Supply Side Inflation Persistence - Supplementary material

Andrea Civelli* Nicola Zaniboni

University of Arkansas Transparent Value, LLC

Abstract

This document provides accessory material to the submitted version of the paper *Supply Side Inflation Persistence*.

The main purpose of this document is to provide some accessory material to the shorter version of the paper *Supply Side Inflation Persistence* as prepared for journal submission. We first provide a brief (but slightly more detailed) review of related literature. We then present some robustness checks for the empirical VAR analysis of the impulse response functions to monetary shocks. In the third section, we describe in detail the model used in the main paper. Finally, we provide some robustness analysis for the theory, showing results for a broader set of calibrations.

1 Literature Review

The empirical relevance of a cost channel of monetary transmission has been argued, for instance, in Barth and Ramey (2001). Looking at industry-level data, they consistently observe negative correlations between output and price-wage ratios following monetary contractions, and they make a case for strong supply-side effects of monetary policy. There is a relatively recent strand of literature that incorporates a cost channel in the New Keynesian framework, following the contributions of Christiano and Eichenbaum (1992) and Christiano, Eichenbaum and Evans (2005), who incorporated working capital considerations in their model. For example, Ravenna and Walsh (2006) study the implications of the presence of a cost channel for optimal monetary policy. Rabanal (2007), Henzel et al. (2009), and Castelnuovo (2011) investigate the relative importance of the cost and demand channel of monetary transmission to determine conditions under which a model-consistent price puzzle is obtained.

*Corresponding author: Economics Department, University of Arkansas, Business Building 402, Fayetteville, AR 72701. E-mail: andrea.civelli@gmail.com.

Several fixes have been proposed to address the problem of inflation persistence (or lack thereof), and they usually imply either some sort of ad-hoc deviation from fully rational, forward looking behavior, or radically different approaches that deviate from the present framework. Gali and Gertler (1999) were among the first to comprehensively analyze the implications of "rule-of-thumb", backward looking firms for the dynamics of inflation, and were followed by other studies that were similar in spirit (although not necessarily in their conclusions).[1] Mankiw and Reis (2002) provide an example of a more radical departure, in that prices are "physically" perfectly flexible and firms are forward looking, but the latter base their pricing decisions on information sets that might be outdated. Other approaches introduce mechanisms to obtain higher degrees of real rigidity/strategic complementarity in pricing: to this extent, for example, Carvalho (2006) incorporates heterogeneous price stickiness in the Calvo model, and Dotsey and King (2005) introduce several "supply side" features, such as produced inputs, variable capacity utilization, and variability in labor supply along the extensive margin. We note here that models of this kind aim at obtaining more prolonged effects on monetary policy on real activity, but do not address the issue of the monotonically decaying response of inflation to monetary shocks.

2 The VAR Analysis

The empirical response of inflation to monetary policy shocks is often at odds with the implications of macroeconomic theory. Following Sims (1991) and Christiano, Eichenbaum, and Evans (1999 and 2005), for example, a very extended empirical VAR literature has documented two main features of the inflation response to identified monetary shocks, both for the U.S. and various other countries. The first is known as price puzzle and indicates a positive response of prices and inflation to a contractionary monetary policy shock. The second is a high persistence of the response of inflation, in the sense that its peak is reached with a substantial lag, and its decay is very gradual.

The price puzzle problem can be at least partially mitigated by introducing a commodity price index in the VAR. Regardless of how the puzzle is ameliorated, the response of inflation to an expansionary monetary policy shock is typically hump-shaped and this naturally increases the persistence of the response. Standard New Keynesian models of inflation, on the other hand, imply a monotonic response to a policy shock. As a consequence, these models have a hard time generating inflation dynamics as persistent as those implied by the empirical estimates.

In this section, we estimate a VAR model with six U.S. macroeconomic variables to study the dynamic response of these variables to monetary policy shocks. In setting up the VAR and the identifying assumptions

[1] See for example Rudd and Whelan (2005).

of the monetary shock, we follow the empirical monetary VAR literature, and specifically Christiano et al. (2005). The variables of interest included in the VAR model are as follows: a commodity price index, the real GDP, the GDP deflator, the federal funds rate, total bank reserves, and non-borrowed reserves. We add a market loan interest rate in a second specification discussed below because the loan rate enters the marginal cost of firms in our model. As for the bank reserves, they seem to add useful information to address the price puzzle and they are explicitly included in our modeling of banks' balance sheets.

The data source for this set of variables is the Federal Reserve Economic Data (FRED), the online dataset maintained by the Federal Reserve Bank of St. Louis. FRED provides the real GDP, the two interest rates, the GDP deflator index, and the commodity price index. We apply a log transformation to real GDP. Each inflation rate is constructed as the annualized quarter-to-quarter log-difference of the corresponding price index. All variables have been seasonally adjusted (if not adjusted at the source).

The impulse response functions are obtained by a Cholesky recursive decomposition of the reduced form covariance matrix of the VAR residuals. The baseline ordering of the variables in the VAR which identifies the structural shocks of the model is based on Christiano et al. (2005). The identification strategy for the credit sector variables mimics that in Lown and Morgan (2006). The commodity price inflation, real GDP, and inflation are assumed not to contemporaneously respond to the monetary shocks and are first in the ordering. The federal fund rate comes fourth and is followed by the financial variables from the banking sector. We then have a three-block structure, where the fed funds rate follows the "real" block, and is followed by a credit sector block.

We also regard this ordering as consistent with the theoretical features of our model in which the policy rate is set by the Central Bank according to a Taylor rule that primarily responds to inflation and output gap. The credit markets are assumed to clear immediately after the observation of the policy rate, but their direct feedback to the monetary policy decision takes place with a delay of one period. While this assumption seems fair for normal economic periods, it may be debatable when the economy is subject to large, negative financial shocks and monetary policy promptly reacts to prevent the spreading of panic and liquidity crisis. The most obvious case would be the latest "Great Recession" of 2008-09. For this reason, we exclude data after the second quarter of 2008 from the sample we use to estimate the VAR. Other parts of the sample could be affected by the same issue, as for instance the aftermaths of the 2001 stock market crisis, but they are definitely shorter and less dramatic. We further discuss this point in the robustness checks section below.

The results reported are based on a Bayesian estimate of the VAR with standard Minnesota priors. The VAR is estimated with quarterly data over the sample 1979:3 to 2008:2 and four lags are included; this sample corresponds to the post-Volcker era. The response functions to a one standard deviation (contractionary) monetary policy shock are reported in Figure 1. For convenience only the response of output, prices, the fed funds rate, and the implied response of inflation are illustrated. The horizon of the responses is 24 quarters; the reported significance bands correspond to the 14th/86th percentiles of the posterior distribution of the responses.

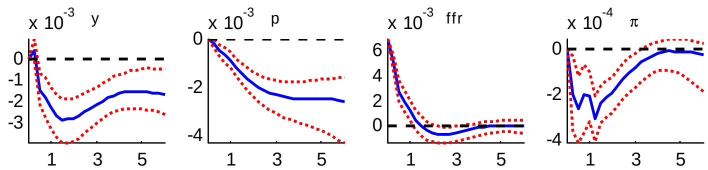

Figure 1: Response functions to a one s.d. contractionary monetary shock.
Notes: Three-block identification scheme. Years from the impulse on the $x-$axis. The dashed-lines are the 14th/86th percentiles of the posterior distributions of the responses.

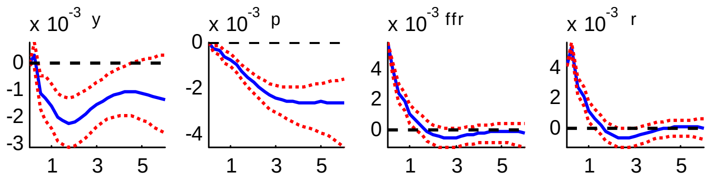

Figure 2: Response functions to a one s.d. contractionary monetary shock. VAR with loan rate.
Notes: VAR specification as in Figure 1, with the addition of the loan rate. Years from the impulse on the $x-$axis. The dashed-lines are the 14th/86th percentiles of the posterior distributions of the responses.

It is worthwhile to stress a few points regarding the results in Figure 1.

- The response function of the inflation rate is hump-shaped. The introduction of the commodity price inflation and of the bank reserves in the VAR reduces the initial positive response of the inflation rate (which is otherwise much stronger). Since the shock is contractionary, the hump of the response is turned upside down and reaches its trough after about one year. As far as the inflation response is concerned, our results are perfectly in line with those in the previous literature.
- The federal funds rate takes about six to eight quarters to return to the pre-shock level.
- The response of GDP is inversely hump-shaped. The GDP falls initially for about two years; it remains quite low for another year, before reverting back to the pre-shock level.

2.1 Robustness of the Response Functions

Figure 2 illustrates the response functions to a monetary shock for a VAR model that includes a loan interest rate; the loan rate is available from 1978 only. The identification of the structural shocks follows the same block strategy, where the loan rate is included in the last block along with the other credit variables and is ordered right after the fed funds rate. The responses of output, prices, and the policy rate do not exhibit noticeable changes. The loan rate's response closely follows the fed funds rate; however, it is slightly smaller for the first periods after the shock. The response of inflation (not reported in the figure) is still hump-shaped and persistent as that in Figure 1, but it displays a somewhat larger short-run volatility for a few periods.

The impulse response functions are robust to changes to the number of lags and different definitions of the variables used in the VAR. Virtually the same results are obtained if we use two, six, or eight lags. When we use six or eight lags, we observe a slightly more persistent and volatile inflation response. Four lags seems to be a satisfactory mid point according to standard optimal lag length tests, and also reflects the choice in Christiano et al. (2005).

The results are more sensitive to changes in the estimation sample, even though the baseline message remains valid. The sample selected for Figure 1 reflects the preference of having results for the period starting with the Volcker era. When we use the full available span from 1960 to 2008, however, the output essentially does not change. Nevertheless, there is a noticeable increase of the persistence of GDP and inflation. If the sample is truncated to 1979, the price puzzle becomes a more evident characteristic of the data. Finally, no substantial differences are observed when more recent sub-samples are considered.

We conclude by checking the robustness of the identification scheme. Many alternative orderings are obviously feasible, but we focus our attention on the position of the credit block, since it is a well established practice in the literature to identify the policy shock by putting the real sector block before the fed funds rate.

Thus, we maintain the position of the real block, and recompute the impulse response functions ordering bank reserves and loan rate before the fed funds rate. The response of inflation is only marginally affected by the adoption of the new ordering. Placing the loan rate before the policy rate seems to affect the response of output, which is now positive for a few periods and follows a declining trajectory. As this result seems especially at odds with conventional wisdom, we are inclined to disregard this ordering. Furthermore, the key element of the ordering is the position of the loan rate relative to the fed funds rate, regardless of the relative position of bank reserves. In fact, moving the loan rate to the last position, while keeping reserves fourth and fifth, does not alter the results of the core identification scheme. It seems, at least, an equivalently plausible story to think that the Central Bank may look at the volume of reserves in the system in order to set its policy, but that, at the same time, the loan rate immediately responds to changes of the funds rate.

Lown and Morgan (2006) study the macroeconomic role of bank standards, and they order them (along with commercial loans) after the fed funds rate. They also add other credit variables and the loan rate is always in the last block of their ordering. While the focus of their analysis is different than ours, we regard this approach as suitable to our context, as all these variables are financial in nature. We then follow it in selecting our core identifying strategy.

3 The model

The model of the paper is a standard New Keynesian general equilibrium model with nominal rigidities only. Both wages and prices are characterized by Calvo (1983)-style stickiness. A bank sector is added in which the financial intermediaries have some monopoly power in allocating funds to firms. We thus assume the loan interest rate is set in a staggered fashion too. Firms are assumed to pay their wage bill in advance, as is standard in the cost channel literature. Finally, complete financial markets are assumed.

In what follows, we use the indices i, j, and h to indicate firms, households and banks, respectively. In each of the three sectors, there is a continuum of individuals of mass one.

3.1 Households

The generic household j has preferences defined over consumption C and labor N, and described by the following isoelastic period utility function, separable in its arguments

$$u\left(C_t\left(j\right), N_t\left(j\right)\right) = \frac{C_t\left(j\right)^{1-\sigma}}{1-\sigma} - \frac{N_t\left(j\right)^{1+\eta}}{1+\eta}$$

The household lifetime utility is represented by

$$\mathbb{E}_t \sum_{k=0}^{\infty} \beta^k u(C_{t+k}\left(j\right), N_{t+k}\left(j\right)) \tag{1}$$

where σ is the inverse of the intertemporal elasticity of substitution and η is the inverse of the Frisch elasticity of labor supply.Utility is maximized subject to a flow budget constraint that reflects market completeness:

$$P_t C_t\left(j\right) + \mathbb{E}_t\left[Q_{t,t+1} B_{t+1}\left(j\right)\right] + D_t\left(j\right) = W_t\left(j\right) N_t\left(j\right) + B_t\left(j\right) + R_{D,t-1} D_{t-1}\left(j\right) + \mathbb{Y}_t \tag{2}$$

where $B_{t+1}\left(j\right)$ is the state-contingent payoff of the portfolio at the beginning of period $t+1$; $Q_{t,t+1}$ is the relevant stochastic discount factor; P_t is the price level in the economy; $W\left(j\right)$ is the nominal wage; $D_t\left(j\right)$ are the bank deposits of household j and $R_{D,t}$ is the average (gross) nominal interest rate paid on deposits by banks. Under complete markets, all households share the same stochastic discount factor. This assumption allows us to separate the consumption from the labor supply decisions. Household are assumed to provide differentiated types of labor, therefore the nominal wage $W_t\left(j\right)$ is specific for each of them. Finally, $\mathbb{Y}_t$

denotes the profits earned by firms and banks and equally distributed to the households (more on this below). Firms employ a composite labor index with constant elasticity of substitution between labor types ϕ_w. The total labor demand for firm i is

$$N_t(i) = \left[\int_0^1 N_t(i,j)^{\frac{\phi_w - 1}{\phi_w}} dj\right]^{\frac{\phi_w}{\phi_w - 1}}$$

This aggregator implies that the total demand for type j labor is given by

$$N_t(j) = \left(\frac{W_t(j)}{W_t}\right)^{-\phi_w} N_t \tag{3}$$

in which $W_t = \left[\int_0^1 W_t(j)^{1-\phi_w} dj\right]^{\frac{1}{1-\phi_w}}$ corresponds to the average wage on the market.[2]

Denoting $\lambda_t(j)$ the Lagrange multiplier of the maximization problem of (1) subject to (2), the relevant first order conditions are as follows:

$$\partial C \quad : \quad C_t(j)^{1-\sigma} = P_t \lambda_t(j) \tag{4a}$$

$$\partial B \quad : \quad Q_{t,t+1} = \beta \frac{\lambda_{t+1}(j)}{\lambda_t(j)} \tag{4b}$$

$$\partial D \quad : \quad \frac{1}{R_{D,t}} = Q_{t,t+1} \tag{4c}$$

These are the standard optimal conditions of the household. In particular, equations (4a) and (4b) link the discount factor to the intertemporal marginal rate of substitution of consumption. Since markets are complete and the discount factor is unique, the households can achieve perfect risk sharing and the consumption decision is the same for any j; we can rewrite (4b) simply as $Q_{t,t+1} = \beta \frac{{C_{t+1}}^{1-\sigma}}{{C_t}^{1-\sigma}} \frac{P_t}{P_{t+1}}$. Equations (4b) and (4c) define the standard consumption Euler equation. Note that (4c) holds for any possible state of the world and, therefore, it must hold in expectations too; this define the deposits rate as a risk-free interest rate: $\frac{1}{R_{D,t}} = \mathbb{E}_t Q_{t,t+1}$. In equilibrium, we will assume that this rate coincides with the policy rate set by the Central Bank.

In order to solve for the equilibrium wage, we assume nominal wage rigidities *à la* Calvo in the wage setting of the households. The probability that an household j is not allowed to adjust his wage at any period t is $\theta_w < 1$; if not allowed to adjust, the worker keeps the wage of period $t-1$. Since they face the same problem, the households allowed to adjust their wages pick the same optimal wage W_t^*, which implies

[2] In obtaining (3), use is made of the following two definitions

$$N_t(j) = \int_0^1 N_t(i,j)\, dj \text{ and } N_t = \int_0^1 N_t(i)\, di$$

that the average wage is $W_t = \left[(1-\theta_w) W_t^{*1-\phi_w} + \theta_w W_{t-1}{}^{1-\phi_w}\right]^{\frac{1}{1-\phi_w}}$. The optimal wage must solve the life utility maximization problem modified to take into account the future probabilities of not adjusting. The household problem becomes

$$\max_{w_t(j)} \mathbb{E}_t \sum_{k=0}^{\infty} (\beta\theta_w)^k u(C_{t+k}(j), N_{t+k}(j))$$

subject to the labor demand in (3) and the budget constraint (2). The solution is

$$W_t^* = \frac{\phi_w}{\phi_w - 1} \frac{\mathbb{E}_t \sum_k (\beta\theta_w)^k N_{t+k}(j)^{\eta+1}}{\mathbb{E}_t \sum_k (\beta\theta_w)^k N_{t+k}(j) C_{t+k}^{-\sigma} P_{t+k}^{-1}} \tag{5}$$

3.2 Firms

3.2.1 Final Good

The final good producer bundles the intermediate goods for the final consumption of households. The aggregation function displays a constant elasticity of substitution between intermediate goods ϕ_p

$$Y_t = \left[\int_0^1 Y_t(i)^{\frac{\phi_p-1}{\phi_p}} dj\right]^{\frac{\phi_p}{\phi_p-1}}$$

where Y_t is the final composite output and $Y_t(i)$ is the intermediate output of firm i.

From cost minimization, we obtain the demand for each variety i

$$Y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\phi_p} Y_t \tag{6}$$

and the corresponding price index $P_t = \left[\int_0^1 P_t(i)^{1-\phi_p} di\right]^{\frac{1}{1-\phi_p}}$

3.2.2 Intermediate Goods

There is a continuum of monopolistically competitive intermediate-good producers of mass 1. They employ a linear technology that uses labor as the sole factor of production

$$Y_t(i) = A_t N_t(i) \tag{7}$$

where A_t is exogenous productivity.

We assume that producers have to finance their working capital, that is, they need to pay part or all of their wage bill $W_t N_t(i)$ before the goods market opens. They do so by borrowing from intermediaries at the gross interest rate $R_{L,t}$. With the assumed technology, cost minimization implies that the real marginal costs are the same for each firm and given by:

$$MC_t = R_{L,t}^{\gamma} \frac{W_t}{P_t A_t}$$

where $\gamma \in [0,1]$ reflects the share of wage bill that must be financed.[3] Literature refers to this type of marginal cost as the *cost channel* of monetary policy.

Firm i's profits, $\mathbb{Y}_t(i)$, are paid to the households each period and are defined as

$$\mathbb{Y}_t(i) = P_t(i) Y_t(i) - R_{L,t}^{\gamma} W_t N_t(i)$$

where the labor demand $N_t(i)$ has been defined above.

The firm chooses its optimal price while taking the market prices as given. We assume Calvo-style nominal price stickiness (Calvo 1983): each period, any firm is able to optimally update its price with probability $1-\theta_p$ (independent of the time elapsed since the last adjustment). Firms that are not able to update leave their price unchanged. The firm's profits maximization problem is defined as

$$\max_{P_t(i)} \mathbb{E}_t \sum_{k=0}^{\infty} \theta_p^k Q_{t,t+k} \left[P_t(i) Y_{t+k}(i) - MC_{t+k} Y_{t+k}(i) \right]$$

subject to the demand for output i in (6). The solution to this problem is the price

$$P_t^* = \frac{\phi_p}{\phi_p - 1} \frac{\mathbb{E}_t \sum_k (\beta\theta_p)^k {C_{t+k}}^{-\sigma} MC_{t+k} P_{t+k}^{\phi_p}}{\mathbb{E}_t \sum_k (\beta\theta_p)^k {C_{t+k}}^{-\sigma} P_{t+k}^{\phi_p - 1}} \tag{8}$$

with a corresponding price index given by $P_t = \left[(1-\theta_p) {P_t^*}^{1-\phi_p} + \theta_p {P_{t-1}}^{1-\phi_p} \right]^{\frac{1}{1-\phi_p}}$.

3.3 Financial Intermediaries

Financial intermediaries in this model are banks that "supply" deposits to households and lend funds to firms to finance their working capital. Based on a simplified balance sheet of a bank h (and abstracting from net worth), we can write the following identity:

$$R_t(h) + L_t(h) = D_t(h)$$

[3] γ can be interpreted as a pass-through parameter from the working capital financing needs to the marginal cost.

where $R_t(h)$ are bank's reserves and $L_t(h)$ is the supply of loans. We assume banks do not hold excess reserves, and they are required to hold reserves as a fixed proportion θ of deposits, so that $R_t(h) = \theta D_t(h)$.[4] Hence:

$$L_t(h) = (1-\theta)D_t(h) \tag{9}$$

which also implies $L_t = (1-\theta) D_t$.

The earnings from the lending activity are paid to the household each period. These earnings are defined as

$$\mathbb{Y}_t(h) = R_{L,t}(h) L_t(h) - (1-\theta) R_{D,t} D_t(h)$$

by construction the total profits distributed by banks and firms to the households are given by $\mathbb{Y}_t = \int_0^1 \mathbb{Y}_t(i)\, di + \int_0^1 \mathbb{Y}_t(h)\, dh$.[5]

Following Henzel et al. (2009), we assume that banks provide differentiated loans to firms, so they have some monopoly power in setting their interest rates $R_{L,t}(h)$. The differentiation of bank loans can be justified with different types of micro-foundations, the most common of which is long-term lending relationships between banks and firms. The elasticity of substitution between loan types, ϕ_L, defines the amount of composite loan that each firm borrows from the bank intermediaries as

$$L_t(i) = \left[\int_0^1 L_t(i,h)^{\frac{\phi_L - 1}{\phi_L}}\, dh\right]^{\frac{\phi_L}{\phi_L - 1}}$$

which yields the total demand of bank h loan from all firms

$$L_t(h) = \left(\frac{R_{L,t}(h)}{R_{L,t}}\right)^{-\phi_L} L_t \tag{10}$$

in which $R_{L,t} = \left[\int_0^1 R_{L,t}(h)^{1-\phi_L}\, dh\right]^{\frac{1}{1-\phi_L}}$ is the market loan rate.[6]

Analogously to the wage and price setting problems, we assume that any bank is allowed to re-set its loan rate with probability $1-\theta_L$ in each period. If the bank is not re-setting, it keeps the loan rate from

[4] The assumptions adopted here define the text-book notion of the simplified money multiplier. We also assume that banks receive an interest rate payment on reserves from the Central Bank which, as explained in Section 3.4, in equilibrium is equal to the deposit rate $R_{D,t}$.

[5] For simplicity in the analysis, households receive even shares of profits at each period. Furthermore, in equilibrium total demand and supply of deposits must be equal too

$$\int_0^1 D_t(h)\, dh = \int_0^1 D_t(j)\, dj$$

.

[6] In obtaining (10), use is made of the following two definitions

$$L_t(h) = \int_0^1 L_t(i,h)\, di \text{ and } L_t = \int_0^1 L_t(i)\, di$$

the previous period. The optimal interest rate solves the problem

$$\max_{R_{L\,t}(h)} \mathbb{E}_t \sum_{k=0}^{\infty} \theta_L^k Q_{t,t+k} \left[R_{L,t}(h) L_{t+k}(h) - (1-\theta) R_{D,t+k} D_{t+k}(h)\right]$$

subject to the demand for loan h in (10) and the balance sheet identity (9). The solution to this problem is

$$R_{L,t}^{*} = \frac{\phi_L}{\phi_L - 1} \frac{\mathbb{E}_t \sum_k (\beta\theta_L)^k {C_{t+k}}^{-\sigma} L_{t+k}(h) R_{D,t+k}}{\mathbb{E}_t \sum_k (\beta\theta_L)^k {C_{t+k}}^{-\sigma} L_{t+k}(h)} \tag{11}$$

with a corresponding market loan defined by $R_{L,t} = \left[(1-\theta_L) {R_{L,t}^{*}}^{1-\phi_L} + \theta_L {R_{L,t-1}}^{1-\phi_L}\right]^{\frac{1}{1-\phi_L}}$.

3.4 Solving the model

The model is log-linearized around a symmetric steady state in which all households set the same wage, all intermediate good producers set the same price, and all banks set the same loan interest rate, with zero steady state inflation. The main paper reports the equations of the linearized model, but here we provide more details on its derivation. In what follows, lower case letters indicate to the log-deviations from the steady state of the corresponding upper case variables. Any exception in the notation will be individually pointed out.

Combining equations (4a)-(4c) in expectations defines the households' Euler equation

$$C_t^{-\sigma} = \beta R_{D,t} \mathbb{E}_t \left(\frac{C^{-\sigma}}{\Pi_{t+1}}\right)$$

where $\Pi_{t+1} = \frac{P_{t+1}}{P_t}$. In log-linear form this equation is

$$c_t = \mathbb{E}_t c_{t+1} - \frac{1}{\sigma}\left(r_{D,t} - \mathbb{E}_t \pi_{t+1}\right) + u_t$$

where inflation is defined as $\pi_t = p_t - p_{t-1}$ and an ad hoc demand side shock $u_t = \rho_u u_{t-1} + \varepsilon_{u,t}$, with $\varepsilon_{u,t}$ *iid*, is added to the equation.

From the production function in (7) for firm i, we integrate over the index i to obtain aggregate production. Since the TFP shock A_t is common across firms, the linear aggregate output is

$$y_t = a_t + n_t$$

where the exogenous process for the TFP shock is $a_t = \rho_a a_{t-1} + \varepsilon_{a,t}$, with $\varepsilon_{a,t}$ *iid*.

From staggered price setting in the three sectors of the economy we obtain the equations for price and wage inflation, and the loan rate dynamics. The derivation of these three equations follow similar steps. First, the optimal prices in equations (5), (8), and (11) are linearized to obtain

$$
\begin{aligned}
p_t^* &= (1-\theta_p\beta)\,\mathbb{E}_t\sum_{k=0}^{\infty}(\theta_p\beta)^k\,(mc_{t+k}+p_{t+k}) \\
w_t^* &= \frac{1-\theta_w\beta}{1+\phi_w\eta}\mathbb{E}_t\sum_{k=0}^{\infty}(\theta_\omega\beta)^k\,(\sigma c_{t+k}+\eta n_{t+k}+\phi_w\eta w_{t+k}+p_{t+k}) \\
r_{L,t}^* &= (1-\theta_L\beta)\,\mathbb{E}_t\sum_{k=0}^{\infty}(\theta_L\beta)^k\,r_{D,t+k}
\end{aligned}
$$

where $mc_t = \gamma r_{L,t} + (w_t - p_t) - a_t$ is the log-deviation of the marginal cost MC_t. Second, the average aggregate price equations are log-linearized to obtain

$$
\begin{aligned}
p_t &= (1-\theta_p)\,p_t^* + \theta_p p_{t-1} \\
w_t &= (1-\theta_w)\,w_t^* + \theta_L w_{t-1} \\
r_{L,t} &= (1-\theta_L)\,r_{L,t}^* + \theta_L r_{L,t-1}
\end{aligned}
$$

Finally, the optimal prices and the aggregate price dynamics are combined to obtain the inflation equations used in the main paper

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \kappa_\pi mc_t \tag{12}$$

$$\pi_t^w = \beta\mathbb{E}_t\pi_{t+1}^w + \kappa_w\,(\sigma c_{t+k} + \eta n_{t+k} - (w_t - p_t)) \tag{13}$$

$$r_{L,t} = \beta\kappa_L\mathbb{E}_t r_{L,t+1} + \kappa_L r_{L,t-1} + \kappa_D r_{D,t} \tag{14}$$

where $\pi_t^w = w_t - w_{t-1}$ is the wage inflation and the κ coefficients are functions of the deep parameters

$$
\begin{aligned}
\kappa_\pi &= \frac{(1-\theta_p)\,(1-\beta\theta_p)}{\theta_p} \\
\kappa_w &= \frac{(1-\theta_w)\,(1-\beta\theta_w)}{\theta_w\,(1+\phi_w\eta)} \\
\kappa_D &= \frac{(1-\theta_L)\,(1-\beta\theta_L)}{1+\theta_L^2\beta} \\
\kappa_L &= \frac{\theta_L}{1+\theta_L^2\beta}
\end{aligned}
$$

Equations (12) and (13) are standard New Keynesian Phillips curves for price and wage inflation under nominal rigidities *à la* Calvo (see for example Woodford, 2003 and Gali, 2008). Equation (14) shows that the loan rate depends on its future and past values with practically identical coefficients (since β is very close to one) and a markup over the deposit rate which represents the bank marginal cost. Note that, in contrast to the price and wage inflation equations, nominal rigidities in the setting of the interest rate are sufficient for the lagged term in $r_{L,t-1}$ to show up in (14).

The model is completed by the marketing clearing condition in the goods and loan markets and by a Taylor rule for the policy rate. The goods market clearing condition implies $y_t = c_t$. On the loan markets, total loan supply must be equal to total deposits from the aggregate balance sheet of the banks $l_t = d_t$. From the cost function of the firms and the funding needs of working capital, total loan demand is equal to the aggregate working capital of firms $l_t = w_t + n_t$. In equilibrium supply and demand of loans are equal.

Finally, to close the model, we assume that the monetary authority sets the policy interest rate according to a Taylor-type rule (see Taylor 1993). For sake of simplicity, we assume that in equilibrium the deposit rate paid by banks to households is equal to the policy rate. The Taylor rule can be expressed as

$$r_{D,t} = \rho_D r_{D,t-1} + \phi_\pi \pi_t + \phi_y y_t + v_t$$

where we potentially allow for some degree of interest rate smoothing (as governed by the parameter ρ_D) and the exogenous monetary shock follows an AR(1) processes $v_t = \rho_v v_{t-1} + \varepsilon_{v,t}$ with the usual $\varepsilon_{v,t}$ *iid*.

4 Calibration

4.1 Baseline scenario

The calibrated parameter values for the baseline scenario are reported in Table 1. Figure 3 shows the response function of the main variables of the model to a contractionary monetary shock for this calibration. It turns out that a relatively broad set of calibrations can generate a hump-shaped response of inflation in our setup, in that different combinations of coefficients return similar results both in qualitative and quantitative terms. In keeping with the idea of sticking to the standard model, we use typical values found in recent literature. Most of the parameters are relatively non-controversial, at least in the context of said literature. We set $\beta = 0.99$ to imply a steady-state annualized risk-less return of about 4% with quarterly data. We use $\sigma = 2$, which is somewhere in between the typical range of 1 to 5 for the coefficient of risk aversion in this family of utility functions. We assume $\eta = 2$, so that the elasticity of labor supply is $\frac{1}{2}$. The coefficients in the interest rate rule are set to $\phi_\pi = 1.5$ and $\phi_y = 0.2$ (in line with Taylor 1993). We experiment with different values for the interest rate smoothing parameter ρ_D between .4 and .8 and we select $\rho_D = .7$. We note here that values for this parameter in recent, related literature are found to be quite high (Steinsson 2008, Castelnuovo 2011).

Key coefficients for the calibration are the three stickiness parameters in the three sectors of the economy. The Calvo non-adjustment probability for the intermediate firms' price is $\theta_p = 0.75$, which gives an average duration of price spells of one year. While this value is on the high end, as argued by recent studies that use micro-level price data (e.g. Bils and Klenow 2004), it is, once again, a widely used standard. The wage rigidity is set at $\theta_w = .6$, again in the normal range of this parameter; while the elasticity of substitution between differentiated labor types is $\phi_w = 6$, which is in the lower end for this type of parameter. There is less guidance for the non-adjusting probability of the loan rate. In a model similar to ours, Henzel et al. (2009) estimate this parameter in the $.4 - .5$ range; therefore, we set $\theta_L . = .4$. The first-order autocorrelation of all exogenous shocks is set at .4.

Finally, a crucial parameter in our exercise is γ, which represents the share of wage bill to be paid in advance: we set $\gamma = .9$. A "full" cost channel would correspond to $\gamma = 1$; we can shut down this channel by setting $\gamma = 0$. We obtain some degree of hump in the inflation response with a relatively broad set of values for γ. Different combinations of γ and the other parameters are studied in the next section.

In Figure 3, we see that inflation response reaches its minimum after 4 or 5 quarters and it reverts to the initial point in about 3 to 4 years. The response is qualitative and quantitative in line with the empirical responses in Figure 1. In particular, the size and the relative response of the loan rate r_L and the policy rate r_D is consistent with the data. We can replicate the hump-shape of the inflation response without relying on unreasonable movements on the loan rate, which is the new key component of firms' marginal costs. The marginal cost and the real wage $w - p$ are illustrated in the last subplot of Figure 3. In the first few periods after the shock, the loan rate has a stronger response than the real wage, which drives the initial positive response of marginal costs and, importantly, the observed shape in the inflation response. This key mechanism of the model is fully explained in Section 4.3. Finally, the response of the price level is decreasing, which shows we do not need to generate a model-consistent price puzzle to obtain the desired shape of the inflation response. The response of consumption (equivalent to income in this closed economy model) is negative, as we should expect, but monotonic. It is known that a convex shape for its response (as observed in the data) is easily obtained with additional features in the model, such as habit formation in the utility function. We abstract from these features in order to isolate the contribution of the cost channel to inflation persistence from other sources of real rigidities (including this extra component in the model, however, does not alter our basic conclusions).

Table 1: Baseline calibration of main parameters

Discount factor	$\beta = 0.99$
Utility	$\sigma = 2;\ \eta = 2$
Taylor rule	$\rho_D = .7;\ \phi_\pi = 1.5;\ \phi_y = 0.2$
Price rigidities	$\theta_p = .75$
Wage rigidities	$\theta_w = .6;\ \phi_w = 6$
Loan rigidities	$\theta_w = .4$
Cost channel	$\gamma = .9$
Innovations AR(1)	$\rho_a = \rho_v = \rho_u = .4$

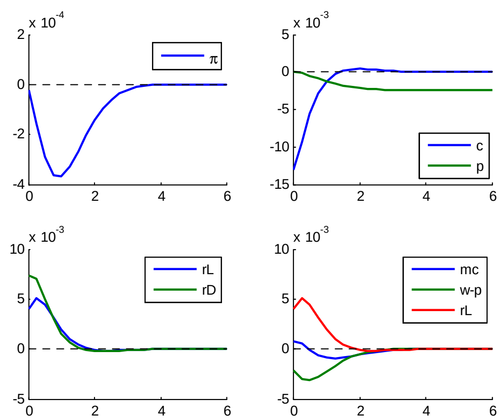

Figure 3: Response functions to a one s.d. monetary shock from the model with baseline calibration.
Notes: Years from the impulse on the $x-$axis.

4.2 Robustness checks

As a first robustness check, we study the implications of different values of γ (given the other parameters in the model). As we reduce the value of γ, the hump progressively disappears, along with the persistence of the inflation response. When the cost channel is shut down ($\gamma = 0$), the response behaves monotonically as in the standard Calvo model with pure nominal rigidities. This is illustrated in Figure 4; the response exhibits a hump for γ values as small as $.45 - .4$. For $\gamma = 0$ the half-life of the response from its trough is about 4 quarters, while under the baseline calibration it is roughly between 7 and 8 quarters. Thus, the increase in persistence due to the cost channel is very substantial, and it is achieved without adding any extra real rigidities to the original Calvo mechanism. We also note that this result does not require a large and implausible trough of the response on impact.

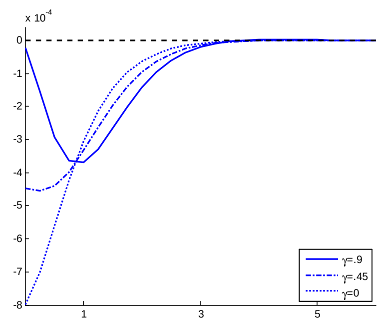

Figure 4: Inflation responses to a monetary shock. Years from the shock.
Notes: Changes in the response of inflation in function of γ. Years from the impulse on the $x-$axis.

Disentangling the effects on the results of the changes in the values of other parameters requires a better understanding of the determinants of the inflation response. Here, we explore two other combinations of parameters that yields similar results and we compare those calibrations to the baseline case. We examine the effects of individual parameters more thoroughly in Section 4.3, where we explicitly link the response of inflation to the response of the components of marginal costs and discuss the mechanism underlying the hump-shaped response.

The first alternative calibration considers the impact of higher price flexibility. The choice of $\theta_p = .75$ implies an average duration of price spells of one year, which we have argued might be too long, in light of more recent micro-level empirical evidence (e.g. Bils and Klenow 2004). Therefore, we lower θ_p to .5. We obtain very similar inflation responses by either lowering the share of working capital paid in advance or increasing the sluggishness of the loan rate. Figure 5 illustrates this point using $\gamma = .65$, but the same result could be obtained setting $\theta_L = .55$. The response of inflation is more negative, but still hump-shaped. The overall degree of persistence generated by the model decreases, as we would expect with less price stickiness.

The relative response of the interest rates, however, is very similar to the baseline case (compare to Figure 3).

The second alternative calibration explores the effects of more persistent innovations by increasing the AR(1) parameter of the monetary shock to $\rho_v = .6$. Focusing only on the inflation response, this change can be compensated by higher wages stickiness, for instance, or more flexible loan rates. However, reducing θ_L makes the response of the loan rate implausibly strong. On the other hand, a small increase in θ_w to .7 is enough to restore the results obtained under the baseline calibration. The impulse response functions for the latter parametrization are reported in Figure 6, where we also observe a very small increase of inflation on impact.

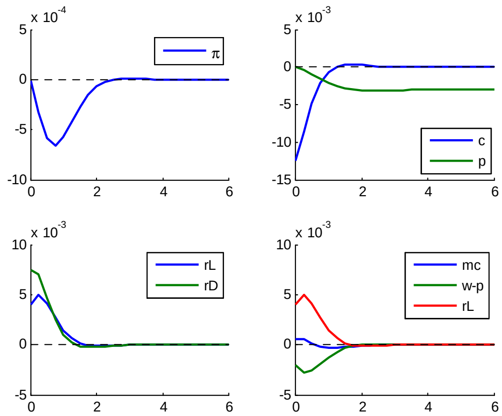

Figure 5: Response functions to a one s.d. monetary shock from the model with first alternative calibration. *Notes:* Calibration with $\theta_p = .75$ and $\gamma = .65$. Years from the impulse on the $x-$axis.

4.3 Discussion

In order to gain a more comprehensive understanding on the determinants of the inflation response in models of this kind, it is useful to write the structural inflation equation (12), the so-called New Keynesian Phillips Curve, in its closed form, conditional on marginal costs. Repeated substitution yields :

$$\pi_t = \kappa_\pi \sum_{i=0}^{\infty} \beta^i \mathbb{E}_t mc_{t+i} \tag{15}$$

which says that current inflation is determined by a weighted average of expected future real marginal costs. This solution makes clear that, in the context of the Calvo model of price stickiness (in its simplest form with nominal rigidities only), inflation is a purely forward looking variable. It also provides the basis for much of the criticism regarding the model's counterfactual implications for inflation dynamics (see for example

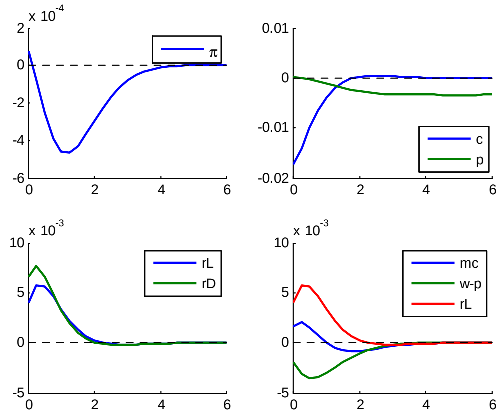

Figure 6: Response functions to a one s.d. monetary shock from the model with second alternative calibration.

Notes: Calibration with $\rho_v = .6$ and $\theta_w = .7$. Years from the impulse on the $x-$axis.

Mankiw 2001 for theoretical considerations, Gali and Gertler 1999 and Rudd and Whelan 2005, for empirical analyses based on the above specification).

Importantly, absent further modifications to the baseline model, inflation displays no "intrinsic" persistence, simply inheriting the persistence properties of its driving process mc.[7] As we have mentioned, another specific aspect of persistence that the basic model fails to match is the inertial, hump-shaped response to monetary policy shocks that is typical in VAR evidence. Our model looks in more detail at the determinants of firms' marginal cost to uncover an additional channel of the monetary transmission mechanism which can generate such a response. In a similar vein to the analysis of the real exchange rate in Steinsson (2008), we write (15) using our expression for marginal costs:

$$\pi_t = \kappa_\pi \sum_{i=0}^{\infty} \beta^i \mathbb{E}_t(\gamma r_{L,t+i} + w_{t+i} - p_{t+i} - a_{t+i}) \tag{16}$$

At each point in time, π_t is given by an infinite sum of terms. If we want its response to a (contractionary) monetary shock to be hump-shaped, we want the first few terms of the summation to be *positive* as a consequence of such shock. As time goes on and these terms drop out of the sum, π_t becomes more negative until it reaches its trough: the crucial point is that, because of this mechanism, the trough is reached later and *not* on impact. In order for this to happen, the response of the real marginal cost to a monetary shock must be itself hump-shaped *and* positive on impact and for a few periods.

We now have a framework to understand the model's impulse response functions shown above. Monetary policy shocks affect marginal costs through two channels: First, a typical *demand* channel that impacts

[7] See also Walsh (2003).

household's intertemporal decisions, in turn affecting labor supply (through the marginal rate of substitution between consumption and leisure) and thus real wages. Second, a *cost* channel that more directly affects marginal costs through the effect of monetary policy on the loan interest rate r_L; this can be defined as a supply side effect. At a basic level, we are looking for a combination of these two channels such that, on impact and for a few periods, the decrease in real wages $w - p$ is more than counteracted by an increase in r_L: according to (16), this would generate the "desired" response of marginal costs and hence inflation. Figure 3, for example, clearly shows this type of response for the mc and its two components.

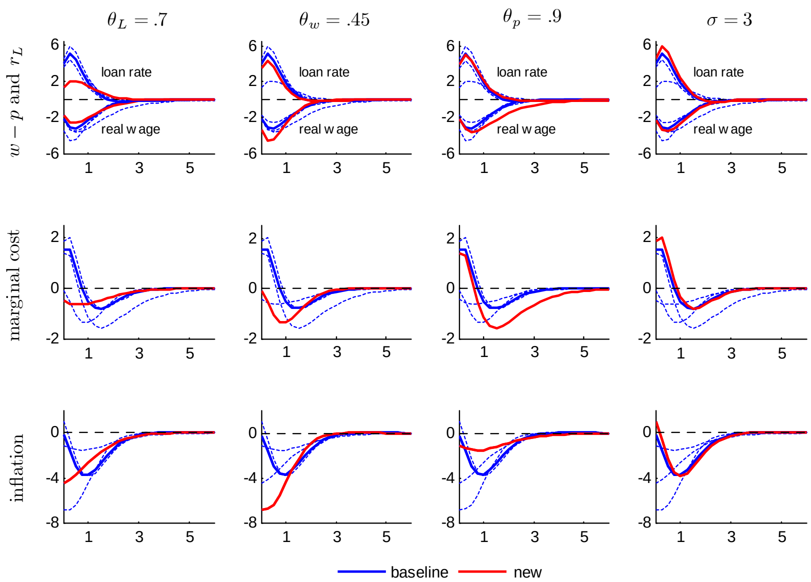

Figure 7: Calibration of selected parameters.
Notes: Blue line: responses under baseline calibration. Red lines: responses under new calibration. Each column (three plots) corresponds to the scenario indicated by the parameter value in the header. The scale of the response functions of inflation is 10^4; the scale of the other responses is 10^3. Years from the impulse on the $x-$axis.

Three aspects of this model are especially important for the resulting properties of the inflation response. First, we need some sort of sluggishness in the response of real wages. Without any kind of wage rigidity, the effect of the cost channel (for reasonable parameters) is not strong enough to counteract the traditional effect of policy shocks on marginal costs. However, even a moderate degree of wage rigidity can produce a hump-shaped response of inflation. Moreover, and importantly, it's worth noting that nominal wage rigidities *alone* cannot produce this type of result within the Calvo model, as made clear in equation (16): the response of marginal costs (and thus inflation) will be more sluggish, but all terms in the summation will be negative.

Second, the stickiness of the loan interest rate matters. It is possible to intensify the effects of the cost channel by increasing the flexibility of the loan rate setting mechanism (decreasing θ_L), and even obtain responses that include an initial model-consistent price puzzle. However, some degree of stickiness in the response of the loan rate to monetary shocks relative to the response of the policy rate seems to be more in line with the empirical evidence provided by the VAR. As explained above, the goal of our model is providing a mechanism that generates higher inflation persistence through a more realistic hump-shaped response to monetary shocks rather than using the cost channel to provide theoretical justifications of the price puzzle.

Third, we can relate these results to the role of strategic complementarities/real rigidities for the dynamics of inflation. It is well understood in the literature that, given reasonable degrees of *nominal* rigidity, it is hard to generate inertial responses of inflation (and thus significant real effects of monetary policy). Hence, mechanisms need to be introduced that slow down price adjustment beyond the effects of nominal stickiness, that is, mechanism that introduce strategic complementarity in pricing (or real rigidities). Examples include heterogeneous labor markets, firm specific capital, and non-constant elasticities of demand. In the context of the class of models we analyze here, the elasticity of marginal costs to output is the relevant parameter to determine whether price setting is characterized by strategic complementarity or, rather, strategic substitutability.[8] Given our assumptions (specifically that of homogeneous labor market), the elasticity in this model is $\zeta = \eta + \sigma$, which is greater than 1 (thus implying strategic *substitutability*) with our calibration. In this sense, our results do not need to rely on a high degree of strategic complementarity.

In light of the above discussion, we conclude by providing a brief analysis of the effects on the results of different calibrations of individual parameters. Figure 7 summarizes this analysis. In general equilibrium, each parameter can have a potentially large impact on both components of the marginal cost. With this exercise, we focus on a subset of the parameters of the model that we regard as especially relevant. The figure compares responses of the real wage and loan rate (first row), the marginal cost (second row), and inflation (third row) in four different cases. We change one parameter at a time, as represented by each of the four columns in the figure. In each scenario, all the remaining parameters are calibrated at their baseline values. A blue line is used to represent the responses for the baseline calibration, while red lines show the response under the new parameter value. The scale of the response functions of inflation is 10^4; the scale of the other responses is 10^3.

In the first scenario, we increase θ_L to .7 which dampens the response of r_L. The response of the loan rate in the first row becomes flatter; $w - p$ is slightly smaller too, but the reduction of r_L dominates. The marginal cost is entirely negative now, and the inflation response doesn't exhibit a hump-shape. Decreasing θ_w to .45 makes for a stronger response of $w - p$ and reinforces the demand channel. This effect dominates over the smaller change in the loan rate; the marginal cost is negative and, again, we do not have a hump-shaped response. Setting $\theta_p = .9$ increases the sluggishness of $w - p$, without a noticeable effect on the loan

[8] This is what Woodford (2003) refers to as the elasticity of the *notional* Short-Run Aggregate Supply (SRAS).

rate. The marginal cost still switches sign after a few periods and, therefore, the hump of π is preserved. However, the change in θ_p also reduces the slope of the Phillips Curve κ_π, which in turn dampens the shape of the inflation response. For high values of θ_w and low values of θ_p, the cost channel dominates on the demand channel, to the point where we have both a hump-shaped response of inflation and an initial price puzzle. Finally, higher risk aversion $\sigma = 3$ implies marginally larger initial responses of the interest rates, r_D and r_L, due to a smaller intertemporal substitution in consumption. This change generates a very mild price puzzle, as illustrated by the fourth scenario in the figure. Higher η similarly causes small changes in the real wage $w - p$ and and a mild price puzzle.

REFERENCES

Barth, Marvin J., and Valerie A. Ramey. 2001. "The Cost Channel of Monetary Transmission." *NBER Macroeconomics Annual:* 199-239.

Bils, Mark, and Peter J. Klenow. 2004. "Some evidence on the importance of sticky prices." *Journal of Political Economy,* 112(5): 947-985.

Calvo, Guillermo A. 1983. "Staggered contracts in a utility maximizing framework." *Journal of Monetary Economics,* 12(3): 383-398.

Carvalho, Carlos. 2006. "Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks." *B.E. Journal of Macroeconomics,* 2(1), Frontiers.

Castelnuovo, Efrem. 2011. "Testing the Structural Interpretation of the Price Puzzle with a Cost Channel Model." *Oxford Bulletin of Economics and Statistics,* 74(3): 425-452..

Christiano, Lawrence J., and Martin Eichenbaum. 1992. "Liquidity Effects and the Monetary Transmission Mechanism." *American Economic Review,* 82(2): 346-353.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary Shocks: What Have We Learned and to What End?" *In Handbook of Macroeconomics,* Edited by John B. Taylor and Michael Woodford. Amsterdam: Elsevier Science.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary." *Journal of Political Economy,* 113(1): 1-45.

Dotsey, Michael, and Robert G. King. 2005. "Pricing, Production, and Persistence." Fed Bank of Philadelphia WP No. 05-4.

Gali, Jordi. 2008. *Monetary Policy, Inflation and the Business Cycle.* Princeton University Press.

Gali, Jordi, and M. Gertler. 1999. "Inflation dynamics: A structural econometric analysis." *Journal of Monetary Economics,* 44(2): 195-222.

Henzel, Steffen, Oliver Hülsewig, Eric Mayer, and Timo Wollmershäuser. 2009. "The price puzzle revisited: Can the cost channel explain a rise in inflation after a monetary policy shock?" *Journal of Macroeconomics,* 31(2): 268-289.

Lown, Cara, and Donald P. Morgan. 2006. "The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey." *Journal of Money, Credit, and Banking,* 38(6)*:* 1575-1597.

Mankiw, N.Gregory. 2001. "The inexorable and mysterious tradeoff between inflation and unemployment." *The Economic Journal,* 111 (May): C45-C61.

Mankiw, N.Gregory and Ricardo Reis. 2002. "Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve." *Quarterly Journal of Economics,* 117(4): 1295-1328.

Rabanal, Pau. 2007. "Does Inflation Increase After a Monetary Policy Tightening? Answers Based on an Estimated DSGE Model." *Journal of Economic Dynamics and Control,* 31(3): 906-937.

Ravenna, Federico, and Carl E. Walsh. 2006. "Optimal Monetary Policy with the Cost Channel." *Journal of Monetary Economics,* 53(2): 199-216.

Rudd, Jeremy, and Karl Whelan. 2005. "New tests of the New-Keynesian Phillips curve." *Journal of Monetary Economics,* 52(6): 1167-1181.

Steinsson, Jon. 2008. "The Dynamic Behavior of the Real Exchange Rate in Sticky Price Models." *American Economic Review,* 98(1): 519-33.

Taylor, John B. 1993. "Discretion versus policy rules in practice." *Carnegie-Rochester Conference Series on Public Policy,* 39: 195-214.

Walsh, Carl E. 2003. *Monetary Theory and Policy.* MIT Press.

Woodford, Michael. 2003. *Interest and Prices.* Princeton University Press.