The cost channel and the persistence of the inflation response to a monetary policy shock.

Andrea Civelli and Nicola Zaniboni*

Very Preliminary Draft - Please Do Not Circulate Without Authors’ Permission.

September 12, 2011

Abstract

The goal of this paper is to generate higher inflation persistence in a DSGE model with Calvo price setting. We explore the cost channel, whereby the marginal cost of firms is affected by the loan rate since firms have to finance their working capital in advance. The results hinges on the loans supply function of banks being more responsive to the output gap relative to the loan rate.

JEL:

Keywords: Inflation Persistence, Marginal Cost Channel,

The persistence of inflation lies at the heart of any monetary policy consideration. Monetary actions have stronger consequences on the real economy the higher the persistence of inflation is. Inflation persistence observed in aggregate data is higher than that generated by the typical micro-founded models used in optimal policy analysis. Disentangling the sources of this persistence is important in order to improve our understanding of the transmission mechanism of monetary policy and of the effects of policy decisions.

Persistence can be defined and studied from different perspectives [more here on empirical definitions of persistence]. In this paper, we focus on one particular dimension of persistence related to the response of inflation to monetary policy shocks.

In a new Keynesian model with Calvo price setting, the workhorse model in the macroeconomic theoretical literature on monetary policy, inflation is purely forward looking and simply inherits the persistence properties of its driving process, be it real marginal costs (as predicted by theory), or more generally some measure of real activity (interpreted as an observable proxy of real marginal costs). Moreover, and importantly, inflation responds monotonically to a monetary shock, peaking on impact. This is strongly at

*Economics Department, University of Arkansas, Business Building 402, Fayetteville, AR 72701. E-mail: andrea.civelli@gmail.com. I am particularly indebted to.... I am also grateful to seminar participants at UARK for their helpful suggestions and comments.
odds with the empirical evidence based on the VAR approach to the identification of monetary shocks, in which inflation responds following the characteristic hump-shaped trajectory. These observations have led to several modifications of the standard new Keynesian approach to the Phillips Curve.

The goal of our paper is to explore the importance and implications of the cost channel for the response of inflation to these shocks in the context of the standard Calvo pricing framework. Firms are assumed to have to finance their working capital in advance by borrowing liquidity from the bank system. The cost channel, then, introduces a dependence of the production costs on the cost of finding funds in the credit markets, the loan rate. Intuitively, this credit mechanism would have a supply side effect on the transmission of monetary policy that combines with the traditional demand-side channel, thus allowing the Calvo pricing mechanism to reproduce the "hump" of empirical responses. Practically, this is obtained by introducing a mechanism that allows for negative responses of the marginal cost to a policy shock at least on impact.

Our results show that this goal can be achieved by adding a loan market to the standard new Keynesian model. A banking sector collects deposits from households and supplies funds to the firms according to a loan supply function that depends on the federal funds rate, the output gap, and the loan rate. The key in the parametrization of the model is the relative magnitude of the elasticities of the loan supply to output and loan rate. The model provides a simple way to introduce the credit market and obtain the desired hump-shape features of the responses for a large set of plausible calibrations of the loan supply function.

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) and Castelnovo (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.

While related to this literature, our modeling approach is quite different. We explicitly build on a very interesting, and slightly older, literature that hasn’t had many applications in the New Keynesian context, and whose main ideas can be found in Bernanke and Blinder (1988), Blinder (1987), and Blinder and Stiglitz (1983). In particular, Bernanke and Blinder present a modification of the basic IS/LM model in which loans and bonds are not longer perfect substitutes and credit markets play an active role in the determination of the equilibrium along the IS curve. We follow them in the way we model the asset side of banks and the
supply of loans, but we limit their distinction between loans and bonds to its effects on the marginal cost.

In this paper, we do not want to achieve a higher inflation persistence at the expense of the Calvo price setting framework. This is an interesting challenge for at least two reasons. First, the modeling and algebraic elegance of this approach have made Calvo-style nominal stickiness, in its "purest" form, what is arguably the most popular alternative in the modern macroeconomic DSGE literature for optimal monetary policy analysis. Second, while numerous other solutions have been proposed to address the problem of inflation persistence (or lack thereof), 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 [add more references here]. 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 produce 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.

We argue that using the Calvo framework, with the introduction of some more realistic features absent in the basic model, we can obtain the "desired" shape for inflation response. Our paper is obviously not in contradiction with the works mentioned above, neither can it be considered a final explanation of the inflation persistence. However, we provide a sharp way to reinterpret and fix this flaw of the Calvo pricing that would definitely enhance the understanding of this problem.

The remaining of the paper is organized as follows: [to be added].

1 The VAR Analysis

The empirical response of inflation to monetary policy shocks is often at odds with the implications of macroeconomic theory. Since Sims (1991), passing through Christiano, Eichenbaum, and Evans (1999 and 2005), a very extended empirical VAR literature has documented two main features of the inflation response both for the U.S. and various other countries. The first is known as price puzzle and indicates a positive

\[1\] Here a good sequence of papers that are most representative of this literature.
response of prices and inflation to a contractionary monetary policy shock. The second is a high persistence of the response of inflation that is hardly matched by the standard New Keynesian model with Calvo pricing.

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. On the contrary, the standard New Keynesian Phillips Curve implies a monotonic response to a policy shock and, as a consequence, generating inflation dynamics as persistent as the empirical estimates is more difficult for this model.

In this section, we estimate a VAR model with seven U.S. macroeconomic variables to study the dynamic response of these variables to monetary policy shocks. We follow the empirical monetary VAR literature, and specifically Christiano et al. (2005), in setting up the VAR and the identifying assumptions of the monetary shock. The variables of interest included in the VAR model are as follows: the inflation rate of the commodity price index, the real GDP, the GDP deflator inflation rate, the federal fund rate, the bank prime loan rate, real bank commercial loans, and real government securities held by commercial banks.

We add to the standard monetary VAR three variables that are important for the theoretical policy transmission mechanism described below. The prime loan rate enters the marginal cost of firms through the cost channel; this is based on the necessity of firms to borrow funds to finance their working capital. Bank commercial loans and government securities respectively reflect the equilibrium on the loans market and the richer modeling of banks’ balance sheets introduced by the working capital needs.

The source for the entire set of variables is the Federal Reserve Economic Data (FRED), the online dataset of 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. Loans and government securities are transformed in real terms by dividing them by the GDP deflator index. We then take the logs of the GDP, loans, and securities. 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 variance-covariance matrix of the VAR residuals. The baseline ordering of the variables in the VAR adopted to identify 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 prime loan rate. The other two credit variables are in the last group, with government securities last and bank loans second-to-last. We have a three-block structure then, with the fed funds rate following the main real sector variables and the
credit sector variables in the last block.

We see this ordering consistent also 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 one lag of delay. This assumption seems to be fair during normal economic periods, it may be arguable when the economy is subject to large, negative financial shocks and the monetary policy promptly reacts to prevent the spreading of panic and liquidity crisis. The most obvious case would be the Great Recession period and, for this reason, we exclude the last three years of data from the sample used 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 crisis, but they are definitely shorter and less dramatic. We further discuss this point in the robustness checks part below.

For the moment, the results reported are based on standard least-square methods to estimate the VAR parameters and the variance of its residuals. The VAR is estimated with quarterly data over the sample 1979:4 to 2008:2 and four lags are included. The response functions to a one standard deviation positive monetary policy shock, corresponding to a contractionary shock, are reported in Figure 1. The horizon of the responses is 20 quarters; the standard error bands are computed by the Monte Carlo integration procedure of EViews with 1000 repetitions and correspond to the dashed lines in the figure.

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

- The response function of the inflation rate displays both a mild price puzzle and the typical hump-shape. The introduction of the commodity price inflation in the VAR allows to restrain the initial positive response of the inflation rate that would have been otherwise much stronger. Since the shock is contractionary, the hump of the response is turned upside down and reaches its trough after about two years. The spike shown at the seventh quarter is quite peculiar and is also found in other specifications of the VAR. From the point of view of the response of inflation, our results are perfectly in line with those in the previous literature.

- The federal funds rate takes about two years to return to the pre-shock level. It is very closely followed by the prime loan rate whose response, however, tends to be slightly larger.

- Also the response of the GDP is inversely hump-shaped. The GDP falls initially for one year, it remains quite low for another year, before reverting back to the pre-shock level.

- Bank loans and bank held government securities show an almost perfectly symmetric behavior. The response of loans is hump-shaped and peaks after less than two years, then it slowly reverts toward
Response to Cholesky One S.D. Innovations ± 2 S.E.

Figure 1: Response functions to a contractionary monetary shock.

Notes: Base ordering. Quarters from the impulse on the $x$–axis. The dashed-lines are the standard error bands.
zero and falls into the negative region in the longer run. The response of government securities displays an inverse hump and it is less persistent than that of loans. The trough is around one year and it moves into the positive region after four years. We will discuss below how this response of the loans is puzzling from the point of view of our theoretical model.

1.1 Some Robustness Checks

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 six lags are selected, the initial response of inflation is more positive. For six and eight lags the response of the government securities gets positive faster. The standard optimal lag length tests point to quite spread out number of lags. Eventually, four lags seems to be a satisfactory mid point and this is also the choice in Christiano et al. (2005).

The results are sixthly 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 1959 to 2008, however, the output essentially does not change. Nevertheless, there is a noticeable increase of the persistence of GDP and inflation and a more remarkable price puzzle. If the sample is truncated to 1979, then the price puzzle becomes a very evident characteristic of the data. By contrast, the initial response of loans is flat and then turns negative after only one year, while the response of government securities follows a similar path and turns positive after about six quarters. These last two responses would definitely find a more consistent theoretical explanation. Finally, no substantial differences are observed when more recent sub-samples are considered.

Furthermore, the results are robust to different transformations of the series. Instead of taking logs, for instance, we repeat the same exercise using de-trended series where the trend is computed by applying an HP filter with parameter set at 1600. This kind of transformation of the data is meant to provide some evidence at the business cycle frequency of the data. We do not find any particular change, except for an even more marked spike in the response of inflation. We also replace the GDP deflator inflation with the CPI inflation. The importance of the price puzzle, even in the baseline specification, increases and the shape of the inflation response loses some of its clarity because it becomes more jagged and it reverts more quickly toward zero after it reaches the trough.

We conclude by checking the robustness of the identification scheme. Many alternative orderings are obviously feasible, but we focus our attention mostly on the position of the credit variables 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 we recompute the impulse response functions ordering loans, government securities, and prime rate fourth to sixth. The response of inflation are strongly affected by the adoption of the new ordering. The price puzzle clearly emerges across different specification of the VAR and sub-samples, but the most important difference is that, for our main sample, the negative portion of the response function is very reduced and weak. This problem is partially corrected estimating the VAR over the full sample, but even in this case the negative response of inflation does not come before two or three years.

This ordering does not allow for contemporaneous responses of the credit variables to the monetary shock. On the other hand, the commercial credit sector is contemporaneously affecting the policy maker’s decision. This may correspond to an economy in which some kind of strong market frictions do not allow a prompt clearing of the loans market and the Federal Reserves actively responds to the shocks in this market. This is a very different view of the economy compared to that underlying our base ordering because the causality of the relation between fed funds rate and credit markets is completely inverted.

Furthermore, the key element of the ordering is the position of the prime loan rate relative to the fed funds rate, regardless of the relative position of loans and securities. In fact, moving the loan rate to the last position, while keeping loans and securities fourth and fifth, respectively, 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 credit in the market in order to set its policy, but that, at the same time, the prime loan rate immediately responds to changes of the funds rate.

Even though they focus more on the macroeconomic role of bank standards, Lown and Morgan (2006) order the standards and the commercial loans variables after the fed funds rate. They also add other credit variables and, in particular, the prime loan rate always to the last block of the ordering. As all these variables are financial in nature, we regard this approach as perfectly acceptable and we follow it in selecting our core identifying strategy.
2 The model

2.1 Households

Assuming complete financial markets, the representative household chooses consumption $C_t$, labor $N_t$ and money balances $D_t$ to maximize lifetime utility

$$E_0 \sum_{t=0}^{\infty} \beta^t u(C_t, N_t, D_t/P_t)$$

As explained below, money in this model is represented by deposits held at banks, hence the notation $D_t$. Preferences are described by the following isoelastic period utility function, separable in its three arguments:

$$u(C, N, D/P) = C^{1-\sigma} \frac{1}{1-\sigma} - N^{1+\phi} \frac{1}{1+\phi} + \left(\frac{D/P}{1-\nu}\right)^{1-\nu}$$

where $\sigma$ is the inverse of the intertemporal elasticity of substitution, $\phi$ is the inverse of the Frisch elasticity of labor supply, and $\nu$ is a parameter related to the interest elasticity of money demand. Utility is maximized subject to a flow budget constraint that reflects market completeness:

$$P_t C_t + E_t \{Q_{t,t+1} A_{t+1}\} + D_t = W_t N_t + A_t + D_{t-1} + T_t + \Pi_t$$

where $A_{t+1}$ is the state-contingent payoff of the portfolio at the beginning of period $t+1$ and $Q_{t,t+1}$ is the relevant stochastic discount factor. $W_t$ is nominal wage, $T_t$ are net transfers from the government, and $\Pi_t$ denotes profits from firms and banks (more on this below). The relevant first order conditions are as follows:

$$Q_{t,t+1} = \beta \frac{C_{t+1}^{-\sigma}}{C_t^{-\sigma}} \frac{P_t}{P_{t+1}}$$

$$E_t Q_{t,t+1} = \frac{1}{1+f_t} \Rightarrow \frac{1}{1+f_t} = E_t \left\{ \beta \frac{C_{t+1}^{-\sigma}}{C_t^{-\sigma}} \frac{P_t}{P_{t+1}} \right\}$$

$$W_t = \frac{N_t^\phi}{C_t^{-\sigma}}$$

$$\frac{(D_t/P_t)^{-\nu}}{C_t^{-\sigma}} = \frac{f_t}{1+f_t}$$

These are the standard optimal conditions of the household. In particular, equation (2a) defines the discount factor as the intertemporal marginal rate of substitution of consumption and equation (2b) is the Euler
equation. Equation (2c) defines the equilibrium wage in absence of any kind of friction in the labor markets\textsuperscript{2} and, finally, equation (2d) is a Fisher relation. Note that $1 + f_t = \frac{1}{\pi_t Q_{t,t+1}}$ is the (gross) nominal risk-free interest rate, i.e. the return on a one-period discount bond. Given the complete markets assumption, it is redundant to explicitly have a risk-free bond in the budget constraint (1). However, we also assume that such a bond exists in the economy and is issued by the government. For the sake of simplicity, we assume that $f_t$ coincides with the interest rate set by the monetary authority and the bonds market is modeled as residual component in the contest of the credit market (more about this in section 2.3). Deposits are assumed to yield no interest.

\subsection*{2.2 Firms}

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

$$Y_t = A_t N_t$$

where $A_t$ is exogenous productivity.

Labor is hired from a homogeneous, economy-wide market. We assume Calvo-style nominal price stickiness (Calvo (1983)): each period, any firm is able to optimally update its price with probability $1 - \theta$ (independent of the time elapsed since the last adjustment). Firms that are not able to update leave their price unchanged. This form of nominal rigidity leads to the familiar (log-linear) relationship between inflation $\pi_t$, expected inflation next period $E_t \pi_{t+1}$ and firms’ real marginal costs $mc_t$ (see for example Woodford (2003))\textsuperscript{3}:

$$\pi_t = \beta E_t \pi_{t+1} + \lambda mc_t$$

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

We assume that producers have to finance their working capital, that is, they need to pay their wage bill $W_t N_t$ before the goods market opens. They do so by borrowing from intermediaries at the gross interest rate $1 + i_t$. With the assumed technology, cost minimization implies that firms’ real marginal costs are given by:

$$MC_t = \frac{(1 + i_t) W_t}{P_t A_t}$$

\textsuperscript{2}We will introduce a degree of real wage rigidity in the section 2.4.

\textsuperscript{3}Here and in what follows, lower case notation denotes logarithmic deviations from the (zero-inflation) steady state.
The firms’ aggregate equilibrium profits, $\Upsilon_{1,t}$, are paid to the household each period and are simply defined as

$$\Upsilon_{1,t} = P_t Y_t - (1 + i_t) W_t N_t$$

2.3 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 representative bank (and abstracting from net worth), we can write the following identity:

$$D_t = R_t + L^S_i + B_t$$

where $R_t$ are bank’s reserves, $L^S_i$ is the supply of loans and $B_t$ are government (one-period) bonds, that yield gross return $1 + f_t$. We assume banks do not hold excess reserves, and they are required to hold reserves as a fixed proportion $\theta_R$ of deposits, so that $R_t = R^R_i = \theta_R D_t$. It is also clear from this balance sheet and the assumption made about deposit holdings, that we are assuming that the household does not hold cash.\(^4\)

Hence:

$$B_t + L^S_i = (1 - \theta_R)D_t$$

Following Bernanke and Blinder (1988), we postulate the following reduced-form functions for banks’ demand for government bonds and banks’ supply of loans:

$$B_t = B(f_t)(1 - \theta_R)D_t$$

$$L^S_i = L(i_t, f_t, Y_t)(1 - \theta_R)D_t$$

where $\frac{\delta B}{\delta f}$, $\frac{\delta L}{\delta f}$, $\frac{\delta L}{\delta Y} > 0$ and $\frac{\delta L}{\delta Y} < 0$. These functions can be rationalized in terms of bank’s portfolio decisions regarding its assets. Note that the loan supply depends negatively on the yield of the alternative investment, namely government bonds (which is also the "cost of money", or reserves, in this model).

\(^4\)The assumptions adopted here define the textbook notion of the simplified money multiplier. A version of the model with excess reserves would read

$$D_t = R_t + L^S_i + B_t$$

$$R_t = R^R_i + R^x$$

$$R^R_i = \theta_R D_t$$

and would allow to add to the model further aspects related to the excess reserves management of banks. In the actual data, however, excess reserves, $R^x$, represent only a small portion of the asset side of a bank balance sheet in normal time and, in a version of the model that includes also these excess reserves, we find the same kind of insight.
Finally, the loan supply depends positively on the level of output; this is a stylized way to capture the bank’s willingness to lend as a function of the state of the economy.\(^5\)

The earnings from the lending activity and the returns from the government bonds in portfolio are paid to the household each period. Assuming constant operative costs for the banks, which are standardized to zero, the earnings are simply defined as

\[ Y_{2,t} = i_t W_t N_t + f_t B_t \]

The model is closed by the government sector, which is modeled residually. The government makes a net transfer payment, \(T_t\), to the household and issues its bonds, \(B_t\), to cover this transfer and the service of the previous period debt. Its budget equations is

\[ T_t = B_t - (1 + f_{t-1}) B_{t-1} \]

since \(B_t\) and \(f_{t-1}\) are set by banks and the monetary authority respectively, this equation define \(T_t\) in a residual way.

### 2.4 Solving the model

After imposing market clearing, \(Y_t = C_t\), and equilibrium in the market for loans, noting that loan demand is simply \(L^D = W_t N_t\), we log-linearize the model around a zero-inflation steady state. The demand side of the economy is summarized by a forward looking IS-type relation\(^6\)

\[ c_t = E_t c_{t+1} - \frac{1}{\sigma} (\hat{f}_t - E_t \pi_{t+1}) \]

while the supply side is, as reported above, described by a Phillips curve of the form

\[ \pi_t = \beta E_t \pi_{t+1} + \lambda mc_t \]

where, importantly, \(mc_t = i_t + w_t - p_t - a_t\): real marginal costs are directly affected by the nominal (loan) interest rate. Regarding the real side of the model, we follow Blanchard and Gali (2007) and introduce real wage rigidities in the form of a wedge between the observed real wage and the value that would arise according to households’ optimal intratemporal choice between consumption and labor, \(\sigma c_t + \phi n_t\). More\(^7\)

\(^5\) As before, and with a similar rationale, we could add a demand function for excess reserves \(R^X_t = R(f_t, Y_t)(1 - \theta R)D_t\).

\(^6\) We can append a shock to this relation and interpret it, for example, as a shock to preferences.

\(^7\) The "hat" notation \(\hat{f}_t\) stands for \(\frac{\log(1+i)}{\log(1+r)}\). The same is of course true for \(\hat{f}_t\), the other interest rate variable in the model.
specifically, real wage evolves according to

$$w_t - p_t = \gamma(w_{t-1} - p_{t-1}) + (1 - \gamma)(\phi n_t + \sigma c_t)$$

where $\gamma \in [0; 1]$. As in Blanchard and Gali (2007), these real imperfections of the labor markets are left unmodeled, with the understanding that they can be obtained from a micro-founded model of wage setting.\(^8\)

As for the credit market, we have\(^9\)

$$(1 - \theta_R)r_t^R = \frac{l}{d}(w_t + n_t) + \frac{b}{d}b_t$$

where we use the fact that $r_t^R = d_t$. Accordingly, money demand is given by

$$r_t^R - p_t = \frac{\sigma}{\nu} c_t - \frac{1}{\nu f} \hat{f}_t$$

Equations for banks’ loan supply and demand and for bond demand are as follows

$$l^S_t = L_i \hat{i}_t + L_{f} \hat{f}_t + L_{g} c_t + r_t^R$$

$$l^D_t = w_t + n_t$$

$$b_t = B_{f} \hat{f}_t + r_t^R$$

where the $L_i$ and $B_{f}$ coefficients are elasticities of the aforementioned functions with respect to the relevant arguments. To close the model, we assume that the monetary authority sets the interest rate $\hat{f}_t$ according to a Taylor-type rule (see Taylor (1993)):

$$\hat{f}_t = \rho_f \hat{f}_{t-1} + (1 - \rho_f)(\phi_x \pi_t + \phi_y c_t) + \varepsilon_t$$

where we potentially allow for some degree of interest rate smoothing (as governed by the parameter $\rho_f$).

Finally, we specify the evolution of the exogenous technology and monetary shocks as AR(1) processes

$$a_t = \rho_a a_{t-1} + \eta_{a,t}$$

$$\varepsilon_t = \rho_{\varepsilon} \varepsilon_{t-1} + \eta_{\varepsilon,t}$$

\(^8\)It must be noted that, since we are interested in the descriptive behavior of the Phillips curve and of inflation in terms of marginal costs, this doesn’t have particular implications for this model. On the other hand, if one is interested in making optimal policy analysis within this framework, it would be important to know the specific implications of this assumption for the relationship between marginal costs and output gap.

\(^9\)Lower case variables with no time subscript denote steady-state values.
where \( \eta_{a,t} \) and \( \eta_{\epsilon,t} \) are zero-mean i.i.d. random variables with variance \( \sigma^2_\eta \) and \( \sigma^2_\epsilon \), respectively (variance equal to 1).

### 3 Calibration

The calibrated parameter values are reported in Table 1. In keeping with the idea of sticking to the standard model, we use typical values found in recent literature. Most of them are relatively non-controversial, at least in the context of said literature. We set \( \beta = 0.99 \) to imply a steady-state annualized riskless return of about 4% with quarterly data (so that \( f = 0.01 \)). 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 \( \phi = 3 \), so that the elasticity of labor supply is \( \frac{1}{3} \) (as in, for example, Steinsson (2008)). The parameter \( \nu \) is 3, so that the interest semi-elasticity of money demand is equal to 4 (Gali (2008)). The Calvo non-adjustment probability is \( \theta = 0.75 \), 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. Finally, the coefficients in the interest rate rule are set to \( \phi_\pi = 1.5 \) and \( \phi_y = 0.125 \) (Taylor (1993)). We will experiment with different values for the interest rate smoothing parameter \( \rho_f \). We note here that values for this parameter in recent, related literature are found to be quite high (Steinsson (2008), Castelnuovo (2011)).

The required reserve ratio \( \theta_R \) is 0.1 (10%). The steady state ratios \( \frac{1}{2} \) (loan to deposits) and \( \frac{5}{4} \) (bonds to deposits) are 0.5 and 0.4, respectively. These are average values based on commercial banks’ balance sheet data at the Federal Reserve.\(^{10}\) It is less straightforward to calibrate the parameters of the banks’ loan supply function.\(^{11}\) We will discuss this more extensively, devising strategies to have some guidance on the value of such parameters. For now, given the importance of this channel in the analysis that follows, we analyze the robustness of results to changes in these parameters. We will also do sensitivity analysis as regards to the real wage rigidity parameter \( \gamma \).

### 4 Implications for inflation persistence

Our main purpose is to verify the implications of the model for the response of the inflation rate to an expansionary monetary policy shock. We do so under alternative parametrizations for the banks’ loan supply function, varying \( L_f, L_f^2 \) and \( L_f \). For each case, we report four set of responses, representing different aspects

\(^{10}\) Averages are computed over the sample 1959:1 – 2008:2. The ratios are appropriately rescaled to take into account the simplified bank’s balance sheet in our model.

\(^{11}\) Note that the bond demand is treated as residual, so we do not need to calibrate the \( B_f \) parameter.
of interest within the model: Inflation and output, interest rates, money and credit, and marginal costs with their components. We then discuss the results, providing some insights on the mechanisms at work.

4.1 Impulse response functions

For the first block of results, we choose to impose a moderate degree of real wage rigidities and past-dependancy for monetary policy. Specifically, figures 2 to 4 use $\rho_f = 0.5$ and $\gamma = 0.5$. Figure 2 reports responses for $L_f = -1$, $L_y = 5$ and $L_i = 1$, to describe a scenario where banks' loan supply responds substantially more to economic conditions than to the loan rate. Interest rates respond negatively, the loan rate $i$ considerably more so than the policy rate $f$ (which we label "ffr" as in Federal Funds Rate). Both inflation and output respond positively, in accord with the conventional wisdom on the monetary transmission mechanism. Importantly, the response of inflation is persistent in that it is hump-shaped, consistently with VAR evidence. Loans, reserves (which correspond to money in the log-linearized version of the model) and bonds all respond positively on impact, reflecting the initial effect of the expansion of the bank's balance sheet, while the subsequent dynamics of monetary/credit variables reflect the adjustment of the bank's "portfolio" as the economy adjusts back. Note that loans' quantity in the model is fully driven by demand (i.e. the wage bill), while the dynamics of loans' price, $i$, is determined by the way the bank's supply reacts to output vs. interest rate. This configuration of parameters thus makes for a substantial response for $i$ which, given the relatively muted movement of real wages, drives the response of marginal costs. The values of the parameters in Figure 3 assume the same relative elasticity $\frac{L_y}{L_f}$, but with higher values for both: the dynamics of inflation are not substantially affected (the response still retains its shape), while the larger value of $L_i$ implies more muted dynamics for $i$. Switching the relative importance of $y$ and $i$ markedly changes the qualitative implications for the response of inflation: as shown in Figure 4 (which uses

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta = 0.99$</td>
</tr>
<tr>
<td>Inverse of EIS</td>
<td>$\sigma = 2$</td>
</tr>
<tr>
<td>Inverse of Frisch elasticity</td>
<td>$\phi = 3$</td>
</tr>
<tr>
<td>Int. rate elasticity of $M^d$</td>
<td>$\frac{1}{\nu_f} = 4$</td>
</tr>
<tr>
<td>Calvo probability</td>
<td>$\theta = 0.75$</td>
</tr>
<tr>
<td>Taylor rule</td>
<td>$\phi_x = 1.5$</td>
</tr>
<tr>
<td></td>
<td>$\phi_y = 0.5/4$</td>
</tr>
<tr>
<td>Tech. shock AR(1)</td>
<td>$\rho_a = 0.5$</td>
</tr>
<tr>
<td>Monetary shock AR(1)</td>
<td>$\rho_u = 0.5$</td>
</tr>
</tbody>
</table>

Notes:
L_y = 1 and L_i = 5), π responds positively and then decreases monotonically. We note that the response of marginal costs now closely follows the real wage and i responds positively to the monetary shock.

We then perform some robustness checks by varying the interest rate smoothing and real wage rigidity parameters. Figures 5 and 6 use the same calibration of Figure 2 for the bank’s parameters and consider scenarios of low and high past-dependancy of monetary policy (ρ_f = 0.2 and 0.8, respectively). The qualitative features of the model, and specifically the inflation response, are unchanged across the two scenarios. The parameter governing real wage rigidities, though, plays an important role in shaping inflation dynamics, because of its (obvious) relevance for the response of w−p and, in turn, marginal costs. Figure 7 once again reverts to the parametrization in Figure 4, while setting γ = 0.2 (i.e. a lower degree of real wage rigidities, as w−p now mostly depends on φn_t + σc_t): with this set of parameters, the substantial response of the real wages implies a monotonically decreasing path for inflation. On the other hand, a slightly larger value of γ (= 0.3) is enough to restore a hump-shaped response for π (see Figure 8).

4.2 Discussion

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

\[ \pi_t = \lambda \sum_{i=0}^{\infty} \beta^i \mathbb{E}_t mc_{t+i} \]

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), 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 (e.g. 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.\(^\dagger\) 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

\(^\dagger\)See also Walsh (2003).
Figure 2: Model response functions to an expansionary monetary shock.

Note: Responses to a one s.d. monetary innovation. $L_f = -1$, $L_y = 5$, $L_i = 1$. $\rho_f = 0.5$, $\gamma = 0.5$. 
Figure 3: Model response functions to an expansionary monetary shock.

Note: Responses to a one s.d. monetary innovation. $L_f = -1$, $L_y = 25$, $L_i = 5$. $\rho_f = 0.5$, $\gamma = 0.5$. 

18
Figure 4: Model response functions to an expansionary monetary shock.

Note: Responses to a one s.d. monetary innovation. \( L_f = -1, L_y = 1, L_i = 5. \rho_f = 0.5, \gamma = 0.5. \)
Figure 5: Model response functions to an expansionary monetary shock.

Note: Responses to a one s.d. monetary innovation. $L_f = -1$, $L_y = 5$, $L_i = 1$. $\rho_f = 0.2$, $\gamma = 0.5$. 
Figure 6: Model response functions to an expansionary monetary shock.

Note: Responses to a one s.d. monetary innovation. $L_f = -1$, $L_y = 5$, $L_i = 1$. $\rho_f = 0.8$, $\gamma = 0.5$. 

21
Figure 7: Model response functions to an expansionary monetary shock.

Note: Responses to a one s.d. monetary innovation. $L_f = -1$, $L_y = 5$, $L_i = 1$. $\rho_f = 0.5$, $\gamma = 0.2$. 
Figure 8: Model response functions to an expansionary monetary shock.

*Note:* Responses to a one s.d. monetary innovation. $L_f = -1$, $L_y = 5$, $L_i = 1$. $\rho_f = 0.5$, $\gamma = 0.3$. 

23
in Steinsson (2008), we write (3) using our expression for marginal costs:

$$\pi_t = \lambda \sum_{i=0}^{\infty} \beta^i \mathbb{E}_t (\hat{\pi}_{t+i} + w_{t+i} - p_{t+i} - a_{t+i})$$ (4)

At each point in time, \(\pi_t\) is given by an infinite sum of terms. If we want its response to an expansionary monetary shock to be hump-shaped, we want the first few terms of the summation to be negative as a consequence of such shock. As time goes on and these terms drop out of the sum, \(\pi_t\) becomes more positive until it reaches its peak: the crucial point is that, because of this mechanism, the peak is reached later and not on impact. In order for this happen, the response of real marginal costs to an expansionary monetary shock must be itself hump-shaped and negative 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 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 \(i\). At a basic level, we are looking for a combination of these two channels such that, on impact and for a few periods, the increase in real wages \(w - p\) is more than counteracted by a decrease in \(i\): according to (4), this would generate the "desired" response of marginal costs and hence inflation. Figure 2, for example, clearly shows this type of response for \(mc\).

Our analysis in the previous section, together with the above discussion, shows that two 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. As shown earlier, we use the formulation in Blanchard and Gali (2007), who rely on unmodeled imperfections in the labor markets to generate a wedge between the marginal rate of substitution of consumption and leisure and the real wage. 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 marginal costs. However, we have shown (see Figure 8) that even a moderate degree of real rigidity (measured by a relatively low value of \(\gamma\)) can produce a hump-shaped response of inflation. Moreover, and importantly, it’s worth noting that real wage rigidities alone cannot produce this type of result within the Calvo model, as made clear in equation (4): the response of marginal costs (and thus inflation) will be more sluggish, but all terms in the summation will be positive (see, for example, Figure 9, which shows the responses of inflation and output in a model with no cost channel, and \(\gamma = 0.8\)).

\[\text{To be added: note on different implications of staggered wage setting.}\]
Figure 9: Model response functions to an expansionary monetary shock. 
*Note:* Inflation and output responses to a one s.d. monetary innovation in the model with no cost channel and real wage rigidity parameter $\gamma = 0.8$.

The second aspect, which is at the core of how we model financial intermediation, lies in the banks’ loan supply function. When banks react more strongly to economic conditions than the loan interest rate, the latter has to move more to bring about equilibrium in the market for loanable funds. Thus, the interest rate is more likely to counteract the real wage. We have also argued that, as far as inflation dynamics are concerned, the *ratio* of the elasticities $\frac{L_i}{L_y}$ matters, more than the individual values themselves. As we have seen, the model fails to replicate the desired shape of inflation response when the interest rate elasticity of loan supply is larger than the output elasticity. It would then be important to learn more about the relative strength of these supply elasticities from macroeconomic data. For now, we note that when $L_i > L_y$, the model produce a *positive* response of $i$ to an expansionary monetary shock (see for example Figure 4), which runs counter conventional wisdom on the effects of monetary policy. However, for moderate degrees of real wage rigidities, a ratio of the elasticities between 2.5 and 7 is successful in generating a hump-shaped response of the inflation rate with results that improve for higher $L_f$.$^{14}$

Finally, we relate these results to the role of strategic complementarities/real rigidities [add references]. It is well understood in the literature that, given reasonable degrees of *nominal* rigidity, it is hard to generate

---

$^{14}$The solution of the system is not well-behaved for ratios $\frac{L_i}{L_y}$ larger than 7 and for values of $L_f$ larger than 3. This is an important point because it implies that the mechanism explored by this paper cannot be implausibly stretched in order to obtain arbitrarily large responses of the loan rate to the monetary shocks.
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 on 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 [add references].

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.\(^{15}\) Given our assumptions (specifically that of homogeneous labor market), the elasticity in this model is \(\zeta = \phi + \sigma\), which is greater than 1 (thus implying strategic substitutability) with our calibration. In this sense, the results do not need to rely on a high degree of strategic complementarity.

5 Conclusions and Future Work

In this paper we explore the cost channel of the transmission of monetary policy and link it to the persistence of inflation responses to monetary shocks. We show that adding a loan market to the standard new Keynesian model with Calvo price setting, it is possible to increase the persistence of these responses in line with the evidence found in the empirical VAR literature.

For the moment, our model reveals the existence of a promising mechanism that could help improving the new Keynesian general equilibrium models used in policy analysis without altering the Calvo pricing framework. However, this work is still essentially a theoretical paper that needs improvements in several directions. The main points are the following:

1. So far, we have explored the calibration only in theoretical terms given the assumptions about the behavior of the response functions of inflation and the other main variables in the model. This allows us to make some interesting considerations on the feasible parameter space for the calibration, but it is not enough. First of all, we need to obtain a better calibration of the elasticities of the loan supply from the data. Then, we will have to introduce a credit market shock and simulate the model in order to calibrate the relative importance of this shock and the usual TFP and demand shocks to better quantify the importance of the credit market for the business cycle and the persistence of inflation in general. In this respect, we also need to have a more formal measurement of the inflation persistence.

2. This paper can be seen as a first step toward a more complete model in which the interaction between banks and firms in the credit markets is more realistic. As argued by Blinder and Stiglitz (1983), the relationship between banks and firms is also based on the information that banks have about

\(^{15}\)This is what Woodford (2003) refers to as the elasticity of the notional Short-Run Aggregate Supply (SRAS).
the firms that demand credit and often a firm can get liquidity only from a small number of banks. Blinder (1987) follows this view in including credit rationing in his model. Instead of rationing, we think that these observations could be translated into some sort of "stickiness" of the response of banks to monetary shocks, similar to those used to model price or wage staggering. Credit lines are not highly substitutable for a firm and this would increase the market power of banks. This view is also supported by the response of loans and bonds to the monetary shock in our VAR analysis. Those response functions are difficult to reconcile with our baseline model and they call for a deeper examination.

3. An estimation of the model will be possible at that point.
REFERENCES


APPENDIX

A Some More Details

[Nothing here for the moment]