Another Look at Sticky Prices and Output Persistence*

Peng-fei Wang
Department of Economics
Cornell University

Yi Wen
Research Department
Federal Reserve Bank of St. Louis

(This Version: June, 2005)

Abstract

Price rigidity is the key mechanism for propagating business cycles in traditional Keynesian theory. Yet the New Keynesian literature has failed to show that sticky prices by themselves can effectively propagate business cycles in general equilibrium. We show that price rigidity in fact can (by itself) give rise to a strong propagation mechanism of the business cycle in standard New Keynesian models, provided that investment is also subject to a cash-in-advance constraint. In particular, we show that reasonable price stickiness can generate highly persistent, hump-shaped movements in output, investment and employment in response to either monetary or non-monetary shocks, even if investment is only partially cash-in-advance constrained. Hence, whether or not price rigidity is responsible for output persistence (and the business cycle in general) may not be a theoretical question, but an empirical one.

Keywords: Business Cycle; Money; Sticky Prices; Output Persistence; New-Keynesian Models; Cash-in-Advance; Money-in-Utility; Financing Constraints and Investment.

JEL Classification: E52, E41, E32.

*We would like to thank Jess Benhabib, Bill Gavin, Patrick Kehoe, Karl Shell, two anonymous referees and the editor, Wouter Denhaan, for helpful comments. The views expressed in the paper and any errors that may remain are the authors' alone. Correspondence: Yi Wen, Research Department, Federal Reserve Bank of St. Louis, St. Louis, MO, 63144. Email: yi.wen@stls.frb.org.
1 Introduction

Sticky prices are the key mechanism assumed in traditional Keynesian theory for propagating the impact of monetary shocks as well as other aggregate shocks throughout the economy. Yet to demonstrate a persistent output effect of sticky prices in a fully-specified new Keynesian dynamic-general-equilibrium model has proven to be very difficult, as recently stressed by Chari, Kehoe and McGrattan (CKM 2000). CKM show that empirically plausible degree of price rigidity generates only a modest degree of output persistence in responding to monetary shocks, far from enough to account for the estimated output persistence in the U.S. economy. The usefulness of the sticky price assumption, one of the cornerstone in traditional Keynesian business cycle theory, is thus under a serious challenge.1

The persistence problem raised by CKM (2000) along with others has led researchers to explore other types of rigidities or economic forces, in conjunction with sticky prices, to explain the persistent effects of monetary shocks. For example, Christiano, Eichenbaum, Evans (2005) obtain more persistent output responses to monetary shock by combining both sticky prices and sticky wages on the nominal side, aided by habit formation, adjustment costs, limited participation in money market and variable capital utilization on the real side. In a model without capital, Jeanne (1998) shows that adding real-wage rigidity into sticky price models can significantly increase the propagation of monetary shocks in output. Dotsey and King (2001) show that output persistence can be increased by features such as a more important role for produced inputs, variable capacity utilization, and labor supply variability through changes in employment. These elements together can reduce the elasticity of marginal cost with respect to output, improving the persistence of output. Bergin and Feenstra (2000) emphasize interactions between input-output production structures and translog preferences to increase output persistence under sticky prices. Similar results based on production chains can also be found in the work of Huang and Liu (2001). Other researchers such as Mankiw and Reis (2002), Woodford (2001), Erceg and Levin (2003) have emphasized the important role of imperfect-information in helping sticky prices to generate persistent output responses to monetary shocks.2

By adding a large number of building blocks, such as real rigidities and complex information

---

1 For a review of the New Keynesian literature, see Clarida, Gali and Gertler (1999). For empirical literature on the persistent effects of monetary shocks, see Sims (1992), Christiano et al. (1995) and Strongin (1995), among others.

2 The literature has also explored the implications of sticky nominal wages for output persistence. Models based on staggering wages such as those in Andersen (1998), Erceg (1997), and Huang and Lin (2001) are still not able to generate sufficient degree of real persistence seen in data, though they do alleviate the problem to some extent. Edge (2002) recently establishes conditions under which wage and price staggering are equivalent regarding their effects on output persistence, thus the persistence problem is similar in both sticky-wage and sticky-price models. Also see Dotsey and King (2005) for the recent new literature on state-dependent pricing in general equilibrium. This literature shows that state-dependent pricing can have dramatically richer propagation mechanisms than time-dependent pricing in generating output and inflation persistence. Benhabib and Farmer (2000) show that externalities can also give rise to output persistence in a monetary model.
structures, into the standard sticky-price model can improve the model’s fit in terms of output persistence, but at the expense of simplicity. Often more than one factors are added to entangle with nominal rigidities such that it becomes hard to distinguish exactly which factor is doing what in generating output persistence. In addition, while sticky or imperfect information proves to be effective in giving rise to output persistence, the way they are modeled in the literature often uses partial equilibrium framework. It is shown recently by Keen (2004), for example, that the business cycle implications of sticky information proposed by Mankiw and Reis (2002) may not be robust to general equilibrium extensions.\(^3\)

This paper takes a step back and asks whether a canonical sticky price model without any additional frictions or real rigidities can generate a reasonable degree of output persistence. Putting it another way, this paper asks why sticky prices by themselves may fail to provide a strong propagate mechanism for the business cycle. This is an intriguing question because intuitively there is no reason price rigidity would not lead to output persistence, since it could turn \(i.i.d\). money shocks into serially correlated movements in the real balance just as effectively as any types of real rigidities. Real balance in turn could affect aggregate spending and production. Yet despite the exploding literature trying to overcome the persistence problem, what exactly fails the Keynesian sticky price propagation mechanism in general equilibrium models remains unclear. Chari, Kehoe and McGrattan (2000), for example, show the inability of sticky prices in generating output persistence mainly via model simulations. The reasons behind the failure are less clearly presented when capital is included.

We show in this paper that sticky prices can in fact by themselves generate highly persistent output movements, contrary to the findings of the existing literature. In particular, we show that empirically plausible degree of price stickiness can generate hump-shaped output responses to monetary shocks in a way very similar to the data. Thus, sticky prices are certainly a useful assumption in explaining the business cycle as far as theory is concerned. Whether they are in fact responsible for the business cycles in the real world, however, is an empirical question.

The key to our finding is a cash-in-advance (CIA) constraint on aggregate demand (consumption plus investment). CIA constraints can significantly limit the initial increase in aggregate demand after a money injection, in sharp contrast to the popular money-in-utility (MIU) specification, because agents are forced to accumulate real balances before they can fully raise spending, leading to more smoothed output responses. Consider a model without capital. Under a money-in-utility (MIU) specification, demand for goods and demand for money are only loosely linked. Households can therefore raise consumption significantly beyond the initial increase in money injection, in anticipation of future money increases, leading to volatile impulse responses in output. But a

\(^3\)Erceg and Levin (2003) is an exception.
CIA constraint limits the rise in consumption to the current rise in money, forcing households to wait until future money injections to fully raise consumption. Hence, CIA can lead to hump-shaped output persistence under serially correlated money shocks while MIU cannot. When there is capital in the model, if both consumption and investment are subject to CIA constraints (there is no equivalent form of this specification to a MIU specification), the increase in aggregate demand is again limited to the current rise in money, giving rise to more smoothed, hump-shaped output responses to shocks.

In addition, with CIA constraint on aggregate spending (consumption plus investment), sticky prices can lead to hump-shaped output persistence not only under monetary shocks, but also under non-monetary shocks, such as technology shocks and preference shocks. The intuition is similar: cash-in-advance postpones the maximum impact of shocks on aggregate demand because agents are forced to intertemporally smooth aggregate spending via real balance accumulation over time. A smoothed aggregate demand thus dictates a smoothed aggregate supply (production).\(^4\)

The key assumption driving our results is that investment must be subject to a CIA constraint. Given the availability of complicated financial markets, this seems a difficult assumption to defend. However, this assumption may be defended at least on three grounds. First, firms' investment projects are often subject to financing constraints due to capital market imperfections (moral hazard, asymmetric information, incomplete markets, and so on). Consequently, firms' internal cash flows are often crucial in determining their investment level. A vast empirical literature in the past twenty years has documented a strong link between firms' internal cash flow and investment. This literature shows that such a close relationship is due to financial constraints rather than to good performances in sales (see, most notably, Fazzari, Hubbard and Petersen, 1988).\(^5\)

Second, the so called “money" in a standard CIA model can be understood in more general terms as \(M1\) or \(M2\), rather than as just cash. Given that the broad money supply (\(M1\) or \(M2\)) is proportional to the monetary base (\(B\)) according to \(Mt = mBt\), the above interpretation is valid as long as the money multiplier (\(m\)) is relatively constant. Thus, a monetary shock in the model can be reinterpreted as a shock to the availability of liquidity or credit in terms of \(M1\) or \(M2\), which is

\(^4\)It is thus not surprising that our findings also contradict a branch of the existing literature that assumes CIA. For example, Yun (1996) studies a CIA constrained sticky price model with capital and finds that money shocks have no persistent effects on output. Ellison and Scott (2000) use the same model and demonstrate that sticky prices not only fail to produce persistent output fluctuations but also generate extremely volatile output at very high frequencies. This is because both papers assume CIA constraint on consumption only. When there is capital in the model, intertemporal substitution between current consumption and future consumption can be achieved through capital accumulation. In this case, imposing CIA constraint only on consumption spending is not effective for generating persistent output, since investment becomes very volatile by serving as the buffer for consumption, and consequently investment will dictate output dynamics. Thus, even if consumption is hump-shaped, output is not. This suggests that a CIA constraint on investment is crucial for generating output persistence, as consumption can always be smoothed by capital accumulation.

\(^5\)See Habbard (1998) for a comprehensive review of this large literature. Also see Fazzari, Hubbard and Petersen (2000) for the recent debate on this issue.
just as likely to affect firms as to consumers.\(^6\)

Third, the subjection of both consumption and investment to money is more consistent with the estimation of aggregate money demand function. For example, Mulligan and Sala-I-Martin (1992) find that aggregate income is a better scale variable than aggregate consumption in estimating money demand. Indeed, most empirical work in money demand estimation has adopted income rather than consumption as a scale variable. Classical examples include Friedman (1959) and Goldfeld (1973, 1976). This may explain why CIA constraints on both consumption and investment are widely used in the theoretical monetary literature, such as Stockman (1981), Abel (1985) and Fuerst (1992), to name just a few.

The above justifications notwithstanding, we show that output continues to be hump-shaped even when only as little as 30\% of firms’ investment is subject to CIA constraint in our model. Therefore, theoretically speaking, sticky prices have no trouble generating output persistence as long as investment is partially constrained by money holdings.

The rest of the paper proceeds as follows. Section 2 demonstrates output persistence under CIA in a simple model without capital. We show in this model that sticky prices can give rise to hump-shaped output responses to money shocks under CIA, but not under MIU.\(^7\) Section 3 studies a fully specified general equilibrium model with capital. It is shown that under either monetary or non-monetary shocks, output exhibits hump-shaped persistence as long as a certain fraction of investment is subject to money-in-advance constraint. Hence, introducing capital into the model does not destroy the persistence mechanism of sticky prices, in contrast to CKM (2000). Section 4 concludes the paper.

2 The Basic Model

2.1 Households

A representative household chooses sequences of consumption, \(\{C_t\}_{t=0}^\infty\), labor supply, \(\{N_t\}_{t=0}^\infty\), and money demand, \(\{M_t\}\), to solve

\[
\max E_0 \sum_{j=0}^\infty \beta^j \left[ \log C_t - aN_t \right]
\]

\(^6\)There is strong evidence that firms hold a substantial fraction of money in the economy. Using data from the Federal Reserve’s Demand Deposit Ownership Survey (DDOS), which separately reports the ownership of demand deposits at commercial banks by financial firms, nonfinancial firms, households and foreigners, Mulligan (1997) found that nonfinancial firms hold at least 50\% more demand deposits than do households. By the 1980s firm had accumulated almost twice as many demand deposits as household had. On the other hand, the Federal Reserve’s Flow of Funds (FOF) reports that households may hold more M1 than firms do, but Mulligan convincingly argued that the DDOS data are more accurate than the FOF data.

\(^7\)Jeanne (1998) shows that real wage rigidity must be introduced into a CIA model without capital in order to generate persistence. Here we show that this is not necessary.
subject to $C_t + \frac{M_t}{P_t} \leq \frac{M_{t-1} + X_t}{P_t} + w_t N_t + \Pi_t$ and the CIA constraint, $C_t \leq \frac{M_t}{P_t}$; where $X$ is money injection, $P$ is the aggregate goods price in terms of money, $w$ is the real wage, and $\Pi$ is the profit income contributed from firms which the household owns. Since the current money holdings, $M_t$, enter the CIA constraint, there is no inflation tax on consumption. Hump-shaped output persistence remains if the inflation tax effect is allowed. Note that a linear leisure function is assumed for simplicity. Making the leisure function nonlinear has little effect on the results. Denoting $\lambda_1$ and $\lambda_2$ as the Lagrangian multipliers for the budget constraint and the CIA constraint respectively, the first order conditions can be summarized by

\[ \frac{1}{C_t} = \lambda_{1t} + \lambda_{2t} \]  
\[ a = \lambda_{1t} w_t \]  
\[ \lambda_{1t} = \beta E_t \frac{P_t}{P_{t+1}} \lambda_{1t+1} + \lambda_{2t}. \]

### 2.2 Firms

The final goods, $Y_t$, are produced by a perfectly competitive firm according to the technology, $Y_t = \left[ \int y_t(i) \frac{d}{\sigma} di \right]^{\frac{1}{\sigma-1}}$, where $\sigma > 1$ measures the elasticity of substitution among the intermediate goods, $y(i)$. Let $p_t(i)$ denote the price of intermediate goods $i$, the demand for intermediate goods is given by $y_t(i) = \left[ \frac{p_t(i)}{P_t} \right]^{-\sigma} Y_t$, and the relationship between final goods price and intermediate goods prices is given by $P_t = \left[ \int p_t(i)^{1-\sigma} d i \right]^{\frac{1}{1-\sigma}}$.

Each intermediate good $i$ is produced by a single monopolistically competitive firm according to the following technology, $y_t(i) = n_t(i)$. Intermediate good firms face perfectly competitive factor markets, and are hence price takers in the factor markets. Profits are distributed to household at the end of each time period. The cost function for firm $i$, can be derived from minimizing $w_t n_t(i)$ subject to $n_t(i) \geq y$. Denoting $\phi_t$ as the Lagrangian multiplier, which is also the real marginal cost, the first order condition for cost minimization is given by $w_t = \phi_t$. Consequently, the real profit in period $t$ is given by $(\frac{p_t(i)}{P_t} - \phi_t) y_t(i)$.

Following Calvo (1983) in assuming that each firm has a probability of $1 - \theta$ to adjust its monopoly price in each period, then a firm’s intertemporal profit maximization problem is to choose the optimal price, $p^*_t$, to maximize

\[ E_t \sum_{s=0}^{\infty} (\beta \theta)^{t+s} \Lambda_{t,t+s} \left[ \frac{p^*_t}{P_{t+s}} - \phi_{t+s} \right] y_{t,t+s}(i), \]
where $\Lambda_{t,t+s} \equiv \left( \frac{C_{t+s}}{C_t} \right)^{-1}$ is the ratio of marginal utilities taken as exogenous by the firm; and $y_{t,t+s}$ denotes the firm’s output level in period $t + s$ given its optimal price in period $t$: $y_{t,t+s}(i) = \left( \frac{p_{t+s}^*(i)}{P_{t+s}} \right)^{-\sigma} Y_{t+s}$. The first order condition for optimal monopoly price implies the following pricing rule:

$$p_t^* = \frac{\sigma \sum_{s=0}^{\infty} (\beta \theta)^{t+s} E_t \Lambda_{t+s} P_{t+s}^{\sigma} Y_{t+s} \phi_{t+s}}{\sigma \sum_{s=0}^{\infty} (\beta \theta)^{t+s} E_t \Lambda_{t+s} P_{t+s}^{\sigma-1} Y_{t+s}}.$$  

Because all firms that can adjust their prices face the same problem, all monopolist firms will set their prices in the same way as indicated above. The average price of firms that do not adjust prices is simply last period’s price level, $P_{t-1}$. Given that only a fraction of $1 - \theta$ can adjust their prices in each period, the final good price index can then be written as $P_t = \left[ \theta P_{t-1}^{1-\sigma} + (1 - \theta) P_t^{1-\sigma} \right]^{\frac{1}{1-\sigma}}$.

### 2.3 Equilibrium Dynamics

In equilibrium, household’s first order conditions and firms’ profit maximization conditions are satisfied, all markets clear, and the CIA constraint binds. We study symmetric equilibrium only. The model is solved by log-linearization around a zero-inflation steady state. Using circumflex lower-case letters to denote percentage deviations around steady state, the log-linearized optimal price and the price index are given respectively by $\hat{p}_t = (1 - \beta \theta) \sum_{s=0}^{\infty} E_t \left( \hat{\phi}_{t+s} + \hat{p}_{t+s} \right)$ and $\hat{p}_t = \theta \hat{p}_{t-1} + (1 - \theta) \hat{p}_t^*$, which together imply the New Keynesian Phillips relationship:

$$\hat{\pi}_t = \beta E_t \hat{\pi}_{t+1} + \frac{(1 - \theta)(1 - \beta \theta)}{\theta} \hat{\phi}_t,$$  

where $\hat{\pi}_t \equiv \hat{p}_t - \hat{p}_{t-1}$ is the inflation rate.

The log-linearized aggregate production function is given by $\hat{y}_t = \hat{n}_t$, hence around the steady state the aggregate production function is the same as individual firm’s production function. Notice that the CIA constraint can be expressed as $\hat{y}_t - \hat{y}_{t-1} = x_t - \hat{\pi}_t$, where $x \equiv \log \frac{X_t}{M_{t-1}}$ denotes the growth rate of nominal money stock. We assume that the monetary authority follows a money growth rule given by $x_t = \rho x_{t-1} + \varepsilon_t$. The household’s first-order conditions are thus reduced to:

$$(2 - \beta) \hat{y}_t - 2 \hat{\phi}_t = -\beta (\hat{\pi}_{t+1} + \hat{\phi}_{t+1}).$$

Substituting out $\pi_t$ in this equation and in the New Keynesian Phillips curve using the CIA constraint, the system of equations for solving $\{\hat{y}_t, \hat{\phi}_t\}$ are given by:

$$x_t + \hat{y}_{t-1} - (1 + \beta) \hat{y}_t = \beta E_t (x_{t+1} - \hat{y}_{t+1}) + \frac{(1 - \theta)(1 - \beta \theta)}{\theta} \hat{\phi}_t,$$  

\[ (7) \]
\[
2\hat{y}_t - 2\hat{\phi}_t = \beta \hat{y}_{t+1} - \beta x_{t+1} - \beta \hat{\phi}_{t+1};
\]
which can be arranged more compactly as

\[
E_t \begin{pmatrix} \hat{y}_{t+1} \\ \hat{y}_t \\ \hat{\phi}_{t+1} \\ x_{t+1} \end{pmatrix} = \begin{pmatrix} \frac{1+\beta}{\beta} & -1 & \frac{(1-\theta)(1-\beta\theta)}{\beta\theta} & -\frac{1+\rho\beta}{\beta} \\ \frac{1+\beta}{\beta} & 0 & \frac{1-\beta\theta}{\beta\theta} & 0 \\ \frac{1+\beta}{\beta} & -1 & \frac{1-\beta\theta+\theta+\beta\theta^2}{\beta\theta} & -\frac{1}{\beta} \\ 0 & 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} \hat{y}_t \\ \hat{y}_{t-1} \\ \hat{\phi}_t \\ x_t \end{pmatrix}.
\]

The eigenvalues of the Jacobian matrix are given by: \( \{ \frac{2}{\beta}, \frac{1}{\beta\theta}, \theta, \rho \} \). Note that the first two of the eigenvalues are larger than unit, hence they can be utilized to solve the system forward to determine \( \{ \hat{y}_t, \hat{\phi}_t \} \) as functions of the state \( \{ \hat{y}_{t-1}, x_t \} \). Clearly, the other two smaller roots, \( \{ \theta, \rho \} \), determine the propagation mechanism of output. The decision rule of output takes the form: \( \hat{y}_t = \theta \hat{y}_{t-1} + \alpha x_t \), where \( \alpha \equiv \frac{2\theta(1-\rho\beta)+\rho\beta(\beta\theta-1+\theta-\beta\theta^2)}{(2-\rho\beta)(1-\rho\beta)} < 1 \) is the elasticity of output with respect to money growth shocks.\(^8\) Clearly, the persistence of output is determined jointly by the degree of price stickiness, \( \theta \), and the persistence of shocks. If monetary shocks follow an AR(1) process: \( x_t = \rho x_{t-1} + \varepsilon_t \), then output becomes an AR(2) process:

\[
\hat{y}_t = (\theta + \rho) \hat{y}_{t-1} - \theta \rho \hat{y}_{t-2} + \alpha \varepsilon_t,
\]
which implies a hump-shaped impulse response function. Suppose that the average price stickiness is about four quarters in the economy, the probability of not adjusting prices is then \( \theta = 0.75 \). Given that money growth shocks have autocorrelation of \( \rho = 0.6 \), as is commonly assumed in the literature (e.g., CKM 2000),\(^9\) then the degree of output persistence implied by the model matches the contract multiplier of the U.S. economy estimated by Chari, Keohoe and McGrattan (2000) almost exactly. The maximum impact of a money injection on output is delayed for three quarters after the shock. The simulated impulse responses of output are graphed in Figure 1 (top window).

---

\(^8\) \( \alpha > 0 \) if \( \theta \) is large enough.

\(^9\) Also see our calibration using post-war data in the next section.
When money demand steps from MIU instead, under standard assumptions regarding the elasticity of substitution between consumption and money, no hump-shaped output persistence can be generated from the model. To demonstrate, let the household solve:

\[
\max E_0 \sum_{j=0}^{\infty} \beta^j \left[ \log C_t + \eta \log \frac{M_t}{P_t} - a N_t \right]
\]

subject to \( C_t + \frac{M_t}{P_t} \leq \frac{M_{t-1} + X_t}{P_t} + w_t N_t + \Pi_t \). Letting all parameters take the same values as in the previous CIA model, the bottom panel in Figure 1 shows that output does not have hump-shaped persistence.\(^{10}\)

Technically speaking, the log-linearized first order conditions of the MIU model can be reduced to the following system:

\[
E_t \begin{pmatrix}
\hat{y}_{t+1} \\
\hat{\pi}_{t+1} \\
\hat{m}_t \\
\hat{x}_{t+1}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{\beta} + \frac{(1-\theta)(1-\beta \theta)}{\theta} & -1 & -1+\beta & -1+\beta \\
-\frac{1}{\beta} (1-\theta)(1-\beta \theta) & \frac{1}{\beta} & 0 & 0 \\
0 & -1 & 1 & 1 \\
0 & 0 & 0 & \rho
\end{pmatrix} \begin{pmatrix}
\hat{y}_t \\
\hat{\pi}_t \\
\hat{m}_{t-1} \\
\hat{x}_t
\end{pmatrix}, \tag{11}
\]

\(^{10}\)Again, the linear leisure function is assumed for simplicity. Making it nonlinear has little effect on the results. CKM (2000) argue that perfect substitutability between consumption and leisure is crucial for generating output persistence under the Taylor (1980) type of price rigidity. Here we find that this requirement is not necessary under the Calvo (1983) type of price rigidity. See Kiley (2002) for discussions regarding the differences between the Taylor type and the Calvo (1983) type of price rigidities.
which has the following analytical solution for the decision rule of output: \[ y_t = \hat{y}_{t-1} + \frac{\theta(1-\beta^2 \rho)}{(1-\beta)(1-\beta \rho)} x_t \]

When \( x_t \) is AR(1), we have

\[ \hat{y}_t = (\theta + \rho) \hat{y}_{t-1} - \theta \rho x_{t-1} + \frac{\theta(1-\beta^2 \rho)}{(1-\beta)(1-\beta \rho)} \epsilon_t - \frac{\rho \theta}{(1-\beta \rho)} \epsilon_{t-1}. \] (12)

Notice that output is no longer an AR(2) process as in equation (10), but an ARMA(2,1) process.

The crucial difference this makes is that one of the autoregressive roots (the poles) and the moving average root (the zeros) almost cancel each other in the MIU model, reducing the ARMA(2,1) process in equation (12) to an AR(1) process. An AR(1) process cannot exhibit hump-shaped dynamics. To see the pole-zero cancellation, let \( \beta = 1.1^1 \) Equation (12) reduces to

\[ \hat{y}_t = \theta \hat{y}_{t-1} + \frac{\theta}{1-\rho} \epsilon_t. \] (13)

This also explains why consumption (output) can be very volatile in MIU models due to the value of \( \rho \) typically assumed in the literature. For example, let \( \theta = 0.75 \) and \( \rho = 0.6 \), then consumption will increase by \( \frac{0.75}{0.4} = 1.875 \) percent when money growth increases by just one percent. This is consistent with the graph in the lower panel in Figure 1.

The intuition is that cash-in-advance prevents consumption from rising too much in the impact period since agents are not able to increase consumption beyond the current cash injections. This smooths demand and hence production. On the other hand, when consumption is not cash-in-advance constrained (as in MIU models), households can raise consumption significantly beyond the initial increase in money injection, in anticipation of future money increases. This mechanism of output smoothing due to a CIA constraint on aggregate demand continues to work in more general models with capital, as the following section shows.

3 The Full Model

3.1 Households

The representative household chooses consumption \( (C) \), hours \( (N) \), capital stock \( (K) \), money demand \( (M) \), and bond holdings \( (B) \) to solve:

\[ \max E_0 \sum_{t=0}^{\infty} \beta^t \left[ \Theta \log C_t - a \frac{N_t^{1+\gamma}}{1+\gamma} \right] \]

\[^11\text{This near pole-zero cancellation will take place regardless the value of } \beta. \text{ The reader can check this by setting } \beta = 0, \text{ for example.}\]
subject to
\[ C_t + [K_{t+1} - (1 - \delta)K_t] + \frac{M_t + B_t/R_t}{p_t} = \frac{M_{t-1} + B_{t-1} + X_t}{p_t} + r_t K_t + w_t N_t + \Pi_t \] (14)
\[ C_t + K_{t+1} - (1 - \delta)K_t \leq \frac{M_t}{p_t} \] (15)

where \( r_t \) and \( w_t \) denote real rental rate and real wage rate that prevail in competitive factor markets; \( R \) denotes nominal returns to bonds, \( \delta \) denotes the depreciation rate of capital. At the end of each period, the household receives wages from hours worked, rental payments from capital lending, and nominal bonds returns as well as profits \( \Pi_t \) from all firms the household owns. If consumption is the only cash goods, then our model reduces to that of Yun (1996) and Ellison and Scott (2000).\(^\text{12}\)

Denoting the Lagrangian multipliers for (14) and (15) as \( \lambda_1 \) and \( \lambda_2 \) respectively, the first order conditions with respect to \( \{C_t, N_t, K_{t+1}, M_t, B_t\} \) can be summarized by

\[ \frac{\Theta_t}{C_t} = \lambda_1 t + \lambda_2 t \] (16)
\[ aN_t^\gamma = \lambda_1 t w_t \] (17)
\[ \lambda_1 t + \lambda_2 t = \beta (1 - \delta) E_t (\lambda_{1t+1} + \lambda_{2t+1}) + \beta E_t \lambda_{1t+1} r_{t+1} \] (18)
\[ \lambda_1 t = \beta E_t \lambda_{1t+1} \frac{P_t}{P_{t+1}} + \lambda_2 t \] (19)
\[ \frac{\lambda_1 t}{R_t} = \beta E_t \lambda_{1t+1} \frac{P_t}{P_{t+1}} \] (20)

### 3.2 Firms

The final good sector is the same as described previously. Hence the demand of intermediate goods is given by \( y(i) = \left[ \frac{p(i)}{T} \right]^{-\sigma} Y \), and the price index for final goods is given by \( P = \left[ \int_0^1 p(i)^{1-\sigma} di \right]^{\frac{1}{1-\sigma}} \).

The production technology for intermediate good \( i \) is given by \( y(i) = Ak(i)^\alpha n(i)^{1-\alpha} \), where \( 0 < \alpha < 1 \) and \( A \) denotes aggregate technology shocks to productivity. The cost function of firm \( i \) is derived by minimizing \( r k(i) + wn(i) \) subject to \( Ak(i)^\alpha n(i)^{1-\alpha} \geq y \). The first order conditions are given by \( r = \phi \frac{y(i)^{\alpha}}{k(i)} \), \( w = \phi (1-\alpha) \frac{y(i)^{\alpha}}{n(i)} \), where \( \phi \) denotes the real marginal cost. Given the production function, the real marginal cost can be written as

\(^{12}\)An alternative specification of the model would be to impose CIA constraints on individual firms’ investment instead of on aggregate investment. However, doing so gives rise to a non-trivial aggregation problem under sticky prices, which is beyond the scope of the current paper to solve. We hope to tackle this aggregation problem in a future project.
\[
\phi_t = \frac{1}{A_t} \left( \frac{r_t}{\alpha} \right)^\alpha \left( \frac{w_t}{1 - \alpha} \right)^{1-\alpha}.
\]  

(21)

Note that, since the total cost equals \( \phi_t y_t \), the marginal cost equals average cost. Let the probability of price adjustment in each period for any intermediate firm be \( 1 - \theta \), a firm’s optimal price is again to choose \( p^* \) to maximize \( E_t \sum_{s=0}^{\infty} (\beta \theta)^{t+s} A_{t,t+s} \left[ p_{t+s}^* - \phi_{t+s} \right] \left[ \frac{p_{t+s}^*}{P_{t+s}} \right]^{-\sigma} Y_{t+s} \), which yields the same pricing rule as before:

\[
p_t^* = \frac{\sigma \sum_{s=0}^{\infty} (\beta \theta)^{t+s} E_t A_{t+s} P_{t+s}^\sigma Y_{t+s} \phi_{t+s}}{(\sigma - 1) \sum_{s=0}^{\infty} (\beta \theta)^{t+s} E_t A_{t+s} P_{t+s}^{\sigma-1} Y_{t+s}}
\]

(22)

### 3.3 Equilibrium and Calibration

In a symmetric equilibrium near the steady state, the aggregate production function can still be expressed as \( Y_t = A_t K_t^\alpha N_t^{1-\alpha} \) and the aggregate profit is still given by \( \Pi_t = Y_t - w_t N_t - r_t K_t \). Hence the equilibrium market clearing conditions and constraints are:

\[
C_t + K_{t+1} - (1 - \delta)K_t = Y_t
\]

(23)

\[
M_t = M_{t-1} + X_t
\]

(24)

\[
B_t = B_{t-1} = 0
\]

(25)

\[
C_t + K_{t+1} - (1 - \delta)K_t = \frac{M_t}{P_t}
\]

(26)

The optimal pricing rule in (22) in conjunction with the law of motion of the aggregate price index, \( P_t = \left[ \theta P_{t-1}^{1-\sigma} + (1 - \theta) P_t^{s-1-s} \right]^{1-\sigma} \), leads to the same relationship for the dynamics of inflation around the steady state as before: \( \tilde{\pi}_t = \beta E_t \tilde{\pi}_{t+1} + \frac{(1-\theta)(1-\beta \delta)}{\theta} \tilde{\phi}_t \), except that the marginal cost function is now different.

In a zero-inflation steady state, it can be shown that the following relationships hold:

\[
\phi = \frac{\sigma - 1}{\sigma}
\]

(27)

\[
\frac{K}{Y} = \phi \frac{\beta \alpha}{(2 - \beta) [1 - \beta (1 - \delta)]}
\]

(28)
Notice that, compared to a standard RBC model in which \( \frac{K}{Y} = \frac{\beta}{1 - \beta(1 - \delta)} \), there are two distortions on the steady state capital-output ratio in the sticky price model. First, monopolistic competition gives rise to a markup of \( \frac{1 - \phi}{\phi} > 0 \), which approaches zero only if the elasticity of substitution \( \sigma \rightarrow \infty \) (i.e., \( \phi \rightarrow 1 \)). A positive markup implies a lower steady state capital-output ratio. Second, due to the fact that money is needed to facilitate transactions, an inflation tax is imposed on investment returns, which lowers the steady state capital-output ratio by a factor of \( 2 - \beta \). If \( \beta = 1 \), this effect disappears.\(^{13}\)

The exogenous shocks are assumed to be orthogonal to each other and follow AR(1) processes in log:

\[
\begin{align*}
xt &= \rho_x xt-1 + \varepsilon_xt \\
\log At &= \rho_A \log A_{t-1} + \varepsilon_A t \\
\log \Theta_t &= \rho_\Theta \log \Theta_{t-1} + \varepsilon_{\Theta} t
\end{align*}
\]

where \( x \equiv \log \frac{X}{M_{t-1}} \) denotes money growth rate. The model is calibrated at quarterly frequency. We choose the time discounting factor \( \beta = 0.99 \), the rate of capital depreciation \( \delta = 0.025 \), the capital elasticity of output \( \alpha = 0.3 \), the inverse labor supply elasticity \( \gamma = 0 \) (Hansen’s indivisible labor), and the elasticity of substitution parameter \( \sigma = 10 \) (implying a markup of about 10%).\(^{14}\) The price rigidity parameter \( \theta \) is set to 0.75, and persistence parameters for technology and preference shocks are set to \( \rho_A = \rho_\Theta = 0.9 \). These parameter values are quite standard in the literature. To calibrate money growth shocks, we estimate an AR(1) model for the growth rate of monetary base \( (M0) \) in the U.S. \( (1950:1 - 2003:4) \), and we obtain \( \rho_x = 0.6 \) and \( \sigma_{\varepsilon_x} = 0.006 \).

### 3.4 Model Evaluation

The impulse responses of output \( (Y) \), consumption \( (C) \), investment \( (I) \) and employment \( (N) \) to a one-standard-deviation shock to money growth are graphed in Figure 2. Several features are worth noticing in Figure 2. First, a monetary growth shock can cause significant increases in economic activity. On impact, investment increases by 2.6 percent and output increases by 0.56 percent, while consumption increases by only 0.06 percent. The overall standard deviation of investment is about four times that of output, and the overall standard deviation of consumption is about half that of output. These different magnitudes suggest that monetary shocks can explain one of the most prominent business cycle facts emphasized by the real business cycle literature; namely, that consumption is less volatile than output and that investment is more volatile than output.

\(^{13}\)See Stockman (1981) for more discussions on this issue.

\(^{14}\)The results are robust to the values of these parameters. For example, very similar results obtain even when the markup is zero and when the utility function on leisure has the log form (i.e., \( \gamma = 0.25 \)).
Second and most strikingly, the impulse responses of output ($Y$), employment ($N$) and investment ($I$) are all hump-shaped, with a peak response reached around the third quarter after the shock. This suggests a richer propagation mechanism of the model than a standard RBC model or a sticky-price model with money-in-utility. This richer propagation mechanism induced by sticky prices and the CIA constraint enables the model to match the observed output persistence in the U.S. economy quite well. For example, if we estimate an AR(2) process for the logarithm of real GDP of the United States (1950:1 - 2003:4) with a quadratic time trend, then the fitted equation is

$$\log(y_t) = 1.3\log(y_{t-1}) - 0.37\log(y_{t-2}) + v_t,$$

where the standard deviation of the residual is $\sigma_v = 0.0088$. Using this estimated standard deviation to simulate the U.S. output by equation (30), Figure 3 (left window) shows that the shape of the impulse response function of the U.S. output looks very much like that implied by the model (where the standard deviation of money shock in the model is $\sigma_{\varepsilon} = 0.006$), except that the volatility of the model output is only about one third of the data.

Chari, Kehoe, and McGrattan (2000) propose to measure the persistence of output by its half life. When the half-life is measured starting from the initial response at impact period, the half-life of output in the model is 10, while that in the data is 11. When it is measured starting from the peak of the response after a shock, the half-life is 8 in the model and 9 in the data.

Ellison and Scott (2000) show that sticky price models cannot explain the business cycle since sticky prices tend to generate too much variations in output at the high frequencies but not enough variations at the business cycle frequencies. Here we show that this is not true if investment spending is subject to cash-in-advance constraint. The right-hand side window in Figure 3 shows that the power spectrum of output growth in the model matches that in the data quite closely in terms of variance distribution across frequencies. However, in terms of total variance (proportional to the area underneath the spectral density function), the model explains only about 16% of the data.\textsuperscript{16}

The intuition for the persistent output effect of sticky prices in the full model with capital is similar to that in the basic model without capital. Cash-in-advance acts to smooth aggregate spending across time; since by requiring cash, the maximum impact of shocks on demand (and hence supply) is postponed until enough real balance is accumulated. Thus the CIA constraint serves essentially like an intertemporal form of adjustment cost, which is well know for generating hump-shaped output dynamics. However, if only consumption goods is subject to CIA, output cannot

\textsuperscript{16}Introducing capacity utilization could improve the model in this regard.
have enough persistence since shocks can immediately impact on investment spending, which will
dictates aggregate demand and supply, making output very volatile at the high frequencies (see,
e.g., Ellison and Scott, 2000).\footnote{Inflation in the model behaves like an AR(1) process, indicating certain degree of persistence, but not hump-shaped persistence. Hence the model cannot explain the well known fact that inflation lags output. However, its volatility relative to output matches the U.S. data quite well. For the issue of inflation persistence and its relation to output, see Fuhrer and Moor (1995), Ireland (2003), Mankiw and Reis (2002) and Wang and Wen (2005), among others.}

Since it is well known that investment is much more volatile than output in the data, to make
sure that a CIA constraint on investment does not lead to too little investment volatility relative to
output, Table 1 reports the standard business cycle statistics of the model. It shows that, among
other things, the model is able to explain the large volatility of investment relative to output despite
investment is subject to CIA constraint.

Table 1. Selected Moments

<table>
<thead>
<tr>
<th>x</th>
<th>$\sigma_x/\sigma_y$</th>
<th>$\text{cor}(x_t, y_t)$</th>
<th>$\text{cor}(x_t, x_{t-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>c_t</td>
<td>t_t</td>
<td>n_t</td>
<td>y_t</td>
</tr>
<tr>
<td>U.S. Data</td>
<td>0.53</td>
<td>3.36</td>
<td>0.97</td>
</tr>
<tr>
<td>Model</td>
<td>0.49</td>
<td>4.18</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Sticky prices under CIA constraint can also effectively propagate non-monetary shocks. Figure
4 plots the impulse responses of output and employment to a one standard deviation technology
shock and a preference shock respectively. It shows that non-monetary shocks can also generate
hump-shaped output persistence in the model (windows A and C). This feature of the model is
worth emphasizing since it is well known that standard RBC models lack an endogenous propagation
mechanism to explain the hump-shaped, trend reverting output response to transitory shocks
(Cogley and Nason, 1995, Watson, 1993). Here it is shown that sticky prices along can do the job.\footnote{For other mechanisms that can also generate hump-shaped output dynamics, see Wen (1998a,b,c) and Benhabib and Wen (2004).}

One more feature of the model to notice is that employment responds negatively to technology
shocks (see Window D in figure 4). Because sticky prices and CIA constraint render aggregate
demand rigid in the short run, higher total factor productivity thus induces cost-minimizing firms
to lower employment. This negative effect of technology shocks on employment as a result of sticky
prices has been empirically documented and explained by Gali (1999).\footnote{Whether empirical data supports the view that technology shocks generate negative employment movements is controversial. See Chari, Kehoe and McGrattan (2005) for a debate.}

However, in a money-in-
utility general equilibrium model, technology shocks generate positive employment even if prices
are sticky, since investment can increase to absorb the shocks.
3.5 Sensitivity Analysis

The assumption that investment is subject to a CIA constraint is crucial in obtaining our results. In reality, not necessarily all firm’s investments are subject to financing constraints and hence they may not all be tight to internal cash flows. We show here that even only as little as 30% of aggregate investment is subject to money-in-advance constraint, aggregate output remains to be hump-shaped.

To demonstrate, modify the CIA constraint in the model to

\[ C_t + \psi [K_{t+1} - (1 - \delta)K_t] \leq \frac{M_t}{P_t}, \] 

where \( \psi \in [0, 1] \) measures the degree of financial constraint on aggregate investment. The model reduces to the previous model if \( \psi = 1 \) and it reduces to the model of Yun (1996) and Ellision and Scott (2000) if \( \psi = 0 \).

<table>
<thead>
<tr>
<th>Table 2. Sensitivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \psi )</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td>Hump (at ( n^{th} ) quarter)</td>
</tr>
<tr>
<td>Half Life</td>
</tr>
</tbody>
</table>
Table 2 shows that output remains hump-shaped and highly persistent under money shocks even for small values of $\psi$. For example, when $\psi = 0.6$, the peak of output response is not reached until three quarters after the shock, consistent with the U.S. data. The half life is 9 quarters, only slightly shorter than the case of $\psi = 1$. As we reduce $\psi$ further to 0.3, the half life is still 6 quarters long and the peak of output response is still postponed beyond the impact period of the shock to the second quarter, indicating a hump shape. Hump-shaped output disappears when $\psi = 0.2$, but the half-life of output is still more than twice as long as the case of $\psi = 0.20$.

4 Conclusion

We showed in this paper that sticky prices alone can generate strong output persistence if the cash-in-advance constraint is extended to investment. Output exhibits a hump-shaped response pattern even when only as little as 30% of aggregate investment is subject to a CIA constraint. In such a model monetary shocks seem capable of explaining a broad range of business cycle facts better than, or at least as well as, a standard RBC model driven by technology shocks. Hence whether or not sticky prices are responsible for the business cycle is not a theoretical question, but rather an empirical one. Given that multiple mechanisms can give rise to hump-shaped output persistence (see, e.g., Wen 1998a, 1998b, 1998c and Benhabib and Wen 2004, among others), it remains to empirically test which mechanism is the main culprit in propagating the business cycle in the real world. Bills and Klenow (2003), for example, find some empirical evidence against the sticky price propagation mechanism of monetary shocks. Baharad and Eden (2004) find that the staggered price setting assumption is not favored by a micro data. Dittmar, Gavin and Kydland (2005) and Wang and Wen (2005) show that endogenous monetary policy, rather than sticky prices, are more likely to be responsible for the inflation dynamics found in the U.S. data. Thus, to establish sticky prices as a key propagation mechanism of the business cycle, more empirical work is obviously needed.

---

$^{20}$Suppose that investment consists of two parts, $I = I_1 + I_2$, where $I_1$ is subject to CIA constraint and $I_2$ is not. Then the elasticity of investment to cash flow is given by $\frac{\partial I}{\partial m} = \frac{I_1}{I} = \psi$. Using annual data, Worthington (1995) found that the elasticity of investment to cash flow is between 0.2 to 0.65. The implied quarterly elasticity of investment to cash flow should be even higher. Thus in a quarterly model, $\psi \geq 0.6$ is a reasonably good approximation.
References


