestic Ricardian households. The Lagrangian of this maximization problem is given by

$$\begin{aligned} MaxV_0^c = {} & \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^{c^t} U(C_t^c, 1 - L_t^c, H_t^c) \\ & - \mathrm{E}_0 \sum_{t=0}^{\infty} \lambda_t^c \beta^{c^t} \Big((1 + t_t^c) p_t^C C_t^c + p_t^H (1 + t_t^H) I_t^{H,c} - B_t^c \\ & + (1 + r_{t-1}) B_t^c) - (1 - t_t^W) w_t L_t^c + adj^W (W_t) + T_t^{LS,c} \Big) \\ & - \mathrm{E}_0 \sum_{t=0}^{\infty} \lambda_t^c \zeta_t^c \beta^{c^t} \Big(H_t^c - J_t^{H,c} - (1 - \delta^H) H_{t-1}^c \Big) \\ & - \mathrm{E}_0 \sum_{t=0}^{\infty} \lambda_t^c \psi_t^c \beta^{c^t} \big((1 + r_t) B_t^c - \chi_t^c p_t^H H_{t-1}^c \big) . \end{aligned} \tag{14}$$

Notice that the collateral constraint increases the shadow price of borrowing as determined by the Lagrange multiplier ψ_t^c of the collateral constraint.

There is a non-fundamental shock to housing investment which is constrained to be equal across household types.

3.2.3 Wage Setting

A trade union is maximizing a joint utility function for each type of labor i where it is assumed that types of labor are distributed equally over constrained and unconstrained households with their respective population weights. The trade union sets wages by maximizing a weighted average of the utility functions of these households. The wage rule is obtained by equating a weighted average of the marginal utility of leisure to a weighted average of the marginal utility of consumption times the real wage of these two household types, adjusted for a wage markup,

$$\frac{s^c U^c_{1-L,t} + s^r U^r_{1-L,t}}{s^c U^c_{c,t} + s^r U^r_{c,t}} = \frac{(1-t^W_t)}{(1+t^C_t)} \frac{W_t}{P^C_t} \eta^W_t, \tag{15}$$

where η^W_t is the wage markup factor, with wage markups fluctuating around $1/\theta$, which is the inverse of the elasticity of substitution between different varieties of labor services. The trade union sets the consumption wage as a markup over the reservation wage. The reservation wage is the ratio of the marginal utility of leisure to the marginal utility of consumption. This is a natural measure of the reservation wage. If this ratio is equal to the consumption wage, the household is indifferent between supplying an additional unit of labor and spending the additional income on consumption and not increasing labor supply. Fluctuation in the wage markup arises because of wage adjustment costs

$$adj^W(W_t) = \frac{\gamma_W}{2} \left(\frac{1}{\Pi^{1-sfw}_{t-1} \bar{\Pi}^{sfw}} \frac{W_t}{W_{t-1}} - 1 \right)^2 Y_t. \tag{16}$$

3.3 Trade and the Current Account

In order to facilitate aggregation, we assume that households, the government, and the corporate sector have identical preferences across goods used for private consumption, public expenditure, and investment. Let $Z^i \in \left\{C^i, I^i, C^{G,i}, I^{G,i}\right\}$ be demand of an individual household, investor, or the government, and then their preferences are given by the following utility function:

$$Z^i = \left[(1 - s^M - u^M_t)^{\frac{1}{\sigma^M}} Z^{d^i \frac{\sigma^M - 1}{\sigma^M}} + (s^M + u^M_t)^{\frac{1}{\sigma^M}} Z^{f^i \frac{\sigma^M - 1}{\sigma^M}} \right]^{\frac{\sigma^M}{(\sigma^M - 1)}}, \tag{17}$$

where the share parameter s^M can be subject to random shocks u^M, and Z^{d^i} and Z^{f^i} are indexes of demand across the continuum of differentiated goods produced, respectively, in the domestic economy and abroad.

Exporters buy final domestic goods X_t and transform them into exportables using a linear technology. Exporters act as monopolistic competitors in export markets and charge a markup over domestic prices. Thus export prices are given by

$$\eta_t^X P_t^X = P_t. \tag{18}$$

Importers buy foreign goods at quantity M_t from foreign exporters and sell them on the domestic market. Importers are monopolistic competitors on the market for imported goods and charge a markup over the purchase price of imports denominated in domestic currency.

$$\eta_t^M P_t^M = E_t P_t^F \tag{19}$$

Markup fluctuations arise because of price adjustment costs. Exports and imports, together with interest receipts/payments, and the exogenous balance of primary incomes and transfers determine the evolution of net foreign assets denominated in domestic currency.

$$B_t^F = (1 + i_t^F) B_{t-1}^F + P_t^X X_t - P_t^M M_t + \varepsilon_t^{B^F} \tag{20}$$

3.4 Policy

Both government expenditure and receipts are responding to business-cycle conditions. On the expenditure side, we identify the systematic response of government consumption, government transfers, and government investment to the annual GDP growth rate. In addition, all three expenditure components are used for stabilizing the debt-to-GDP ratio, where b^T is the government debt target and def^T is the associated deficit target. For government consumption and government investment, we specify the following rules for detrended c^G and i^G (removing trend productivity growth):

$$c_t^G - \overline{c^G} = \tau_{Lag}^{CG}(c_{t-1}^G - \overline{c^G}) + \tau^{CG}\left(\sum_{i=1}^{4} \Delta y_{t-i} - 4\overline{\Delta y}\right)$$

$$
\begin{aligned}
&- \tau^{CGB} \left(\frac{B_{t-1}}{Y_{t-1} P_{t-1}} - b^T \right) \\
&- \tau^{CGDEF} \left(\left(\frac{\Delta B_{t-1}}{Y_{t-1} P_{t-1}} \right) - def^T \right) + u_t^{CG} \qquad (21)
\end{aligned}
$$

$$
\begin{aligned}
i_t^G - \overline{i^G} = {} & \tau_{Lag}^{IG} (i_{t-1}^G - \overline{i^G}) + \tau^{IG} \left(\sum_{i=1}^{4} \Delta y_{t-i} - 4\overline{\Delta y} \right) \\
&- \tau^{IGB} \left(\frac{B_{t-1}}{Y_{t-1} P_{t-1}} - b^T \right) \\
&- \tau^{IGDEF} \left(\left(\frac{\Delta B_{t-1}}{Y_{t-1} P_{t-1}} \right) - def^T \right) + u_t^{IG}. \qquad (22)
\end{aligned}
$$

Government consumption and government investment can temporarily deviate from their long-run targets $\overline{c^G}$ and $\overline{i^G}$ in response to fluctuations in growth rates. In addition, government expenditure is used for stabilizing the debt-to-GDP ratio, where b^T is the government debt target and def^T is the associated deficit target. The shocks u^{CG} and u^{IG} are white noise.

The transfer system consists of two parts: unemployment benefits *UBEN* and other transfers *TR*. The former provides income for the unemployed $(POP_t^W - POP_t^{NPART} - L_t)$. Other transfers *TR* consists of transfers to pensioners POP_t^P and other transfer payments, and is used for stabilizing the debt-to-GDP ratio. We assume that unemployment benefits and pensions are indexed to wages with replacement rates b^U and b^R, respectively.

$$
\begin{aligned}
tr_t = {} & b^U w_t (POP_t^W - POP_t^{NPART} - L_t) + b^R w_t POP_t^P \\
&- \tau^{TRB} \left(\frac{B_{t-1}}{Y_{t-1} P_{t-1}} - b^T \right) \\
&- \tau^{TRDEF} \left(\left(\frac{\Delta B_{t-1}}{Y_{t-1} P_{t-1}} \right) - def^T \right) + u_t^{TR}, \qquad (23)
\end{aligned}
$$

where u^{TR} is an autocorrelated shock. Government revenues R_t^G consist of taxes on consumption as well as capital and labor income.

$$
\begin{aligned}
R_t^G = {} & (ssc_t + t_t^w) W_t L_t + t_t^c P_t^c C_t + t_t^c P_t^H I_t^H \\
&+ t_t^K [(Y_t - W_t L_t) - \delta K_t P_t^I] \qquad (24)
\end{aligned}
$$

We assume consumption and capital income tax follow a linear scheme but a progressive labor income tax schedule

$$t_t^w = \tau_0^w Y_t^{\tau_1^w}, \tag{25a}$$

where τ_0^w measures the average tax rate and τ_1^w measures the degree of progressivity. A simple first-order Taylor expansion around a steady-state growth rate yields

$$t_t^w = \tau_0^w + \tau_0^w \tau_1^w \left(\sum_{i=0}^{3} \Delta y_{t-i} - 4\overline{\Delta y} \right)_t . \tag{25b}$$

Government debt (B_t) evolves according to

$$B_t = (1 + i_t^B) B_{t-1} + P_t^C C_t^G + P_t^C I_t^G + TR_t - R_t^G - T_t^{LS}, \tag{26}$$

where i_t^B is the implicit interest rate the government pays on its debt, which depends on the average maturity structure of sovereign debt $(1/(1-\rho^B))$ and the policy rate augmented by a markup made up of a sovereign risk premium, which is assumed to depend on the government-debt-to-GDP ratio and an autoregressive term.

$$i_t^B = \rho^B i_{t-1}^B + (1 - \rho^B) \left[i_t + mup^B + rprem^B (B_t/Y_t - \overline{B/Y}) + \varepsilon_t^{rpb} \right] \tag{26$'$}$$

Monetary policy is modeled exogenous, with interest rates i_t^{EA} set by the European Central Bank (ECB).

$$i_t = i_t^{EA} + u_t^{RPREM} \tag{27}$$

In the years prior to EMU, the differential between the policy rate in Spain and the (synthetic) euro-area average was gradually eliminated.

Finally, we define a monetary policy as the deviation of i_t^{EA} from a synthetic interest rate determined by a Taylor rule for Spain that responds to consumer price inflation and the annual growth rate of output, with weights based on estimates for the euro area (Ratto, Roeger, and in 't Veld 2009):

$$z_t^M = i_t^{EA} - \left[\begin{array}{c} \tau_{lag}^M i_{t-1}^{EA} + (1 - \tau_{lag}^M)[r^{EQ} + \pi_t^T + \tau_\pi^M (\pi_t^C - \pi_t^T) \\ + \tau_y^M (gy_t + gy_{t-1} + gy_{t-2} + gy_{t-3} - 4\overline{gy})/4] \end{array} \right]. \tag{28}$$

3.5 Equilibrium

Equilibrium in our model economy is an allocation, a price system, and monetary and fiscal policies such that both non-constrained and constrained households maximize utility; final-goods-producing firms, firms in the construction sector, and investment goods producers maximize profits; and the following market clearing conditions hold for final domestic goods:

$$Y_t = C_t^d + J_t^{inp,d} + J_t^{Constr} + C_t^{G,d} + I_t^{G,d} + X_t, \tag{29}$$

and final imported goods

$$M_t = C_t^f + J_t^{inp,f} + C_t^{G,f} + I_t^{G,f}, \tag{30}$$

where total domestic and imported consumption C_t^i is the sum of savers' and borrowers' consumption, with their per capita consumption multiplied by the respective population shares s^r and s^c:

$$C_t^i = s^r C_t^{r,i} + s^c C_t^{c,i}, \text{ with } i = d, f. \tag{31a}$$

Similarly, total housing investment is defined as

$$J_t^H = s^r J_t^{H,r} + s^c J_t^{H,c} \tag{31b}$$

and equilibrium in the labor market is given by

$$L_t = s^r L_t^r + s^c L_t^c \text{ with } L_t^r = L_t^c. \tag{31c}$$

Credit-constrained households only engage in debt contracts with Ricardian households, i.e.,

$$B_t^c = \frac{s^r}{s^c} B_t^r. \tag{32}$$

4. Model Estimation

The model is estimated on quarterly data over the period 1995:Q1 to 2012:Q4, using Bayesian inference methods to estimate model

Table 1. Calibrated Structural Parameters

Structural Parameters	Calibrated Value
χ^c	0.5
α	0.60
δ	0.025
δ^{house}	0.01
Target Government Debt to GDP	0.6
Target Government Deficit to GDP	−0.025
ρ^B	0.95
ssc	0.14
τ_0^W	0.13
τ_1^W	0.8
t^K	0.29
t^C	0.15
$1/\beta^c - 1$	0.04
$1/\beta^{r,EUR} - 1$	0.005
$(1 - s^M)$	0.75
b^R	0.22
τ_{lag}^M	0.9
τ_π^M	1.5
τ_y^M	0.4
mup^B	0.004

parameters and shocks.[7] Concerning the steady-state calibration, parameters shown in table 1 have been calibrated to match ratios of main economic aggregates (corporate investment, construction investment, and government consumption and investment) to GDP over the period 1995–99.[8] The two general exceptions to this are the calibration of the labor market and the steady-state debt ratios. For the former, the 1995–99 averages can no longer be considered

[7] We use the Dynare toolbox for MATLAB (Adjemian et al. 2011) to conduct the first-order approximation of the model, to the calibrated steady state and to perform the estimation. We run four Metropolis-Hastings chains of 100,000 draws to estimate the posterior distribution. A detailed description of data sources and estimation is described in appendices 1 and 2.

[8] This period was chosen to exclude the more turbulent 2000s, in which several variables (for example, construction investment share in GDP) could have diverged from their historical level.

representative of the Spanish labor market, as labor market reforms have led to a regime shift in the Spanish economy. Instead we base the calibration of labor market parameters on the full sample 1995:Q1–2012:Q4. The steady-state employment rate as share of total population is set to 0.41, the wage share to 57 percent. Concerning the government debt ratio, we impose the debt target of 60 percent of GDP, which is close to the sample average. This target implies, given the nominal growth rate in the steady state, a deficit target of 2.5 percent of GDP. The average maturity structure of sovereign debt is set at five years. Tax rates are calibrated on sample averages. Government transfers to households are set to 12.9 percent of GDP, benefit indexation (b^U) is estimated, and pension indexation (b^R) is set to match this steady-state calibration. For construction of the monetary policy shock, Taylor-rule coefficients are imposed based on estimates for the euro area (Ratto, Roeger, and in 't Veld 2009). The constant in the markup of sovereign bond interest rates (mup^B) is set at 0.4 percent quarterly. Based on the whole sample, the quarterly GDP trend growth rate was set to 0.55 percent, while the inflation trend growth rate is set to 0.5 percent. Credit-constrained households are calibrated with a high rate of time preference, 4 percent quarterly, while the discount rate for non-constrained households is estimated (see below). The euro-area discount rate is set at 0.5 percent quarterly; openness is calibrated at 0.25.

The estimation results of the main structural parameters are summarized in table 2.[9] The population share of Ricardian households s^r is estimated at 0.49, implying the share of credit-constrained households s^c of 0.51. Concerning consumption, the intertemporal elasticity of substitution is set to 1, habit persistence h is estimated to be 0.75, and the substitution elasticity for housing services σ_H is estimated at 0.62. The discount factor for Spanish households β^r is estimated close to 0.992, reflecting a higher propensity to consume than for the rest of the euro area. The estimate for *rprem* implies a highly inelastic interest rate with regard to external indebtedness,

[9] "HPD Inf." and "HPD Sup." denote the bounds of the 90 percent highest probability density interval. The prior distributions used and posterior estimates of all parameters can be found in a supplementary appendix on the *IJCB* website (http://www.ijcb.org).

Table 2. Estimation Results for Main Structural Parameters

	Prior Distribution	Prior Mean	Prior S.D.	Posterior Mean	Posterior S.D.	HPD Inf.	HPD Sup.
h	Beta	0.700	0.1000	0.751	0.0519	0.6749	0.8400
κ	Gamma	1.000	0.4000	0.892	0.3231	0.3641	1.3872
σ^H	Gamma	0.500	0.1000	0.623	0.1252	0.4196	0.8192
σ^X	Gamma	1.250	0.5000	1.864	0.3286	1.3198	2.3870
σ^M	Gamma	1.250	0.5000	1.153	0.2458	0.7359	1.5503
σ^L	Beta	0.500	0.2000	0.319	0.1272	0.1006	0.5089
s^s	Beta	0.500	0.1500	0.486	0.0971	0.3340	0.6490
$rprem$	Beta	0.0025	0.0010	0.00098	0.00029	0.00052	0.00145
$1/\beta^r - 1$	Beta	0.007	0.0010	0.008	0.0007	0.0067	0.0091
τ_{lag}^{CG}	Beta	0.500	0.2000	0.933	0.0135	0.9097	0.9549
τ_1^{CG}	Beta	−0.100	0.0400	−0.027	0.0124	−0.0461	−0.0071
τ^B	Beta	0.020	0.0100	0.021	0.0034	0.0157	0.0266
τ^{DEF}	Beta	0.020	0.0100	0.012	0.0055	0.0029	0.0197
τ_{lag}^{IG}	Beta	0.500	0.2000	0.899	0.0666	0.8056	0.9974
τ_1^{IG}	Beta	−0.100	0.0400	−0.037	0.0165	−0.0613	−0.0094
$\tau^{B,IG}$	Beta	0.020	0.0100	0.010	0.0032	0.0044	0.0147
$\tau^{DEF,IG}$	Beta	0.020	0.0100	0.024	0.0087	0.0097	0.0375
$rprem^B$	Beta	0.003	0.0012	0.0031	0.0012	0.0012	0.0051
b^U	Beta	0.300	0.1000	0.466	0.0683	0.3618	0.5715

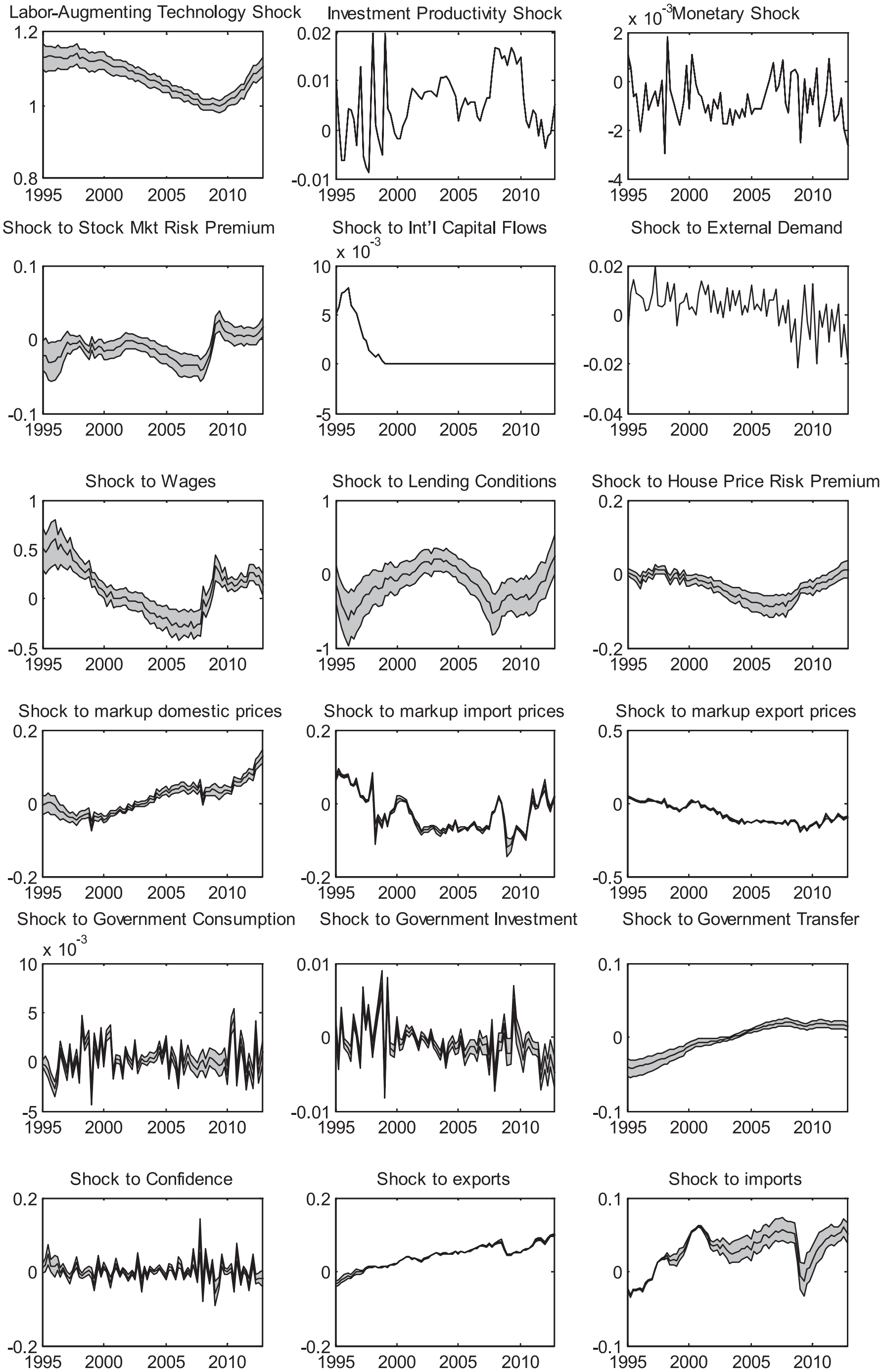
an increase in the risk premium of 4 basis points for every 10-percentage-point increase in net foreign liabilities. This low estimate reflects the persistent buildup in net foreign liabilities over the sample period, reaching almost 100 percent of GDP by the end of the sample period. Fiscal policy reactions are generally countercyclical, while government consumption also contains a debt- and deficit-stabilizing response. The estimated elasticity of the sovereign risk premium with regard to the government-debt-to-GDP ratio implies an increase in the risk premium of 12 basis points for a 10-percentage-point increase in the debt ratio.

4.1 Historical Evolution of the Main Shocks

We now turn to an evaluation of the main shocks identified in the estimation procedure.[10] Figure 2 shows the estimated historical evolution of these fundamental and non-fundamental shocks of the model over the period 1995:Q1–2012:Q4. Some of the shocks can be directly observed; other shocks can be identified as residuals to specific structural equations of the DSGE model. We capture *labor-augmenting technology shocks* to final goods by the terms U_t^Y in the production function and we specify it as a random-walk process. We identify *investment-specific technical change* U_t^{PI} by differences in growth rates between a weighted average of the GDP deflator and the import price deflator on the one hand and the investment deflator on the other. Labor-augmenting technological progress shows a strongly declining trend up to 2008 and then increases again in the recession. The evolution of investment-specific technological progress is less uniform over time and shows less variation, falling in 2000, further declining in 2005, and again experiencing a sharp decline in 2010. Shocks to *monetary policy* z_t^M are identified as stationary deviations of the ECB policy rate from the interest rate implied by a Taylor rule for Spain. The monetary policy shock shows a loose monetary stance for Spain over the whole sample, with the exception of a brief period in 2000 when there were some large increases in the ECB policy rate in quick succession, and again in 2007–8 when the ECB policy rate was raised while conditions in Spain deteriorated.

[10]In total, the model is estimated with twenty-nine shocks, but we focus here on the most important shocks over the estimation period.

Figure 2. Profile of Selected Shocks 1995–2012



Following the further reductions in the policy rate, the monetary policy shock turns negative again in the last observations.

By adding exogenous shocks to the discount factors of the various asset market arbitrage equations, we allow for non-fundamental shocks (bubbles) in the model. In particular, we identify *stock market bubbles* as a shock to the discount factor for corporate investment (equation (12a)), and a *house price bubble* as shocks to residential investment (equation (12b)). We find a declining trend in the stock market risk premium that could indicate a bubble building up between 2002 and 2008. This is followed by a sharp increase in the risk premium in 2009–10. The house price risk premium shows a stronger declining trend in the years up to 2008, with a gradual reversal after that, suggesting a bubble that built up between 1999 and 2008 and then slowly burst.

We identify the *shock to international capital flows* u_t^{RPREM} as the differential between nominal interest rate in Spain and the (synthetic) euro-area average rate pre-1999. The shock to international capital flows shows the elimination of the interest rate risk premium when Spain joined EMU in 1999 and which led to the inflow of cheap capital. The *shock to external demand* combines the shock to foreign demand and the shock to foreign prices and is predominantly positive over much of the period up to 2007 but largely negative in the years since then.

The *labor market shock* is identified as the residual to the wage-setting equation, and this shock to the wage markup shows a trend decline up to 2007–8, corresponding to a declining NAIRU over that period, and an increase in the years after 2008. Shifts in *lending conditions* are shocks to the collateral constraint of debtor households χ_t. The shock shows a loosening of lending conditions between 2003 and 2007 and a tightening since 2008.

The shock to the *markup* on domestic prices is increasing over the sample, partly reflecting a sectoral shift towards the non-tradable sector (services) which is generally characterized by higher markups. The shock to the *markup on import prices* shows a decline in 1999, while that on export prices shows a general declining trend over much of the sample. The shocks to the government spending categories represent residuals to estimated responses to the economic cycle. Hence, they cannot be interpreted as discretionary policy shocks, but should be seen as non-systematic innovations, i.e., deviations from "normal" cyclical responses. Government spending is estimated as countercyclical over the sample and with an active response to debt and deficit developments. The shock to government consumption u^{CG} indicates periods of fiscal stimulus and contraction relative to this estimated rule, with some evidence of a positive deviation in 2009–10. Government investment is also estimated to have been countercyclical over the sample, with also some evidence of a positive deviation in u^{IG} in 2009 but a negative deviation in the last two years (consolidation). The shock to government transfers u^{TR} shows a trend increase reflecting increasing generosity in the transfer system, which is reversed in 2011. The *shock to confidence* is the consumption preferences shock to the Euler equation, which suggests some decline in confidence in the crisis. The shocks to exports u^{X} and imports u^{M} represent shocks to trade shares and show trend increases over the sample, with noticeable declines in the crisis. In the following section we show the impact of these shocks on the main economic aggregates.

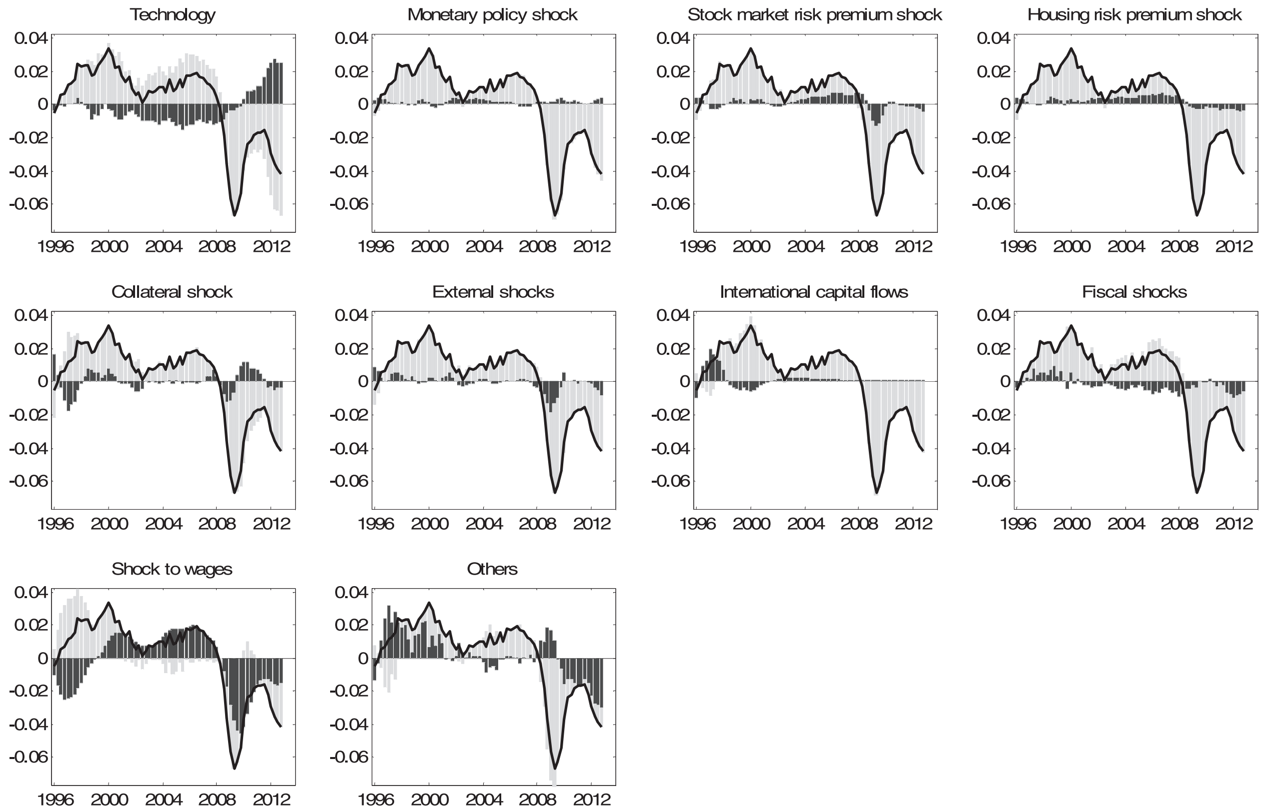
5. Historical Shock Decompositions

The estimated shocks can be used to provide a historical decomposition of the data, by decomposing the quarter-on-quarter or year-on-year growth rates or domestic demand shares (in deviations from their steady-state levels) into the different shocks (cf. Christiano, Motto, and Rostagno 2008). The following figures in this section do this for the profiles for GDP growth, respective domestic demand shares, and external position. To keep the analysis tractable, we only focus on the main shocks highlighted in the previous section, and group the contribution of all other shocks into a residual category labeled "others."[11]

Figure 3 shows the additive decomposition of year-on-year real GDP growth (in deviation from steady-state growth) into the different shocks. Real GDP growth was above trend in the early years

[11]In the figures, the category "others" combines all other shocks included in the model (e.g., shocks to markups) and includes the effects of the initial conditions. Detailed shock decompositions of other variables including employment, unit labor costs, and real exchange and interest rates are available in the supplementary appendix.

Figure 3. Shock Decomposition: GDP Growth



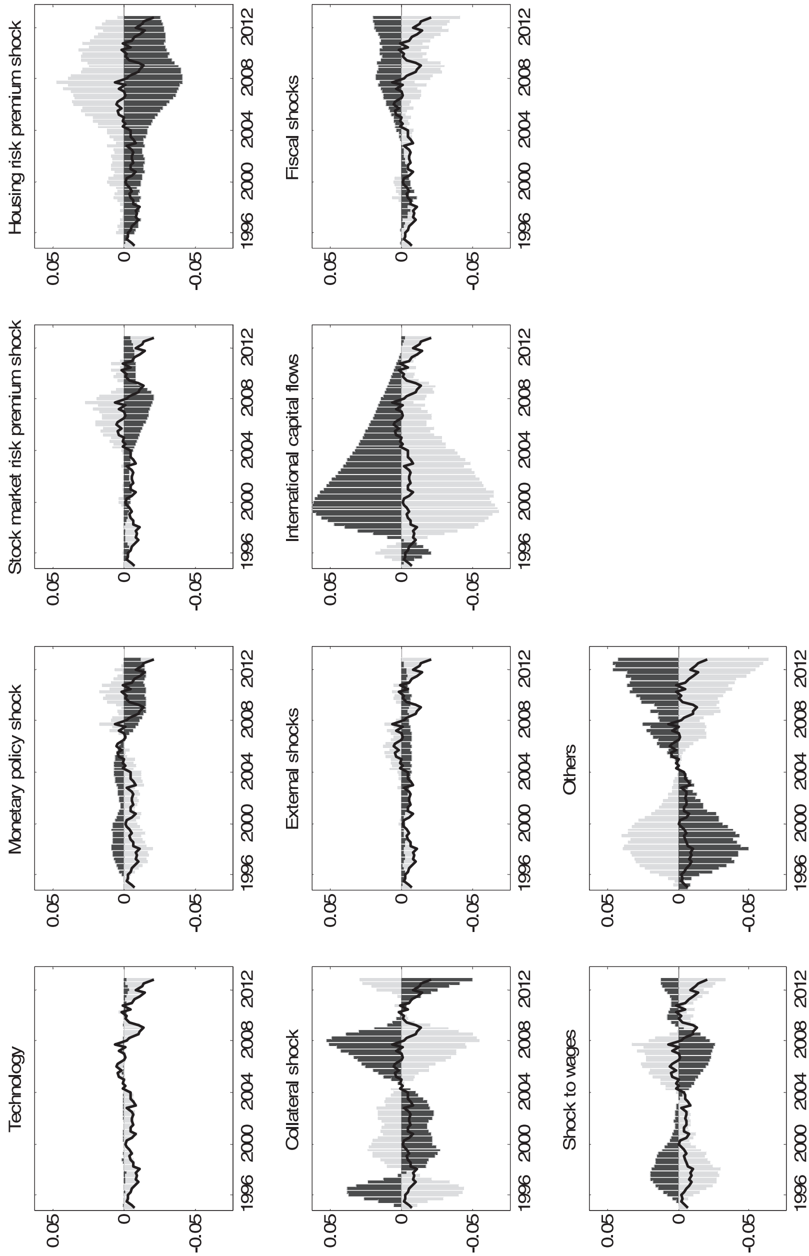
Notes: The following notes apply to figures 4–8 as well. The black lines show historical data (from which steady-state values have been subtracted). In each sub-plot, the vertical dark bars show the contribution of a different group of shocks to historical data, while stacked light bars show the contribution of all remaining shocks and initial conditions. Bars above the horizontal axis represent positive shock contributions, while bars below the horizontal axis represent negative contributions. The sum of shock contributions equals the historical data.

of the sample, the buildup and first years of EMU membership, and fell by a maximum of around 6 percent below trend in the trough of the financial crisis. In terms of shock contributions, the technology shock (the combination of the labor-augmenting productivity shock U_t^Y and the investment-specific productivity shock U_t^{PI}) contributed negatively to GDP growth throughout the decade but had a positive contribution in the last years of the sample. The (elimination of the) risk premium shock (international capital flows) u_t^{RPREM} boosted growth in the years prior to accession of EMU. For much of the decade, the labor market shock contributed positively to growth, but this turned around in the crisis starting in 2008. Stock market and housing risk premium shocks as well as looser lending conditions (collateral shock) χ_t contributed positively to growth rates, and these reversed in 2008 when they started to drag down growth.

The main driving factors of the collapse in real GDP growth in the crisis are the tightening in lending conditions (collateral shock), stock market crash (negative contribution of stock market risk premium), the bursting house price bubble (housing risk premium shock), the decline in world trade (negative contribution of external demand shock), and the reversal in the labor market shock (insufficient wage flexibility). Real GDP growth was supported in the crisis by a positive contribution from the productivity shocks (decline of the construction sector, a relatively low-tech sector, raised average productivity) and fiscal stimulus measures (2009–10). Fiscal consolidation in 2011–12 contributed negatively to growth.

Figure 4 shows the shock decomposition of the deviations in the consumption-to-output ratio relative to its steady-state level. Most striking is the large and persistent positive contribution of the financial capital flow shock. The disappearance of the interest rate risk premium upon accession to EMU led to an inflow of cheap capital and boosted consumption. Substitution between residential investment and consumption implies a negative contribution of the housing boom (housing risk premium). In addition, there is some negative effect from the stock market risk premium (substitution to investment). Looser lending conditions (collateral shock) boosted consumption between 2005 and 2008, but reversed into a negative contribution post-2008 (deleveraging). Other dampening effects on the consumption share came from the labor market shock (wage moderation). This turned positive in the crisis years and helped to

Figure 4. Shock Decomposition: Consumption-to-GDP Ratio



support consumption in the last years of the sample. Fiscal policy (transfers) boosted consumption spending in the years after 2004 and helped to support consumption in the crisis. The positive effect from low policy rates came to an end by 2007 and turned negative in 2008, as the policy rate did not respond sufficiently to the deterioration in conditions in Spain.

The shock decomposition of the deviations in the corporate investment share in GDP from its steady-state level is shown in figure 5. Like for consumption, there is a persistent positive contribution of the capital flow shock (lower interest rates). Non-fundamental factors play an important role in the shock decomposition of investment, with a large contribution from the residual term in the investment decision equation (stock market risk premium). Productivity shocks boosted the investment share, as did the loose monetary policy stance implied by the ECB policy rate for Spain. Lending conditions to households (collateral shock) relaxed between 2004 and 2008 and show as an increasing negative contribution to the investment share, indicating substitution effects towards consumption. But the sharp collapse in investment spending in the crisis is mainly linked to the shock to the stock market risk premium.

The boom-bust cycle in the residential investment share in GDP is also largely associated with what is in the model a non-fundamental shock. As figure 6 shows, the increase in residential investment up to 2007 and the subsequent collapse is primarily driven by the housing risk premium shocks. The interest rate risk premium shock (financial capital flows) also had a positive contribution to housing investment up to 2005, while looser lending conditions also added to the housing boom. Overall, the main driving factor behind the boom-bust cycle in residential investment appears to have been a bubble that, when it burst, led to a sharp reduction in construction investment.

We now turn to a shock decomposition of the external position. Figure 7 shows the trade balance as share of GDP. The trade balance has been in persistent deficit since 1998, widening to more than 6 percent of GDP in 2007. The crisis forced an adjustment, and the trade balance moved into surplus in 2012. The main driver for the trade deficit in the model is the financial capital flows shock. The elimination of the interest rate risk premium in EMU and interest rates determined by a common monetary policy in EMU boosted

Figure 5. Shock Decomposition: Corporate-Investment-to-GDP Ratio

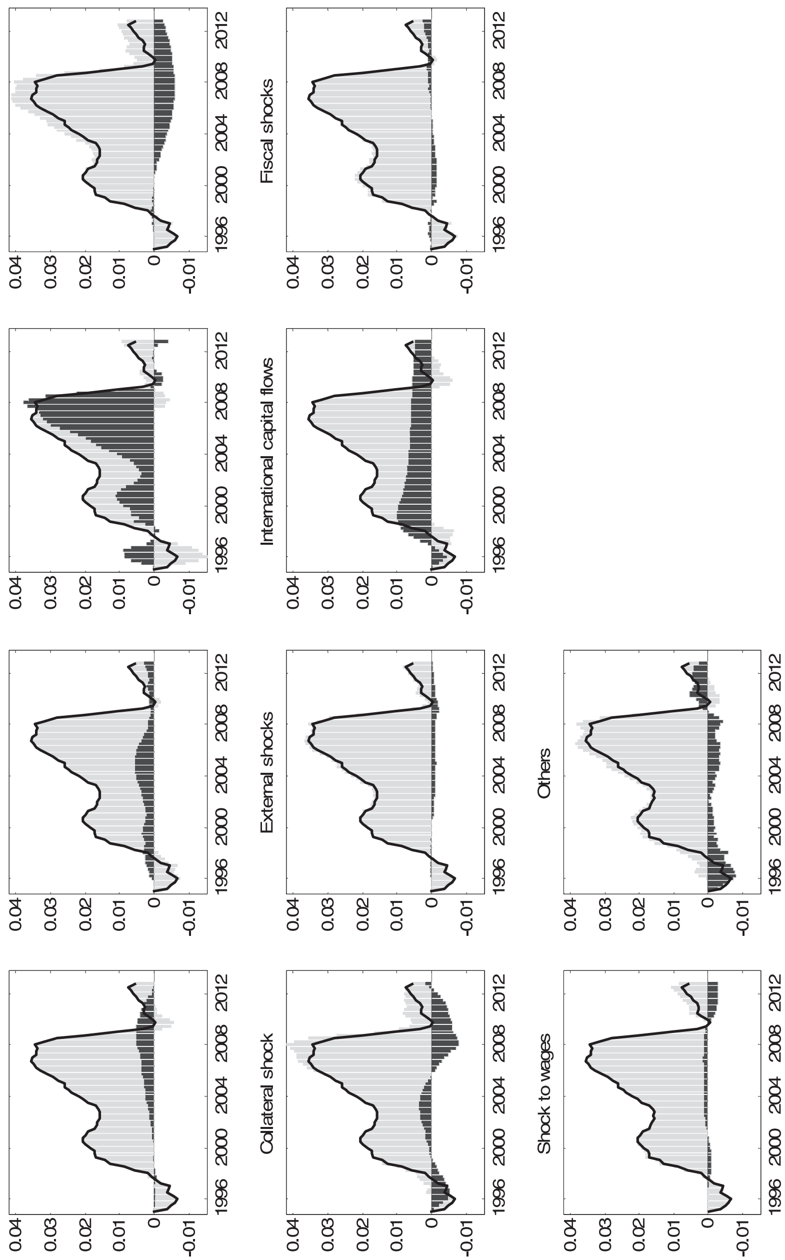


Figure 6. Shock Decomposition: Residential-Investment-to-GDP Ratio

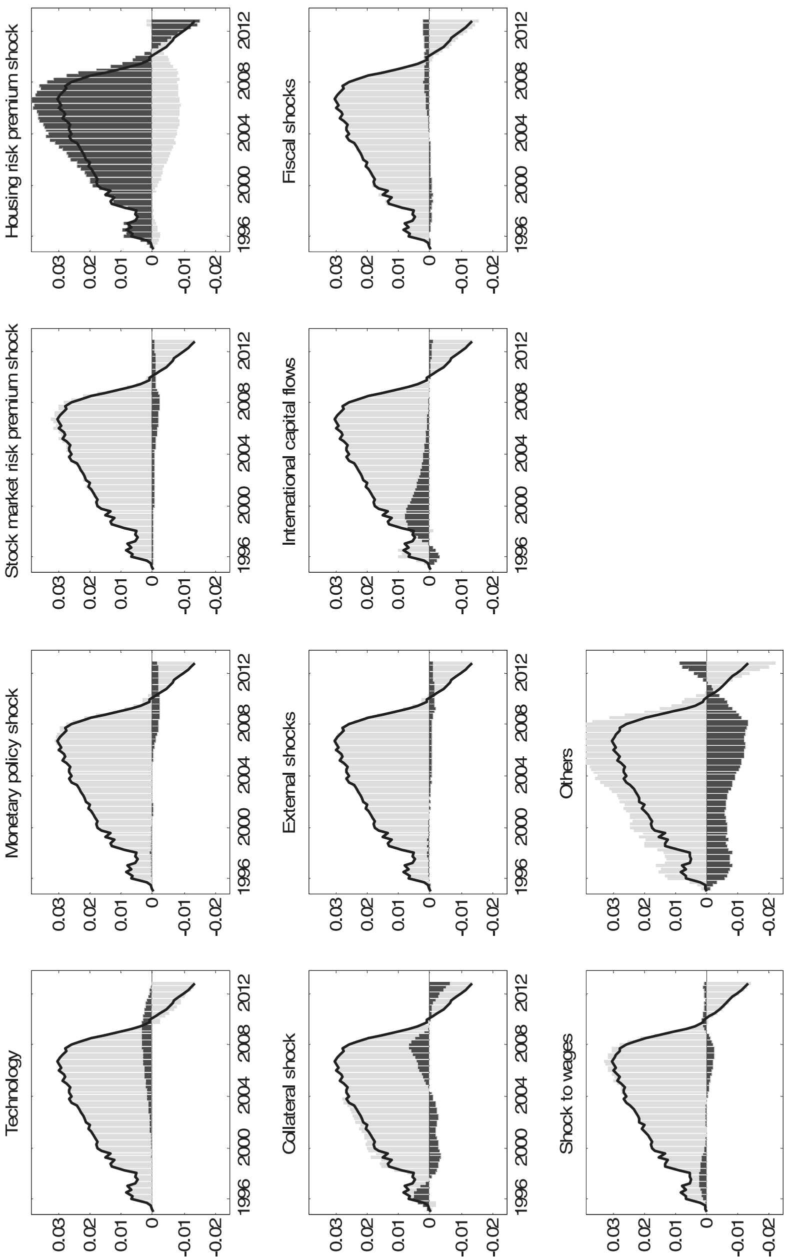
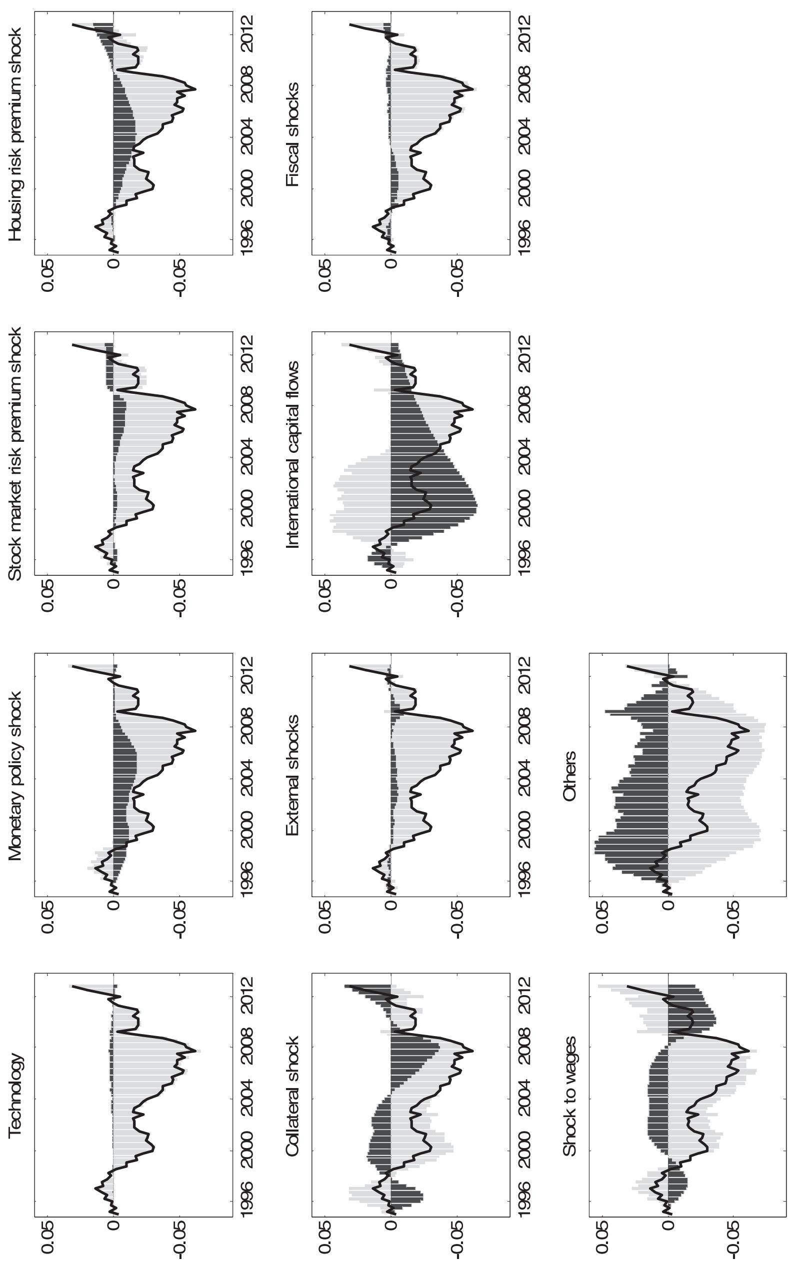


Figure 7. Shock Decomposition: Trade-Balance-to-GDP Ratio



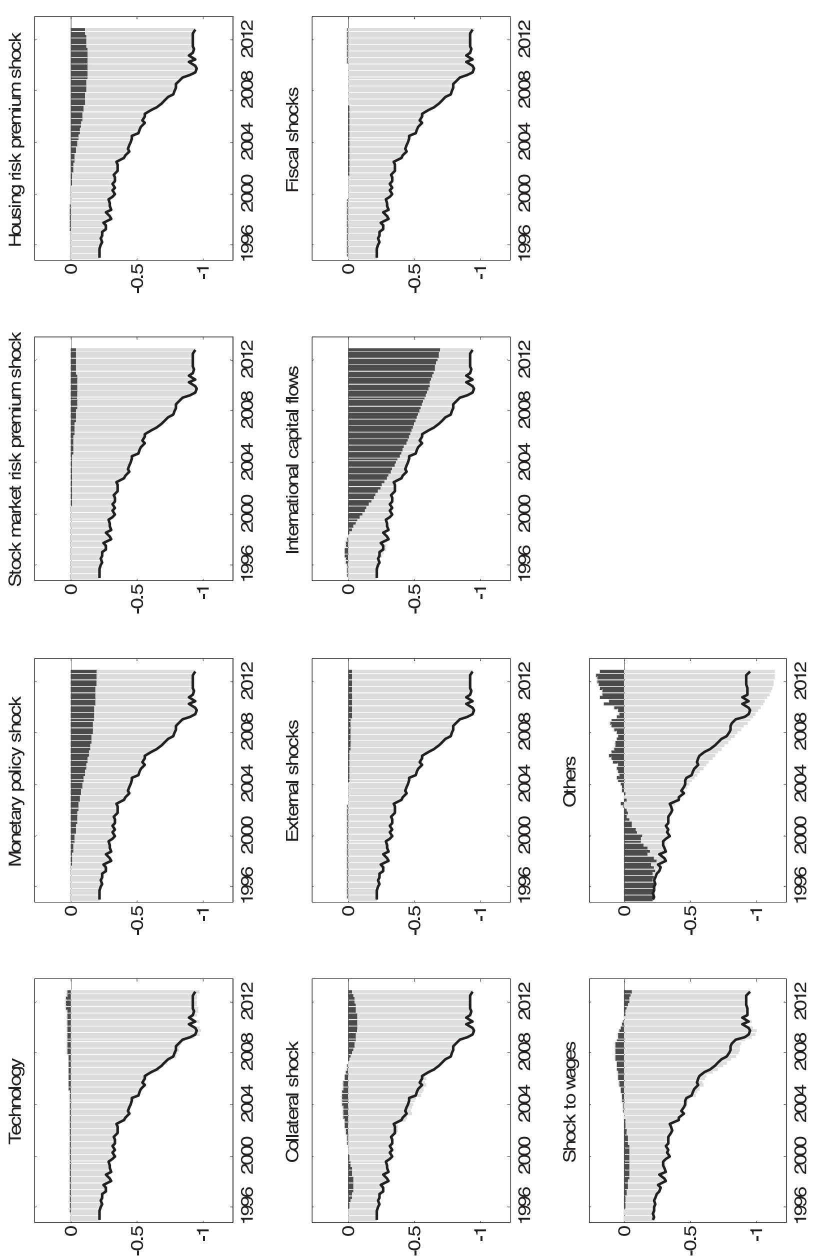
domestic demand and worsened the trade balance (this is the mirror image of the shock decompositions of consumption and investment). The stock market risk premium, the housing risk premium, and, since 2004, looser lending conditions (collateral) contributed further to a widening of the trade balance in the boom years. On the positive side, there was a positive contribution from the labor market shock up to 2007 (wage developments improving competitiveness and supporting exports), but this was reversed in the crisis. The stock market risk premium shock (collapse in corporate investment) and tightening of lending conditions contributed to the improvement in the trade balance in 2010–11.

Figure 8 shows the extent to which each of these factors contributed to the accumulation of net foreign liabilities. All in all, up to three-quarters of the buildup in foreign debt can be attributed to the capital flow shock in the model. Looser monetary policy added further to the buildup, and additional smaller contributing factors were the stock market and housing shocks, and the loosening of lending conditions.

6. Concluding Remarks

In this paper we have shown how an open-economy DSGE model with financial frictions and exogenous time-varying risk premia can be used to analyze the driving factors behind internal and external imbalances that have built up over past years. Our estimates for Spain indicate that the main source behind the buildup in debt has been low real interest rates, linked to the inflow of cheap capital due to the disappearance of the risk premium and monetary policy set at the euro-area level. The subsequent adjustment in the trade balance is mainly driven by the collapse in the bubbles in the stock and housing markets and tightening of lending conditions. The correction relies more on expenditure reduction than on expenditure switching, with a decline in domestic absorption. However, the accompanying deflation and decline in unit labor costs also brings about a further depreciation of the real effective exchange rate. We find no strong evidence of excessive wage growth as a driver of external imbalances. However, the dual labor market structure in Spain may create a degree of wage rigidity that hinders any adjustment in wage growth required to improve competitiveness.

Figure 8. Shock Decomposition: NFA Position



This analysis highlights the need for an integrated analysis of external imbalances and a proper framework for addressing such imbalances. As the main driver identified in the model is related to low real interest rates, the correction of the imbalances also has to come from an adjustment to the borrowing costs for economic agents. This could lead to a protracted period of depressed demand and a painful correction to past excess demand growth. This underlines the importance of avoiding the buildup of such imbalances in future. The EU's macroeconomic imbalances procedure (MIP) is intended to deal with this by stricter surveillance and introducing early-warning mechanisms that can alert policymakers to the buildup of unsustainable imbalances so they can take timely action. Possible policy responses include timely fiscal adjustment. Although this may be an appropriate strategy for other deficit countries, it should be noted that public-sector borrowing does not appear to have been a factor behind Spain's imbalances, and its fiscal position was, at least up to 2007, on a sustainable path. Policies that improve competitiveness may be called for—in particular, reforms reducing wage costs in the tradable sector—but productivity-enhancing matters will also raise demand, partly offsetting the net effect on external balances.[12] To the extent that non-fundamental shocks have played a role in the buildup phase, the policy recommendation would be to avoid such bubbles building up in the first place. But whether the inflow of cheap capital in a monetary union can or should be avoided is a more challenging question. Prudential policies in both recipient countries of capital flows (Spain) and source countries (surplus countries like Germany) might have helped to prevent the buildup of imbalances. Restrictions on the access to credit for households and firms, or policies to promote savings could have been introduced to avoid excessive demand growth in the early years in EMU, but it is not a priori clear whether such policies would be desirable from a welfare perspective.

[12]The direction of the impact of structural reforms on the current account is ambiguous from a theoretical point of view (see, e.g., Fournier and Koske 2010, Vogel 2011). Empirical evidence is also mixed. Jaumotte and Sodsriwiboon (2010) report a large positive effect of labor productivity on current accounts, while the empirical results in Kerdrain, Koske, and Wanner (2010) imply that such reforms have a negative impact on the current account position.

Appendix 1. Data

In total, twenty-nine variables have been used in the estimation, and their sample range is 1995:Q1–2011:Q4. GDP and national account data (consumption, government consumption, government investment, gross fixed capital formation, gross fixed capital formation construction (housing), social benefits other than social transfers in kind, as well as the corresponding price deflators) are based on Eurostat data. Residential property prices (new and existing dwellings) are from the ECB. The nominal effective exchange rate, the world price index, and the world output are based on own calculations.[13] They are trade-weighted averages across Spain's main trade partners: Argentina, Australia, Brazil, Canada, China, the euro area, Hong Kong, India, Israel, Japan, Malaysia, Mexico, Norway, Russia, Singapore, South Korea, Switzerland, Taiwan, Turkey, the United Kingdom, and the United States, altogether forty-one countries. General government consolidated gross debt, net government lending, and the implicit interest rate faced by general government are from AMECO. Three-month money-market rates for Spain and the United States are taken from Eurostat. The source for data on international investment position is the Bank of Spain.

The model is estimated in growth rates and GDP shares. Specifically, the following twenty-nine series are treated as observed:

- GDP growth,
- GDP shares (twelve): consumption, government consumption, government investment, transfers, imports, exports, world demand, construction investment, total investment, government deficit and debt, net foreign asset.
- Prices (nine): GDP, consumption, import, export, construction, house, government purchases, total investment, world.
- Spain and euro-area nominal interest rate; government interest rate (three).
- Exchange rate, wages, employment, and non-active population (four).

[13] Available on http://ec.europa.eu/economy_finance/db_indicators/competitiveness/data_section_en.htm.

Appendix 2. Priors and Estimated Parameters

Table 3. Results for Posterior Parameters

	Prior Distribution	Prior Mean	Prior S.D.	Posterior Mean	Posterior S.D.	HPD Inf.	HPD Sup.
dcc	Beta	0.500	0.1500	0.692	0.0863	0.5489	0.8328
$\gamma_{ucap,2}$	Gamma	0.020	0.0080	0.042	0.0109	0.0235	0.0593
τ_1^{CG}	Beta	−0.100	0.0400	−0.027	0.0124	−0.0461	−0.0071
γ_H	Gamma	30.000	20.0000	51.433	22.4214	16.3837	82.9471
$\gamma_{J,H}$	Gamma	30.000	20.0000	75.882	26.1220	35.1673	113.9595
γ_K	Gamma	30.000	20.0000	22.858	9.6797	7.6487	37.2564
γ_I	Gamma	15.000	10.0000	27.631	10.8292	9.7072	43.8167
γ_L	Gamma	30.000	20.0000	15.170	4.9011	7.4165	22.7198
γ_P	Beta	4.000	2.0000	6.423	1.4369	3.9558	8.6485
$\gamma_{Pconstr}$	Gamma	30.000	20.0000	30.066	12.2682	10.4622	49.1661
γ_{Phouse}	Gamma	30.000	20.0000	7.049	6.1026	0.5685	15.9709
γ_{PM}	Gamma	4.000	2.0000	1.128	0.4860	0.3754	1.8466
γ_{PX}	Gamma	4.000	2.0000	1.239	0.7338	0.2797	2.3550
γ_W	Gamma	12.000	4.0000	26.907	5.8333	17.3277	36.2256
τ_{lag}^{CG}	Beta	0.500	0.2000	0.933	0.0135	0.9097	0.9549
h	Beta	0.700	0.1000	0.751	0.0519	0.6749	0.8400
τ_1^{IG}	Beta	−0.100	0.0400	−0.037	0.0165	−0.0613	−0.0094
τ_{lag}^{IG}	Beta	0.500	0.2000	0.899	0.0666	0.8056	0.9974
κ	Gamma	1.000	0.4000	0.892	0.3231	0.3641	1.3872
ρ^L	Beta	0.850	0.0750	0.924	0.0321	0.8749	0.9775
ρ^η	Beta	0.500	0.2000	0.934	0.0364	0.8827	0.9916
$\rho^{\eta,Constr}$	Beta	0.500	0.2000	0.824	0.0691	0.7162	0.9445
$\rho^{\eta,M}$	Beta	0.850	0.0750	0.850	0.0524	0.7648	0.9363
$\rho^{\eta,X}$	Beta	0.850	0.0750	0.966	0.0179	0.9406	0.9943
ρ^X	Beta	0.500	0.2000	0.989	0.0049	0.9816	0.9969
ρ^M	Beta	0.500	0.2000	0.928	0.0209	0.8957	0.9603
ρ^N	Beta	0.500	0.2000	0.904	0.0534	0.8364	0.9997
ρ^{PCPM}	Beta	0.500	0.2000	0.375	0.1369	0.1526	0.6043
ρ^{PWPX}	Beta	0.500	0.2000	0.298	0.0776	0.1681	0.4213
ρ^{er}	Beta	0.500	0.2000	0.254	0.0110	0.2355	0.2717
ρ^{rpk}	Beta	0.850	0.0750	0.910	0.0344	0.8571	0.9670
$\rho^{rph,c}$	Beta	0.850	0.0750	0.940	0.0146	0.9162	0.9640
ρ^{rpland}	Beta	0.850	0.0750	0.950	0.0196	0.9187	0.9808
ρ^{tax}	Beta	0.850	0.0750	0.885	0.0468	0.8098	0.9557
τ_{lag}^L	Beta	0.500	0.2000	0.862	0.0327	0.8091	0.9150
sfp	Beta	0.700	0.1000	0.827	0.0803	0.6990	0.9559
$sfpconstr$	Beta	0.700	0.1000	0.896	0.0538	0.8123	0.9808
$sfphouse$	Beta	0.700	0.1000	0.772	0.0934	0.6215	0.9260
$sfpm$	Beta	0.700	0.1000	0.819	0.0842	0.6862	0.9561
$sfpx$	Beta	0.700	0.1000	0.818	0.0926	0.6749	0.9666
sfw	Beta	0.700	0.1000	0.731	0.0966	0.5698	0.8872

(*continued*)

Table 3. (Continued)

	Prior Distribution	Prior Mean	Prior S.D.	Posterior Mean	Posterior S.D.	HPD Inf.	HPD Sup.
σ^{H}	Gamma	0.500	0.1000	0.623	0.1252	0.4196	0.8192
σ^{X}	Gamma	1.250	0.5000	1.864	0.3286	1.3198	2.3870
σ^{M}	Gamma	1.250	0.5000	1.153	0.2458	0.7359	1.5503
σ^{L}	Beta	0.500	0.2000	0.319	0.1272	0.1006	0.5089
s_L	Beta	0.300	0.1000	0.274	0.0872	0.1351	0.4116
s^s	Beta	0.500	0.1500	0.486	0.0971	0.3340	0.6490
τ^{B}	Beta	0.020	0.0100	0.021	0.0034	0.0157	0.0266
τ^{DEF}	Beta	0.020	0.0100	0.012	0.0055	0.0029	0.0197
$\tau^{B,IG}$	Beta	0.020	0.0100	0.010	0.0032	0.0044	0.0147
$\tau^{DEF,IG}$	Beta	0.020	0.0100	0.024	0.0087	0.0097	0.0375
$\tau^{DEF,TR}$	Beta	0.020	0.0100	0.021	0.0101	0.0047	0.0359
rpe	Beta	0.003	0.0010	0.001	0.0003	0.0005	0.0015
θ^{s}	Beta	0.007	0.0010	0.008	0.0007	0.0067	0.0091
$rpdebt$	Beta	0.003	0.0012	0.003	0.0012	0.0012	0.0051
b^{U}	Beta	0.300	0.1000	0.466	0.0683	0.3618	0.5715
ρ^{TR}	Beta	0.850	0.0750	0.946	0.0216	0.9116	0.9799

Table 4. Results from Posterior Parameters (standard deviation of structural shocks)

	Prior Distribution	Prior Mean	Prior S.D.	Posterior Mean	Posterior S.D.	HPD Inf.	HPD Sup.
ε^{UC}	Gamma	0.020	0.0080	0.036	0.0082	0.0225	0.0486
ε^{L}	Gamma	0.040	0.0160	0.065	0.0140	0.0417	0.0865
ε^{η}	Gamma	0.020	0.0080	0.015	0.0032	0.0103	0.0207
$\varepsilon^{\eta Constr}$	Gamma	0.100	0.0400	0.122	0.0396	0.0547	0.1833
$\varepsilon^{\eta,M}$	Gamma	0.020	0.0080	0.031	0.0053	0.0223	0.0393
$\varepsilon^{\eta,X}$	Gamma	0.020	0.0080	0.020	0.0045	0.0136	0.0273
ε^{X}	Gamma	0.005	0.0020	0.006	0.0006	0.0051	0.0070
ε^{CG}	Gamma	0.005	0.0020	0.002	0.0002	0.0015	0.0023
ε^{IG}	Gamma	0.005	0.0020	0.003	0.0004	0.0027	0.0040
ε^{M}	Gamma	0.005	0.0020	0.008	0.0008	0.0069	0.0095
ε^{N}	Gamma	0.050	0.0200	0.067	0.0172	0.0367	0.0932
$\varepsilon^{M,EA}$	Gamma	0.003	0.0010	0.001	0.0001	0.0011	0.0014
ε^{PC}	Gamma	0.003	0.0010	0.004	0.0003	0.0036	0.0047
ε^{rpe}	Gamma	0.005	0.0020	0.000	0.0000	0.0003	0.0003
ε^{rpk}	Gamma	0.005	0.0020	0.008	0.0019	0.0052	0.0111
$\varepsilon^{rph,c}$	Gamma	0.010	0.0040	0.009	0.0026	0.0048	0.0128
ε^{rpland}	Gamma	0.010	0.0040	0.016	0.0048	0.0084	0.0237
ε^{T}	Gamma	0.010	0.0040	0.020	0.0029	0.0153	0.0247
ε^{TR}	Gamma	0.005	0.0020	0.003	0.0008	0.0021	0.0047
ε^{W}	Gamma	0.020	0.0080	0.030	0.0078	0.0179	0.0426
ε^{z}	Gamma	0.010	0.0040	0.007	0.0008	0.0061	0.0088

References

Adjemian, S., H. Bastani, M. Juillard, F. Mihoubi, G. Perendia, M. Ratto, and S. Villemot. 2011. "Dynare: Reference Manual, Version 4." Dynare Working Paper No. 1, CEPREMAP. Available at http://www.dynare.org.

Andres, J., S. Hurtado, E. Ortega, and C. Thomas. 2010. "Spain in the Euro: A General Equilibrium Analysis." *SERIEs, Journal of the Spanish Economic Association* 1 (1): 67–95.

Bernanke, B., and M. Gertler. 1999. "Monetary Policy and Asset Price Volatility." *Economic Review* (Federal Reserve Bank of Kansas City) (4th Quarter): 17–51.

Burriel, P., J. Fernández-Villaverde, and J. F. Rubio-Ramírez. 2010. "MEDEA: A DSGE Model for the Spanish Economy." *SERIEs, Journal of the Spanish Economic Association* 1 (1): 175–243.

Catao, L., and G. M. Milesi-Ferretti. 2013. "External Liabilities and Crises." IMF Working Paper No. 13/113.

Christiano, L., R. Motto, and M. Rostagno. 2008. "Shocks, Structures or Monetary Policies? The Euro Area and US after 2001." *Journal of Economic Dynamics and Control* 32 (8): 2476–2506.

European Commission. 2006. "The EU Economy, 2006 Review: Adjustment Dynamics in the Euro Area—Experiences and Challenges. *European Economy* 2006/6.

———. 2008. "EMU@10: Successes and Challenges after 10 Years of Economic and Monetary Union." *European Economy* 2008/2.

Fournier, J., and I. Koske. 2010. "A Simple Model of the Relationship between Productivity, Saving and the Current Account." OECD Economics Department Working Paper No. 816.

Gorton, G. B. 2010. *Slapped by the Invisible Hand. The Panic of 2007.* Oxford University Press.

Hall, R. E. 2010. "The High Sensitivity of Economic Activity to Financial Frictions." Mimeo, Stanford University.

Iacoviello, M. 2005. "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle." *American Economic Review* 95 (3): 739–64.

Iacoviello, M., and S. Neri. 2010. "Housing Market Spillovers: Evidence from an Estimated DSGE Model." *American Economic Journal: Macroeconomics* 2 (2): 125–64.

In 't Veld, J., R. Raciborski, M. Ratto, and W. Roeger. 2011. "The Recent Boom-Bust Cycle: The Relative Contribution of Capital Flows, Credit Supply and Asset Bubbles." *European Economic Review* 55 (3): 386–406.

Jaumotte, F., and P. Sodsriwiboon. 2010. "Current Account Imbalances in the Southern Euro Area." IMF Working Paper No. 10/139.

Kerdrain, C., I. Koske, and I. Wanner. 2010. "The Impact of Structural Policies on Saving, Investment and Current Accounts." OECD Economics Department Working Paper No. 815.

Kiyotaki, N., and J. Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105 (2): 211–48.

Lane, P. R., and G. M. Milesi-Ferretti. 2011. "External Adjustment and the Global Crisis." IMF Working Paper No. 11/197.

Martinez-Mongay, C., L. A. Maza Lasierra, and J. Yaniz Igal. 2007. "Asset Booms and Tax Receipts: The Case of Spain, 1995–2006." European Economy Economic Paper No. 293.

Monacelli T. 2007. "New Keynesian Models, Durable Goods and Collateral Constraints." CEPR Discussion Paper No. 5916.

Ratto, M., W. Roeger, and J. in 't Veld. 2009. "QUEST III: An Estimated Open-Economy DSGE Model of the Euro Area with Fiscal and Monetary Policy." *Economic Modelling* 26 (1): 222–33.

Schmitt-Grohe, S., and M. Uribe. 2003. "Closing Small Open Economy Models." *Journal of International Economics* 61 (1): 163–85.

Vogel, L. 2011. "Structural Reforms and External Rebalancing in the Euro Area: A Model-Based Analysis." European Economy Economic Paper No. 443.

Supplementary Appendix. Shock Decompositions

The appendix shows historical decomposition of other series in addition to what is included in the main text.

Employment Rate

Figure S.1 shows the shock decomposition of the deviations of the employment rate, as share of total population, from its steady state. The employment share increased over the sample by almost 20 percentage points by 2007 before falling back again by more than half.[1] The contribution of the wage markup shock was negative in the first years of the estimation period but gradually reduced in size and turned into a positive contribution in 2001. It continued to boost employment up to 2007–8 but then declined.[2] The positive effect mainly reflects moderate wage growth and an increase in the participation rate. Other shocks that played a role are a positive contribution from the housing risk premium shock and a negative contribution from fiscal shocks.

Real Exchange Rate

The real exchange rate shows a strong appreciation trend since 2001, to a large extent driven by the positive contribution of the capital flows shock (lower interest rates), which boosted demand and inflation (figure S.2). Positive contributions also came from technology and loosening lending conditions. Lower productivity growth raised inflation in Spain and shows up as a positive contribution to the

[1] The accumulated shocks and the data line show a large gap at the beginning of the sample (contribution of Others). The reason for the initial gap is the calibration of the steady-state employment rate at 60 percent, or 41 percent as share of total population, which is above the level in the beginning of the sample. We assume a higher steady-state level in the calibration, as the 1995–99 averages can no longer be considered representative for the Spanish labor market, as labor market reforms have led to a regime shift in the Spanish economy. Instead we base the calibration of labor market parameters on the full sample 1995–2011 (see section 4).

[2] The profile of this shock corresponds to European Commission estimates of the NAIRU for Spain that show a decline from 18 percent in 1995 to 8 percent in 2007, and a rise again since the onset of the crisis.

Figure S.1. Shock Decomposition: Employment

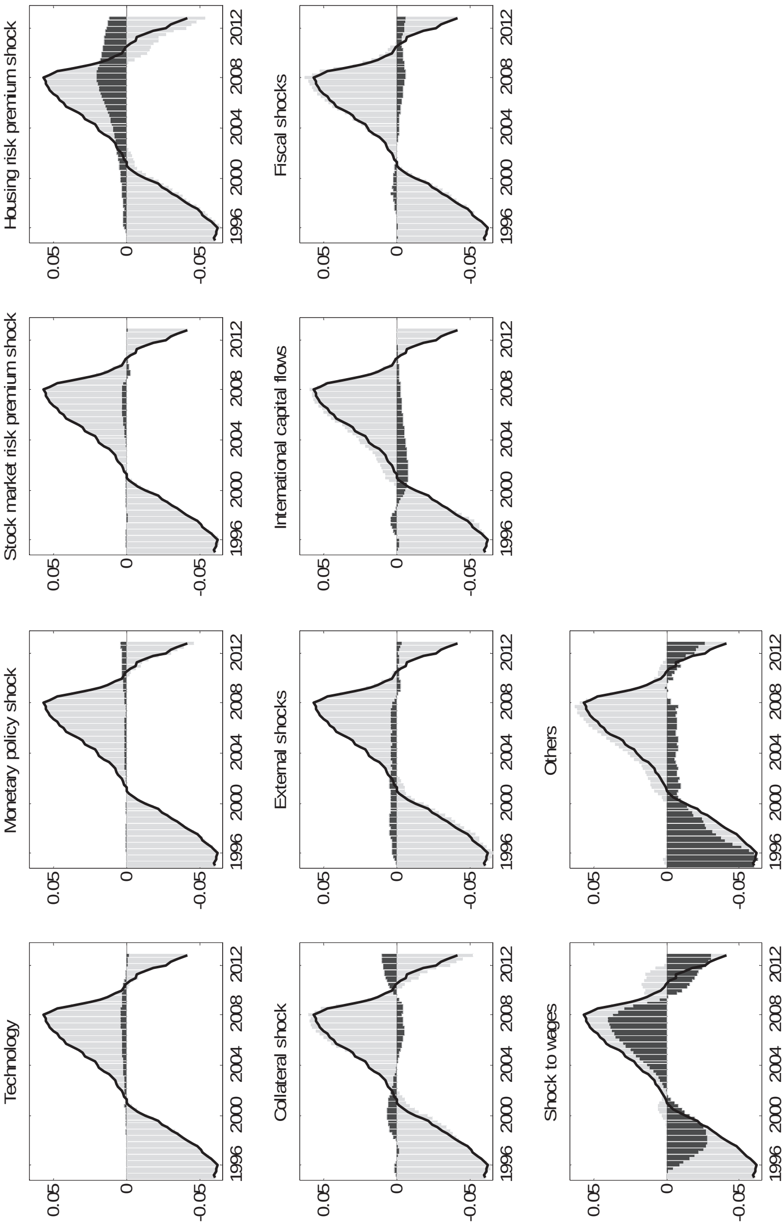
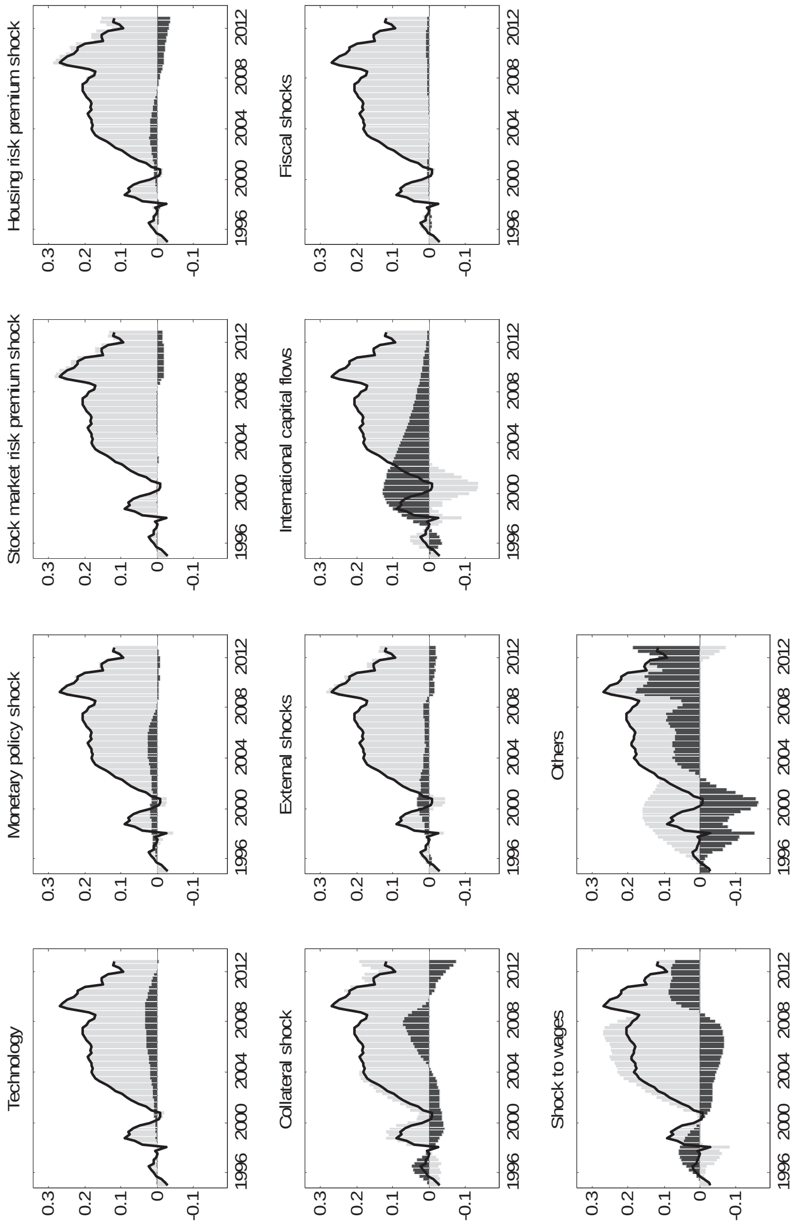


Figure S.2. Shock Decomposition: Real Exchange Rate



real exchange rate. The main negative factor reducing the appreciation trend was the labor market shock (labor market developments helped to reduce inflationary pressures in the period up to 2008). Insufficient wage flexibility in the crisis led to a further overvaluation of the real effective exchange rate. In recent years some correction has taken place, but there is still a sizable overvaluation of around 10 percent by the end of the sample period.

Unit Labor Costs

The shock decomposition of the year-on-year growth rate in unit labor costs shows, over most of the sample, no strong deviation of growth in unit labor costs from steady-state levels, reaching a peak in 2008 but falling significantly in recent years (figure S.3). The capital flows shock boosted demand and inflation and contributed to growth in unit labor costs up to 2003, and the collateral shock played a similar role in following years. The shock to wages helped to offset these two factors, but turned into a positive contribution in 2007–9. In the last years, growth in unit labor costs has been kept down by negative contributions of the technology shock and the collateral shock (tighter lending conditions), the negative stock market risk premium shock, and the external demand shock (decline in world trade).

House Prices

The shock decomposition of real house prices (deflated by the GDP deflator; figure S.4) shows that the sharp rise in house prices is mainly driven by non-fundamental factors in the model (housing risk premium shock). This house price bubble appears not to have completely burst yet, and there is still a significant positive contribution of the risk premium shock on house prices at the end of 2011. In addition, there was also a small positive contribution of the capital flows shock. Technological progress put some downward pressure on house prices, but this effect has become smaller in recent years.

Real Interest Rate

The shock decomposition of the real interest rate (figure S.5) shows the dominant effect of the financial shocks (elimination of risk

Figure S.3. Shock Decomposition: Growth of Unit Labor Cost

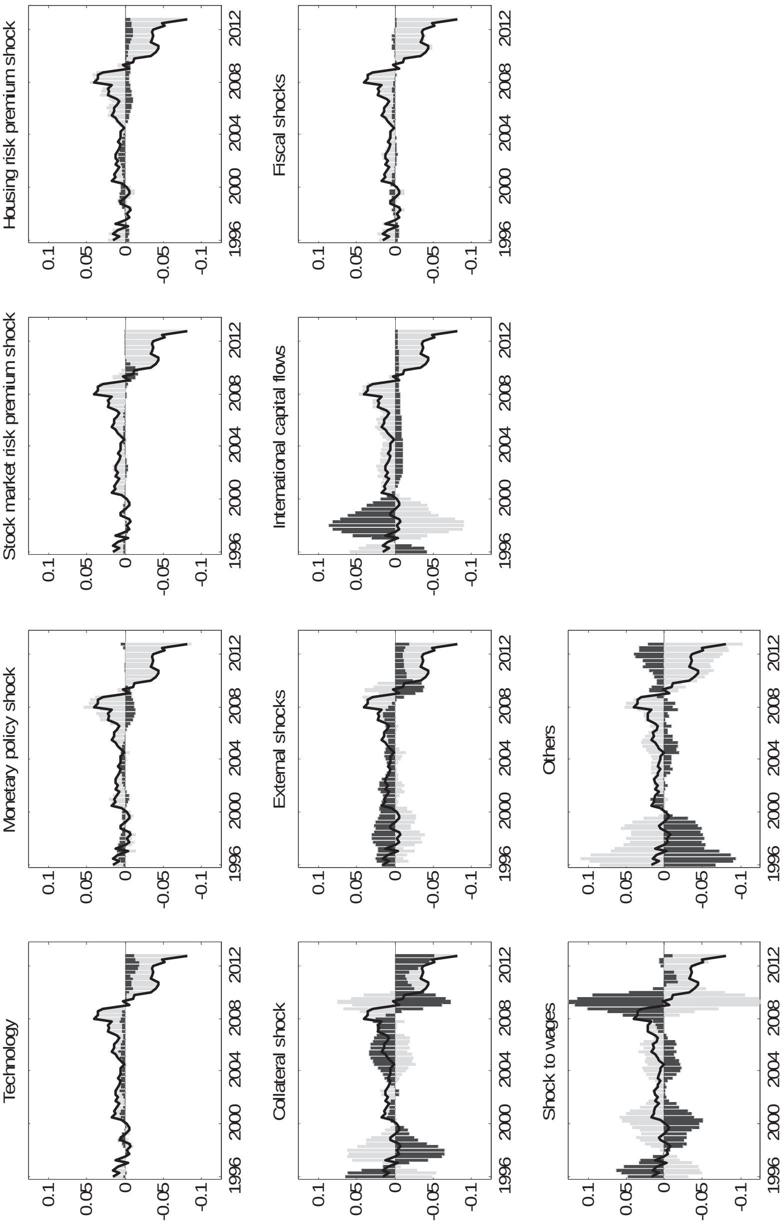


Figure S.4. Shock Decomposition: House Prices

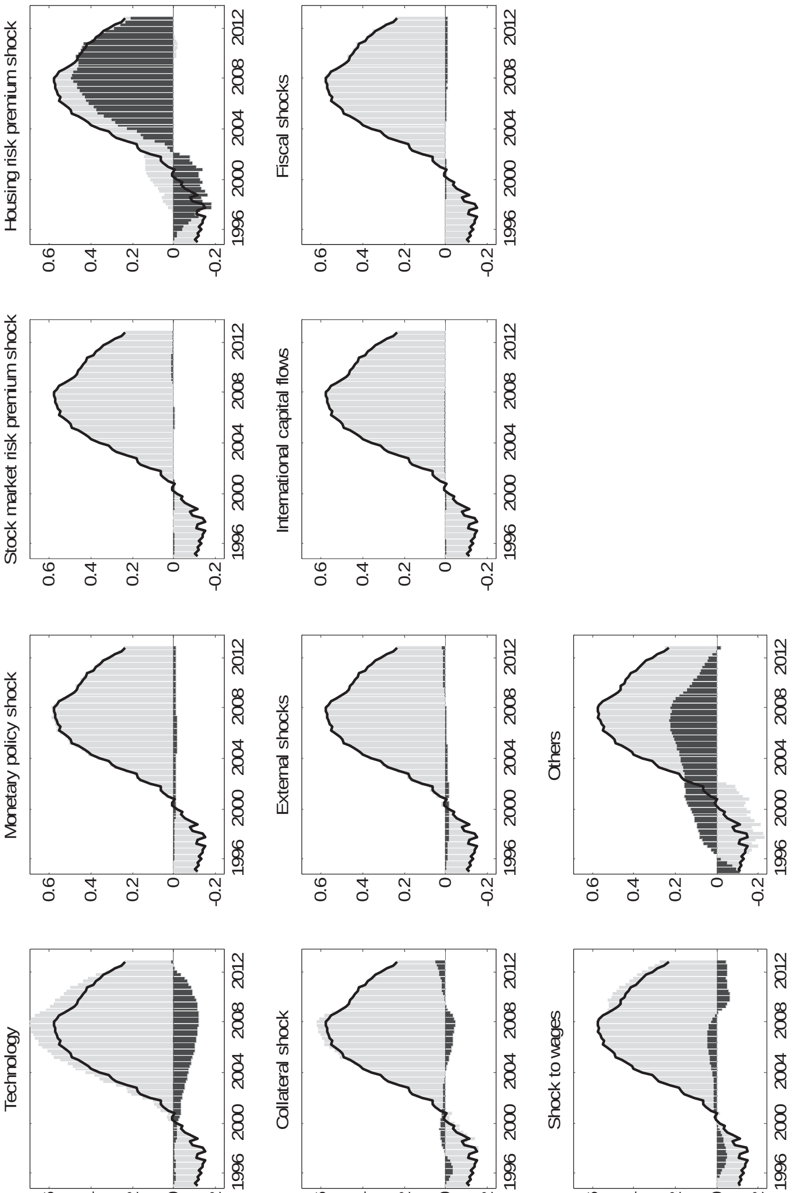
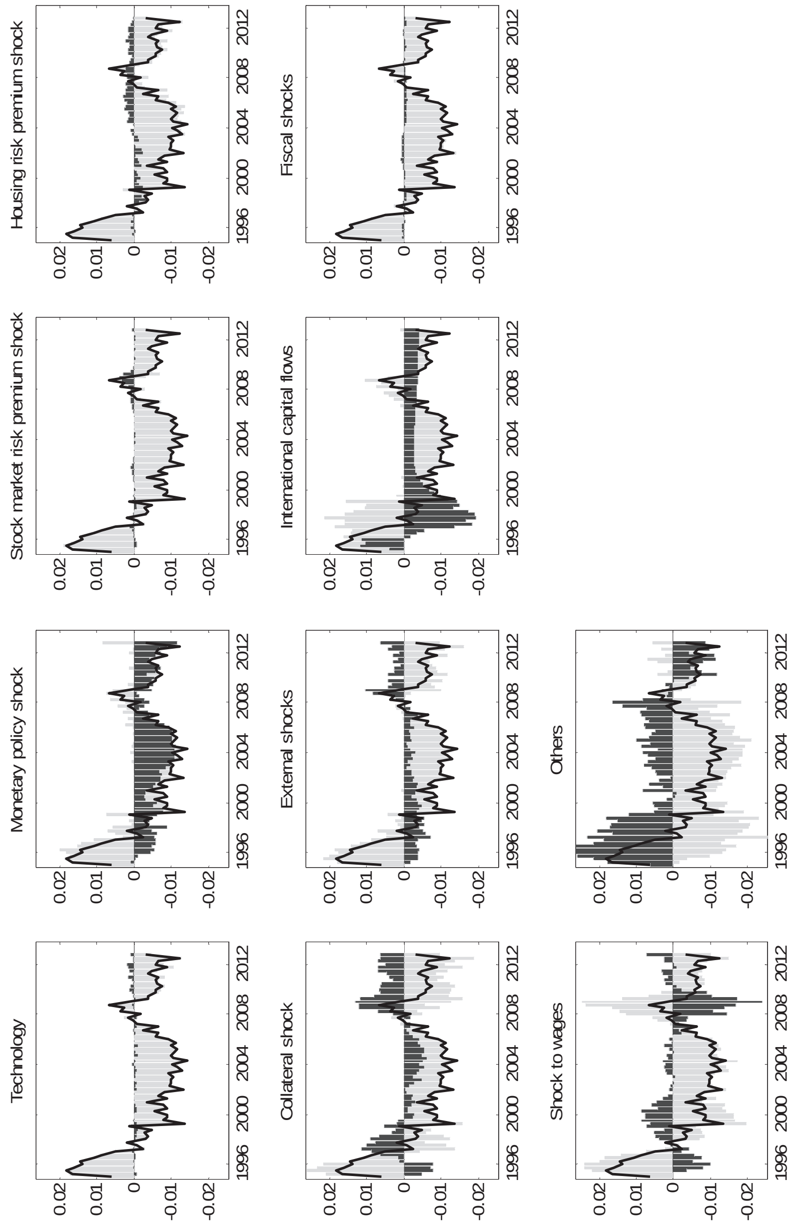


Figure S.5. Shock Decomposition: Real Interest Rate



premium in EMU and relative monetary stance). Real interest rates were below their steady-state level for most of the period since 1997, and were only briefly above it in 2008, when the ECB kept interest rates high while conditions in Spain deteriorated and Spain experienced a short period of deflation. Then, in late 2008, the ECB policy rate came down and the real interest rate declined again.

The Comeback of Inflation as an Optimal Public Finance Tool*

Giovanni Di Bartolomeo,[a] Patrizio Tirelli,[b] and Nicola Acocella[c]
[a]Department of Economics and Law, Sapienza University of Rome
[b]DEMS, University of Milan Bicocca
[c]MEMOTEF, Sapienza University of Rome

We challenge the widely held belief that New Keynesian models cannot predict optimal positive inflation rates. In fact, interest rates are justified by the Phelps argument that monetary financing can alleviate the burden of distortionary taxation. We obtain this result because, in contrast with previous contributions, our model accounts for public transfers as a component of fiscal outlays. We also contradict the view that the Ramsey policy should minimize inflation volatility and induce near-random-walk dynamics of public debt in the long run. In our model it should instead stabilize debt-to-GDP ratios in order to mitigate steady-state distortions. Our results thus provide theoretical support to policy-oriented analyses which call for a reversal of debt accumulated in the aftermath of the 2008 financial crisis.

JEL Codes: E52, E58, J51, E24.

1. Introduction

Optimal monetary policy analyses (Khan, King, and Wolman 2003; Schmitt-Grohé and Uribe, SGU henceforth, 2004a) identify two key

*The authors are grateful to the IJCB editor Pierpaolo Benigno, three anonymous referees, Fabio Canova, Martin Ellison, Jordi Galì, Luca Onorante, Paolo Paesani, Paul Levine, Frank Smets, and seminar participants at EC2 2011 Conference (EUI), MMF2012 (Trinity College, Dublin), National Bank of Poland, and Universities of Milan Bicocca, Cork, and Crete for useful comments on earlier drafts. The authors acknowledge financial support from MIUR (PRIN). Patrizio Tirelli also gratefully acknowledges financial support from EC project 320278 - RASTANEWS.

frictions driving the optimal level of long-run (or trend) inflation. The first one is the adjustment cost of goods prices, which invariably drives the optimal inflation rate to zero. The second one is monetary transaction costs that arise unless the central bank implements the Friedman rule, i.e., a zero nominal inflation rate in steady state. In their survey of the literature, SGU (2011) argue that the optimality of zero inflation is robust to other frictions, such as nominal wage adjustment costs, downward wage rigidity, hedonic prices, the existence of an untaxed informal sector, and the zero bound on the nominal interest rate. This latter result is broadly confirmed by Coibion, Gorodnichenko, and Wieland (2012), who find that the optimal inflation rate is low, typically less than 2 percent, even when the economy is hit by costly but infrequent episodes at the zero lower bound. A consensus therefore seems to exist that monetary transaction costs are relatively small at zero inflation, and that implementing low and stable inflation is the proper policy.

This theoretical result is in sharp contrast with empirical evidence. For instance, both in the United States and in the euro area, average inflation rates over the 1970–99 period have been close to 5 percent. Even the widespread central bank practice of adopting inflation targets between 2 percent and 4 percent is apparently at odds with theories of the optimal inflation rate (SGU 2011).

Furthermore, following the buildup of large stocks of debt in the aftermath of the 2007–8 financial crisis, some economists have argued that the public debt surge should be reversed and that a temporary increase in inflation might be necessary to achieve this goal. For instance, Rogoff (2010) suggests that "two or three years of slightly elevated inflation strikes me as the best of many very bad options." Blanchard, Dell'Ariccia, and Mauro (2010) point at the potential role of the inflation tax as one among several distortionary taxes which are available to policymakers. Aizenman and Marion (2011) predict that a 6 percent inflation rate would reduce the debt-to-GDP ratio by 20 percent within four years. These contributions are in line with the well-known Phelps (1973) argument that to alleviate the burden of distortionary taxation, it might be optimal for governments to resort to monetary financing, driving a wedge between the private and the social cost of money.

The Phelps argument has been widely investigated in the framework of general equilibrium models, and never found sufficient to warrant the optimality of a significantly positive inflation rate. Two main results have been established. The first one is that distortionary taxation does not warrant deviations from the Friedman rule unless factor incomes are sub-optimally taxed (see SGU 2011 and references cited therein). The underlying intuition is that since all resources are eventually used for consumption, then the inflation tax, which affects consumption transaction costs, is desirable only to the extent that other taxes have a sub-optimal effect on consumption. The second main result is that when the goods market characterization is modified to account for (sub-optimally taxed) monopolistic distortions, numerical simulations suggest that the optimal inflation rate is negative and very close to zero, even accounting for the Phelps effect (SGU 2004a). This conclusion carries over to the optimality of near-zero volatility of inflation and near-random-walk behavior in government debt and tax rates in response to shocks, implying that the recent increase of public debt in developed economies should be regarded as a tax-smoothing device in response to the financial crisis.

Our paper reconsiders the importance of the Phelps effect and obtains results that challenge the optimality of near-zero inflation rates when the tax system is incomplete. We show that a non-negligible inflation rate might indeed be optimal and that inflation (and tax rates) volatility should be exploited in order to stabilize debt-to-GDP ratios in the long run.

The starting point in our analysis is that the optimal zero-inflation result obtained in dynamic stochastic general equilibrium (DSGE) models with incomplete tax systems is the consequence of unrealistic assumptions about the size and composition of public expenditure. In the literature, standard calibrations of public expenditures focus on public-consumption-to-GDP ratios, typically set at 20 percent (SGU 2004a; Aruoba and Schorfeide 2011). This follows a long-standing tradition in business-cycle models, where only public consumption decisions have real effects. In our framework this choice is not correct, because the focus here is on distortionary financing of public expenditures in steady state, where also other components of public expenditure matter. To the best of our knowledge, the only exception is SGU (2006), who determine the optimal inflation rate in a medium-scale model where public consumption and transfers respectively amount to about 20 percent and 9 percent of GDP.

They find that the optimal inflation rate is positive but very small, half a percentage point, and that the inclusion of public transfers accounts for a 0.7 percent increase in the optimal inflation rate. Their intuition for the inflationary effect of public transfers is that only transfers are pure rents to households and inflation is an indirect tax on those pure rents.

As a matter of fact, public consumption accounts for a limited component of the overall public expenditures in OECD countries, and transfers are relatively large (table 1).

We show that just allowing for a plausible parameterization of public consumption and transfers in the SGU (2004a) model reverses the standard conclusion about the optimal inflation rate, which now monotonically increases from 2 percent to 12 percent as the transfers-to-GDP ratio goes from 10 percent to 20 percent. Further, our calculations contradict the claim that public transfers per se require an inflation tax (SGU 2006). In fact, we also find that, absent public transfers, very large public-consumption-to-GDP ratios are also associated with a positive inflation rate. For instance, we obtain that the optimal inflation rate monotonically increases from 2 percent to 12 percent as the public-consumption-to-GDP ratio grows from 40 percent to 47 percent. Given the historically observed public consumption ratios, these latter results are not empirically relevant, but they challenge received wisdom about the reasons why level and composition of public expenditures should matter for the identification of the optimal inflation rate.

By working with a simplified version of our model, we are able to show that changes in public consumption and public transfers would generate identical variations in the optimal inflation rate if public consumption did not affect the aggregate resource constraint. The limited incentive to inflate that we observe in response to a public consumption variation is due to its contemporaneous effects that operate through the aggregate resource constraint and impact on (i) inflation- and labor-tax revenues, and (ii) the planner's desired marginal rate of substitution between consumption and leisure.

We also investigate the optimal fiscal and monetary policy responses to shocks. The issue is admittedly not new, but we are able to provide new contributions to the literature. When prices are

Table 1. Government Expenditures and Revenues, 1998–2008 (average ratios to GDP)

	Public Consumption	Other Public Expenditures	Total Revenues
Australia	18.00	16.97	36.26
Austria	19.10	32.29	49.71
Belgium	22.13	27.82	49.39
Canada	19.49	21.56	42.08
Czech Republic	21.24	22.81	40.12
Denmark	25.84	27.88	55.96
Finland	21.75	27.74	53.12
France	23.39	29.21	49.90
Germany	18.96	27.58	44.61
Greece	16.52	28.32	40.19
Hungary	21.98	27.42	43.20
Ireland	15.11	19.40	44.16
Italy	19.10	28.94	45.25
Japan	17.07	21.28	31.81
Netherlands	23.57	22.19	45.34
New Zealand	17.97	20.89	42.01
Norway	20.76	23.54	56.63
Poland	17.95	25.34	39.20
Portugal	19.57	25.48	41.59
Slovak Republic	20.24	21.35	36.55
Spain	17.75	21.52	38.67
Sweden	26.67	29.03	57.21
Switzerland	11.4	23.48	34.40
United Kingdom	19.83	22.28	40.38
United States	15.26	20.51	33.47
Euro Area	20.17	27.11	45.39

Source: OECD.

flexible and governments issue non-contingent nominal debt (Chari, Christiano, and Kehoe 1991), it is optimal to use inflation as a lump-sum tax on nominal wealth, and the highly volatile inflation rate allows to smooth taxes over the business cycle. This result is intuitive insofar as taxes are distortionary whereas inflation volatility is costless. SGU (2004a) show that when price adjustment is costly, optimal inflation volatility is in fact minimal and long-run debt adjustment allows to obtain tax smoothing over the business cycle. In our paper, the SGU result is reversed, even when the amount of public transfers is relatively small (12 percent of GDP). In this case, tax and inflation volatility are exploited to limit debt adjustment in the long run.

The interpretation of our result is simple. As discussed above, public transfers increase the tax burden in steady state. In this case, the accumulation of debt in the face of an adverse shock—which would work as a tax-smoothing device in SGU (2004a)—is less desirable, because it would further increase long-run distortions. To avoid such distortions, the policymaker is induced to front-load fiscal adjustment and to inflate away part of the real value of outstanding nominal debt. Consumption smoothing is therefore reduced relative to SGU (2004a).

To the best of our knowledge, this is the first study of the optimal interaction between inflation and tax policies when transfers account for the relatively large proportion of public expenditures that is documented in the data. A number of recent papers have analyzed the macroeconomic implications of public transfer schemes, but their focus is different from ours. Alonso-Ortiz and Rogerson (2010) investigate the labor supply response and the welfare implications of an optimal public transfer scheme in the context of a model with idiosyncratic productivity shocks, incomplete financial markets, and flexible prices. Oh and Reis (2011) analyze the role of transfers for consumption stabilization in the context of heterogeneous agents, incomplete markets, and sticky prices—when taxes are lump sum, no public debt accumulation is allowed and the central bank is constrained to implement a zero-inflation policy. Angelopoulos, Philippopoulos, and Vassilatos (2009) maintain the representative-agent hypothesis and incorporate an uncoordinated redistributive struggle for transfers into an otherwise standard DSGE model. Zubairy (2014) investigates the consequences of temporary public transfer shocks in an estimated representative-agent DSGE model.

The remainder of the paper is organized as follows. The next section describes the model. Section 3 introduces the Ramsey policy and illustrates our main results. Section 4 discusses optimal monetary and fiscal stabilization policies. Section 5 concludes.

2. The Model

We consider a simple infinite-horizon production economy populated by a continuum of households and firms whose total measures are normalized to one. Monopolistic competition and nominal rigidities characterize product markets. The labor market is competitive. A demand for money is motivated by assuming that money facilitates transactions. The government finances an exogenous stream of expenditures by levying distortionary labor income taxes and by printing money. Optimal policy is set according to a Ramsey plan.

As discussed by SGU (2011), positive inflation may be a desirable instrument if some part of income is sub-optimally taxed. In the narrow framework of our model, the choice of inflating the economy depends on untaxed monopolistic profits in the goods market, and the introduction of a uniform income tax would reduce the incentive to inflate. However, the tax system might be incomplete or sub-optimal for other reasons. For instance, one might take into account the existence of an informal sector of the economy, or introduce monopolistic competition in the labor market. Here the model is deliberately simple to highlight the theoretical challenge to the claim that price stability is indeed optimal even when the tax system is incomplete. Providing a complete quantitative analysis of the optimal inflation rate is beyond the scope of the paper.

2.1 Households

The representative household (i) maximizes the following utility function:

$$U = E_{t=0} \sum_{t=0}^{\infty} \beta^t u\left(c_{t,i}, l_{t,i}\right);$$

$$u\left(c_{t,i}, l_{t,i}\right) = \ln c_{t,i} + \eta \ln\left(1 - l_{t,i}\right), \tag{1}$$

where $\beta \in (0,1)$ is the intertemporal discount rate, $c_{t,i} = \left(\int_0^1 c_{t,i}(j)^{\rho} dj\right)^{\frac{1}{\rho}}$ is a consumption bundle, and $l_{t,i}$ denotes the individual labor supply. The consumption price index is $P_t = \left(\int_0^1 p_t(j)^{\frac{\rho}{\rho-1}} di\right)^{\frac{\rho-1}{\rho}}$.

The flow budget constraint in period t is given by

$$c_{t,i}\left(1+s(v_{t,i})\right)+\frac{M_{t,i}}{P_t}+\frac{B_{t,i}}{P_t}=\left(1-\tau_t\right)w_{t,i}l_{t,i}+\frac{M_{t-1,i}}{P_t}+\theta_t+\frac{R_{t-1}B_{t-1,i}}{P_t}+t_t, \qquad (2)$$

where $w_{t,i}$ is the real wage; τ_t is the labor income tax rate; t_t denotes real fiscal transfers; θ_t is firms' profits; R_t is the gross nominal interest rate; and $B_{t,i}$ is a nominally riskless bond that pays one unit of currency in period $t+1$. $M_{t,i}$ defines nominal money holdings to be used in period $t+1$ in order to facilitate consumption purchases.

Consumption purchases are subject to a transaction cost[1]

$$s(v_{t,i}), \quad s'(v_{t,i})>0 \text{ for } v_{t,i}>v^*, \qquad (3)$$

where $v_{t,i}=\frac{P_{t,i}c_{t,i}}{M_{t,i}}$ is the household's consumption-based money velocity. The features of $s(v_{t,i})$ are such that a satiation level of money velocity ($v^*>0$) exists where the transaction cost vanishes and, simultaneously, a finite demand for money is associated with a zero nominal interest rate. Following SGU (2004a) the transaction cost is parameterized as[2]

$$s(v_{t,i})=Av_{t,i}+\frac{B}{v_{t,i}}-2\sqrt{AB}. \qquad (4)$$

The first-order conditions of the household's maximization problem are[3]

$$c_t(j)=c_t\left(\frac{p_t(j)}{P_t}\right)^{\frac{1}{\rho-1}} \qquad (5)$$

[1]See Sims (1994), SGU (2004a, 2011), Guerron-Quintana (2009), and Altig et al. (2011).

[2]Our results are robust to the alternative specification for the transaction cost used by Brock (1989) and Kimbrough (2006), which implies a Cagan (1956) money demand function. A proof is available upon request. The model is also compatible with Baumol (1952) demand for money (see SGU 2004a).

[3]When solving its optimization problem, the household takes as given goods and bond prices. As usual, we also assume that the household is subject to a solvency constraint that prevents it from engaging in Ponzi schemes.

$$\lambda_t = \frac{u_c\left(c_t, l_t\right)}{1 + s(v_t) + v_t s'(v_t)} \tag{6}$$

$$\lambda_t = \beta E_t\left(\frac{\lambda_{t+1} R_t}{\pi_{t+1}}\right) \tag{7}$$

$$\lambda_t\left(1 - \tau_t\right) w_t = -u_l\left(c_t, l_t\right) \tag{8}$$

$$\frac{R_t - 1}{R_t} = s'(v_t) v_t^2. \tag{9}$$

Equation (5) is the demand for the good j. As in SGU (2004a), condition (6) states that the transaction cost introduces a wedge between the marginal utility of consumption and the marginal utility of wealth that vanishes only if $v = v^*$. Equation (7) is a standard Euler condition where $\pi_{t+1} = P_{t+1}/P_t$ denotes the gross inflation rate. Equation (8) defines the individual labor supply condition. Finally, equation (9) implicitly defines the money demand function, such that

$$\frac{M_t}{P_t} = \left(\frac{R_t - 1}{R_t A} + \frac{B}{A}\right)^{-\frac{1}{2}} c_t. \tag{10}$$

2.2 Firms' Pricing Decisions

Each firm (j) produces a differentiated good:[4]

$$y_t(j) = z_t l_{t,j}, \tag{11}$$

where z_t denotes a productivity shock.[5]

We assume a sticky-price specification based on Rotemberg (1982) quadratic cost of nominal price adjustment:

$$\frac{\xi_p}{2} y_t\left(\pi_t - 1\right)^2, \tag{12}$$

[4]We abstract from capital accumulation and assume constant returns to scale of employed labor. The consequences of these two assumptions are discussed in SGU (2006) and SGU (2011), respectively. Our results are not affected by the introduction of diminishing returns to scale for labor (simulation results available upon request).

[5]We assume that $\ln z_t$ follows an $AR(1)$ process.

where $\xi_p > 0$ is a measure of price stickiness. In line with Ascari, Castelnuovo, and Rossi (2011), we assume that the reoptimization cost is proportional to output.[6]

In a symmetrical equilibrium, the price adjustment rule satisfies

$$\frac{z_t\left(\rho - mc_t\right)}{1-\rho} + \xi_p \pi_t\left(\pi_t - 1\right) = E_t \beta \frac{y_{t+1}\lambda_{t+1}}{y_t \lambda_t}\xi_p\left[\pi_{t+1}\left(\pi_{t+1}-1\right)\right], \tag{13}$$

where

$$mc_t = \frac{1}{z_t} w_t. \tag{14}$$

From (5) it would be straightforward to show that $\frac{1}{\rho} = \mu^p$ defines the price markup that obtains under flexible prices.

2.3 Government Budget and Aggregate Resource Constraints

The government supplies an exogenous, stochastic, and unproductive amount of public good g_t and implements exogenous transfers t_t. Government financing is obtained through a labor income tax, money creation, and issuance of one-period, nominally risk-free bonds. The government's flow budget constraint is then given by[7]

$$R_{t-1}b_{t-1} + g_t + t_t = \tau_t w_t l_t + \frac{M_t - M_{t-1}}{P_t} + b_t, \tag{15}$$

where $b_t = \frac{B_t}{P_t}$ defines real debt.

The aggregate resource constraint closes the model:

$$y_t = c_t\left(1 + s(v_t)\right) + g_t + \frac{\xi_p}{2} y_t \left(\pi_t - 1\right)^2. \tag{16}$$

[6] Our results are independent of this assumption. A proof is available upon request.

[7] As in SGU (2004a), $\ln(g_t/y_t)$, is assumed to evolve exogenously following an independent $AR(1)$ process. We assume instead that the level of the real transfer (t_t/y_t) is non-stochastic.

3. Ramsey Policy

3.1 Optimal Fiscal and Monetary Policy

The Ramsey policy is a set of plans $\{c_t, l_t, \lambda_t, mc_t, \pi_t, v_t, R_t, \tau_t, b_t\}_{t=0}^{+\infty}$ that maximizes the expected value of (1) subject to the competitive equilibrium conditions (6), (7), (8), (9), (11), (13), (14), (15), and (16), and to the exogenous fiscal and technology shocks. Given (6), (8), and (14), labor-tax revenues may be written as

$$\tau_t w_t l_t = \left(z_t mc_t + \frac{u_l\left(c_t, l_t\right)\left(1 + s(v_t) + v_t s'(v_t)\right)}{u_c\left(c_t, l_t\right)} \right) l_t. \tag{17}$$

Condition (17) simply states that government fiscal revenues are equivalent to the wedge between the firm's wage cost and the household's desired wage rate.

The Lagrangian of the Ramsey planner problem can be written as follows:

$$\begin{aligned}
\mathcal{L} = E_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ & u\left(c_t, l_t\right) + \lambda_t^{AR} \left[z_t l_t - c_t\left(1 + s_t\right) - g_t - \frac{\xi_p z_t l_t \left(\pi_t - 1\right)^2}{2} \right] \\
& + \lambda_t^{B} \left[\lambda_t - \beta \frac{\lambda_{t+1} R_t}{\pi_{t+1}} \right] \\
& + \lambda_t^{GBC} \left[\frac{c_t}{v_t} + \frac{B_t}{P_t} + \left(z_t mc_t + \frac{u_l\left(c_t, l_t\right)\left[1 + s(v_t) + v_t s'(v_t)\right]}{u_c\left(c_t, l_t\right)} \right) l_t \right. \\
& \left. - R_{t-1} \frac{B_{t-1}}{P_{t-1}} - \frac{c_{t-1}}{\pi_t v_{t-1}} - g_t - t_t \right] \\
& + \lambda_t^{Ph} \left[\frac{\beta y_{t+1} \lambda_{t+1} \xi_p \pi_{t+1}\left(\pi_{t+1} - 1\right)}{y_t \lambda_t} - \frac{z_t\left(\rho - mc_t\right)}{1 - \rho} \right. \\
& \left. - \xi_p \pi_t \left(\pi_t - 1\right) \right] + \lambda_t^{MUC} \left[\frac{u_c\left(c_t, l_t\right)}{1 + s(v_t) + v_t s'(v_t)} - \lambda_t \right] \Bigg\},
\end{aligned}$$

where R and $s(v)$ are defined in (4) and (9), respectively.

The solution requires numerical simulations.[8] For the sake of comparison, we calibrate our model as SGU (2004a). The time unit

[8]These are obtained implementing SGU (2004b) second-order approximation routines.

is meant to be a year; the subjective discount rate $\beta = 0.96$ is consistent with a steady-state real rate of return of 4 percent per year. Transaction cost parameters A and B are set at 0.011 and 0.075, the debt-to-GDP ratio is set at 0.44 percent, the benchmark level for the public-consumption-to-GDP ratio is 0.20, the gross price markup is 1.2, and the annualized Rotemberg price adjustment cost is 4.375 (this implies that firms change their price on average every nine months; see SGU 2004a, p. 210). The preference parameter η is set so that in the flexible-price steady state households allocate 20 percent of their time to work when public transfers are nil.

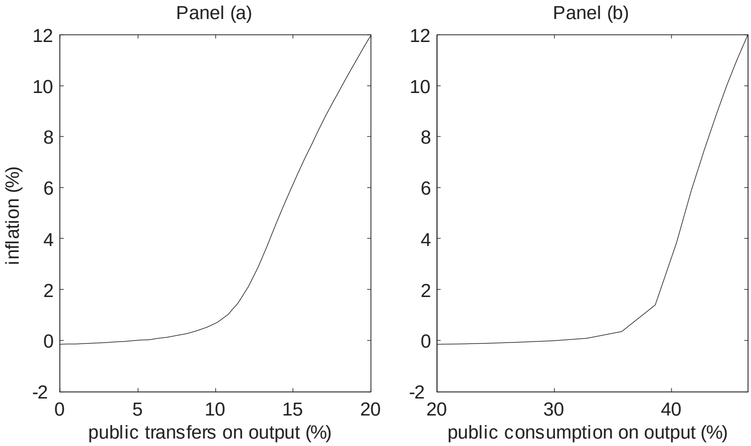
In figure 1 we describe the steady-state optimal inflation response to the transfer increase and to a corresponding variation in public consumption in addition to the benchmark 20 percent value. Both public consumption and transfers are defined as GDP ratios: $g_{PC} = g_t/y_t$ and $g_{PT} = t_t/y_t$. Simulations show that steady-state inflation rapidly increases when g_{PT} grows beyond the 8 percent threshold. For instance, the optimal inflation rate is close to 3 percent when g_{PT} is 10 percent, and exceeds 12 percent when the transfer ratio is 20 percent. When public expenditure is confined to public consumption, a 40 percent public-consumption-to-GDP ratio is associated with a 2 percent optimal inflation rate, and optimal inflation monotonically grows up to 12 percent as the public consumption share reaches 47 percent.[9]

Note that when different public consumption and transfers levels induce the Ramsey planner to choose identical inflation rates, we also obtain identical consumption, labor market, and inflation wedges, $s(v)$, $\frac{1+s(v)+vs'(v)}{(1-\tau)}$, and $\frac{\xi_p}{2} l\,(\pi - 1)^2$, respectively (figure 2). It is also interesting to note that when either g_{PT} or g_{PC} reach the levels which trigger the optimality of positive inflation, the optimal policy generates an almost identical consumption pattern. In both cases, abandoning price stability allows to stabilize consumption in spite of the increasing burden of fiscal revenues.

It is interesting to compare our results with the interpretation of the inflationary outcome generated by the need to finance transfers offered by SGU (2006, p. 385). They claim that when the private

[9]As pointed out above, in our model untaxed monopolistic profits are necessary to generate the planner's incentive to inflate. For instance, if one sets $\mu^p = 1.1$, the optimal inflation rate remains close to zero for $g_{PT} \leq 15$ percent.

Figure 1. Public Expenditure Components and Optimal Inflation

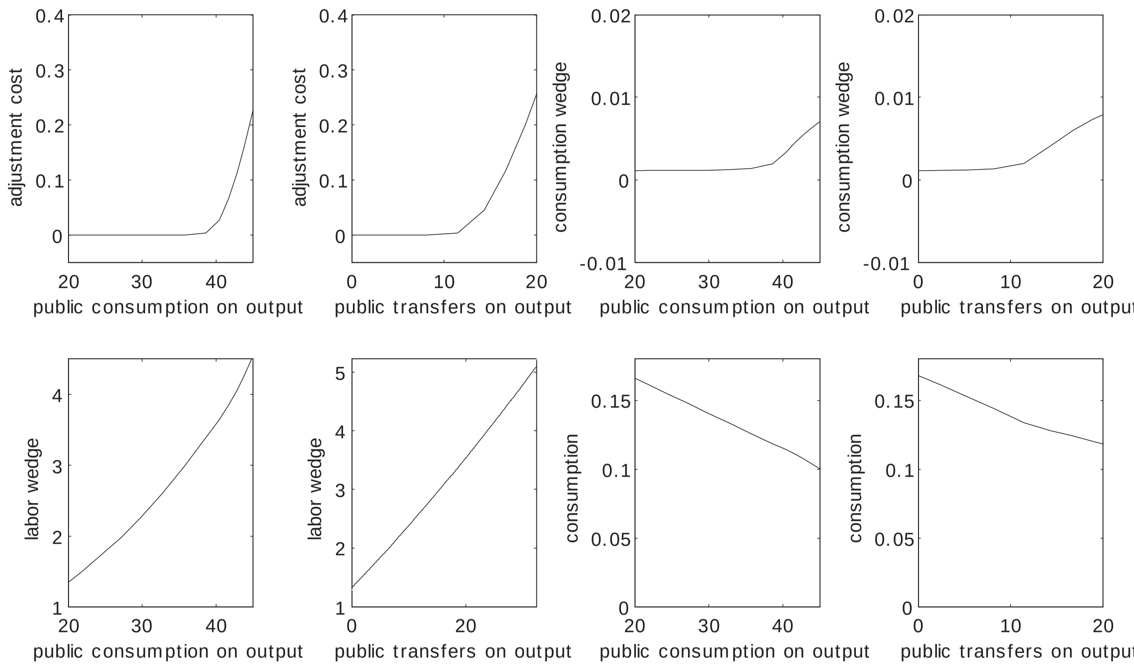


sector must receive an exogenous amount of after-tax transfers, it is optimal to exploit the inflation tax on money balances in order to impose an indirect levy on the (transfers-determined) source of household income. Given our finding that relatively large levels of public consumption exist such that the planner chooses identical inflation rates, this intuition must be incorrect.

One mechanism driving the choice of the optimal policy mix might be related to the distortionary taxation necessary to finance the additional transfers, which adversely affects the labor supply and reduces the tax base, whereas the increase in public consumption generates a negative wealth effect that triggers a positive labor supply response and expands the tax base. In this case, the incentive to increase inflation should be much reduced. To check the importance of the wealth effect of g_{PC} on the labor supply, we solved the Ramsey problem under a different specification of the utility function, such as the Greenwood-Hercowitz-Huffman (1988; GHH henceforth) preferences,

$$ {}_{GHH}u\left(C_{t,i}, l_{t,i}\right) = \frac{\left(C_{t,i} - \eta l_{t,i}^{1+\phi}\right)^{1-\sigma}}{1-\sigma}. \tag{18} $$

Figure 2. Policy Wedges and Consumption



Notes: The wedges are computed as follows: Price adjustment cost is (12). The consumption wedge is (4). The labor wedge is divided by λ.

Under (18) the marginal rate of substitution $-\frac{u_l(c_t,l_t)}{u_c(c_t,l_t)} = \eta l_{t,i}^{\phi}$ is independent of consumption, i.e., there is no wealth effect on the labor supply, and the labor market equilibrium condition in steady state is $\eta l^{\phi} = w$. Simulations contradict our conjecture. In fact, under GHH preferences we obtain almost identical inflation rates in response to the variations in g_{PC} and g_{PT} considered in figure 1. In particular, an increase in public consumption is met by an expansion in the labor tax, whereas inflation remains very close to zero unless public consumption is relatively large.[10]

3.2 The Ramsey Solution in a Simplified Model

Further insights can be obtained by imposing restrictions on some parameter values, which allow to simplify the Ramsey solution in the steady state.[11] To begin with, we set $\beta = 1$, $\xi_p = 0$. In this case, the Friedman rule is satisfied for $\pi = 1$ and price adjustment frictions do not matter, restricting the policymaker's trade-off to two dimensions: the Friedman rule calls for complete price stability, whereas the public finance motive calls for positive inflation because the tax system is incomplete. From (13) it is easy to see that in steady state $w = \rho$ irrespective of the inflation rate. We also assume that steady-state debt is nil, and set $B = 0$ in (4). Under this latter assumption[12] we obtain $s(v) = Av$ and, since $R = \frac{\pi}{\beta}$, $v = \left(\frac{R-1}{AR}\right)^{\frac{1}{2}} = \left(\frac{\pi-1}{A\pi}\right)^{\frac{1}{2}}$.[13]

By setting $g_{PC} = 0.2$ and $g_{PT} = 0$, the optimal inflation rate in the simplified model is $\pi = 1.3$ percent, whereas for the full model we obtained $\pi = -0.16$ percent. This result is obviously due to the assumed reduction in inflation costs, but our focus here is on obtaining a better understanding of the reason why similar levels of public transfers and public consumption are associated with different optimal inflation rates in steady state. In this regard, note that a 5 percent increase in g_{PC} now is matched by $\pi = 1.8$ percent, whereas an identical variation in g_{PT} is associated with $\pi = 2.6$ percent.

---

[10] Results are available upon request.

[11] See the appendix for a derivation of our results.

[12] With $B = 0$, the model is characterized by a standard Tobin money demand.

[13] The result described in this section can also be obtained by removing the assumptions $\xi_p = B = 0$, but in this case the algebra is rather cumbersome and it is more difficult to support the intuition.

In this simplified model, the steady-state Ramsey solution is characterized by $\lambda^{Ph} = \lambda^{MUC} = 0$ and $\lambda^{B} = -\lambda_t^{GBC} \frac{c}{v\pi}\frac{1}{\lambda}$. This latter condition implies that the marginal effect of inflation on the Euler equation constraint must equal the marginal effect of π on the government budget constraint. Further, we obtain that the marginal effect of money velocity on the aggregate resource constraint must equal its marginal effect on the government budget constraint, i.e.,

$$\lambda^{AR} c s'(v) = \lambda^{GBC} \left[\frac{R'(v)}{v\pi^2} - \frac{\pi - 1}{v^2 \pi} - \frac{2\delta A l}{1 - l} \right] c. \tag{19}$$

Finally, the solution for λ^{GBC} is

$$\lambda^{GBC} = u_c(c, l) \left[\left(\frac{R'(v)}{v\pi^2} - A - \frac{2\delta A l}{1 - l} \right) \frac{1 + s(v)}{s'(v)} + \frac{1}{v}\frac{\pi - 1}{\pi} - \frac{\delta l \gamma(v)}{1 - l} \right]^{-1}. \tag{20}$$

The Ramsey planner's choice of c takes into account effects on the marginal utility of consumption; on the aggregate resource constraint, $\lambda^{AR}[1 + s(v)]$; and on the government budget constraint, where $\frac{1}{v}\frac{\pi-1}{\pi}$ and $-\frac{\delta l \gamma(v)}{1-l}$ define consumption effects on revenues from inflation and from labor taxes, respectively. In a sense, just like monetary transaction costs drive a wedge between the marginal utility of consumption and the marginal utility of wealth in the representative-household first-order condition (6), here monetary transaction costs and the need to enforce distortionary taxation drive a wedge between the Ramsey planner's marginal utility of consumption and the marginal utility of revenues.

The optimal labor supply condition is

$$-u_h(c, l) = \lambda^{AR} + \lambda^{GBC} \left(\rho - \frac{\delta c \gamma(v)}{(1 - l)^2} \right) \tag{21}$$

$$= \left(\frac{R'(v)}{A v \pi^2} - 1 - \frac{2\delta l}{1 - l} + \rho - \frac{\delta c \gamma(v)}{(1 - l)^2} \right) \lambda^{GBC}. \tag{22}$$

The right-hand side of (21) accounts for the marginal effects of l on on the aggregate resource constraint (23), which is proportional to the multiplier λ^{AR}, and on the government budget constraint (24), which is determined by the multiplier λ^{GBC} and by the marginal effect of l on tax revenues, $\left(\rho - \frac{\delta c \gamma(v)}{(1-l)^2}\right)$. This is the Ramsey planner's equivalent of the representative-agent first-order condition (8).

Using (19), (20), and the explicit functional forms for v, $s(v)$, and $s'(v)$, the Ramsey planner's problem collapses to the following conditions:

$$c = \frac{1-g_{PC}}{1+\left(A\frac{\pi-1}{\pi}\right)^{\frac{1}{2}}} l \tag{23}$$

$$\frac{c}{lv}\frac{\pi-1}{\pi} + \left\{\rho - \frac{\delta c}{1-l}\left[1+2\left(A\frac{\pi-1}{\pi}\right)^{\frac{1}{2}}\right]\right\} = g_{PC} + g_{PT} \tag{24}$$

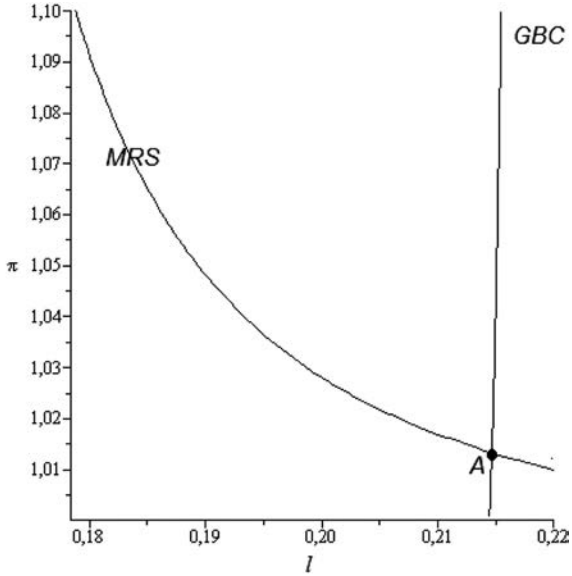
$$\frac{\delta c}{1-l} = \frac{\frac{1-\pi^2}{\pi^4} - \frac{\delta l}{1-l} + \frac{\rho}{2} - \delta c\left(\frac{1}{2} + \left(A\frac{\pi-1}{\pi}\right)^{\frac{1}{2}}\right)/(1-l)^2}{\left(\frac{1-\pi^2}{\pi^4} - \frac{\delta l}{1-l}\right)\left[1+\left(A\frac{\pi-1}{\pi}\right)^{\frac{1}{2}}\right] + \frac{1}{2}\left(A\frac{\pi-1}{\pi}\right)^{\frac{1}{2}} - \frac{\delta l\left(\frac{1}{2}+\left(A\frac{\pi-1}{\pi}\right)^{\frac{1}{2}}\right)}{1-l}}. \tag{25}$$

Conditions (23) and (24), respectively, are the aggregate resource and government balance constraints that the Ramsey planner solution must satisfy. Condition (25) simply rearranges (21). It is easy to see that changes in g_{PC} and g_{PT} would generate identical variations in consumption hours and inflation if public consumption did not enter the aggregate resource constraint (23). Therefore, the smaller incentive to inflate that we observed in response to a public consumption variation is due to its contemporaneous effects that operate through the aggregate resource constraint and impact on (i) inflation- and labor-tax revenues, i.e., on the left-hand side of (24), and (ii) the planner's desired marginal rate of substitution in (25).

In figure 3 we present a graphical solution. Substituting for c from (23) into (24), we obtain the *GBC* schedule which defines combinations of l and π that are consistent with a balanced government budget constraint for given values of ρ, g_{PC}, and g_{PT}. It is upward sloping[14] because an increase in employment can be obtained through a labor-tax reduction. This, in turn, requires an increase in

[14]It is very steep for the small inflation range (1.00–1.10) used in figure 3.

Figure 3. Graphical Illustration of the Ramsey Equilibrium

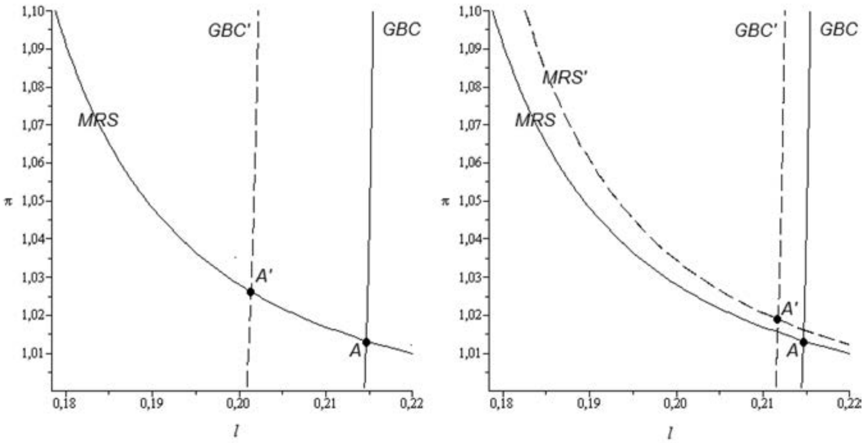


inflation in order to compensate for the revenue loss. By substituting (23) into (25), we obtain the *MRS* schedule which defines combinations of l and π such that the Ramsey planner's desired marginal rate of substitution obtains for given values of ρ and g_{PC}. It is downward sloping[15] because an increase in l brings the consumption-to-labor ratio below its desired level. A fall in inflation is therefore necessary to raise desired relative consumption. The *MRS* and *GBC* schedules are plotted by assuming $g_{PC} = 0.2$ and $g_{PT} = 0$. The corresponding Ramsey equilibrium is then point A.

Figure 4 describes the effects on the Ramsey equilibrium of a 5 percent increase in public expenditure. Specifically, panel A

[15]Due to its complex functional form, the slope of (25) is not trivial. Note that, in addition to g_{PC}, (25) includes only two calibrated parameters, A and δ, which take values 0.011 and 2.9 as in SGU (2004a). We experimented for values of A and δ in the ranges 10^{-4}–10 and 0.5–8, respectively. In all cases, we obtained a downward-sloping *MRS* schedule around realistic values for l and π (including the Ramsey equilibrium). Note that the different values for δ imply an equilibrium for l between about 0.1 and 0.9. Therefore, we explored the range of all possible values.

Figure 4. The Effects of a Change in Government Expenditures



A. Public transfers variations. B. Public consumption variation.

illustrates the effects of an increase in public transfers, whereas panel B shows the effects of an equivalent increase in public consumption. The common initial equilibrium in the two panels is described by point A, where $g_{PC} = 0.2$ and $g_{PT} = 0$.

Starting from point A in panel A, the 5 percent increase in g_{PT} shifts the GBC locus to the left to GBC' because, holding inflation constant, the increase in the tax rate necessary to balance the budget inevitably reduces employment. As pointed out above, the MRS schedule is not affected by g_{PT} and the new equilibrium A' is characterized by a relatively large increase in inflation.

The effects of a 5 percent increase in g_{PC}, panel B, are more complex. Consider the GBC locus; in this case, for any given value of l, private consumption must fall, causing a twofold effect on government revenues. On the one hand, the reduction in real money holdings lowers inflation-tax proceedings. On the other hand, from (17) we know that for any given value of l, the lower private consumption is associated with larger fiscal revenues. This latter effect unambiguously dominates, limiting the leftward shift of GBC.[16]

[16]This latter effect also helps to explain why incentives to inflate remain limited when the utility function is characterized by GHH preferences.

Turning to the *MRS* locus, we find that an increase in g_{PC} now also causes a rightward shift to MRS'. This happens because for any given level of inflation, the planner seeks a reduction in leisure to partly offset the reduction in consumption determined by the increase in g_{PC}. Thus, relative to the increase in g_{PT}, the shift in *MRS* would cause larger inflation, but this is dominated by the corresponding shift in *GBC*.

4. Optimal Monetary and Fiscal Stabilization Policies

In this section we investigate whether our characterization of steady-state public expenditures also bears implications for the conduct of macroeconomic policies over the business cycle. SGU (2004a) show that, when public transfers are nil, costly price adjustment induces the Ramsey planner to choose a minimal amount of inflation volatility and to select a permanent public debt response to shocks in order to smooth taxes over the business cycle. Benigno and Woodford (2004), who emphasize the complementarity between fiscal and monetary policies, substantially confirm the optimality of near-zero inflation volatility for a plausible degree of nominal price stickiness.

We discuss how the optimal fiscal and monetary stabilization policies[17] change when, in steady state, g_{PT} is 0.1 instead of zero, whereas other fiscal figures are assumed to be 0.44 for debt-to-GDP ratio and $g_{PC} = 0.2$. In table 2 we show that the volatility of both taxes and inflation dramatically increases, whereas the strong persistence of taxes vanishes.[18] Thus, even if we still obtain a unit root in the dynamic process for debt accumulation, a more realistic calibration of fiscal outlays has important implications for the dynamic pattern of fiscal and monetary stabilization policies. To grasp this intuition, consider the impulse response functions to a temporary increase in government purchases (figure 5).[19]

[17] We consider a productivity shock and a public consumption shock. Properties of stochastic processes are described in table 2. We compute the second-order approximation using SGU (2004b) routines (see also SGU 2004a, section 7).

[18] To sharpen the analysis, we assume the shock is not serially correlated.

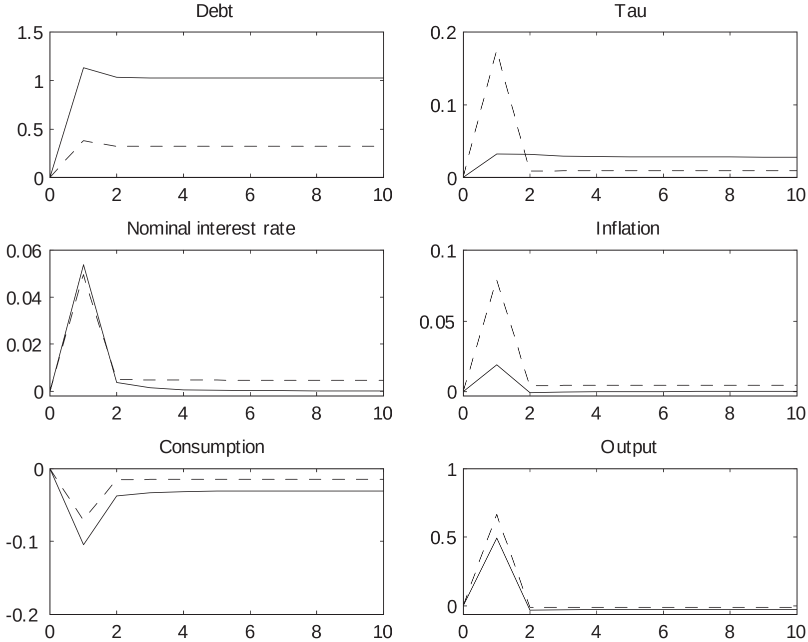
[19] Additional experiments are reported in the working paper version of the paper (downloadable from Ideas). Specifically, there we report the impulse response functions for different composition of the public expenditure and levels of government debt.

Table 2. Dynamic Properties of the Ramsey Allocation (second or approx.)

	Mean	**St. Dev.**	**Auto. Corr.**	**corr(x,y)**	**corr(x,g)**	**corr(x,z)**
			$g_{PT} = 0, g_{PC} = 0.2$			
τ	25.19	1.062	0.759	−0.305	0.436	−0.236
π	−0.16	0.177	0.034	−0.108	0.374	−0.275
R	3.82	0.566	0.863	−0.942	−0.044	−0.962
y	0.21	0.007	0.820	1.000	0.204	0.938
l	0.21	0.003	0.823	−0.085	0.590	−0.402
c	0.17	0.007	0.824	0.940	−0.123	0.954
			$g_{PT} = 0.1, g_{PC} = 0.2$			
τ	42.69	2.860	−0.053	−0.110	0.284	−0.356
π	1.46	0.962	−0.054	−0.062	0.304	−0.309
R	5.50	0.489	0.775	−0.790	0.142	−0.926
y	0.17	0.005	0.823	1.000	0.408	0.884
l	0.17	0.003	0.714	−0.237	0.699	−0.651
c	0.13	0.005	0.783	0.851	−0.091	0.985

Notes: In the table, τ, π, R, y, l, and c stand for the tax rate, inflation rate, nominal interest rate, output, hours, and consumption, respectively.

Figure 5. Fiscal Shock Impulse Response Functions under Different Levels of Public Transfers



Notes: The solid line shows no transfers and the dashed line shows 10 percent of the transfers-to-GDP ratio. The figure shows impulse responses to an i.i.d. government purchases shock. The size of the innovation in government purchases is one standard deviation (a 3 percent increase in g). The shock takes place in period 1. Public debt, consumption, and output are measured in percent deviations from their pre-shock levels. The tax rate, the nominal interest rate, and the inflation rate are measured in percentage points.

Under both scenarios, the permanent debt adjustment allows to smooth tax distortions. However, the different magnitudes of the permanent debt and tax adjustments associated with the two cases ($g_{PT} = 0$ and $g_{PT} = 0.1$) are also evident. When $g_{PT} = 0.1$, the long-run debt adjustment is reduced by 70 percent. In this case, long-run tax and inflation distortions are already relatively large, and the steady-state accumulation of debt in the face of an adverse shock becomes less desirable. Instead, the planner finds it optimal to front-load tax adjustment and to inflate away part of the real value of outstanding nominal debt. In addition, the increase in inflation has a positive impact on seigniorage revenues. This explains the surge in inflation volatility reported in table 2. Our model is also able to match the positive empirical correlation between average inflation and inflation variability.[20] For the sake of fairness, it is worth noticing that inflation volatility still appears to be substantially limited relative to the case of flexible prices, which is the main point of SGU (2004a). Our contribution here is that a substantial complementarity exists between inflation and taxes in response to the public consumption shock.

5. Conclusions

Incompleteness of the tax system is a necessary condition for the existence of a public finance justification for inflation. The strong point of SGU (2004a, 2011) was to argue that irrespective of the incompleteness of the tax system, optimal inflation should be between zero and the Friedman rule.

The point of this paper is that for the same incompleteness of the tax system, a non-negligible inflation rate in steady state is indeed optimal if one adopts a realistic calibration for fiscal outlays, including public transfers. Differently from SGU (2011), who argue that central bank inflation targets are too high, our contribution shows that a 2 percent target might indeed be too low.

However, to obtain an empirically relevant assessment of the optimal inflation rate, the model should be extended to account for a number of country-specific factors, such as governments' ability to optimally tax factor incomes, composition of public expenditures, monetary transaction costs, other frictions such as nominal wage stickiness, and the existence of an informal sector. All this should be done bearing in mind that the tax system incompleteness probably is an inherent feature of modern economies. Similar considerations can be made concerning inflation costs. For instance, Calvo pricing, which implies price dispersion, might generate higher inflation costs than Rotemberg pricing, but one should also take into account inflation indexation and its correlation with the underlying

[20]See, e.g., Friedman (1977), Ball and Cecchetti (1990), Caporale and McKiernan (1997).

inflationary regime, as shown in Fernández-Villaverde and Rubio-Ramírez (2008). All this is left for future research.

Further, our analysis of the optimal fiscal and monetary stabilization policies strengthens the Benigno and Woodford (2004) argument that the two policy tools should be seen as complements and that the monetary authority should consider the consequences of their actions for the government budget. In this regard, we show that a substantial amount of inflation volatility is indeed desirable to deflate nominal debt and to limit the accumulation of real debt in the long run. Our results thus provide theoretical support to policy-oriented analyses which call for a reversal of debt accumulated in the aftermath of the 2008 financial crisis and for a reconsideration of the role of inflation in facilitating debt reductions.

Appendix

The steady-state solution of the Ramsey problem defined in section 3 is characterized by the following set of first-order conditions:

$$l = [1 + s(v)]\, c + g_c l + \frac{\xi_p}{2} l(\pi - 1)^2 \tag{26}$$

$$1 = \beta r(v) \frac{1}{\pi} \tag{27}$$

$$\frac{c}{v} + b + [mc + Z\gamma(v)]\, l = \frac{r(v)b}{\pi} + \frac{c}{v\pi} + (g_{PC} + g_{PT})\, l \tag{28}$$

$$\xi_p(1 - \beta)\pi(\pi - 1) = \frac{mc - \rho}{1 - \rho} \tag{29}$$

$$u_c(c, l) = \lambda\gamma(v) \tag{30}$$

$$u_c(c, l) - \lambda^{AR}[1 + s(v)] + \left[\frac{1}{v}\left(1 - \frac{\beta}{\pi}\right) - \frac{\delta}{1 - l} l\gamma(v)\right] \lambda^{GBC} + \lambda^{MUC} u_{cc} = 0 \tag{31}$$

$$u_l(c, l) + \lambda^{AR}\left(1 - \frac{\xi_p}{2}(\pi - 1)^2\right) + \lambda^{GBC}\left[mc - \left(\frac{\delta c}{1 - l} + \frac{\delta c}{(1 - l)^2} l\right)\gamma(v)\right] = 0 \tag{32}$$

$$\left(1-\frac{r(v)}{\pi}\right)\lambda^{B}+(1-\beta)\frac{\lambda^{Ph}}{\lambda}\pi(\pi-1)-\lambda^{MUC}\gamma(v)=0 \tag{33}$$

$$-\lambda^{AR}s'(v)c-\lambda^{B}\frac{\beta r'(v)\lambda}{\pi}$$
$$-\lambda^{GBC}\left[\left(1-\frac{\beta}{\pi}\right)\frac{c}{v^2}+\frac{\delta c}{1-l}l\gamma'(v)+\frac{\beta b r'(v)}{\pi}\right]$$
$$-\lambda^{MUC}\lambda\gamma'(v)=0 \tag{34}$$

$$-\lambda^{AR}\xi_p(\pi-1)l+\frac{1}{\pi^2}\left[\lambda^{B}r(v)\lambda+\lambda^{GBC}\left(r(v)b+\frac{c}{v}\right)\right]=0 \tag{35}$$

$$\pi=\beta r(v) \tag{36}$$

$$\xi_p\lambda^{GBC}l=-\frac{\lambda^{Ph}}{1-\rho}, \tag{37}$$

where we have expressed public consumption and transfers as GDP ratio (i.e., $g = g_{PC}l$ and $t = g_{PT}l$, recall that $y = l$) and $r(v)=\frac{1}{1-s'(v)v^2}$ from (9).

As said in the main text, we impose $\beta = 1$, $b = 0$, $\xi_p = 0$, and $B = 0$. In that case, $v=\left(\frac{R-1}{AR}\right)^{\frac{1}{2}}=\left(\frac{\pi-1}{A\pi}\right)^{\frac{1}{2}}$, $s(v) = Av$, $s'(v) = A$, and $\gamma(v) = 1 + s(v) + vs'(v) = 1 + 2A^{\frac{1}{2}}\left(\frac{\pi-1}{\pi}\right)^{\frac{1}{2}}$. From (37) we get $\lambda^{Ph} = 0$. Then from equation (33) we get $\lambda^{MUC} = 0$. Equation (29) implies $mc = \rho$. From (35) we obtain $\lambda^{B}\frac{\lambda}{\pi}=-\left[\lambda^{GBC}\frac{c}{v\pi^2}\right]$.[21] Then substitute $\lambda^{B}\frac{\lambda}{\pi}=-\lambda^{GBC}\frac{c}{v\pi^2}$ into (34) to obtain $\lambda^{AR}=\frac{\lambda^{GBC}}{s'(v)}\left[\frac{\rho'(v)}{v\pi^2}-A-\frac{2\delta Al}{1-l}\right]$. Substituting for λ^{AR} in (31), we obtain

$$\lambda^{AR}=\frac{\frac{U_c}{s'(v)}\left[\frac{\rho'(v)}{v\pi^2}-A-\frac{2\delta lA}{1-l}\right]}{\left\{\left[\frac{\rho'(v)}{v\pi^2}-A-\frac{2\delta lA}{1-l}\right]\frac{[1+s(v)]}{s'(v)}+\frac{1}{v}\left(\frac{\pi-1}{\pi}\right)-\frac{\delta l}{1-l}\gamma(v)\right\}} \tag{38}$$

$$\lambda^{s}=u_c(c,l)\left\{\left[\frac{\rho'(v)}{v\pi^2}-A-\frac{2\delta lA}{1-l}\right]\frac{[1+s(v)]}{s'(v)}\right.$$
$$\left.+\frac{1}{v}\left(\frac{\pi-1}{\pi}\right)-\frac{\delta l}{1-l}\gamma(v)\right\}^{-1}. \tag{39}$$

[21]This latter condition implies that the marginal effect of π on savings must equal the marginal effect of π on the government budget constraint.

Then substituting for λ^{GBC}, λ^{AR} into (32), we get

$$\frac{\left[\frac{2}{\pi^2}\frac{1}{\pi^2}-1-\frac{2\delta l}{1-l}\right]+\left\{\rho-\frac{\delta c\gamma(v)}{1-l}-\frac{\delta c\gamma(v)l}{(1-l)^2}\right\}}{\left[\frac{2}{\pi^2}\frac{1}{\pi^2}-1-\frac{2\delta l}{1-l}\right][1+s(v)]+\frac{1}{v}\left(\frac{\pi-1}{\pi}\right)-\frac{\delta l}{1-l}\gamma(v)}=\frac{c}{1-l}. \quad (40)$$

The model is then solved using (26), (28), and (40).

References

Aizenman, J., and N. Marion. 2011. “Using Inflation to Erode the U.S. Public Debt.” *Journal of Macroeconomics* 33 (4): 524–41.

Alonso-Ortiz, J., and R. Rogerson. 2010. “Taxes, Transfers and Employment in an Incomplete Markets Model.” *Journal of Monetary Economics* 57 (8): 949–58.

Altig, D., L. Christiano, M. Eichenbaum, and J. Linde. 2011. “Firm-Specific Capital, Nominal Rigidities and the Business Cycle.” *Review of Economic Dynamics* 14 (2): 225–47.

Angelopoulos, K., A. Philippopoulos, and V. Vassilatos. 2009. “The Social Cost of Rent Seeking in Europe.” *European Journal of Political Economy* 25 (3): 280–99.

Aruoba, S. B., and F. Schorfheide. 2011. “Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-Offs.” *American Economic Journal: Macroeconomics* 3 (1): 60–90.

Ascari, G., E. Castelnuovo, and L. Rossi. 2011. “Calvo vs. Rotemberg in a Trend Inflation World: An Empirical Investigation.” *Journal of Economic Dynamics and Control* 35 (11): 1852–67.

Ball, L., and S. Cecchetti. 1990. “Inflation and Uncertainty at Short and Long Horizons.” *Brookings Papers on Economic Activity* 1: 215–54.

Baumol, W. J. 1952. “The Transactions Demand for Cash: An Inventory Theoretic Approach.” *Quarterly Journal of Economics* 66 (4): 545–56.

Benigno, P., and M. Woodford. 2004. “Optimal Monetary and Fiscal Policy: A Linear-Quadratic Approach.” In *NBER Macroeconomics Annual 2003*, ed. M. Gertler and K. Rogoff, 271–364. Cambridge, MA: MIT Press.

Blanchard, O. J., G. Dell’Ariccia, and P. Mauro. 2010. “Rethinking Macroeconomic Policy.” IMF Staff Position Note No. 2010/03.

Brock, P. L. 1989. “Reserve Requirements and the Inflation Tax.” *Journal of Money, Credit and Banking* 21 (1): 106–21.

Cagan, P. 1956. “The Monetary Dynamics of Hyperinflation.” In *Studies in the Quantity Theory of Money*, ed. M. Friedman, 25–117. Chicago: University of Chicago Press.

Caporale, T., and B. McKiernan. 1997. “High and Variable Inflation: Further Evidence on the Friedman Hypothesis.” *Economics Letters* 54 (1): 65–68.

Chari, V., L. Christiano, and P. Kehoe. 1991. “Optimal Fiscal and Monetary Policy: Some Recent Results.” *Journal of Money, Credit and Banking* 23 (3): 519–39.

Coibion, O., Y. Gorodnichenko, and J. Wieland. 2012. “The Optimal Inflation Rate in New Keynesian Models: Should Central Banks Raise Their Inflation Targets in Light of the Zero Lower Bound?” *Review of Economic Studies* 79 (4): 1371–1406.

Fernández-Villaverde, J., and J. Rubio-Ramírez. 2008. “How Structural Are Structural Parameters?” In *NBER Macroeconomics Annual 2007*, ed. D. Acemoglu, K. Rogoff, and M. Woodford, 83–137. Cambridge, MA: MIT Press.

Friedman, M. 1977. “Nobel Lecture: Inflation and Unemployment.” *Journal of Political Economy* 85 (3): 451–72.

Greenwood, J., Z. Hercowitz, and G. Huffman. 1988. “Investment, Capacity Utilization, and the Real Business Cycle.” *American Economic Review* 78 (3): 402–17.

Guerron-Quintana, P. A. 2009. “Money Demand Heterogeneity and the Great Moderation.” *Journal of Monetary Economics* 56 (2): 255–66.

Khan, A., R. G. King, and A. L. Wolman. 2003. “Optimal Monetary Policy.” *Review of Economic Studies* 70 (4): 825–60.

Kimbrough, K. P. 2006. “Revenue Maximizing Inflation.” *Journal of Monetary Economics* 53 (8): 1967–78.

Oh, H., and R. Reis. 2011. “Targeted Transfers and the Fiscal Response to the Great Recession.” NBER Working Paper No. 16775.

Phelps, E. S. 1973. “Inflation in the Theory of Public Finance.” *Swedish Journal of Economics* 75 (1): 67–82.

Rogoff, K. S. 2010. “Why America Isn’t Working.” Online Project Syndicate Article. Available at http://www.project-syndicate.org/.

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

Schmitt-Grohé, S., and M. Uribe. 2004a. "Optimal Fiscal and Monetary Policy under Sticky Prices." *Journal of Economic Theory* 114 (2): 198–230.

———. 2004b. "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function." *Journal of Economic Dynamics and Control* 28 (4): 755–75.

———. 2006. "Optimal Fiscal and Monetary Policy in a Medium-Scale Macroeconomic Model." In *NBER Macroeconomics Annual 2005*, ed. M. Gertler and K. Rogoff, 383–425. Cambridge, MA: MIT Press.

———. 2011. "The Optimal Rate of Inflation." In *Handbook of Monetary Economics*, ed. B. M. Friedman and M. Woodford, 723–828. Amsterdam: Elsevier.

Sims, C. A. 1994. "A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy." *Economic Theory* 4 (3): 381–99.

Zubairy, S. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *International Economic Review* 55 (1): 169–95.

Risk Shifting with Fuzzy Capital Constraints*

Simon Dubecq,[a] Benoit Mojon,[b] and Xavier Ragot[c]
[a]European Central Bank
[b]Banque de France
[c]CNRS - Paris School of Economics

We construct a model where risk shifting can be moderated by capital requirements. Imperfect information about the level of capital per unit of risk, however, introduces uncertainty about the risk exposure of intermediaries. Over-estimation of the capital held by financial intermediaries, or the extent of regulatory arbitrage, may induce households to wrongly infer from higher asset prices that the fundamentals of risky assets have improved. This mechanism can notably explain the low risk premia paid by U.S. financial intermediaries between 2000 and 2007 in spite of their increased exposure to risk through higher leverage. Moreover, the lower the level of the risk-free interest rate, the more risk is under-estimated.

JEL Codes: G14, G21, E52.

1. Introduction

The goal of this paper is to explain why, in the run-up to the sub-prime crisis, U.S. financial intermediaries were able to pay

*We wish to thank L. Meneau and O. Prato for helpful comments, and for providing information on Basel regulation; V. Acharya, G. Barlevy, J.-P. Benassy, C. Borio, E. Challe, S. Cecchetti, G. Demange, E. Farhi, D. Gale, E. Kharroubi, A. Kashyap, N. Kiyotaki, O. Loisel, A. Martin, H. Pages, F. Portier, J.-C. Rochet, J. Tirole, C. Upper, and M. Woodford and seminar participants at the Federal Reserve Banks of Boston, Chicago, and New York, the BIS, and conference participants at the BIS, CEMFI, and Tilburg University for comments. We wish to express special thanks to C. Hellwig, B. Barsky, F. Velde, the editor Carl Walsh, and an anonymous referee for their suggestions on the structure of the paper, and Allen Monks and Tony Dare for drafting suggestions. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Banque de France, the European Central Bank, or the Eurosystem.
Corresponding author: Benoit Mojon, benoit.mojon@banque-france.fr.

non-increasing risk premia while their leverage increased.[1] To do so, we introduce the capital of financial intermediaries into the risk-shifting model developed by Allen and Gale (2000). In their model, households can invest in risky assets only indirectly, by lending to financial intermediaries. Households require a risk premium on this loan because they anticipate that financial intermediaries will default in the bad state of the world. However, intermediaries that have limited liability will take too much risk. A bubble results, as the price of the risky asset will be higher than in the case in which households, which do not have limited liability, can directly invest in the risky asset.

In our model, the amount of risk taken by households through lending to intermediaries—i.e., the amount and the interest rate at which they lend—will crucially depend on the level of capital held by intermediaries. Households, which cannot observe the degree of risk of the risky asset ex ante, try to infer it from the price of the risky asset and from their assumption of the level of capital held by financial intermediaries.

Our contribution is twofold. First, we show that the patterns of risk premia and leverage ratios observed in the United States between 2000 and 2007 can be understood only if investors under-estimated the intermediaries' incentives to take risks. This is likely when capital constraints are fuzzy, meaning that households can form wrong beliefs on the level of banks' capital. We show how investors may wrongly infer from rising asset prices that the aggregate risk is decreasing, and thus charge a low risk premium on their loans to intermediaries. This will be the case if investors under-estimate the degree of regulatory arbitrage, which allows intermediaries to minimize the capital they pledge on risky assets. As argued by Acharya and Schnabl (2009) and Rochet (2008), one of the reasons why the risk-weighted regulatory capital ratio may be opaque is that intermediaries use off-balance-sheet conduits to "play" the level of capital. Uncertainty about the level of capital of banks then implies uncertainty about the risk characteristics of their assets. We show that investors under-estimate the risk of some assets (and thus charge low risk premia) if they over-estimate the level of capital of

[1] See Acharya and Schnabl (2009), Blanchard (2008), Brender and Pisani (2010), Brunnermeier (2009), and Greenlaw et al. (2008).

intermediaries. This model can therefore show why risk premia did not increase before the crisis, which is one of the most puzzling stylized facts of this period. In addition, the mechanism we formalize might be relevant for other periods of major financial innovation or deregulation, when risk-weighted capital positions are likely to be opaque.

Second, the model points to a risk-taking channel for the impact of low real interest rates: the misperception of risk is greater at lower levels of interest rates. This is because the impact of the interest rate on asset prices is larger when the leverage of financial intermediaries is high. This implies that changes in the level of capital have a larger impact on the price of risky assets and, in turn, on the perception of risk by investors at lower levels of interest rates.[2] We discuss in this paper how this channel may translate into a risk-taking channel of monetary policy.

1.1 Related Literature

This article focuses on the link between the leverage of financial intermediaries, asset prices, and interest rates. It draws on the results of Adrian and Shin (2010) and Geanakoplos (2009), who have highlighted the impact of financial intermediaries' leverage on asset prices. It also provides a theoretical underpinning for the empirical results of Adrian and Shin (2010), Altunbas, Gambacorta, and Marques-Ibanez (2010), Ciccarelli, Maddaloni, and Peydró (2010), Ioanidou, Ongena, and Peydró (2008), Maddaloni, Peydró, and Scopel (2008), and Shin (2010), who showed that accommodative monetary policies leading to low real interest rates are associated with increased risk taking by banks. We hence provide a theory for what Adrian and Shin (2010) and Borio and Zhu (2009) call the "risk-taking" channel of monetary policy.

Among the literature on risk shifting, our paper relates first to the contribution of Allen and Gale (2000), where they showed how limited liability on the part of debt issuers leads to over-investment

[2]In fact, the model we use is real, and the interest rate is also real, rather than nominal. We assume that monetary policy can affect, possibly only temporarily, the level of the real interest rate on the storage asset. Section 4 discusses the impact of monetary policy and the savings glut on the level of real and nominal interest rates in the run-up to the crisis.

in risky assets. Barlevy (2008) proved that risk shifting also implies bubbles within more general frameworks of financial intermediation (i.e., when the formation of financial contracts is endogenous). He also generalized risk shifting to a continuous-time dynamic framework. Challe and Ragot (2011) expand the risk-shifting model to the case in which the supply of loans is endogenous.[3]

Finally, two recent papers developed models on similar issues. Dell'Ariccia, Laeven, and Marquez (2010) developed another model of the risk-taking channel of monetary policy framed in a moral hazard setup for banks' capital. Challe, Mojon, and Ragot (2012) show that the proportion of banks that prefer a risky investment portfolio over a diversified, less risky, one decreases with the level of interest rates.

The paper proceeds as follows. Section 2 documents the stylized facts about the sub-prime crisis. Section 3 presents the model. Section 4 solves the model and shows in turn the implications of different assumptions on households' beliefs on risk-weighted capital. Section 5 discusses robustness of the main conclusions of the paper for alternative specifications of the model. Section 6 reports alternative explanations of low risk premia in the run-up to the crisis. Section 7 concludes.

2. Stylized Facts on the Period Preceding the Sub-Prime Crisis

We underline three major stylized facts from the period that preceded the sub-prime crisis: the U.S. banking sector increased its exposure to credit risk and liquidity risk; the perceived riskiness of U.S. financial intermediaries did not increase; and the effective level of banks' capital was difficult to assess during the period.

[3]It is also important to underline the difference between the risk-shifting literature and the literature on endogenous credit constraints. The latter analyzes how asymmetric information introduces external finance premia and collateral constraints. This literature effectively accounts for the financial accelerator, either in the boom phase, when the rising price of collateral relaxes credit constraints (Kiyotaki and Moore 1997), or in the bust phase, when the collapse in asset prices tightens the credit constraints considerably (Holmstrom and Tirole 1997). However, these models face some difficulties in explaining why there are equilibria with too much credit and over-investment in the risky asset.

2.1 Risk Taking in the U.S. Banking Sector

There is now a consensus view that U.S. financial intermediaries increased their risk exposure during the decade leading up to the crisis. This took the form of an expansion of balance sheets and increased leverage on the part of U.S. investment banks. For instance, the Security and Exchange Commission (SEC) reports that, between 2003 and 2007, the mean leverage ratio (defined as the ratio between overall debt and bank's equity) of the five major investment banks[4] jumped from 22 to 30. Among these five investment banks, only one survived the crisis as a stand-alone institution.

This expansion in the size of banks' balance sheets was accompanied by an increase in "off-balance-sheet leverage," as documented in Acharya and Schnabl (2009). This allowed financial intermediaries to generate higher profits without additional capital, in spite of increased potential future losses: in ex ante terms, the unit of risk borne by each dollar of the U.S. banking system's equity increased markedly.

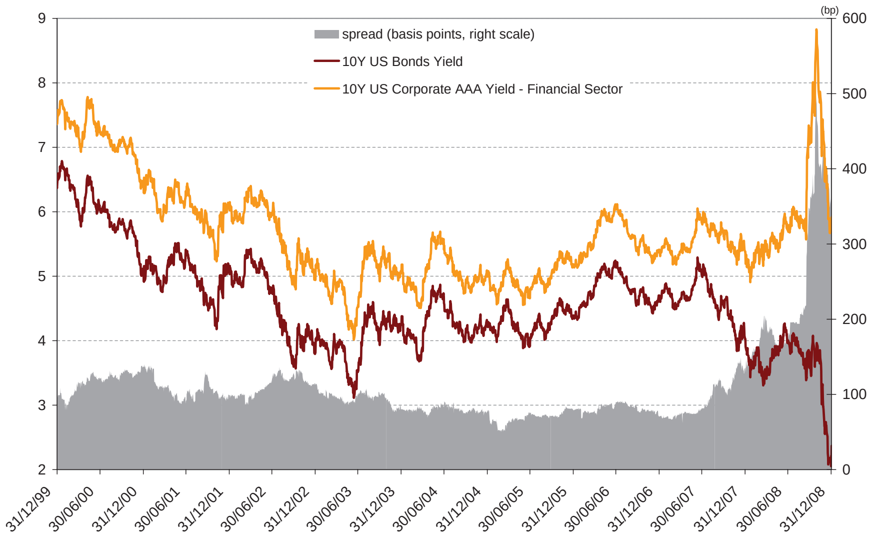
2.2 The Perceived Riskiness of Financial Intermediaries Was Stable

U.S. banks, however, benefited from very low risk premia paid on their debt. The spread between bonds of U.S. financial companies and government bonds (figure 1) shows that the premia paid on banks' default risk did not increase from 2000 to mid-2007. The price of credit risk for banks even declined somewhat between 2002 and 2007.

The change in banks' expected default frequencies (EDFs) is another indicator of the ease with which banks accessed market funding between 2002 and 2007. Banks' EDFs decreased worldwide between 2002 and 2007 (see figure 2), suggesting that market investors either assigned lower probabilities to defaults in the banking sector or required lower risk premia to invest in banks' debt instruments. The same observation that credit risk for banks

[4]Lehman Brothers, Bear Stern, Merrill Lynch, Goldman Sachs, and Morgan Stanley.

Figure 1. Spread between Ten-Year U.S. T-Bonds and Ten-Year Bonds of U.S. AAA Financial Companies



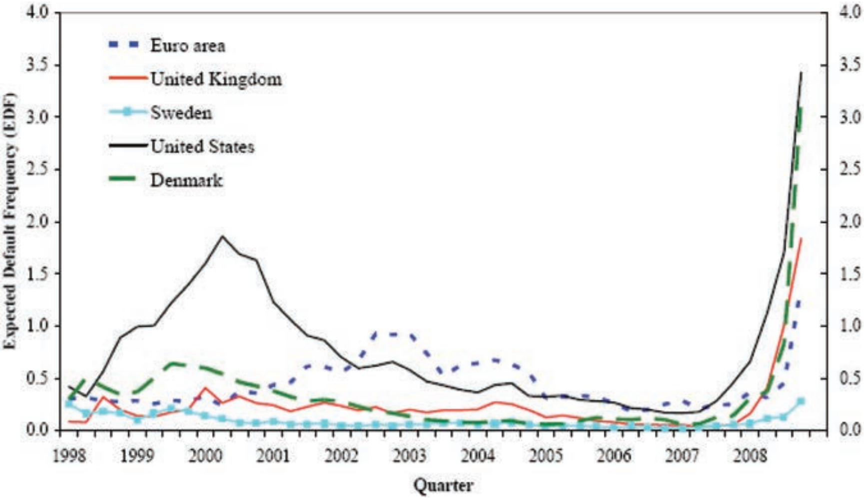
Source: Bloomberg.

was perceived to be negligible can be made by comparing CDS contracts on U.S. banks intermediaries and on issuers in others economic sectors.[5]

To sum up, we observe that, during the years leading up to the crisis, the U.S. banking sector experienced very favorable funding conditions and faced very low risk premia on its debt, while at the same time increasing its leverage (for investment banks) or, more generally, its exposure to credit risk and liquidity risk via off-balance-sheet vehicles. We also note that the increase in banks' assets was concentrated among assets that require very low capital funding. These products, considered to be quite safe by regulatory standards, were however at the root of significant losses for banks after the beginning of the financial crisis.

[5]This is also particularly striking when comparing banks and non-financial corporate credit spreads in the euro area between 2000 and 2007 (Gilchrist and Mojon 2014).

Figure 2. Expected Default Frequency of Banks, Over a One-Year Horizon (averages by country and groups of countries)



Source: Moody's KMV, from Altunbas, Gambacorta, and Marques-Ibanez (2010).

2.3 Changes in Capital Requirements

Several factors explain why capital requirements and capital norms were weak during this period. Most of all, one of the very purposes of Basel II was to authorize banks to reduce their capital base through the use of internal risk models. Blundell-Wignall and Atkinson (2008) and Rochet (2008) highlight the difficulty for outsiders to obtain extensive information on the level of risk borne by financial intermediaries. Such complexity must have led financial intermediaries to minimize risk per unit of risk in line with the vested interest of the industry. Finally, the accounting rules concerning the consolidation of off-balance-sheet entities were singled out by the Financial Stability Forum (2008) for creating "a belief that risk did not lie with arrangers and led market participants to underestimate firms' risk exposures."[6]

[6]This issue is actually on the agenda of the G20, and similar concerns about off-balance-sheet vehicles have been brought up by academics (see Acharya and Schnabl 2009), official regulators, and central bankers (see, for instance, speeches of Christian Noyer and Ben Bernanke in 2008).

3. The Model

There are two dates $t = 1, 2$. The economy comprises four types of agents: households, financial intermediaries, entrepreneurs, and initial sellers. We first describe the assets available to these agents and then their investment decision.

3.1 Assets

Agents make their investment choices at date 1 and get asset returns at date 2. Four financial assets are available in the economy:

(i) A storage asset F, which has a constant return τ. This asset is available in infinite supply.

(ii) A safe asset whose supply X_S is variable, and whose return is r_S. The safe asset will be issued by entrepreneurs who have access to an iso-elastic production function $f(.) = X^{1-\eta}/(1-\eta), \eta < 1$.

(iii) A risky asset in fixed supply X_R, whose return is R^*. R^* equals R with probability π and 0 with probability $(1-\pi)$, which is the level of "economic risk" in the model. The price of the risky asset in period 1 is denoted as P. The assumption of fixed supply simplifies the model and is considered to be the benchmark case. It is relaxed in section 5.1, as a robustness check.

Moreover, we make the following technical assumption, which relates the concavity of the production function η and the extent of economic risk $1-\pi$:

$$\eta > \frac{1-\pi}{\pi}. \tag{1}$$

This assumption, which ensures the uniqueness of the equilibrium, is satisfied for reasonable values of the parameters.[7]

---

[7]In equilibrium we will obtain $f'(X_S) = r_S$ or $X_S = r_S^{-\frac{1}{\eta}}$. If η is too small, the volume (and the share) of safe assets held by financial intermediaries is quite sensitive to interest rates. Both the riskiness and the ex ante return of the entire

(iv) Debt B issued by financial intermediaries and acquired by households.

Financial assets in this economy can be interpreted in the following way:

- The storage asset may, for example, include deposit facilities at the central bank or cash. Indeed, it allows agents to invest without limit at a low and constant rate. In what follows, we will use the return on the storage asset as a proxy for the interest rate set by the monetary policy authority.
- The safe asset represents investment-grade bonds. It can be interpreted as a loan to the "real" sector in order to finance investment or production.
- Finally, the risky asset encompasses all types of investments whose expected returns are higher than the return on the safe asset. It can be either real estate mortgages, junk bonds, or stocks.

3.2 Agents

3.2.1 Financial Intermediaries

There is a unit mass of financial intermediaries (which we also designate as "banks") that are risk neutral and receive an endowment W^f at the beginning of date 1. Agents maximize their consumption over the two periods with a discount factor β, such that

$$\beta < 1/\tau. \tag{2}$$

This assumption implies that the intermediaries are comparatively impatient that they want to borrow in period 1. In addition, they enjoy a private benefit U from being intermediaries. This benefit guarantees that these agents agree to operate as intermediaries rather than consuming all their endowment in period 1. They thus seek to maximize $c_1^f + \beta E\left[c_2^f\right] + U$, where c_1^f and c_2^f are the period 1

portfolio could decrease when r_S decreases. It implies that the ex ante return on the whole portfolio could increase when r_S decreases, which would generate multiple equilibria. (1) is a sufficient condition relating the curvature η and the risk π to ensure that it is not the case.

and period 2 consumption levels. Financial markets open in period 1 after goods markets. This implies that financial intermediaries can bring to financial markets as equity only K of their wealth that is not consumed in period 1:

$$c_1^f \leq W^f - K. \tag{3}$$

Financial intermediaries can invest in all existing assets. They do not invest in the storage asset, because they have access to the safe asset, which yields a higher return. Thus, their balance sheet is composed of a risky asset, PX_R, and a safe asset, X_S, on the asset side, whereas their liabilities are either equity, K, or debt, B. The amount K stands for the fraction of resources invested by the intermediaries themselves in their business. The resource constraint of financial intermediaries is

$$PX_R + X_S = B + K. \tag{4}$$

We assume that financial intermediaries are subject to a norm of "risk coverage" or "risk-weighted capital requirements" by either financial regulation or market discipline.[8] They have to invest from their endowment at least Δ per unit of risky asset:

$$K \geq \Delta PX_R. \tag{5}$$

Following Allen and Gale (2000), we assume that financial intermediaries raise some funds using debt contracts, and that households, who lend to them, are not able to fully observe the investment decisions of financial intermediaries. Hence they will demand the same interest rate r irrespective of the size of the loan they grant to the financial intermediary.

Financial intermediaries can default on their debt. Default occurs when the intermediary's wealth is negative in period 2. Their second-period consumption c_2^f is thus

$$c_2^f \leq \max\{R^* X_R + r_S X_S - rB, 0\}. \tag{6}$$

[8]Indeed, the level of capital requirements need not exclusively be the one set by regulators. It can also be the market norm on the acceptable level of capital for a given level of risk taking.

They therefore choose their debt level B, the equity K, the composition of their portfolio (X_S, X_R), and their consumption profile c_1^f and c_2^f to solve the following program:

$$\max_{K,B,X_R,X_S,c_1^f,c_2^f} c_1^f + \beta E\left[c_2^f\right] + U$$

subject to the conditions (4)–(6).

3.2.2 Households

There is a unit mass of households who are risk neutral and who are uniformly distributed on the $[0,1]$ interval. Each household j receives an endowment W_j^H at the beginning of date 1. To simplify the algebra, and without any loss of generality, households maximize their date 2 consumption only.

As in Allen and Gale (2000), we introduce the following form of market segmentation. Households cannot invest directly in the risky asset or in the safe asset, and they can only either invest in the storage asset or lend to financial intermediaries an amount B at the interest rate r. This assumption captures the advanced skills and accumulated rents (asset-management abilities, private information, and so on) needed to trade sophisticated financial products.

Households do not know the level of risk in the economy, summarized by π. Moreover, they cannot observe the composition of the liabilities of financial intermediaries, K and B. This assumption reflects the inability of each household to observe the extent of the total indebtedness of financial intermediaries. As argued above, the rationale for this assumption is the complexity of their liability structure due to off-balance-sheet liabilities. Technically, households know their portfolio, but they do not observe the portfolio of other households and we do not restrict our analysis to symmetric equilibria.

The key assumption of our model is that households do not know the true value of capital requirement Δ, but their belief Δ^H may differ from Δ. This assumption is meant to reflect that households, investors, or even rating agencies have a hard time assessing the degree of risk effectively borne by financial intermediaries. This opaqueness can be due to regulatory arbitrage or financial innovations. In brief, the information set of households comprises prices and both risky and safe assets held by financial intermediaries (X_R and X_S). However, households cannot observe the composition of banks' liabilities (K and B).

Each household j chooses the composition of its financial portfolio in order to maximize its consumption:

$$\max_{F_j, B_j^H, c_j^H} \beta E\left[c_j^H\right] \tag{7}$$

$$F_j + B_j^H \leq W_j^H \text{ (at date 1)} \tag{8}$$

$$c_j^H \leq \rho B_j^H + \tau F_j \text{ (at date 2),} \tag{9}$$

where $E[\cdot]$ is the expectation operator and β is the discount factor. The expectations are formed on R^* for a given value Δ^H. In the budget constraints, F_j is the amount invested in the storage asset, and B_j^H is the amount lent to intermediaries. We further denote $B^H = \int_j B_j^H dj$, $F = \int_j F_j dj$, $W^H = \int_j W_j^H dj$, and $c = \int_j c_j^H dj$ the aggregate amount of loans to intermediaries, the aggregate investment in the storage asset, the aggregate households endowment, and the aggregate households consumption, respectively. The stochastic interest rate ρ that all households receive ex post on their loans to financial intermediaries is uncertain and depends on the probability of default of intermediaries. Intermediaries default in the bad state of the world, in which the return on the risky asset is 0, because their debt burden will be greater than their portfolio invested in the safe asset. In the event of default, households receive the residual value of the portfolio of financial intermediaries $r_S X_S$, so that ρ in case of default is $r_S X_S / B$. When intermediaries do not default, households get the return r.

$$\rho = \begin{cases} r \text{ if no default, with probability } \pi \\ \frac{r_S X_S}{B} \text{ if default, with probability } (1-\pi) \end{cases}$$

3.2.3 Entrepreneurs and Initial Sellers

There is a unit mass of entrepreneurs who maximize period 2 consumption, denoted as c^e. They have no wealth and they need to borrow in period 1 to produce in period 2. Their production function is $f(X) = X^{1-\eta}/(1-\eta)$. They borrow an amount denoted as

X_S^e from financial intermediaries at a rate r_S, to maximize their period 2 consumption $c^e = f\left(X_S^e\right) - r_S X_S^e$. This maximization over X_S^e yields the simple relationship

$$r_s = f'\left(X_S^e\right). \tag{10}$$

Initial sellers are agents who sell risky assets to intermediaries at period 1, consume, and leave the economy. These agents are only introduced as a simple way of creating a supply of the risky asset, and thus an observable asset price, at the beginning of period 1. Initial sellers have no choices to make and simply consume in period 1 the amount obtained from the sale of the risky asset:

$$c^i = PX_R.$$

In the benchmark model, the quantity of the risky asset is fixed and equal to 1. We relax this assumption below.

3.2.4 Equilibrium

For given parameters, and a given value of households' belief Δ^H, an equilibrium of this economy is a set of prices r, r_S, P and quantities $F_j, B_j^H, c_j^H, K, B, c_1^f, c_2^f, X_S, X_R, X_S^e$, and a risk assessment by households π^H, such that (i) quantities solve the program of all agents at given prices and given households' belief Δ^H, and (ii) markets clear $X_R = 1$, $X_S = X_S^e$ and $B = B^H$.

3.3 Pareto-Efficient Equilibria

We first derive the set of Pareto-efficient allocations. In order to do this, we maximize a general welfare function, which weights the utility of the four types of agents. This can be written, with obvious notations for the Pareto weights, as follows:

$$W = \omega^H \beta E\left[c^H\right] + \omega^f\left(c_1^f + \beta E\left[c_2^f\right]\right) + \omega^i c^i + \omega^e \beta c^e \tag{11}$$
$$\text{with } \omega^H, \omega^f, \omega^i, \omega^e > 0 \text{ and } \omega^H + \omega^f + \omega^i + \omega^e = 1.$$

Expectations in the objective function are only taken with respect to the economic risk, $R^* = R$ with probability π and $R^* = 0$ with probability $1 - \pi$. The feasibility constraints are

$$W^f + W^H = c_1^f + F + X_S + c^i \tag{12}$$

$$\tau F + f(X_S) + RX_R = c^H + c_2^f + c^e. \tag{13}$$

As $\tau < 1/\beta$, forming the Lagrangian for the maximization of (11) subject to the constraints (12) and (13), one can check that the solution has the following properties:

$$F = 0 \text{ and } f'(X_S) = \frac{1}{\beta}.$$

The allocation of the central planner can be achieved in the decentralized economy if we remove any market segmentation and allow for lump-sum transfers. In this case, households can directly lend to entrepreneurs and buy the safe asset. In this equilibrium, the interest rate on the safe asset is $1/\beta$ and the price of the risky asset is equal to its fundamental value $P^* = \beta\pi R$.

4. Model Solution

4.1 Asset Prices and Households' Beliefs

In this section we derive the price of the risky asset by solving the program of financial intermediaries, and provide the intuitions for the main results of the paper.

We solve the program of financial intermediaries under two assumptions. The first is that the capital requirement constraint is always binding, hence $K = \Delta PX_R$. This case is, of course, the one of interest for this model. The capital norm is binding if financial intermediaries are sufficiently impatient, i.e.,

$$\pi r < 1/\beta. \tag{14}$$

This inequality stipulates that the expected cost of the debt πr (because the debt is repaid only outside the bad state, which occurs with probability π) must not be too high. If the expected cost of the debt is too high, intermediaries would want to invest all of their wealth to decrease their expected debt burden, and the capital norm constraint would therefore not bind. As r is determined in equilibrium, we show below that the condition (14) is fulfilled for a wide range of parameter values.

The solution of the program of intermediaries yields the equilibrium price of the risky asset:

$$P = \frac{\beta \pi R}{\Delta + \beta \pi r (1 - \Delta)}. \tag{15}$$

This asset price equilibrium is the main equation of the model. First note that when there is no capital requirement ($\Delta = 0$), the price is simply $P = R/r$, which is the case studied by Allen and Gale (2000).[9] As intermediaries default in the bad state, their demand for the risky asset is always higher than under the first-best equilibrium. Indeed, since $\pi r < 1/\beta$, one finds $P > P^*$. Asset prices are thus too high. Second, when capital requirements increase, the price of the risky asset decreases. Taking r as given, increasing Δ implies a cost in the form of additional foregone consumption in period 1, an effect that dominates the reduction in size of the loan that needs to be repaid with probability π.

Thus, in partial equilibrium, the price of the risky asset can increase for two reasons: either because π increases, which means that the expected return of the risky asset is higher, i.e., "fundamentals are better," or because Δ decreases (the amount of ex ante risk shifting increases).

Maximization with respect to the demand for the safe asset X_S implies that the funding cost of financial intermediaries is equal to the return on the safe asset, as in Allen and Gale (2000): $r = r_s$. This is necessary and sufficient in order to avoid infinite riskless profit opportunities on the part of financial intermediaries, while guaranteeing a positive demand in equilibrium.

The demand for the safe asset implies

$$f'(X_S) = r_S = r \Longrightarrow X_S = [f'(r)]^{-1}. \tag{16}$$

[9]In their model, Allen and Gale show how incomplete debt contracts limit debtors' losses in the bad state of the world (losses fall on lenders). In other words, debt contracts act as call options for borrowers. This implies that borrowers only focus on the good state of the world when deciding the composition of their portfolio: the share of the portfolio at risk is higher and the price of risky assets is inflated above its level in a world without segmentation or incomplete contracts.

The previous equalities are valid irrespective of the beliefs on the part of households about the economic environment. These beliefs will, however, determine the interest rate charged by households.

The basic assumption of the model is that households infer the probability of default from their observation of the risky asset price and their belief on risk-weighted capital Δ^H. The price of the asset is given by (15) for the true value of Δ and π, because it results from a no-arbitrage condition for intermediaries, who know the real value π and the real Δ. Households deduce a value of π^H that is consistent with price P and their belief Δ^H. It therefore follows that

$$P = \frac{\beta\pi R}{\Delta + r\beta\pi\left(1-\Delta\right)} = \frac{\beta\pi^H R}{\Delta^H + r\beta\pi^H\left(1-\Delta^H\right)}. \tag{17}$$

We deduce the following inference for households:

$$\pi^H = \pi\frac{\Delta^H}{\Delta + r\beta\pi\left(\Delta^H - \Delta\right)}. \tag{18}$$

Due to the condition $r\beta\pi < 1$, if $\Delta^H > \Delta$, then $\pi^H > \pi$, and if $\Delta^H < \Delta$, then $\pi^H < \pi$. In addition, if $\Delta^H = \Delta$, then $\pi^H = \pi$. In other words, if households over-estimate risk-weighted capital, they under-estimate the risk, and if they under-estimate risk-weighted capital, they over-estimate the risk. Moreover, when households have correct beliefs concerning risk-weighted capital (i.e., $\Delta^H \equiv \Delta$ irrespective of the value of Δ), they correctly infer the right level of aggregate risk.

We now solve the model for the various cases concerning the relationship between Δ^H and Δ.

4.2 Symmetric Information Over Δ

If both households and financial intermediaries have correct beliefs on Δ, households can deduce the level of aggregate risk π, as shown in the discussion of equation (18). In addition, knowing Δ, they deduce $K = \Delta P X_R$. They can also infer the amount of aggregate debt B from the budget constraint of financial intermediaries.

With the expression for X_S given by (16) and $r = r_S$, the no-arbitrage condition for household can be written as

$$\pi r + (1 - \pi) \frac{r f'^{-1}(r)}{B} = \tau. \tag{19}$$

This condition states that average return for each unit invested in financial intermediaries (taking into consideration the possibility of default) should be equal to the return on the storage technology. We introduce the main conclusions of the paper as propositions. All proofs are presented in the appendix.

PROPOSITION 1. *If* $\Delta^H \equiv \Delta$, *households' expectations of the aggregate risk are correct* ($\pi^H = \pi$) *and*

$$\frac{\partial B}{\partial \Delta} < 0 \textit{ and } \frac{\partial (r - \tau)}{\partial \Delta} < 0.$$

Proposition 1 states that decreasing capital per unit of the risky asset increases both the volume of debt of intermediaries and the credit risk premium, $r - \tau$. This result is due to two effects. First, the overall general equilibrium effect of a decrease in Δ is an increase in the intermediaries' debt level, as financial intermediaries have a greater incentive to increase their exposure to risk by issuing debt. Second, when Δ decreases, households understand that the residual value of the assets they receive in the event of default decreases. They hence request a larger default risk premium $r - \tau$ to compensate for the increased cost of default. This version of the model is therefore *not* consistent with the stylized facts of the sub-prime cycle. As shown in figure 1, banks have been able to borrow at lower risk premia during the five years leading up to the crisis, in spite of increasing leverage and decreasing risk-weighted capital with respect to, for instance, U.S. housing loans.

To summarize, the change in Δ can account for an increase in the debt level of banks, but it cannot explain the path of the risk premia between 2000 and 2007. We therefore assert that risk shifting, per se, is not sufficient to replicate the stylized fact of the *sub-prime crisis*. Before the crisis, banks and financial intermediaries benefited in fact from extremely favorable funding conditions, which would not be the case if changes to risk-weighted capital were fully understood by investors.

The next sections will discuss the more likely cases where capital constraints are fuzzy. In view of the complexity of financial intermediaries' balance sheets, and off-balance-sheets transactions, this case is more likely to reflect the real world. Households may either over-estimate or under-estimate Δ.

4.3 Over-Estimation of Δ

We now assume that households believe that the level of risk-weighted capital is higher than that actually faced by financial intermediaries: $\Delta^H > \Delta$. In this case, the amount of capital pledged by financial intermediaries is lower than that expected by households, and households over-estimate the probability of success of the risky asset $\pi^H > \pi$, due to the relationship (18).

Households form their inference about the residual value of their portfolio, $\frac{rX_S}{B^H}$, in the following way. First, from the observation of the amount of risky assets in the economy X_R, and from their belief about Δ^H, households infer that the level of the capital of financial intermediaries is

$$K^H = X_R \Delta^H P.$$

Second, from the budget constraint of financial intermediaries, households form the following expectation about the amount of debt:

$$B^H = X_S + PX_R \left(1 - \Delta^H\right). \tag{20}$$

Third, the no-arbitrage condition for households must now be written according to their expectations:

$$\pi^H r + \left(1 - \pi^H\right) \frac{rX_S}{B^H} = \tau. \tag{21}$$

Using equation (18) in order to substitute for π^H, the expressions for X_S given by (16), the value of B implied by the balance sheet constraint of the intermediary, and the fact that $r_S = r$, we obtain an equation for the equilibrium interest rate r which depends only on known parameters and functional forms.

In order to obtain analytical insight, we assume that households' belief about capital requirement is not too far away from the true one, i.e., we assume that $\varepsilon \equiv \Delta^H - \Delta$ is small. In this case, we can perform first-order approximations.

PROPOSITION 2. *If* $\varepsilon \equiv \Delta^H - \Delta$ *is small, we have*

$$\frac{\partial \pi^H}{\partial \varepsilon} > 0, \frac{\partial (r - \tau)}{\partial \varepsilon} < 0 \text{ and } \frac{\partial B}{\partial \varepsilon} > 0.$$

Proposition 2 summarizes the effect of an increase in households' estimation of risk-weighted capital $\Delta^H - \Delta$ (or a decrease in risk-weighted capital Δ keeping belief Δ^H constant). Households become more optimistic about the risk of the asset π^H. They hence charge a lower risk premium, which allows financial intermediaries to borrow more. This proposition illustrates how unexpected regulatory arbitrage might explain why, before the crisis, banks increased the risk they took without being sanctioned by higher risk premia.

PROPOSITION 3. *If* $\varepsilon \equiv \Delta^H - \Delta$ *is small and positive, we have*

$$\frac{\partial \pi^H}{\partial \tau} < 0, \frac{\partial (r - \tau)}{\partial \tau} > 0.$$

When households over-estimate risk-weighted capital, a decrease in the risk-free rate τ exacerbates their optimistic bias about the risky asset. The reason stems from equation (17). When the level of the risk-free rate τ decreases, the lending rate to financial intermediaries r also decreases under general conditions. The asset price P increases and households assign part of this increase to a decrease in the riskiness of the asset, leading to an increase in borrowing by financial intermediaries. The model can therefore explain one of the channels through which monetary policy might affect risk taking by financial intermediaries.[10]

The predictions of the model for risk perception are actually consistent with the empirical results produced by Altunbas, Gambacorta, and Marques-Ibanez (2010). They found that the expected

[10]Borio and Zhu (2009) coined the term "the risk taking channel of monetary policy" that they define as "the impact of changes in policy rates on either risk perceptions or risk-tolerance and hence on the degree of risk in the portfolios, on the pricing of assets, and on the price and non-price terms of the extension of funding."

default frequencies, and other market-based measures of bank's risks as perceived by financial market participants, react positively to changes in interest rates: a lower interest rate leads investors to perceive banks as less risky. Turning to banks' risk taking, which may be interpreted as banks exploiting their ability to borrow cheaply from financial markets, a number of recent studies—including Ciccarelli, Maddaloni, and Peydró (2010), Ioannidou, Ongena, and Peydró (2008), and Jimenez et al. (2007)—show that credit standards are correlated to the level of interest rates. Lower interest rates therefore imply lower credit standards, including for customers who are perceived as presenting a higher credit risk.

It is important to stress, however, that in our model the impact of the level of interest rates on risk perception and risk taking does not depend on the source of variation in interest rates. The interest rate in the model is real and can, therefore, be influenced by several factors. During the decade leading up to the crisis, several explanations were put forth in order to explain the low level of nominal and real interest rates. According to Taylor and Williams (2009), U.S. monetary policy was overly accommodative. Bernanke (2010), however, stressed instead that China's excess savings have played a major role in keeping the long end of the U.S. yield curve at comparatively low levels. Either of these factors may in turn have been amplified by the phenomenon of "search for yield," as emphasized by Rajan (2005). We do not take a position on these alternative possible drivers of the level of interest rates, and only stress that the endogenous mechanism described in our model would hold for either of them.

What the model highlights, however, is that the search for yield and risk taking can in part result from the wrong inference of risks from asset prices. This is because interest rates are central in the valuation of assets and the inference on risk incentives. It points to the interdependence of interest rates, asset prices, and capital-based prudential policies in a world where risk incentives and exposure cannot be assessed with certainty.

4.4 Under-Estimation of Δ

The previous section focused on the case in which households overestimate capital and wrongly infer the level of collapse risk. We show that this is consistent with the stylized facts on the pre-crisis period. The symmetric case is, however, also interesting.

The case of excess caution, where the risk-weighted capital of banks is believed to be too low, may help understand other periods of history. In a recent paper, Malmendier and Nagel (2011) show that households who experienced the Great Depression are less likely to invest in stock markets or participate in financial markets. Our model is able to rationalize this behavior by an under-estimation of the constraints imposed on banks after the Glass-Steagall Act. Anticipating that both the banking system and risky assets are more risky than they really are, households ask for higher returns or guarantees to compensate for the perceived risk. Empirical support for the view that investors' appetite for risk varies over time can also be found in Gilchrist and Zakrajsek (2012), who show large and persistent swings in the price of risk, defined as the part of bond risk premia that are not explained by "fundamentals" on the risk of default, where the latter is derived from the Merton's valuation of firms' stocks as an option to default. The price of credit risk was consistently negative from 2003 to 2007. It has also remained positive for several periods, such as around 2000, around 2008, and throughout the 1980s (see figure 1 in Gilchrist and Zakrajsek 2012). Again, the periods of over-estimation of risk may be due to an under-estimation of the capital constraints imposed on financial intermediaries.

5. Alternative Specifications of the Model

5.1 Elastic Supply of the Risky Asset

In the baseline model, we assume that the supply of the risky asset was fixed, i.e., $X_R = 1$. This section shows that the results are robust even if this supply may respond to prices, provided this response is not too large.

Let us now assume that instead of being sold in period 1 by initial sellers, the risky asset is produced by a new class of entrepreneurs. There is a unit mass of such entrepreneurs. They have access to a risky technology and consume in period 2. The risky technology yields $g(Y) = \lambda Y^{1-\theta}/(1-\theta)$, with a probability π in period 2, or fails to produce anything with probability $1-\pi$. We assume that risks are perfectly correlated among these entrepreneurs.

They sell the risky asset to financial intermediaries in period 1, at a unit price P. One unit of risky asset costs P in period 1 and pays off R with a probability π and 0 with a probability $1-\pi$ in period 2. Entrepreneurs choose how many units X_R of risky asset to sell. And they have no alternative to investing it. Their objective function is to maximize their period 2 consumption, denoted as c^R, with $c^R = \pi \left[g\left(PX_R\right) - RX_R \right]$. It yields

$$X_R = \left(\frac{\lambda}{R}\right)^{\theta} P^{\frac{1-\theta}{\theta}}.$$

Our baseline model is a special case where $\theta = 1$, and $\lambda = R$. It can be shown that our results on the effects of wrong beliefs (proposition 2) and the change in the risk-free rate τ (proposition 3) are still valid when θ is not too low. For low values of θ, P cannot deviate enough from its fundamental value for obvious reasons, so wrong beliefs concerning Δ have negligible effects.

This extension is important because financial innovation is likely to be stimulated when investors are optimists, or when either interest rates or risk premia are low. In the case of the U.S. sub-prime crisis, the supply of several forms of risky assets increased. More houses and condos were constructed, especially in states where housing prices increased the fastest (Florida, Nevada, etc.). The "originate-to-distribute" business model exposed mortgages to U.S. households that had decreasing creditworthiness. And more mortgage-backed securities and collateralized debt obligations that packaged these mortgages were sold to investors. The under-estimation of credit risk, as shown in figure 1, carried on for as long as the price of risky assets increased, i.e., the supply could not catch up with demand.

5.2 Risk about the Effectiveness of Capital Requirements

Our baseline model assumes that households may have wrong beliefs about the effectiveness of capital requirements, i.e., how much loss would be absorbed by holders of banks' stocks in the bad state of the world. Relaxing this assumption, while feasible, would greatly increase the complexity of the model. It is indeed possible to introduce an additional shock on the level of capital requirement. This shock is known by banks but unknown to the households. In this framework households form on average correct expectations about the average level of capital requirements. For an unexpected negative shock to this level, we would find the same results as in the current model where $\Delta^H > \Delta$. We chose a simpler structure to derive theoretical results in a transparent way.

5.3 Uncertain Return

In the baseline model, we assumed that the return in case of success R was known but that the probability distribution of risk π was unknown. An alternative modeling strategy would be to consider π as known but that the return $R = R(e)$ is uncertain and affected by the private actions of financial intermediaries. In this case, the main result would be preserved: if households know capital requirements Δ, they can infer from asset prices the real return $R(e)$ and thus the private action e. Changes in capital requirements may drive changes in private actions, but these changes would be anticipated and thus reflected in risk premia. Alternatively, when capital requirements are unknown, changes in capital requirements, and thus in equilibrium prices, will be partly understood to be a higher return and would thus bias the estimate of credit risk. Although it may be hard to distinguish between uncertain return and uncertain probabilities for specific assets, our modeling choice is motivated by the direct evidence of a sharp change in the expected probability of default during the crisis, as mentioned in section 2. Our model is thus designed to explain this bias in expectations of default.

5.4 Return and Risk

In our model, a change in the probability of default π affects both the mean and the variance of the return on the risky asset. As agents are risk neutral, the effect on the variance does not affect prices, but it would still be useful to express the model in order to analyze the effect of a change in the mean return keeping its variance constant. It is possible to do so by introducing an additional risk. Let us assume that the risky asset is equal to 0 with a probability $1 - \pi$ and equal to a stochastic variable $\tilde{R}$ with a probability π. $\tilde{R}$ has a mean R and is uniformly distributed in the support $[R - \delta; R + \delta]$. If the support is small enough, default will occur only when R is equal to 0. It is then possible to jointly choose π and δ to study the effect of a change in mean which keeps the variance constant.

6. Alternative Explanations of Low Risk Premia

6.1 Expectations of Bailouts

The model focuses on uncertainty of capital requirements to explain why risk premia were low before the crisis. An alternative explanation is that investors expected to be collectively bailed out by governments and central banks. Farhi and Tirole (2012) propose a model in which financial institutions coordinate their exposure to risks in order to increase systemic risk, and therefore the likelihood that public authorities will bail them out. Their model explores the issue of risk shifting from investors to taxpayers. We focus instead on the shifting of risk from banks to bondholders.

It is clear that the concepts of "too big to fail" or "too interconnected to fail" are likely to have influenced investors. The expectations of bailouts may have played a role in the evolution of risk premia. In particular, the big change in expected default frequency after the collapse of Lehman in September 2008 may have been due partly to a revision in the perceived probability of a public bailout and partly to a reassessment of the underlying risk. To the best of our knowledge, there is no evidence that allows us to rule out that these phenomena may have influenced risk premia.

It should also be stressed, however, that the political process that leads to bailing out financial institutions is uncertain. Financial institutions usually pay a credit risk premia with respect to the Treasury. These premia vary over time for a number of reasons, including the collective moral hazard hypothesis of Farhi and Tirole (2012) and the one we propose in this paper.

6.2 General Under-Estimation of Risk

Another explanation of low risk premia would be that all agents, including financial intermediaries, under-estimated the risk of default on housing assets. This is, for instance, the view held by Schleifer (2011). He stresses that the ex post losses of the financial industry were so large as to discard the "expectation of bailouts" hypothesis as well as hypotheses, like the one we present in this paper, which highlight asymmetry of information between the financial industry and non-financial agents. Again, a global underestimation of risk cannot be rejected, and it may be hard to claim that banks correctly anticipated the real risk of all assets. However, some evidence, such as legal actions against financial intermediaries, suggests that financial intermediaries and households did not all have the same information. In 2007 Chuck Prince, then chairman of Citigroup, declared: "When the music stops, in terms of liquidity, things will be complicated. But, as long as the music is playing, you've got to get up and dance. We are still dancing." (*Financial Times* 2007)

This is highly suggestive that banks knew they were taking risks.[11]

7. Concluding Remarks: Can the Model Explain the Buildup of Financial Fragility?

In this paper we show that the combination of risk shifting and fuzzy capital requirements may explain one of the sub-prime crisis puzzles, i.e., that financial intermediaries were able to increase their exposure to risk without having to pay higher risk premia on their debt.

In an opaque banking system where regulatory constraints are difficult to observe, an increase in asset prices can be interpreted as a lower aggregate risk in the economy while, in fact, asset prices are driven by greater risk taking on the part of financial intermediaries. We also showed that this model gives rise to a risk-taking channel of low interest rates, which reduces the perceived risk of some agents and increase the exposure to risk of others.

Our result resonates with the popular notion that financial markets participants can draw incorrect inferences about risks. In particular, when the effectiveness of capital requirements is not observable by agents, the signal extracted from market prices is contaminated by noise coming from excessive risk-taking behavior. In our model,

[11] Another paper elaborating on the information asymmetry between financial intermediaries and households is Shleifer and Vishny (2010). In this paper, the expectations of households are taken as given. Instead, we endogenize them through an inflation-extraction problem.

market forces, by themselves, do not result in an optimal allocation of capital, because risk incentives are not correctly understood.

We see two obvious extensions to our model. First, it is possible to endogenize the expectations of households within a dynamic setting in which households learn about the relevant parameters. The results of our model would still hold if the priors of the households were far enough from the true parameters. The resulting dynamics of their learning process should generate useful patterns. Second, it would be interesting to study the political economy aspects associated with the assessment of risk within such an economy. Sellers of the assets have an incentive to under-estimate the degree of risk, or to generate complexity in order to increase the cost of signal extraction. This should be anticipated by households, who would then look for other sources of information. It is in this context, for example, that we interpret the current discussion about rating agencies to be part of the debate on the management of risk expectations in economies where intermediaries play an important role.

Appendix

Proof of Proposition 1

Since $\pi \in]0,1[$ and $X_S < B$, the no-arbitrage condition (19) implies $r > \tau$. Equality (19) can be written as

$$B(r) = \frac{(1-\pi)\, r X_S}{\tau - \pi r}. \tag{22}$$

We can substitute K, X_S, and P with their equilibrium values given by equations (15) and (16) to obtain an expression $B\,(r)$:

$$B\,(r) = \frac{(1-\Delta)\, R}{\frac{\Delta}{\beta\pi} + r\,(1-\Delta)} X_R + f'^{-1}\,(r)\,. \tag{23}$$

Let us define: $\Theta \equiv \frac{\Delta}{1-\Delta}$. Then from (16), (23), and (19), we find that the real interest rate r satisfies the equality

$$\tau = \pi r + (1-\pi)\, \frac{(\Theta + \beta\pi r)\, r^{1-1/\eta}}{\beta\pi R X_R + \Theta r^{-1/\eta} + \beta\pi r^{1-1/\eta}}.$$

This last equality implicitly defines the interest rate by equality $M(r) = \Theta$, where

$$M(r) \equiv \beta\pi \left(\frac{(\tau - \pi r) r^{1/\eta} \mu X_R}{r - \tau} - r \right).$$

In the equilibrium under consideration, $\pi r < \tau < 1$ and $r > \tau$. As a consequence, we can check that a sufficient condition for $M'(r) < 0$ is $\eta > \frac{1-\pi}{\pi}$, which is (1). In this case, the equality $M(r) = \Theta$ implies that r is decreasing with Δ.

From equality (19), one finds $B(r) = \frac{1-\pi}{\tau - \pi r} r^{1-\frac{1}{\eta}}$. After some algebra, we find that $B(r)$ increases with r when (1) is fulfilled. As a consequence, $\frac{\partial B}{\partial r} > 0$, $\frac{\partial (r-\tau)}{\partial \Delta} < 0$, and $\frac{\partial B}{\partial \Delta} < 0$.

Proof of Proposition 2

Denote $\varepsilon = \Delta^H - \Delta$. From equations (16), (21), (17), (18), and (20) we find that the real interest rate satisfies $G(r, \tau, \Delta, \varepsilon) = 0$, where

$$G(r, \tau, \Delta, \varepsilon) \equiv \beta\pi \left[\frac{\left(\tau - \pi r \frac{\Delta + \varepsilon}{\Delta + r\beta\pi\varepsilon}\right) R r^{\frac{1}{\eta}}}{r - \tau} - r \left(1 - \frac{\varepsilon}{1 - \Delta}\right)^{-1} \right] - \frac{\Delta}{1 - \Delta - \varepsilon}. \tag{24}$$

As $\tau, \Delta, \varepsilon$ are given parameters, the equality $G(r, \tau, \Delta, \varepsilon) = 0$ defines the equilibrium interest rate as a function of the parameters of the model. Studying the derivative of the function G, we find the following signs for the derivatives (with obvious notations):

$$G\left(\underset{-}{r}, \underset{+}{\tau}, \underset{-}{\Delta}, \underset{-}{\varepsilon}\right) = 0.$$

As a consequence and by the implicit function theorem, we find that r has the following variations: $r = r\left(\underset{+}{\tau}, \underset{-}{\Delta}, \underset{-}{\varepsilon}\right)$. This proves $\frac{\partial (r-\tau)}{\partial \varepsilon} < 0$.

Next, the anticipated probability can be written as $\pi^H = \pi(\varepsilon + \Delta) / (\Delta + r\beta\pi\varepsilon)$ from (18). As a consequence, one finds the following variations: $\pi^H = \pi^H \left(\underset{-}{r}, \underset{+}{\varepsilon} \right)$. This proves $\frac{\partial \pi^H}{\partial \varepsilon} > 0$ (as r decreases when ε increases).

Finally, the budget constraint of financial intermediaries, together with the price (17), gives the debt level B:

$$B = r^{-\frac{1}{\eta}} + \frac{\beta \pi R}{\frac{\Delta}{1-\Delta} + r\beta\pi}.$$

One can easily deduce the variation $B = B\left(\underset{-}{r}\right)$ and $r = r\left(\underset{+}{\tau}, \underset{-}{\Delta}, \underset{-}{\varepsilon}\right)$. We have thus the variations $B = B\left(\underset{-}{\tau}, \underset{+}{\varepsilon}\right)$.

Proof of Proposition 3

From the proof of proposition 2, we obtain $r = r\left(\underset{+}{\tau}, \underset{-}{\Delta}, \underset{-}{\varepsilon}\right)$ and $\pi^H = \pi^H \left(\underset{-}{r}, \underset{+}{\varepsilon} \right)$. As a consequence, we find $\frac{\partial \pi^H}{\partial \tau} < 0$. The proof of the inequality $\frac{\partial (r-\tau)}{\partial \tau} > 0$ requires more algebra. This inequality is first proven for $\varepsilon = 0$. Then, a continuity argument is invoked. From the definition (24) and the equality $G(r, \tau, \Delta, \varepsilon) = 0$, we find

$$r - \tau = r \frac{1 - \pi \frac{\Delta+\varepsilon}{\Delta + r\beta\pi\varepsilon}}{\frac{1}{\mu_1 r^{\frac{1}{\eta}}} \left[\frac{\Delta}{1-\Delta} \frac{1}{\beta\pi} + r \right] \frac{1}{1-\frac{\varepsilon}{1-\Delta}} + 1}.$$

Taking the derivative with respect to τ and setting $\varepsilon = 0$, we easily find $\frac{\partial (r-\tau)}{\partial \tau} > 0$. By continuity, this inequality is fulfilled when ε is small.

References

Acharya, V., and P. Schnabl. 2009. "How Banks Played the Leverage 'Game'." NYU Stern White Paper. Executive summary of *Restoring Financial Stability: How to Repair a Failed System*, ed. V. V. Acharya and M. Richardson, chapter 2. John Wiley and Sons, Inc.

Adrian, T., and H. S. Shin. 2010. "Liquidity and Leverage." *Journal of Financial Intermediation* 19 (3): 418–37.

Allen, F., and D. Gale. 2000. "Bubbles and Crises." *Economic Journal* 110 (460): 236–55.

Altunbas, Y., L. Gambacorta, and D. Marques-Ibanez. 2010. "Bank Risk and Monetary Policy." *Journal of Financial Stability* 6 (3): 121–29.

Barlevy, G. 2008. "A Leverage Based Model of Speculative Bubbles." Working Paper No. 2008-01, Federal Reserve Bank of Chicago.

Bernanke, B. 2010. "Monetary Policy and the Housing Bubble." Speech given at the Annual Meeting of the American Economic Association, Atlanta, Georgia, January 3.

Blanchard, O. 2008. "The Crisis: Basic Mechanisms, and Appropriate Policies." MIT Working Paper No. 09-01.

Blundell-Wignall, A., and P. Atkinson. 2008. "The Sub-prime Crisis: Causal Distortions and Regulatory Reform." Paper presented at the Reserve Bank of Australia conference on "Lessons from the Financial Turmoil of 2007," Sydney, Australia, July 14–15.

Borio, C., and H. Zhu. 2009. "Capital Regulation, Risk-Taking and Monetary Policy: A Missing Link in the Transmission Mechanism?" BIS Working Paper No. 26.

Brender, A., and F. Pisani. 2010. "Global Imbalances and the Collapse of Globalized Finance." CEPS Monography. Available at http://www.ceps.eu.

Brunnermeier, M. 2009. "Deciphering the Liquidity and Credit Crunch 2007-08." *Journal of Economic Perspectives* 23 (1): 77–100.

Challe, E., B. Mojon, and X. Ragot. 2012. "Equilibrium Risk Shifting and Interest Rate in an Opaque Financial System." Working Paper No. 391, Banque de France.

Challe, E., and X. Ragot. 2011. "Bubbles and Self-fulfilling Crisis." *B.E. Journal of Macroeconomics* 11 (1): ISSN (Online) 1935-1690.

Ciccarelli, M., A. Maddaloni, and J.-L. Peydró. 2010. "Trusting the Bankers: A New Look at the Credit Channel of Monetary Policy." ECB Working Paper No. 1228.

Dell'Ariccia, G., L. Laeven, and R. Marquez. 2010. "Monetary Policy, Leverage and Bank Risk-Taking." IMF Working Paper No. 10/276.

Farhi, E., and J. Tirole. 2012. "Collective Moral Hazard, Maturity Mismatch and Systemic Bailouts." *American Economic Review* 102 (1): 60–93.

Financial Stability Forum. 2008. "Report of the Financial Stability Forum on Enhancing Market and Institutional Resilience." (April).

Financial Times. 2007. "Citigroup Chief Stays Bullish on Buy-Outs." July 7.

Geanakoplos, J. 2009. "The Leverage Cycle." Discussion Paper No. 1715R, Cowles Foundation.

Gilchrist, S., and B. Mojon. 2014. "Credit Risk in the Euro Area." Banque de France Working Paper No. 482 and NBER Working Paper No. 20041.

Gilchrist, S., and E. Zakrajsek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–1720.

Greenlaw, D., J. Hatzius, A. K. Kashyap, and H. S. Shin. 2008. "Leveraged Losses: Lessons from the Mortgage Market Meltdown." U.S. Monetary Policy Forum Report No. 2.

Holmstrom, B., and J. Tirole. 1997. "Financial Intermediation, Loanable Funds and the Real Sector." *Quarterly Journal of Economics* 112 (3): 663–91.

Ioannidou, V. P., S. Ongena, and J. L. Peydró. 2008. "The Impact of Short-Term Interest Rates on Risk-Taking: Hard Evidence." VoxEU CEPR Publication.

Jiménez, G., S. Ongena, J.-L. Peydró-Alcalde, and J. Saurina. 2007. "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk?" ECB Working Paper No. 835.

Kiyotaki, N., and J. Moore. 1997. "Credit Cycles." *Journal of Political Economy* 105 (2): 211–48.

Maddaloni, A., J.-L. Peydró, and S. Scopel. 2008. "Does Monetary Policy Affect Bank Credit Standards?" Mimeo, European Central Bank (August).

Malmendier, U., and S. Nagel. 2011. "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *Quarterly Journal of Economics* 12 (1): 373–416.

Rajan, R. 2005. "Has Financial Development Made the World Riskier?" In *The Greenspan Era: Lessons for the Future*, 314–69.

Proceedings of the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming, August 25–27.
Rochet, J.-C. 2008. "The Future of Banking Regulation." Toulouse School of Economics Notes No. 2.
Schleifer, A. 2011. "The Transformation of Finance." Presentation at the American Finance Association Congress, Denver, Colorado, January 7–9.
Schleifer, A., and R. Vishny. 2010. "Unstable Banking." *Journal of Financial Economics* 97 (3): 306–18.
Shin, H. S. 2010. *Risk and Liquidity.* Clarendon Lectures in Finance. Oxford University Press.
Taylor, J., and J. Williams. 2009. "A Black Swan in the Money Market." *American Economic Journal: Macroeconomics* 1 (1): 58–83.

Ordering Policy Rules with an Unconditional Welfare Measure*

Tatiana Damjanovic,[a] Vladislav Damjanovic,[a] and Charles Nolan[b]
[a]University of Exeter
[b]University of Glasgow

The unconditional expectation of social welfare is often used to assess alternative macroeconomic policy rules in applied quantitative research. This paper provides a detailed analysis of such policies. It sets out the unconditionally optimal (UO) policy problem and derives a linear-quadratic (LQ) version of that problem that approximates the exact non-linear problem. The properties of UO policies are analyzed through a series of examples and contrasted with the timeless perspective (TP), exposited in Benigno and Woodford (2012). Some substantive implications for optimal monetary policy are explored.

JEL Codes: E20, E32, F32, F41.

1. Introduction

DSGE models can be difficult to analyze in their non-linear form, so a linear-quadratic (LQ) approximation is often adopted. That approach has been extended by Sutherland (2002), Benigno and Woodford (2004, 2005, 2006b), and many others in the context of specific models. More recently, Benigno and Woodford (2012)

*This paper was substantially improved following comments by the editor, Pierpaolo Benigno. We should also like to thank Jinill Kim and two anonymous referees for helpful comments. We also thank the participants at the 2008 North American meeting of the Econometric Society at Carnegie Mellon University, the 2008 Norges Bank Workshop on Optimal Monetary Policy in Oslo, and the 2007 Royal Economic Society Conference and research seminars at the University of Exeter, the Bank of England, and the Dutch Central Bank for helpful comments and discussions. The usual disclaimer, of course, applies. Author e-mails: t.damjanovic@exeter.ac.uk; v.damjanovic@exeter.ac.uk; Charles.Nolan@glasgow.ac.uk.

demonstrate that the LQ approach can be applied quite generally to optimal policy problems. They show that if that approximation is performed around the optimal steady state, the solution to the LQ problem represents a first-order approximation to the (non-linear) optimal policy. The steady state is optimal in the sense that it maximizes the original non-linear objective function with respect to the original non-linear structural constraints of the model. In particular, Benigno and Woodford (2012) show that it is possible to apply the LQ method to a model where the government searches for an optimal policy from a timeless perspective (TP). Moreover, their method delivers a pure second-order approximation to TP social welfare which permits one to rank alternative policies from a timeless perspective.[1]

One objective of this paper is to show that it is also possible to apply an LQ approach to the case where a policymaker maximizes the unconditional expectation of social welfare. Precisely, we will show that an LQ approximation to the unconditional expectation of social welfare around the unconditionally optimal (UO) steady state will also generate a welfare function with purely second-order terms. The solution of that LQ-UO problem yields a first-order approximation to the policy which maximizes the unconditional expectation of social welfare. Moreover, a pure second-order welfare function can be obtained and used to measure the performance of simple policy rules from an UO perspective.

The more important objective of the paper is to offer a detailed analysis of UO policies since that is lacking in the literature, despite the prevalence of the UO perspective in quantitative macroeconomic research. For example, Taylor (1979) suggested that, in quantitative theoretical investigations under rational expectations, macroeconomic stabilization policies ought to optimize the unconditional expectation of the policymaker's objective function. That perspective on policy assessment has been popular; some prominent examples include Whiteman (1986), Rotemberg and Woodford (1998), Clarida, Gali, and Gertler (1999), Woodford (1999b), Erceg, Henderson, and Levin (2000), Kollman (2002), Kim and Henderson (2005), and Schmitt-Grohe and Uribe (2007). More recently, many

[1]See also Kim and Kim (2007).

researchers have adopted the TP in quantitative and theoretical research. Consequently, it will be useful to compare UO and TP approaches in what follows in order to highlight the key features of these differing perspectives on optimal policy.

The potential attractions of unconditionally optimal policy, as opposed to the timelessly optimal policy, are pursued in Jensen and McCallum (2010). The timeless perspective minimizes the variance but does not account for the initial conditions (see Woodford 2002, p. 508). However, the distribution of initial conditions depends on the policy adopted in preceding periods. That relation is explicitly taken into account by the unconditionally optimal approach. We demonstrate through a number of examples that adopting the unconditionally optimal methodology internalizes the initial distribution and delivers a policy which is best on average. That is, UO policy performs better on average precisely because it takes into consideration the initial conditions. UO policy is optimal given that a similar policy choice was made by past policymakers. Hence, the principal difference between TP policy and UO policy is that the former takes the distribution of initial conditions as given and ignores impact of policy on them in its design. In section 3 we explore in depth the fundamental reasons why UO and TP policies differ. We also extend Jensen and McCallum (2010) by applying the UO methodology to a DSGE model with no forward-looking constraints on the policy problem. That provides a very clear example of what internalizing initial conditions actually means.

This paper also highlights important differences between UO and TP policies when the steady state is distorted. Jensen and McCallum (2010) compare UO (what they call "optimal continuation" policies) with TP policies, when the steady state is efficient. In that case, it is shown that the form of the welfare function to be optimized is the same across programs. Here we show that the corresponding LQ problems can be significantly different when the steady state is distorted. We find that UO and TP approaches imply different steady states, different arguments in the quadratic social welfare function, and different linear dynamic constraints. Even the number of nondegenerate dynamic constraints may differ across the TP and UO policy problems.

To design our algorithm, we extend the methodology of Damjanovic, Damjanovic, and Nolan (henceforth DDN) (2008), which derives the first-order necessary conditions for the policy optimizing the unconditional expectation of welfare.[2] Then, similar to Judd (1999) and Benigno and Woodford (2012), the linear-quadratic approximation is done around the optimal deterministic steady state; in our case, around the unconditionally optimal steady-state, and in Benigno and Woodford's case, the timeless-perspective (TP) steady state.[3]

As a result, we develop a useful approach for constructing the pure second-order approximation to unconditional welfare. Since this measure can be presented in the form of a linear combination of the second moments, one can apply the Anderson et al. (1996) algorithm, which has good convergence properties. Consequently, it is also straightforward numerically to analyze policies from a UO perspective.

A specific application of the approach is provided employing the canonical New Keynesian model. A number of insights emerge. First, unconditionally optimal monetary policy is characterized by trend inflation. That trend in inflation complicates the linear quadratification.[4] That explains a second insight: The second-order approximate loss function is no longer defined solely over terms in output and inflation as found in DDN (2008) for the non-distorted steady-state case. However, the loss function that one obtains is easily interpreted in light of the underlying distortions in the economy. The approximate loss function is used to evaluate and rank different simple rules for monetary policy (i.e., the nominal interest rate). The welfare implications of nominal income targeting versus inflation targeting are explored, and our results are contrasted with some of those of Kim and Henderson (2005).

The rest of the paper is organized as follows. In section 2 we recall the basic setup of the problem and the necessary first-order conditions for the optimal steady state.[5] In section 3 we discuss further and contrast the attractions of UO and TP policies. It is

---

[2]See also Whiteman (1986), Blake (2001), and Jensen and McCallum (2002), (2010).

[3]See also Sutherland (2002), Debortoli and Nunes (2006), and Levine, Pearlman, and Pierse (2008).

[4]As shown in Ascari and Ropele (2007) and Damjanovic and Nolan (2010).

[5]That is the only section which overlaps with DDN (2008). Moreover, here we provide a new derivation of the UO policy problem.

shown in section 4 that one can derive a purely quadratic approximation to the unconditional expectation of the objective function. In section 5 a canonical New Keynesian, Calvo price-setting model is set up. Section 6 formalizes the policy problem and demonstrates the application of the various steps in the approach of section 4. There is then a brief discussion of the implications for optimal monetary policy when the steady state is distorted and the authorities are optimizing over the unconditional loss function. We also contrast the linear-quadratic form of UO and TP approaches, including the difference in the point of approximation. In section 7 we use the unconditional welfare criterion to explore briefly the impact of different simple rules for monetary policy. There is also a short discussion of optimal monetary policy following productivity and markup shocks under TP and UO policies. Section 8 offers some conclusions. Appendices contain proofs and details of key derivations.

2. The General UO Problem

Consider a discounted loss function of the form

$$L_t = (1-\beta)\, E_t \sum_{j=0}^{\infty} \beta^j l(x_{t+j}), \tag{1}$$

where E_t is the expectations operator *conditional on information up through date* t, β is the time discount factor, $l(x_{t+j})$ is the period loss function, and x_t is a vector of target variables. Specifically, $x_t = [Z_t, z_t, i_t]$, where Z_t is a vector of predetermined endogenous variables (lags of variables that are included in z_t and i_t); z_t is a vector of non-predetermined endogenous variables (including "jump" variables), the value of which will generally depend upon both policy actions and exogenous disturbances at date t; and i_t is a vector of policy instruments, the value of which is chosen in period t. Let μ_t denote a vector of exogenous disturbances. For simplicity, assume that μ_t is a function of primary i.i.d. shocks, $(e_i)_{-\infty}^{t}$.

Further, let the evolution of the endogenous variables z_t and Z_t be determined by a system of simultaneous equations,

$$E_t F\left(x_{t+1}, x_t, \mu_t\right) = 0. \tag{2}$$

Let us further assume, following Taylor (1979), that the policymaker seeks to minimize the unconditional expectation of the loss function (1), subject to constraints, (2).[6] That is, he or she searches for a policy rule

$$\varphi\left(E_t x_{t+1,} x_t, \mu_t\right) = 0, \tag{3}$$

such that

$$\varphi^* = \arg\min EL_t(\varphi), \tag{4}$$

where E is the unconditional expectations operator. We call such a policy "unconditionally optimal" and denote it "UO policy."

Formally, the unconditional expectation of any function $l(x)$ can be represented in Lebesgue integral form as

$$El_t(x_t(\varphi)) = \int l_t(x_t(\varphi, e))de,$$

where de is the Cartesian product probability measure of i.i.d. primary shocks with history, $(de_{t-k})_{k=0}^{\infty}$. We emphasize that de is given exogenously and does not change with policy.

2.1 Solution

The first step is to formulate the non-linear policy problem and identify the non-stochastic steady state around which approximation needs to take place. For this purpose we will use constraints (2) and necessary first-order conditions.

2.1.1 Necessary Conditions for an Optimum

Consider the following Lagrangian function which is implied by the above optimal policy problem:

$$\mathcal{L}^{UO}\left(\{y_{t,} x_t, \mu_t\}\right) = E\left(l\left(x_t\right) + \xi_t F\left(y_{t,} x_t, \mu_t\right) + \rho_t\left(x_{t+1} - y_t\right)\right), \tag{5}$$

[6] Taylor's approach may be interpreted as a recommendation: Policymakers *ought* to seek to minimize the unconditional value of the loss function. This appears partly, perhaps largely, in response to the issue of time inconsistency. See Taylor (1979) for further discussion. McCallum (2005) is an interesting discussion of these, and related, issues.

where we define

$$y_t = x_{t+1}. \quad (6)$$

We introduce definition (6) solely for presentational purposes. The formulation of (5) is justified as follows. The Lagrangian is defined as a function of all variables in all periods of time. Constraints (2) and (6) should be satisfied for all time periods and for any realization of the history of the shocks, $(e_i)_{-\infty}^{t+1}$. In each period t for each realization of the history e, we will have a pair of constraints corresponding to (2) and (6) and a corresponding pair of Lagrange multipliers $(\xi_t(e); \rho_t(e))$. We need to sum up all these constraints across all possible histories of the shocks' realizations and across time. To that end, first define $\mathcal{L}_t$, which is the sum of all constraints at time t, multiplied by their Lagrange multipliers:

$$\mathcal{L}_t := \int \xi_t(e) F(y_t, x_t, \mu_t, e) + \rho_t(e)(x_{t+1}(e) - y_t(e))\, de. \quad (7)$$

By definition, $\mathcal{L}_t$ is the unconditional expectation, and its value does not depend on time. Thus, one may write

$$\mathcal{L}_t = E(\xi_t F(y_t, x_t, \mu_t) + \rho_t(x_{t+1} - y_t)). \quad (8)$$

Now formally one can sum over time periods for any discount rate; we use β for consistency. Thus, $(1-\beta) E_t \sum_{j=0}^{\infty} \beta^j \mathcal{L}_t = \mathcal{L}_t$, as $\mathcal{L}_t$ represents the unconditional expectation which is independent of time. Therefore expression (5) is the sum of the objective and all the constraints multiplied by corresponding Lagrange multipliers.

DDN (2008) show that the necessary conditions for the optimality of policy, φ, is that it implies a path for the endogenous variables, x_t and y_t, and that there exist Lagrange multipliers, (ξ_t, ρ_t), that together satisfy the first-order conditions (9), (10) and constraints (2),[7]

$$\frac{\partial H}{\partial x_t} = \frac{\partial l(x_t)}{\partial x} + \xi_t \frac{\partial F(y_t, x_t, \mu_t)}{\partial x} + \rho_{t-1} = 0; \quad (9)$$

$$\frac{\partial H}{\partial y_t} = \xi_t \frac{\partial F(y_t, x_t, \mu_t)}{\partial y_t} - \rho_t = 0, \quad (10)$$

[7] The notation ξF is a shorthand for the tensor product, $\sum_{i=1}^{n} \xi_i F_i$.

where $H(y_{t,}x_t, \mu_t)$ is the Hamiltonian for (5), such that $\mathcal{L}(y_{t,}x_t, \mu_t) = E(H(y_{t,}x_t, \mu_t))$.

Judd (1999), Woodford (2002), and Benigno and Woodford (2005) demonstrate very clearly that the choice of the steady state is crucial (along with the solution concept for forward-looking policy problems) in being able to obtain LQ approximations to general nonlinear, forward-looking policy problems. To choose the deterministic steady state around which log-linearization takes place, one needs to solve the system of first-order conditions (9), (10) and constraints (2). This leads to the following proposition.

PROPOSITION 1. *The steady state* (X, ξ) *is defined by the system (11)–(12):*

$$F(X, X, \mu) = 0; \tag{11}$$

$$\frac{\partial l(X)}{\partial x_t} + \xi \frac{\partial F(X, X, \mu)}{\partial x} + \xi \frac{\partial F(X, X, \mu)}{\partial y} = 0, \tag{12}$$

where X, ξ, *and* μ *indicate the vectors of steady-state values of endogenous variables, Lagrange multipliers, and the average value of shocks, respectively.*

We refer to (X, ξ) as the "unconditionally optimal steady state."[8] In the absence of shocks, solution (11) shows that unconditionally optimal policy delivers the steady state with the highest level of steady-state welfare "on average," where the averaging is with respect to the unconditional measure. It is worth emphasizing that the TP approach discussed in, e.g., Woodford (2002) implies different first-order conditions and therefore a different center of approximation. That difference will be shown to lead to a different optimal monetary policy.

3. Comparing UO and TP

In this section we begin our analysis of UO and TP policies. Before proceeding, it is noted that comparing TP and UO policies should be done with care[9]: TP policies are optimal (and time consistent)

[8]It is assumed throughout that system (11) has a unique solution.
[9]The editor has emphasized this point to us.

when one employs a TP welfare measure, and the same is true vis-à-vis UO policy. However, we will on occasion ask whether agents would prefer to be situated in economies with TP or UO policies.[10]

3.1 TP Program and Social Time Discounting

The TP approach was defined first in Woodford (1999b), where a mathematical description of the policy was provided via the first-order conditions for TP policy. One first forms an appropriate Lagrangian for the TP program:

$$\mathcal{L}^{TP}\left(y_t, x_t, \mu_t\right) = E_t \sum_{j=0}^{\infty} \beta^j \big(l\left(x_{t+j}\right) + \xi_t F\left(y_{t+j}, x_{t+j}, \mu_{t+j}\right) + \rho_t \left(x_{t+1+j} - y_{t+j}\right) \big). \tag{13}$$

The necessary first-order conditions with respect to future variables will be

$$\frac{\partial \mathcal{L}^{TP}}{\partial x_t} = \frac{\partial l\left(x_t\right)}{\partial x} + \xi_t \frac{\partial F\left(y_t, x_t, \mu_t\right)}{\partial x} + \frac{1}{\beta}\rho_{t-1} = 0; \tag{14}$$

$$\frac{\partial \mathcal{L}^{TP}}{\partial y_t} = \xi_t \frac{\partial F\left(y_t, x_t, \mu_t\right)}{\partial y_t} - \rho_t = 0. \tag{15}$$

Based on these relations, one concludes that TP policy coincides with fully optimal Ramsey policy conditional on the economy starting from the TP-optimal steady state. Thus, (14), (15), and (2) determine the TP optimal steady state. It is also the convergence state of the fully optimal Ramsey policy, conditional on the government possessing appropriate commitment technology.

Comparing the first-order conditions for the UO program (9) and (10) with those of the TP (14) and (15), one concludes that TP policy coincides with UO policy if the policymaker's discount factor is equal to unity, $\beta = 1$. Hence, the policymaker's time discounting is central to the differing perspectives of UO and TP policies. We explore these differences in a little more detail now.

[10]In much the same way as public finance theorists might inquire whether agents would prefer to live in a world with Benthamite or Rawlsian tax and redistribution policies.

3.2 Policies Compared

One may compare TP and UO policies along two related dimensions. First, which concept of optimality is more appropriate in normative study of policy problems? Second, if actual policymakers are able to commit, which of the two policies are preferable? Jensen and McCallum (2010) pursue the first of these, concluding that UO policies are more appropriate in many instances. We refer the reader to their insightful discussion. Assuming both policies are achievable, which policy ought a policymaker adopt? That is perhaps something of a philosophical question.

Unconditionally optimal policy maximizes average utility, given that all generations follow one rule. To design that rule, one takes into account not only current actions but also the actions one would have wished our predecessor policymakers to have followed; such a rule would bequeath preferable initial conditions to the current policymakers. By the same token, whatever current policymakers would have asked of our predecessors is what current policymakers undertake to provide to future policymakers. Therefore the benefits to the current generation of the rule, due to past commitments, are traded against costs associated in passing benefits to future generations. In other words, the UO methodology entails treating future generations the same as the current generation, or without any discounting ($\beta = 1$). Of course, that idea is not new in economic philosophy.

According to Ramsey (1928), discounting future generations' welfare is unethical. Harrod (1948) recommends that government "correct" individuals' savings decisions because they reflect positive time discounting and a resulting "palpable improvidence." In this he agrees with Pigou (1932), who comments on agents' "defective telescopic faculty" as a reason why private discount rates are excessive. In particular, Harrod argues that an individual will ex post be grateful to a government which induces him or her to invest the amount corresponding to a decision predicated on a zero discount rate. Solow (1974) also argues that the social discount factor perhaps should be higher than the private one.

More concretely, a recent example of the importance of time discounting is in the area of environmental policy, where the current generation is expected to invest in environmental protection for the benefit of future generations. The size of that investment crucially depends on the discount factor attached to the welfare of future generations. Stern (2007, p. 31) proposed to "treat the welfare of future generations on a par with our own," which on a strict interpretation is to set the discount factor equal to 1. The same view has been expressed in Anand and Sen (2000) and many others.

Another reason for a "low" social discount factor follows from the assumption that utility depends on past as well as present and future variables. For example, good memories may make one happier. Strotz (1956) introduced that idea and Caplin and Leahy (2004) show that, in such a setting, policymakers should be more patient than private individuals.

In the next section a simple example is provided that highlights, we think, the attractions, and difficulties, in implementing UO policy relative to TP policy. We show that it may be not altruism towards future generations but rather considerate behavior of the predecessors that makes UO policy more attractive. In particular, in the following example, it is preferable for agents to live in a UO world given that the previous generations acted under UO policy, rather than living in a TP world inherited from TP predecessors. These benefits notwithstanding, there are costs in a transition to UO policy, and some generation will have to bear them.

3.3 Treatment of the Initial Conditions

The central difference between UO and TP policies is in the treatment of initial conditions. TP policy takes the initial conditions as given and optimizes future losses accordingly. UO policy acknowledges the fact that policy affects the distribution of initial conditions as well. To show the difference, we apply these differing policies to a simple model with only backward-looking constraints on policy—in that case, the value of the state variable is the relevant "initial condition." In this "backward-looking" model, the TP policy will coincide with optimal Ramsey policy (i.e., one with discounting). UO policy will give a different outturn, which we describe below. Consider the following example.

Example 1. *The technological process is the following. Generation t makes a costly investment N_t (like planting edible seeds). Generation $t+1$, consumes the fruit from the crop*

$$C_{t+1} = A_{t+1} N_t, \tag{16}$$

and makes an investment for the next generation. Here A_{t+1} *is an index of productivity. Society is altruistic and cares about succeeding generations applying a discount rate* β *to their welfare. So utility of the current generation* t *is*

$$U_t = \log(C_t) - kN_t + \beta U_{t+1}, \quad k > 0.$$

Upon integrating forward, social welfare is found to be

$$U_t = \sum \beta^{t+s} \left(\log(C_{t+s}) - N_{t+s}\right).$$

Then, on substituting in the production technology (16), the policy maximand is

$$U_t = \sum \beta^{t+s} \left(\log(A_{t+s}) + \log(N_{t+s-1}) - N_{t+s}\right). \tag{17}$$

The first-order condition with respect to N_t *is*

$$\frac{\partial U_t}{\partial N_t} = \beta \frac{1}{N_t} - 1 = 0.$$

Clearly, therefore, the Ramsey solution and the TP solution coincide and $N_t = \beta$*.*

Now consider a rule for N *that has applied for all time; it has been followed by past generations and it will be followed by the current and future generations. In that case, one seeks an optimal choice,* $N_t = N$ *for all* t*. Thus, utility becomes*

$$U = \sum \beta^{t+s} \left(\log(A_{t+1}) + \log(N) - N\right) \tag{18}$$

and the first-order condition gives $N = 1$*. One might inquire whether or not one would wish to live in an economy with a TP policy or a UO policy. In fact, the UO policy generates much larger utilities for all generations. In this particular case, the gain in consumption equivalent is* $\exp(-\log(\beta) - (1 - \beta))$*, which can vary from 0.5 percent if* β *is relatively large, say* $\beta = 0.9$*, to 6 percent for* $\beta = 0.7$*.*

So one may assume that, given the choice, an individual would prefer to live in an economy where the government runs a policy corresponding to a higher social discount factor.

The difference in TP and UO policies can be explained by the treatment of initial conditions. TP methodology considers N_{t-1} as given and treats it as a "term independent of policy." UO methodology is designed as the best policy, conditional that it is accepted by all generations, including generation $t-1$. That is why N_{t-1} is taken into account.

However, in the present example, there are two major drawbacks of the UO approach relating to transition and time consistency. If a country decides to switch from TP to UO policy, the transition generation will be worse off; they will inherit a lower investment left by the previous generation but will be asked to invest more for the sake of the next generation. Moreover, every generation *irrespective of their inheritance* will have an incentive to deviate from UO policy, switching to Ramsey policy (which coincides in this instance with the TP). Therefore, UO policy is time inconsistent in the sense of Kydland and Prescott (1977).

It is an open question whether one is able to identify actual policies that may have been (approximately) optimal from an unconditional perspective. However, there is circumstantial evidence that sometimes transition costs are incurred following major policy changes, that is, when the current generation is forced to suffer for a better future. It happens during wartime. It also often appears to happen during pension reform; all future generations will benefit from a less distorted economy if the current generation sacrifices part of their pension benefits. Despite the costly transition, more than eighty countries recently undertook some degree of pension reform (see Holzmann and Hinz 2005). Moreover, although the transition cost is clear in a deterministic environment, it may be less visible, and therefore more politically implementable, when an economy is subject to stochastic shocks and the favorable shocks can compensate the present generation for lower investment made by the previous generation.

So, with only backward-looking constraints, TP policy, unlike UO policy, is in a sense "stable"; it is time consistent in the sense of Kydland and Prescott (1977) and credible. However, matters are somewhat different when forward-looking structural equations are present, as we now discuss.

3.4 Forward-Looking Models

As Soderlind (1999) shows, the best time-invariant policy depends on initial conditions at the time when the decision is made. Moreover, the initial conditions are changing over time so that there is always an incentive to deviate from any time-invariant policy. DDN (2008) show that UO policy maximizes objectives over all possible initial conditions, which is presented as the history of the realization of the exogenous shocks. UO policy acknowledges that the initial conditions depend on the policy run by predecessors and internalizes this. Another feature of the UO policy is that it maximizes over all initial conditions implied by the policy run by the same policymaker acting optimally in the past. In contrast, TP policy commits to time-zero expectations in the same way (that is, with the same functional form) as in the future. However, initial conditions are not necessarily the steady state, but can be any state. What is key is that TP does not internalize initial conditions. We now set these differences out explicitly.

Any policy ϕ together with constraints (2) define choice variables x_t as a function of initial values and shocks e_t:

$$\phi : \{x_{t-1,} e_t\} \rightarrow x_t.$$

Therefore, policy ϕ generates a distribution of initial conditions, $F_\phi(x_t)$. The timeless-perspective methodology takes that distribution as given and ignores the fact that policy influences the distribution of initial conditions. In particular, Woodford (2002, p. 509) decomposes the objective function into two components:

$$L = L^{\mathrm{det}} + L^{\mathrm{stab}},$$

where L^{det} depends on initial conditions, and L^{stab} depends only on the responses to unexpected shocks. He explains that the TP method minimizes L^{stab} and so does not internalize the influence of policy on the distribution of initial conditions. Hence,

$$\phi^{TP} = \arg\min L^{stab}(\phi).$$

UO policy takes the influence of ϕ on the distribution of initial conditions into account. The simplest way to see this is by noting that the distribution of $L^{\det}$ depends on the policy which was implemented by predecessors. In particular, $L^{\det}$ can be completely defined once we know the complete history of shocks and the policy: $L^{\det} = L^{\det}(\phi, e_{t-})$. Therefore, UO policy is defined as

$$\phi^{UO} = \arg\min\left(\int L^{\det}(\phi, e_{t-})de_{t-} + L^{stab}(\phi)\right).$$

In other words, UO methodology internalizes the initial distribution and delivers the policy which is best on average.

3.5 Stationarity

In contrast to the TP, UO policy induces stationarity. For example, under the TP it is optimal to permit permanent increases in debt and taxes following structural shocks under nominal rigidity (see Benigno and Woodford 2004, 2006a, and Schmitt-Grohe and Uribe 2004). However, Horvath (2011) finds that in the log-linearized unconditionally optimal economy, public debt converges to its steady state following a shock. Clearly, the UO approach is not applicable to non-stationary policies. Any policy which causes some variable to evolve as a unit root would generate unlimited unconditional losses, and such a policy would not be adopted by a policymaker with a UO perspective.

An alternative measure of unconditionally optimal policy is proposed in Benigno and Woodford (2012) in which the unconditional expectations operator is applied to a stationary sub-space of all variables. The values and expected values of unit-root variables are treated as initial conditions and therefore are classified as "terms independent of policy," or "t.i.p." Specifically, all predetermined variables are split into "trend" and "cyclical" components, where the "trend" consists of all non-stationary variables. The unconditional measure is then applied to the "cyclical" component only. How does this alternative UO policy differ from the one developed in the current paper? Consider two policies. Assume that they generate the same volatility of inflation. However, policy 1 ($P1$) induces a unit root in output, while policy 2 ($P2$) induces stationary output with finite volatility. According to UO policy as proposed in Benigno and Woodford (2012), $P1$ is to be preferred to $P2$ since the volatility of output will be counted as part of trend output and will not have an impact on the ranking of alternative, feasible policies. Therefore, an economy with infinitely volatile output will be preferred to an economy with low output volatility, for a given volatility of inflation.[11]

Alternatively, under the standard UO measure (as defined in section 2), $P1$ will be rejected in favor of $P2$. That is because the rank of the "trend" sub-space depends on policy design and will be internalized by the policymaker optimizing unconditional losses.

4. The Possibility of Pure Second-Order Approximation

In this section we show that it is possible to construct a pure second-order approximation to a general unconditional optimization problem (1), subject to constraints (2). We formulate it in the following proposition.

PROPOSITION 2. *It is always possible to approximate unconditional welfare up to second order around the UO steady state* (X, ξ)*, defined by the system (11)–(12).*

Proof. The value of the loss function $El(x_t)$ should not change if combined with the unconditional expectation of the constraints $EF(y_t, x_t, \mu_t)$. Thus, appendix 1 demonstrates that the second-order approximation to this combination has a pure second-order form. That is,

$$\begin{aligned} El(x_t, \mu_t) &= E\left[l(x_t) + \xi F(y_t, x_t, \mu_t)\right] \\ &= EQ_l + \xi EQ_F + t.i.p + O_3. \end{aligned} \tag{19}$$

The notation O_3 denotes third- or higher-order terms. Q_l and Q_F are pure second-order terms of the log-approximation, around the

[11] To be clear, what matters for welfare is the discounted value of second moments, which might be finite even if volatility is infinite. This is why unit-root processes might be allowed.

unconditionally optimal steady state, to the loss function $l(x_t)$ and dynamic constraints $EF(y_t, x_t, \mu_t)$:

$$Q_l = \frac{1}{2}\left(X^2 \frac{\partial^2 l}{\partial x^2}\widehat{x}_t\widehat{x}_t\right);$$

$$Q_F = \frac{1}{2}X^2\left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}\right)\widehat{x}_t\widehat{x}_t + XX\frac{\partial^2 F}{\partial x \partial y}\widehat{x}_t\widehat{x}_{t+1} + X\mu\frac{\partial^2 F}{\partial x \partial \mu}\widehat{x}_t\widehat{\mu}_t$$
$$+ X\mu\frac{\partial^2 F}{\partial y \partial \mu}\widehat{x}_{t+1}\widehat{\mu}_t,$$

where we use $\widehat{x}_t$ to denote a log-deviation from steady state.

It is straightforward to show that the maximization of the unconditional objective (19) subject to the linearized analogues of equations (2) yields the same solution as log-linearization of the first-order conditions (9). This latter approach is proposed by Khan, King, and Wolman (2003) in the context of conditional optimization, and is extended in DDN (2008) to unconditional optimization. Implementing the above result turns out to be fairly straightforward both from a numerical perspective and an analytical perspective.

4.1 Numerical Methodology

Any model economy can be described by the agents' behavioral dynamics (2) and the policy rule (3). Log-linearization of those two equations can be presented in the form of

$$E_t V_{t+1} = AV_t + B\varepsilon_{t+1},$$

where V_t is the vector of endogenous variables and ε_{t+1} is the vector of exogenous shocks. In this form it is straightforward to construct the variance-covariance matrix, $R \equiv EV_tV_t'$, using standard software.[12] That is, R is recovered by solving the following matrix equation:

$$R = ARA' + B\Upsilon B', \tag{20}$$

[12]For example, Dynare gives the variance-covariance matrix as part of its standard output.

where $\Upsilon = E\varepsilon_t\varepsilon_t'$ is the unconditional variance-covariance matrix of the underlying shock processes. Equation (20) can be solved numerically using a doubling algorithm as described in Anderson et al. (1996) using an equivalent form

$$R = \sum_{j=0}^{+\infty} A^j B \Upsilon B' A'^j.$$

The social welfare function can be computed as a linear combination of the elements of matrix R.

4.2 Substitution Techniques for UO and TP Policies

Although the method of pure second-order approximation (19) is straightforward and quite efficient, it may be useful to show how one can replicate the same welfare analysis by substituting variables employing the dynamic constraints (2). In particular, it demonstrates that even though UO policy cannot ignore initial conditions, that does not prevent one from using a substitution approach for UO policy analysis. Consider a second-order approximation to the dynamic constraint equations,

$$\widehat{x}_{t+1} = \alpha \widehat{x}_t + \widehat{y}_t + Q_t + O_3, \tag{21}$$

where Q_t is a pure quadratic form.

We first discuss the TP methodology which has been important for recovering the relationship between means and variances in second-order approximations useful for welfare analysis.[13] The TP substitution methodology expresses the discounted sum of $\{\widehat{x}_{t+s}\}_{s=0}^{+\infty}$ as a function of $\{\widehat{y}_{t+s}\}_{s=0}^{+\infty}$. In that case, equation (21) is integrated forward to yield

$$\sum_{s=0}^{+\infty} \beta^s \widehat{x}_{t+1+s} = a \sum_{s=0}^{+\infty} \beta^s \widehat{x}_{t+s} + \sum_{s=0}^{+\infty} \beta^s \widehat{y}_{t+s} + \sum_{s=0}^{+\infty} \beta^s Q_{t+s} + O_3.$$

[13] Sutherland (2002) was the first to apply this approach to the case of an economy with a distorted steady state and a particular policy rule. See also Kim and Kim (2003).

That expression can be simplified as

$$\left(\beta^{-1} - a\right) \sum_{s=0}^{+\infty} \beta^s \widehat{x}_{t+s} - \beta^{-1} \widehat{x}_t = \sum_{s=0}^{+\infty} \beta^s \widehat{y}_{t+s} + \sum_{s=0}^{+\infty} \beta^s Q_{t+s} + O_3.$$

Then an initial value, $\widehat{x}_t$, is ignored as a "t.i.p." and the final expression appears as

$$\sum_{s=0}^{+\infty} \beta^s \widehat{x}_{t+s} = \frac{1}{\beta^{-1} - a} \sum_{s=0}^{+\infty} \beta^s \widehat{y}_{t+s} + \frac{1}{\beta^{-1} - a} \sum_{s=0}^{+\infty} \beta^s Q_{t+s} + O_3.$$

This expression is then used to calculate approximate utility.

To deliver the analogous expression in the case of UO policy, one applies the unconditional expectations operator to (21),

$$E\widehat{x}_{t+1} = E\alpha\widehat{x}_t + E\widehat{y}_t + EQ_t + O_3. \tag{22}$$

Then, one uses the fact that $E\widehat{x}_{t+1} = E\widehat{x}_t$, which transforms (22) into

$$E\widehat{x}_t = \frac{1}{1-a} E\widehat{y}_t + \frac{1}{1-a} EQ_t + O_3, \tag{23}$$

which is the desired expression.

5. Example: Calvo Model with Distorted Steady State

A more or less canonical dynamic New Keynesian model is now developed and two issues in particular are pursued. First, which model variables appear in the approximate loss function under UO policy? Second, some insight is sought into the nature of UO monetary policy compared with TP policy. That comparison is pursued further in section 7.

5.1 The Households

There is a large number of identical agents in this (closed) economy where the only input to production is labor. Each agent evaluates utility using the following criterion:

$$E_0\sum_{t=0}^{\infty}\beta^t U(Y_t, N_t(i)) = E_0\sum_{t=0}^{\infty}\beta^t\left(\log(Y_t) - \frac{\lambda}{1+\upsilon}\left(\int_i N_t(i)di\right)^{1+\upsilon}\right). \tag{24}$$

E_t denotes the conditional expectations operator at time $t \geq 0$, β is the discount factor, Y_t is consumption, and $N_t(i)$ is the quantity of labor supplied to industry i; labor is industry specific. $\upsilon \geq 0$ measures the labor supply elasticity, while λ is a preference parameter.

Consumption is defined over a Dixit-Stiglitz basket of goods,

$$Y_t = \left[\int_0^1 Y_t(i)^{\frac{\theta-1}{\theta}}di\right]^{\frac{\theta}{\theta-1}}. \tag{25}$$

The average price level, P_t, is known to be

$$P_t = \left[\int_0^1 p_t(i)^{1-\theta}di\right]^{\frac{1}{1-\theta}}. \tag{26}$$

The demand for each good is given by

$$Y_t(i) = \left(\frac{p_t(i)}{P_t}\right)^{-\theta} Y_t^d, \tag{27}$$

where $p_t(i)$ is the nominal price of the final good produced in industry i and Y_t^d denotes aggregate demand.

Agents face the flow constraint

$$P_tY_t + B_t = (1 + i_{t-1})\,B_{t-1} + (1-\tau)\,W_tN_t + \Pi_t. \tag{28}$$

As all agents are identical, the only financial assets traded in equilibrium will be those issued by the fiscal authority. Here B_t denotes the nominal value of government bond holdings, at the end of date t; $1 + i_t$ is the nominal interest rate on this "riskless" one-period nominal asset; W_t is the nominal wage in period t (our assumptions mean that we do not need to index wages on i); and Π_t indicates any profits remitted to the individual. It is assumed that labor income is taxed at rate τ. The usual conditions are assumed to apply to the consumer's limiting net savings behavior. Hence, necessary conditions for an optimum include

$$-\frac{U_N'(Y_t, N_t)}{U_Y'(Y_t, N_t)} = \lambda N_t^v Y_t = (1-\tau)\, w_t, \tag{29}$$

$$w_t = \frac{\lambda}{1-\tau} N_t^v Y_t, \tag{30}$$

and

$$E_t \left\{ \frac{\beta U_Y'(Y_{t+1}, N_{t+1})}{U_Y'(Y_t, N_t)} \frac{P_t}{P_{t+1}} \right\} = \frac{1}{1+i_t}. \tag{31}$$

Here w_t denotes the real wage. The complete-markets assumption implies the existence of a unique stochastic discount factor,

$$Q_{t,t+k} = \beta \frac{Y_t P_t}{Y_{t+k} P_{t+k}}, \tag{32}$$

where

$$E_t \left\{ Q_{t,t+k} \right\} = E_t \prod_{j=0}^{k} \frac{1}{1+i_{t+j}}.$$

5.2 Representative Firm: Factor Demand

As noted, labor is the only factor of production. Firms are monopolistic competitors who produce their distinctive goods according to the following technology:

$$Y_t(i) = A_t \left[N_t(i) \right]^{1/\phi}, \tag{33}$$

where $N_t(i)$ denotes the amount of labor hired by firm i in period t, A_t is a stochastic productivity shock, and $1 < \phi$.

The demand for output determines the demand for labor. Hence one finds that

$$N_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\theta\phi} \left(\frac{Y_t}{A_t} \right)^{\phi}. \tag{34}$$

There is an economy-wide labor market so that all firms pay the same wage for the same labor. As a result, as asserted above, one may write $w_t(i) = w_t, \forall i$. All households provide the same share of labor to all firms. The total amount of labor will then be

$$N_t = \int N_t(i)di = \left(\frac{Y_t}{A_t}\right)^{\phi} \int \left(\frac{P_t(i)}{P_t}\right)^{-\theta\phi} di = \left(A_t^{-1} Y_t\right)^{\phi} \Delta_t, \quad (35)$$

where Δ_t is the measure of price dispersion:

$$\Delta_t \equiv \int_0^1 \left(\frac{P_t(i)}{P_t}\right)^{-\theta\phi} di. \quad (36)$$

5.3 Representative Firm: Price Setting

As in Calvo (1983), each period a fixed proportion of randomly chosen firms is allowed to adjust prices. Those firms choose the nominal price which maximizes their expected profit given that they have to charge the same price in k periods' time with probability α^k. As usual, we assume that firms are cost takers. Let $p_t'(i)$ denote the choice of nominal price by a firm that is permitted to reprice in period t.

Real profits can be written as

$$\frac{P_t(i)}{P_t} Y_t(i) - \mu_t w_t N_t(i), \quad (37)$$

where μ_t is a cost-push shock so that the total cost facing firm i will be $TC(i) = \mu_t w_t N_t(i)$. In combination with the production function, $TC(i) = \mu_t w_t \left(\frac{Y_t(i)}{A_t}\right)^{\phi}$, and therefore marginal cost for a particular firm i is $c(i) = \phi \mu_t w_t \left(\frac{Y_t(i)}{A_t}\right)^{\phi} / Y_t(i)$.

As all firms that are permitted to reprice will choose the same price, optimal repricing implies

$$\left(\frac{p_t'}{P_t}\right)^{1+\theta(\phi-1)} = \frac{\left(\frac{\theta}{\theta-1}\right) \sum_{k=0}^{\infty} (\alpha\beta)^k Y_{t+k}^{-1} \times \left[\phi \mu_{t+k} w_{t+k} A_{t+k}^{-\phi} Y_{t+k}^{\phi} (P_t/P_{t+k})^{-\theta\phi}\right]}{\sum_{k=0}^{\infty} (\alpha\beta)^k (P_t/P_{t+k})^{1-\theta}}. \quad (38)$$

The price index then evolves according to the law of motion,

$$P_t = \left[(1-\alpha)\, p_t'^{1-\theta} + \alpha P_{t-1}^{1-\theta}\right]^{1/(1-\theta)}. \tag{39}$$

Because the relative prices of the firms that do not change their prices in period t fall by the rate of inflation, the law of motion for the measure of price dispersion is

$$\Delta_t = \alpha \Delta_{t-1} \pi_t^{\theta\phi} + (1-\alpha)\left(p_t'/P_t\right)^{-\theta\phi}. \tag{40}$$

6. UO Monetary Policy

Proposition 3 sets out the relevant UO Ramsey problem.

PROPOSITION 3. *The UO Ramsey plan is a choice of state-contingent paths for the endogenous variables* $\{\pi_{t+k}, \Delta_{t+k}, p_{t+k}, c_{t+k}, X_{t+k}, Z_{t+k}\}_{k=0}^{\infty}$ *from date* t *onwards given* $\{E_t A_{t+k}, E_t \mu_{t+k}\}_{k=0}^{\infty}$*, so as to maximize social welfare function (41) subject to constraints (42)–(45):*

$$\max E E_t \sum_{k=0}^{\infty} \beta^k \left(\frac{\log\left(c_{t+k}\right)}{\phi} - \frac{v}{\phi} \log \Delta_{t+k} - (1-\tau) \frac{c_{t+k}\Delta_{t+k}}{\mu_{t+k}} \right), \tag{41}$$

subject to the following:

- *the Phillips block*

$$\left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi-\theta+1}{1-\theta}} X_t = \frac{\theta}{\theta-1} Z_t; \tag{42}$$

$$X_t = 1 + \alpha\beta E_t X_{t+1} \pi_{t+1}^{\theta-1}; \tag{43}$$

$$Z_t = c_t + \alpha\beta E_t Z_{t+1} \pi_{t+1}^{\theta\phi}. \tag{44}$$

- *the law of motion of prices*

$$\Delta_t = \alpha \Delta_{t-1} \pi_t^{\theta\phi} + (1-\alpha) \left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi}{\theta-1}}. \tag{45}$$

Here c_t is real marginal cost for the firm which produces output Y_t,

$$c_t = \phi \frac{\lambda}{1-\tau} \mu_t \Delta_t^v \left(A_t^{-1} Y_t \right)^{(v+1)\phi}; \tag{46}$$

discounted marginal revenue is $X_t := E_t \sum_{k=0}^{\infty} (\beta\alpha)^k \left(\frac{P_t}{P_{t+k}} \right)^{1-\theta}$; and discounted marginal cost is $Z_t := E_t \sum_{k=0}^{\infty} (\beta\alpha)^k c_{t+k} \left(\frac{P_t}{P_{t+k}} \right)^{-\theta\phi}$.

6.1 The Steady State

We now turn in more detail to steady-state analysis. In appendix 1 we solve the policymaker's problem defined in proposition 3 and establish the following result.

PROPOSITION 4. *The steady-state inflation is positive,* $\pi \geq 1$. *Price stability is only optimal if either* $\beta = 1$ *or if* $\frac{\theta-1}{\theta}\frac{1-\tau}{\mu} = 1$ *(which corresponds to the non-distorted steady state). Moreover,* π *is unique and bounded:* $\pi \leq \min(\beta^{1/(\theta-1-\phi\theta)}, \alpha^{-1/(\phi\theta)})$.

Proof. See appendix 1.

Our result shows that UO policy delivers a different equilibrium inflation than TP optimal policy. It is well known (see Benigno and Woodford 2005) that TP optimal policy requires price stability in the steady state. The UO policy delivers a trend inflation. The intuition follows from the fact that whilst higher inflation induces price-setting firms to choose a higher markup, firms holding prices constant will see their markup erode more quickly; one effect acts to boost demand and the other to reduce it. King and Wolman (1999) show that a slightly positive inflation rate maximizes steady-state welfare by reducing the markup distortion (inverse of marginal cost). In our model the steady-state value of real marginal cost as a function of inflation is obtained from the Phillips relation (42)–(44):

$$c = \frac{1-\alpha\beta\pi^{\theta\phi}}{1-\alpha\beta\pi^{\theta-1}} \frac{\theta-1}{\theta} \left(\frac{1-\alpha\pi^{\theta-1}}{1-\alpha} \right)^{\frac{\theta\phi-\theta+1}{1-\theta}}.$$

It is easy to see that it increases with inflation at $\pi = 1$, and it follows that

$$\frac{d \log c}{d\pi}\pi = (\theta - 1)\,\phi\left(\frac{\alpha\pi^{\theta-1} - \alpha\beta\pi^{\theta\phi}}{(1-\alpha\pi^{\theta-1})\,(1-\alpha\beta\pi^{\theta\phi})}\right)$$
$$= \frac{(\theta-1)\,\phi\alpha(1-\beta)}{(1-\alpha)\,(1-\alpha\beta)} > 0.$$

This shows that a small increase in inflation will reduce the price-to-marginal-cost ratio and reduce the monopolistic distortion in the economy. In fact, the following proposition is true:

Proposition 5. *Steady-state inflation increases with the distortion, measured as* $1 - \frac{\theta-1}{\theta}\frac{1-\tau}{\mu}$, *and declines in the discount factor* β *and the labor elasticity,* v.

Proof. See appendix 1.

Using parameter values typically found in the literature, we find that the optimal steady-state inflation is of the order of 0.2 percent a year.

So, on the one hand, a small amount of inflation can boost demand, as it partially offsets the markup distortion. On the other hand, price dispersion, which is rising in inflation, acts rather like a cost shock on firms, for reasons analyzed in Damjanovic and Nolan (2010). Hence, one finds that optimal trend inflation has a U-shaped relation to price stickiness, α; it is increasing in α when initial price dispersion is relatively small, and declines once initial price dispersion is sufficiently large. Optimal inflation declines in the discount factor, β. As discussed DDN (2008) and emphasized earlier, UO policy—in contrast to timeless-perspective policy—gives some weight to the distribution of initial conditions. In particular, it considers the distribution of the initial output gap, which is a time-invariant ergodic distribution imposed by chosen policy. That is partly why some stimulation of output via inflation is desirable. So the smaller the discount factor, the higher is the relative weight on initial conditions and the higher the optimal inflation rate.

6.1.1 Contrast to TP Policy

In the above example, UO policy exploits the Phillips curve to reduce the markup. To understand the difference with the TP approach, consider the following example.

EXAMPLE 2. *Assume that the government problem can be formulated in terms of the following social objective:*

$$\max E_t \sum_{k=0}^{\infty} \beta^k u(c_t, \pi_t),$$

and a price-setting constraint

$$E_t f(c_t, \pi_t, \pi_{t+1}) = 0, \tag{47}$$

where c_t *is real marginal costs,* π_t *is inflation, and* $u(c_t, \pi_t)$ *is the period social objective. Social welfare increases with* $c, u_c > 0$ *and achieves its maximum when prices are stable:*

$$u_\pi(1) = 0; \tag{48}$$

$u_{\pi\pi}(1) < 0$. *The second equation, (47), is a Phillips curve and has the following properties:*

$$-\frac{\partial f}{\partial \pi_t} = \beta \frac{\partial f}{\partial \pi_{t+1}}; \tag{49}$$

$$\frac{\partial f}{\partial \pi_t} < 0; \quad \frac{\partial f}{\partial c_t} > 0. \tag{50}$$

Equation (49) reflects the intertemporal trade-off between current and future inflation, while (50) suggests a positive correlation between inflation and marginal cost. Those properties are general and satisfied in Calvo, Rotemberg, and discounted Taylor frameworks.

The UO program is set out and solved:

$$H^{UO} = u(c, \pi) - \lambda f(c, \pi, \pi);$$

$$\frac{\partial H^{UO}}{\partial c} = \frac{\partial u}{\partial c} - \lambda \frac{\partial f}{\partial c} = 0; \tag{51}$$

$$\frac{\partial H^{UO}}{\partial \pi} = \frac{\partial u}{\partial \pi} - \lambda\left(\frac{\partial f}{\partial \pi_t} + \frac{\partial f}{\partial \pi_{t+1}}\right) = 0. \tag{52}$$

Since $\frac{\partial u}{\partial c} > 0$, and $\frac{\partial f}{\partial c} > 0$, the Lagrange multiplier, λ, is positive and the last equation can be rewritten as

$$\frac{\partial H^{UO}}{\partial \pi} = \frac{\partial u}{\partial \pi} + \lambda(1-\beta)\frac{\partial f}{\partial \pi_{t+1}} = 0.$$

Therefore, in equilibrium there is a positive inflation, $\pi > 1$, since $\frac{\partial u}{\partial \pi} < 0$.

Unconditional optimization uses positive inflation to reduce the markup, which is equivalent to increasing steady-state marginal cost. The implicit function theorem is applied to the price-setting equation to show that

$$\frac{dc}{d\pi} = -\left(\frac{\partial f}{\partial \pi_t} + \frac{\partial f}{\partial \pi_{t+1}}\right)\Big/\frac{\partial f}{\partial c_t} = (1-\beta)\,\frac{\partial f}{\partial \pi_{t+1}}\Big/\frac{\partial f}{\partial c_t} > 0,$$

and this is exploited in order to improve welfare.

The TP approach, on the other hand, ignores the relation between inflation and marginal costs in the Phillips curve and does not use inflation to stimulate the economy by reducing the inefficient markup. The TP program and solution is

$$L^{TP} = E_t\sum_{k=0}^{\infty}\beta^k\left(u(c_{t+k},\pi_{t+k}) - \lambda_{t+k}E_t f(c_{t+k},\pi_{t+k},\pi_{t+1+k})\right);$$

$$\frac{\partial L^{TP}}{\partial c_t} = \frac{\partial u}{\partial c} - \lambda_t\frac{\partial f}{\partial c} = 0;$$

$$\frac{\partial L^{TP}}{\partial \pi_{t+1}} = \frac{\partial u}{\partial \pi} - \lambda_{t+1}\beta\frac{\partial f}{\partial \pi_t} - \lambda_t\frac{\partial f}{\partial \pi_{t+1}} = 0. \tag{53}$$

Combining (49) with (53) in steady state shows that TP policy chooses inflation to optimize intertemporal utility in the steady state:

$$\frac{\partial u}{\partial \pi} = 0.$$

It is apparent that the effect of inflation on the reduction of the price markup is not utilized when TP optimization is applied.

Finally, we note that price stability only occurs under the TP if conditions (48) and (49) are satisfied when the price setting is one period forward looking. In a more general case, if the Phillips relation should satisfy $\sum_{k=-\infty}^{\infty}\beta^{-k}\frac{\partial f}{\partial \pi_{t+1}}=0$, that may not be so for a wide range of models. For example, it may not be the case for an open economy (see Benigno and Lopez-Salido 2006).

6.2 The Quadratic Form

Having recovered the optimal steady state, one can obtain a quadratic loss function of the form (19). The quadratic welfare derived from (41) and constraints (42)–(45) is

$$EU = -\frac{1}{2}E\left(h\Phi\widehat{u}_t^2 + (1-h\Phi)\widehat{c}_t^2 + \Lambda_x\widehat{X}_t^2 + \Lambda_\pi\widehat{\pi}_t{}^2 + \Lambda_\Delta\widehat{\Delta}_t^2\right), \tag{54}$$

where $\widehat{u}_t = \widehat{c}_t + \widehat{\Delta}_t - \widehat{\mu}_t$ is the log-linearized disutility of labor, $u_t = -\frac{\lambda N_t^{v+1}}{v+1}$.

For the model at hand, one can show that it can be written as follows:

$$EU= -\frac{1}{2}E\left[\phi\left(1+v\right)\left(\widehat{Y}_t - \widehat{Y}_t^*\right)^2 + G\widehat{g}_t^2 + \Lambda_x\widehat{X}_t^2 + \Lambda_\Delta\widehat{\Delta}_t^2 + \Lambda_\pi\widehat{\pi}_t{}^2\right], \tag{55}$$

where $\widehat{g}_t = \widehat{\mu}_t - \widehat{\Delta}_t$. $\widehat{g}_t$ has an intuitive interpretation as the log-deviation of the ratio of natural output to labor cost. To see that, note that if prices were flexible ($\alpha = 0$), the price-setting condition (38) will result in the following level of output:

$$Y_t^n = \frac{\theta}{\theta-1}\frac{\mu_t}{\Delta_t}\phi w_t N_t. \tag{56}$$

Therefore, if we define $g_t := \frac{Y_t^n}{w_t N_t} = \phi\frac{\theta}{\theta-1}\frac{\mu_t}{\Delta_t}$, that becomes a relationship between the log-deviation of the cost-push shock and relative price dispersion.

The term $\widehat{Y}_t^*$ represents the "target" level of output $Y_t^* := \widehat{A}_t - v_\mu\widehat{\mu}_t - v_\Delta\widehat{\Delta}_t$. Details concerning coefficients are given in appendix 2. The "target" rate is increasing in productivity and declining in the cost-push shock; it is also declining in price dispersion. The variable $\widehat{X}_t$ represents the losses to firms forced to charge sub-optimal prices due to price stickiness and expected inflation, to which they may not be able to react.

This form of the loss function can easily be nested to familiar cases, either the non-distorted steady state where $\Phi = 1$, or where the steady state of the model economy remains distorted but where the social discount rate is equal to the private rate of discount, $\beta = 1$ (in which case the UO policy and the timeless-perspective policies coincide). In both special cases, optimal monetary policy corresponds to price stability, and the loss function (55) reduces to a familiar form defined simply over inflation and output. Specifically, if the optimal steady state is characterized by price stability, then $\Lambda_x = 0$. Moreover, one can easily show that price dispersion, $\widehat{\Delta}_t$, is a second-order term in that case. Lastly, the labor wedge $\widehat{g}_t$ is then simply a cost-push shock, $\widehat{\mu}_t$, and can be considered as a term independent of policy.

7. Application: Unconditional Ordering of Simple Rules

The foregoing approach is easily used to evaluate simple rules for monetary policy and to highlight the potential significance for policy design of a distorted steady state. First, write the model in vector autoregressive form as follows:

$$E_t\widehat{\pi}_{t+1} + E_t\widehat{Y}_{t+1} = \widehat{Y}_t + (1-\beta)\widehat{i}_t; \tag{57}$$

$$\beta\alpha\pi^{\theta\phi}E_t\left(\widehat{Z}_{t+1} + \theta\phi\widehat{\pi}_{t+1}\right) = \widehat{Z}_t - \left(1-\alpha\beta\pi^{\theta\phi}\right)\widehat{c}_t; \tag{58}$$

$$\beta\alpha\pi^{\theta-1}E_t\left(\widehat{X}_{t+1} + (\theta-1)\,\widehat{\pi}_{t+1}\right) = \widehat{X}_t; \tag{59}$$

$$\widehat{Z}_t - (\theta\phi - \theta + 1)\,\frac{\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta-1}}\widehat{\pi}_t - \widehat{X}_t = 0; \tag{60}$$

$$-\widehat{g}_t + \widehat{\Delta}_t - \widehat{\mu}_t = 0; \tag{61}$$

$$-\widehat{c}_t + v\widehat{\Delta}_t + (v+1)\,\phi(\widehat{Y}_t - \widehat{A}_t) + \widehat{\mu}_t = 0; \tag{62}$$

$$\widehat{\Delta}_{t+1} + \theta\phi\frac{\alpha\pi^{\theta-1} - a\pi^{\theta\phi}}{1-\alpha\pi^{\theta-1}}\widehat{\pi_{t+1}} = a\pi^{\theta\phi}\widehat{\Delta}_t; \tag{63}$$

$$-Y_t^* + \widehat{A}_t - v_\mu \widehat{\mu}_t - v_\Delta \widehat{\Delta}_t = 0; \tag{64}$$

$$-\widehat{i}_t = \phi_\pi \widehat{\pi}_t + \phi_f \widehat{f}_t + m_t. \tag{65}$$

In the above linearized system of equations, the final equation (65) is the policy rule, where $\widehat{i}_t$ is the gross nominal interest rate, $\widehat{i}_t = \log\left(\frac{\beta}{\pi}(1+i_t)\right)$, and $\widehat{f}_t$ represents a linear combination of policy feedback variables, while m_t is a policy shock.[14]

It is clear that steady-state distortions complicate the policy problem so far as the policymaker's objective function is concerned.[15] However, does it make any difference so far as the design of simple rules are concerned?[16]

First, a simple interest rate feedback rule is considered, where the interest rate responds to current and lagged inflation only. The feedback on current inflation is fixed at $\phi_\pi = 1.5$. Given this, the optimized weight on lagged inflation, $f = \widehat{\pi}_{t-1}$, is computed. It is about 15 in the distorted steady-state case and 14 for the non-distorted steady state. However, the difference in welfare between responding and not responding to lagged inflation is up to 0.16 percentage points in terms of consumption-equivalent units (see the top-right graph in figure 1; ϕ_f is at its optimal value). As in the TP approach, relative price distortion is very costly, and the optimal simple rule may be very close to price stability ($\phi_\pi = +\infty$). However, if for any reason the policy reaction on current inflation is restricted, the economy may significantly benefit from a response to lagged inflation.

One can also show that the optimal feedback on output should be slightly *negative*, $\phi_f = -0.015$. Furthermore, inclusion of real output targeting leads to very modest welfare improvements, in the

[14]The following parameterization is used in the quantitative investigation: $\beta = 0.9, v = 1.1, \theta = 7, \alpha = 0.5$, and $\phi = 1.3$. It is assumed that shocks A_t, μ_t, and m_t follow $AR(1)$ processes with $\rho_A = 0.98, \sigma_A = 0.008, \rho_m = 0.9, \sigma_m = 0.005$, and $\rho_\mu = 0.9, \sigma_\mu = 0.02$.

[15]That is, complicates it relative to the objective function in the non-distorted case.

[16]In the particular model developed above, the UO trend inflation is rather small and the policy ordering across distorted and non-distorted steady states is often the same for given simple rules. However, in simulations not reported, it was possible to find simple, plausible rules that result in welfare "reversals"; that is, where rule 1 welfare dominates rule 2 in the distorted economy, but where the ranking switched in the non-distorted economy.

order of 10^{-3} compared with targeting inflation alone. This result is consistent with Schmitt-Grohe and Uribe (2007).

The results are summed up in figure 1 (where the broken line is the non-distorted economy).

7.0.1 Targeting Nominal Income Growth

Finally, inflation targeting and nominal income targeting are compared under a UO policy criterion as in Kim and Henderson (2005). Kim and Henderson suggest, in a model with one-period price stickiness, that nominal income targeting may have superior welfare properties to inflation targeting. Two rules are compared:

$$\text{Nominal income growth targeting: } i_t = 0.05\,(y_t - y_{t-1} + \pi_t) + (\phi_\pi - 0.05)\,\pi_t + m_t; \quad (66)$$

$$\text{Inflation targeting: } i_t = \phi_\pi \pi_t + m_t. \quad (67)$$

In the case of a non-distorted steady state and a "low" feedback on inflation, the findings are similar to some of Kim and Henderson's findings. Specifically, in the case of a distorted steady-state model, the net welfare gain from targeting nominal income growth over inflation targeting is positive. In the non-distorted case, inflation targeting is rarely dominated by nominal income targeting. In figure 2, the relative welfare gain (over inflation targeting) in targeting nominal income growth is plotted against ϕ_π.

The precise position of these net welfare schedules is quite sensitive to parameterization of the model (in particular, the persistence of shocks), but in general one finds that as the feedback on inflation rises, inflation targeting is likely to dominate nominal income targeting.

7.1 Comparing Optimal Policies

In this section we compare impulse responses when UO and TP policies are applied. The linear approximation to UO policy looks rather complex. It has three predetermined variables; besides inflation, there is price dispersion Δ_t and the Lagrange multiplier associated with one of the dynamic constraints (43). The complete system is presented in appendix 2. The TP policy system breaks naturally

Figure 1. Sub-Optimal Simple Policies

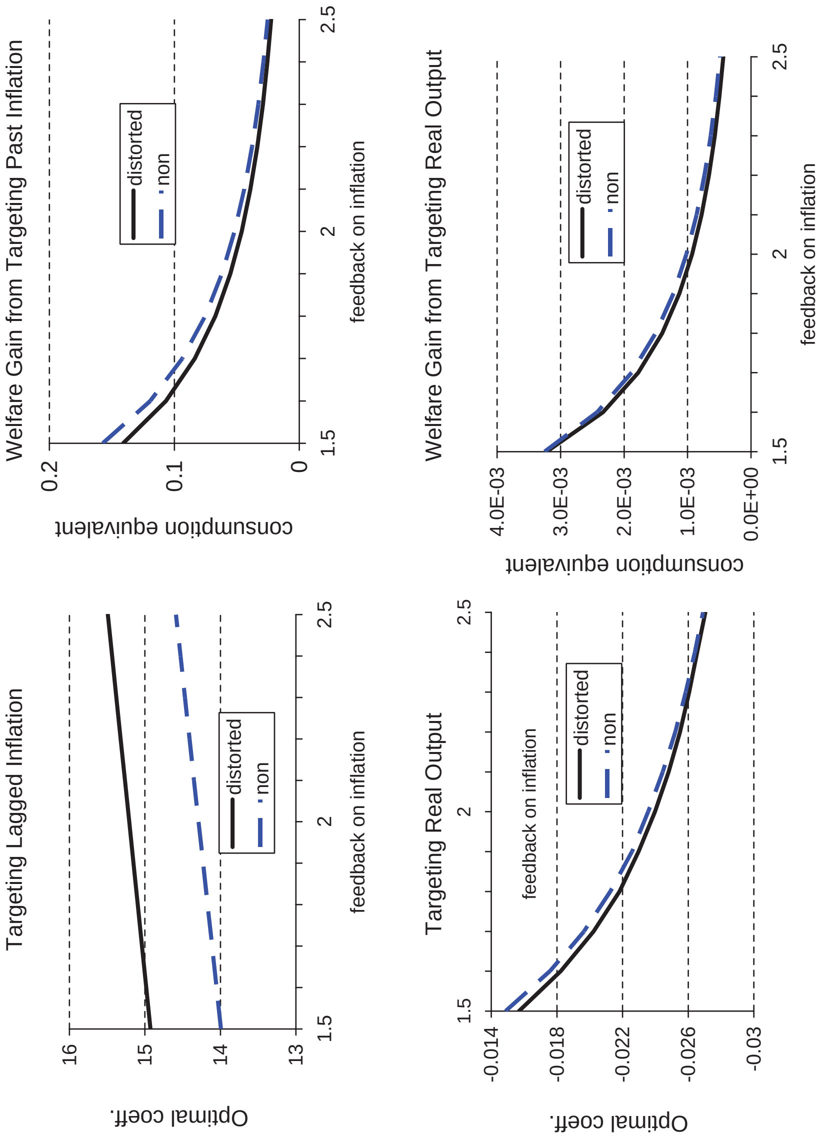
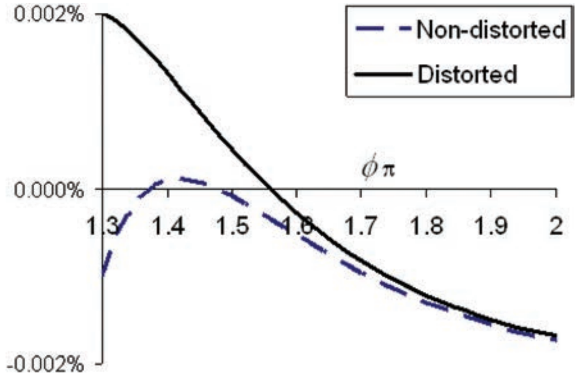


Figure 2. Relative Welfare Gain in Targeting Nominal Income Growth



into blocks, and the dynamics for inflation and marginal cost can be solved from these two-equation blocks as shown in appendix 2. However, due to the fact that β is very close to 1, we do not expect to see quantitatively a large difference across TP and UO policy.

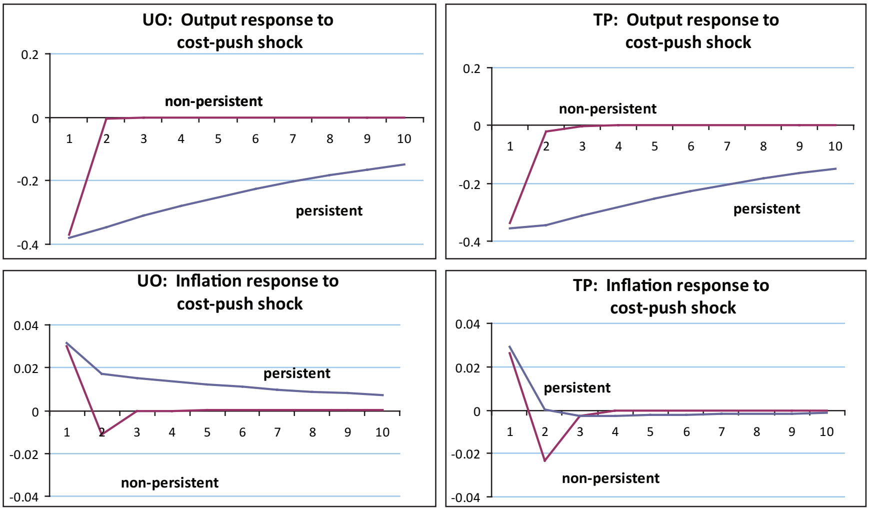
Note that {(41), (42)–(45)} does not include the productivity shock. Hence, one may conclude that neither UO nor TP policy should react to such shocks and that price/inflation stability will be optimal under both policies. Below we report the impulse responses from TP and UO policies following a cost-push shock. We ran two experiments, one with persistent ($\rho_\mu = 0.9$) and one with non-persistent ($\rho_\mu = 0$) shocks.[17]

Strict inflation targeting is sub-optimal under both approaches when a markup shock hits the economy. When the shock has low persistence, one observes overshooting in inflation, although this is absent when the shock is persistent. Perhaps the key point to observe is that when the shock is persistent, UO policy implies a slow rate of convergence of inflation.

Not surprisingly, perhaps, for the baseline New Keynesian model we find that the differences between TP and UO policies are small. In figure 3, they generate very similar responses in output to a cost-push shock, but slightly different responses in inflation. When the

[17]Let ρ_μ denote the autocorrelation coefficient; then the shock process that is posited is described by $\mu_t = \rho_\mu \mu_{t-1} + \varepsilon_t$, where $\varepsilon_t \tilde{} N(0, 1)$ is an i.i.d. shock.

Figure 3. UO and TP-Optimal Policies



shock is persistent, it affects the economy for a longer period of time, and the difference in social time discounting becomes more important in that case.

Finally, one observes that the price level returns to its initial value after the shocks under TP policy but not following UO policy. Thus, UO policy induces infinite variability in the price level (although inflation is stationary), whilst TP gives a stationary price.[18]

8. Conclusion

In this paper we analyze UO policies and compare them with TP policies. UO policy in spirit is actually very similar to the definition given when TP policy was introduced in Woodford (1999a, p. 19): "under the timeless perspective, one chooses to act as one believes one would have wished to commit oneself to act at a date far in the past." The TP acknowledges that commitment results in more favorable expectations (which are considered, in effect, as state variables).

[18]We thank the editor for highlighting this finding in our results.

But whilst the TP recognizes that expectations can be changed by a particular policy implementation and that that effect should be internalized, it rejects the notion that policy affects the distribution of the other state variables (like capital) and that that influence ought to be internalized also. So we think of the UO perspective pushing the original intuition for the TP one step further: One chooses to act as one believes one would have wished to commit oneself to act at a date far in the past not only with respect to private expectations but with respect to other state variables like capital.

The paper also demonstrates that one can formulate a purely quadratic approximate unconditional loss function to a model economy with a distorted steady state. It develops a straightforward, efficient approach to implementing the UO approach. It contrasts this approach to policy formulation with the TP approach, giving a number of examples where policies and objectives differ. It explores reasons why one may be interested in pursuing UO policies and the difficulties so encountered. In an application, it is shown that the loss function may be somewhat more complex than in a model with no steady-state distortions; inflation and output are no longer the sole arguments in the loss function. However, the loss function so obtained is easily interpreted in terms of the underlying distortions in the economy.

Appendix 1. A Second-Order Approximation of the Welfare Function

The first part of this appendix demonstrates the key result in section 4, namely the existence of the quadratic form, (19). The first line of the following block of equations corresponds to the top line of (19), the subsequent lines being its quadratic approximation:

$$\begin{aligned}
El\left(x_t\right) &= E\left[l\left(x_t\right)+\xi F\left(y_t,x_t,\mu_t\right)\right]\\
&= E\left(l+X\frac{\partial l}{\partial x}\widehat{x}_t+\frac{1}{2}\left(X^2\frac{\partial^2 l}{\partial x^2}+X\frac{\partial l}{\partial x}\right)\widehat{x}_t\widehat{x}_t\right)\\
&\quad+E\xi\left(F+X\frac{\partial F}{\partial x}\widehat{x}_t+X\frac{\partial F}{\partial y}\widehat{y}_t+\mu\frac{\partial F}{\partial \mu}\widehat{\mu}_t\right)\\
&\quad+\frac{1}{2}\xi\left(X\frac{\partial F}{\partial x}+X^2\frac{\partial^2 F}{\partial x^2}\right)E\widehat{x}_t\widehat{x}_t+\frac{1}{2}\xi\left(X\frac{\partial F}{\partial y}+XX\frac{\partial F}{\partial y^2}\right)
\end{aligned}$$

$$\times E\widehat{y}_t\widehat{y}_t + \frac{1}{2}\xi\left(\mu\frac{\partial F}{\partial x} + \mu^2\frac{\partial^2 F}{\partial x^2}\right)E\widehat{\mu}_t\widehat{\mu}_t$$
$$+ \xi E\left(XX\frac{\partial F}{\partial x\partial y}\widehat{x}_t\widehat{y}_t + X\mu\frac{\partial^2 F}{\partial x\partial \mu}\widehat{x}_t\widehat{\mu}_t + X\mu\frac{\partial F}{\partial y\partial \mu}\widehat{y}_t\widehat{\mu}_t\right)$$
$$+ O3.$$

Using the constraints $E_t x_{t+1} = y_t$, and the property of unconditional expectations that $Ez_{t+1} = Ez_t$, this can be rewritten as

$$El\left(x_t\right) = XE\widehat{x}_t\left(\frac{\partial l}{\partial x} + \xi\frac{\partial F}{\partial x} + \xi\frac{\partial F}{\partial y}\right)$$
$$+ \frac{1}{2}XE\widehat{x}_t\widehat{x}_t\left(\frac{\partial l}{\partial x} + \xi\frac{\partial F}{\partial x} + \xi\frac{\partial F}{\partial y}\right) + EQ_l + \xi EQ_F \quad (68)$$
$$+ l + \xi F + \xi\mu\frac{\partial F}{\partial \mu}E\widehat{\mu}_t + \frac{1}{2}\xi\left(\mu\frac{\partial F}{\partial x} + \mu^2\frac{\partial^2 F}{\partial x^2}\right)E\widehat{\mu}_t\widehat{\mu}_t$$
$$+ O3. \quad (69)$$

Here Q_l and Q_F are pure second-order terms:

$$Q_l = \frac{1}{2}X^2\frac{\partial^2 l}{\partial x^2}\widehat{x}_t\widehat{x}_t;$$
$$Q_F = \frac{1}{2}X^2\left(\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2}\right)\widehat{x}_t\widehat{x}_t + XX\frac{\partial^2 F}{\partial x\partial y}\widehat{x}_t\widehat{x}_{t+1} + X\mu\frac{\partial^2 F}{\partial x\partial \mu}\widehat{x}_t\widehat{\mu}_t$$
$$+ X\mu\frac{\partial^2 F}{\partial y\partial \mu}\widehat{x}_{t+1}\widehat{\mu}_t.$$

Furthermore, using the steady-state conditions (11), one can show that the first line of expression (68) equals zero. Moreover, expression (69) consists of $l + \xi F = l$, the steady-state value of the loss function and shocks. These are terms independent of policy (t.i.p.). Thus, it is proved that the loss function can be represented in a pure quadratic form.

$$El\left(x_t\right) = EQ_l + \xi EQ_F + t.i.p + O3.$$

UO Policy in a Distorted Calvo Model

In this section we apply the algorithm designed in the main text to the Calvo model summarized in section 6. One can set up the Hamiltonian for this problem, as proposed in section 2.1.1, as follows:

$$H = \left(\log\left(c_t\right) - v\log\Delta_t - \Delta_t\frac{\theta}{\theta-1}\Phi c_t\right) + \rho_t\left(X_t - 1 - \beta\alpha\pi_{t+1}^{\theta-1}X_{t+1}\right) + \varphi_t\left(-Z_t + c_t + \alpha\beta E_t Z_{t+1}\pi_{t+1}^{\theta\phi}.\right) + \xi_t\left(\left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi-\theta+1}{1-\theta}}X_t - \frac{\theta}{\theta-1}Z_t\right) + \eta_t\left(\Delta_t - \alpha\Delta_{t-1}\pi_t^{\theta\phi} - (1-\alpha)\left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi}{\theta-1}}\right).$$

The necessary conditions for an optimum include

$$\frac{\partial H}{\partial c_t} = \frac{1}{c_t} - \frac{\theta}{\theta-1}\Phi\Delta_t + \varphi_t;$$

$$\frac{\partial H}{\partial\Delta_t} = -\frac{v}{\Delta_t} - \frac{\theta}{\theta-1}\Phi c_t + \eta_t - E_t\alpha\eta_{t+1}\pi_{t+1}^{\theta\phi};$$

$$\frac{\partial H}{\partial X_t} = \rho_t - \rho_{t-1}\beta\alpha\pi_t^{\theta-1} + \xi_t\left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi-\theta+1}{1-\theta}};$$

$$\frac{\partial H}{\partial Z_t} = -\varphi_t + \varphi_{t-1}\beta\alpha\pi_t^{\theta\phi} - \xi_t\frac{\theta}{\theta-1};$$

$$\pi_t\frac{\partial H}{\partial\pi_t} = -\left(\theta-1\right)\rho_{t-1}\beta\alpha\pi_t^{\theta-1}X_t + \varphi_{t-1}\beta\alpha\theta\phi\pi_t^{\theta\phi}Z_t + \xi_t X_t\left(\theta\phi-\theta+1\right)\left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi}{1-\theta}}\frac{\alpha}{1-a}\pi_t^{\theta-1} - \eta_t\alpha\theta\phi\Delta_{t-1}\pi_t^{\theta\phi} + \eta_t\theta\phi\alpha\pi_t^{\theta-1}\left(\frac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi-\theta+1}{\theta-1}}. \tag{70}$$

Optimal Steady State

The value of the endogenous variables in steady state should solve the system of constraints (42)–(45) and the first-order conditions, (70). As a result, one obtains the following steady-state equations:

$$\begin{aligned}
X &= \tfrac{1}{1-\beta\alpha\pi^{\theta-1}}; & \left[\Phi\Delta - \varphi\tfrac{\theta-1}{\theta}\right]\tfrac{\theta}{\theta-1}c &= 1;\\
Z &= \tfrac{\theta-1}{\theta}X\left(\tfrac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi-\theta+1}{1-\theta}}; & \eta\Delta\left(1-\alpha\pi^{\theta\phi}\right) &= \left(v + \tfrac{\theta}{\theta-1}\Phi\Delta c\right);\\
c &= \left(1-\alpha\beta\pi^{\theta\phi}\right)Z; & \xi &= -\varphi\tfrac{\theta-1}{\theta}\left(1-\alpha\beta\pi^{\theta\phi}\right);\\
\Delta &= \left(\tfrac{1-\alpha}{1-\alpha\pi^{\theta\phi}}\right)\left(\tfrac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi}{\theta-1}}; & \rho &= -\xi\left(\tfrac{1-\alpha\pi_t^{\theta-1}}{1-\alpha}\right)^{\frac{\theta\phi-\theta+1}{1-\theta}}X.
\end{aligned} \tag{71}$$

For further convenience, we will compute the steady-state value of the Lagrange multipliers

$$\begin{aligned}
\rho &= \Phi h - 1 < 0;\\
\eta\Delta\left(1-\alpha\pi^{\theta\phi}\right) &= v + \Phi h,
\end{aligned}$$

where we defined $h := \frac{1-\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta\phi}}\frac{1-\alpha\beta\pi^{\theta\phi}}{1-\beta\alpha\pi^{\theta-1}}$. The last first-order condition (70) provides an equation to find optimal inflation,

$$\begin{aligned}
0 = {} & [1-\Phi h]\,\phi\theta\left(\frac{\beta\alpha\pi^{\theta\phi}}{(1-\alpha\beta\pi^{\theta\phi})} - \frac{\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta-1}}\right)\\
& + [1-\Phi h](\theta-1)\left(\frac{\alpha\pi_t^{\theta-1}}{1-\alpha\pi_t^{\theta-1}} - \frac{\beta\alpha\pi^{\theta-1}}{1-\beta\alpha\pi^{\theta-1}}\right)\\
& + [v+\Phi h]\,\theta\phi\left(\frac{\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta-1}} - \frac{\alpha\pi^{\theta\phi}}{(1-\alpha\pi^{\theta\phi})}\right).
\end{aligned} \tag{72}$$

Using these equations, one can infer certain properties of the optimal steady-state inflation rate.

Proof of Proposition 4: Existence

One may rewrite (72) as (73):

$$F(\pi) = vg(\pi) + [g(\pi) - f(\pi)] + \Phi h(\pi)f(\pi) = 0, \tag{73}$$

where $h(\pi) = \frac{1-\alpha\beta\pi^{\theta\phi}}{1-\alpha\beta\pi^{\theta-1}} \frac{1-\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta\phi}} > 0$, and $g(\pi) = \frac{\theta\phi}{1-\alpha\pi^{\theta\phi}} - \frac{\theta\phi}{1-\alpha\pi^{\theta-1}}$; $f(\pi) = \left[\frac{\theta\phi}{1-\alpha\pi^{\theta\phi}} - \frac{\theta\phi}{1-\alpha\beta\pi^{\theta\phi}}\right] - \left[\frac{\theta-1}{1-\alpha\pi^{\theta-1}} - \frac{\theta-1}{1-\beta\alpha\pi^{\theta-1}}\right]$; $\Phi = \frac{\theta-1}{\theta}\frac{1-\tau}{\mu}$.

It is easy to see that $g(1) = 0$; $h(1) = 1$ and $f(1) = \frac{(\theta\phi-\theta+1)\alpha(1-\beta)}{(1-a)(1-\alpha\beta)} > 0$, which implies that $F(1) = -(1-\Phi)\frac{(\theta\phi-\theta+1)\alpha(1-\beta)}{(1-a)(1-\alpha\beta)} \leq 0$. The strict equality obtains in three cases only: first, when prices are flexible, $\alpha = 0$; second, when the future is not discounted by firms, $\beta = 1$; and finally, when there are no distortions in steady state, $\Phi = 1$.

Define $\pi_h = \alpha^{-1/(\phi\theta)}$ and note that the functions g, f, and h are defined on an interval $[1, \pi_h)$. The difference $[g(\pi_h) - f(\pi_h)]$ is bounded while $g(\pi)$, $h(\pi)$, and $f(\pi)$ tend to positive infinity as π approaches π_h. Hence, $\lim_{\pi \longrightarrow \pi_h} F(\pi) = +\infty$. Since $F(\pi)$ is a continuous function, one can conclude that there is a solution to (73) on the interval $[1, \alpha^{-1/(\phi\theta)})$. One may easily show then that if $\pi_m = \beta^{1/(\theta-1-\phi\theta)}$, then it follows that $F(\pi_m) > 0$, since $g(\pi_m) - f(\pi_m) > 0$. Therefore, optimal inflation is smaller than π_m.

Proof of Proposition 4: Uniqueness

The proof is by contradiction. First it is proved that if $\beta < 1$, for any $\pi_1 < \pi_m$ such that $F(\pi_1) = 0$, it is necessary that $F'(\pi_1) > 0$. By direct differentiation, it follows that

$$F'(\pi_1) = (v+1)\, g'(\pi_1) + (\Phi h(\pi_1) - 1)\, f'(\pi_1) + \Phi h'(\pi_1) f(\pi_1).$$

Moreover, since $F(\pi_1) = 0$, it follows that $\Phi h(\pi_1) - 1 = -(v+1)\, g(\pi_1)/f(\pi_1)$. Therefore,

$$F'(\pi_1) = \frac{(v+1)}{f(\pi_1)}\left[g'(\pi_1)f(\pi_1) - f'(\pi_1)g(\pi_1)\right] + \Phi h'(\pi_1) f(\pi_1),$$

and it is easy to show that for any $\pi_1 < \pi_m$, $g'(\pi_1)f(\pi_1) - f'(\pi_1)g(\pi_1) > 0$, and therefore $F'(\pi_1)$ is positive.

Since F is continuously differentiable, if a solution of (73) is not unique, there will be at least one solution such that $F'(\pi_1) \leq 0$. It has been demonstrated that such a solution is impossible and the necessary contradiction is obtained.

Proof of Proposition 5

By the implicit function theorem, one concludes that $\frac{d\pi}{d\Phi} = -\frac{\partial F}{\partial \pi} / \frac{\partial F}{\partial \Phi}$. From the proof of proposition 4 (uniqueness), we know that $\frac{\partial F}{\partial \pi} > 0$, while $\frac{\partial F}{\partial \Phi} = h(\pi) f(\pi) > 0$. Therefore $\frac{d\pi}{d\Phi} < 0$, and equilibrium inflation increases with steady-state distortions, measured as $1 - \Phi$.

Similarly, $\frac{d\pi}{dv} = -\frac{\partial F}{\partial \pi} / \frac{\partial F}{\partial v}$, where $\frac{\partial F}{\partial v} = g(\pi) > 0$ for $\pi > 1$, therefore $\frac{d\pi}{dv} < 0$, and optimal inflation declines with the elasticity of labor.

Moreover, $\frac{d\pi}{d\beta} = -\frac{\partial F}{\partial \pi} / \frac{\partial F}{\partial \beta}$, where $\frac{\partial F}{\partial \beta} = -(1 - \Phi h(\pi)) \frac{\partial f}{\partial \beta} + \Phi f(\pi) \frac{\partial h}{\partial \beta}$, and one may prove by direct differentiation that $\frac{\partial f}{\partial \beta} < 0$, $\frac{\partial \ln h}{\partial \beta} > 0$, and $(1 - \Phi h(\pi)) = (v + 1)\, g(\pi)/f(\pi) > 0$. Therefore, $\frac{\partial F}{\partial \beta} > 0$, and $\frac{d\pi}{d\beta} < 0$.

Finally, it is worth noting that steady-state inflation can both increase or decrease in price stickiness, since the sign of $\frac{\partial F}{\partial \alpha}$ may be positive or negative.

Appendix 2. The Second-Order Approximation to Unconditional Welfare

In section 6.2 of the main text, we asserted the existence of the following quadratic equation:

$$EU = E\left(Q_l + \rho Q_X + \varphi Q_Z + \xi Q_{ZX} + \eta Q_{\Delta}\right),$$

where Q_l is the second-order term of the loss function and Q_X, Q_Z, Q_{ZX}, Q_{Δ}, and Q_p are the second-order terms of the log-linear approximation to constraints in the above Hamiltonian. We outline here the key manipulations required to derive these expressions. A more detailed appendix is available upon request. For optimal π, they can be written as

$$Q_l = -\frac{1}{2}\widehat{c}_t^2 + \frac{1}{2} v \widehat{\Delta}_t^2 - \frac{1}{2}\frac{\theta}{\theta - 1}\Phi c \Delta \left(2\widehat{c}_t \widehat{\Delta}_t - 2\widehat{c}_t \widehat{\mu}_t - 2\widehat{\mu}_t \widehat{\Delta}_t\right) + tip; \tag{74}$$

$$Q_x = -(\theta - 1)\,\beta\alpha\pi^{\theta-1} X \left[\widehat{X}_t \widehat{\pi}_t + \frac{1}{2}(\theta - 2)\,\widehat{\pi}_t^2\right]; \tag{75}$$

$$Q_z = \theta\phi\beta\alpha\pi^{\theta\phi} Z \left(\widehat{Z}_t \widehat{\pi}_t + \frac{1}{2} (\theta\phi - 1) \widehat{\pi}_t^2 \right); \tag{76}$$

$$\begin{aligned} Q_{zx} = {} & (\theta\phi - \theta + 1) X \left(\frac{1 - \alpha\pi^{\theta-1}}{1-\alpha} \right)^{\frac{\theta\phi}{1-\theta}} \frac{\alpha\pi^{\theta-1}}{1-\alpha} \widehat{X}_t \widehat{\pi}_t \\ & + \frac{1}{2} X \frac{(\theta\phi - \theta + 1) \alpha\pi^{\theta-1}}{1-\alpha} \left(\frac{1 - \alpha\pi^{\theta-1}}{1-\alpha} \right)^{\frac{\theta\phi}{1-\theta}} \\ & \times \left[\frac{\theta\phi\alpha\pi^{\theta-1}}{1 - \alpha\pi^{\theta-1}} + \theta - 2 \right] \widehat{\pi}_t^2; \end{aligned} \tag{77}$$

$$\begin{aligned} Q_\Delta = {} & -\theta\phi\alpha\Delta\pi^{\theta\phi} \left(\widehat{\Delta}_{t-1} \widehat{\pi}_t + (\theta\phi - 1) \frac{1}{2} \widehat{\pi}_t^2 \right) \theta\phi\alpha\pi^{\theta-1} \left(\frac{1 - \alpha\pi_t^{\theta-1}}{1-\alpha} \right)^{\frac{\theta\phi}{\theta-1}-1} \\ & \times \left(-\frac{(\theta\phi - \theta + 1) \alpha\pi^{\theta-1}}{1 - \alpha\pi_t^{\theta-1}} + \theta - 2 \right) \frac{1}{2} \widehat{\pi}_t^2. \end{aligned} \tag{78}$$

The linear relations are

$$\widehat{X}_t - \beta\alpha\pi^{\theta-1} \left(\widehat{X}_{t+1} + (\theta - 1) \widehat{\pi}_{t+1} \right) = O2; \tag{79}$$

$$\widehat{Z}_t - \left(1 - \alpha\beta\pi^{\theta\phi}\right) \widehat{c}_t - \beta\alpha\pi^{\theta\phi} \left(\widehat{Z}_{t+1} + \theta\phi\widehat{\pi}_{t+1} \right) = O2; \tag{80}$$

$$\widehat{Z}_t - \left(\widehat{X}_t + (\theta\phi - \theta + 1) \frac{\alpha\pi^{\theta-1}}{1 - \alpha\pi_t^{\theta-1}} \widehat{\pi}_t \right) = O2; \tag{81}$$

$$-\widehat{\Delta}_t + \alpha\pi^{\theta\phi} \widehat{\Delta}_{t-1} + \theta\phi \left(\frac{\alpha\pi^{\theta\phi} - \alpha\pi^{\theta-1}}{1 - \alpha\pi^{\theta-1}} \right) \widehat{\pi}_t = O2. \tag{82}$$

Simplification of Q_Δ

Use (82) to derive an expression for $\widehat{\Delta}_t^2$. From that we find an expression for the cross term $\widehat{\pi}_t \widehat{\Delta}_{t-1}$:

$$\begin{aligned} E\alpha\pi^{\theta\phi}\theta\phi\widehat{\pi}_t\widehat{\Delta}_{t-1} = {} & E\frac{1}{2} \left(\frac{\alpha\pi^{\theta\phi} - \alpha\pi^{\theta-1}}{1 - \alpha\pi^{\theta-1}} \right)^{-1} \left(1 - \left(\alpha\pi^{\theta\phi}\right)^2 \right) \widehat{\Delta}_t^2 \\ & - \frac{1}{2} (\theta\phi)^2 \left(\frac{\alpha\pi^{\theta\phi} - \alpha\pi^{\theta-1}}{1 - \alpha\pi^{\theta-1}} \right) \widehat{\pi}_t^2. \end{aligned}$$

Then, substitute that into (78) to conclude that

$$EQ_{\Delta}\eta = -\frac{1}{2}\frac{\left(1-\alpha\pi^{\theta-1}\right)\left(1+\alpha\pi^{\theta\phi}\right)}{\alpha\pi^{\theta\phi}-\alpha\pi^{\theta-1}}\left(v+\Phi h\right)E\widehat{\Delta}_t^2 + \theta\phi\frac{\left(v+\Phi h\right)}{1-\alpha\pi^{\theta-1}} \times\left[\left(\theta-1-\theta\phi\right)\frac{\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta-1}}+\frac{\alpha\pi^{\theta\phi}-\alpha\pi^{\theta-1}}{1-\alpha\pi^{\theta\phi}}\right]\frac{1}{2}E\widehat{\pi}_t^2. \tag{83}$$

Simplification of Q_z

Using (81) in (76) results in

$$\varphi Q_z = \rho\frac{\theta\phi\beta\alpha\pi^{\theta\phi}}{1-\alpha\beta\pi^{\theta\phi}}\left(\widehat{X}_t\widehat{\pi}_t + \left(\theta\phi-\theta+1\right)\frac{\alpha\pi^{\theta-1}}{1-\alpha\pi_t^{\theta-1}}\widehat{\pi}_t^2 + \frac{1}{2}\left(\theta\phi-1\right)\widehat{\pi}_t^2\right). \tag{84}$$

Computing the Final Quadratic Form

We use the steady-state relation among Lagrange multipliers (71) and quadratic forms (74), (75), (77), (83), and (84) to write

$$EU = -E\frac{1}{2}\Lambda_{\Delta}\widehat{\Delta}_t^2 - \frac{1}{2}(1-h\Phi)\widehat{c}_t^2 - \frac{1}{2}h\Phi\left(\widehat{c}_t+\widehat{\Delta}_t-\widehat{\mu}_t\right)^2; - E\left(\Lambda_{x\pi}\widehat{X}_t\widehat{\pi}_t + \frac{1}{2}\Lambda_{\pi\pi}\widehat{\pi}_t{}^2\right). \tag{85}$$

In that expression, parameters are gathered in the Λ-terms. For example, $\Lambda_{\Delta} = \left(\frac{1-\alpha\pi^{\theta-1}\alpha\pi^{\theta\phi}}{\alpha\pi^{\theta\phi}-\alpha\pi^{\theta-1}}\right)(v+\Phi h)$. Finally, we exploit the linear relation (79) to find an expression for the cross term $E\widehat{X}_t\widehat{\pi}_t$: $E\widehat{X}_t^2 = \left(\beta\alpha\pi^{\theta-1}\right)^2 E\left(\widehat{X}_{t+1}+(\theta-1)\widehat{\pi}_{t+1}\right)^2$ and $E\widehat{X}_t\widehat{\pi}_t = -\frac{1}{2}(\theta-1)E\widehat{\pi}_{t+1}^2 + \frac{1}{2}\frac{1-\left(\beta\alpha\pi^{\theta-1}\right)^2}{(\theta-1)\left(\beta\alpha\pi^{\theta-1}\right)^2}E\widehat{X}_t^2$. Hence, one recovers

$$EU = -\frac{1}{2}E\left(h\Phi\left(\widehat{c}_t+\widehat{\Delta}_t-\widehat{\mu}_t\right)^2 + (1-h\Phi)\widehat{c}_t^2 + \Lambda_x\widehat{X}_t^2 + \Lambda_{\pi}\widehat{\pi}_t{}^2 + \Lambda_{\Delta}\widehat{\Delta}_t^2\right). \tag{86}$$

Alternative Presentation

There are alternative ways to compose the quadratic criterion. Recall that c_t is marginal cost defined in (46):

$$\widehat{c}_t = \widehat{\mu}_t + v\widehat{\Delta}_t + (v+1)\,\phi\left(\widehat{Y}_t - \widehat{A}_t\right). \tag{87}$$

The first two terms in the quadratic loss function (86) can be simplified, defining Y_t^*, $\widehat{g}_t$, and G as

$$Y_t^* := \widehat{A}_t - v_\mu \widehat{\mu}_t - v_\Delta \widehat{\Delta}_t;$$
$$G := h\Phi\,(1 - h\Phi)\,; \quad v_\mu := \frac{1 - h\Phi}{(v+1)\,\phi}; \quad v_\Delta := \frac{v + h\Phi}{(v+1)\,\phi};$$
$$\widehat{g}_t : = \widehat{\Delta}_t - \widehat{\mu}_t.$$

Thus (86) becomes

$$EU = -\frac{1}{2}E\left[\phi\,(1+v)\left(\widehat{Y}_t - \widehat{Y}_t^*\right)^2 + G\widehat{g}_t^2 + \Lambda_x \widehat{X}_t^2 + \Lambda_\Delta \widehat{\Delta}_t^2 + \Lambda_\pi \widehat{\pi}_t{}^2\right]. \tag{88}$$

Linear Approximation to UO Policy

This is constructed by approximating the relevant first-order conditions around the steady state defined in section 6.1. Details are available upon request.

Linear Approximation to TP-Optimal Policy

For the most part, these derivations, although a little involved in places, are straightforward. Details are available upon request.

References

Anand, S., and A. K. Sen. 2000. "Human Development and Economic Sustainability." *World Development* 28 (12): 2029–49.

Anderson, E. W., E. R. McGrattan, L. P. Hansen, and T. J. Sargent. 1996. "Mechanics of Forming and Estimating Dynamic Linear Economies." In *Handbook of Computational Economics*, Vol. 1, ed. H. M. Amman, D. A. Kendrick, and J. Rust, 171–252 (chapter 4). Elsevier.

Ascari, G., and T. Ropele. 2007. "Optimal Monetary Policy under Low Trend Inflation." *Journal of Monetary Economics* 54 (8): 2568–83.

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

Benigno, P., and M. Woodford. 2004. "Optimal Monetary and Fiscal Policy: A Linear-Quadratic Approach." In *NBER Macroeconomics Annual 2003,* ed. M. Gertler and K. Rogoff, 271–333. Cambridge, MA: MIT Press.

———. 2005. "Inflation Stabilization and Welfare: The Case of a Distorted Steady State." *Journal of the European Economic Association* 3 (6): 1185–1236.

———. 2006a. "Optimal Inflation Targeting under Alternative Fiscal Regimes." NBER Working Paper No. 12158.

———. 2006b. "Optimal Taxation in an RBC Model: A Linear-Quadratic Approach." *Journal of Economic Dynamics and Control* 30 (9–10): 1445–89.

———. 2012. "Linear-Quadratic Approximation of Optimal Policy Problems." *Journal of Economic Theory* 147 (1): 1–42.

Blake, A. P. 2001. "A Timeless Perspective on Optimality in Forward-Looking Rational Expectations Models." Discussion Paper No. 154, National Institute of Economic and Social Research.

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

Caplin, A., and J. Leahy. 2004. "The Social Discount Rate." *Journal of Political Economy* 112 (6): 1257–68.

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

Damjanovic, T., V. Damjanovic, and C. Nolan. 2008. "Unconditionally Optimal Monetary Policy." *Journal of Monetary Economics* 55 (3): 491–500.

Damjanovic, T., and C. Nolan. 2010. "Relative Price Distortions and Inflation Persistence." *Economic Journal* 120 (547): 1080–99.

Debortoli, D., and R. Nunes. 2006. "On Linear Quadratic Approximations." MPRA Paper No. 544, University Library of Munich (revised July 2006).

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.

Harrod, R. 1948. *Towards a Dynamic Economics.* London: Macmillan and Company.

Holzmann, R., and R. Hinz. 2005. *Old-Age Income Support in the 21st Century: An International Perspective on Pension Systems and Reform.* Washington, D.C.: The World Bank.

Horvath, M. 2011. "Alternative Perspectives on Optimal Public Debt Adjustment." *B.E. Journal of Macroeconomics* 11 (1): 34.

Jensen, C., and B. T. McCallum. 2002. "The Non-Optimality of Proposed Monetary Policy Rules under Timeless Perspective Commitment." *Economics Letters* 77 (2): 163–8.

———. 2010. "Optimal Continuation versus the Timeless Perspective in Monetary Policy." *Journal of Money, Credit and Banking* 42 (6): 1093–1107.

Judd, K. L. 1999. *Numerical Methods in Economics.* MIT Press.

Khan, A., R. G. King, and A. L. Wolman. 2003. "Optimal Monetary Policy." *Review of Economic Studies* 70 (4): 825–60.

Kim, J., and D. Henderson. 2005. "Inflation Targeting and Nominal-Income-Growth Targeting: When and Why Are They Suboptimal?" *Journal of Monetary Economics* 52 (8): 1463–95.

Kim, J., and S. H. Kim. 2003. "Spurious Welfare Reversals in International Business Cycle Models." *Journal of International Economics* 60 (2): 471–500.

———. 2007. "Two Pitfalls of Linearization Methods." *Journal of Money, Credit and Banking* 39 (4): 995–1001.

King, R., and A. L. Wolman. 1999. "What Should the Monetary Authority Do When Prices Are Sticky?" In *Monetary Policy Rules*, 349–404. National Bureau of Economic Research, Inc.

Kollmann, R. 2002. "Monetary Policy Rules in the Open Economy: Effect on Welfare and Business Cycles." *Journal of Monetary Economics* 49 (5): 989–1015.

Kydland, F., and E. C. Prescott. 1977. "Rules rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85 (3): 473–90.

Levine, P., J. Pearlman, and R. Pierse. 2008. "Linear-Quadratic Approximation, External Habit and Targeting Rules." *Journal of Economic Dynamics and Control* 32 (10): 3315–49.

McCallum, B. T. 2005. "What Is the Proper Perspective for Monetary Policy Optimality?" *Monetary and Economic Studies* 23 (S1): 13–24.

Pigou, A. C. 1932. *The Economics of Welfare.* 4th ed. London: Macmillan.

Ramsey, F. P. 1928. "A Mathematical Theory of Saving." *Economic Journal* 38 (152): 543–59.

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.

Schmitt-Grohe, S., and M. Uribe. 2004. "Optimal Fiscal and Monetary Policy under Sticky Prices." *Journal of Economic Theory* 114 (2): 198–230.

———. 2007. "Optimal Simple and Implementable Monetary and Fiscal Rules." *Journal of Monetary Economics* 54 (6): 1702–25.

Soderlind, P. 1999. "Solution and Estimation of RE Macromodels with Optimal Policy." *European Economic Review* 43 (4–6): 813–23.

Solow, R. M. 1974. "The Economics of Resources or the Resources of Economics." *American Economic Review* 64 (2): 1–14.

Stern, N. H. 2007. *The Economics of Climate Change: The Stern Review.* Cambridge University Press.

Strotz, R. H. 1956. "Myopia and Inconsistency in Dynamic Utility Maximization." *Review of Economic Studies* 23 (3): 165–80.

Sutherland, A. 2002. "A Simple Second-Order Solution Method for Dynamic General Equilibrium Models." CEPR Discussion Paper No. 3554.

Taylor, J. B. 1979. "Estimation and Control of a Macroeconomic Model with Rational Expectations." *Econometrica* 47 (5): 1267–86.

Whiteman, C. H. 1986. "An Analytical Policy Design under Rational Expectations." *Econometrica* 54 (6): 1387–1405.

Woodford, M. 1999a. "Commentary: How Should Monetary Policy Be Conducted in an Era of Price Stability?" In *New Challenges* *for Monetary Policy*, 277–316. Proceedings of the Federal Reserve Bank of Kansas City Economic Symposium, Jackson Hole, Wyoming, August 26–28.

———. 1999b. "Optimal Monetary Policy Inertia." *The Manchester School* 67 (s1): 1–35.

———. 2002. "Inflation Stabilization and Welfare." *Contributions to Macroeconomics* 2 (1): Article 1.

Responses to the Financial Crisis, Treasury Debt, and the Impact on Short-Term Money Markets*

Warren B. Hrung[a] and Jason S. Seligman[b]
[a]Federal Reserve Bank of New York
[b]John Glenn School of Public Affairs, The Ohio State University

The United States introduced several programs in response to the financial crisis. We examine responses involving Treasury debt—the Term Securities Lending Facility (TSLF), Supplementary Financing Program (SFP), Treasury issuance, open-market operations—and associated impacts on collateralized funding markets. We find the TSLF uniquely effective, due primarily to its introduction during the financial crisis. We find some evidence that the SFP helped alleviate funding market stress. This is notable, as the SFP actually drained bank reserves. Our results show that the proper policy response to a financial crisis can involve options beyond an increase in the level of bank reserves.

JEL Codes: G01, E52, E63.

1. Introduction

This study examines whether and to what degree policy responses involving supplies of Treasury debt alleviated stress in collateralized

*The authors thank James Choi, Michael Kim, and Matthew Wieler for excellent research assistance. We thank Alan Auerbach, Chris Burke, Michael Fleming, Gary Gorton, Mark Jensen, Frank Keane, Debby Perelmuter, Brian Sack, and seminar participants at the 2011 Federal Reserve Day Ahead Conference on Financial Markets & Institutions, 2011 ASSA meetings, 2011 NBER summer institute, the Federal Reserve Bank of New York, the Federal Reserve Bank of Cleveland, the Federal Reserve Bank of Richmond, the Federal Reserve Bank of Dallas, and the U.S. Department of the Treasury for helpful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Author e-mails: warren.hrung@ny.frb.org and seligman.10@osu.edu.

funding markets over the period of the recent financial crisis. The response to a disruption in the traditional banking system typically follows the widely appreciated concept attributed to Walter Bagehot (1873): to avert a panic, central banks should lend early and freely at high rates, to solvent firms, against good collateral.[1] By 2007, short-term collateralized funding was central to the operations for many of the institutions which came under duress in the recent financial crisis. As these funding markets became impaired, a policy response based on a narrow interpretation of Bagehot's dictum would not have effectively relieved money-market stresses. The impairment in collateralized funding markets called for more high-quality collateral such as Treasury securities, not more cash or bank reserves. We focus on this aspect of the financial crisis in this study.

Since the fall of 2007, various programs together involving trillions of dollars have been introduced in the United States by both fiscal and monetary authorities in response to the financial crisis. In contrast to many other studies, we examine a number of policies in tandem. This paper controls for the full spectrum of concurrent monetary and fiscal policies that had direct impacts on Treasury supplies.

One such program, the Term Securities Lending Facility (TSLF), was introduced in March 2008, as money markets became severely impaired. The TSLF was specifically designed to address dislocations in money markets by exchanging Treasury securities for less-liquid and lower-quality collateral held by market participants. A second program, the Supplementary Financing Program (SFP), introduced in the fall of 2008, was designed to help the Federal Reserve manage bank reserves through the issuance of special Treasury debt, with proceeds held at the Federal Reserve Bank of New York (FRBNY). To the best of our knowledge, this study is the first to examine the SFP.

A third, *fiscal*, policy change occurred as other Treasury debt issuance increased from late 2007 onward, as a result of increased expenditures and lower tax receipts. Other Treasury debt issuance was also tied to the financial crisis through programs such as the Troubled Asset Relief Program (TARP) and Treasury's agency mortgage-backed security (MBS) purchase program. Finally,

[1] As summarized by Tucker (2009).

open-market operations (OMOs)—both temporary and permanent—which increase or decrease holdings of Treasury debt in the Federal Reserve's System Open Market Account (SOMA), also impacted market supplies of Treasury collateral. Over the course of the financial crisis, the Federal Reserve first sold Treasury holdings to maintain the size of its balance sheet to better manage the federal funds rate, and then later bought Treasury securities as part of its Large-Scale Asset Purchase (LSAP) program. While the SFP, OMOs, and programs such as TARP were not aimed directly at dislocations in short-term money markets, they did change publicly available supplies of Treasury securities.

In general, collateralized rates, such as the overnight Treasury general collateral (GC) repurchase (repo) rate, should be similar to but lower than uncollateralized rates, such as the federal funds rate. Repo rates can be of various terms and are backed by various types of collateral. We focus on the overnight Treasury GC repo rate because it represents the rate on the shortest term for the safest and most liquid type of collateral. As such, this rate is a benchmark for other repo rates. Bartolini et al. (2011) document that Treasury GC collateral forms the highest asset class in repo markets.

Over the course of the recent financial crisis, as the quality of non-Treasury collateral came into question, demand for Treasury collateral increased. An outward shift in demand for high-quality Treasury collateral caused Treasury GC rates to plunge relative to other collateralized and uncollateralized rates. While low interest rates are generally considered to be desirable, the low relative Treasury GC rates observed during the crisis were evidence of impaired market functioning for collateralized funds.

The most basic hypothesis that we examine involves the impact of changes in the supply of Treasury collateral on the GC rate.[2] To the extent that market functioning improves as the supply of Treasury collateral increases, a simple supply-demand framework predicts that overnight Treasury repo rates should increase, and the spread to other rates should narrow. Our results confirm this basic intuition. Furthermore, each program we study had different transmission channels, different initiation periods, and different

[2]More formally, our null hypothesis is that changes in the supply of Treasury collateral have no impact on the GC rate.

patterns of changes in supply, allowing us to identify each program's relative impact on the overnight Treasury GC repo market. We therefore refine our analysis to examine the changes in Treasury collateral due to the various programs noted above.[3] The TSLF is found to have the largest impact, while other sources of Treasury collateral yield smaller impacts. Our results further suggest that the TSLF would have a much smaller effect on repo rates during a period of normal market functioning.[4] As well, and of further note, the wind-down of the TSLF does not appear to have resulted in market disruptions. When we allow for non-constant volatility, we also find evidence that the SFP helped alleviate funding market stress. Additional results are provided regarding the impact of Treasury supply on other collateralized funding markets. In particular, our results are consistent with the TSLF helping to normalize the asset-backed commercial paper (ABCP) market.

Our results show that the proper policy response to a financial crisis may include options which do not increase bank reserves. The TSLF program was reserve neutral, while increases in Treasury collateral via the SFP actually drained reserves from the banking system.

The remainder of this paper is structured as follows: section 2 provides background on secured funding markets and policy responses to the financial crisis that involved Treasury debt, highlighting relevant literature; section 3 describes data and methods; section 4 presents regression results; and section 5 concludes.

2. Background

Secured funding markets allow for collateralized borrowing by participants. In these markets, the most common type of transaction is a repurchase agreement, or repo. In a repo, a sale of securities is combined with an agreement to repurchase the same securities at a later date, typically at a higher price. The higher price represents an

[3]Here, our null hypothesis is that Treasury supply impacts on the GC rate are the same across all the programs examined.

[4]This result rejects a null hypothesis of no difference in the relative effectiveness of policy channels for increasing Treasury supplies in normal versus crisis periods.

interest rate paid to the lender of the cash (buyer of the security), from the borrower of the cash (lender of the security). The lender of funds takes possession of the borrower's securities over the term of the loan and can resell them in the event of a borrower default. The borrower retains the spread between the interest rate on the security and the interest rate paid to the lender of cash. Gorton and Metrick (2012) give an excellent description of the repo securitization process. Also, the amount of cash borrowed is a percent of the value of the collateral pledged. Therefore, tightening collateral requirements can cause rapid contractions in repo market activity. In fact, this type of contraction occurred in the recent financial crisis, as shown in Adrian and Shin (2010).

Repo markets display segmentation, as some contracts specify particular collateral to be used while others are "general"; for a general collateral (GC) repo, any given security within an asset category is acceptable as collateral by the lender. For example, a *Treasury GC repo* contains any Treasury security as collateral.[5] Overnight GC repo rates tend to track rates on uncollateralized overnight federal funds loans; the spread between the overnight GC repo rate and the federal funds target rate is typically less than 10 basis points (bps). This reflects the use of GC repos as a mechanism for lending and borrowing money. In recent years, primary dealers have used repos to finance \$2–5 trillion in fixed-income securities.[6]

As a general rule, there should be a positive relationship between the supply of collateral and the interest rate that the borrower must pay to obtain funds (this is because scarce collateral is more valuable, and thus the borrower need pay less to borrow funds).[7] In fact, a body of literature on specialness and segmentation has evolved along with the repo market itself, defined both narrowly (as with Duffie 1996; Fleming and Garbade 2004, 2007; and Jordan and Jordan 1997) and broadly to generic bond market demand and supply (as seen in Greenwood and Vayanos 2014). Moreover, demand for particular bonds as collateral is a function of their liquidity, such

[5]For a special collateral repo, the lender of funds seeks a specific security—identified by its particular CUSIP number.

[6]See http://www.newyorkfed.org/markets/primarydealers.html for information on primary dealer financing.

[7]See Fleming, Hrung, and Keane (2009, 2010b) for more details regarding secured financing markets.

that on-the-run issues (the latest issues) hold premium collateral status, as documented in Keane (1996) and Longstaff (2004), and is also related to demand to hedge interest rate risk as found in Graveline and McBrady (2011).

Our study examines both monetary and fiscal policy responses involving Treasury debt supplies simultaneously, so that we examine a number of policies instead of the impact of just a single policy response. Our results also highlight the need to carefully consider the interaction between various policies.[8]

2.1 The Term Securities Lending Facility (TSLF)

The TSLF was introduced on March 11, 2008 "to promote liquidity in the financing markets for Treasury and other collateral and thus to foster the functioning of financial markets more generally."[9] As the financial crisis progressed, funding markets came under unprecedented stress; liquidity and counterparty concerns led many money-market participants to seek out the safety of Treasury securities, and term funding became scarce. As a result, Treasury overnight GC was in high demand, causing its rates to plunge, and the spread between the federal funds target rate and Treasury GC repo rates widened to extraordinary levels, as seen in figure 1.[10] Other repo spreads such as the rate spread between agency and Treasury GC collateral also widened markedly over this period.

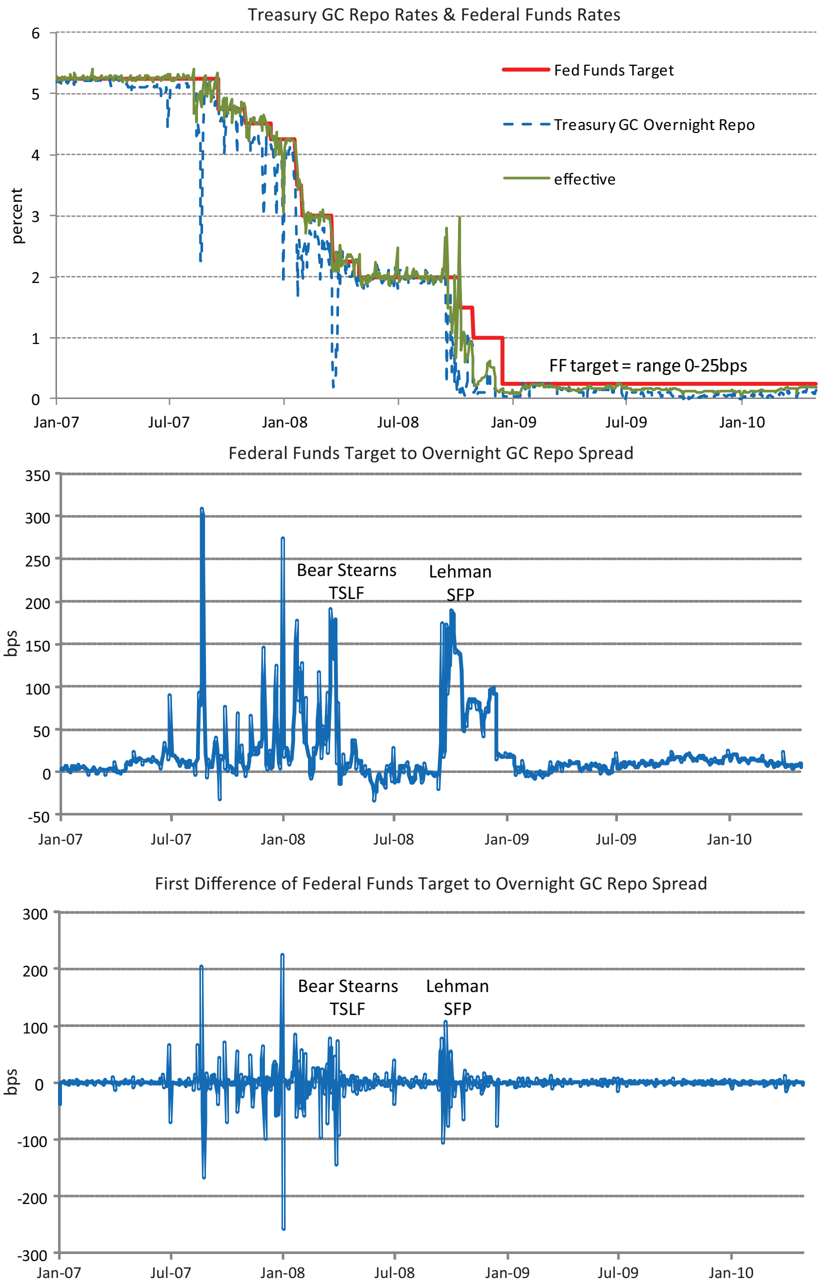
The TSLF addressed widening spreads by increasing the supply of Treasury collateral, which would be expected to increase Treasury GC rates and decrease repo rate spreads. Primary dealers with a trading relationship with the FRBNY were eligible to swap their holdings of less-liquid collateral for Treasury securities held in the System Open Market Account (SOMA) for around twenty-eight

[8]Brunetti, di Filippo, and Harris (2011) considered European Central Bank (ECB) interventions over the financial crisis and found that those failing to target counterparty risk also failed to improve liquidity.

[9]See the Federal Reserve press release announcing the TSLF, at http://www.federalreserve.gov/newsevents/press/monetary/20080311a.htm.

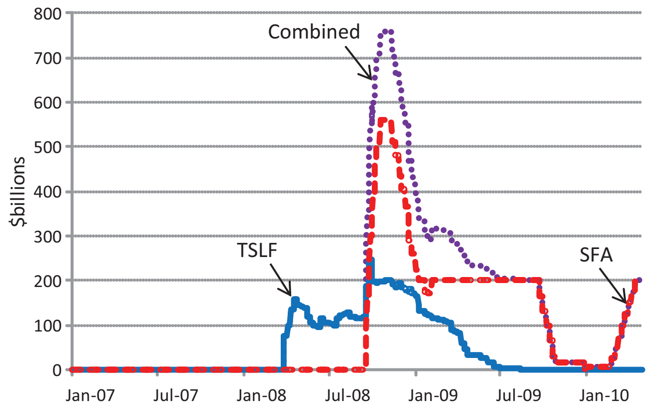
[10]The Treasury GC rates presented in figure 1 are the publicly available rates from Bloomberg. Longstaff (2004) documents pre-crisis flight-to-liquidity premiums somewhat in line with the time t/time s transmission mechanism suggested by Krishnamurthy (2010), though whether these were priced correctly at market circa 2002–7 is debatable—especially in light of the TSLF as a policy innovation.

Figure 1. Daily Repo-Federal Funds Rates and Spread: 2007–10



Source: Federal Reserve Bank of New York and Bloomberg.

Figure 2. The Term Securities Lending Facility and Supplementary Financing Account (Daily)



Source: Federal Reserve Bank of New York and Daily Treasury Statement.

days.[11] The dealers bid a fee via a single-price auction to access the TSLF, with a minimum fee set by the FRBNY.[12]

The TSLF was specifically designed to directly address money-market stresses as it increased the supply of high-quality Treasury collateral and removed less-liquid collateral from the marketplace.[13] Also worth noting, the program's policy design is uniquely elegant, as it involves a *security-for-security* exchange and so does not expand the Federal Reserve's sheet; it is therefore reserve-neutral. Thus there was no need to sterilize the impact of the TSLF and, as a result, the program was able to grow to a substantial size very quickly. This is documented in figure 2.

By mid-April—one month after the first TSLF auction—the facility reached $150 billion in size.[14] The facility briefly peaked

[11]Term lengths ranged from fourteen to thirty-five days, with most ranging between twenty-seven and twenty-nine days.

[12]For more on the TSLF, see Fleming, Hrung, and Keane (2009).

[13]The Federal Reserve also conducted twenty-eight-day single-tranche open-market operations with primary dealers which involved agency MBS collateral. These operations were also targeted at stresses in money markets. We do not examine this program, as it did not involve Treasury collateral.

[14]Note that the maximum amount of Treasury collateral that can be supplied via TSLF is limited to Treasury holdings in the SOMA account. In early March

at almost $250 billion in the fall of 2008 and wound down to zero by early August 2009 as rate spreads in the market contracted and rendered the facility uncompetitive (i.e., expensive relative to market alternatives). The TSLF officially expired on February 1, 2010.[15]

2.2 Supplementary Financing Program (SFP)

While the TSLF impacted supplies of Treasury collateral and less-liquid collateral in the marketplace, the other programs we examine only impacted Treasury collateral. Figure 2 also documents SFP balances over the policy period from 2008 to 2010. The U.S. Treasury announced the SFP on September 17, 2008, two days after the collapse of Lehman Brothers. In just over one month's time, the SFP reached its peak scale of $560 billion. The program was initiated to help the Federal Reserve drain bank reserves accumulating through liquidity facilities introduced during the crisis.[16] Because the level of bank reserves tends to impact the effective federal funds rate, such an offset to the increase in reserves was needed to help the Open Market Trading Desk meet the target federal funds rate set by the Federal Open Market Committee (FOMC).

The program consisted of the issuance of a series of Treasury bills, which were separate and distinct from regular Treasury debt issuance. SFP bills are essentially cash management bills (CMBs). But whereas pricing of CMBs has tended to be punitive as they potentially drain liquid reserves from primary dealers as documented in Seligman (2006) and Simon (1991), SFP proceeds are less likely to be disruptive because bank reserves accumulating at the time were in excess of what would normally have been productive. Further,

2008, the Federal Reserve held around $700 billion in Treasury securities. By the end of April 2008, the Federal Reserve held around $550 billion. We describe the evolution of the SOMA account over our sample period in greater detail below.

[15] The amounts presented and studied include amounts exercised in the TSLF Options Program. For more information on this program, see http://www.federalreserve.gov/newsevents/press/monetary/20080730a.htm.

[16] See http://www.treasury.gov/press-center/press-releases/Pages/hp1144.aspx and http://www.newyorkfed.org/markets/statement_091708.html. The liquidity facilities include the Term Auction Facility and swap line agreements to ease dollar funding stress. A full set of Federal Reserve policies over the financial crisis can be found at http://www.newyorkfed.org/research/global_economy/Crisis_Timeline.pdf.

an incidental by-product of the program was that it increased the amount of high-quality collateral available in the market, helping to alleviate the very same supply-side stresses in money markets that the TSLF was designed to address.

Another way in which SFP transactions differ from CMBs is in the utilization of funds from issuance and, therefore, the impact on the composition of Federal Reserve liabilities. CMB proceeds, like regular Treasury issuance and certain classes of tax payments, are deposited in Treasury's General Account (TGA) at the FRBNY, the account that pays most federal outlays; the TGA can be thought of as Treasury's "checking account." In contrast, SFP proceeds are held in the Supplementary Financing Account (SFA), an account that does not accept tax receipts or pay outlays. Since the TGA and SFA are both liability items on the Federal Reserve's balance sheet, increases in either account will drain reserves from the banking system.[17]

Comparing the TSLF and SFP in terms of sheer magnitude, note that the peak amount of Treasury collateral supplied by the SFP was more than double the peak amount supplied by the TSLF ($560 billion versus $223 billion). But while the SFP is a very effective method for quickly draining bank reserves, one drawback to the SFP as a policy instrument is that SFP bills count against the federal debt ceiling; as such, balances were soon reduced.[18] Figure 2 also displays the combined impact of both programs over the period of observation; at their peak in October 2008, the combined magnitude of the two programs exceeded $750 billion.

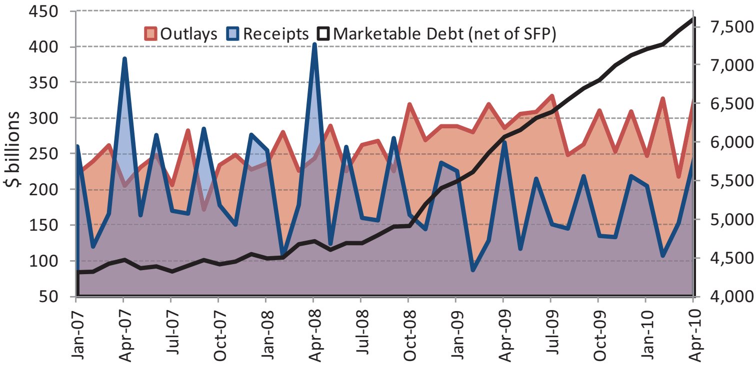
2.3 Treasury Issuance

As in previous recessions, federal tax revenue declines contributed to countercyclical fiscal policy. U.S. federal tax receipts began to

[17]Amounts held in the TGA and SFA can be found on the Daily Treasury Statement and on the Federal Reserve's weekly H4.1 release.

[18]The SFA decreased to $200 billion by early February 2009 and remained at that level into the third quarter of 2009. In mid-September 2009, again driven by concerns related to the debt ceiling, the Treasury announced a further decrease in the SFP balance to $15 billion by the fourth quarter of 2009. The SFA briefly had a zero balance, but after the federal debt ceiling was increased in February 2010, the SFA again increased to $200 billion by mid-April 2010 and remained at that level through the end of our sample period.

Figure 3. Monthly Federal Receipts and Outlays and Daily Marketable Treasury Debt



Source: Daily and Monthly Treasury Statements.
Notes: Outlays and receipts are scaled on the left vertical axis while marketable debt is scaled on the right vertical axis.

fall in late 2007. This required increased debt issuance to cover budgetary shortfalls. In addition, federal outlays increased, widening the budget gap and necessitating a further increase in debt issuance. Beyond both of these traditional "automatic stabilizer" channels, increased outlays due to programs directly related to the financial crisis, such as the Troubled Asset Relief Program (TARP) and Treasury's agency mortgage-backed security (MBS) purchase program, enhanced federal funding requirements. TARP expended around $380 billion and Treasury's agency MBS purchase program purchased a total of $221 billion from September 2008 through December 2009.[19] Figure 3 presents monthly federal receipts and outlays, as well as the daily quantity of marketable outstanding Treasury obligations (net of SFP) from January 2007 to April 2010.

The U.S. Treasury responded to funding needs by increasing the number of different types of securities, as well as increasing the frequency of auctions. Table 1 compares 2009 and 2006 auction policy, documenting the addition of a fifty-two-week bill and a seven-year

[19] Information on TARP and Treasury's agency MBS purchase program can be found at http://www.financialstability.gov.

Table 1. Treasury Issuance: 2006 vs. 2009

Marketable U.S. Treasury Securities				
	2006		**2009**	
Type	**Maturities**	**Schedule**	**Maturities**	**Schedule**
Bills				
	Cash Mgmt. Bills	As Needed	Cash Mgmt. Bills	As Needed
	4 Weeks	Weekly	4 Weeks	Weekly
	13 Weeks	Weekly	13 Weeks	Weekly
	26 Weeks	Weekly	26 Weeks	Weekly
			52 Weeks	Every 4 Weeks
Notes				
	2 Years	Monthly	2 Years	Monthly
	3 Years	Quarterly	3 Years	Monthly
	5 Years	Monthly	5 Years	Monthly
			7 Years	Monthly
	10 Years	8 Times a Year	10 Years	Monthly
Bonds				
	30 Years	2 Times a Year	30 Years	Monthly
Inflation Indexed				
	5-Year Notes	2 Times a Year	5-Year Notes	2 Times a Year
	10-Year Notes	4 Times a Year	10-Year Notes	4 Times a Year
	20-Year Bonds	2 Times a Year	20-Year Bonds	2 Times a Year

Source: U.S. Department of the Treasury.

note. Auction frequencies increased for the three-year, ten-year, and thirty-year issues.

Further, as highlighted in figure 3, the level of outstanding marketable Treasury debt (excluding SFP) increased substantially over the course of 2008–9. Note in the figure that there are seasonal fluctuations in the level of outstanding Treasury debt, so that the level does not monotonically increase. For example, April tax season typically results in net paydowns of Treasury debt and a decrease in the level of outstanding Treasury securities.

2.4 Open-Market Operations (OMOs)

In this section we detail temporary and permanent open-market operations over the period of observation, beginning first with temporary operations.

2.4.1 Temporary Open-Market Operations

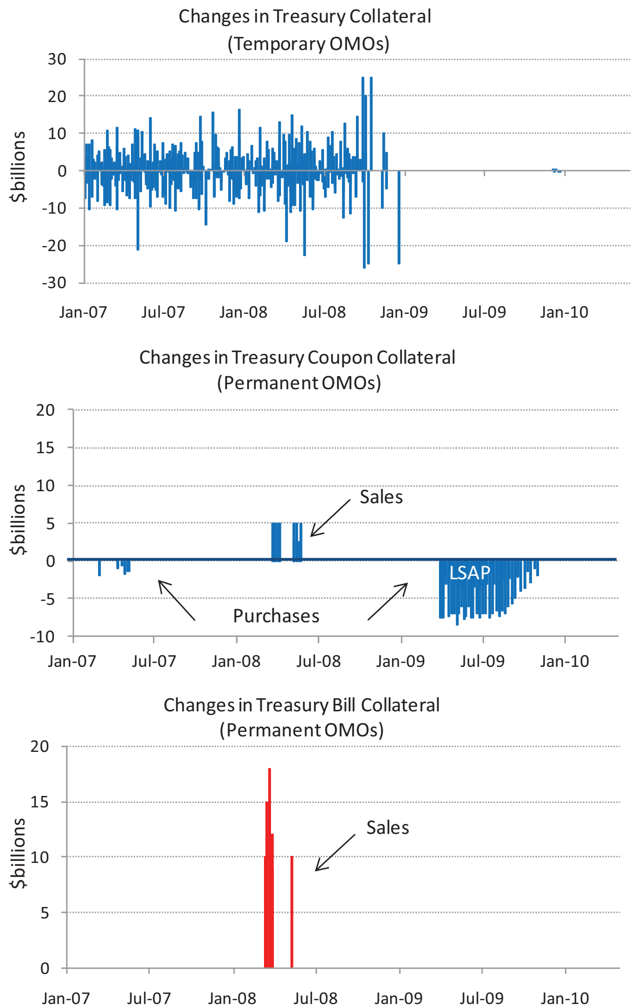
The top panel of figure 4 details the magnitude and frequency of temporary operations impacting Treasury collateral.[20] Temporary OMOs are conducted by the Open Market Trading Desk of the FRBNY to adjust the aggregate supply of bank reserves to foster conditions in the market consistent with the FOMC's policy directive for the federal funds rate. These operations consist of short-term repurchase and reverse repurchase agreements impacting daily trading in the federal funds market. An OMO that drains reserves will add OMO-eligible collateral (Treasury, agency debt, and agency MBS) to the market, and vice versa.[21] Upon maturity of the operation, the movement of collateral is reversed. The term of these operations typically ranges from overnight to twenty-eight (business) days. For more on temporary OMOs, see Carpenter and Demiralp (2006), Friedman and Kuttner (2010), and Hilton and Hrung (2010).

As the top panel of figure 4 highlights, the active daily management of bank reserves via temporary OMOs by the Trading Desk is concentrated prior to and through the initial phases of the crisis. By

[20] Excluded are operations involving agency debt and MBS.

[21] Operations during our sample period that drained reserves only involved Treasury collateral.

Figure 4. Daily Open-Market Operation Impacts on Treasury Collateral, 2007–10



Source (all three graphs): Federal Reserve Bank of New York.

the end of 2008, when the FOMC adopted a target range of 0–25 bps for the federal funds rate instead of an explicit target rate, the Trading Desk stopped conducting temporary OMOs for the remainder of the sample period, aside from some small-scale operations at the end of 2009. More detailed information on the breakdown of Treasury collateral provided for OMOs (e.g., bills vs. notes and bonds) is not publicly available.

2.4.2 Permanent Open-Market Operations

The Federal Reserve's SOMA portfolio traditionally consists primarily of Treasury securities, and these holdings tend to grow over time so as to roughly match growth in currency demand. A permanent OMO to purchase Treasury securities decreases the amount of Treasury collateral available for private parties to utilize in Treasury-secured repo finance. Figure 4 shows that prior to the crisis in the fall of 2007, the Federal Reserve conducted a number of OMOs, of which the permanent OMOs were all confined to be purchases under $5 billion in size.

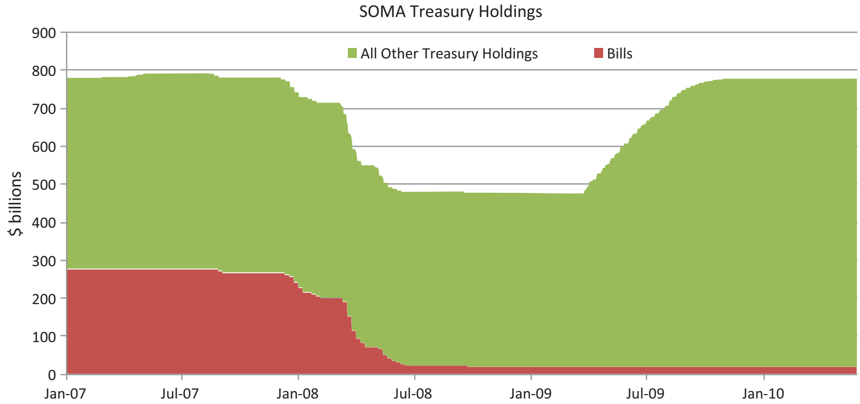
As the crisis intensified, the Federal Reserve's balance sheet began to take on riskier assets as emergency liquidity facilities were introduced. These assets collateralized the funds provided to financial institutions via the liquidity facilities. In an effort to maintain the size of its balance sheet, the Federal Reserve began to allow its Treasury holdings to mature and to sell its holdings. This increased the supply of Treasury collateral available to the public, and decreased bank reserves. As the bottom two panels of figure 4 reveal, the Federal Reserve sold a greater amount of its Treasury-bill holdings than coupon holdings.[22] In the fall of 2008, the Federal Reserve no longer sought to maintain the size of its balance sheet, and Treasury redemptions/sales were discontinued.

Within our observation period, there are only *seven* OMOs involving bill sales and *no* transactions involving bill purchases, so it is difficult to identify the full relationship between repo rates and changes in bills availability due to SOMA transactions. By contrast, the SOMA both purchased and sold meaningful quantities of Treasury coupon holdings over our sample period.

In March 2009, the FOMC announced that it would purchase $300 billion in longer-dated Treasury securities as part of

[22]Coupon holdings refer to holdings of both notes and bonds.

Figure 5. Daily SOMA Treasury Holdings: 2007–10



Source: Federal Reserve Bank of New York.

its Large-Scale Asset Purchase (LSAP) program.[23] The purpose of these purchases was to "help improve conditions in private credit markets," not the repo market.[24] These purchases, which removed Treasury collateral from market, commenced later that month and were completed by the end of October 2009. By the end of the purchases, total SOMA Treasury holdings were similar to their pre-crisis levels, albeit with a different maturity composition weighted more toward coupon holdings. Figure 5 documents the changes in relative composition over this period. As seen in the figure, bills declined from comprising roughly three-eighths of Treasury holdings pre-crisis to a much smaller component.

Table 2 summarizes the above programs, the various sources of Treasury collateral, their primary motivation, and our related hypotheses.

[23]See http://www.federalreserve.gov/newsevents/press/monetary/20090318a.htm for the announcement. The Federal Reserve also purchased $1.25 trillion in agency MBS and around $172 billion in agency debt.

[24]See http://www.newyorkfed.org/markets/funding_archive/lsap.html. Gagnon et. al (2011) examine the impact of LSAPs on domestic interest rates, and Neely (2010) examines their impact on foreign interest rates and exchange rates.

Table 2. Summary of Programs Impacting Treasury Collateral

Program	Primary Purpose	Expected Impact on Federal Funds-Repo Spread (regression coefficient)
TSLF	Alleviate Stresses in Funding Markets	$\beta_{TSLF} < 0$
SFP	Drain Reserves from the Banking System	$\lvert\beta_{TSLF}\rvert \geq \lvert\beta_{SFP}\rvert$
Treasury Issuance	Fund Government Expenditures	$\lvert\beta_{TSLF}\rvert \geq \lvert\beta_{bills}\rvert \geq \lvert\beta_{notes\ and\ bonds}\rvert$
Open-Market Operations		
Temporary	Target Federal Funds Rate	$\lvert\beta_{TSLF}\rvert \geq \lvert\beta_{TOMO}\rvert$
Permanent (includes large-scale asset purchases)	Fund Currency Demand Adjust Federal Reserve Balance Sheet Reduce Long-Term Interest Rates	$\lvert\beta_{TSLF}\rvert \geq \lvert\beta_{POMO\ bills}\rvert \geq \lvert\beta_{POMO\ notes\ and\ bonds}\rvert$

The final column of this table presents our prediction of the respective program's impact on the effective federal funds-GC rate spread relative to the TSLF. These will be discussed further in the next section. Since all of the programs in table 2 impacted the supply of Treasury collateral, the table highlights the need to carefully consider the impact of policies beyond their intended target. For example, consider the SFP and LSAP programs. The SFP was primarily intended to the help drain the level of bank reserves, while LSAP purchases helped lower long-term U.S. interest rates. Regarding collateral impacts, the SFP reinforced collateral injections from the TSLF, while LSAP purchases work to remove Treasury collateral from market.

3. Data and Methods

We analyze daily (business day) data from January 2007 through May 2010. This time frame extends from a period pre-crisis through the period over which the several direct and indirect policies described in the last section manifest: the TSLF and LSAP program, the initiation of the SFP, and the rapid expansion of outstanding publicly held Treasuries from below $5 trillion to close to $8 trillion dollars.

Our dependent variable is the change in the spread between the overnight Treasury GC repo rate and the effective federal funds rate ("the spread," or the "FF-repo spread"). Employing a Dickey-Fuller test over the time period of study, we reject the null of a unit root. Differenced data is employed for economic rather than for econometric reasons, because changes in market actions and policy actions are likely to impact the magnitudes of changes in rates and spreads.[25] Examining spreads, rather than simply the GC repo rates alone accounts for the role the federal funds rate typically serves—as a ceiling for repo rates. (Federal funds transactions are uncollateralized, and collateralized borrowing is typically less expensive.[26]) So the federal funds rate affects repo rates irrespective of the level of relevant collateral.

Data for GC rates come from the Federal Reserve's primary dealer survey, and the effective federal funds rates are available from the Federal Reserve Bank of New York. Overnight GC rates are impacted by the amount of collateral available on a given day, meaning expectations and other potential sources of endogeneity are less of a concern.[27]

[25]Consider the dependent variable as depending on states of the world (as reflected by amounts in the TSLF, for example); then the dependent variable will only change if the state of the world changes. This motivates a specification based on differences instead of levels.

[26]The different markets for the various types of borrowing can result in cases when collateralized borrowing is more expensive than uncollateralized borrowing. For example, money-market mutual funds tend to invest in repo, while banks operate in the federal funds market, changing relative dynamics as the composition of demand for liquidity varies across intermediaries. Over the course of the crisis, as money-market mutual funds faced massive redemptions and pulled away from the repo market, repo rates for all types of collateral rose and even GC rates occasionally traded above the effective federal funds rates.

[27]For example, results of TSLF auctions were released a day or two before settlement, so that amounts settling were known in advance. The overnight GC rate is only impacted by the TSLF settlement, not the announcement of the auction results.

The change in the rate spread is related to changes in Treasury collateral, broken into TSLF, SFP, Treasury bills, and Treasury coupon securities, temporary OMOs, SOMA bills, and SOMA coupon securities.[28] Data on the TSLF, temporary OMOs, SOMA bills, and SOMA coupon securities are available from the Federal Reserve Bank of New York, and data on Treasury debt are available from the Daily Treasury Statement.

While all Treasury securities are eligible to serve as collateral in a Treasury GC repo, the different types of securities could have different impacts on GC rates. For example, comparing TSLF and SFP, the TSLF was targeted at and introduced during a time of great stress in funding markets. As a result, Treasury securities lent out via the TSLF were very likely to have been used as collateral in repo transactions. Similarly, the SFP was initiated in the fall of 2008, when funding markets faced unprecedented stress following the bankruptcy of Lehman Brothers. While SFP was not directed at stresses in funding markets, at its peak it provided more than twice the amount of Treasury collateral as the TSLF. By virtue of magnitude and time of introduction, the SFP may have also impacted FF-repo spreads.

Also worth considering, bills (including SFP bills) may impact funding markets more than notes or bonds. Previous research has shown that primary dealers purchase over 90 percent of CMBs and nearly 85 percent of four-week Treasury bills, while the percentage for longer-term Treasury securities is around 60 percent (Fleming 2007). As dealers tend to hold CMB purchases, it is likely that shorter-maturity securities are more likely to be pledged as collateral in funding markets (Fleming and Rosenberg 2007). Also, some investors, such as money-market mutual funds, need to hold down the weighted-average maturity of their portfolios. Therefore, they typically invest in short-term instruments such as repo or Treasury bills, but not Treasury notes and bonds. As a result, an increase in bills can divert funds away from repo markets and drive up repo

[28]The TSLF auctions alternated in terms of the types of collateral which could be exchanged for Treasury securities. Previous studies (Fleming, Hrung, and Keane 2010a, 2010b) have examined the two types, or "schedules," separately. However, we are concerned only with the amount of Treasury collateral supplied, not the type of collateral withdrawn from the market, so we do not distinguish between the different schedules.

rates. On the other hand, a corresponding increase in notes and bonds will likely not result in a direct diversion of funds from repo markets. And finally, buy-and-hold investors, such as insurance companies and sovereign entities, may prefer longer-duration Treasury securities, and these investors are unlikely to use their Treasury securities as collateral for repo transactions.

As controls, we include data documenting daily changes in measures of stress such as the Chicago Board Options Exchange Volatility Index (VIX), which measures the implied volatility of the S&P 500 index, the Merrill Lynch Global Financial Bond index option-adjusted spread (OAS), the change in the one-month spread between AA financial and non-financial commercial paper (CP), and the change in the one-month LIBOR-OIS (LOIS) spread. The CP rate data comes from the Federal Reserve Board, and the remaining variables were extracted from Bloomberg. We further include calendar dummy variables for the beginning and end of quarters and years—times when demand for collateral may be impacted by reporting requirements.[29]

Table 3 presents summary statistics for the variables studied. Note the wide disparities between the mean values and the minimum and maximum values for the variable levels as well as changes of the variables in the table. This range reflects the extreme distortions in financial markets experienced over our sample period,

[29] The year-end and year-start dummy variables are additive to the quarter-end and quarter-start dummy variables, respectively. Sundaresan and Wang (2009) discuss seasonality in the spread between overnight repo rates and the federal funds target rate. LIBOR stands for the London Interbank Offered Rate, which is a daily reference rate for interbank unsecured borrowing. OIS stands for overnight indexed swap, which is referenced to the daily federal funds rate.

Taylor and Williams (2009) employ a LOIS spread as a dependent variable; however, they express some concern about LIBOR validity due to the self-reported nature of rates by surveyed banks. McAndrews, Sarkar, and Wang (2008) document LIBOR reports in line with expected market reactions. Similarly, Gorton and Metrick (2012) devote a good deal of work to documenting LOIS and several other asset-class spreads and include documentation of exploding haircuts in their descriptive analysis of several dimensions of the 2007–8 period. As compared to our current work, all three papers focus primarily on the early 2007–8 time period, and in the cases of the first two papers, the Term Auction Facility, which was introduced by the Federal Reserve in late 2007. Because of debate regarding the veracity of LIBOR rates, as a robustness check for our results, we have run specifications that simply omit the LOIS variable—results do not fundamentally vary.

Table 3. Summary Statistics

Variables	Mean	Std. Dev.	Min.	Max.
(Effective FF-GC Rate) (bps)	13.230	35.63	−46.7	273.3
OAS (bps)	266.5	169.6	59.0	686.0
VIX (%)	26.5	12.6	9.9	80.9
One-Month AA Financial/ Non-Financial CP (bps)	16.2	26.7	−14.0	236.0
One-Month LIBOR-OIS (bps)	35.9	49.2	3.7	337.8
Δ(Effective FF-GC Rate) (bps)	−0.041	24.53	−174.0	191.3
ΔTSLF ($b)	0.000	4.99	−47.2	75.0
ΔTsy Bills ($b)	0.833	10.54	−55.0	70.0
ΔTsy Notes and Bonds ($b)	3.167	14.54	−54.8	99.0
ΔSFP ($b)	0.234	10.01	−75.0	60.0
ΔSOMA Bills ($b)	0.104	1.18	0.0	18.0
ΔSOMA Notes and Bonds ($b)	−0.297	1.54	−8.5	5.0
ΔShort-Term OMOs ($b)	0.024	4.38	−26.0	25.0
ΔOAS (bps)	0.218	4.58	−37.0	41.0
ΔVIX (percentage points)	0.024	2.56	−17.4	16.5
ΔOne-Month AA Financial/ Non-Financial CP (bps)	0.008	14.69	−106.0	146.0
ΔOne-Month LIBOR-OIS (bps)	0.007	6.20	−44.2	50.4

Notes: Sample: January 2, 2007–May 28, 2010. Observations = 854 (diffs obs. = 853).

during which large aberrations are fairly common. In fact there are eighty-three occurrences (nearly 10 percent of our observations) for which the absolute value of the *change* in spread was greater than 25 bps.[30]

Employing the data represented in table 3, we consider our hypotheses regarding the impacts of generic and program-specific Treasury supply changes for money markets. We estimate the following regression, and the results are presented in table 4:

$$\Delta[FF_{effective} - r_{GC}]_t = \alpha + \beta \Delta USTR_t + \gamma \Delta X_t + \varepsilon_t, \qquad (1)$$

where $FF_{effective}$ represents the effective federal funds rate and r_{GC} represents the collateralized private repo market rate for general U.S. Treasury collateral. The variable(s) $USTR$ take on three variant designs—first as a single variable that combines all sub-types of collateral in keeping with our most generic hypothesis, second as a vector of differentiated policy sources of Treasury collateral. (This disaggregation will allow us to test the null hypothesis that all public policies were essentially equivalent in terms of their impact on the FF-repo spread.) Third and finally, we include the various sources of Treasury collateral interacted with the one-month Treasury GC-agency MBS spread to control for market stress. If the degree of impact varies with increases in market spread, the interaction terms will systematically absorb impacts on spreads when market stress is relatively high, allowing us to distinguish between generic collateral impacts and any additional crisis-period impacts.

Vector X contains the controls listed above. We employ the VIX and the other interest rate spreads as controls due to their associations with funding market stress. We focus on the one-month spreads because term funding became particularly scarce as counterparty and liquidity concerns escalated. We expect that changes in the VIX and the various interest rate spreads will be positively related to the change in the spread. Parameters $\{\alpha, \beta, \gamma\}$ are the subjects of estimation and ε represents an error term.

[30] Of the eighty-three cases, in forty-one we observe changes less than –25 bps, while in the remaining forty-two cases, the observed change was greater than 25 bps.

Table 4. Effective Federal Funds-Treasury General Collateral Repo Rate Spread Analysis

Variables	1		2		3	
	Full Period January 2007–May 2010					
U.S. Treasury Issuance, Total	−0.167***	(0.057)				
Term Securities Lending Facility			−.667**	(0.324)	0.095	(0.064)
Supplemental Financing Program			−0.124	(0.173)	0.014	(0.012)
U.S. Treasury Issuance, Bills			−0.194**	(0.0760)	−0.0171*	(0.00969)
U.S. Treasury Issuance, Notes and Bonds			−0.0903*	(0.0511)	−0.0387***	(0.00656)
Temporary Open-Market Operations			−0.326	(0.303)	0.214***	(0.0431)
System Open Market Account Transactions, Bills			2.625***	(0.763)	−1.565	(8.353)
SOMA Transactions, Notes and Bonds			0.190	(0.304)	0.172***	(0.0532)
TSLF by (GC-MBS)					−0.00701**	(0.00356)
SFP by (GC-MBS)					−0.000660	(0.00166)
T-Bill by (GC-MBS)					0.000425	(0.000736)
T.NB by (GC-MBS)					0.00215*	(0.00129)
STOMO by (GC-MBS)					−0.0105***	(0.00239)
SOMA.Bill by (GC-MBS)					0.0396	(0.0978)
SOMA.NB by (GC-MBS)					−0.0518***	(0.0168)
General Collateral–Mortgage-Backed Security Repo Spread					0.0256***	(0.00846)

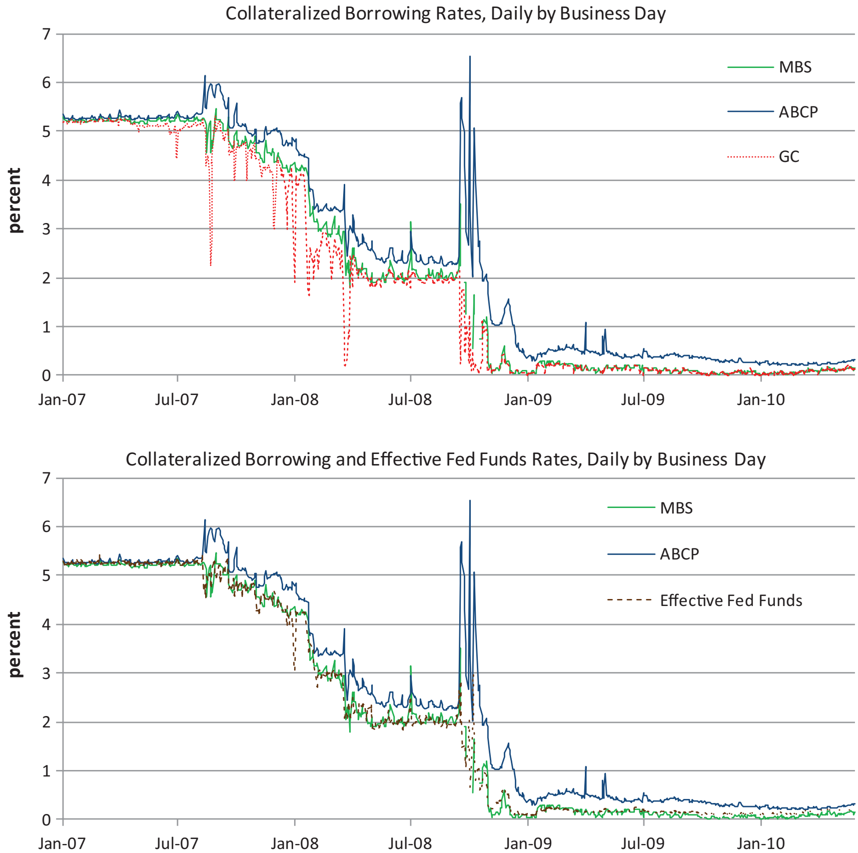
(continued)

Table 4. (Continued)

Variables	1		2		3	
	Full Period January 2007–May 2010					
Global Financial Bond Index Option-Adjusted Spread	0.776**	(0.346)	0.774**	(0.387)	0.00398	(0.0274)
Options Exchange Volatility Index	−0.01	(0.312)	−0.0737	(0.307)	0.0824**	(0.0395)
AA (Financial/Non-Financial) Commercial Paper	0.0546	(0.063)	0.0397	(0.0685)	−0.00123	(0.0154)
London Interbank Offered Rate–Overnight Index Swap	0.495*	(0.252)	0.560**	(0.248)	0.464***	(0.0211)
Lagged Federal Funds GC Repo Spread	0.227***	(0.054)	−0.234***	(0.0492)	−0.318***	(0.0179)
Quarter End	52.95***	(13.230)	50.24***	(12.70)	9.670***	(0.805)
Quarter Start	−37.36***	(9.905)	−36.94***	(9.990)	−7.441***	(1.360)
Year End	−0.268	(34.680)	−1.240	(36.21)	−8.194	(7.446)
Year Start	−2.595	(24.120)	−3.582	(23.80)	0.901	(1.885)
Constant	3.295***	(0.606)	3.030***	(0.567)	0.725***	(0.108)
ARCH					1.163***	(0.109)
TARCH					0.0916	(0.147)
GARCH					0.429***	(0.0196)
Constant					0.329**	(0.145)
Number of Observations	853		853		853	

Notes: All variables are expressed as first differences unless noted. Newey-West standard errors are in parentheses. ***, **, and * denote $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.

Figure 6. Selected Daily Collateralized Borrowing Rates: 2007–10



Source (top graph): Bloomberg and Federal Reserve Board of Governors.
Source (bottom graph): Bloomberg, Federal Reserve Bank of New York, and Federal Reserve Board of Governors.

3.1 Other Funding Markets

It is possible that the various policies that we investigate were able to alleviate stresses beyond the funding market for Treasury collateral. Figure 6 presents overnight MBS repo, ABCP, and GC rates, as well as the effective federal funds rate. Agency repo rates are not shown, as they exhibit similar dynamics to MBS rates.

Ideally, we would be able to offer strong priors on impacts in other funding markets. However, theoretic predictions in this area can be confounded by countervailing forces. For example, consider agency debt and agency MBS. To the extent that lenders of funds became wary of housing-related collateral, agency and MBS repo rates would have increased.[31] Therefore, if increases in Treasury collateral helped heal funding markets in general, these rates would have decreased and we would expect spreads to GC to narrow more than the FF-GC spread. However, to the extent that agency securities are substitutes for Treasury securities and OMO collateral in general was subject to flight-to-quality flows, agency and MBS repo rates would have fallen along with Treasury repo rates as the crisis deepened, and risen with Treasury repo rates as funding market stresses abated.[32] In terms of the spread of agency rates to GC, the compression in response to increases in Treasury collateral would then be smaller in magnitude than the FF-GC spread. So unfortunately, theory provides little guidance in terms of prediction—almost any result would be consistent with Treasury collateral-related policies having an impact on funding markets for agency securities.

Funding markets for less-liquid collateral will likely provide clearer results. Therefore, we also investigate the overnight AA-rated ABCP rate, as ABCP is not OMO eligible.[33] As shown in figure 6, this rate increased substantially as the financial crisis deepened and

[31]Comparing the MBS repo rates and GC rates in figure 6 (top panel), the dynamics for GC rates differ from MBS repo rates, as the MBS rates do not exhibit the steep drops that characterize GC rate movements starting in the second half of 2007. The dynamics for MBS rates are more similar to those for the effective federal funds rate (bottom panel), but there are still periods of noticeable divergence between the two rates. The same characterization generally applies to the ABCP rates in figure 6, though more pronounced. Spreads between the ABCP and effective federal funds rate after the fall of 2007 are notably wider through the end of 2009.

[32]Krishnamurthy, Nagel, and Orlov (forthcoming, figure 7) suggests that both factors were present over the course of the crisis for agency debt, with agency rates exceeding the federal funds rate in the fall of 2007 and mid-2008, but falling below the federal funds rate after the Lehman bankruptcy, while corporate debt and private-label asset-backed security repo rates spiked substantially after the Lehman bankruptcy.

[33]Commercial paper rates are compiled by the Federal Reserve Board. See Adrian, Kimbrough, and Marchioni (2011) for a discussion of the commercial paper market during the financial crisis.

the commercial paper market came under immense stress. For this rate spread to GC, if increases in Treasury collateral helped stabilize this funding market, the ABCP rate would be expected to fall while the GC rate increases. Therefore, the ABCP-GC spread would narrow to a greater extent than the FF-GC spread. Note that if we utilized the FF-ABCP spread instead, we would anticipate a positive coefficient for Treasury collateral as evidence of an abatement of stresses in the ABCP market, and we focus on the ABCP-GC spread to ease comparability with our other results.

4. Results

4.1 Effective FF-GC Repo Spread

Table 4 presents results over the full sample period from January 2007 through May 2010. Columns 1 and 2 present ordinary least squares (OLS) regression results. The first column combines all sources of Treasury collateral, and in line with our first hypothesis, the observed relationship with our dependent spread variable is negative and statistically significant. Each billion dollars of Treasury supply are associated with a 0.167 bps reduction in the FF-GC repo spread.

The second column breaks out the sources of Treasury collateral into seven categories: TSLF, SFP, Treasury bills, Treasury notes and bonds, temporary OMOs, SOMA bills, and SOMA notes and bonds. We can reject the more refined null hypothesis of equal coefficients between types of collateral at the 95 percent confidence level. The largest estimated reductions in spread are associated with the TSLF program, a program designed to relieve funding market stresses. Every $1 billion increase in Treasury collateral due to TSLF is correlated with a 0.667 bps narrowing of the FF-repo spread. As the TSLF was reserve neutral, our results show that a proper policy response to a financial crisis includes options which do not increase bank reserves.

Treasury issuance was another channel for increasing available supplies of Treasury instruments in repo markets. Bills in particular are robustly (99 percent confidence level) associated with reductions of 0.194 bps per $billion. The strong and positive SOMA bills coefficient is not easily interpreted, prima facia. As noted earlier, SOMA bills sales occur on only seven dates—between March 10 and May 8, 2008—and there are no bills purchases during our sample period (see figure 4, bottom panel). The coefficient is likely a spurious artifact.[34]

The lack of significance for the SFP coefficient may not be altogether surprising given that the SFP also drained bank reserves. Because the SFP placed upward pressure on the effective federal funds rate while increasing the supply of Treasury collateral, the program moved the effective federal funds rate and the GC repo rate in the same direction, and its net impact depends on the relative magnitude of each effect.[35]

As regards other coefficients in column 2, the OAS and LOIS spread coefficients are positive, consistent with a flight to quality. The coefficient for changes in the VIX is small and not statistically significant. The lagged spread coefficient suggests some degree of reversion in these data so that, for example, a widening of the spread on any given day is followed by a somewhat mitigating reduction on the following day, all else equal.[36]

The various sources of Treasury collateral might be expected to behave differently outside of a collateral crisis; column 3 differentiates impacts by the degree of funding market stress. We interact the Treasury supply channels with the level of the one-month Treasury GC-to-agency MBS repo rate spread—a measure of market stress which compares two term-collateralized funding rates. Column 3's specification also incorporates the threshold ARCH (TARCH) model of Glosten, Jagannathan, and Runkle (1993). Threshold ARCH

[34] The results for this variable found in column 3 supports this view. Additionally of note, virtually no temporary OMOs were conducted in the latter half of our period of observation. In regressions not presented here, the temporary OMO variable is dropped as a generic robustness check on our remaining estimated coefficients. No coefficients change in terms of magnitude or statistical significance in any meaningful way. Another robustness check of the specification addresses the concern in Taylor and Williams (2009) regarding LIBOR. As noted in footnote 29, when we simply omit the LOIS variable, results do not fundamentally vary.

[35] Since excess reserves can affect trading in the federal funds market, we include the change in the level of excess reserves as an additional independent variable in the table 4 specifications. The coefficient for this variable was positive but not statistically significant. The other variable coefficients were not notably impacted, so the results are not presented.

[36] In the appendix, table 8 employs public data to replicate column 2 of table 4, presenting results using publicly available repo data. Results are generally consistent with those found in table 4.

procedures accommodate asymmetric variations in our dependent variables' volatility over our sample period (see figure 1, bottom panel). Specifically, we utilize the TARCH model to account for any asymmetric responses to positive and negative innovations in volatility.[37] Our motivation for employing this model rests in an attempt to distinguish whether the impacts of TSLF and other supply responses were more or less specific to the circumstances under which they were implemented, while also taking into account non-constant volatility.

Regarding the timing of TSLF implementation, the TSLF coefficient in column 2 embeds both a crisis and a general collateral impact, whereas the same coefficient in column 3 estimates just a general collateral impact (with the GC-MBS spread set to a de-minimus level), while the TSLF*(GC-MBS spread) coefficient reports a crisis impact.

The TSLF stand-alone coefficient (0.095) is not significant at any standard confidence level. By comparison, the TSLF interaction coefficient (−0.007) is significant at the 95 percent confidence level. The TSLF interaction coefficient suggests a sizable impact of the program during times of funding market stress. To illustrate, a $70 billion TSLF settlement with the GC-MBS spread at 200 bps (the values around the time of the Lehman bankruptcy) implies a reduction in the FF-repo spread of almost 100 bps.

The results in column 3 suggest that the TSLF program had a large impact on the FF-repo spread because it was introduced and operated at a time of stress when spreads were unusually wide. The TSLF would have had a much smaller impact if it had been introduced during times of normal market functioning. Note that this impact is not present for the SFP, which was introduced at the height of the financial crisis and reached a peak size that was double that of the TSLF.

[37]The test statistic for ARCH effects had a value of 118.03, so we can reject the null hypothesis of no ARCH effects at the 99 percent level. Since the dependent variable is composed of two market rates, an asymmetric response may be less likely in this specification if both rates respond in a similar manner. However, in unreported results utilizing publicly available repo rates, the TARCH coefficient is statistically significant at the 95 percent level. Moreover, an asymmetric response may be more likely when we investigate the spread between the federal funds target and the GC rate (table 6), since the target rate will not respond to innovations in volatility. We retain the TARCH specification here for consistency across the analyses.

4.2 Other Funding Markets

In table 5, we investigate the impact of Treasury collateral on other funding rates relative to the GC rate.

The first column in table 5 reports the results from column 2 of table 4 as a reference. Column 2 shows corresponding results for the agency-GC spread and column 3 reports results for the MBS-GC spread.[38] Compared with column 1, the TSLF coefficient in these columns remains negative, but is smaller in magnitude and no longer statistically significant. This result is consistent with agency securities being closer substitutes for Treasury collateral than less-liquid collateral. The results for the other coefficients are generally similar.

Column 4 reports results for the ABCP-GC spread. Compared with column 1, the TSLF coefficient is negative, larger in magnitude, and still statistically significant. This suggests that the TSLF did have an impact beyond funding markets for Treasury collateral by helping lower ABCP rates while increasing GC rates. Some of the coefficients for other sources of Treasury collateral are also larger in magnitude than the corresponding coefficients in column 1. Notably, the coefficient for temporary OMOs is much larger and now statistically significant at the 10 percent level.

4.3 Target FF-GC Repo Spread

Table 6 employs the target federal funds rate in place of table 4's effective rate. While employing the federal funds target better isolates the impact of Treasury collateral, as the target rate is set by the FOMC and is not impacted by the level of bank reserves, the dependent variable is no longer a spread between two market rates. For the sub-period where the federal funds target was the range of 0–25 bps (mid-December 2008 forward), we set the target rate to 25 bps.[39]

As shown in table 6, the results are similar when employing the target rate. One notable difference from the results in table 4 is that the SFP stand-alone coefficient now exhibits the expected negative

[38] Data for these rates also come from the primary dealer survey.

[39] In an alternate specification not reported here for the sake of brevity, we employ a midpoint of 12.5 bps as the target rate in the target-range period from December 16, 2008 to May 28, 2010. Results are essentially equivalent.

Table 5. Other Collateral Spreads

Dependent Variable	(Table 4, Col. 2) **Δ(Effective – GC)**	**Δ(Agency – GC)**	**Δ(MBS – GC)**	**Δ(ABCP – GC)**
Variables	**Full Observation Range of January 2007–May 2010**			
Term Securities Lending Facility (TSLF)	−0.667**	−0.156	−0.281	−0.738**
	(0.32)	(0.12)	(0.24)	(0.36)
Supplemental Financing Program (SFP)	−0.124	−0.130	−0.104	−0.303
	(0.17)	(0.11)	(0.13)	(0.20)
U.S. Treasury Issuance, Bills (T.Bill)	−0.194***	−0.226***	−0.248***	−0.393***
	(0.08)	(0.07)	(0.08)	(0.15)
U.S. Treasury Issuance, Notes and Bonds (T.NB)	−0.09*	−0.092	−0.096	−0.083
	(0.05)	(0.06)	(0.07)	(0.09)
Temporary Open-Market Operations (STOMO)	−0.326	−0.250	−0.288	−0.865*
	(0.30)	(0.20)	(0.22)	(0.53)
System Open Market Account Transactions, Bills (SOMA.Bill)	2.62***	1.92*	2.40***	1.66*
	(0.76)	(1.11)	(0.92)	(1.00)
SOMA Transactions, Notes and Bonds (SOMA.NB)	0.190	0.044	−0.036	−0.025
	(0.30)	(0.28)	(0.31)	(0.54)
Number of Observations	853	853	853	853

Notes: All variables are expressed as first differences unless noted. Newey-West standard errors are in parentheses. ***, **, and * denote $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. Other coefficients are not presented in order to conserve space.

Table 6. Target Federal Funds-Treasury General Collateral Repo Rate Spread Analysis

Variables	1		2		3	
	Full Period January 2007–May 2010					
U.S. Treasury Issuance, Total	-0.157^{***}	(0.048)				
Term Securities Lending Facility			-0.731^{**}	(0.290)	0.109	(0.0677)
Supplemental Financing Program			−0.0936	(0.104)	-0.0838^{***}	(0.0263)
U.S. Treasury Issuance, Bills			-0.148^{**}	(0.0697)	-0.0194^{*}	(0.0111)
U.S. Treasury Issuance, Notes and Bonds			-0.123^{**}	(0.0575)	-0.0555^{***}	(0.00575)
Temporary Open-Market Operations			0.148	(0.144)	0.0293	(0.0285)
System Open Market Account Transactions, Bills			1.600^{**}	(0.679)	−1.453	(23.11)
SOMA Transactions, Notes and Bonds			−0.0661	(0.230)	−0.0172	(0.0674)
TSLF by (GC-MBS)					-0.0106^{**}	(0.00504)
SFP by (GC-MBS)					0.00212	(0.00149)
T-Bill by (GC-MBS)					-0.00658^{***}	(0.000760)
T.NB by (GC-MBS)					0.000955	(0.000740)
STOMO by (GC-MBS)					0.00477^{***}	(0.00177)
SOMA.Bill by (GC-MBS)					0.0218	(0.273)
SOMA.NB by (GC-MBS)					-0.0535^{***}	(0.0123)
General Collateral–Mortgage-Backed Security Repo Spread					-0.0474^{***}	(0.00853)

(continued)

Table 6. (Continued)

Variables	**1**		**2**		**3**	
	Full Period January 2007–May 2010					
Global Financial Bond Index Option-Adjusted Spread	0.738**	(0.238)	0.755***	(0.264)	0.0234	(0.0228)
Options Exchange Volatility Index	−0.064	(0.253)	−0.0959	(0.270)	−0.0707*	(0.0387)
AA (Financial/Non-Financial) Commercial Paper	0.0121	(0.047)	−0.00571	(0.0473)	0.00299	(0.0122)
London Interbank Offered Rate–Overnight Index Swap	0.387*	(0.210)	0.414*	(0.212)	0.00167	(0.0202)
Lagged Federal Funds GC Repo Spread	−0.145***	(0.044)	−0.149***	(0.0447)	−0.0963***	(0.00685)
Quarter End	42.46***	(8.468)	41.47***	(8.523)	15.61***	(0.972)
Quarter Start	−32.42***	(7.717)	−32.39***	(7.737)	−12.35***	(2.522)
Year End	43.07	(59.410)	42.86	(60.47)	−8.779	(12.49)
Year Start	−36.03	(43.730)	−35.51	(43.38)	4.225	(4.433)
Constant	3.534***	(0.726)	3.332***	(0.777)	1.343***	(0.126)
ARCH					0.964***	(0.0890)
TARCH					1.100***	(0.158)
GARCH					0.265***	(0.0117)
Constant					1.494***	(0.236)
Number of Observations	853		853		853	

Notes: All variables are expressed as first differences unless noted. Newey-West standard errors are in parentheses. ***, **, and * denote $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively.

(and significant) coefficient in column 3. Previously, the impact of the SFP on the FF-repo spread was confounded by the fact that SFP bills simultaneously drained bank reserves and increased the supply of Treasury collateral. Here, the SFP has no impact on the federal funds target rate set by the FOMC, so we observe just the GC repo rate impact. Based on the results in table 6, we conclude that there is some evidence that the SFP was able to help alleviate funding market stresses.

4.4 Sub-Sample Analysis

Table 7 presents results over three sub-periods of our observation period, early (January–July of 2007), mid (August 2007–December 2008), and late (January 2009–May 2010). Panel A replicates Treasury collateral impacts over the sub-periods. Panel B distinguishes between injections and extractions of Treasury collateral. Panel C demonstrates an ability to reject the null hypothesis of a unit root over each sub-period.

The first and fifth columns in panel A replicate full sample results presented in tables 4 and 6, for the reader's convenience. Notice that the middle and late periods are most associated with statistically significant effects. Amplitudes in the middle period were higher than in the late period by an order of magnitude. This is not surprising, as the late period coincides with the target federal funds range of 0–25 bps; there is less scope for large movements in the repo rate spread in either direction during this period. Note that a more refined breakdown of Treasury supply by program is not possible, as some sources of Treasury collateral were not available in all sub-periods. For example, the TSLF was not initiated until the middle period.

When injections and extractions are treated separately, injections are generally statistically significant and economically meaningful, especially over the mid-crisis period (August 2007–December 2008). Extraction impacts are statistically indistinguishable from zero. These results suggest that injections generated the desired impacts, while extractions caused little to no money-market disruption.

Finally, concerning stationarity of these data, panel C considers an additional time period. This allows the reader to be assured

Table 7. Alternate Periods for Analysis and Dickey-Fuller Test Results

	Effective minus GC				Target minus GC			
	1/07–5/10	**1/07–7/07**	**8/07–12/08**	**1/09–5/10**	**1/07–5/10**	**1/07–7/07**	**8/07–12/08**	**1/09–5/10**
Variables	**Full**	**Early**	**Mid**	**Late**	**Full**	**Early**	**Mid**	**Late**
A. Replication of Results over Sub-Periods								
U.S. Treasury Issuance, Total	-0.167^{***} (0.057) (Table 4, Col. 1)	−0.115 (0.074)	$-.241^{*}$ (0.124)	$-.0343^{***}$ (0.007)	-0.157^{***} (0.048) (Table 6, Col. 1)	−0.043 (0.059)	$-.227^{**}$ (0.105)	$-.0538^{***}$ (0.007)
B. Injections and Extractions Treated Separately								
Injections of U.S. Treasury Collateral	$-.0195^{**}$ (0.080)	$-.403^{**}$ (0.188)	$-.324^{*}$ (0.194)	$.0465^{***}$ (0.007)	$-.170^{**}$ (0.067)	−0.248 (0.185)	$-.297^{*}$ (0.164)	$.0667^{***}$ (0.008)
Extractions of U.S. Treasury Collateral	−0.08 (0.078)	0.115 (0.071)	−0.03 (0.137)	0.012 (0.016)	−0.116 (0.081)	0.117 (0.074)	−0.052 (0.124)	−0.006 (0.016)
C. Stationarity of Data								
	1/07–5/10	**1/07–7/07**	**8/07–12/08**	**1/09–5/10**	**1/07–5/10**			
	Full	**Early**	**Mid**	**Late**	**Mid-Late**			
Dickey-Fuller Test for Unit Root	-10.7^{***}	-6.613^{***}	-7.349^{***}	-8.277^{***}	-9.675^{***}			
Number of Obs.	853	146	355	352	707			
1% Critical Values	−3.430	−3.452	−3.452	−3.452	−3.430			

Data Source: Federal Reserve Bank of New York.

Notes: All variables are expressed as first differences unless noted. Panels A and B: Other coeffiecients are not presented. Newey-West standard errors are in parentheses. ***, **, and * denote $p < 0.01$, $p < 0.05$, and $p < 0.1$, respectively. Panel C: Results for Effective minus GC spread, t-statistics based on asymptotic Dickey-Fuller distribution; *** denotes $p < 0.01$.

that over the whole of the TSLF policy period (over the mid-to-late period, combined), the data are suitable for the econometric analysis undertaken.

5. Conclusion

This study has investigated the impact of policies involving the supply of Treasury debt on overnight Treasury GC repo rates, an important benchmark rate in collateralized funding markets. In general, we find that increases in Treasury collateral supplies increase repo rates and narrow the spread between repo rates and the effective federal funds rate. Since flight-to-quality flows over the course of the recent financial crisis drove GC rates to relatively low levels, this narrowing of the spread is interpreted as a reduction in market stress. Broadly, we find U.S. Treasury supplies were useful, the method of injection to market mattered, and policies targeting short-term stress were effective both in terms of introduction and exit.

More specifically, we find that the TSLF, which was introduced to address stresses in short-term funding markets, was effective in alleviating the dislocations due to the increased demand for Treasury collateral as the crisis progressed. We also provide evidence that the TSLF had an impact beyond the funding market for Treasury collateral. In particular, our results are consistent with the TSLF having alleviated stress in the ABCP market. When we isolate changes in the GC-repo market from changes in the effective federal funds rate and allow for non-constant volatility, we also find that the SFP, which was designed to drain bank reserves, contributed to funding market stabilization. In addition, we find that the TSLF would likely have a much smaller effect on repo rates if the program had been introduced in a period of normal market functioning. Finally, our results suggest that the extraction of Treasury collateral did not disrupt funding markets in any significant way.

An implication of our results is that a proper policy response to a financial crisis may require responses other than increases in bank reserves. The TSLF program was very effective at alleviating short-term money-market stresses during the financial crisis and this program was reserve neutral. Furthermore, the SFP actually drained reserves from the banking system.

Our results also suggest a modified interpretation of Bagehot's dictum. During the financial crisis, securitized lending markets experienced an upheaval not only because of a lack of cash but also because of a lack of high-quality collateral, and thus increasing supplies of this type of collateral facilitated market functioning.

Appendix

Table 8 presents regression results with the dependent variable set as the change in the spread between the overnight Treasury GC repo rate via Bloomberg and the effective federal funds rate.

Table 8. Effective Federal Funds-(Public) Treasury General Collateral Repo Rate Spread Analysis

Variables		
Term Securities Lending Facility	−0.925**	(0.38)
Supplemental Financing Program	−0.071	(0.17)
U.S. Treasury Issuance, Bills	−0.140**	(0.07)
U.S. Treasury Issuance, Notes and Bonds	−0.092**	(0.04)
Temporary Open-Market Operations	−0.412	(0.37)
System Open Market Account Transactions, Bills	2.515**	(0.92)
SOMA Transactions, Notes and Bonds	0.387	(0.28)
Global Financial Bond Index Option-Adjusted Spread	0.508	(0.37)
Options Exchange Volatility Index	−0.077	(0.30)
AA (Financial/Non-Financial) Commercial Paper	0.0252	(0.06)
London Interbank Offered Rate–Overnight Index Swap	0.59**	(0.27)
Lagged Federal Funds GC Repo Spread	−.242***	(0.05)
Quarter End	44.49***	(11.75)
Quarter Start	−36.06***	(13.55)
Year End	−9.914	(26.36)
Year Start	4.23	(21.28)
Constant	4.11***	(0.66)
Number of Observations	853	

Notes: All variables are expressed as first differences unless noted. Newey-West standard errors are in parentheses. ***, **, and * denote p < 0.01, p < 0.05, and p < 0.1, respectively.

References

Adrian, T., K. Kimbrough, and D. Marchioni. 2011. "The Federal Reserve's Commercial Paper Funding Facility." *Economic Policy Review* (Federal Reserve Bank of New York) 17 (1): 25–39.

Adrian, T., and H. S. Shin. 2010. "Liquidity and Leverage." *Journal of Financial Intermediation* 19 (3): 418–37.

Bagehot, W. 1873. *Lombard Street: A Description of the Money Market*. London: King. Reprint, Gloucester, UK: Dodo Press, 2006.

Bartolini, L., S. Hilton, S. Sundaresan, and C. Tonetti. 2011. "Collateral Values by Asset Class: Evidence from Primary Securities Dealers." *Review of Financial Studies* 24 (1): 248–72.

Brunetti, C., M. di Filippo, and J. H. Harris. 2011. "Effects of Central Bank Intervention on the Interbank Market During the Subprime Crisis." *Review of Financial Studies* 24 (6): 2053–83.

Carpenter, S., and S. Demiralp. 2006. "The Liquidity Effect in the Federal Funds Markets: Evidence from Daily Open Market Operations." *Journal of Money, Credit and Banking* 38 (4): 901–20.

Duffie, D. 1996. "Special Repo Rates." *Journal of Finance* 51 (2): 493–526.

Fleming, M. J. 2007. "Who Buys Treasury Securities at Auction?" *Current Issues in Economics and Finance* (Federal Reserve Bank of New York) 13 (1).

Fleming, M. J., and K. D. Garbade. 2004. "Repurchase Agreements with Negative Interest Rates." *Current Issues in Economics and Finance* (Federal Reserve Bank of New York) 10 (5).

———. 2007. "Dealer Behavior in the Specials Market for US Treasury Securities." *Journal of Financial Intermediation* 16 (2): 204–28.

Fleming, M. J., W. B. Hrung, and F. M. Keane. 2009. "The Term Securities Lending Facility: Origin, Design, and Effects." *Current Issues in Economics and Finance* (Federal Reserve Bank of New York) 15 (2).

———. 2010a. "Repo Market Effects of the Term Securities Lending Facility." Staff Report No. 426, Federal Reserve Bank of New York.

———. 2010b. “Repo Market Effects of the Term Securities Lending Facility.” *American Economic Review Papers and Proceedings* 100 (2): 591–96.

Fleming, M. J., and J. V. Rosenberg. 2007. “How Do Treasury Dealers Manage Their Positions?” Staff Report No. 299, Federal Reserve Bank of New York.

Friedman, B. M., and K. N. Kuttner. 2010. “Implementation of Monetary Policy: How Do Central Banks Set Interest Rates?” In *Handbook of Monetary Economics*, ed. B. Friedman and M. Woodford, 1345–1438. North-Holland.

Gagnon, J. E., M. Raskin, J. Remache, and B. P. Sack. 2011. “The Financial Market Effects of the Federal Reserve’s Large-Scale Asset Purchases.” *International Journal of Central Banking* 7 (1): 3–43.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. “On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks.” *Journal of Finance* 48 (5): 1779–1801.

Gorton, G., and A. Metrick. 2012. “Securitized Banking and the Run on the Repo.” *Journal of Financial Economics* 104 (3): 425–51.

Graveline, J. J., and M. R. McBrady. 2011. “Who Makes On-the-Run Treasuries Special?” *Journal of Financial Intermediation* 20 (4): 620–32.

Greenwood, R., and D. Vayanos. 2014. “Bond Supply and Excess Bond Returns.” *Review of Financial Studies* 27 (3): 663–713.

Hilton, S., and W. B. Hrung. 2010. “The Impact of Banks’ Cumulative Reserve Position on Federal Funds Rate Behavior.” *International Journal of Central Banking* 6 (3): 101–18.

Jordan, B. D., and S. D. Jordan. 1997. “Special Repo Rates: An Empirical Analysis.” *Journal of Finance* 52 (5): 2051–72.

Keane, F. 1996. “Repo Rate Patterns for New Treasury Notes.” *Current Issues in Economics and Finance* (Federal Reserve Bank of New York) 2 (10).

Krishnamurthy, A. 2010. “Amplification Mechanisms in Liquidity Crises.” *American Economic Journal: Macroeconomics* 2 (3): 1–30.

Krishnamurthy, A., S. Nagel, and D. Orlov. Forthcoming. “Sizing Up Repo.” *Journal of Finance*.

Longstaff, F. A. 2004. "The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices." *Journal of Business* 77 (3): 511–26.

McAndrews, J., A. Sarkar, and Z. Wang. 2008. "The Effect of the Term Auction Facility on the London Inter-Bank Offered Rate." Staff Report No. 335, Federal Reserve Bank of New York.

Neely, C. J. 2010. "The Large-Scale Asset Purchases Had Large International Effects." Working Paper No. 2010-018C, Federal Reserve Bank of St. Louis.

Seligman, J. S. 2006. "Does Urgency Affect Price at Market? An Analysis of U.S. Treasury Short-Term Finance." *Journal of Money, Credit and Banking* 38 (4): 989–1012.

Simon, D. 1991. "Segmentation in the Treasury Bill Market: Evidence from Cash Management Bills." *Journal of Finance and Quantitative Analysis* 26 (01): 97–108.

Sundaresan, S., and Z. Wang. 2009. "Y2K Options and Liquidity Premium in Treasury Markets." *Review of Financial Studies* 22 (3): 1021–56.

Taylor, J. B., and J. C. Williams. 2009. "A Black Swan in the Money Market." *American Economic Journal: Macroeconomics* 1 (1): 58–83.

Tucker, P. 2009. "The Repertoire of Official Sector Interventions in the Financial System: Last Resort Lending, Market-Making, and Capital." Remarks at the Bank of Japan 2009 International Conference on the Financial System and Monetary Policy: Implementation, Tokyo, May 27–28. Available at http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2009/speech390.pdf.

Lessons from the Historical Use of Reserve Requirements in the United States to Promote Bank Liquidity*

Mark Carlson
Board of Governors of the Federal Reserve System

Efforts in the United States to promote bank liquidity through reserve requirements, a minimum ratio of liquid assets relative to liabilities, extend as far back as 1837. Despite such requirements, banking panics and suspensions of deposit convertibility continued to occur. Eventually, policymakers created a central bank to ensure bank liquidity. This paper reviews the historical debates about reserve requirements, supplemented by empirical evidence, to provide insights relevant today about using reserve requirements to regulate liquidity. The insights are related to convincing institutions to use the reserve during stress events and the ways reserve requirements for banks affect interactions with other financial firms before and during a panic.

JEL Codes: G21, N21, E58, B00.

1. Introduction

Shortly after the Panic of 1837, states began instituting reserve requirements which mandated that banks had to hold some minimum ratio of liquid assets relative to their liabilities. When Congress passed the National Bank Acts in the 1860s, banks receiving

*Lindsay Boersma provided valuable research assistance. I thank Burcu Duygan-Bump, Brandon DuPont, Galina Hale, John James, Tom King, Jamie McAndrews, Bill Nelson, Enrico Perotti, Jonathan Rose, and Eugene White as well as seminar participants at the Federal Reserve Bank of Chicago and the Federal Reserve Bank of New York for helpful comments. The views expressed in this paper are solely those of the author and do not necessarily reflect those of the Federal Reserve Board or its staff. Author Contact: Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue, Washington, DC 20551. E-mail: mark.a.carlson@frb.gov.

national bank charters also faced a reserve requirement. These rules were part of an effort to promote liquidity and soundness by ensuring that each individual bank had a pool of liquid assets that it could draw on during times of stress.[1]

Despite these efforts, the banking system remained vulnerable to banking panics and suspensions of convertibility in which banks temporarily stopped or restricted withdrawals of funds (Calomiris and Gorton 1991, Sprague 1910, Wicker 2000). These suspensions of convertibility disrupted economic activity and demonstrated that the reserve requirements were not sufficient to ensure that the financial system remained liquid during periods of stress (James, Weiman, and McAndrews 2013). Partly to address these concerns, Congress established the Federal Reserve to create an "elastic" currency that could add liquidity to the banking system and serve as a lender of last resort.[2]

This paper reviews the historical thinking and experience in the United States regarding reserve requirements to provide insights for policymakers today regarding efforts to promote individual bank liquidity and the relation of those efforts with a lender of last resort. Following the recent financial crisis, there has been increased interest in liquidity requirements among policymakers, such as the liquidity coverage ratio proposed by the Basel Committee on Banking Supervision (2010) or in countries such as the United Kingdom, the Netherlands, and others, and in academic research such as Buiter (2009), Cao and Illing (2011), Perotti and Suarez (2011), and Rochet (2008). Historical policymakers, academics, and bankers thought deeply about reserve requirements, and their discussions provide insights into broad theoretical concerns and practical issues regarding the effectiveness of reserve requirements. Moreover, important

[1] The United States was one of a few countries to have a formal reserve requirement (Comptroller 1895). Other countries, such as Canada and Scotland, had informal, and more flexible, best practices about the appropriate level of reserves (Bordo, Redish, and Rockoff 1993, Kroszner 1995, and MacKenzie 1932).

[2] There were other sources of instability in the financial system at this time, such as seasonal fluctuations in interest rates, the inelasticity of the money supply, and a fragmented banking system. These issues have been discussed extensively elsewhere (see, for example, Bordo and Wheelock 2010, Calomiris 2009, and Friedman and Schwartz 1963). This paper focuses more on how reserve requirements affected bank behavior during and around panics, and the other issues are discussed only to the extent they bear on this issue.

lessons regarding how reserve requirements perform during stress situations can be drawn from the historical narrative and by analyzing changes in bank balance sheets during the panics of the era in which reserve requirements were in use.

One important lesson that is apparent from historical experience is that it is quite difficult to convince banks to use their reserve. Banks appear to have been willing to close their doors and temporarily suspend operations, which had notable reputational consequences, well before they had exhausted their reserve. This tendency appears to have been exacerbated by uncertainties regarding the rules about when use of the reserve was permitted. Relatedly, concerns about the dependability of the usual liquidity backstop impact panic dynamics. Banks often depend on other banks or particular financial markets to meet their regular liquidity needs. During a panic, the ability of those other banks or markets to furnish that liquidity support may become impaired. Uncertainty about how a liquidity backstop will operate during a crisis can lead banks to hoard liquid assets during stress events, a dynamic that may exacerbate the problems (see also Allen and Gale 2007). The importance of having a credible and dependable lender of last resort to prevent this dynamic was a key argument made by proponents of a central bank (Warburg 1916).

The historical experience also provides a lesson on the relationship between reserve requirements and moral hazard issues. Some of the modern literature has focused on using liquidity requirements to reduce banks' reliance on the central bank as a backstop (Cao and Illing 2011, Rochet 2008). However, the historical record indicates that reserve requirements may lead other institutions, not subject to the requirements, to opt to rely more on the banks as a source of liquidity. This tendency could put additional pressure on the banks during a crisis.

Another lesson is that the liquidity of assets being used in the reserve can change during a crisis. Moreover, if the same assets are used as a store of liquidity by both regulated and non-regulated institutions, shifts in asset liquidity can transmit financial pressures from one group to the other.

The paper also reviews the compliance of banks with the reserve requirement. Compliance was generally high, but it was regularly the case that a notable number of banks were deficient. The historical experience illustrates how compliance can be related to the discipline provided by the bank examiners.

Many of these lessons were discussed by policymakers as they debated the creation of the Federal Reserve. This paper reviews the debates and describes how policymakers expected a central bank to address some of the inadequacies of reserve requirements in mitigating panics. These discussions provide insights into how the founders of the Federal Reserve expected it to act in response to a financial crisis. Fairly soon after the founding of the Federal Reserve, reserve requirements ceased to be viewed as a tool to manage individual bank liquidity and instead were seen as a tool for managing overall credit conditions and financial market liquidity. This shift in thinking is also discussed.

This paper is organized as follows. Section 2 reviews the introduction of reserve requirements and some alternative ways they were constructed. The purpose of the reserve requirements is discussed in section 3. Section 4 describes historical lessons regarding the use of reserve requirements as liquidity-management regulation, in normal times and especially during panics. Section 5 presents the arguments made for establishing a central bank and indicates how these arguments stemmed from some of the issues raised in previous sections. This section also notes the decline in the use of reserve requirements as a tool for regulating liquidity following the establishment of the Federal Reserve. Section 6 concludes.

2. Introduction and Calculation of Historical Reserve Requirements

The first reserve requirements were introduced in the United States shortly following the Panic of 1837 by the states of Virginia, Georgia, and New York (Rodkey 1934). These requirements were generally intended to ensure that banks had ready access to resources that would enable them to meet their liability obligations. (When reserve requirements were first adopted, banks were chartered exclusively by the states, so it was state laws that mattered.) The adoption of reserve requirements by other states occurred slowly; only ten states had such laws by 1860. However, after the Panic of 1857 there were a number of journal articles and pamphlets advocating reserve requirements.[3]

When reserve requirements were first enacted, the main bank liability was bank notes, privately issued currency that banks promised to redeem for specie (gold or silver coin), and state laws referred to those liabilities as the base for determining the appropriate reserve. As the liability base of banks shifted toward deposits, the reference point for the reserve requirements shifted as well. In 1842, Louisiana passed a law requiring banks to maintain a reserve in specie equal to one-third of their total liabilities to the public, which included both notes and deposits (White 1893). By 1895, twenty-one states had reserve requirements for commercial banks; at this time, all such laws included deposits in liability base (Comptroller 1895). For states that enacted reserve requirements, the laws regarding the ratio of reserves that had to be held relative to the liability base ranged from between 10 percent to 33 percent.

There was some variation in the types of deposits included in the base of the reserve. By the end of the 1800s, a majority of states required the reserve to be calculated against all deposits. However, some policymakers argued that banks ought to maintain a greater reserve only against more volatile deposits, and a few states just required a reserve against demand deposits (Comptroller 1895, Welldon 1910). A handful of states mandated reserves against both demand and time deposits, but specified that the amount of liquid resources that needed to be held against each dollar of time deposits was smaller than that required for demand deposits.

State laws also differed with respect to what assets could be included in the reserve held against liabilities. Some states allowed deposits in other banks to count; this feature was likely due to the fact that many banks in smaller communities maintained balances at banks in larger cities to clear payments. As many bank notes, and later checks, were redeemed at these clearing banks, interbank deposits played an important part in a bank's liquidity profile (James 1978, White 1983). A few states required that the entire reserve be carried as specie in the bank's vault. A handful of states allowed short-term loans to count as part of the reserve.

[3]See also Miller (1927) for discussion of shifts in thinking around this time.

When the U.S. Congress passed the National Banking Acts in the early 1860s and provided for national bank charters, the legislation included reserve requirements for national banks. These reserve requirements were tiered and depended on the location of the banks. For much of the National Banking Era, banks located outside major cities—referred to as "country banks"—were required to hold reserves equal to 15 percent of deposits, three-fifths of which could be held as deposits in banks located in reserve cities while the rest was required to be held as vault cash.[4] Banks in reserve cities—generally larger cities—were required to hold reserves equal to 25 percent of deposits, half of which could be carried as balances in central reserve cities. Banks in central reserve cities—at first just New York but later Chicago and St. Louis as well—held significant amounts of interbank deposits. These banks were required to maintain a reserve equal to 25 percent of deposits which needed to be held in gold or in legal tender.[5] One reason that banks in reserve and central reserve cities were expected to hold a higher portion of their assets as reserves was that they held more interbank deposits; these deposits were seen as more volatile and, in particular, more likely to be withdrawn during banking panics (Federal Reserve 1927).

Although it was typical for interbank deposits to be included in the reserve, this choice was subject to some criticism, as it was understood that these assets were generally not an effective source of liquidity during banking panics. Noyes (1894) noted that when demand during a panic was for physical currency, reserves held elsewhere were not particularly useful. More fundamentally, allowing interbank deposits to count as reserves created a pyramid structure. A bank could deposit cash in another bank and count that deposit in its reserve while the second bank counted the cash in its reserve. The second bank could then deposit the cash in a third bank and compound the process. A withdrawal of reserves by the bottom of the pyramid during a panic would thus result in a rapid depletion of reserves within the banking system (*Bankers Magazine* 1907a).

[4]Prior to 1874, national banks had to hold reserves against both notes and deposits.

[5]A number of states also adopted a tiered system in which state-chartered banks in larger cities faced more stringent reserve requirements.

The Comptroller of the Currency (Comptroller)—the chief regulator of national banks—argued that reserves held in other banks had been ineffective in protecting depositors during the panics of 1873 and 1893 and encouraged Congress to increase the portion of the reserve that banks had to carry in their vaults (Comptroller 1900, pp. 25–27).

3. The Purpose of the Reserve

Reserve requirements were implemented as a prudential requirement meant to ensure that banks maintained the resources to meet their obligations. This goal is very broad and has both solvency and liquidity attributes. Indeed, proponents of reserve requirements often blended the two or spoke of the benefits in terms of both the safety of the banks and the promptness with which banks could meet withdrawals.

An example of arguments framing the reserve as a tool for supplying liquidity comes from the 1873 report of the Comptroller, where it was noted that "the question is not whether a reserve shall be held which shall insure the *payment*, merely, of the note, for that is unnecessary, but what amount of reserve shall be held by the banks to insure the *prompt* payment of all their liabilities?" (p. 19). With respect to solvency benefits, Tucker (1858) suggested that bank failures, such as during the Panic of 1857, were the result of "imprudence" as banks over-extended themselves and did not maintain a reserve of at least one-third of their liabilities.[6] Most advocates tended to blend these ideas. Hooper (1860) provided one of the most interesting blends. He argued that a bank could reduce its riskiness by increasing either its capital or its reserve, and institutions with a greater fraction of their assets as reserves could sustain a higher

[6] Capital also played an important role in prudential regulation, and it was well understood that an adequate capital base was necessary to establish the safety of the bank. Tucker (1839) maintained that "to secure the requisite solidity, two things seem essential. One is that the bank should have a sufficient amount of capital; and the other is, that such capital should be real, not nominal or borrowed [i.e., paid in installments or with borrowed funds]" (p. 192). Interestingly, and in contrast to the reserve requirements, capital requirements were often expressed in terms of a fixed dollar amount rather than as a ratio. Nevertheless, a few states in the early 1800s did limit the amount of loans that could be extended to some multiple of capital.

leverage ratio.[7] Having a strong reserve meant that the bank would be able to avoid being forced to access emergency funds from other banks or rapidly call in their loans (or presumably be forced to sell assets in fire sales) and thus be stronger overall.

Commentators also noted that the reserve was important in promoting confidence and could prevent troubles from starting. Hooper (1860) suggested that confidence about the reserve likely affected banks' willingness to use it. In particular, he argued that banks in New Orleans were required by law to maintain a higher reserve than those in other large cities (Boston in particular) and that the populace of New Orleans, knowing the strength of the reserve, had greater confidence in their banks and thus the New Orleans banks were better able to use their reserve times of financial trouble. Westerfield (1921) summarizes this debate regarding a minimum reserve requirement succinctly: "The psychological effect of the known large reserve is undoubtedly good, allaying the depositor's fears as to the bank's ability to pay on demand; but if the depositor also realized that this reserve could not be actually used when required, his faith would be less strong" (p. 146).

While much of the discussion focused on the micro-prudential benefits to the individual banks of requiring a minimum level of reserves, some commentators did suggest that there were systemic benefits of ensuring banks retained sufficient liquid resources on hand. Opdyke (1858) argued that excessive credit growth led to a boom-and-bust cycle and that a reserve requirement could be useful in restraining credit growth.[8]

[7]While he does not use these terms, Hooper appears to argue that banks should target a particular overall level of risk and can hit that risk target by adjusting either their capital-to-asset ratio (leverage) or the balance between riskless cash and risky loans. Calomiris (2011) makes a fairly similar point.

[8]Opdyke advocated limiting loans to one and a half times capital plus specie, which also suggests that reserves blended solvency and liquidity benefits. He further argued that the importance of restraining credit growth, and hence of a reserve requirement, had more than an economic motivation. "Whenever our currency is thrown into one of its paroxysms of extreme expansion, by the undue enlargement of bank loans, it literally intoxicates commerce, and drives it into all kinds of excesses. The desire of gain is stimulated to an unwonted degree, and manifests itself in over-trading, imprudent credits, reckless speculations, and numerous enterprises of questionable utility and still more questionable morality" (p. 8).

Other systemic benefits were described by Coe (1873) and Hooper (1860), who suggested that there were collective-action reasons to mandate minimum reserves, especially for banks in the main money center of New York City. Hooper noted that the reserve of banks in New York was a common good benefiting all the banks in the city as well as the rest of the country.[9] He further argued that the managers of those banks might not internalize the social benefit they provided and, as a consequence, the law needed to require them to hold a larger reserve than the banks would otherwise have chosen.

> It is out of the question for the banks of the city of New York to hold that relation of the entire confidence through the country, so long as the action of each bank, in regard to the amount of its reserve of specie, is dependent upon the peculiar views or character of its board of managers. The *law* must secure the uniform ability of the banks to meet their engagements by making it imperative upon each one of them to hold the requisite amount of specie as a condition of their power to discount. (Hooper 1860, p. 44)

Coe noted that banks in New York City were linked through their mutual dependence on the call loan market for liquidity and that interior banks tended to react to troubles at one bank in New York as a signal of troubles at all the banks. (The call loan market was a money market in which banks and others would provide short-term loans to brokers secured by equity collateral. As equities were liquid, when a bank needed to call in a loan, the broker was assumed to be able to sell the collateral to repay the loan. However, if all the banks called these loans at the same time, the brokers would be forced to liquidate their collateral at fire-sale prices and might be unable to repay the loans.) Thus, during a panic the strong banks needed to support the weak institutions to contain liquidity drains and prevent problems from cascading. This linkage, Coe argued, was a reason that all banks needed to hold a strong reserve and was a

[9]The centrality of banks in New York City reflected their place in the payment system, as a typical source of liquidity through the provision of discounting, and as an access point to the New York money markets. This importance was enhanced when New York City was designated as a central reserve city at the top of the reserve pyramid in the National Banking Era (see also James 1978).

motivation for the New York Clearinghouse to establish a reserve requirement in 1857.

4. Lessons from History Regarding Reserve Requirements

This section reviews four lessons regarding reserve requirements. The lessons concern the enforcement of the requirement during normal times, use of the reserve during panics, the complexities of moral hazard issues, and the changing liquidity of assets during panics.

4.1 Enforcing the Reserve Requirement

Modern discussions about enforcing the reserve requirement involve efforts to balance the need to ensure that banks meet the requirement with the desire not to push banks to maintain too substantial a buffer. Goodhart (2008) favors a tiered structure based on the severity of the violation. The Basel Committee on Banking Supervision (2010) lists extensive considerations regarding enforcement rules.

The historical debate about how to ensure that banks met the reserve requirement started soon after the first requirements were introduced. Tucker (1839) advocated enforcing the requirement using a moderate penalty proportional to any deficiency of the reserve. He maintained that the penalty should be high enough to dissuade banks from running below the reserve in good times but not so high that banks were unwilling to use the reserve during a crisis.[10]

In the National Banking Era, the law provided that in the event the Comptroller found that a national bank was deficient in its reserve, the bank could be required to cease making loans and stop paying dividends until the reserve was restored.[11] If the reserve was

[10]Opdyke (1858) argued for requiring a minimum reserve somewhat below what was desired, as he maintained that banks would hold a buffer stock above the requirement and that the buffer could then be used: "A legal minimum of 20 per cent. will, it is believed, give a practical minimum of not less than 25 to 30 per cent., for no prudent bank will voluntarily occupy a position on the verge of legal death" (pp. 15–16).

[11]The Comptroller (1893) stated that in the event that the bank had loaned out too great a portion of its funds or depositors had withdrawn a significant

not restored within thirty days, the Comptroller could, with the concurrence of the Secretary of the Treasury, appoint a receiver for the bank. It was well noted that both the finding by the Comptroller that the bank was deficient and the decision to seek a receiver were discretionary on the part of the Comptroller. Moreover, the Comptroller stated that he only had the opportunity to learn about the bank's balance sheet from one of the biannual bank examinations or the report of condition filed five times a year (1893). The actual ability to monitor was slightly more complicated. In their examination reports, examiners were asked to review the bank's books and comment on the adequacy of the bank's reserve for the past thirty days (or more if deemed appropriate). Thus the examination reports allowed the Comptroller more than just a single day's observation.

Carter Glass (1913) asserted that this particular penalty regime was not very successful. In the debates related to the Federal Reserve Act, he maintained that the penalties for holding inadequate reserves for an extended period were so severe that they were not applied and that in some cases banks had been allowed by regulators to have deficient reserves for several years. Examination reports for a sample of banks indicate that the examiners took note of the condition of the bank's reserves and used this information, along with other aspects of the bank's condition, to make recommendations about whether the bank should be required to suspend dividend payments or take other remedial actions. An inadequate reserve alone may have resulted in some scolding by the examiner but does not appear to have been sufficient to cause the examiner to recommend strong corrective action. However, if the bank had other problems, such as too many large loans or a rising number of overdue loans, then a deficient reserve was used by the examiners as an additional lever to push the bank to take remedial actions.

Information from Welldon (1910) suggests that, as of 1909, many states had similar, though slightly less severe, penalties for banks falling short of their reserve requirement. Out of the thirty-nine states that had reserve requirements at that time, Welldon mentions

amount of funds, the only "safe and prudent course for the bank to pursue is to cease paying out money in any direction except to depositors until either through the collection of demand or maturing loans on the one hand, or the receipt of deposits on the other, the required portion has been restored" (p. 18).

a penalty for failing to meet that reserve for twenty-five states. In every case, that penalty involved a prohibition on extending new loans. In fifteen cases, there was also a prohibition on issuing dividends. For only one state, Arizona, does Welldon mention an explicit provision that failure to restore the reserve could result in a bank being declared insolvent.

Looking at the status of reserves for a sample of 208 national banks in both reserve cities (82 banks) and larger country towns (126 banks) using data from the September 1892 Call Report provides some further information about the level of bank reserves.[12] Most banks appear to have held reserve in excess of the required reserve; the average reserve ratio was around 31 percent and quite similar for both country banks and those in reserve cities (table 1).[13] (These ratios are similar to those reported by the Comptroller in 1887.) Moreover, the ratio of reserves to deposits exceeded the legal requirement (15 percent for country banks and 25 percent for reserve city banks) by 10 or more percentage points for two-thirds of country banks and one-fifth of reserve city banks. Banks may have preferred

[12]The sample used here consists of 208 national banks from cities located predominantly in the West and South; these parts of the country were most affected by the Panic of 1893, a fact that will become useful below. The sample was limited in size in order to allow for more in-depth analysis of these banks. Importantly, the examiner reports used in the paper are also from this group of institutions. The sample includes most of the banks in this part of the country that would have stood between the country banks and the banks in the money centers and thus the banks most affected by developments in interbank markets during the panic. The particular cities used are Birmingham and Mobile, Alabama; Los Angeles and San Diego, California; Denver and Pueblo, Colorado; Indianapolis, Indiana; Des Moines and Dubuque, Iowa; Lexington and Louisville, Kentucky; New Orleans, Louisiana; Minneapolis, Rochester, St. Paul, and Stillwater, Minnesota; Kansas City and St. Joseph, Missouri; Helena, Montana; Lincoln and Omaha, Nebraska; Fargo, North Dakota; Cincinnati, Ohio; Portland, Oregon; Knoxville, Memphis, and Nashville, Tennessee; Dallas, El Paso, and San Antonio, Texas; Salt Lake City, Utah; Spokane and Tacoma, Washington; Milwaukee and Racine, Wisconsin; and Cheyenne, Wyoming. Data are from Comptroller (1892).

[13]While the reserve ratio was roughly cash and due from reserve agents divided by net due to banks and individual deposits, the precise calculation includes various offsets and adjustments such as for cash items in the process of collection, clearinghouse exchange, and unpaid dividends. These adjustments increased the reserve ratio by 2 percentage points on average for the banks in sample. The calculation here follows the procedure used by the national bank examiners and described in Coffin (1896).

Table 1. Reserve Ratios in 1892 (actual reserves as a percent of reservable liabilities)

	Country Banks			Reserve City Banks		
	Reserve in Bank	**Held at Reserve Agents**	**Total Reserve Ratio**	**Reserve in Bank**	**Held at Reserve Agents**	**Total Reserve Ratio**
Mean	15.5	16.6	32.1	15.0	15.2	30.2
75th Percentile	18.7	20.5	38.4	19.0	18.3	35.1
Median	14.0	14.5	29.3	14.0	14.4	27.8
25th Percentile	9.2	8.9	20.8	11.6	9.9	24.6
Standard Deviation	9.7	12.8	9.6	5.6	7.6	8.5
Observations	126			82		

Source: Call Report for September 30, 1892 (Comptroller 1892).

to hold reserve ratios in excess of what was required simply because they preferred being more liquid, as is suggested by *Bankers Magazine* (1908), or because they viewed the required reserve ratio as a minimum they did not want to breach and desired to maintain a buffer.[14]

Reserves held in the bank (as opposed to with reserve agents) accounted for about half the total reserve. Relative to deposits, reserves at the bank averaged about 14 percent for both groups, also well above the legal requirements of 6 percent for country banks and 12.5 percent for reserve city banks. The finding that about half the reserve was held in cash matches similar findings by the Comptroller a decade or so later (Comptroller 1907).

While many banks appear to have preferred to hold reserves well in excess of what was legally required, some banks had deficiencies in their reserve ratios. Of the banks in the sample, 10 percent of the country banks and 25 percent of reserve city banks had a deficient

[14]Baer and McElravey (1993) provide (and test) a model of a bank's optimal capital stock that illustrates how the presence of asymmetric information that raises the costs of issuing equity could prompt banks to maintain a capital ratio in excess of the regulatory requirement. Similar arguments could apply to the excess reserve ratios found here.

reserve. The fact that banks had deficient reserves suggests that they did not see the reserve ratio as something that had to be met at all times (perhaps especially because they had thirty days to restore it upon notice by the Comptroller). Nevertheless, most of these banks did not sink too far below the legal limit—many of them being within 3 percentage points of the requirement.[15]

4.2 Use of the Reserve Requirements during Panics

Whether the banks would actually use the reserve is very important. As noted earlier, a key motivation for the historical reserve requirements was to have a store of liquid assets that could be used during a panic. Modern liquidity regulations also presume that the reserve will be used (Basel Committee on Banking Supervision 2010, Stein 2013). This section describes the legal ability of banks to use their reserves, the alternatives they had, and some evidence on actual use. Banks appear to have been quite reluctant to use their reserve, and this section discusses a particular feature of the banking system during the National Banking Era that may have contributed to this reluctance.

It should be noted that some recent work suggests that reserves may be beneficial if they are *not* usable. For instance, Calomiris (2011) and Calomiris, Heider, and Hoerova (2012) argue that an unusable reserve limits liability holder losses in the event of default, as more safe assets put a floor on possible losses. Moreover, an unusable reserve may also reduce risk shifting. As both modern and historical rules intend(ed) the reserve to be used, I focus here on lessons regarding usability.

4.2.1 Legal Ability of Banks to Use Their Reserve

The concern that legally required reserves held by banks would not be helpful if the banks had to maintain these reserves at all times

[15]Comparing reserve deficiencies over time suggests only limited persistence. Looking at banks with Call Report data for both 1892 and 1894, of the ten banks with the lowest reserve ratios in 1892, only one was still in the ten with the lowest reserve ratios in 1894 and only three were in the bottom thirty banks.

and could not use them was stated clearly early on. In 1848, Kettell argued that "This *keeping* of 15 per cent. of specie on *hand* has been tried in New York, in Alabama, and elsewhere, and its gross absurdity always made manifest. Of what use is it that a bank has the gold and silver, if the law forbids it to part with it?"

The debate about whether banks could legally use their reserves continued during the National Banking Era. In his annual report for 1894, the Secretary of the Treasury argued against the reserve requirement, saying that, as the law was silent on when the national banks could use their reserves, it created a situation in which they were unusable: "Among these are the requirements . . . that a fixed reserve, which cannot be lawfully diminished, shall be held on account of deposits. The consequence of this last requirement is that when a bank stands most in need of all its resources it cannot use them without violating the law" (p. LXXIX).

Proponents of reserve requirements responded that the reserve was established with the intent that it be used during stress periods. As noted above, the decision to find a bank deficient in its reserve was discretionary on the part of the Comptroller. This discretion allowed the Comptroller to effectively waive the requirement during a panic and allow banks time to rebuild their reserves subsequently (Comptroller 1893).[16]

Others viewed the vagueness of the law regarding the use of the reserve to be a notable impediment to banks' willingness to use the reserve.[17] *Bankers Magazine* (1907b) argued that the vagueness of the law regarding when the reserve could be drawn down meant that many bankers felt that the reserve could not be used during a crisis. The president of the American Bankers Association expressed similar sentiments in 1908 (see *Bankers Magazine* 1908).

[16]It is unknown whether the examiners were given explicit instructions to waive the requirement. No mention of any changes in standards appear in the Comptroller's annual reports in which responses to panics are discussed, nor is anything apparent from the examiner reports reviewed.

[17]Providing certainty about when the reserve could be used was seen as inherently difficult. Coe (1873) argued that it is very challenging to prescribe rules regarding the circumstances or timing in which the reserve should be allowed to be used or rebuilt.

4.2.2 Alternative Options for Banks

Banks during this period had some alternatives to running down their reserve during a crisis. One option was to try to expand the supply of liquid assets. Banks in New York and other large cities formed clearinghouses to facilitate the settlement of payments between members. These clearinghouses also provided a way to supply liquidity to member banks during a panic by allowing banks to deposit securities with the clearinghouse and receive clearinghouse loan certificates that could be used to make payments to other members of the clearinghouse. Using clearinghouse notes allowed specie or other forms of cash to be used to satisfy the heightened demand from others for liquid assets (Comptroller 1873 and 1890, Nash 1908). Clearinghouse notes were issued extensively in the Panic of 1907. In New York, these notes continued to be large-denomination notes, but in many smaller cities small-denomination notes were issued and circulated with other currency in the general public market. The clearinghouse notes worked for interbank and sometimes local transactions, but were not well suited for interregional payments and were thus an imperfect remedy.

Another alternative to using the reserve during this period was for the bank to suspend convertibility of deposits or to completely suspend operations. If a single national bank in the town closed its doors, it would be reviewed by the bank examiners before being allowed to reopen. Examiner reports indicate the examiners would require actions on the part of both owners and depositors (such as capital injections and agreements to refrain from large withdrawals, respectively) before the bank would be reopened. Moreover, when only one bank in the town suspended, there were often notable reputational consequences for that institution. In a severe situation, the clearinghouse might coordinate the closing of all the banks in the community. In these cases, the clearinghouse might be able to coordinate a reopening without all banks needing to be examined.

4.2.3 Evidence on Use of the Reserve during a Panic

Evidence on use of the reserve during panic situations suggests some reluctance to do so. Of the banks in the sample described in section 4.1, forty-eight closed temporarily or permanently during the Panic of 1893, and the examiner reports conducted when these institutions were closed provide useful information about the willingness of banks to utilize their reserve when under pressure.[18] Among the banks that closed, either permanently or temporarily, the median cash reserve ratio was about 2 percent, well below the legal requirement. Nevertheless, the median cash ratio was still decidedly positive, and fourteen institutions had cash ratios of 6 percent or more. (Balances due from reserve agents represented about 60 percent of the total reserve for banks on the day they were closed. Such balances were notably higher at banks that suspended but were allowed to reopen, which suggests that an inability to access these funds in a timely fashion may have contributed to the suspensions of some banks during the panic.)

As their reserves were depleted during banking panics, banks in the central reserve city of New York would suspend or curtail shipments of currency to other parts of the country.[19] Sprague (1913) argued that the New York banks tended to do so well before they had exhausted their reserve. In 1907, the *Wall Street Journal* noted that reserves were about 21 percent of deposits around the time of suspension, below the legal requirement but still fairly high.[20] It was noted in the *Journal* that use of the reserve during the panic was appropriate:

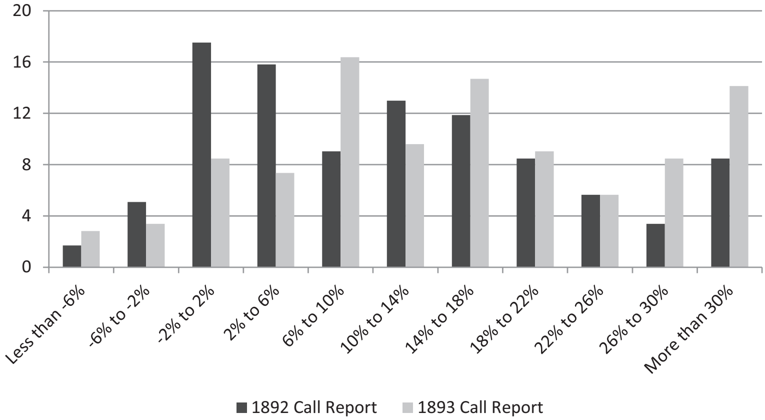
> [T]here is a deficit of the bank reserve of $38,838,825. It should be remembered, however, that a reserve is for use. There is no wisdom in locking up immense sums of money in bank vaults

[18] For the histories of the Panic of 1893, as well as other panics of the National Banking Era, see Sprague (1910) and Wicker (2000).

[19] Banks in reserve cities would in turn suspend payments to country banks. A lack of available currency gave rise to the use of scrip and other currency substitutes (see Andrew 1908 and Warner 1895).

[20] Typically, all the banks in the clearinghouse would publish their individual balance sheets each week. During the periods when currency shipments were suspended, it was common for the clearinghouse association to provide a common balance sheet. Doing so avoided revealing whether a particular institution was deficient in its reserve and may have reflected concerns about revealing a deficiency. However, such periods also corresponded to periods in which clearinghouse loan certificates were being issued and it may have been more appropriate to consider the collective liquidity position of the clearinghouse members rather than looking at them individually.

Figure 1. Distribution of Total Reserve Ratios in Excess of Legal Requirement

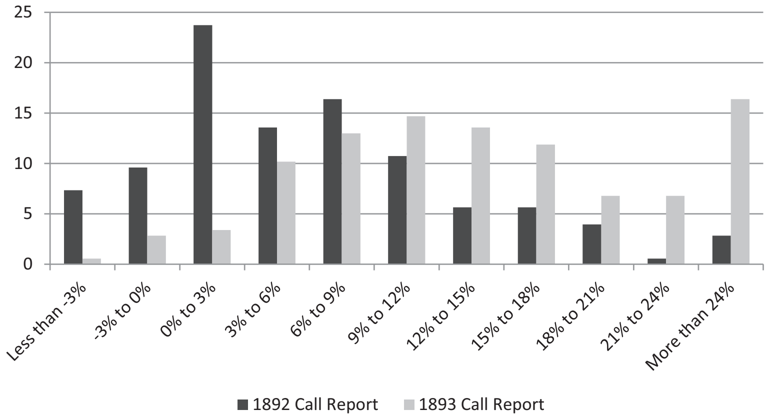


Source: Call Report for September 30, 1892 and October 3, 1893 (Comptroller 1892, 1893).

> unless they can be employed in times of emergency. (*Wall Street Journal,* Nov. 4, 1907)

Detailed information on bank balance sheets and reserves is available at a time shortly after the panic from the October 3, 1893 Call Report. Comparing reserve ratios in excess of the legal requirement in 1893 to such ratios in 1892 for national banks operating in both periods (using the same cities as before) suggests that reserve ratios rose at both country banks and reserve city banks (figure 1). The change in the composition of the reserves is even more dramatic as banks shifted from holding their reserve with agents to holding it as cash. The average ratio of cash to liabilities subject to the reserve rose by nearly 10 percentage points; for many banks, cash holdings alone became enough to satisfy the reserve requirement (figure 2). By contrast, average reserves held at agent banks relative to reservable liabilities fell by 5 percentage points. As might be expected, when faced with the possibility of deposit runs and large-scale withdrawals, banks preferred to hold more liquid assets and, in particular, to have cash on hand. Given the pyramiding of reserves that occurred through interbank deposits, this shift to holding reserve in

Figure 2. Distribution of Cash Reserve Ratios in Excess of Legal Requirement



Source: Call Report for September 30, 1892 and October 3, 1893 (Comptroller 1892, 1893).

cash at the bank would have contracted the supply of reserves for the banking system as a whole.

The increase in the cash reserve ratio is due largely to a drop in the liability base. Individual deposits declined 28 percent between 1892 and 1893 at the median bank while net due to banks dropped more than 50 percent. By contrast, cash balances decreased only a little (about 5 percent). Cash balances may have been maintained in part as banks called in funds from their reserve agents; such balances also decreased more than 50 percent. The declines in the various measures of interbank activity again illustrate the collapse in the general liquidity of the financial system as banks pulled back from each other. Loans contracted 17 percent, possibly as banks raised cash by calling in loans (of course, loans may have declined for other reasons such as lack of demand).

Comparing reserves in 1894 to those in 1892 provides some information about longer-term changes to reserve holdings following a panic. It appears that for country banks there was little persistent impact on the reserve ratios, either in the overall ratio or in the composition of the reserve. For reserve city banks, there may have been somewhat more persistence in the effect of the panic, as these institutions continued to maintain slightly higher reserve ratios a year later (33 percent in 1894 versus 30 percent in 1892). That increase appears evenly split between cash and deposits with reserve agents.

4.2.4 Importance of the Liquidity Backstop

Banks in New York City were vital liquidity providers to the financial system (James 1978). Balances held at these institutions were key for processing payments and could be drawn down if banks needed currency at home. As banks in different parts of the country ran currency surpluses or deficits, specie would be shipped in or out of New York and redistributed around the country. Moreover, New York City banks were important providers of short-term interbank funds through the practice of rediscounting bills.

The centrality of New York City banks to the system had serious repercussions during panics. When New York City suspended convertibility, banks around the country lost access to their typical liquidity backstop. During a panic, individual banks sought to pull their funds out of the banks in the reserve cities and bolster their liquid resources (*Bankers Magazine* 1908); several observers, such as the Comptroller (1907) and Herrick (1908), reported that declines in interbank deposits contributed at least as much to the panic as the actions of individual depositors.[21] Knowing that banks in New York had suspended payments to out-of-town banks during prior banking panics, by the Panic of 1907, banks were reported to be preemptively trying to pull their funds out of New York City before the banks there could suspend. The country banks were viewed as acting out of self-preservation because there was no guarantee that their regular source of liquidity would be able to furnish that liquidity should the crisis intensify (Roberts 1908, Sprague 1913). Thus, the uncertainty about the dependability of the typical source of liquidity appears to have exacerbated the panic dynamics (such run dynamics are also described in Allen and Gale 2007).

[21]Stickney (1901) makes a similar point comparing the U.S. experience during panics with that of England. He argues that during times of stress, the U.S. system resulted in banks having to compete for reserves, which increased their scarcity, while in England the Bank of England faced no competition and could thus act to make reserves more readily available.

4.3 Moral Hazard Issues

One of the typical motivations for reserve requirements in modern literature is that it reduces the moral hazard concerns associated with a lender-of-last-resort backstop (Cao and Illing 2011, Ratnovski 2009, Rochet 2008). The historical record indicates that the moral hazard issue is more complicated; it suggests that when some institutions are subject to reserve requirements, other institutions may reduce their reserves and depend on liquidity being provided by those institutions subject to the requirement.

Following the Panic of 1893, there were discussions about whether some state banks were free-riding on the national banks. *Bankers Magazine* (1894) indicated that there was an expectation by some state banks that the national banks would use their reserves to provide liquidity and support to state banks and trust companies not subject to the reserve requirements, even if this support was only provided by the national banks to stem local unrest and thus protect themselves. The article indicated that this expectation was the source of some tension among bankers and resulted in some lack of cooperation during the panic.

Another example comes from the trust companies in New York, which played a key role in the Panic of 1907. These institutions took deposits and were similar to banks, but the state laws allowed them to operate with smaller reserve requirements; indeed, these institutions established themselves as trust companies partly to avoid capital and reserve requirements. Reportedly, the trust companies operated with cash on hand of only about 2 percent of their deposits, much less than the banks (Noyes 1901). Instead, as their store of liquidity, the trust companies held deposits at the commercial banks that earned interest and could be drawn upon in the event there were withdrawals. Noyes argues that the trust companies did so in part because the reserves maintained by the banks gave the trust companies confidence that the banks would be liquid in a crisis. Consequently, during the panic, the withdrawals by the trust companies became a serious drain on the banks (Sprague 1908).

4.4 Lessons Regarding the Assets Selected to Serve as the Reserve

As noted extensively above, the interbank deposits that were often included as part of the reserve tended to become illiquid during crises. This is a useful reminder that the liquidity properties of some assets are not the same in normal times as in stress episodes.[22]

A related point is also demonstrated by the dynamics of the call loan market. As noted above, call loans were short-term loans to stockbrokers to finance stock purchases and were collateralized by stocks. These loans could be called by the bank when funds were needed, and it was assumed the stockbrokers would be easily able to sell the stock to repay the loan. While not technically part of the reserve, these loans were viewed as highly liquid and served as a secondary reserve for both banks and trust companies. Moreover, a significant portion of the funding for the call loan market came from the banks and trusts.

The Panic of 1907 involved both runs on the trust companies and the banks and a plunge in the stock market. The call loan market quickly came under immense pressure as both banks and trust companies sought to utilize this secondary reserve and borrowers in that market that were unable to find alternative funding faced the prospect of selling their stocks in a fire sale and possibly defaulting. Consequently, institutions were not able to tap the call loan market as a secondary source of liquidity as they might normally do, and the functioning of that market deteriorated significantly.[23]

As noted by Moulton (1918) and highlighted by the call loan market example, securities holdings did not function well as a secondary reserve during a crisis in this period, as banks could often only sell their securities to other banks or similar institutions. If all banks were seeking to sell their securities holdings at the same

[22]Given the biases against interbank transactions and in favor of assets associated with safe-haven demands, modern liquidity requirements may be less subject to some of the concerns identified here.

[23]This dynamic was well known. In discussing the use of call loans as stores of short-term liquidity, Dwight (1858) noted that "the causes which alarm one bank alarm the whole. Upon any shock to confidence, they all call in at once. The stock collaterals are forced upon the market at the same moment that its ability to take them is almost destroyed by the total cessation of new loans" (p. 159). The argument was repeated following the panic of 1893, where it was noted that when all institutions used the same market as a source of liquidity, that market would be unable to provide liquidity during a panic (*Bankers Magazine* 1894). This dynamic is also the one warned against by Coe (1873) when he indicated that the size of banks' reserves had systemic implications and warned that strong banks would need to support weaker institutions.

time, those securities would not function as a source of liquidity. For a security to function as a source of liquidity during a crisis, ready purchasers from outside the banking system are needed (similar ideas also appear in Allen and Gale 2007 and Brunnermeier 2009).[24]

5. Reserve Requirements and the Federal Reserve

As it became clear that reserve requirements were not successful in preventing panics or in keeping banks liquid during panics, discussions turned to the need for a central bank. Many of the associated debates dealt with the issues raised in section 4. Subsequent to the establishment of the Federal Reserve, the purpose of reserve requirements shifted from being used to maintain individual bank liquidity to providing a tool for the central bank to control the cost of liquidity and credit.

5.1 Discussions Regarding the Establishment of a Central Bank

There were several lessons that policymakers took from the Panic of 1907 that prompted them to work toward establishing a central bank. One lesson was that when the instrument used as a reserve and primary source of liquidity—in this case, the supply of gold and Treasury notes—was fairly inelastic in the short run, demand for that instrument would exceed the available supply during a panic.[25] Banks would compete for these funds and withdraw funding from each other; the competition triggered some of the negative dynamics discussion in section 4.2.3. If banks used other markets as secondary reserves, the associated scramble for liquidity during panics would impair functioning in these other markets, as was the case in the call loan market noted in section 4.4.

[24] Around this time there was a broader debate about what it meant for assets and for banks to be liquid. Traditionally the self-liquidating nature of short-term loans had been emphasized as a source of liquidity, and banks were more liquid when they had more of these loans. This view was in the process of being challenged. See Suviranta (1933) for a more in-depth discussion.

[25] Gold could, and did, flow into the United States from abroad in response to rising interest rates. These inflows boosted liquidity but did take some time to arrive in quantities sufficient to meet demand.

To address these issues, many policymakers concluded that an "elastic" currency that could increase in quantity was required (Vanderlip 1908). The notion that an elastic currency was needed was not new.[26] However, following the Panic of 1907, legislative action seemed considerably more likely. Some proposals provided for an emergency currency that could be issued by a central authority only during a crisis; as a temporary palliative, such a currency was included in the Aldrich-Vreeland Act of 1908. Under this Act the Secretary of the Treasury could, during a crisis, authorize the issuance of currency backed by any securities held by banks instead of the usual requirement that the currency be backed by U.S. government bonds.[27] Ultimately, policymakers chose instead to create the Federal Reserve as a permanent solution where the discount window could be used to turn bank assets into central bank reserves and would thus provide an elastic currency that could be used to respond to changing stringencies in money markets more flexibly and continuously than could the issuance of emergency currency.

A closely related argument made by advocates of a central bank was that only central bank notes or reserves are certain to be liquid during a financial crisis (Sprague 1911).[28] Other assets were argued to be liquid only to the extent that they could be converted into central bank reserves:

> In countries where these notes of the central banks are generally accepted in settlement of debts by business men and banks,

[26] As early as 1868, the Comptroller argued in favor of providing some elasticity to the currency for use during times of stress, and at various times bankers had also argued for an elastic currency (Hamilton 1906 and Pugsley 1902), but these prior efforts had not resulted in any significant changes. White (1983) describes these and other initiatives.

[27] Some, such as Silber (2007), have argued that the issuance of Aldrich-Vreeland currency in 1914 helped prevent a panic in U.S. financial markets associated with the onset of World War I.

[28] In particular, Sprague (1911) and Warburg (1916) argued against thinking of interbank deposits as reserves even though these assets had traditionally been a part of banking reserves. They argued that interbank deposits tended to concentrate risk in a few large institutions, so that the entire system was affected if issues arose at those institutions and that there was no guarantee that those institutions in which the interbank deposits were vested would themselves remain liquid. Vanderlip (1908) noted that the interbank deposits created interdependence among financial centers and that once New York closed, other centers were compelled to suspend payments to out-of-town banks.

> the "banking reserves" of the stock banks may safely consist of the central bank currency, or of a balance kept with the central bank, convertible into such currency. These form the first line of banking reserves. The second line consists of those assets which, with certainty and promptness, may be converted into credit balances with the central bank. (Warburg 1916, p. 9)

Central bank reserves can also be expanded rapidly by the central bank during a stress episode. Moulton (1918) noted that the expansion of liquidity is essential during a crisis, as banks are expected to be the source of liquidity for their non-financial customers during a crisis, and if banks are required to bolster their own liquidity to support their reserve by demanding repayment of, or even refusing to renew, loans during a crisis, then financial strains can be significantly exacerbated.

As a central bank would be able to provide a guaranteed liquidity backstop, individual banks would not need to hoard liquidity at the first sign of stress because they would know that the backstop would still be available in a crisis (again addressing the negative dynamics discussed in section 4.2.3). Warburg (1914) goes a bit further and argues that, in order to prevent hoarding, the backstop and ability to supply cash must have the absolute credibility that only a central bank can provide.[29] It was expected that the existence of the central bank would prompt a change in behavior during a panic and would stop minor stresses from escalating into full blown crises (Warburg 1916).

A third lesson was that the liquidity requirements that tried to strike a balance between ensuring that the liquidity of the banking system was maintained yet not hampering banks in providing credit were likely to be overwhelmed during a panic. Even critics of central banks sought ways to allow private market participants to expand the supply of liquid assets during a panic.

One other aspect of the Panic of 1907 that was not lost on policymakers was that institutions outside the normal banking system

[29] However, it should be noted that the Federal Reserve was subject to a reserve requirement of its own in that each Reserve Bank had to maintain a gold reserve equal to 40 percent of its note issue and reserves in the form of gold or other eligible securities equal to the remaining 60 percent. Bordo and Wheelock (2010) argue that this requirement restrained the Federal Reserve actions at some times.

that were not required to maintain a reserve, in this case the trust companies, could precipitate a drain on the liquidity of the banking system (the moral hazard issue from section 4.3). The realization that these outside institutions could threaten the stability of the system may have prompted some large, influential New York Clearinghouse Association members to support a central bank (see Moen and Tallman 1999, White 1983).

5.2 Reserve Requirements after the Founding of the Federal Reserve

With the establishment of the Federal Reserve, required reserves were reduced, as it was expected that the liquidity backstop from the central bank provided individual commercial banks with a ready means of meeting extraordinary liquidity demands.[30] As noted by Rodkey (1934, p. 64):

> With the advent of the Federal Reserve System in 1914, we entered upon an era of central banking . . . It is clear that the presence of a central bank, prepared to make advances on eligible assets, places the individual bank in a less vulnerable position with respect to demands of its depositors. It tends to lessen the need for primary reserves. The Federal Reserve Act recognized this fact by reducing materially the percentage of required reserves.

Westerfield (1921) noted that the reduction in reserves was appropriate for several reasons, including because the reserves were concentrated (as opposed to dispersed across banks throughout the system), because the reserves were "located in a central bank which feels its responsibility," and because their "availability is now unquestioned."[31] Lunt (1922), who provided instructions to insurers on how to assess the quality of a bank from its balance sheet, noted that prior to the founding of the Federal Reserve, the statement of cash and cash items "was regarded as extremely important, and

[30]See also Feinman (1993), who provides more detail on reserve requirements from the founding of the Federal Reserve until the 1990s.

[31]White (1983) suggests that further lowering the reserve requirements also enhanced the attractiveness of joining the Federal Reserve System.

banks that habitually carried larger reserves than those required by law were thought to be exceptionally safe" (p. 217). However, with the Federal Reserve, the "point seems far less important now, since any bank that has a proper loan account can replenish its reserve at will by the simple process of rediscounting."

While reserve requirements continued to be viewed as a tool to promote bank liquidity for some time, there was a gradual shift away from this view. Indeed, by the late 1930s, reserve requirements were no longer seen as playing an important role in providing liquidity:

> The committee [Federal Reserve System Committee on Bank Reserves] takes the position that it is no longer the case that the primary function of legal reserve requirements is to assure or preserve the liquidity of the individual member bank. The maintenance of liquidity is necessarily the responsibility of bank management and is achieved by the individual bank when an adequate proportion of its portfolio consists of assets that can be readily converted into cash. Since the establishment of the Federal Reserve System, the liquidity of an individual bank is more adequately safeguarded by the presence of the Federal Reserve banks, which were organized for the purpose, among others, of increasing the liquidity of member banks by providing for the rediscount of their eligible paper, than by the possession of legal reserves. (Federal Reserve 1938)

It is useful to note that during this period, the Federal Reserve was a regular lender to the banking system. Burgess (1936) notes that in a typical month during the mid-1920s, about one-third of member banks obtained at least one loan or advance from their Reserve Bank. As a regular lender, it would be fairly easy for the Federal Reserve to provide additional liquidity to individual banks. The discount window was seen by Federal Reserve staff as the primary source of emergency liquidity for the banking system, especially after the range of eligible collateral was significantly expanded in 1932.

Rather than promoting individual bank liquidity, reserve requirements became seen as a tool to manage credit growth and facilitate the use of monetary policy. This development occurred as the Federal Reserve began to use open-market operations to adjust available reserves in the banking system as its primary monetary policy tool;

it was seen as impractical to have reserves both serve as a source of liquidity and be manipulated for monetary policy purposes.

> The two main functions of legal requirements for member bank reserves under our present banking structure are, first, to operate in the direction of sound credit conditions by exerting an influence on changes in the volume of bank credit, and secondly, to provide the Federal Reserve banks with sufficient resources to enable them to pursue an effective banking and credit policy. (Federal Reserve 1938)

6. Lessons and Concluding Remarks

From the late 1830s until 1913, regulatory efforts aimed at promoting bank liquidity consisted primarily of reserve requirements that mandated that individual institutions hold liquid assets. However, these requirements were not sufficient to provide liquidity and prevent banks from suspending deposit withdrawals during banking panics. It became clear to many observers that the reserve requirements created, at best, a static pool of liquidity that banks would compete over during a crisis. To enable the expansion of that pool of liquid assets to meet the extraordinary liquidity demands experienced during a crisis, the Federal Reserve was established.

Policymakers today are considering various liquidity requirements for banks. For instance, under the Basel III requirements, banks will be subject to a liquidity coverage ratio (LCR). Under this requirement, banks will be required to maintain a stock of high-quality and liquid assets as a buffer that is sufficient to cover potential net cumulative cash outflows at all times during a thirty-day period. The LCR is similar to a reserve requirement in that it effectively requires liquid assets to be held against certain classes of liabilities (and lines of credit).

Several lessons from the historical reserve requirements for liquidity regulation today are apparent. One lesson is that it is extremely challenging to convince institutions to switch from maintaining a reserve to drawing down that reserve, especially during a period when their inclination is to take steps to bolster their liquidity situation. The historical experience highlights how the tendency to hoard liquidity can be heightened by uncertainty about when it is appropriate to use the reserve and by uncertainties regarding the availability of other liquidity backstops. These uncertainties may in fact contribute to the strength and speed of the run and point to the importance of implementing modern regulations in ways that minimize these uncertainties.

The historical experience also shows that requiring some institutions to carry a reserve can cause other institutions to reduce their own reserves and rely instead on the institution subject to the requirement. Such responses have important implications for who benefits from and who bears the burdens of the liquidity requirements. In several historical episodes, other institutions put additional pressures on banks subject to reserve requirements to support them during panics. In at least one case, institutions not subject to the reserve appear to have made a conscious choice to reduce their own liquidity buffer because they believed the banks would need to use theirs.

References

Allen, F., and D. Gale. 2007. *Understanding Financial Crises.* Oxford: Oxford University Press.

Andrew, A. P. 1908. "Substitutes for Cash in the Panic of 1907." *Quarterly Journal of Economics* 22 (4): 497–516.

Baer, H., and J. McElravey. 1993. "Capital Shocks and Bank Growth—1973–1991." *Economic Perspectives* (Federal Reserve Bank of Chicago) 17 (4): 2–21.

Bankers Magazine. 1894 (April). New York: Homans Publishing.

———. 1907a (July). New York: The Bankers Publishing Company.

———. 1907b (August). New York: The Bankers Publishing Company.

———. 1908 (November). New York: The Bankers Publishing Company.

Basel Committee on Banking Supervision. 2010. *Basel III: International Framework for Liquidity Risk Management, Standards and Monitoring.* Bank for International Settlements.

Bordo, M., A. Redish, and H. Rockoff. 1993. "A Comparison of the United States and Canadian Banking Systems in the Twentieth Century: Stability vs. Efficiency." NBER Working Paper No. 4546.

Bordo, M., and D. Wheelock. 2010. "The Promise and Performance of the Federal Reserve as Lender of Last Resort 1914–1933." Working Paper No. 2010-036A, Federal Reserve Bank of St. Louis.

Brunnermeier, M. 2009. "Deciphering the Liquidity and Credit Crunch 2007–2008." *Journal of Economic Perspectives* 23 (1): 77–100.

Buiter, W. 2009. "Central Banks and Financial Crises." In *Maintaining Stability in a Changing Financial System.* Proceedings from an Economic Policy Symposium sponsored by the Federal Reserve Bank of Kansas City, held in Jackson Hole, Wyoming, August 21–23, 2008.

Burgess, W. R. 1936. *The Reserve Banks and the Money Market.* New York: Harper and Brothers Publishers.

Calomiris, C. 2009. "Banking Crises and the Rules of the Game." NBER Working Paper No. 15403.

———. 2011. "Identifying the Right Mix of Capital and Cash Requirements in Prudential Bank Regulation." Mimeo, Columbia University.

Calomiris, C., and G. Gorton. 1991. "The Origins of Banking Panics: Models, Facts, and Bank Regulation." In *Financial Markets and Financial Crises*, ed. G. Hubbard, 109–75. Chicago: University of Chicago Press.

Calomiris, C., F. Heider, and M. Hoerova. 2012. "A Theory of Bank Liquidity Requirements." Mimeo, Columbia University.

Cao, J., and G. Illing. 2011. "Endogenous Exposure to Systemic Liquidity Risk." *International Journal of Central Banking* 7 (2): 173–216.

Coe, G. 1873. *Report to the New York Clearing House Association.* New York: W. H. Arthur & Company.

Coffin, G. 1896. *Handbook for Bank Officers.* Washington, DC: McGill and Wallace.

Comptroller of the Currency. Various years. *Annual Report.* Washington, DC: U.S. Government Printing Office.

Dwight, E. 1858. "The Financial Revulsion and the New York Banking System." *Hunt's Merchants Magazine and Commercial Review* 38 (February): 157–63.

Federal Reserve. 1927. *Report on Member Bank Reserve Requirements.* Memorandum prepared by Mr. Smead of the Division of Bank Operations.

———. 1938. *Member Bank Reserve Requirements – Analysis of Committee Proposal.* Memorandum prepared by Woodlief Thomas.

Feinman, J. 1993. "Reserve Requirements: History, Current Practice, and Potential Reform." *Federal Reserve Bulletin* (June): 569–89.

Friedman, M., and A. Schwartz. 1963. *A Monetary History of the United States, 1867–1960.* Princeton, NJ: Princeton University Press.

Glass, C. 1913. Speech in the House of Representatives on September 10.

Goodhart, C. 2008. "Liquidity Risk Management." *Financial Stability Review* (Banque de France) 11 (February): 39–44.

Hamilton, J. 1906. "Currency Reform." Address before the Michigan Bankers Association in Grand Rapids, Michigan, June. Reprinted in *Practical Problems in Banking and Currency*, ed. W. Hull. New York: Macmillan Company, 1907.

Herrick, M. 1908. "The Panic of 1907 and Some of Its Lessons." *Lessons of the Financial Crisis*, *Annals of the American Academy of Political and Social Science* 31 (2): 8–25.

Hooper, S. 1860. *An Examination of the Theory and the Effects of Laws Regulating the Amount of Specie in Banks.* Boston: Little, Brown and Company.

James, J. 1978. *Money and Capital Markets in Postbellum America.* Princeton, NJ: Princeton University Press.

James, J., D. Weiman, and J. McAndrews. 2013. "Wall Street and Main Street: The Macroeconomic Consequences of New York Bank Suspensions 1866–1914." *Cliometrica* 7 (2): 99–130.

Kettell, T. P. 1848. "The Money of Commerce." *De Bow's Review* 6 (October): 243–64.

Kroszner, R. 1995. "Free Banking: The Scottish Experience as a Model for Emerging Economies." Policy Research Working Paper No. 1536, World Bank.

Lunt, E. 1922. *Surety Bonds.* New York: The Ronald Press Company.

MacKenzie, K. 1932. *The Banking Systems of Great Britain, France, Germany, and the United States of America.* London: McMillan and Company.

Miller, H. 1927. *Banking Theories in the United States Before 1860.* Cambridge, MA: Harvard University Press.

Moen, J., and E. Tallman. 1999. "Why Didn't the United States Establish a Central Bank Until After the Panic of 1907?" Working Paper No. 99-16, Federal Reserve Bank of Atlanta.

Moulton, H. G. 1918. "Commercial Banking and Capital Formation, part III." *Journal of Political Economy* 26 (7): 705–31.

Nash, W. 1908. "Clearing-House Certificates and the Need for a Central Bank." In *Lessons of the Financial Crisis*, *Annals of the American Academy of Political and Social Science* 31 (2): 61–66.

Noyes, A. 1894. "The Banks and the Panic of 1893." *Political Science Quarterly* 9 (1): 12–30.

———. 1901. "The Trust Companies: Is There Danger in the System?" *Political Science Quarterly* 16 (2): 248–61.

Opdyke, G. 1858. *A Report on the Currency.* New York: John Trow, Printer.

Perotti, E., and J. Suarez. 2011. "A Pigovian Approach to Liquidity Regulation." *International Journal of Central Banking* 7 (4): 3–41.

Pugsley, C. 1902. "Emergency Circulation." Address to the American Bankers Association in New Orleans, Louisiana, November. Reprinted in *Practical Problems in Banking and Currency*, ed. W. Hull. New York: Macmillan Company, 1907.

Ratnovski, L. 2009. "Bank Liquidity Regulation and the Lender of Last Resort." *Journal of Financial Intermediation* 18 (4): 541–58.

Roberts, G. 1908. "The Need of a Central Bank." In *Lessons of the Financial Crisis, Annals of the American Academy of Political and Social Science* 31 (2): 45–54.

Rochet, J.-C. 2008. "Liquidity Regulation and the Lender of Last Resort." *Financial Stability Review* (Banque de France) 11 (February): 45–51.

Rodkey, R. 1934. *Legal Reserves in American Banking.* Ann Arbor, MI: University of Michigan, School of Business Administration, Bureau of Business Research.

Secretary of the Treasury. 1894. *Report of the Secretary of the Treasury.* Washington, DC: U.S. Government Printing Office.

Silber, W. 2007. "The Great Financial Crisis of 1914: What Can We Learn from Aldrich-Vreeland Emergency Currency?" *American Economic Review: Papers and Proceedings* (May): 285–89.

Sprague, O. M. W. 1908. "The American Crisis of 1907." *Economic Journal* 18 (71): 353–72.

———. 1910. *History of Crises Under the National Banking System.* Washington, DC: National Monetary Commission.

———. 1911. "The Aldrich Plan for Monetary Legislation." *American Economic Review* 1 (2): 262–71.

———. 1913. *Banking Reform in the United States.* Cambridge, MA: Harvard University Press.

Stein, J. 2013. "Liquidity Regulation and Central Banking." Address at the Credit Markets Symposium sponsored by the Federal Reserve Bank of Richmond, held in Charlotte, North Carolina, April 19. Available at http://www.federalreserve.gov/newsevents/speech/stein20130419a.htm.

Stickney, A. B. 1901. "The Medium of Exchange and the Banking Function." Address to the American Bankers Association in Milwaukee, Wisconsin, October. Reprinted in *Practical Problems in Banking and Currency*, ed. W. Hull. New York: Macmillan Company, 1907.

Suviranta, B. 1933. "The Shiftability Theory of Bank Liquidity." In *Economic Essays in Honour of Gustav Cassel.* London: G. Allen & Unwin.

Tucker, G. 1839. *The Theory of Money and Banks Investigated.* Boston: Charles Little and James Brown.

———. 1858. "Banks or No Banks." *Hunt's Merchants Magazine and Commercial Review* 38 (February): 147–57.

Vanderlip, F. 1908. "The Panic as a World Phenomenon." In *Lessons of the Financial Crisis*, *Annals of the American Academy of Political and Social Science* 31 (2): 2–7.

Wall Street Journal. 1907. Various issues.

Warburg, P. 1914. "A United Reserve Bank of the United States." *Proceedings of the Academy of Political Science in the City of New York* 4 (4): 75–115.

———. 1916. "The Reserve Problem and the Future of the Federal Reserve." Speech before the Convention of the American Bankers Association, Kansas City, Missouri, September 29.

Warner, J. D. 1895. "The Currency Famine of 1893." *Sound Currency* 2 (6): 337–56.

Welldon, S. 1910. *Digest of State Banking Statues.* Washington, DC: National Monetary Commission.

Westerfield, R. 1921. *Banking Principles and Practice,* Vol. 1 and II. New York: The Roland Press Company.

White, E. 1983. *The Regulation and Reform of the American Banking System, 1900–1929."* Princeton, NJ: Princeton University Press.

White, H. 1893. "National and State Banks." *Annals of the American Academy of Political and Social Science* 3 (May): 1–30.

Wicker, E. 2000. *Banking Panics of the Gilded Age.* New York and Melbourne: Cambridge University Press.

How Monetary Policy Is Made: Two Canadian Tales*

Pierre L. Siklos[a] and Matthias Neuenkirch[b]
[a]Wilfrid Laurier University and Balsillie School of International Affairs
[b]University of Trier

We examine policy rate recommendations of the Bank of Canada's Governing Council (GC) and its shadow, the C.D. Howe Institute's Monetary Policy Council (MPC). Individual recommendations of the MPC are observed but not those of the GC. Differences in the two committees' recommendations are small but persistent. The MPC is more responsive to the output gap than its GC counterpart. Both committees respond similarly to inflation. Disagreement within the MPC and with the GC is more likely when rates are rising. Finally, the Bank's forward guidance had a significant influence on the MPC's views about the future inflation path.

JEL Codes: E43, E52, E58, E61, E69.

*Research for this paper was carried out in part while the first author visited the Viessmann Centre. Siklos gratefully acknowledges financial assistance from a CIGI-INET research grant. Comments on an earlier draft by four anonymous referees, Michael B. Devereux (the editor), Angelo Melino, Philippe Bergevin, and Alessandro Riboni are much appreciated. Earlier drafts were presented at the 2012 Public Choice Society Conference in New Orleans, LA; the RWTH Workshop on Central Bank Transparency and Communication in Aachen, Germany; the 2013 European Public Choice Society Meeting in Zurich, Switzerland; and the 2013 Money, Macro, and Finance Research Group Meeting in London, United Kingdom. An unpublished appendix as well as the data will be posted on the Central Bank Communication network website (http://www.central-bank-communication.net/). The first author has been a member of the C.D. Howe Institute's Monetary Policy Council since 2008 but receives no direction or funding from the C.D. Howe Institute for participating in this group. All members of the Monetary Policy Council provide an independent opinion on monetary policy issues. Corresponding author: Matthias Neuenkirch, University of Trier, Department of Economics, Universitätsring 15, 54286 Trier, Germany. E-mail address: neuenkirch@uni-trier.de.

1. Introduction

Understanding how monetary policy decisions are made became easier, at least until the global financial crisis that began in late 2007. Most central banks gravitated towards reliance on a single instrument of policy, usually a short-term interest rate. This approach is more transparent and easier for markets and the public to follow. These developments popularized reliance on the Taylor (1993) rule—which relates the central bank's policy rate to indicators of inflationary pressure and slack in economic activity—as a straightforward device observers can rely on to determine whether policy is too loose or tight. If the actual policy rate is below that which Taylor's rule recommends, policy is deemed too loose and vice versa when the policy rate is above what the rule suggests.[1] The simplicity of this approach generates regular commentary in the press (e.g., see Davies 2011) and central bankers frequently comment about the stance of monetary policy in relation to what it would be if the central bank followed some monetary policy rule.

Enhanced central bank transparency has also stimulated interest in second guessing the decisions of the monetary policy authorities. In response, "shadow" committees have emerged. They provide independent recommendations on what the appropriate policy rate ought to be.[2] Canada pioneered their introduction when, in 2002, the C.D. Howe Institute (CDHI) created the Monetary Policy Council (MPC) to provide "the Bank of Canada, financial-market participants and economic policy commentators with a regular independent assessment of the appropriate stance of Canadian monetary

---

[1]In light of the events since 2008, there has been a shift towards identifying and explaining persistent deviations from the Taylor rule, termed by some the "Great Deviation" (e.g., Taylor 2013). Indeed, it was this kind of analysis which prompted disagreement between the rule's creator and the former Chairman of the Federal Open Market Committee (FOMC), Ben Bernanke, about whether monetary policy was to blame for the U.S. financial crisis of 2007–9 (Taylor 2007, Bernanke 2010). Of course, monetary policy cannot be reduced to a simple equation as Poole (2006), among others, reminds us. Central bankers are also at pains to point out that there is considerable uncertainty, for example, about the size of any output gap.

[2]As this is written, there are shadow monetary policy committees in the United States, the euro area, the United Kingdom, Australia, and New Zealand. See also Neuenkirch and Siklos (2013).

policy."[3] The adoption of inflation targeting and a more transparent and accountable Bank of Canada also facilitated the formation of a shadow committee that could engage in the kind of second-guessing exercise examined in this paper.

This paper examines the record of the MPC and asks whether we can identify differences between its recommendations and the decisions of the Bank's Governing Council (GC). Second, we obtain insights about the sources of disagreement about the appropriate stance of Canadian monetary policy. In particular, the MPC records individual recommendations, whereas we cannot observe individual members' views inside the GC. Finally, the sample period under study is of specific interest since it covers the Bank's conditional commitment to keep the policy rate at the zero lower bound. This period marks the Bank's foray into providing forward guidance. How the Bank's announcement impacted the MPC's recommendations and views about the path for inflation in Canada is also investigated.

Our results suggest, first, that differences in the two committees' recommendations are small but persistent. Second, the MPC is more responsive to the output gap than its GC counterpart. Both committees respond similarly to inflation. Third, disagreement within the MPC and with the GC is more likely when rates are rising. Finally, the Bank's forward guidance had a significant influence on the MPC's views about the future inflation path.

The rest of the paper is organized as follows. Section 2 describes the functioning of the CDHI's MPC and compares it with the Bank's GC. Section 3 introduces the econometric methodology. Section 4 presents some stylized facts, the empirical results, and a counterfactual experiment. Section 5 concludes with some policy implications.

2. The Monetary Policy Council and the Governing Council

2.1 The C.D. Howe Institute's Monetary Policy Council

Since 2002, the C.D. Howe Institute has convened a Monetary Policy Council that includes academic and professional economists.

[3]See http://www.cdhowe.org/monetary-policy-council-2.

The MPC consists of twelve members named by the CDHI based on the expertise of potential candidates to comment and provide recommendations on the appropriate stance of monetary policy.[4] Its aim is to "discuss the Bank of Canada's policy toward the overnight rate ... shortly before each of the Bank's interest-rate announcements." In other words, the MPC provides *independent* advice about what the appropriate policy rate ought to be to reach the Bank's inflation target and *not* a forecast of how the Bank is likely to set future policy rates. Moreover, in the last few years, the MPC, unlike its counterpart at the Bank, has also provided information about the appropriate interest rate path for up to a year ahead.[5]

MPC meetings are usually chaired by the president and CEO of the CDHI or, in his absence, the vice-president of research. Meetings are normally held fivc days before the Bank's announcement of the overnight rate.[6] For a member's vote to be recorded, he or she must either be present in person or participate via teleconference. When the Bank sets the overnight rate on a Tuesday,[7] the MPC meets

[4]The list of current members, along with their background and affiliations, can be found at http://www.cdhowe.org/monetary-policy-council-2. Members do not receive any financial support from the institute, nor are they asked to adopt a particular ideology in making monetary policy recommendations. Invitations to join the MPC normally come from the president of the CDHI, currently William Robson. He is also the one who first convened the committee when he was the CDHI's director of research. In correspondence with Robson, the MPC is deliberately structured to ensure a "diversity of approaches" about how monetary policy ought to be conducted. Former Bank of Canada employees have never been invited to join the MPC.

[5]An evaluation of this recommended forward interest rate path can be found in Neuenkirch and Siklos (2014).

[6]Clearly, the Bank may receive new information after the CDHI's MPC decision. Whether this delay has a material impact on potential differences between the MPC's recommendation and the Bank's decision is unclear. While the volume of information received by decision makers at the Bank is clearly greater than what the CDHI's MPC likely considers, it is not obvious—unless there is (i) a major crisis between the MPC and GC meeting dates or (ii) some decisive private information available only to the GC—that the gap between MPC and GC meeting days is significant. The timing gap was examined as part of our robustness tests but did not affect our conclusions (results are relegated to the appendix).

[7]In 2013 the regular policy rate announcement was rescheduled to Wednesday to parallel the release of the quarterly Monetary Policy Report.

the previous Thursday. Occasionally, when the Bank announces the policy rate on a Thursday, the MPC will meet the preceding Tuesday. Meetings follow an agenda which has remained unchanged since the MPC's inception.[8]

Each meeting lasts approximately one hour, begins at 11:30 a.m., and ends around 12:30 p.m.[9] The Chair opens the meeting with a roll call, then an introduction that may include a brief overview of the outcome of the last meeting, or some other item of interest to the MPC to start the discussion. From time to time, the Chair will also point out if a committee member has separately submitted some information in advance intended to highlight an area of concern or interest to the MPC.

Members are assigned a rank according to the first letter of their last name. A die is then thrown and this determines which member is asked, at the outset, to express a "bias" concerning the direction of the next policy rate setting. This is referred to as the "straw poll" and it serves to motivate the deliberations that follow. At this stage, there has been no discussion of policy issues or any debate about the appropriateness of the present stance of monetary policy. Put simply, each member provides a numerical value, expressed in fractions of 25 basis points (bps), indicating whether their recommendation at the beginning of the meeting leans toward a rise or a fall in the upcoming policy rate setting. Ostensibly, the objective of the exercise is to give all participants an idea of where each member stands, prior to any group influence. The straw poll is not taken as a commitment but rather helps focus the discussion and provides an ex ante signal of the likely consensus, or lack thereof, inside the MPC.

Next, the meeting considers the latest economic forecasts and outlook presented by the professional economists on the MPC. Members are then encouraged to address questions to the professional economists about their outlook and views concerning the Canadian

[8]A copy of the agenda is reproduced in the appendix.

[9]Each meeting should be viewed as the conclusion of a process. In between each MPC meeting, there is an implicit expectation that each member, in their own manner, will prepare for the next meeting. Clearly, unlike their counterparts at the Bank (see below), MPC members likely devote relatively less time thinking about the future course of monetary policy.

economy.[10] Much of the remainder of the meeting is then devoted to a discussion during which all members express their views about what issue, domestic and foreign, is likely to influence their position about the current and future direction of the policy rate. The Chair ensures that all members' views are represented. In addition, he encourages everyone to query the views of others and to articulate the principal motivations in support of their recommendations. Generally, the focus of the discussion is on the next interest rate setting decision even though MPC members are subsequently also asked to take a stand on the policy rate settings up to one year into the future.

Just before the discussion ends, the Chair calls for each member's recommendations. Once again, a die is thrown.[11] Each MPC member must recommend the overnight rate setting the Bank should announce at (i) the upcoming meeting, (ii) the meeting thereafter, (iii) the meetings in six months' time, and (iv) the meeting in twelve months' time, keeping in mind the calendar of announcements that the Bank publishes well in advance (July of the previous year).

Each member is also permitted to make a very brief statement to emphasize the reasoning behind their recommendations and to offer advice that the Bank ought to consider. Once all the votes are cast, the Chair reviews the recommendations to ensure they were accurately recorded, announces the recommendations, and concludes the meeting. Members are asked to keep silent until the statement and individual recommendations are published on the CDHI's website at 2:00 p.m. While the statement explaining the outcome of the meeting is prepared by the CDHI, the text reflects the views of the committee as a whole, with any differences of opinion also being noted.

2.2 The Bank of Canada's Governing Council

The Bank of Canada's Governing Council publishes eight times a year a decision about the overnight interest rate target on

[10]It should be noted that several of the professional economists on the MPC also meet and regularly exchange forecasts and other views about the Canadian economy with Bank of Canada officials.

[11]It is only fairly recently that the identity of the first member to vote has been recorded. Hence, we are unable to control for any "first mover" type of effect in the empirical work reported below.

Wednesdays at 9:00 a.m. (see also footnote 7). The GC consists of six members, namely the Governor, the Senior Deputy-Governor, and four Deputy-Governors. The Governor and Senior Deputy-Governor are appointed by a Board, subject to government approval. The remaining Deputy-Governors are appointed by the Bank. The Bank also publishes, four times a year, a Monetary Policy Report, on the same day as the policy rate announcement. It contains the latest assessment of domestic and global economic conditions as well as the GC's projections, in particular, for inflation and real GDP growth.

A statement accompanies each GC decision that briefly outlines its views and outlook. Decisions are reached through consensus. The GC is a creation of former Governor Gordon Thiessen intended to enhance central bank transparency but has no basis in statute. Accordingly, no votes or minutes are released. It was also under Thiessen's governorship, in November 2000, that fixed announcement dates for the overnight rate were introduced. It was precisely this innovation in policymaking, together with inflation targeting, that inspired the creation of the CDHI's MPC.

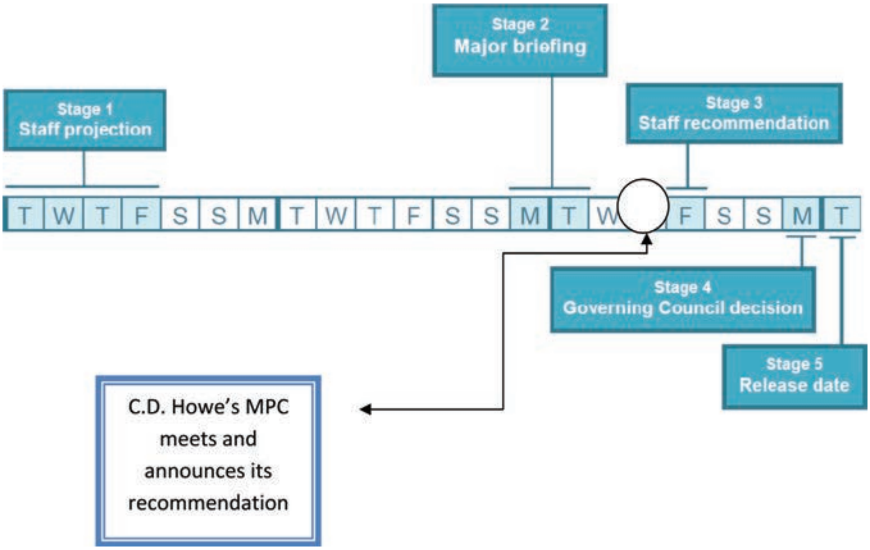
The Bank of Canada Act stipulates that monetary policy decisions are communicated by the Governor of the Bank, who is accountable for these decisions. Figure 1 reproduces the illustration used by Murray (2012) to explain the timing and stages in the Bank's decision-making process against the timing of the MPC's recommendations.

The staff recommendations are discussed by the Bank's Monetary Policy Committee—which consists of twelve to twenty senior officers—on the Friday prior to the decision. Every member of this MPC makes an individual recommendation for the policy rate. On the following Monday, the GC goes in camera to make the decision which is released on Wednesday.

In April 2009, the Bank made a commitment to keep the policy rate at the effective zero lower bound until the end of the second quarter of 2010, conditional on the outlook for inflation.[12] At the time, the Bank announced that "with monetary policy now

[12]Evaluations of this episode suggest that the Bank was effective in communicating the conditionality of the commitment (e.g., see He 2010, Siklos and Spence 2010).

Figure 1. Stages in the Bank of Canada's Decision-Making Process



Source: Murray (2012) and authors.

operating at the effective lower bound for the overnight policy rate, it is appropriate to provide more explicit guidance than is usual regarding its future path so as to influence rates at longer maturities. ... The Bank will continue to provide such guidance in its scheduled interest rate announcements as long as the overnight rate is at the effective lower bound."[13] The commitment was repeated each time the Bank set the overnight rate until it was withdrawn in April 2010, ahead of schedule. The target rate was actually raised at the next meeting in May 2010. This episode is considered to be the first example of an explicit form of forward guidance later adopted by other central banks. The original, calendar-based, form of forward guidance introduced by the Bank has since been replaced by state-dependent forms of forward guidance as in, for example, the variants adopted by the Federal Reserve and the Bank of England (e.g., see Filardo and Hofmann 2014).

[13]See http://www.bankofcanada.ca/2009/04/fad-press-release-2009-04-21.

3. Policy Rate Setting, Consensus, and Disagreement: Specifications and Econometric Issues

3.1 Monetary Policy Reaction Functions

A useful starting point to evaluate potential differences between the MPC and the GC is the Taylor rule (TR), though this rule does have its limitations due to the ultra-low interest rate environment of the past few years. The general specification is written as follows:

$$i_t^{\tau} = (1 - \rho_0)(\alpha r_t + \beta_0 \tilde{\pi}_{t+12|t} + \beta_1 \tilde{y}_{t+12|t}) + \rho_0 i_{t-1}^{CB} + \rho_1 i_t^{Fed} + \theta \boldsymbol{X}_t + \epsilon_t. \quad (1)$$

Equation (1), for the most part, incorporates a standard TR where the policy rate i_t^{τ} is set either by the MPC or the GC, together with an interest rate smoothing parameter ρ_0. Both committees must set the current-period policy rate according to the level set by the central bank in period $t-1$.[14] Given the potential role the United States plays in Canadian economic performance, we also control for the impact of the Federal Reserve's current target rate (ρ_1).[15] Following Clarida (2012), the real interest rate r_t is assumed to be time varying and is based on the real return yield on ten-year Canadian Government bonds.[16]

The determinants of the TR include an expected inflation gap $\tilde{\pi}_{t+12|t}$ defined as the twelve-month-ahead inflation forecast minus the stated inflation target of 2 percent. The headline consumer price index forecast by *The Economist* poll of forecasters is used to measure inflation expectations.[17] Furthermore, in proxying the

[14]If the Bank decides on a policy rate that differs from the MPC's recommended setting, the former is the starting point of discussions at the next MPC meeting.

[15]Like the Federal Reserve, the GC meets eight times a year. Usually, FOMC and GC meetings are held in the same month.

[16]Most estimates of the Taylor rule do not permit a time-varying real interest rate. However, in view of the events during the sample period (i.e., the end of the Great Moderation, the global financial crisis, and the experiment with forward guidance), it is unreasonable to assume that the "neutral" real rate is constant throughout.

[17]This seems appropriate, since the published forecasts are made by financial institutions. Several of their chief economists are, or have been, members of the MPC.

unobservable expected output gap $\tilde{y}_{t+12|t}$, we follow past practice by employing a Hodrick-Prescott (HP) (1997) filter with the standard smoothing parameter of 14,400 for the one-year-ahead real GDP growth forecast by *The Economist*.[18] We chose not to add an exchange rate variable.[19]

We also consider several other determinants of i_t^T, summarized by the vector $\boldsymbol{X}_t$. These are motivated by events during the global financial crisis and thereafter. First, the VIX volatility index[20] is employed as an indicator measuring financial market stress which is also monitored by the GC and the MPC. Second, the period of credit easing and quantitative easing in the United States is captured by a dummy variable. Third, and most significantly, we include a dummy variable set to 1 during the period of the conditional commitment, that is, when the Bank provided an explicit form of forward guidance (also see section 2.2).[21]

Equation (1) relies on data publicly available at the time of the MPC proposal and the GC decision. As a consequence, the specification is estimated via least squares.[22]

[18]We also used the Bank's own output-gap series. However, the specification employing forecasts by *The Economist* yields more plausible results, which might reflect the fact that the Bank's forecasts are not updated monthly. The same explanation holds for the Bank's quarterly inflation forecasts.

[19]Inclusion of a U.S. interest rate can be said to indirectly capture any exchange rate motive in policy rate setting. In addition, research on estimated as well as optimal TR (e.g., Clarida 2001, Collins and Siklos 2004) suggests that adding this series does not make much difference to inferences based on the standard or extended TR specifications.

[20]The VIX is the Canadian version of the well-known U.S. VIX indicator. The indicator estimates the thirty-day volatility of the Toronto Stock Exchange that is implied by the Toronto Stock Exchange index options.

[21]As noted in the Introduction, there are claims that persistent deviations from TR-based recommendations may be partly explained by the failure of some central banks to respond to excessive credit growth. Accordingly, we also considered growth rates in two indicators of credit, namely consumer credit and residential mortgage credit. The results are relegated to the appendix.

[22]A referee has suggested that the time-series properties of some of the data may pose problems. In particular, the policy rates (Canada and United States) and inflation may display some evidence of non-stationarity. Conventional unit-root tests suggest that the series are either stationary or non-stationary, depending on the inclusion of a trend or the lag length in the test equation. For this reason, we supplement our results with R^2 values for non-stationary series. See Harvey (1989).

3.2 Consensus within the MPC and Disagreement with the GC

A principal concern in this paper is the information content of individual policy rate recommendations of MPC members. Individual behavior within the MPC effectively amounts to asking whether one can detect any evidence of the concerns sometimes raised about how committee members interact with each other, i.e., "free riding," "groupthink," or the risk of "information cascades" and whether committee decisions are necessarily superior to other forms of decision making (e.g., Mahadeva and Sterne 2000, Morris and Lybek 2004, Sibert 2006, Visser and Swank 2007, Maier 2010, Swank and Visser 2013).

Similarly, disagreements between the MPC and the GC raise a host of questions about the extent to which the two groups see eye to eye and whether the respective composition, size, timing of decisions, and possibly differences in information sets can account for different policy rate recommendations. Clearly, some of the potential determinants of consensus and disagreement are unobservable. Hence, any specification considered is only able to partially deal with the complexities of committees and policy rate setting behavior. The following two specifications are estimated:

$$\begin{aligned} \mathit{Consensus}_t &= \delta_0 + \delta_1 \mathit{Consensus}_{t-1} + \delta_2 \boldsymbol{Y}_t \\ &\quad + \delta_3 \mathit{Disagreement}_{t-1} + \mu_t \qquad (2) \\ \mathit{Agreement}_t^* &= \Pr[i_t^{SC} - i_t^{CB} = 0 | Z_t] = \kappa_0 + \kappa_1 \boldsymbol{Y}_t \\ &\quad + \kappa_1 \mathit{Consensus}_t + \eta_t \qquad (3) \\ \boldsymbol{Y}_t &= \{\mathit{Pros}_t, \mathit{In\text{-}Person}_t, \mathit{Cond.Commit.}_t, \mathit{Cut}_t, \mathit{Rise}_t, \\ &\quad \mathit{VIX}_t, \mathit{Infl.\,Volatility}_t\}. \qquad (4) \end{aligned}$$

Equation (2) considers sources of consensus in committee decisions. We rely on two metrics to measure consensus. First, we rely on the fraction of MPC members who vote in favor of the committee's proposal (*Cons. Share*). Second, we rely on the simple standard deviation of proposals as an alternative since this gives a measure of the range of recommendations (*SD (Ind. Prop.)*).

As explanatory variables, we first employ the respective lagged consensus variable to test for persistence in the degree of consensus within the MPC. Second, we consider the representation of professionals on the committee and the share of members attending the meeting in person. Third, we control for the conditional commitment period since, during this period, interest rate changes were potentially put on hold, subject to the Bank's inflation outlook.[23] Fourth, the specification reflects whether consensus is asymmetric, that is, whether there are differences between a rate rise or a fall. Fifth, we proxy macroeconomic uncertainty by the VIX volatility index and the conditional volatility of inflation obtained by estimating a GARCH(1,1) model. Finally, we test if past disagreements in policy recommendations between the MPC and the GC influences consensus within the MPC.[24] Equation (2) is estimated via OLS.[25]

Equation (3) transforms non-zero differences between the actual GC decision and the MPC recommendation into a binary variable (0 = disagreement, 1 = agreement). The purpose is to obtain, via probit estimation, estimates of the probability of agreement between the shadow and formal monetary policy committees, conditional on the set of observables in (4) and the current degree of consensus within the MPC.

4. Stylized Facts and Empirical Results

4.1 Stylized Facts

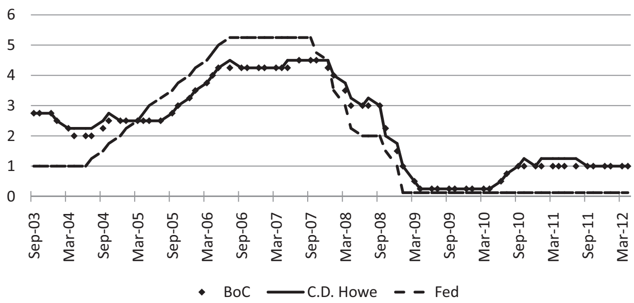
Figure 2 plots the GC's policy rate, the MPC's recommendation, and the federal funds rate. It clearly reveals that there have been persistent differences between U.S. and Canadian policy rates. Until early 2005 the Bank's overnight rate was usually higher than its

[23]As previously noted (see section 2), MPC members are not aware of the contents of the Bank's upcoming projections. These are published in the Bank's Monetary Policy Report released on the same day as the policy rate announcement is made.

[24]We also estimate a version of (2) where consensus in the actual poll is a function of MPC members' consensus in the straw poll. Put differently, adding this variable could be construed as reflecting the "distance" between an initial and a final position.

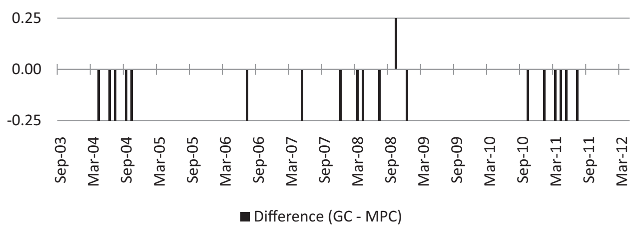
[25]We do not report the results for consensus in the straw poll as a left-hand-side variable since participants are forced to give a 25 bps interval rather than a single value. As suggested by one of the referees, this adds a lot of noise to each participant's outlook going into the meeting and leads to mostly insignificant coefficients for the explanatory variables. However, the results are relegated to the appendix for interested readers.

Figure 2. MPC Recommendation, GC Target Rate, and Federal Funds Target Rate



Source: CDHI, Bank of Canada, and FRED II (Federal Reserve Bank of St. Louis). The vertical axis is in percent.

Figure 3. Differences between the GC Target Rate and the MPC Recommendation

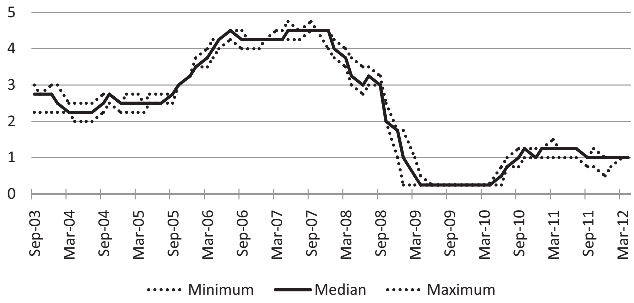


Source: Authors' calculations. The vertical axis is in fractions of 1 percent (i.e., basis points).

U.S. counterpart. A reversal takes place until the global financial crisis hits the U.S. economy in late 2007. Since then, the federal funds rate remains below the Bank's policy rate. Figure 2 also suggests that while the MPC and the GC are not too far apart in their views about the appropriate setting for the policy rate, differences do emerge and can remain persistent for some time.

Figure 3 plots the differences between the GC policy rate and the median MPC proposal. The MPC has tended to almost always recommend a policy rate that is higher than the one actually set by the GC in the case of disagreement between both committees. The only exception is in October 2008, right after the internationally coordinated interest rate cut of 50 bps. At the time, there were

Figure 4. Minimum, Median, and Maximum Policy Rate Recommendations from MPC Members



Source: CDHI and authors' calculations. The vertical axis is in percent.

several announcements from the United States aimed at stemming the impact of the worsening financial crisis. This may have influenced the MPC to recommend a 25 bps lower policy rate target than the GC later adopted.

Figure 4 provides an indication of the range of policy rates based on the individual recommendations of MPC members. It is interesting that gaps between the most hawkish and most dovish recommendations persist over time except during the period of the conditional commitment and, more recently, when central banks, including the Bank of Canada, have emphasized that policy rates are likely to remain low for some time.

4.2 Monetary Policy Reaction Functions

Next, we compare the two committees through the device of the TR. Table 1 presents estimates for the MPC (top panel) and the GC (bottom panel) for several variants of equation (1).[26]

[26]In addition to the determinants previously discussed, we also considered (i) changes in the TED spread (i.e., the change in the difference between the yield on three-month Canadian prime corporate papers and Treasury bills of the same maturity), (ii) interaction terms wherein the proportion of professionals in the shadow committee affects the reaction to inflation and output forecasts, and (iii) the bias of the MPC based on future policy rate recommendations. The latter two variations appear to play no statistically significant role in explaining the policy rate recommendations. A rise in the TED spread has a significantly positive effect on the recommendation. These results are relegated to the appendix.

Table 1. Taylor-Rule Estimates for the CDHI's MPC and the Bank of Canada's GC

	(1)	(2)	(3)	(4)
CDHI's MPC				
Lag BoC Target Rate	0.837***	0.877***	0.833***	0.790***
Lag Fed Target Rate	0.091***	0.052**	0.070**	0.115***
Real Bond Yield	1.438***	2.366***	1.581***	1.479***
Inflation Forecast Gap	1.177*	1.070*	0.800	−0.392
GDP Forecast Gap	1.192**	0.515	0.804**	0.508
VIX-20 Index		−0.008***		
Credit Easing/ Quantitative Easing			−0.146**	
Cond. Commitment				−0.286**
S.E. of Regression	0.180	0.158	0.169	0.174
Observations	105	105	105	105
R^2	0.985	0.988	0.987	0.986
Harvey's R^2	0.202	0.397	0.303	0.261
Bank of Canada's GC				
Lag BoC Target Rate	0.808***	0.836***	0.805***	0.787***
Lag Fed Target Rate	0.122***	0.095***	0.103***	0.133***
Real Bond Yield	1.097***	1.773***	1.321***	1.163***
Inflation Forecast Gap	1.131*	1.053*	0.777	0.304
GDP Forecast Gap	0.836**	0.339	0.549**	0.520
VIX-20 Index		−0.006***		
Credit Easing/ Quantitative Easing			−0.135***	
Cond. Commitment				−0.128*
S.E. of Regression	0.149	0.135	0.137	0.148
Observations	105	105	105	105
R^2	0.989	0.991	0.991	0.990
Harvey's R^2	0.334	0.455	0.441	0.348

Notes: Estimates are for equation (1). OLS is used. ***/**/* indicates significance at the 1 percent/5 percent/10 percent level. Newey-West (1987) standard errors are used. The table reports estimates for the steady-state coefficients for α (real bond yield), β_0 (inflation gap), and β_1 (output gap). The inflation and output forecast gaps are based on one-year-ahead forecasts of inflation and real GDP growth from *The Economist*. An HP filter is used to estimate the gap with the standard monthly smoothing filter.

Focusing on the steady-state parameter estimates, we find that the (time-varying) real interest rate of the MPC is consistently higher than the one implicit in the GC's TR.[27] Next, the MPC and GC are equally dovish in response to inflation surprises,[28] as the steady-state estimate of the inflation parameter is well below the 1.5 recommended by Taylor as reflecting best practice in maintaining inflation on "target."[29] Finally, the MPC is considerably more responsive to the output gap than its GC counterpart.

If the specifications are conditioned on periods when "unconventional" policies are in place, these clearly replace inflation as the primary driver of lower policy rate recommendations. In particular, the Bank's conditional commitment has a significantly larger negative effect on the MPC's recommendation than on the policy rate set by the GC. This might imply that the credibility of the Bank's conditional commitment has convinced the MPC to pause proposing interest rate changes as well. Alternatively, members of the MPC may have felt that, once announced, it was very important for the Bank to honor its conditional commitment, irrespective of whether they agreed with it or not. This would also partly explain the disappearance of the gap for the most hawkish and most dovish MPC recommendation shown in figure 4.

As previously stressed, possibly the most notable difference between the MPC and the GC is that the individual recommendations of the shadow committee are published. Consequently, we are also able to investigate if there are differences in the reaction functions across MPC members.

The first column of table 2 reproduces the baseline TR estimate found in table 1. Next, we examine the recommendations of two MPC members who, based on the historical experience, have consistently been either below the MPC recommendation ("dovish") or above the median recommendation ("hawkish"). As seen from

[27]Note that this and the following conclusions are based on the comparison of point estimates unless otherwise stated.

[28]Statistical testing confirms that there are no significant differences in the MPC's and GC's reaction to inflation.

[29]Wald tests confirm that the inflation coefficient is statistically different from 1.5 for specification (4) in the case of the MPC (at the 1 percent significance level) and specifications (3) and (4) in the case of the GC (at the 5 and 1 percent significance levels, respectively).

Table 2. Taylor-Rule Estimates for the MPC Dove and Hawk

	(1)	Dove	Hawk
Lag BoC Target Rate	0.837***	0.718***	0.918***
Lag Fed Target Rate	0.091***	0.169**	0.035
Real Bond Yield	1.438***	0.993**	3.024***
Inflation Forecast Gap	1.177*	−0.255	3.586**
GDP Forecast Gap	1.192**	1.517*	0.833
Average Diff. to Median Proposal		−0.091	0.152
S.E. of Regression	0.180	0.334	0.196
Observations	105	69	46
R^2	0.985	0.949	0.983
Harvey's R^2	0.202	−0.384	0.699

Notes: Estimates are for equation (1). OLS is used. ***/**/* indicates significance at the 1 percent/5 percent/10 percent level. Newey-West (1987) standard errors are used. The table reports estimates for the steady-state coefficients for α (real bond yield), β_0 (inflation gap), and β_1 (output gap). The dove (hawk) is a time series of the recommendations of a member who consistently votes for a policy rate below (above) the recommendation of the MPC. The inflation and output forecast gaps are based on one-year-ahead forecasts of inflation and real GDP growth from *The Economist*. An HP filter is used to estimate the gap with the standard monthly smoothing filter.

table 2, the dove pays less attention to the GC's previous setting, while the opposite is true for the hawk on the committee. Further, the dove seems unresponsive to inflation, while the hawk not only responds positively to inflation shocks but does so by a magnitude that exceeds the Taylor recommendation for good practice in monetary policy by a wide margin. In contrast, the dove responds strongly to output fluctuations, while the hawk is unresponsive to real shocks emanating from GDP forecasts. Finally, the dove is clearly distinguished from the hawk based on the estimated steady-state real interest rate, with the dove's estimate being a third lower than the hawk's estimate. This suggests considerable diversity of opinion inside the MPC. Interestingly, the professional/academic distinction appears to matter less than the difference between hawks versus doves when viewed through the lens of the TR (see the appendix posted on the Central Bank Communication network website, http://www.central-bank-communication.net/).

Table 3. Consensus within the CDHI's MPC

Dependent Variable	Cons. Share	SD (Ind. Prop.)	Cons. Share	SD (Ind. Prop.)
Constant	0.516***	0.066**	0.450**	0.048
Lagged Dependent Variable	0.311*	0.283**	0.280	0.206
Consensus Straw Poll	–	–	0.071	0.344*
Professional Share	0.282	−0.056	0.383**	−0.093
In-Person Voting Share	−0.124	−0.043	−0.120	−0.017
Cond. Commitment	0.142**	−0.086***	0.117*	−0.072***
Lag (Diff (MPC – GC))	−0.232	0.141*	−0.215	0.142*
Proposal: Cut	−0.152**	0.049**	−0.149**	0.040*
Proposal: Hike	−0.094*	0.012	−0.107**	0.018
VIX Index	−0.001	0.002**	−0.001	0.001
Cond. Infl. Volatility	0.002	−0.001	−0.013	0.009
Observations	68	68	61	61
S.E. of Regression	0.145	0.049	0.146	0.048
R^2	0.547	0.649	0.520	0.650

Notes: Estimates are for equation (2). OLS is used. ***/**/* indicates significance at the 1 percent/5 percent/10 percent level. White (1980) standard errors are used.

4.3 Consensus within the MPC and Disagreement with the GC

Table 3 displays estimates of various specifications for equation (2). The results apply only to the MPC since the Bank does not release information about the individual positions taken by GC members.

Focusing on the results in the first and second column, consensus (or the lack thereof) is persistent, as the first lag is statistically significant for both indicators, that is, the share of votes in favor of the MPC proposal and the standard deviation of individual proposals. Second, policy rate cuts and hikes yield less consensus.[30] Third, past disagreement between the GC and the MPC also significantly reduces consensus in the MPC's current meeting. When overall

[30]Note that a positive sign indicates greater consensus in the case of the indicator based on the share of votes in favor of the proposal, whereas in the case of the standard deviation indicator, a negative sign implies more consensus, i.e., a decrease in the standard deviation.

economic signals are uncertain, leading to disagreement between both committees, this is also reflected in a wider variety of views within the MPC. Fourth, in line with the previous interpretation, higher stock market volatility leads to less consensus inside the MPC. Finally, and perhaps most interestingly in terms of whether certain forms of central bank communication can help anchor expectations, the Bank's conditional commitment had a statistically significant impact on the degree of consensus inside the MPC. It raises the share of members who support the recommendation and reduces the variability of policy rate proposals.[31]

Next, we turn our attention to investigating the determinants of the agreement between the MPC and the GC. The results are shown in table 4.

When consensus in the MPC is higher, this increases the likelihood that both committees agree on the preferred level of interest rates. In addition, the period of the conditional commitment contributed to raising the agreement between the MPC and the GC as well. Hence, differences between the respective stances taken by the GC and the MPC dissipated during the roughly one-year period when the conditional commitment was in place. Finally, allowing for asymmetry between rate hikes and cuts proves to be important. While proposed rate hikes significantly increase differences of opinion between the MPC and the GC, rate cuts have no significant impact on disagreement. This may well reflect the relatively greater degree of hawkishness of the MPC's median member compared with the GC's position.

4.4 Counterfactual Analysis

In view of the influence of the conditional commitment on the MPC's median recommendation, we felt it was useful to investigate its

[31] An interesting finding is that, when we add the degree of consensus in the straw poll as an additional determinant (third and fourth columns), the variable is significant only for one of the two indicators, the standard deviation of individual proposals (and only at the 10 percent level). Thus, there is no robust relationship between consensus in the straw poll and consensus in the actual poll. Whether this difference reflects the impact of groupthink or conveys the effect that committee discussions have on the outcome of individual decision makers is unclear. Nevertheless, the outcome suggests that there may be interesting information content in the differences between the actual published polls relative to the straw poll taken.

Table 4. Determinants of Disagreement between the MPC and the GC

Consensus Indicator	Cons. Share 1.162***	SD (Ind. Prop.) −1.833**
Professional Share	−0.256	0.041
In-Person Voting Share	−0.120	−0.443*
Cond. Commitment	0.730***	1.109***
Proposal: Cut	−0.088	−0.125
Proposal: Hike	−0.264***	−0.371***
VIX-20 Index	−0.001	0.001
Cond. Infl. Volatility	−0.140	−0.281*
Observations	69	69
LR Statistic	321.02***	315.07***
Pseudo Log-Likelihood	−19.128	−25.248
Pseudo R^2	0.529	0.378

Notes: The dependent variable is a dummy variable measuring if the BoC's GC was following the CDHI's MPC proposal or not (1 = Yes, 0 = No). The table shows average marginal effects for probit estimations of equation (3). ***/**/* indicates significance at the 1 percent/5 percent/10 percent level. Huber (1967)/White (1980) robust standard errors are used.

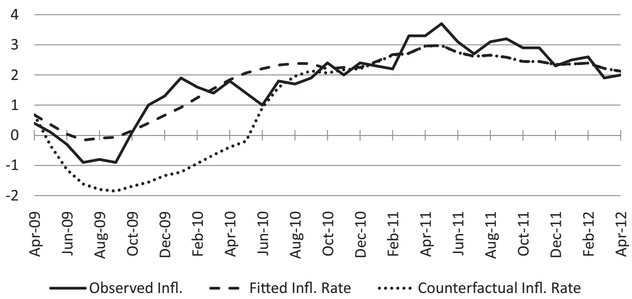
impact on the shadow committee's views about inflationary developments. Consequently, we consider a counterfactual experiment which asks the following: supposing that the conditional commitment essentially represented the GC's attempt to temporarily set the neutral real rate to zero,[32] what inflation path for the median MPC member would have been consistent with such an outcome?

For that purpose, we estimate a simple autoregressive distributed lag (ADL) model for inflation which includes lags of inflation, the output gap, the GC's target rate, and the real interest rate.[33] The solid line in figure 5 shows the observed inflation rate. The dashed line represents the fitted values from the ADL model with a time-varying real interest rate and relying on the GC's policy rate. The dashed line shows the counterfactual path for inflation when two

[32]Presumably, as a device to communicate, as clearly as possible, a commitment to ease policy for an extended period of time.

[33]Lag length is determined based on a joint significance test of each additional lag.

Figure 5. Inflation Rates: Observed, Fitted, and Counterfactual



Notes: The vertical axis is in percent. The solid line is CPI inflation. The dashed line is the MPC's inflation path derived from the ADL model described in the text. The dotted line is the MPC's inflation path, assuming that the neutral real interest rate is set to zero and the MPC's median policy rate recommendation is followed but only during the period of the conditional commitment (April 2009–April 2010). The model is estimated over the period July 2003–April 2012.

variables from the ADL model have been replaced with their counterfactual values. This means that the neutral real interest rate is restricted to zero and the MPC's median recommendation replaces the GC's policy rate setting during the period April 2009–April 2010.

Figure 5 shows what inflation path the median MPC member had in mind under the counterfactual against observed inflation. The counterfactual effectively implies a temporary deflation of almost 2 percent by October 2009, thereafter rising sharply in early 2010 to around the Bank's inflation target of 2 percent by April 2010, when the commitment was removed by the Bank. Thus, the Bank's conditional commitment had the desired impact, and the MPC effectively treated the episode as leading back to the inflation target by the time the commitment ended.[34] This reinforces the potential impact of forward guidance in monetary policy. Whether guidance beyond the next interest rate decision is, in fact, a successful device is not immediately clear, however. After all, the timing of the

[34]Standard-error bands (not shown) confirm that the counterfactual inflation path was statistically significantly lower throughout the December 2009–April 2010 period. Hence, it took some time for the MPC to adjust its view of the likely inflation path.

conditional commitment, arguably introduced at the height of the crisis, may also reflect a response by members of the MPC, and the Bank, to the unfolding of events in financial markets as well as to the ongoing poor economic environment.

5. Conclusions and Policy Implications

In this paper, we examine the policy rate recommendations of the Bank of Canada's Governing Council relative to its shadow, the C.D. Howe Institute's Monetary Policy Council. In addition, we examine what determines the likelihood of consensus inside the MPC. Finally, we consider several factors in explaining disagreement between the GC and the MPC. The sample consists of seventy interest rate decisions between September 2003 and April 2012 and covers the Bank's conditional commitment to keep the policy rate at the zero lower bound. Our principal conclusions are as follows.

First, differences between both committees are small but persistent. In the case of disagreement between the committees, the MPC has tended to almost always recommend a policy rate that is higher than the one actually set by the GC. These differences are partly driven by the fact that the MPC assumes a higher steady-state real interest rate than its GC counterpart. In contrast, there are few differences in both committees' responses to inflation and output shocks. Finally, the period of the Bank's conditional commitment had a measurable impact on the MPC's views about future inflationary developments. This result is supported by a counterfactual experiment which finds that the median MPC member's inflation path decreases quickly once the conditional commitment policy is announced.

While comparisons between the two committees can yield useful insights, there are interesting additional results to be gained by investigating individual recommendations from the MPC since these are observable, while those of its counterparts in the GC are not. There is relatively less consensus inside the MPC when rates are rising or falling than when they remain unchanged. Differences between the MPC's and the GC's recommendations are least notable when there is consensus inside the MPC.

It appears that we can learn something from the comparison of a shadow committee and its counterpart formally responsible for the conduct of monetary policy in Canada. Recently, there has been a resurgence in outsiders expressing independent opinions about the appropriate stance of monetary policy. This kind of development is helpful to central banks who seek to be transparent and accountable. Transparency and accountability also demand that central banks explain their actions and provide sufficient information to enable the public to understand why certain policy decisions are made. One way of independently assessing the value of monetary authorities' decisions is simply to provide the public with a second opinion, which is the MPC's task.

Our findings also suggest some policy implications. In spite of the variety of views and backgrounds of the members of the MPC and the GC, it is likely that similarity in training and experience contribute to narrowing the differences between the two monetary policy committees. In addition, there is the over-arching constraint imposed by inflation targeting and the transparency associated with such a regime. If the policy regime faced by the MPC was different, there may well be more disagreement about the appropriate policy stance.[35]

Conditional commitments, or forward guidance, of the kind the Bank of Canada introduced in 2009 influence the views of those who shadow monetary policy. Nevertheless, it remains to be seen whether calendar-driven forms of forward guidance are superior to threshold-based attempts, such as the variants introduced by the Federal Reserve or the Bank of England. While such forms of policymaking may be temporary, they ought to be used sparingly and in crisis conditions only since the risk of reputational loss may well prove significant and the task of communicating such commitments is difficult.

Finally, diversity within the MPC does not appear to pose any particular difficulties in setting the appropriate stance of monetary policy, nor do differences in views appear so large as to threaten the ability of diverse opinions to provide useful informed opinion about monetary policy actions. Inflation targeting, together with adequate transparency and accountability, provides the necessary constraint to ensure that there is some value in airing differences in opinion.

[35] Neuenkirch and Siklos (2013) provide such evidence in the case of the European Central Bank.

Consequently, consideration should be given to formally recognizing and defining the responsibilities of the Governing Council. Explicit recognition of the committee structure as a means of delivering monetary policy has spread around the world, and there is no reason why this approach should not be followed in Canada.

References

Bernanke, B. 2010. “Monetary Policy and the Housing Bubble.” Remarks delivered at the Annual Meeting of the American Economic Association, Atlanta, GA, January 3. Available at http://www.federalreserve.gov/newsevents/speech/bernanke20100103a.htm.

Clarida, R. 2001. “The Empirics of Monetary Policy Rules in Open Economies.” *International Journal of Finance and Economics* 6 (4): 315–23.

———. 2012. “What Has—and Has Not—Been Learned about Monetary Policy in a Low Inflation Environment? A Review of the 2000s.” *Journal of Money, Credit and Banking* 44 (s1): 123–40.

Collins, S., and P. Siklos. 2004. “Optimal Monetary Policies and Inflation Targets: Are Australia, Canada and New Zealand Different from the US?” *Open Economies Review* 15 (4): 347–62.

Davies, G. 2011. “The Timing of the Fed’s Exit.” *Financial Times,* March 15. Available at http://blogs.ft.com/gavyndavies/2011/03/15/the-timing-of-the-feds-exit/#axzz25tXCNoKD.

Filardo, A., and B. Hofmann. 2014. “Forward Guidance at the Zero Lower Bound.” *BIS Quarterly Review* (March): 37–53.

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

He, Z. 2010. “Evaluating the Effect of the Bank of Canada’s Conditional Commitment Policy.” Discussion Paper No. 2010-11, Bank of Canada.

Hodrick, R., and E. Prescott. 1997. “Postwar US Business Cycles: An Empirical Investigation.” *Journal of Money, Credit and Banking* 29 (1): 1–16.

Huber, P. 1967. “The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions.” In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, ed. L. M. Le Cam and J. Neyman, 221–33. Berkeley, CA: University of California Press.

Mahadeva, L., and G. Sterne. 2000. *Monetary Policy Frameworks in a Global Context.* London: Routledge.

Maier, P. 2010. "How Central Banks Take Decisions: An Analysis of Monetary Policy Meetings." In *Challenges in Central Banking,* ed. P. Siklos, M. Bohl, and M. Wohar, 320–56. Cambridge University Press.

Morris, J. A., and T. Lybek. 2004. "Central Bank Governance: A Survey of Boards and Management." IMF Working Paper No. 04/226.

Murray, J. 2012. "Monetary Policy Decision-Making at the Bank of Canada." Remarks delivered to the Mortgage Brokers Association of British Columbia, Vancouver, May 7. Available at http://www.bankofcanada.ca/2012/05/speeches/monetary-policy-decision-making/.

Neuenkirch, M., and P. Siklos. 2013. "What's in a Second Opinion? Shadowing the ECB and the Bank of England." *European Journal of Political Economy* 32 (December): 135–48.

———. 2014. "When Is Lift-off? Evaluating Forward Guidance from the Shadow." *Open Economies Review* 25 (5): 819–39.

Newey, W., and K. West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–8.

Poole, W. 2006. "The Fed's Monetary Policy Rule." *Review* (Federal Reserve Bank of St. Louis) 88 (1): 1–11.

Sibert, A. 2006. "Central Banking by Committee." *International Finance* 9 (2): 145–68.

Siklos, P., and A. Spence. 2010. "Faceoff: Should the Bank of Canada Release Its Projections of the Interest Rate Path? The Cases For and Against." *C.D. Howe Institute Backgrounder* 134.

Swank, O., and B. Visser. 2013. "Is Transparency to No Avail?" *Scandinavian Journal of Economics* 115 (4): 967–94.

Taylor, J. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy* 39 (1): 195–214.

———. 2007. "Housing and Monetary Policy." NBER Working Paper No. 13682.

———. 2013. "International Monetary Coordination and the Great Deviation." *Journal of Policy Modeling* 35 (3): 463–72.

Visser, B., and O. Swank. 2007. "On Committees of Experts." *Quarterly Journal of Economics* 122 (1): 337–72.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38.

Large Excess Reserves in the United States: A View from the Cross-Section of Banks*

Huberto M. Ennis and Alexander L. Wolman
Research Department, Federal Reserve Bank of Richmond

Bank reserves in the United States increased dramatically at the end of 2008. Subsequent asset purchase programs in 2009 and 2011 more than doubled the quantity of reserves outstanding. We study the cross-sectional distribution of reserves in that period, and the relationship between holdings of reserves and other components of banks' balance sheets. We find that reserves were widely distributed, increasing the liquidity position of many banks which, at the same time, were far from facing tight capital constraints. Our findings have implications for assessing the importance of large quantities of excess reserves for monetary policy.

JEL Codes: G21, E44, E58.

1. Introduction

In the final months of 2008, the quantity of bank reserves in the United States increased by almost a factor of twenty, to more than $850 billion. During the second half of 2009 reserves increased again

*We would like to thank Carl Walsh (the editor) and two referees, as well as Doug Diamond, Marvin Goodfriend, Todd Keister, Ellis Tallman, Steve Williamson, and the discussants of the paper, Morten Bech, Jagjit Chadha, Oreste Tristani, and Larry Wall, for their comments and suggestions. We would also like to thank participants at the Bob King Conference, the Swiss National Bank conference on Policy Challenges and Developments in Monetary Economics, the ECB conference on "The Post-Crisis Design of the Operational Framework for the Implementation of Monetary Policy," the Federal Reserve's 2011 Day Ahead Conference and at seminars at the Bundesbank, the Federal Reserve Bank of New York, the Federal Reserve Board, and the Swedish Riksbank for helpful comments, and Tim Hursey and Sam Marshall for excellent research assistance. The views in this paper do not represent the views of the Federal Reserve Bank of Richmond, the Board of Governors of the Federal Reserve, or the Federal Reserve System. Authors' e-mail addresses: huberto.ennis@rich.frb.org; alexander.wolman@rich.frb.org.

by approximately $300 billion, and in the first half of 2011 another significant increase in reserves took the level outstanding to approximately $1.6 trillion. These increases in reserves were the result of Federal Reserve policies aimed at mitigating the financial crisis and stabilizing the economy.

On October 1, 2008, the Federal Reserve began paying interest on reserves and, to maintain its interest rate target, no longer had to sterilize the creation of reserves associated with the various credit programs in place at the time. In late November 2008, the Federal Reserve also announced an asset purchase program that would increase its holdings of agency debt and mortgage-backed securities (MBS) by a total of $600 billion. As its lending programs started to wind down, the Federal Reserve increased the size of its planned asset purchases. The expanded program involved the purchase of $300 billion in Treasury securities beginning in March 2009 and a significant increase in the total purchases of agency debt and MBS. This program ended in March 2010. After a period of relative inactivity, the Federal Reserve undertook a new round of Treasury purchases between December 2010 and June 2011, for a total of $600 billion.[1]

These facts about the origin and evolution of aggregate bank reserves in the United States are well known. Much less attention has been devoted to the distribution of reserves across banks and to the interaction of reserves with other components of banks' balance sheets at the aggregate and individual levels. Studying these issues is the objective of this paper. We use cross-sectional data on banks' balance sheets from the regulatory filings commonly referred to as the Call Reports and other complementary data sources that provide aggregated information. For the period between mid-2008 and mid-2011, we document the relationship between reserve holdings and relevant bank-level measures of liquidity, capital, lending capacity, and lending opportunities.[2]

[1]Currently (as of April 2014), the Federal Reserve is engaged in an open-ended asset purchase program that entails acquiring $40 billion of agency mortgage-backed securities per month (starting in September 2012) and $45 billion of long-term Treasury securities per month (starting in January 2013). Since this current program is still under way and is open ended, it is not included in our study.

[2]In a recent related paper, Chang, Contessi, and Francis (2013) use regression analysis to study the cash holdings (net of required reserves) of all reporting

Why is studying the cross-sectional distribution of bank reserves interesting? We think that bank holdings of excess reserves can matter, under certain conditions, for the effect of monetary policy on the real economy (Ennis and Wolman 2010). In particular, large quantities of excess reserves make a strengthening economy potentially more sensitive to delays in (ex post) proper adjustments of monetary policy.

Our working hypothesis has points of contact with the bank lending view of monetary transmission (see Kashyap and Stein 1994). First, we proceed under the assumption that banks prefer to fund their lending activities with deposits instead of large (brokered) CDs or short-term interbank loans and that it takes time for banks to increase their deposit base. Second, we adhere to the view that some borrowers cannot find perfect substitutes for bank loans, especially in the short run.

Under these premises, a banking system that is holding a larger quantity of excess reserves is able to expand lending faster in response to changes in economic conditions (such as changes in real expected rate-of-return differentials). A way to understand this process is to think that holding reserves is a way for banks to "store" deposits that could eventually be used to finance lending. A banking system with a higher amount of "stored deposits" (i.e., excess reserves) can adjust lending more quickly than one that has already used most of its deposits to finance existing loans.

Of course, by paying interest on reserves (IOR), a policymaker could control how much banks want to move away from excess reserves and into loans (Dudley 2011). Higher IOR, everything else constant, will induce banks to refrain from making marginal loans with risk-adjusted rates of return comparable to the one being offered on reserves. However, determining when to adjust monetary policy is a difficult issue, and the possibility of policy being (on occasion) behind the curve is a real one (Plosser 2011 and Levin and Taylor 2013). When policy delays do occur, then the level of reserves and the distribution of reserves across banks can matter.

Aside from the availability of funding, the propensity of banks to expand lending quickly depends on other factors, such as the

institutions below the twenty largest, from the third quarter of 2008 to the second quarter of 2010.

state of their balance sheets.[3] In this paper, we study the cross-section of U.S. banks' balance sheets in an effort to better understand the general financial condition of the banks holding (most of) the reserves.

Five main findings arise from our analysis. First, reserves were widely distributed across the banking system, with both domestic and foreign institutions holding significant portions of the outstanding reserves. Second, reserves did not substitute for other liquid assets on banks' balance sheets: as total reserves increased, so did the banking system's total liquidity. Third, using data on leverage ratios, we conclude that the increase in reserves did not appear to constrain banks' ability to expand their balance sheets. Fourth, a significant proportion of reserves were held by banks apparently not restricted by capital requirements. Finally, there is little evidence that idiosyncratic changes in the return on lending induced banks to adjust their reserve holdings. We stated earlier that a large quantity of excess reserves gives a bank the potential to expand lending quickly. Of course, from 2008 through 2011, a rapid lending expansion did not materialize. These five findings seem to suggest that the lending behavior of banks during this period more likely reflected poor lending opportunities than limits induced by their financial conditions.

The paper is organized as follows. In the next section we discuss details of the data we use. After that, sections 3–7 each discuss one of the five main findings from our work. Section 8 concludes.

2. The Data

Our panel data covers commercial banks, savings banks, and trust companies, as well as uninsured branches and agencies of foreign banks. These institutions file quarterly supervisory reports, Call Reports, which are our primary data source. Our study does not cover credit unions and some other thrift institutions that were not required to report reserve holdings in their regulatory filings during the period of our study.

[3]The classic example of this interaction between different components of banks' balance sheets is the limitations that bank capital can impose on lending (Van den Heuvel 2002). But, as we discuss in later sections, there are others.

In the second quarter of 2011, there were 2,366 reporting institutions with reserve accounts, 237 of which were classified as uninsured.[4] Most of the uninsured institutions are affiliates of foreign banks, and in the second quarter of 2011 they held 98 percent of the reserves held by uninsured institutions.[5] For this reason, we refer to the uninsured group as FBOs (foreign banking organizations) for short, and we call the rest of the (insured) institutions domestic banks.

FBOs can hold accounts with Federal Reserve Banks, and thus reserves. However, since they do not hold insured deposits, they file a Call Report form, FFIEC 002, that is somewhat less detailed than the form FFIEC 031 or 041 filed by insured domestic institutions. For domestic institutions, we will aggregate banks up to the bank holding company level; decisions about reserve holdings presumably are made in the interest of the owners of the holding company.[6]

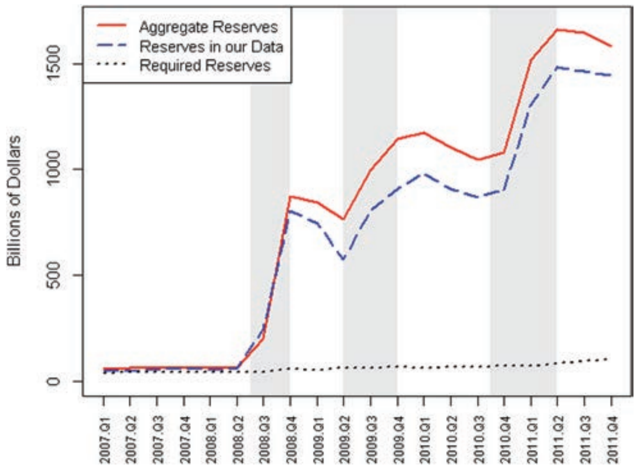
To complement the Call Reports when we discuss aggregate data for the banking system, we use the Federal Reserve's H.3 and H.4.1 statistical releases, which report aggregate reserves and other aspects of the Federal Reserve's balance sheet. This aggregate data released by the Federal Reserve is not derived from the Call Reports. Therefore, judgment is required when relating such data to

[4]In the second quarter of 2011, there were 7,147 reporting institutions in our data, 6,814 of which were insured. The 4,781 institutions that do not have reserve accounts are small banks that hold their reserves with correspondent banks. The correspondents report their respondents' reserves together with their own reserves. Thus, we have a measurement issue. With the Call Report data alone it is not possible to resolve this issue, but the non-account-holding banks hold a very small amount of assets—approximately 5 percent—so we are not too concerned about this source of mismeasurement.

[5]As of the second quarter of 2011, uninsured institutions held $678 billion of reserves. Among the uninsured institutions, there were 228 U.S. branches and agencies of foreign banks, 34 Edge and Agreement corporations, and 71 non-deposit trust companies.

[6]Note that some banking corporations have both U.S. insured affiliates and uninsured affiliates. An example of this is Deutsche Bank, which had both insured U.S. banks and an uninsured U.S. affiliate. We do not incorporate uninsured foreign affiliates when we aggregate U.S. insured banks at the holding company level. In principle, however, reserves in foreign affiliates could be transferred to a related U.S. bank relatively quickly if the parent company decided to do so. In the second quarter of 2011, Deutsche's domestic insured institutions reported reserves of about $21 billion, and its uninsured affiliate reported reserves of about $67 billion.

Figure 1. Total Reserves and Our Data



the aggregates constructed using the Call Reports. For example, the thrift institutions mentioned above do not report reserve holdings at the individual level, but those holdings are included in the measure of aggregate reserves in the H.3 tables. Thus, there will inevitably be a gap between the total reserves accounted for by our cross-sectional data and the total reserves reported on the H.3 statistical release. In the appendix, we explain in detail other issues with reconciling these two sources of data and how we deal with them in our analysis.

Figure 1 displays the time series for end-of-quarter total reserves as reported on the Federal Reserve's H.3 statistical release. Required reserves over the period in question and total reserves in our sample are also shown in the figure (dotted and dashed lines, respectively).[7] The gap between total reserves in our data and total reserves from the H.3 release amounted to $176.6 billion in the second quarter of 2011. Reserve holdings by credit unions and some other thrifts account for this gap.

[7]The Federal Reserve's H.3 statistical release typically does not cover the last day of the quarter, and for this reason the H.3 aggregates can show some inconsistencies with those produced using the Call Reports—our main data source. This is especially a factor during the periods when reserve balances change significantly from one day to the next, as was the case at the end of the third quarter of 2008.

Figure 1 also depicts three shaded areas that demarcate three distinct periods of noticeable increases in total reserves. Prior to the autumn of 2008, total reserves had been fluctuating between approximately \$40 billion and \$60 billion, and for the previous five years required reserves had never accounted for less than 80 percent of total reserves. Starting in mid-September 2008—the time of the Lehman Brothers bankruptcy—this situation changed dramatically and the level of reserves increased rapidly, to reach a level of approximately \$850 billion by the end of the year. We call this surge in reserves the *first wave* of increases in reserves. As the Federal Reserve's credit programs were winding down, the level of reserves actually decreased during the first two quarters of 2009. A *second wave* of reserve increases came in the last two quarters of 2009, as the run-off of the credit programs no longer compensated for increases associated with the Federal Reserve's first asset purchase program. Total reserves increased by approximately \$300 billion in this second wave, to reach a level of about \$1.1 trillion. During most of 2010 the level of reserves fluctuated around that level. In November 2010, the Federal Reserve embarked on a new program of asset purchases for \$600 billion, which resulted in a *third wave* of increases in total reserves lasting until the end of the second quarter of 2011. Reserves reached a level of \$1.6 trillion at that time.[8]

The initial increase in reserves in September 2008 occurred in an environment of falling market interest rates on low-risk debt, and amid crisis conditions in financial markets. This combination of factors meant that there was both a lower opportunity cost and a higher perceived benefit of holding excess reserves. As a result, the demand for reserves likely increased.

In mid-October 2008, the Federal Reserve began paying interest on all reserve balances held by depository institutions. By mid-November 2008 the interest rate paid on reserves became essentially equal to the target for the federal funds rate, allowing the Federal Reserve to continue to increase the quantity of reserves independently of its interest rate target. With IOR close to the federal funds

[8]Ennis and Wolman (2012) discuss in more detail the Federal Reserve's policy actions and the resulting evolution of reserves. For a good taxonomy of the programs, see the Federal Reserve Board website, http://www.federalreserve.gov/monetarypolicy/bst_openmarketops.htm

rate, the costs of holding reserves were significantly reduced, and this remained the case for the rest of the period under study.[9] Keister and McAndrews (2009) explain that the increases in reserves can be viewed as an artifact of the credit and asset purchase programs that the Federal Reserve undertook. It should be stressed that while the asset purchase programs resulted in reserves increasing, the stated intention of those programs was to reduce long-term interest rates. Many papers have assessed their effectiveness in this dimension. See, for example, Gagnon et al. (2011).

3. The Distribution of Reserves

A natural first step toward understanding the potential implications of a large quantity of reserves is to determine who were the main holders of those reserves and to what extent the distribution of reserves changed over time. For this purpose, we provide an overview of the main features of the distribution of reserves across banks in our sample period.[10]

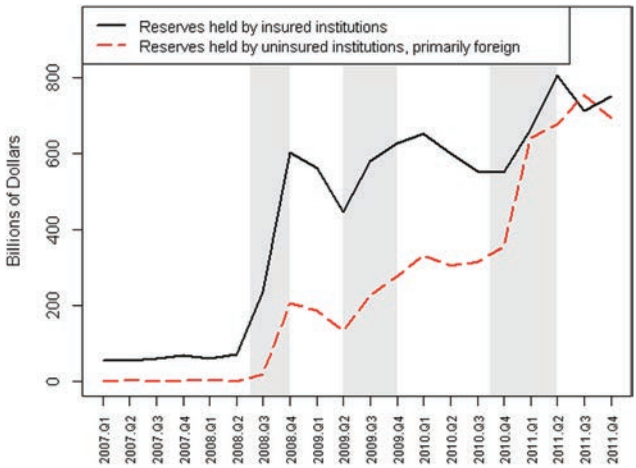
A first broad decomposition of aggregate reserves is between domestic banks and FBOs. Figure 2 shows that FBOs held a significant amount of reserves beginning in the third quarter of 2008. In fact, by 2011 the quantity of reserves was roughly the same in the two groups of institutions, even though total assets in FBOs are between 10 and 20 percent of total assets in domestic institutions. While both domestic banks and FBOs absorbed reserves during the first and second wave of increases in reserves, during the third wave FBOs increased their participation in reserve holdings significantly more than domestic banks. In fact, FBOs absorbed the lion's share of the reserves in the third wave.

Because FBOs are somewhat atypical banking organizations, we will discuss them in more detail below. We then turn our attention to domestic institutions and, in particular, to the distribution of

[9]The effective federal funds rate has generally been below IOR. It is widely believed that this reflects the fact that the government-sponsored enterprises cannot earn interest on reserves and only deal with a small set of counterparties, which are not willing to expand their balance sheets enough to compete away the spread. See, for example, Bech and Klee (2011).

[10]For brevity, we will often use the term "banks" when we mean "banks aggregated to the holding company level."

Figure 2. Insured and Uninsured Institutions



reserves across individual banks within this group. To the extent that there is a non-negligible idiosyncratic component to the arrival of lending opportunities, it is important to know whether reserves tend to be concentrated in a few banks or spread out across institutions.[11]

3.1 Foreign Banking Organizations

Uninsured FBOs increased their reserve holdings by a factor of more than 1,000 from the middle of 2008 to the middle of 2011, going from \$563 million on June 30, 2008 to \$666 billion on June 30, 2011.[12] As a proportion of total reserves, FBOs' reserves increased from 6 percent to close to 50 percent. Relative to FBOs' assets, reserves rose from 0.03 percent to more than 30 percent. Meanwhile, their

[11]While at any point in time the total amount of reserves in the system is set by the monetary authority, *an individual bank* can choose to hold whatever level of reserves it wishes. And, during the period under study here there was, in fact, substantial reallocation of reserve across banks from quarter to quarter (see Ennis and Wolman 2012 for calculations of gross flows of reserves across institutions).

[12]Reserve holdings at these institutions increased in the last two quarters of 2008, to reach a level of around \$200 billion, and increased another \$150 billion in the second half of 2009. But the biggest increase in reserve balances held by foreign institutions came during the Federal Reserve's second asset purchase program in the first half of 2011, when they increased their reserve balances by more than half (\$363 billion) of the total size (\$600 billion) of the program.

total assets rose by less than 2.5 percent, so the increase in reserves was accompanied by a significant reduction in the *levels* of some other asset categories.

There are several factors that can help explain why FBOs increased their reserves so much. In late 2008, many foreign institutions with large-dollar assets experienced difficulty in rolling over short-term dollar funding (Fleming and Klagge 2010, and Goldberg, Kennedy, and Miu 2011). One way these institutions responded was by drawing down their U.S. affiliates' deposits with the parent company to build up reserve accounts with the Federal Reserve, creating a pool of precautionary dollar balances.

Another element in explaining the increase in reserves is the decline in market interest rates that occurred over the period: with the Federal Reserve holding fixed the interest rate on reserves, it made sense for foreign (and domestic) institutions to hold a larger share of their assets in reserves than in securities bearing lower market interest rates.

In mid-2011, amid the European sovereign debt crisis, dollar funding concerns at foreign banks again became a major issue. In addition, in April 2011 a change in the FDIC premium for deposit insurance lowered the effective return on reserves for insured institutions.[13] With the Federal Reserve's asset purchases driving up aggregate reserves, it was predictable for them to flow disproportionately to uninsured FBOs that were not exposed to the increased premium (Kreicher, McCauley, and McGuire 2013).

Table 1 summarizes the changes in FBOs' balance sheets during the three waves of increases in reserves. For a more comprehensive description of these changes, see Ennis and Wolman (2012). The asset categories in the table, in addition to reserves, are some of the categories that experienced significant changes as reserves changed. During the first wave, reserves increased by more than 10 percent of assets, and this change was more than offset by a change in

[13]Insured institutions contribute a quarterly assessment to fund the Federal Deposit Insurance Fund. An institution's assessment is calculated by multiplying its assessment rate by its assessment base. From the beginning of the FDIC until 2010, a bank's assessment base was about equal to its total domestic deposits. The 2010 financial reform legislation (the Dodd-Frank Act) required that the FDIC amend its regulations effective April 2011 to define a bank's assessment base as its average consolidated total assets minus its average tangible equity.

Table 1. Balance Sheets of Uninsured FBOs

Time Period	Reserves	Net Due from Related DIs	Loans and Leases
	(measured as % of assets)		
First Wave	10.06	−10.91	3.37
Second Wave	7.71	−1.57	−3.08
Third Wave	15.59	−9.38	−1.11

FBOs' deposits at other related institutions ("net due from related depository institutions").[14] During the second wave, reserves rose by almost 8 percent of assets, with offsetting reductions in deposits at related institutions (1.57 percent), loans (3.08 percent), and various other asset categories (not included in the table). Finally, in the third wave (the Federal Reserve's second asset purchase program), reserves increased by more than 15 percent of assets, and deposits at related institutions and loans again offset a significant portion of that change.

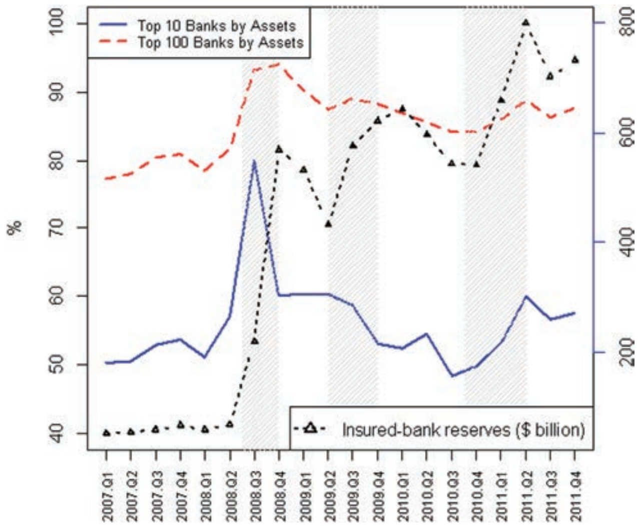
It should be stressed here that reserves at foreign institutions are by no means "stuck" there. Just as the foreign institutions rapidly increased their reserve holdings from 2008 to 2011, they could rapidly decrease those balances. These institutions hold significant quantities of both loans and securities, and they could be sensitive to economic conditions and market interest rates in choosing their reserve positions.

3.2 Domestic Banks

The high concentration of assets in the domestic banking sector is a well-documented fact (Janicki and Prescott 2006). For this reason, reserve holdings are also likely to be highly concentrated. Figure 3 plots the time series for the percentage of total reserves of U.S. insured banks held by the top 10 and top 100 banks by assets on the left axis, and the level of insured-bank reserves on the right axis.

[14]These foreign institutions were major participants in the Federal Reserve Term Auction Facility, which lent significant amounts to banks between 2008 and 2010 (see Benmelech 2012 for details).

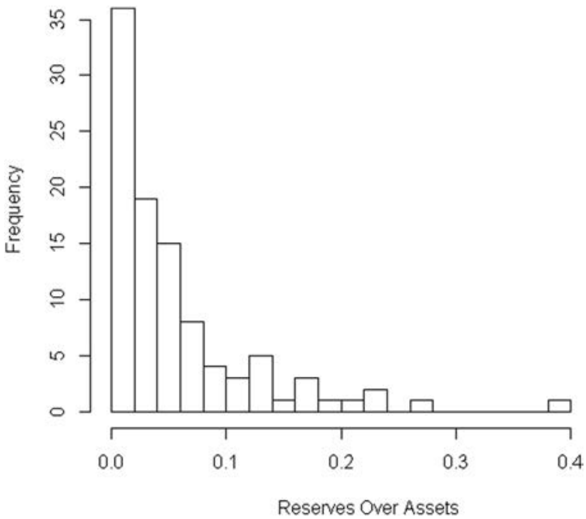
Figure 3. Reserves Held by Top 10, Top 100 Insured Banks



The 10 largest banks tend to hold roughly half of the reserves (and 55 percent of the assets—not shown in figure 3) and the top 100 banks hold roughly 80 percent of the reserves (and also 80 percent of the assets). It seems that the proportion of reserves held by the top 100 banks moved to a permanently higher level (approximately 85 percent) after the level of reserves increased at the end of 2008. This change in the proportion of reserves held by the top 100 banks did not correspond to a similar change in the proportion of assets. In fact, the proportion of assets held by the largest banks was fairly stable during the entire period under consideration.

We can also see in figure 3 that the percentage of reserves held by the ten largest banks increased significantly in the third quarter of 2008 when the level of reserves was increasing rapidly for the first time. This is also true for the top 100 banks. Initially the top 10 banks were the main drivers of the increase in the proportion of reserves held by the top 100 banks. From the fourth quarter of 2008 onward, however, the banks in the top 100 group which were not in the top 10 group increased their reserve holdings more markedly (as shown by the increasing distance between the solid and the dashed lines in figure 3). Starting in the third quarter of 2010, the top 10

Figure 4. Reserves Over Assets, 2010:Q2 (top 100 institutions by assets)

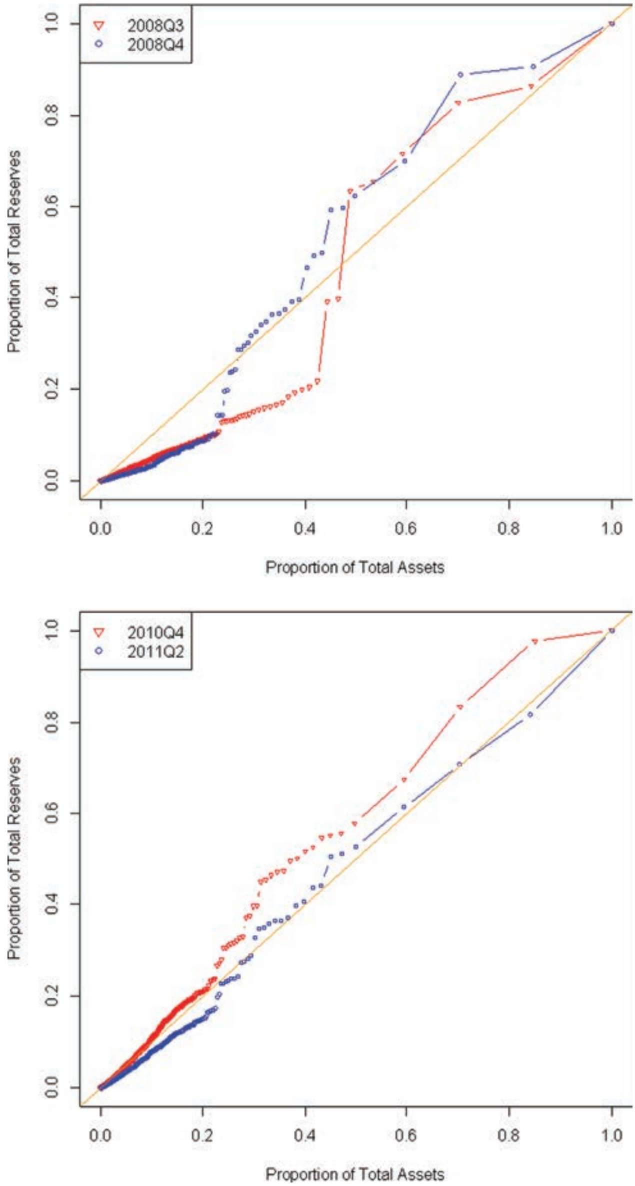


banks again increased their reserve holdings faster than the rest of the large banks in the system. Note that the periods when aggregate reserves grew rapidly tended to coincide with the periods when the top 10 banks increased their holdings of reserves faster than the rest.

According to figure 3, large institutions held a large portion of total reserves. But, were reserve holdings across these large institutions proportional to assets? Figure 4 displays the histogram of the ratio of reserves to assets in the second quarter of 2010 for the 100 largest insured institutions by assets. We choose this quarter because it is representative of a time when total reserves were not increasing rapidly and financial conditions were relatively calm. The figure shows that there was in fact wide dispersion in the ratio of reserves to assets across institutions.

So far we have discussed the distribution of reserve holdings among the largest banks. To get a broader view of the distribution of reserves across the entire domestic banking system, we construct a plot *similar* to a Lorenz curve for reserves. Figure 5 presents those plots. Instead of ranking banks by reserves on the horizontal axis, they are ranked by assets, and the horizontal axis measures fraction

Figure 5. Lorenz-Type Curve of Concentration of Reserves



of assets, not fraction of banks.[15] As is clear from the top chart in figure 5, the largest banks held a disproportionately high amount of reserves at the peak of the crisis in late 2008.

In the curve that corresponds to the end of the third quarter of 2008 (triangles), we see two institutions that held a disproportionate amount of reserves relative to assets. These institutions are State Street Bank (with $52 billion, amounting to 19 percent of its assets) and BONY Mellon (with $38 billion, amounting to 16 percent of its assets), both banks that primarily provide services to other financial institutions. During the peak of the financial crisis, State Street and BONY likely faced significant uncertainty about their needs for payment-related liquidity, which would justify holding a disproportionately large amount of reserves.

At the end of 2008, banks holding the bottom 20 percent of assets held less than 10 percent of reserves. During the second wave, reserves slowly became more evenly distributed across banks. If anything, large banks held a lower proportion of the reserves than they held of the assets in the system. By mid-2010, this same group of banks held almost exactly 20 percent of the reserves, and in fact, the Lorenz-type curve for the second quarter of 2010 (not included in the figures) roughly overlaps with the 45-degree line for the first 25 percent of assets (and reserves) in the system. Perhaps the normalization of financial conditions contributed to a more even distribution of reserves and a higher participation of smaller banks in the holding of those reserves. Also, the fact that the Federal Reserve conducts open-market operations with only a very small set of counterparties may tend to create some concentration of reserves in large institutions when those open-market operations are large. Since by mid-2010 the quantity of reserves had not been growing for some time, this force towards concentration of reserves in large banks was not operative at the time, allowing the reserves to become more evenly distributed.[16]

[15] Note that because the horizontal axis measures fraction of assets, the curves can lie above the 45-degree line.

[16] It is interesting to note that prior to September 2008, small banks held a disproportionately large share of reserves because reserve holdings were driven by required reserves, and small banks, on average, hold relatively large reservable deposits. In this sense, although the level of reserves was much higher, the distribution of reserves among small banks in 2010 had moved back close to its pre-crisis state.

The bottom chart in figure 5 shows the Lorenz-type curves for the fourth quarter of 2010 and the second quarter of 2011. The figure indicates that the third wave of increases in reserves, which happened mostly during the first half of 2011, resulted in a shift in the distribution of reserves towards large banks. In other words, by the end of the second quarter of 2011 the largest banks in the system held a higher proportion of total reserves than they held at the end of 2010, when the 2010–11 Federal Reserve asset purchase program started.

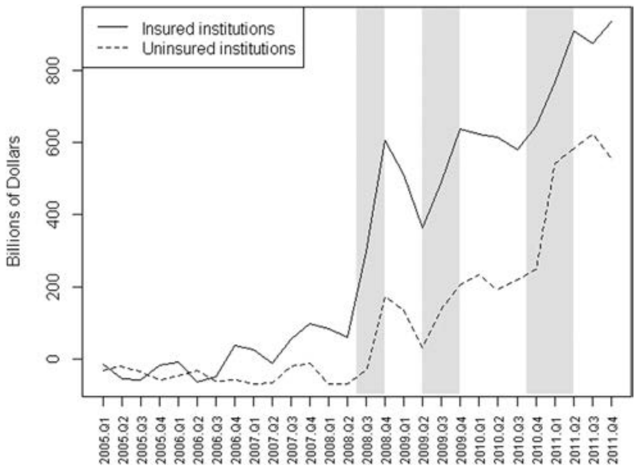
The degree to which reserves were concentrated among banks is informative for the banking system's ability to adjust to changing conditions. Bank lending opportunities likely arise in an idiosyncratic and partly unexpected manner. Even in situations when general economic conditions are improving, some banks may experience more active demand for loans than others. Anticipating which banks will be the ones making more loans seems challenging. At the same time, according to our working hypothesis, the ability of banks to quickly take advantage of new lending opportunities depends in part on the quantity of reserves they are holding. The fact that the reserves were fairly widespread across banks in the data, as we have shown in this section, suggests that many of those banks could have been in a favorable position to tap their reserve holdings had lending opportunities arrived.

4. Bank Liquidity

Interest-bearing reserves are a close substitute for short-term low-risk securities. An increase in aggregate bank reserves brought about by Federal Reserve policies can, in principle, be offset by a reduction in banks' holdings of short-term low-risk securities. If the offset is complete and total effective liquidity in the banking system does not change, then changes in total reserves are unlikely to make a difference for lending and monetary policy more generally. In this section, we look at the data on bank liquid assets and reserves to assess the extent of the substitution between reserves and other liquid assets during our sample period.

In figure 6, we display time series for measures of aggregate liquidity held by banks. For insured domestic institutions, the solid line comprises reserve balances, vault cash, short-term securities (one

Figure 6. Aggregate Bank Liquidity



year or less to maturity), and net repo and federal funds sold to non-banks. We view this measure as a reasonable proxy for low-risk assets on the banking system's balance sheet. Given the large changes in levels that we are concerned with, alternative measures of liquidity are unlikely to modify the main conclusions. Note, however, that we have excluded from aggregate bank liquidity in figure 6 federal funds sold to banks, reverse repo with banks, and balances due from other depository institutions. These items represent liquidity to individual institutions but not to the banking system as a whole. For this reason, our measure of aggregate liquidity seems most relevant when considering aggregate shocks to the banking system, instead of idiosyncratic shocks to certain banks that could be accommodated by changes in interbank lending.

FBOs do not report short-term securities. For those institutions the dashed line in figure 6 plots a measure of liquidity that instead uses U.S. Treasury securities and U.S. government agency obligations as the measure of liquid securities.

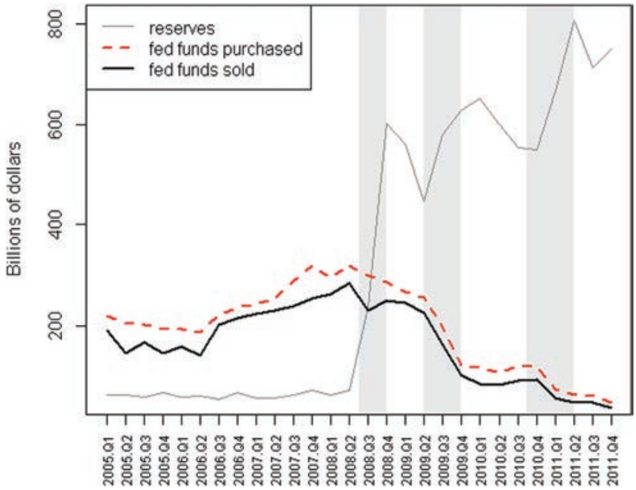
According to figure 6, prior to the crisis aggregate liquidity in the banking system was actually negative in some periods: banks were borrowing short-term funds from the non-bank sector in excess of the liquid assets they held. The initial increase in liquidity in the fall of 2008 corresponds to a period when banks' demand for liquid assets is likely to have increased significantly (Ashcraft, McAndrews, and Skeie 2011). However, in subsequent quarters financial market conditions tended to normalize, and still the quantity of liquid assets remained high—increasing with both of the Federal Reserve asset purchase programs. We interpret these patterns as saying that contrary to what one might have expected, once financial conditions normalized, the banking system did not primarily substitute reserves for other forms of liquid assets. In fact, from 2009 to 2011 aggregate liquidity in insured institutions increased by more than the increase in reserves (reserves increased by $350 billion and liquidity increased by $560 billion) and the increase in liquidity occurred at both insured and uninsured institutions.[17]

Having looked at aggregate bank liquidity, we now turn our attention to the cross-section of domestic banks. Aside from reserves and securities, individual banks hold liquidity in the form of deposits with other banks or short-term loans to other banks. For this reason, we modify our measure of liquidity to include all the components in the aggregate measure used in figure 6 plus balances due from other depository institutions. While this modification in the definition of liquidity is conceptually important, it is empirically less consequential given the policy environment under consideration. By paying interest on reserves at a near-market rate and dramatically increasing the quantity of reserves, the Federal Reserve changed the calculus of individual banks' liquidity management. Banks that previously economized on holding reserves and then borrowed when needed in the federal funds market no longer faced a significant cost of, instead, holding a high level of precautionary reserves. For this reason, the volume of activity in the federal funds market declined considerably. Figure 7 shows that among the insured institutions in our sample, trading in the federal funds market fell from around $300 billion in the second quarter of 2008 to around $50 billion in the second quarter of 2011.

We denote by OLq_i the liquidity, other than reserves, held by bank i. Thus total liquidity of bank i is the sum of OLq_i and reserves R_i. To understand the relationship between increases in reserves and increases in liquidity at the level of individual banks, we run the following regression *separately* for each of the three waves:

[17]It is of course true that because reserves were introduced by purchasing Treasury and mortgage-related securities, the consolidated private sector did substitute reserves for other forms of liquid assets. The extent to which the increase in reserves matters depends in large part on the extent to which banks matter.

Figure 7. Interbank Activity at Insured Institutions



$$\frac{\Delta OLq_{it}}{A_{it}} = \alpha_t + \beta_t \frac{\Delta R_{it}}{A_{it}} + \varepsilon_{it},$$

where ΔOLq_{it} is the change in OLq_i between the beginning and the end date of a given wave t and A_{it} is the level of total assets of institution i at the beginning of wave t, with $t = 1, 2, 3$. We normalize the change in liquidity by assets to avoid an automatic positive correlation driven by bank size. Similarly, we exclude reserves from the dependent variable to avoid generating any spurious correlation.

Table 2 reports the dollar change in (total) liquidity per dollar increase in reserves implied by the estimated coefficients (i.e., the value of $1 + \widehat{\beta}_t$, where $\widehat{\beta}_t$ is the estimated value of β_t). We also report in the table the standard error of $\widehat{\beta}_t$ and the R-squared from the regression.[18]

[18] As a robustness check, we run the regression excluding the two clearing banks, JP Morgan Chase and BONY, and State Street, since these are banks with potentially very different exposure to the financial crisis. As we saw in section 3, BONY and State Street were indeed evident outliers for the change in reserves during the first wave of increases in reserves. Excluding these three banks, however, does not change the results in any significant way.

Table 2. Reserves and Liquidity

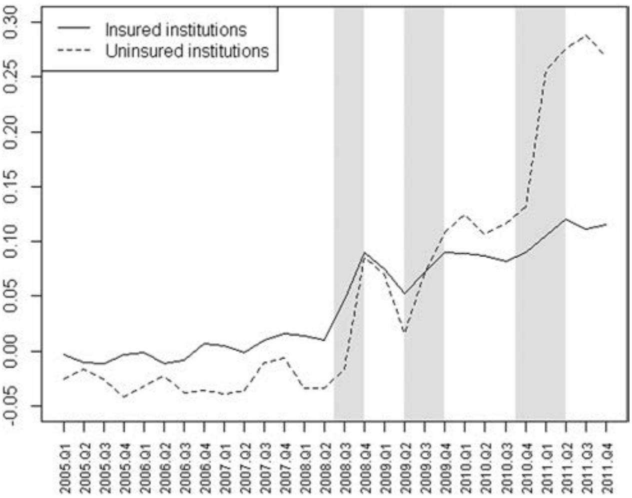
Time Period	$1 + \widehat{\beta}_t$	Standard Error of $\widehat{\beta}_t$	R-squared
First Wave	0.662	0.040	0.42
Second Wave	0.777	0.060	0.11
Third Wave	0.945	0.076	0.00

We see in table 2 that changes in reserves are not offset by changes in other liquid assets at the level of individual banks (that is, $\widehat{\beta}_t$ is significantly different from -1). Rather, (total) liquidity changes almost one for one with reserves, especially during the second and third waves. During the first wave, a crisis period, the correlation between reserves and other sources of bank liquidity is stronger, as is evident from the table ($\widehat{\beta}_t$ is close to -0.4). This stronger correlation may in part be the result of banks stocking up reserves for precautionary reasons as they were losing access to other sources of liquidity during the crisis (Ashcraft, McAndrews, and Skeie 2011, and Berrospide 2013).

So far, we have considered only absolute measures of bank liquidity. Often, macroeconomic models with money and banking posit an aggregate technology whereby reserves, in their role as the ultimate liquid asset, are an input into the production of deposits.[19] In these models, the focus of attention is generally the ratio of aggregate liquid assets to aggregate deposits. Figure 8 plots the ratios of the measures of aggregate liquidity in figure 6 to total deposits at domestic banks and FBOs. The figure could be considered an empirical counterpart of the assumptions in the aforementioned models. There are two points worth noting from the figure. First, the ratio of FBOs' liquid assets to deposits went from being negative to almost 30 percent. A significant portion of the increase came in 2011, not surprising since these institutions do not offer insured deposits and the terms for holding reserves shifted in their favor after the change in FDIC insurance premiums in April of 2011. Second, for insured domestic institutions, the increase in liquidity relative to deposits

[19] An early example of a macro model with deposits and reserves is Chari, Christiano, and Eichenbaum (1995). More recently, versions of this approach have been taken by Canzoneri et al. (2008), Hornstein (2010), and Ireland (2014).

Figure 8. Ratio of Aggregate Bank Liquidity to Total Deposits



was dramatic in late 2008 but relatively modest since then. In other words, while figure 6 already showed that domestic banks were not simply substituting reserves for other liquid assets, figure 8 shows that the increases in liquidity after 2008 were accompanied by nearly proportional increases in deposits.

In their recent contributions, Hornstein (2010) and Ireland (2014) are explicitly concerned with monetary policy when interest is paid on reserves, and thus deserve some special attention here. In Hornstein (2010), reserves are a perfect substitute for bonds in serving as liquidity, which must be held as a constant proportion of deposits (a technological assumption). In contrast, Ireland assumes that reserves serve a unique function in supporting deposits, so that even when there is interest on reserves at the market interest rate, banks have a determinate demand for reserves. Hornstein's model would require a large shift in the deposit-taking technology to accommodate the increase in liquidity relative to deposits since mid-2008. However, from 2009 through 2011 the behavior of domestic banks does not seem grossly at odds with Hornstein's model: with the interest rate on reserves essentially equal to the short-term market rate, banks' ratio of liquid assets to deposits fluctuated in a relatively narrow range.

With reserves having a unique character in Ireland's model, it is appropriate instead to look at the ratio of *reserves* to deposits. Although the interest rate on reserves has been constant since late 2008, short-term market rates have exhibited some fluctuations. Ireland's model predicts that the ratio of reserves to deposits would co-move negatively with the difference between market rates and IOR. Using weekly data from the Federal Reserve's H.15 and H.8 statistical releases, covering January 2009 through June 2011, the correlation between the three-month Treasury-bill rate and the ratio of aggregate reserves to deposits was −0.74. Unfortunately, it is not possible to compute this correlation only for insured institutions using weekly data.[20]

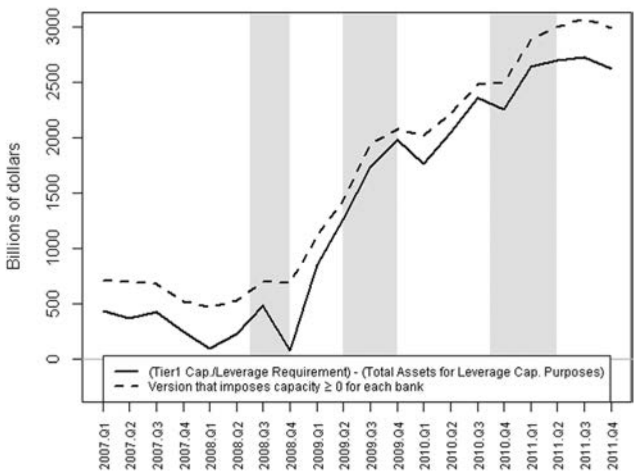
5. Bank Balance Sheet Capacity

Since reserves must be held by banks, a significant increase in the stock of reserves could, potentially, put pressure on banks' balance sheet capacity (Martin, McAndrews, and Skeie 2011). *For a given level of bank capital*, increases in the outstanding stock of reserves must be met with a decrease in other asset categories or in banks' leverage ratios. Regulatory requirements may restrict the latter as a margin of adjustment. To the extent that banks' leverage ratios are close to their regulatory minimum, increasing the stock of reserves implies that some other category of assets (for example, loans) must fall, and in that sense increases in reserves may become contractionary.

To assess the impact of the increases in reserves on banks' balance sheet capacity during our sample period, we start by computing an aggregate measure of balance sheet capacity for the largest 100 domestic banks. Denote by ρ the required leverage ratio—that is, the required ratio of tier 1 capital to assets, following the definition used in the Call Reports. Given capital holdings K, a bank could

[20]Kashyap, Rajan, and Stein (2002) emphasize the role of liquid assets in supporting not only deposit taking but also lending that is done through loan commitments. In the fall of 2008 there was a large increase in lending as commitments were drawn upon (see, for example, Ivashina and Scharfstein 2010). Surely that phenomenon must have played some role in the observed willingness of banks to hold higher liquidity in late 2008 and early 2009. However, for much of our sample period lending growth was practically non-existent and therefore cannot explain the increase in bank liquidity.

Figure 9. Balance Sheet Capacity of Top 100 Banks (based on leverage requirement = 0.07)



increase its assets up to K/ρ and still satisfy the maximum leverage restriction. We define as the bank's balance sheet capacity the difference between K/ρ and its actual level of assets for leverage capital purposes (as reported in the Call Reports).

Using a value of ρ equal to 7 percent, we aggregate in figure 9 the balance sheet capacity of the top 100 banks in two alternative ways: one in which we allow capacity to be negative (solid line) and one where we only consider positive values for the purpose of aggregation (dashed line).[21] We see in the figure that balance sheet capacity in fact increased significantly during this period. Of course, there are several factors that could explain this increase. For example, as a result of the crisis, many banks might have decided to strengthen their capital position at the same time that good lending opportunities were hard to come by. Also, as we saw in figure 2, a significant portion of the increase in reserves associated with the third wave did not flow onto the balance sheets of domestic banks.

To get a better sense of whether the increases in reserves put pressure on banks' balance sheets, we exploit the cross-sectional

[21] Taking $\rho = 0.07$ is a conservative choice. According to banking regulations at the time, to be considered well capitalized a bank had to have a ratio of capital to assets not lower than 5 percent.

dimension of our data set.[22] Figure 10 presents two charts, one corresponding to the first wave of increases in reserves and the other to the third wave; the chart for the second wave is similar and is not included for brevity.

The structure of the figures deserves some explanation. For each wave, we plot the leverage ratios for the largest 100 domestic banks against their changes in total assets net of reserves.[23] Furthermore, we split the sample into two sub-samples—banks that experience a large change in reserves (greater than 0.5 percent of the total change in reserves for these banks—represented with a dark circle in the figure) and the rest (represented with a light cross). If the increase in reserves put pressure on banks' balance sheets, then those banks that increased their reserve holdings significantly would have shown a positive correlation between their leverage ratio and the change in their assets net of reserves; banks that increased their reserves and had a low leverage ratio would have had to adjust their balance sheet by decreasing some other asset, and hence would have shown a decrease in their assets net of reserves.

Figure 10 shows little evidence of a positive relationship between the leverage ratio and the change in assets net of reserves during the first and third wave of increases in reserves. If anything, figure 10 seems to indicate that those banks that increased their reserves significantly and had a relatively low leverage ratio (between 5 percent and 10 percent) actually increased their holdings of other assets. We conclude from these charts that reserves, as they were created, did not crowd out other kinds of banking assets (such as securities, loans, and leases).[24]

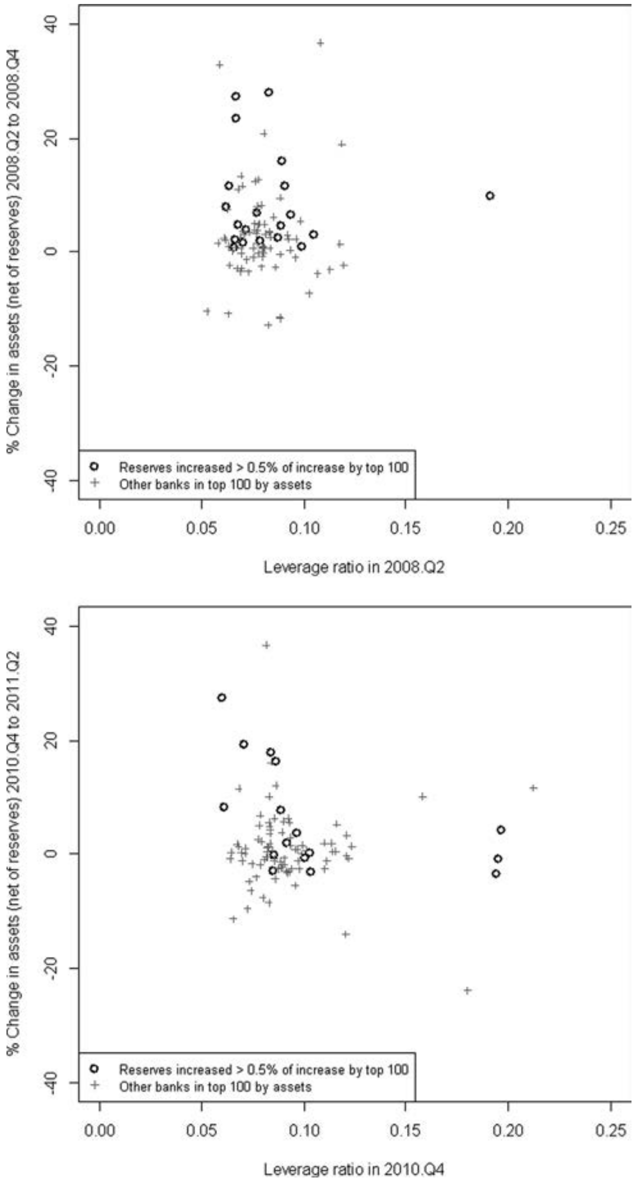
There is thus no clear evidence that the observed increases in reserves during the 2008–11 period systematically reduced the balance sheet capacity of banks. If anything, the balance sheet

[22]Aside from increases in the stock of aggregate reserves, the balance sheet capacity of domestic banks could play a role in a situation where the incentives for FBOs to hold reserves change abruptly and large quantities of reserves start to move into the balance sheet of insured domestic institutions.

[23]The two charts in figure 10 omit outliers with asset growth of more than 45 percent. This eliminates five observations from the top chart and two observations from the bottom chart.

[24]This finding is robust to different thresholds for what is considered a large increase in reserves and also to the sub-sample of large banks that is chosen.

Figure 10. Leverage Ratio and Net-of-Reserves Asset Growth (top 100 banks by assets)



capacity of (large) banks appears to have increased during that period. In principle, excess balance sheet capacity would allow banks to increase lending even without decreasing their holdings of reserves. More germane to the subject of this paper is the case in which banks fund new lending with their holdings of excess reserves. Such activity, however, has no implications for the balance sheet capacity of banks: it constitutes just a swap of one asset (reserves) for another asset (loans). Obviously, swapping reserves for loans has implications for the riskiness of the bank's balance sheet and impacts the bank's position regarding existing risk-sensitive capital regulations. We turn to this issue next.

6. Bank Capital

As a readily available source of funding, high levels of reserves provide flexibility to a bank that is looking to expand its loan portfolio. However, loans (and risky securities) are associated with higher capital requirements than reserves. A bank that is holding reserves but is facing a binding capital constraint is thus unlikely to engage in a sudden expansion of lending. As with deposits, raising capital quickly can be costly. For this reason, even a bank that holds a high level of excess reserves may not be able to take advantage of new lending (or investment) opportunities (see, for example, Van den Heuvel 2002 and Carlson, Shan, and Warusawitharana 2011).

During the first wave of reserve increases around the peak of the crisis, the possibility of banks increasing their lending activity seemed quite remote. However, by the end of 2009 and the beginning of 2010, U.S. economic conditions had become more stable and an increase in bank lending was not beyond the realm of possibility. In such a situation, it seems useful to assess the extent to which banks held capital that would have allowed them to "convert" excess reserves into loans or other risky assets.[25]

For the 100 largest (insured) banks, the charts in figure 11 plot the ratio of reserves to assets on the vertical axis and the risk-based capital ratio for each bank on the horizontal axis. The top chart in

[25] Relatedly, Bliss and Kaufman (2003) argue that the effects of reserve injections by the Federal Reserve depend on whether or not capital requirements are binding.

Figure 11. Reserves vs. Total Risk-Based Capital Ratio in 2009:Q4 (top 100 banks by assets)

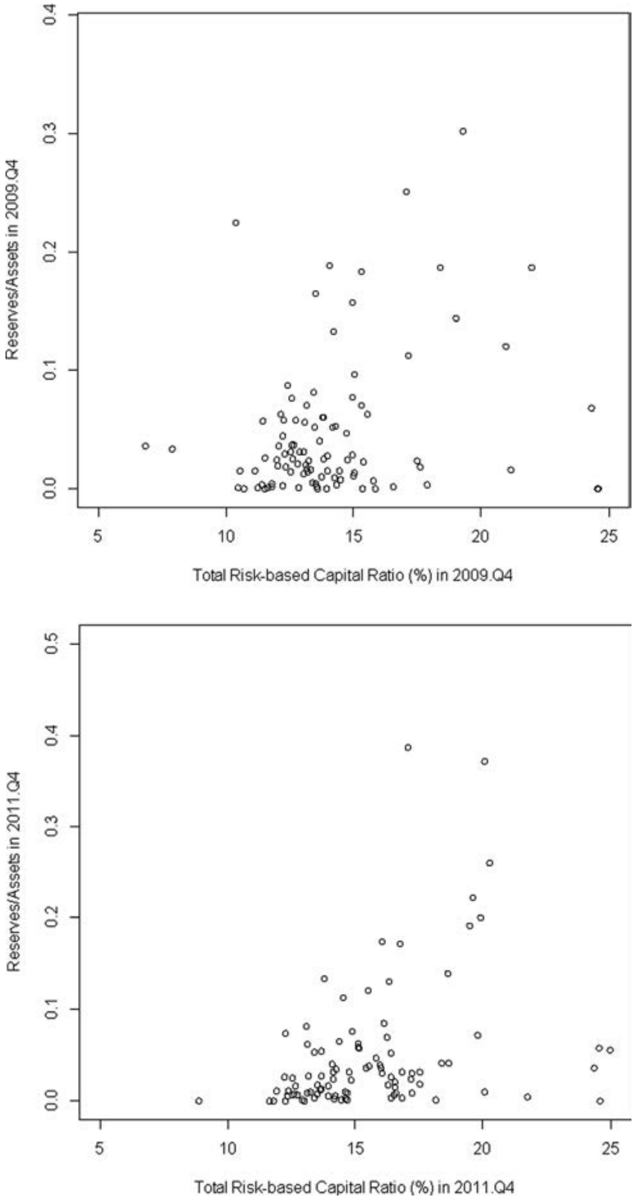


Table 3. Loanable Reserves at Large Insured Institutions (billions of dollars)

		Loanable Reserves	
	Reserves	Standard	Conservative
2009:Q4	510	485 (95%)	317 (62%)
2011:Q4	594	543 (91%)	452 (76%)

figure 11 corresponds to the last quarter of 2009 and the bottom chart is for the end of 2011 (the last data point in our sample). The figures show that in both periods there were a number of banks with fairly high capital ratios (above, say, 12 percent) that were also holding significant levels of reserves.[26]

Based on the figure alone, it is difficult to quantify banks' immediate ability to expand lending. To get a better sense of the quantitative importance of this factor, we compute an aggregate measure of *loanable reserves* that adjusts for the fact that each bank's new lending has to be consistent with satisfying its capital requirement and other regulatory requirements (the calculations are explained in the appendix). In the fourth quarter of 2009, the top 100 banks held $510 billion of reserves, and we calculate that $485 billion were loanable reserves (see table 3). We also calculate a more conservative measure of loanable reserves that takes into account that, during the later part of our sample period, U.S. regulators were in the process of raising capital requirements, and banks may have changed their behavior in anticipation thereof.[27] Even with our more conservative measure, $317 billion of reserves could be considered loanable. In summary, while binding capital requirements are likely to limit the

[26] At the time, to be considered well capitalized, a bank had to have a total capital ratio of at least 10 percent.

[27] The new rules on bank capital were approved by U.S. regulators in July 2013. The rules increased capital requirements and introduced a new minimum requirement on common equity capital. Our conservative measure is broadly consistent with the new rules. We assume that banks were targeting a 10.5 percent tier 1 capital ratio and a 12.5 percent total capital ratio. While we do not make adjustments to account for banks' common equity positions (which were not part of the regulation during our sample period), our measure is conservative enough that such an adjustment is unlikely to have any significant effect on the results.

ability of certain banks to transform reserves into loans, it is clear from the top chart in figure 11 and our measure of loanable reserves that in late 2009 several banks would have been able to use reserves to accommodate a significant increase in loan demand without facing binding capital constraints.

The bottom chart in figure 11 shows that again in the last quarter of 2011 large banks with significant holding of reserves were generally well capitalized, and that in fact there were many large banks with both high levels of reserves and high capital ratios. According to our measure of loanable reserves, \$543 billion of the \$594 billion of reserves held by these banks were loanable given existing capital requirements, and \$452 billion were loanable using the more conservative calculation. In other words, a significant proportion of the reserves held by large banks in the last quarter of 2011 could have been quickly used to fund loans without pushing these banks against their minimum regulatory capital levels. In line with the findings in Berrospide and Edge (2010), these numbers suggest that, rather than bank capital, loan demand was likely the main driving force behind banks' lending behavior.

The two charts in figure 11 also show that large banks in general had higher capital ratios in 2011 than at the end of 2009. Additionally, from figure 3 we know that they held more reserves. For these two reasons, as table 3 clearly indicates, the potential for large banks to create loans funded with reserves increased considerably between 2009 and 2011.

So far, we have demonstrated that the large outstanding level of bank reserves in 2011, in combination with the liquidity and capital position of large banks, could have supported a sudden increase in bank credit if conditions would have warranted. Evidently, aggregate loan demand did not materialize during the period under consideration. However, it is possible that some individual banks might have experienced an outbreak of lending opportunities. By looking at their responses, then, we could gain some insight into the possible interaction of reserves and bank credit in the aggregate. In principle, good lending opportunities should get reflected in the bank's return on lending. For this reason, we turn now our attention to the cross-section of interest income for insured banks in our sample.

7. Bank Interest Income

In the months prior to the autumn of 2008, market interest rates were over 2 percent and bank reserves yielded effectively zero interest. At the time, most banks held minimal excess reserves, with the potential benefits from holding reserves unable to compensate for the unfavorable rate-of-return differential. By the end of 2008, once the Federal Reserve had started to pay interest on reserves and to let the federal funds rate fluctuate in a 0- to 25-basis-points range, the decision of any given bank to hold reserves had changed dramatically.

For example, for the majority of the time between late 2008 and the end of our sample period, the interest rate on Treasury bills was below the 25-basis-points interest rate on reserves. For a bank choosing how to allocate its liquid assets, this became an obvious reason to favor reserve holdings over Treasury bills. Because the Treasury bill return is widely available, this shift in the rate-of-return differential affected all banks equally. However, the rates of return on lending opportunities vary across banks and, presumably, affect the opportunity cost of holding reserves for each bank. The question then arises whether those banks experiencing an unusually high return on their lending activities tended to accumulate fewer reserves.

To address this question, we concentrate our attention on the later part of the sample period: the years 2010 and 2011. Of course, one could also study this issue using data from 2008 and 2009. However, unusual economic circumstances during that period directly undermine the possibility of identifying any meaningful relationship between reserve holdings and rates of return on lending opportunities.

We construct a balanced panel with the largest 100 insured banks by assets. We approximate each bank's return on loans in the third quarter of 2010 (2011) by the forward-looking ratio of average loan interest income in the third and fourth quarter of 2010 (2011) to the level of loans in the third quarter of 2010 (2011). The first row of table 4 summarizes the distribution of rates of return on lending. We observe substantial heterogeneity in the level of rates, which is not surprising given the wide array of lending strategies pursued by different banks in the United States. When interpreting this table, it is important to keep in mind that the returns shown in the first row do not make any correction for risk.

Table 4. Distribution of Quarterly Rates of Return on Loans Across Top 100 Banks (annualized, in percentage terms)

	Min.	1st Quart.	Median	Mean	3rd Quart.	Max.	St. Dev.
Level 2011	−0.00	4.58	5.17	5.56	5.70	18.59	2.26
Change 2010 to 2011	−5.23	−0.39	−0.17	−0.15	−0.00	5.73	0.10

The second row of table 4 shows the change in returns from the third quarter of 2010 to the third quarter of 2011. These numbers give us an approximate measure of the change in each bank's rate of return on lending. Furthermore, differencing is a crude way to control for fixed variation in strategies across banks (with respect to riskiness, for example). The main feature to highlight from these calculations is that in general during this period changes in rates of return on lending were small relative to the level of those rates.

If banks with better lending prospects perceive a higher opportunity cost of holding reserves and hence hold less of them, then we should find a negative correlation between the changes in reserves and the changes in rates of return on lending across banks. We run a set of cross-section regressions to investigate this issue, using as the dependent variable the change in reserves (divided by assets) for each of the top 100 banks between the second half of 2010 and the second half of 2011. To focus on banks with significant lending activities, we only include in the regression those banks with a loan-to-asset ratio greater than 0.25. Our main results are reported in table 5. The coefficient on the change in the rate of return on lending is actually positive and not significant even when we add several controls, such us the change in deposits (over assets), the relative size of the institution, and the percentage change in assets (and even when we control for possible outliers). We conclude from this evidence that without much idiosyncratic variation in returns, it is difficult to determine whether an individual bank would choose to reduce their reserve holdings when experiencing especially good lending opportunities: during the period under study, meaningful changes in lending opportunities simply do not appear to have occurred.

Table 5. Regression of Change in Reserves (Divided by Assets) for Large Banks, from Second Half of 2010 to Second Half of 2011

	All Banks with Loan-to-Asset Ratio > 0.25				Without Outliers	
	(1)	**(2)**	**(3)**	**Without Top 4 Banks**	**All**	**Only Banks with Positive Loan Growth**
ΔReturn on Loans	0.838	0.837	0.875	0.826	1.298	1.916
ΔDeposits over Assets	0.333***	0.333***	0.330**	0.332***	0.233***	0.209**
ΔAssets over Assets	−0.127	−0.127	−0.125	−0.126	−0.092	−0.084
Assets as % of Total			0.379			
Dummy Top 4 Banks		0.035	−5.723			
Intercept	−1.220**	−1.222**	−1.362**	−1.222*	−0.868**	−0.750
Adjusted R^2	0.29	0.28	0.27	0.29	0.25	0.27
Number of Observations	96	96	96	92	85	43

Notes: *** indicates a p-value of less than 1 percent, ** a p-value less than 5 percent, and * a p-value less than 10 percent. Outliers are defined as those observations where the absolute value of the change in reserves was greater than 10 percent of bank assets and the absolute value of the change in return on loans was greater than 4 percent.

8. Conclusion

We see the research described in this paper as shedding light on the cross-sectional consequences of the dramatic increase and sustained high level of bank reserves in the United States between the last months of 2008 and the end of 2011. Our empirical investigation provides several elements that are relevant for assessing the policy implications of a banking system with large levels of excess reserves.

We uncovered five main patterns in the data. First, we found that reserves were widely distributed across banks and appeared to get relatively more concentrated in large banks only during periods of high growth in the total amount of reserves outstanding. When the total level of reserves stabilized for some time, reserves became more evenly distributed among institutions. Uninsured institutions (mainly U.S. branches and agencies of foreign banks) played a significant role in absorbing the changes in aggregate reserves during the sample period. In fact, in the second quarter of 2011 these institutions, numbering around 200, held more than 40 percent of total reserves.

Second, we observed an increase in the level of total liquidity in the banking system concurrent with the increase in aggregate reserves. In other words, reserves did not substitute for liquid securities at banks but rather complemented them, resulting in a material increase in total bank liquidity. With the Federal Reserve paying interest on reserves, banks had less of an incentive to economize on reserve holdings. They appear to have shifted from a policy of holding a low level of liquidity and regularly borrowing in the interbank market, to holding a permanently high level of outside liquidity with a significant share of excess reserves.

Third, we do not find evidence that the increase in reserves put pressure on insured banks' balance sheet capacity. If anything, between 2008 and 2011 large domestic banks increased their potential for balance sheet growth. While this fact does not have direct implications for banks' ability to fund lending with reserves, it does suggest that the banking system could expand significantly without confronting tight regulatory constraints.

Fourth, after the height of the financial crisis had passed and as the Federal Reserve increased aggregate reserves by undertaking large asset purchase programs, reserves flowed to the balance sheets of banks with relatively abundant capital. As a result, once crisis conditions abated, the lion's share of reserves in most of the large banks in our sample could have been converted into loans without creating any substantial pressure on their capital ratios.

Finally, we consider the possibility that changes in the rate of return on lending could drive the decision of individual banks to hold reserves. We find that during our sample period changes in rates of return on lending were small and not tightly linked to changes in the reserve allocation across large banks. The evidence, though, is far from conclusive. It may just be that the effect on reserves becomes evident only when more significant changes in rates of return occur.

This paper has concentrated on the period 2008 to 2011, when the Federal Reserve's second asset purchase program was completed. Since then, total reserves have increased by more than $1 trillion, making it increasingly important to understand whether and to what extent reserves matter for monetary policy and economic outcomes. The objective of our work was to contribute to that process.

Appendix

Integrating Aggregate and Cross-Sectional Data

Aggregate reserve balances with Federal Reserve Banks are reported in the Federal Reserve's H.3 release. Because banks can also use vault cash to satisfy reserve requirements, to arrive at a number for "Total Reserves" we sum depository institutions' reserve balances with Federal Reserve Banks and vault cash, both from the H.3 release.[28] To construct the analogous number for an individual bank in our sample, we sum the Call Report entries "balances due from federal reserve banks" and "currency and coin." Required reserves for an

[28] Note that there is a significant quantity of deposits at Federal Reserve Banks that is not included in our definition (or the Federal Reserve's definition) of reserves. On June 29, 2011 the U.S. Treasury held $106 billion on deposit at the Federal Reserve. Other institutions such as the International Monetary Fund, the United Nations, the World Bank, Fannie Mae, and Freddie Mac also have accounts with the Federal Reserve. Neither the Treasury's balances nor these other institutions' balances are included in reserves.

individual bank can be estimated using Call Report deposits and the Federal Reserve's formula for required reserves, though these calculations are less reliable because of the complications in the accounting rules used for reserves requirements. Excess reserves are the amount by which balances due from the Federal Reserve and vault cash used to satisfy reserve requirements exceeds required reserves. Unfortunately, while "vault cash used to satisfy reserve requirements" is reported in the H.3 release, it is not a Call Report entry. We assume that all vault cash is used to satisfy reserve requirements. To compare aggregated Call Report reserves plus currency and coin to the H.3 release's reserves plus vault cash, we need to make an adjustment for required clearing balances. For aggregated Call Report reserves held by insured institutions, we subtract required clearing balances, as reported in the Federal Reserve's H.4.1 release; these balances are included in balances due from Federal Reserve banks in the Call Report, but they are excluded from reserve balances in the H.3 release. This adjustment is minor. Required clearing balances had been trending down for several years and by the end of 2010 amounted to only $2 billion in total. It is important to note that this adjustment is not possible at the level of individual banks and hence we simply use the Call Report items "balances due" and "currency and coin" to proxy for the reserve position of individual banks at each point in time.

Calculating Total Loanable Reserves

Capital requirements in the United States mandate that banks satisfy several minimum ratios of capital to assets, based on different measures of capital and of assets. The leverage ratio, for example, is a simple ratio of capital to assets (without any significant adjustments). Transforming reserves into loans on the asset side of the balance sheet does not change this ratio in any material way. For this reason, the leverage ratio does not play a role in our calculation of loanable reserves.

The tier 1 capital ratio is the ratio of tier 1 capital to risk-adjusted assets.[29] The risk charge for reserves is lower than the risk charge for loans. Hence, transforming reserves into loans results in an

[29] Tier 1 capital consists of common equity and some types of preferred stock.

increase in risk-adjusted assets and, given tier 1 capital, a decrease in the tier 1 capital ratio. A bank with a low tier 1 capital ratio will then be less able to increase lending (or investment), even if it is holding sufficient excess reserves to fund the loans.

The total capital ratio is the ratio of the sum of tier 1 and tier 2 capital to risk-adjusted assets.[30] As with tier 1 capital, a bank with a relatively low total capital ratio (such that approaching the regulatory minimum becomes a concern) will tend to limit its expansion of credit, even when funding could be readily provided with the holdings of excess reserves.

In summary, some of the potential lending capacity associated with holding excess reserves should be "discounted" to the extent that those reserves are being held by banks with (effectively) binding capital constraints. It is impossible to determine precise levels of capital ratios at which the requirements become effectively binding. Some banks may be willing to make certain loans even if their capital requirement is relatively low. Others may take a more conservative approach to capital management and lending.

To obtain a simple estimate, we take the view that each bank's loanable reserves are given by the amount of new loans that could be funded by excess reserves while keeping the bank "well capitalized" for regulatory purposes. Reserves have zero weight in the computation of risk-weighted assets, and we take a conservative approach and assume all loans have 100 percent risk weight. During our sample period, a "well-capitalized" bank needed a tier 1 capital ratio higher than 6 percent and a total capital ratio higher than 10 percent. Using this criterion, we compute for each quarter in our data set the following measure of loanable reserves for the largest 100 insured banks in the sample:

$$LR(t) = \sum_i \min\left\{ \frac{\left[K_i^{T1}(t) - K_{iR}^{T1}(t)\right]^+}{0.06}, \frac{\left[K_i^{T}(t) - K_{iR}^{T}(t)\right]^+}{0.1}, \left[R_i(t) - R_{iR}(t)\right]^+ \right\},$$

[30]Tier 2 capital consists of allowance for loan losses, subordinated debt, and other convertible debt securities.

where K_i^{T1} is the dollar amount of tier 1 capital held by bank i and K_{iR}^{T1} is the amount of tier 1 capital that would allow the bank to have a ratio equal to 6 percent. Similarly, K_i^T is the dollar amount of total capital held by bank i and K_{iR}^T is the amount of total capital that would allow the bank to have a ratio equal to 10 percent. Finally, R_i is the level of reserves held by bank i and R_{iR} is the required reserves (net of vault cash) given its average level of transaction accounts liabilities in the quarter. The superscript + sign means that we are only considering non-negative values of these terms.

In 2011, capital ratios were regarded as likely to increase in the near future, especially for large banks. For this reason, we also compute a more conservative measure of loanable reserves based on the Basel III proposal. In particular, we fix the required tier 1 capital ratio to 10.5 percent and the total capital ratio to 12.5 percent. We use these ratios as a way to obtain a relatively conservative measure of loanable reserves, given the uncertainty with respect to capital requirements at the time.[31]

References

Ashcraft, A., J. McAndrews, and D. Skeie. 2011. "Precautionary Reserves and the Interbank Market." *Journal of Money, Credit and Banking* 43 (s2): 311–48.

Bech, M. L., and E. Klee. 2011. "The Mechanics of a Graceful Exit: Interest on Reserves and Segmentation in the Federal Funds Market." *Journal of Monetary Economics* 58 (5): 415–31.

Benmelech, E. 2012. "An Empirical Analysis of the Fed's Term Auction Facility." *CATO Papers on Public Policy* 2: 57–91.

Berrospide, J. M. 2013. "Bank Liquidity Hoarding and the Financial Crisis: An Empirical Evaluation." FEDS Paper No. 2013-03, Board of Governors of the Federal Reserve System.

Berrospide, J. M., and R. M. Edge. 2010. "The Effects of Bank Capital on Lending: What Do We Know, and What Does It Mean?" *International Journal of Central Banking* 6 (4): 5–54.

[31] The new capital regulation approved in July 2013 requires that a well-capitalized bank hold 10 percent total capital, 8 percent tier 1 capital, and 6.5 percent equity capital.

Bliss, R. R., and G. G. Kaufman. 2003. "Bank Procyclicality, Credit Crunches, and Asymmetric Effects of Monetary Policy: A Unified Model." *Journal of Applied Finance* 13 (2): 23–31.

Canzoneri, M., R. Cumby, B. Diba, and D. Lopez-Salido. 2008. "Monetary Aggregates and Liquidity in a Neo-Wicksellian Framework." *Journal of Money, Credit and Banking* 40 (8): 1667–98.

Carlson, M., H. Shan, and M. Warusawitharana. 2011. "Capital Ratios and Bank Lending: A Matched Bank Approach." FEDS Paper No. 2011-34, Board of Governors of the Federal Reserve System.

Chang, S.-H., S. Contessi, and J. L. Francis. 2013. "Understanding the Accumulation of Bank and Thrift Reserves during the U.S. Financial Crisis." Forthcoming in *Journal of Economic Dynamics and Control*.

Chari, V. V., L. J. Christiano, and M. Eichenbaum. 1995. "Inside Money, Outside Money, and Short-Term Interest Rates." *Journal of Money, Credit and Banking* 27 (4, Part 2): 1354–86.

Dudley, W. C. 2011. "Prospects for the Economy and Monetary Policy." Speech given at New York University's Stern School of Business, New York City, February 28. Available at www.newyorkfed.org/newsevents/speeches/2011/dud110228.html.

Ennis, H. M., and A. L. Wolman. 2010. "Excess Reserves and the New Challenges for Monetary Policy." *Economic Brief* (Federal Reserve Bank of Richmond) 10 (3).

———. 2012. "Large Excess Reserves in the U.S.: A View from the Cross-Section of Banks." Working Paper No. 12-05, Federal Reserve Bank of Richmond (August).

Fleming, M. J., and N. J. Klagge. 2010. "The Federal Reserve's Foreign Exchange Swap Lines." *Current Issues in Economics and Finance* (Federal Reserve Bank of New York) 16 (4).

Gagnon, J., M. Raskin, J. Remache, and B. Sack. 2011. "The Financial Market Effects of the Federal Reserve's Large-Scale Asset Purchases." *International Journal of Central Banking* 7 (1): 3–43.

Goldberg, L. S., C. Kennedy, and J. Miu. 2011. "Central Bank Dollar Swap Lines and Overseas Dollar Funding Costs." *Economic Policy Review* (Federal Reserve Bank of New York) 17 (1): 3–20.

Hornstein, A. 2010. "Monetary Policy with Interest on Reserves." *Economic Quarterly* (Federal Reserve Bank of Richmond) 96 (2): 153–77.

Ireland, P. N. 2014. "The Macroeconomic Effects of Interest on Reserves." *Macroeconomic Dynamics* 18 (6): 1271–1312.

Ivashina, V., and D. Scharfstein. 2010. "Bank Lending during the Financial Crisis of 2008." *Journal of Financial Economics* 97 (3): 319–38.

Janicki, H., and E. S. Prescott. 2006. "Changes in the Size Distribution of U. S. Banks: 1960–2005." *Economic Quarterly* (Federal Reserve Bank of Richmond) 92 (4): 291–316.

Kashyap, A. K., R. Rajan, and J. C. Stein. 2002. "Banks as Liquidity Providers: An Explanation for the Coexistence of Lending and Deposit-Taking." *Journal of Finance* 57 (1): 33–73.

Kashyap, A. K., and J. C. Stein. 1994. "Monetary Policy and Bank Lending." In *Monetary Policy,* ed. N. Gregory Mankiw, 221–56. University of Chicago Press.

Keister, T., and J. McAndrews. 2009. "Why Are Banks Holding So Many Excess Reserves?" *Current Issues in Economics and Finance* (Federal Reserve Bank of New York) 15 (8).

Kreicher, L. L., R. N. McCauley, and P. McGuire. 2013. "The 2011 FDIC Assessment on Banks' Managed Liabilities: Interest Rate and Balance-Sheet Responses." BIS Working Paper No. 413 (May).

Levin, A., and J. Taylor. 2013. "Falling Behind the Curve: A Positive Analysis of Stop-Start Monetary Policies and the Great Inflation." In *The Great Inflation*, ed. Michael Bordo and Athanasios Orphanides, chapter 4. University of Chicago Press.

Martin, A., J. McAndrews, and D. Skeie. 2011. "Bank Lending in Times of Large Bank Reserves." Staff Report No. 497, Federal Reserve Bank of New York (May, revised June 2013).

Plosser, C. 2011. "Exit." Speech given at the Shadow Open Market Committee Symposium, New York City, March 25. Available at www.philadelphiafed.org/publications/speeches/plosser/2011/03-25-11_shadow-open-market-committee.cfm.

Van den Heuvel, S. J. 2002. "Does Bank Capital Matter for Monetary Transmission?" *Economic Policy Review* (Federal Reserve Bank of New York) 8 (1): 259–65.