Is Less More?: Menu Costs and Estimated Thresholds in the Response of Output to Monetary Policy

Luiggi Donayre[1]
Department of Economics
Washington University in St. Louis

This version: April 2009

Abstract

This paper investigates asymmetries in the response of output to monetary policy shocks of different magnitude. Traditionally in the literature, the threshold determining which shocks are 'large' and which are 'small' has been imposed exogenously. To the extent that such a threshold is misspecified, tests for asymmetry will have low power. In this context, an unobserved components (UC) model of output with a TAR-driven transitory component augmented by a monetary policy variable is estimated. This framework makes it possible to introduce the threshold determining the size of monetary policy shocks as an additional parameter, instead of imposing an ad hoc definition. When the threshold is estimated from the data, there is strong statistical evidence that the response of output to monetary shocks of different size behaves asymmetrically. Moreover, the results found show some support for the implications of menu-cost models.

JEL Classification Code: C32, E32

Keywords: Asymmetry, Monetary Policy, Regime Switching, Threshold Autoregressions, Unobserved Components Model.

[1] I am very thankful to James Morley for valuable insights and suggestions. Helpful comments and discussion from Yunjong Eo, Steven Fazzari, Kyu Ho Kang, Bing Li, Rody Manuelli, Werner Ploberger, Mark Vaughan, and the participants at the Applied Time-Series Research Group and the Graduate Student Conference at Washington University, as well as the participants at the Missouri Valley Economic Association Meeting are gratefully acknowledged. All remaining errors are my own.

1. Introduction

The study of business cycles fluctuations dates back to the early decades of the twentieth century. In this ongoing effort, menu cost models have played an important role providing some insight to the issue of infrequent price adjustments as an explanation to the existence of such fluctuations (Ball and Romer, 1990; Ball and Mankiw, 1994). More recently, the implications of these models have been related to the literature studying asymmetries in the response of output to monetary policy shocks of different size (Ravn and Sola, 2004).

In order to evaluate those implications, this paper focuses on the study of asymmetries in the response of output with respect to the magnitude of monetary shocks. In particular, the contribution of this paper is twofold. First, unlike other documents in the literature, it proposes an empirical framework that allows for the introduction of the threshold that classifies monetary shocks in terms of their size as an additional parameter to be estimated. Second, it links theory and data by providing a way to directly test the implications of menu-cost models.

Theoretically, menu-cost models provide a motivation for distinguishing between monetary shocks that are 'large' or 'small'. Ball and Romer (1990) and Ball and Mankiw (1994) present standard menu-cost models whose results imply that only 'small' monetary shocks have effects on output. When such shocks are 'big', firms find it optimal to adjust their prices[2], leaving real output unchanged.

Despite these theoretical results, the empirical evidence in the literature is mixed. Based on Markov-switching models and multivariate smooth transition vector autoregressions respectively, the results found in Ravn and Sola (2004) and Weise (1999) suggest that the response of output to 'large' and 'small' monetary shocks are different. Sensier (1996) and Lo and Piger (2005), on the other hand, find that this type of asymmetry is not statistically significant.

One possible explanation is that the threshold determining the size of monetary shocks has been set exogenously. Previous authors have defined 'large' shocks as those greater than one standard deviation and, conditional on this exogenous definition of size, have been unable to reject the null hypothesis of linearity (i.e., symmetry). The effects of the 'standard deviation' approach on linearity tests are unclear at best, since there is no economic model supporting this assumption. Indeed, the past mixed evidence reported above can be interpreted as a direct result of the imposition of this exogenous threshold. To the extent that the latter is misspecified, tests for asymmetry will have low power, leading to the inability to reject the null hypothesis. Consistent with that hypothesis, this paper argues for the estimation of the threshold.

A natural way to include the threshold as an additional parameter to be estimated is by means of the Threshold Autoregressive (TAR) processes introduced by Tong (1983). In this paper, the approach proposed unifies the models that have dealt with asymmetries related to the magnitude of the monetary shock. It is similar in spirit to the model proposed by Weise (1999), who estimates a logistic smooth transition vector autoregression (LSTVAR). Nonetheless, it introduces the TAR process within the Unobserved Components (UC) framework from Lo and Piger (2005), LP henceforth. By doing so, the monetary shocks affect only the transitory component of output to be consistent with the notion of money neutrality in the long-run.

---

[2]/ When the monetary shock is 'big', the menu cost becomes relatively small. That is, firms think of it as a secondary cost and, therefore, are willing to incur the menu cost to adjust their prices.

Moreover, the model proposed addresses two issues left unexplored in the existing literature. The first one is motivated by the great moderation. It is widely known that most macroeconomic series have experienced a reduction in volatility since the mid 1980s. To take this phenomenon into account, the model allows for a break in variance. Additionally, it accounts for the possibility that the error terms of the transitory and permanent components of output are correlated (Morley, Nelson and Zivot, 2003, MNZ henceforth) [3].

This paper considers two measures of monetary policy. The first one proxies monetary shocks on the basis of the Federal Funds rate (FFR), whereas the second one employs M1 as the primary monetary instrument. In both cases, there is strong statistical evidence of asymmetries with respect to the size of monetary shocks once the threshold is estimated from the data. Furthermore, when linearity is tested imposing the threshold to be one standard deviation of the shock instead of the estimated threshold from the model, linearity cannot be rejected.

As it is implied by menu-cost models, the results from the estimated model suggest that the response of output to 'small' monetary policy shocks is larger than the response to 'big' shocks. Using generalized impulse response functions (GIRF), the dynamics of the model show that output drops twice as much when a 'small' monetary shock hits the system. Additionally, the analysis of the GIRFs shows that the response of output to 'large' monetary shocks is statistically significant –albeit smaller than the response to 'small' shocks. This result is at odds with the implications of menu-cost models.

The remainder of the paper is organized as follows. The second section motivates the distinction between small and large monetary shocks by exploring the possibilities for asymmetries on output arising in theoretical menu-cost models. In the third section, the empirical methodology is formally described. In section four, the results of the model and linearity tests are presented for both measures of monetary shocks. Some concluding remarks are presented in the fifth section.

2. Theoretical Background

Why should we expect a different behavior in real activity when the size of monetary shocks differs? To motivate this distinction and make the empirical analysis clear, a menu-cost model is presented in this section whose results imply that output responds differently, depending on whether monetary shocks are below or above certain threshold.

The model follows those of Ball and Romer (1990) and Ravn and Sola (2004). The economy consists of N producer-consumer agents, each of whom produces a differentiated good. The price of each good is set by each agent and equals p_j, $j = 1,...,N$. The only friction in the model, which gives rise to nominal rigidities, is introduced through the assumption that each agent must pay a menu-cost if his price is adjusted. After observing the current price level p, each agent sets his price to p_j. However, this price will not necessarily prevail throughout the entire period. By paying a menu cost $c > 0$, any agent j can adjust his price.

Agent j's profit function π_j can be described according to:

---

[3]/ Traditionally, unobserved components models have imposed a zero covariance assumption between the errors of the permanent and transitory components for identification purposes. Morley, Nelson and Zivot (2003) show that, under certain circumstances, it is possible to estimate this covariance.

$$\pi_j = V\left(\frac{m}{p}, \frac{p_j}{p}\right) - cd_j \qquad \ldots (1)$$

where d_j is a dummy variable that equals 1 if agent j adjusts his nominal price and 0 otherwise, $V(.)$ is increasing in both arguments, and m is the money stock in the economy.

In a symmetric equilibrium[4] in which prices are not changed, the optimal price p_j^* for agent j is implicitly defined by the first-order condition:

$$\frac{\partial \pi_j}{\partial p_j} \equiv \frac{\partial V(\frac{m}{p}, \frac{p_j^*}{p})}{\partial(\frac{p_j}{p})} = 0$$

Since the equilibrium is symmetric, any shock to the stock of money m is innocuous[5]. For simplicity, m is normalized to 1 and, thus, the equilibrium corresponds to $m = p = p_j^* = 1$.

To show how the size of the shock to the stock of money can matter, we compare agent j's profit function when his price is left unchanged to his profit when he decides to adjust his price and incur in the menu-cost c, given that all other agents do not.

If the government unexpectedly changes the stock of money to $m \neq 1$, then agents need to evaluate whether it is optimal to leave their prices unchanged or adjust them. In the first case, profit for agent j when prices are not changed (sticky prices) is given by $\pi_{sp} = V(m,1)$. However, if agent j decides to adjust his price, then he must pay the menu-cost c and his profit is given by $\pi_{ap} = V(m, p_j^* / p) - c$. Hence, agents will not adjust their prices if $\pi_{ap} < \pi_{sp}$, which is equivalent to:

$$V(m, p_j^* / p) - V(m,1) < c$$

Ball and Romer (1990) show that, making a second-order Taylor expansion around $m = 1$, the range of money stock for which agents will not adjust their prices (that is, the monetary shock is so small that incurring in the menu-cost c is not optimal) is given by:

$$m \in (1 - \overline{m}, 1 + \overline{m}) \qquad \ldots (2)$$

4/ In a symmetric equilibrium, $p_j = p_0$ for all $j = 1,\ldots,N$.

5/ That is, all agents set the price to $p_j^* = p_0$, for all $j = 1,\ldots,N$, so that their optimal price is consistent with the price level, leaving relative prices, and thus output, unchanged.

where $\overline{m} = \left(\frac{-2cV_{22}}{V_{12}^2} \right)^{1/2}$ and V_{22} (V_{12}) represents the second (mixed) derivative of $V(.)$ with respect to the second argument (first and second arguments). Similarly, the range of the money stock for which agents will adjust their prices (so that changes are neutral) is given by:

$$m \in (-\infty, \overline{\overline{m}}) \cup (\overline{\overline{m}}, \infty) \qquad \text{… (3)}$$

where $\overline{\overline{m}} = \left(\frac{-2c}{V_{22}} \right)^{1/2}$. Hence, 'small' monetary shocks can have real effects if $m \in (1-\overline{m}, 1+\overline{m})$ and 'large' monetary shocks can be neutral if $m \in (-\infty, \overline{\overline{m}}) \cup (\overline{\overline{m}}, \infty)$.

3. Empirical Approach

This section describes the model, the approach undertaken to estimate it and the bootstrap procedure to test whether the TAR-driven transitory component is statistically significant in comparison to a linear transitory component.

3.1. The model

In this section, the empirical approach undertaken to estimate the model is described. Previous models testing for asymmetries regress output growth on measures of 'large' and 'small' policy actions. Thus, the presence of asymmetries involves determining whether the coefficients associated with such measures are statistically different from each other. This approach is at odds with the notion of money neutrality in the long-run.

The model proposed here builds on the UC framework considered by LP. Within this structure, monetary shocks are assumed to affect only the transitory component of output[6]. The main difference with the LP approach is that, while they model the nonlinear relationship between money and output as a Markov-switching process, the regime-switching is driven by estimated, but observable thresholds here. The model can then be described by:

$$y_t = y_t^T + y_t^P \qquad \text{… (4)}$$

$$y_t^p = \mu + y_{t-1}^p + \nu_t \qquad \text{… (5)}$$

$$y_t^T = \sum_{p=1}^{P} \phi_p y_{t-p}^T + \sum_{j=1}^{J} \alpha_j^S x_{t-j} I(|x_t| \le \gamma) + \sum_{j=1}^{J} \alpha_j^L x_{t-j} I(|x_t| > \gamma) + \varepsilon_t \qquad \text{… (6)}$$

[6]/ In a previous version of the model, the monetary shock was allowed to enter the permanent component of output. The coefficients associated with it were not significantly different than zero, however. Therefore, the model considered in this paper only allows the monetary shock to enter the transitory component.

where y_t is a measure of output, y_t^P is the permanent component of output, y_t^T is the transitory component of output, and x_t is a measure of monetary policy.

The system (4)-(6) is a modified version of the simple UC decomposition of real output into the permanent (or stochastic trend) and transitory (or cyclical) components, as in Watson (1986). Following the original model, the permanent component of output is modeled as a random walk with a drift term, given by equation (5). Unlike LP, who allow for the drift term to evolve according to a random walk process as in Clark (1987), μ is treated as a constant in this model.

Equation (6) describes the dynamics of the transitory component of output, y_t^T. It is modeled as an autoregressive process, augmented by the monetary policy variable, x_t. $I(.)$ denotes the indicator function; $|x_t|$ is the threshold variable, since we are interested in how the response of output changes depending on the magnitude of the monetary policy; and γ is the threshold parameter. When $|x_t| \leq \gamma$, the slope-coefficients are captured by the J x *1* vector α^S and when $|x_t| > \gamma$, they are captured by the J x *1* vector α^L. In order to be consistent with the measures of monetary policy considered below, where the monetary variable does not affect output contemporaneously, only lags of x_t enter equation (6).

Note that the coefficients ϕ_p , $p = 1, \ldots, P$ are not state-dependent. Provided that the question of interest concerns the response of output to monetary shocks of different size (i.e, in different regimes), the autoregressive dynamics are assumed to be the same in both regimes. The innovations ε_t and v_t are assumed to follow a normal distribution with mean zero and variance-covariance matrix Ω. Traditionally, UC models have assumed that the covariance between ε_t and v_t is zero (i.e., that Ω is a diagonal matrix), in order to identify the model. Nonetheless, MNZ show that, under certain conditions, UC models can be estimated without imposing such restriction[7]. Thus, in order to explore the possibility that the innovations from the transitory and permanent components of output are correlated, the covariance term is estimated. That is, $Cov(\varepsilon_t, v_t) \neq 0$.

An additional feature of the model estimated here is that it allows for a one time break in the variance-covariance matrix. Output, as well as many other macroeconomic aggregates, has experienced a reduction in volatility since the mid 1980s, an episode known as the great moderation. To account for this fact, an exogenous break date[8] is set to the last quarter of 1983 to split the sample accordingly.

The motivation to account for the great moderation does not exclusively arise from the need to model the reduction in volatility exhibited by output. Koop and Potter (2001) have argued that apparent findings of threshold-type nonlinearities could be due to structural instability instead. Using Monte Carlo simulations, they showed that traditional procedures, which do not consider

[7]/ In an univariate ARMA (*p*,*q*) process, the identification condition to estimate a nonzero covariance between the trend and transitory innovations is that $p \geq q + 2$. See MNZ for further details.

[8]/ This paper is about thresholds, not break dates. Given that many authors have estimated the great moderation to begin in the mid 1980s, the break date is set to the last quarter of 1983, consistent with these findings.

structural instability, led them to incorrectly conclude that threshold-type nonlinearities were present in two of their three data-generated processes. Hence, this criticism is explicitly addressed here by explicitly modeling the break in variance.

3.2. Estimation

Following the approach discussed in Hansen (1997), the estimation of the coefficients of the system (4)-(6) involves an iterative procedure. The model is estimated sequentially for each possible value of the threshold, yielding a γ-dependent loglikelihood function $\log L(\gamma)$ in each iteration. Thus, the maximum likelihood (ML) estimate of γ is the value of this parameter that maximizes $\log L(\gamma)$. Formally, the ML estimate of γ is defined as:

$$\hat{\gamma} = \arg\max_{\gamma \in \Gamma} \{\log L(\gamma)\} \quad \dots (7)$$

where $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ is defined a priori to contain the middle 70% of all possible threshold values to ensure that the model is well identified[9].

The ML estimates of $\alpha^S, \alpha^L, \sigma_\varepsilon, \sigma_\nu, \sigma_{\varepsilon\nu}$ are then given by the parameters associated with $\log L(\hat{\gamma})$. That is, $\hat{\theta} = \hat{\theta}(\hat{\gamma})$ where $\theta = (\alpha^{S\,\prime} \quad \alpha^{L\,\prime} \quad \sigma_\varepsilon \quad \sigma_\nu \quad \sigma_{\varepsilon\nu})'$.

3.3. Testing for a TAR-driven transitory component

Computationally, the estimation of the model is cumbersome due to the sequential iteration of the threshold parameter. Letting $|\Gamma|$ denote the cardinality of Γ, there are $|\Gamma|$ potential thresholds and, therefore, the same number of models to be estimated. This routine becomes even more time-intensive when it is used to test whether the TAR-driven transitory component is statistically significant relative to a linear one. In particular, as it will be explained below, the linearity test involves a bootstrap procedure in which data is generated under the null hypothesis *B* times. For each bootstrap sample, the grid search across possible threshold parameters is carried out. As a consequence, *B* x $|\Gamma|$ potential models need to be estimated. With quarterly data, $|\Gamma|$=135 and for $B = 99$ bootstrap samples, 13,365 potential models would need to be estimated using numerical optimization. Even if convergence for each model was achieved after only 10 seconds, the bootstrap procedure would require 38 hours.

The procedure developed by Hansen (1996) to test TAR processes against linear ones is modified to fit the UC framework (4)-(6). Considering these modifications, the relevant null hypothesis is given by $H_0 : \alpha^S = \alpha^L$. Since this problem is tainted by the existence of nuisance parameters (specifically, the threshold γ is not identified under the null hypothesis), a test with near-optimal

[9]/ It is customary to exclude the 15% of each end of the vector of ordered thresholds to avoid distortions in inference. If possible thresholds that are too close to the beginning or the end of the ordered data were considered, there would not be enough observations to identify the subsample parameters.

power against a wide range of alternative hypotheses is given by the following likelihood ratio statistic:

$$LR = \sup_{\gamma \in \Gamma} \{ LR_n(\gamma) \} \qquad \ldots (8)$$

where

$$LR_n(\gamma) = 2\{ \log \hat{L}_1(\gamma) - \log \hat{L}_0 \} \qquad \ldots (9)$$

is the LR statistic against the alternative $H_1 : \alpha^1 \neq \alpha^2$ when γ is known. $\log \hat{L}_0$ and $\log \hat{L}_1(\gamma)$ correspond to the values of the likelihood functions under the null hypothesis and under the alternative hypothesis for each γ, respectively.

Because γ is not identified, the distribution of the LR statistic (8) is non-standard. Hansen (1996) shows that its asymptotic distribution can be approximated through a bootstrap procedure[10]. Following his work, his approach is modified to fit the system (4)-(6) and the asymptotic distribution of (8) is approximated by a bootstrap experiment in which $y_t^* = y_t^{NH}$, $t = 1, \ldots, n$ where y_t^{NH} is a new dependent variable generated under the null hypothesis. Using y_t^{NH}, a new LR statistic is calculated for this new dependent variable –as in (7)– to form $LR^* = \sup_{\gamma \in \Gamma} \{ 2(L\hat{F}_1^*(\gamma) - L\hat{F}_0^*) \}$.

The procedure described here is similar in spirit to the one detailed in Hansen (1996). The asymptotical equivalence of the likelihood ratio statistic and the original F-statistic in his paper guarantees that his results are carried out to this framework[11].

Concisely, the test to determine whether the TAR-drive transitory component is statistically significant with respect to one in which output responds symmetrically to monetary shocks can be summarized in the following steps:

1. Estimate the model under $H_0 : \alpha^S = \alpha^L$ and obtain the likelihood function $L\hat{F}_0$.
2. Estimate the model under $H_1 : \alpha^S \neq \alpha^L$ and obtain the likelihood function $L\hat{F}_1(\gamma)$.
3. Form the LR statistic $LR = \sup_{\gamma \in \Gamma} \{ 2(L\hat{F}_1(\gamma) - L\hat{F}_0) \}$
4. Bootstrap distribution

    a. Generate a new independent variable $y_t^* = y_t^{NH}$ under the null hypothesis.
    b. Estimate the model under $H_0 : \alpha^S = \alpha^L$ and obtain the likelihood function $L\hat{F}_0^*$.

---

[10]/ See Hansen (1996, 1997) for further details.

[11]/ To evaluate the power and size of the test, a Monte Carlo experiment is also conducted. Data were generated according to the system (4)-(6) and, among the 100 Monte Carlo simulations, only one bootstrapped p-value exceeded 0.05, suggesting that the test has good power at the 5 percent significance level.

c. Estimate the model under $H_1 : \alpha^S \neq \alpha^L$ and obtain the likelihood function $L\hat{F}_1^*(\gamma)$.

d. Form LR statistic $LR^* = \sup_{\gamma}\left\{2(L\hat{F}_1^*(\gamma) - L\hat{F}_0^*)\right\}$.

5. Obtain the bootstrapped p-value as the percentage of bootstrap samples for which $LR^* > LR$.

4. Results

The estimation results from the model specified in the previous section are discussed here. The iterative procedure to obtain the threshold parameter described in (7) involves casting the model (4)-(6) in state-space form and applying the Kalman filter (see the appendix for a general state-space representation of the model). For further details about the Kalman filter, refer to Hamilton (1994) and Kim and Nelson (1999).

Output is measured as 100 times the natural logarithm of real Gross Domestic Product (GDP). This series, as well as all others, is taken from the Federal Reserve Economic Data (FRED) database. To approximate the monetary policy variable, two alternative measures are considered. The first one involves a shock where the Federal Funds rate (FFR) is the monetary instrument, whereas, for the second one, the monetary instrument is given by M1. The results for these two measures are described in sections 4.1. and 4.2., respectively.

4.1. Results for FFR

For the first measure of monetary policy, an interest rate-based monetary shock is constructed from the residuals of an identified VAR, which contains three variables: the Federal Funds rate (FFR), the logarithm of real GDP and the logarithm of the GDP deflator. To identify the shock, the policy variable is ordered last in the VAR (i.e., monetary shocks do not affect output contemporaneously) and four lags of each variable are included. The estimation period for the VAR goes from 1959:Q1 through 2007:Q4, corresponding to 196 observations.

Once the monetary shock is identified, a second step involves the determination of P and J, the number of autoregressive coefficients for y_t^T and the number of lags for the monetary shock, respectively. To do so, the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC) are employed to select among models with different numbers of parameters. Given that quarterly data are used in the estimation, the maximum number of lags for each coefficient is set to 4. Both criteria select the model in which $P = J = 2$. After the model is specified, the estimation approach described in section 3 is applied. For the estimation of the model, the first 20 observations are used as a training sample to avoid the effects of the starting values for the initialization of the Kalman filter[12]. Table 1 reports the estimated coefficients of the model (4)-(6) when monetary policy shocks are measured as the residuals from an FFR-based identified VAR.

[12]/ Since there is no unconditional expectation to initialize the Kalman filter, a high variance is placed on the initial values. To avoid distortions arising from this initialization procedure, and to prevent sensitivity of the model to the initial values, the first 20 observations are disregarded. For further details about the initialization of the Kalman filter, refer to Kim and Nelson (1999).

Table 1
Parameter estimates: UC model with TAR-drive transitory component
Threshold variable: FFR-based VAR residuals

Parameter	Estimate	St. Error	Parameter	Estimate	St. Error
ϕ_1	1.5670	0.0370	σ_ε	0.1047	0.0436
ϕ_2	-0.6138	0.0290	σ_υ	1.0442	0.0977
α_1^S	0.5906	0.1535	$\sigma_{\varepsilon\nu}$	-0.1094	0.0523
α_2^S	-0.5481	0.1282	λ	0.2908	0.0677
α_1^L	0.0147	0.0152	μ	0.7523	0.0531
α_2^L	-0.2111	0.0552	γ	0.1247	-
$\lvert \alpha_1^S \rvert + \lvert \alpha_2^S \rvert$	1.1387				
$\lvert \alpha_1^L \rvert + \lvert \alpha_2^L \rvert$	0.2257		Loglikelihood	-175.23	

Estimated coefficients from the model given in (4) – (6) when the monetary instrument is the FFR. The threshold variable was set to contain the 70 percent middle part of the observations to avoid overfitting. Sample ranges from the first quarter of 1964 through the fourth quarter of 2007 after discarding the first 20 observations to avoid distortions due to the initial effects.

In this table, regime 1 corresponds to the situation in which monetary shocks are ‘small’, as defined by the estimated threshold, while regime 2 corresponds to that in which monetary shocks are defined as ‘large’. The coefficients linking monetary policy to output suggest that the estimated threshold divides policy shocks that have relatively small effects form those that have larger effects. For instance, in regime 1, a change in the FFR of one percent at time *t* reduces output by 0.55 percent two quarters later. In regime 2, however, the same change in the FFR at time *t* reduces output by 0.21 percent two quarters later. In fact, both response coefficients in regime 1, α_1^S and α_2^S, are larger in absolute value than α_1^L and α_2^L. Note, also, that the response coefficient α_1^L is not statistically significant. Furthermore, the sum of the magnitude of the response coefficients in regime 1 is 1.1387, whereas the sum of the magnitude of the response coefficients in regime 2 is 0.2257, around one fifth of the former. As discussed in section 2, this result is consistent with the implications of the menu-cost models from Ball and Romer (1990) and Ball and Mankiw (1994).

The estimated threshold is $\hat{\gamma} = 0.125$, obtained as in equation (7). The standard deviation of the vector of shocks is 0.302, about 2.5 times larger than $\hat{\gamma}$. Thus, the data show that using one standard deviation as the threshold would classify monetary shocks as small when, optimally, they should have been considered large shocks.

From table 1, the estimated coefficients suggest that the variance of output is driven by changes in both the transitory and permanent components. Note, however, that σ_ν is larger than σ_ε, suggesting that the stochastic trend is characterized by low frequency shocks that account for most of the variation in the permanent component of real GDP. To account for the break in variance, the variance-covariance matrix Ω was rescaled after the last quarter of 1983 by a factor λ. Hence, the fact that this factor is less than 1 and significant supports the notion that output growth volatility has reduced since the mid 1980s. In particular, it suggests that volatility has reduced to a third after 1984.

The covariance of the innovation terms, $\sigma_{\varepsilon v}$, is negative and statistically significant, although relatively small. The correlation coefficient, $\rho = -0.9999$, implies that the permanent and transitory components of output are (almost perfectly) negatively correlated, consistent with the findings in the literature (see MNZ, Sinclair (2007), among others).

To test whether the TAR-driven transitory component is significantly better than a linear one, the bootstrap procedure described in section 3.3. is applied using 99 bootstrap samples[13]. The bootstrapped *p-value* that the procedure yields is 0.05. Thus, linearity is rejected at the 5% level and equation (6) is favored in detriment of a linear one. Interestingly, when the same bootstrap procedure is applied to test linearity after imposing the threshold to be one standard deviation of the monetary shocks, linearity cannot be rejected. This supports the hypothesis that previous authors have not been able to reject linearity because they imposed an ad hoc threshold, instead of estimating it directly from the data.

If the Fed ought to implement small changes in the FFR to have a large impact on output, a natural question that arises in this context refers to whether such a change must be carried out all at once or gradually. That is, it is important to understand whether the size of the shock matters for itself, or whether it depends on the frequency of the data. This question is closely linked to the interest-rate smoothing literature and can have important policy implications. To address this issue, the model in (4) – (6) is estimated using annual data from 1959 through 2007. If the size matters relative to the frequency of the data, the estimated threshold using annual data should be expected to be, approximately, four times the estimated threshold using quarterly data. When annual data is used, the estimated threshold is $\hat{\gamma}^{A} = 0.443$, slightly less than four times the estimated threshold using quarterly data, $\hat{\gamma}^{Q} = 0.125$.

Although the relationship is not perfect, it suggests that the size of the shock matters when it is relative to the frequency of the data. This finding is in line with the notion that the Fed, and many other central banks, seeks to 'smooth' interest rates to minimize the volatility of the FFR changes. Because economic agents are forward-looking, changes in the monetary instrument today will be more effective if they are expected to persist over time. Hence, by smoothing interest rates, the size of the change in the FFR required to reduce fluctuations in the economy can be smaller than it would otherwise be necessary.

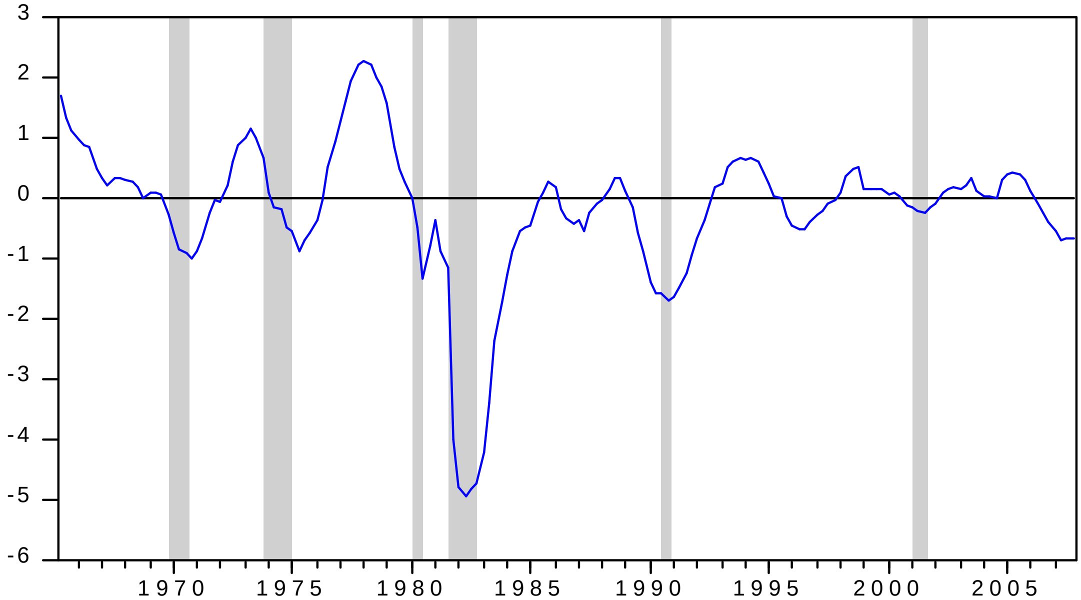
Figure 1 below depicts the estimated transitory component of real GDP. Recessions as defined by the NBER, are shown in shaded areas. Even though the covariance term is statistically significant, as observed in table 1, it is relatively small. Because of this fact, the estimated transitory component is similar to those from the UC literature (see MNZ, LP, Sinclair (2007), among others). Indeed, it corresponds closely to recessions in the economy and exhibits sharp declines followed by more gradual increases.

Figure 1

Estimated TAR-driven transitory component

[13]/ Given that a grid-search over all possible values of the threshold parameters is necessary to estimate the model, bootstrapping the asymptotic p-value is very time-consuming. As a consequence, only 99 bootstrap samples were used to test linearity. Even though a small number of bootstrap samples weaken power, this should not be a problem as linearity is rejected.

Threshold: FFR-based VAR residuals



Estimated TAR-driven transitory component from the model in equations (4) – (6), when the monetary policy variable is measured as the residuals from an FFR-based VAR. The sample goes from the second quarter in 1965 through the fourth quarter 2007. NBER recession dates are shaded.

Even though the response coefficients are larger, on impact, when 'small' monetary shocks hit the economy, it is important to evaluate these responses over time, given the nonlinear nature of the model. Simple impulse-response functions (IRF) are a convenient way to analyze the difference in the response of output to monetary shocks of different size over time. However, when the model is nonlinear, such as the one in equations (4) – (6), the IRFs are sensitive to the history of the system and the future shocks assumed to hit it. Koop *et al.* (1996) examine these issues in detail.

In order to address these problems, generalized impulse-response functions (GIRF) are constructed, following Koop *et al.* (1996). The model is assumed to be known, so sample variability is not taken into account. Moreover, attention is restricted to the transitory component of output using the estimated parameters from table 1. To compute the GIRFs, the following procedure is implemented (see appendix for a detailed description). First, monetary shocks and idiosyncratic shocks for periods 1 to q are drawn, with replacement, from the residuals of the identified VAR and the estimated transitory component, respectively and, for a given history of the system[14], fed through equation (6) to produce a simulated data series. This originates a forecast of the transitory component conditional on a particular history and sequence of shocks (both monetary and idiosyncratic) for q periods ahead. Second, the same procedure is carried out, given the same particular history and sequence of shocks, with the exception that the monetary shock to the transitory component of output in period 0 is fixed at a particular value. The shocks are fed through equation (6) and a forecast is produced as explained above. Third,

---

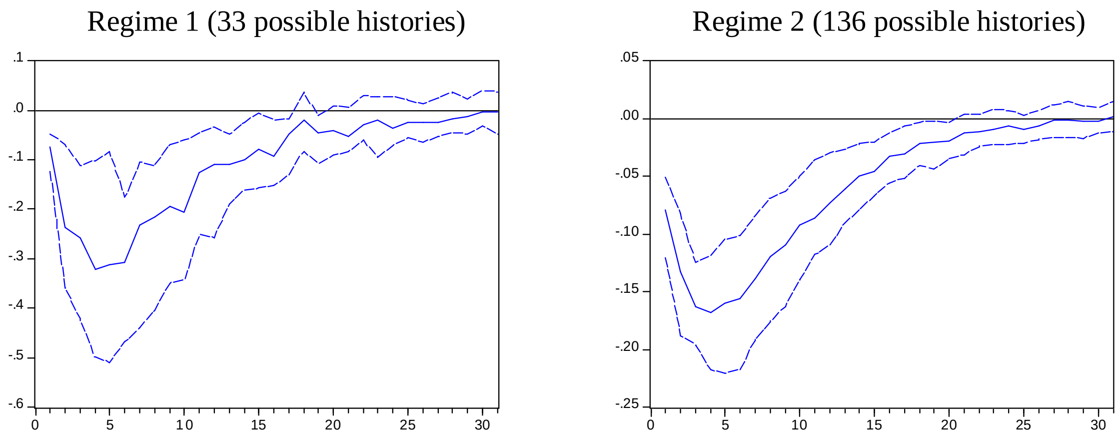
[14]/ The GIRFs are averaged over different histories taken from subsamples of the data. For instance, the GIRFs for the 'small' monetary shock regime are computed averaging out over histories (or initial values for $y_{t-1}^{T} y_{t-2}^{T}$,) corresponding to all dates in which the given monetary shock was smaller than $\hat{\gamma}$.

these steps are repeated 100 times and averaged out across individual Monte Carlo repetitions. Fourth, given the arbitrary shock and particular history, the difference between the averaged forecasts is taken to form a Monte Carlo estimate of the GIRF.

For regimes 1 and 2, respectively, positive 'small' and 'large' monetary shocks to the transitory component of output in period zero are fixed so that they fall below and above the estimated threshold $\hat{\gamma} = 0.125$[15]. In particular, they are set to 0.08 and 0.23. Figure 2 presents the GIRFs for the transitory component of output when the monetary shock is measured as the residuals from an FFR-based VAR.

The left panel of figure 2 plots the response of y_t^T for $q = 30$ periods ahead in regime 1, that is when the monetary shock hitting the system is 'small', corresponding to 33 possible histories. The right panel plots the response of y_t^T for the same number of periods ahead in regime 2, corresponding to 136 possible histories. Since each particular history generates a given forecast of y_t^T, we report the median of these forecasts, together with the 25th and 75th quantiles (dashed lines), which serve as bands for the GIRFs.

Figure 2
Generalized Impulse-Response Functions
Threshold: FFR-based VAR residuals



Generalized impulse-response functions of y_t^T to a positive shock to the monetary policy variable measured as the residuals from an FFR-based VAR, computed as described in section 4. The size of the shocks corresponds to a standard deviation difference between the small and large shocks, with the estimated threshold as the middle point.

From figure 2, three facts can be observed. First, on impact, monetary shocks do not have an effect on output since only lags of x_t enter equation (6). In period 1, however, the response of the transitory component of output to 'small' monetary shocks is similar than its response to 'large' monetary shocks. Because the 'large' shock is almost three times larger than the 'small' shock,

[15]/ This guarantees that the 'small' ('large') shock is below (above) the estimated threshold, triggering a response of output captured by the $\alpha^S (\alpha^L)$ coefficients.

this fact tends to compensate the difference in the magnitude of the response coefficients in regimes 1 and 2, as shown in table 1.

A second fact is more supportive of the results found in table 1. Over time, the response of the transitory component of output in regime 1 (when 'small' monetary shocks hit the economy) is larger than that in regime 2. Despite similarities in the response of the transitory component of output to 'small' and 'large' shocks on impact, the dynamics of the model are such that the difference in the response coefficients under each regime prevails over time. Graphically, this is easily seen by comparing the magnitude of the median response of the transitory component of output in each regime. Such median responses reach their maxima in period 4 for regime 1 (-0.32) and period 5 for regime 2 (-0.22). The difference in the magnitude of the response of the transitory component becomes even bigger if the 75^{th} quantiles are considered. Their maxima are reached in period 5 for regime 1 (-0.51) and period 4 for regime 2 (-0.22). Furthermore, the accumulated median response of the transitory component of output in regime 1 is almost twice that in regime 2 (-3.38 and -1.82, respectively). Based on the evidence in these two facts, the transitory response of output exhibits an overall larger response when 'small' monetary shocks hit the economy, even after controlling for future monetary and idiosyncratic shocks hitting the system and for the history of the economic conditions. That is, the implications of menu-cost models regarding these facts hold true.

Nonetheless, not all implications of menu-cost models are supported. The third fact show that, contrary to the predictions of menu-cost models, the response of output to 'large' monetary shocks is statistically different than zero, as can be observed from figure 2 (the bands do not contain the zero line). An explanation to this inconsistency between theory and data resides in the implicit assumption behind the implications of menu-cost models. According to these models, when 'large' monetary shocks disturb the economy, the menu-cost becomes secondary and relatively small and, thus, agents adjust their prices, leaving real balances and, thus, real output, unaffected. However, this result assumes that all firms adjust their prices and that they adjust it to the optimal price level. Neither of these assumptions seems to hold true as shown in the data. Firms may be heterogeneous in the way they interpret monetary shocks. Moreover, they face imperfect information in the sense that, even when the menu cost is relatively small, their adjusted prices need not match the optimal price level.

Ss models, on the other hand, can generate results whose implications are consistent with 'large' shocks triggering significant effects on output. These models link infrequent price adjustment at the microeconomic level with aggregate price stickiness by means of fixed-cost inventory adjustments. More importantly, they do not assume that firms face perfect information. Thus, the findings in this paper support the implications of menu-cost models only to a certain extent: they do generate a larger response of output to 'small' monetary shocks but they do not support their implications regarding the neutrality of 'large' monetary shocks.

4.2. Results for M1

For the second measure of monetary policy, an M1-based monetary shock is constructed from the residuals of an identified VAR, which contains four variables: the FFR, the logarithm of real GDP, the logarithm of the GDP deflator and the logarithm of M1. To identify the shock, the policy variable M1 is ordered last so that monetary shocks do not affect output contemporaneously –as in the previous case– and four lags of each variable are included. The estimation period for the VAR goes from 1959:Q1 through 2007:Q4, corresponding to 196 observations. As before, the first 20 observations are used as a training sample to avoid the effects of the starting values for the initialization of the Kalman filter.

The number of autoregressive coefficients for y_t^T, *P*, and the number of lags for the monetary shock, *J*, are selected as before, setting $P = 1$ and $J = 2$ for this case.

Once the model is specified, it is estimated as in section 3. Table 2 reports the estimated coefficients of the model (4)-(6) when monetary policy shocks are measured as the residuals from an M1-based identified VAR.

Like in the previous case, regime 1 corresponds to the situation in which monetary shocks are 'small', as defined by the estimated threshold, while regime 2 corresponds to that in which monetary shocks are defined as 'large'. In general, the results where M1 is the monetary instrument are similar than the case in which the monetary instrument is the FFR. In particular, both α_1^S and α_2^S are larger in absolute value than either α_1^L or α_2^L. Moreover, the sum of the magnitude of the response coefficients in regime 1 is 26.13, whereas the sum of the response coefficients in regime 2 is 4.37. Here too, the response of output to 'small' monetary shocks is larger than the response to 'large' shocks, as evidenced in the response coefficients. The estimated threshold is $\hat{\gamma} = 0.0047$, obtained as in equation (7), and is about half of the standard deviation of the vector of M1-based monetary shocks (0.0089). Hence, using the one standard deviation approach tends to overstate the optimal threshold parameter.

Table 2
Parameter estimates: UC model with TAR-driven transitory component
Threshold variable: M1-based VAR residuals

Parameter	Estimate	St. Error	Parameter	Estimate	St. Error
ϕ_1	0.6229	0.1210	σ_ε	2.0727	0.3843
α_1^S	4.3122	7.9978	σ_υ	1.6727	0.3034
α_2^S	21.8193	7.9087	$\sigma_{\varepsilon\nu}$	-3.4665	1.2490
α_1^L	-3.3714	3.6984	λ	0.1981	0.0438
α_2^L	-0.9966	2.6991	μ	0.7379	0.0717
$\lvert\alpha_1^S\rvert + \lvert\alpha_2^S\rvert$	26.1315		γ	0.0047	-
$\lvert\alpha_1^L\rvert + \lvert\alpha_2^L\rvert$	4.3680		Loglikelihood	-172.442	

Estimated coefficients from the model given in (4) – (6) when the monetary instrument is M1. The threshold variable was set to contain the 70 percent middle part of the observations to avoid overfitting. Sample ranges from the first quarter of 1964 through the fourth quarter of 2007 after discarding the first 20 observations to avoid distortions due to the initial effects. Standard errors in parentheses.

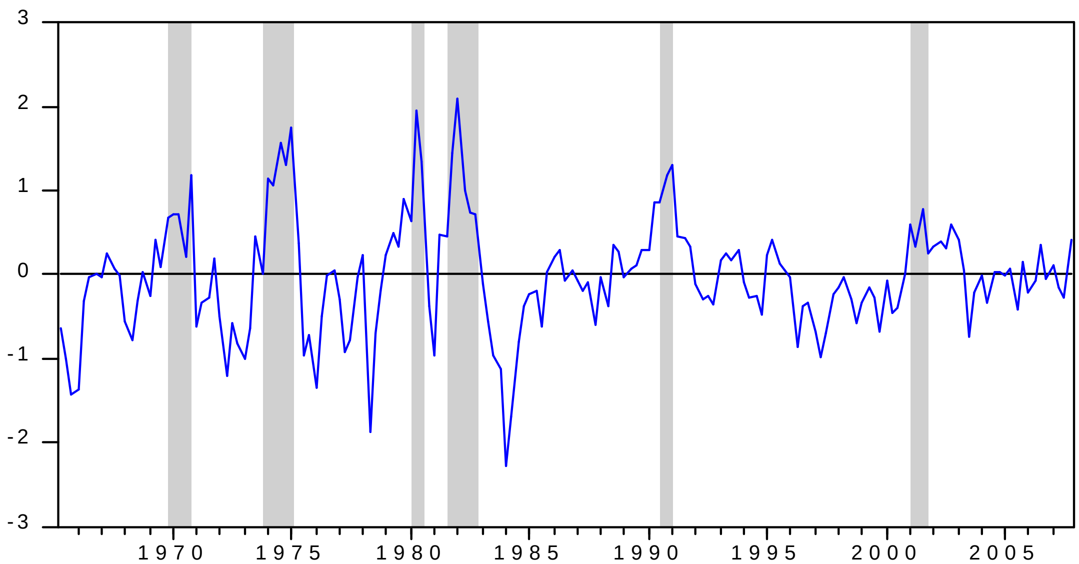
The bootstrap procedure to test whether the transitory component is significantly different than a linear one is applied as explained in section 3. The p-value that the procedure yields is 0.03, providing statistical evidence to the existence of asymmetries related to the size of monetary shocks when the monetary instrument is M1. Furthermore, when the same bootstrap procedure was carried out to test linearity imposing the threshold to be one standard deviation of the vector of monetary shocks, the null hypothesis could not be rejected, like in the previous case.

The model was then estimated using annual data to determine whether the size of the shock matters for itself or relative to the frequency of the data. When annual data is used, the estimated threshold is $\hat{\gamma}^A = 0.0185$, almost perfectly four times the estimated threshold using quarterly data, $\hat{\gamma}^Q = 0.0047$. Therefore, the results using M1 as the monetary instrument also support the notion that the size of monetary shocks matter when relative to the frequency of the data.

Figure 3 depicts the estimated transitory component of real GDP. Recessions, as defined by the NBER, are shown in shaded areas, as before. Unlike the case in which x_t was FFR-based, however, the transitory component is much noisier and does not correspond with the NBER-dated recession dates. The reason for this is that the covariance between the permanent and transitory errors is significantly different than zero, and large. According to MNZ, once the order condition for the identification of an unrestricted ARMA(*p*,*q*) model is met, the only difference between UC cycles and Beveridge-Nelson cycles is the assumption that $Cov(\varepsilon_t, v_t) = 0$. This implies that when the covariance between innovations is significantly different than zero, the transitory component of a UC model is identical to the transitory component from the Beveridge-Nelson decomposition.

If this is the case, then the transitory component ought to be much smaller in amplitude and less persistent than the transitory component from a restricted UC model. In the Beveridge-Nelson decomposition, the variation in any particular univariate model is driven, mainly, by the permanent component. As a consequence, the transitory component is much noisier. Consistently, figure 3 shows a smaller and noisier transitory component than that depicted in figure 1. As before, the correlation coefficient between the transitory and permanent innovations is highly negative ($\rho = -0.99$).

Figure 3
Estimated TAR-driven transitory component
Threshold: M1-based VAR residuals



Estimated TAR-driven transitory component from the model in equations (4) – (6), when the monetary policy variable is measured as the residuals from an M1-based VAR. The sample goes from the second quarter in 1965 through the fourth quarter 2007. NBER recession dates are shaded.

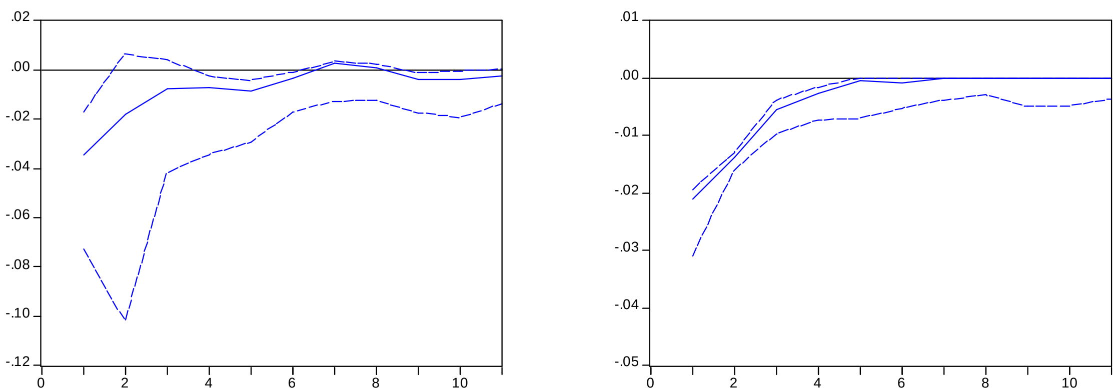
To determine the effects of ‘small’ and ‘large’ monetary shocks on the transitory component of output over time, GIRFs are computed for the case in which the monetary policy instrument is M1. The functions are calculated following the same steps described in the previous subsection.

Figure 4 plots the response of y_t^T for $q = 10$ periods ahead. The left panel shows the response of the transitory component of output to a negative monetary shock in regime 1 (i.e., the response of output to a ‘small’ monetary shock) for all 75 possible histories. The right panel displays the response of the transitory component of output to a negative monetary shock in regime 2, when the system undertakes a ‘large’ monetary shock, corresponding to 94 possible histories. As before, the 25^{th}, 50^{th} and 75^{th} quantiles (dashed lines) are reported.

The results from the GIRFs in figure 4 are consistent with those found for the FFR-based monetary policy shock shown in figure 2. Unlike the previous case, however, the response of the transitory component of output to ‘small’ monetary shocks (i.e, in regime 1) is larger than the response to ‘large’ monetary shocks, even on impact (with one lag since x_t does not enter the transitory component). Over time, this difference persists. The accumulated median responses of output are -0.0869 and -0.0446 in regimes 1 and 2, respectively.

Figure 4
Generalized Impulse-Response Functions
Threshold: M1-based VAR residuals

Regime 1 (33 possible histories) Regime 2 (136 possible histories)



Generalized impulse-response functions of y_t^T to a positive shock to the monetary policy variable measured as the residuals from an M1-based VAR, computed as described in section 4. The size of the shocks corresponds to a standard deviation difference between the small and large shocks, with the estimated threshold as the middle point.

Finally, the response of output to ‘large’ monetary shocks is significantly distinct from zero, as in figure 2, up to the sixth period. Thus, as suggested in the previous subsection, the implications of menu-cost models regarding the response of output to ‘large’ monetary shocks might be at odds with the data.

5. Concluding Remarks

An asymmetric relationship between real aggregate economic activity and monetary policy can arise with respect to different characteristics of these variables. This paper focuses on the asymmetric relationship between output and monetary policy that refer to the size of the monetary innovations. Provided that such asymmetric effects could have strong implications not only for the way economic agents think about the economy, but also for the way economic policy is conducted, it then seems important to find an empirical consensus.

It is claimed in this paper that the potential source of mixed results in the literature studying the asymmetric effects of monetary policy shocks of different size on output is the fact that the threshold determining which shocks are 'small' and which are 'large' has been set exogenously. To overcome this situation, an unobserved components model is proposed in which the monetary threshold in the transitory component is introduced as an additional parameter to be estimated.

Once this threshold is estimated from the data, there is strong evidence of such asymmetry. In particular, the null hypothesis of linearity is rejected for either measure of monetary policy employed. Furthermore, when the bootstrap procedure to test linearity is carried out using the one-standard deviation approach for the threshold determining the size of monetary shocks instead of the estimated threshold from the model, linearity cannot be rejected. This supports the hypothesis that the optimal threshold should be estimated from the data and not imposed ad hoc.

The estimated coefficients suggest that the response of output to 'small' monetary shocks is larger than the response to 'large' monetary shocks, as it is implied by menu-cost models. As explained in section 4, the sum of the magnitude of the coefficients in regime 1 –when 'small' monetary shocks hit the economy– is larger than that sum in regime 2.

GIRFs were computed to analyze the dynamics of the model, given its nonlinear nature. Three facts were observed. In the first place, the response of the transitory component of output is larger on impact in regime 1. That is, when 'small' monetary shock hit the economy (especially, in the case in which the monetary instrument is M1).

Over time, this difference becomes more pronounced, consistent with the implications of menu-cost models. When the monetary instrument is the FFR, the response of the transitory component of output in regime 1 reaches a maximum of -0.32 in period 4, while the maximum response of output when monetary shocks are 'large' is -0.22, reached in period 5. When the monetary instrument is M1, the maximum response of output when monetary shocks are 'small' is -0.035, while that when monetary shocks are 'large' is only -0.021, both reached in the first period. Moreover, the accumulated response of the transitory component of output after a 'small' monetary shock hits the system is, approximately, twice the accumulated response after a 'large' shock for both measures of monetary policy.

Finally, the path of the response of output to 'large' monetary shocks reveals that such responses are significantly different than zero. This result is at odds with the implications of menu-cost models. A potential explanation to this fact arises from the implicit assumptions in these models. In particular, they assume that when a shock is 'large', all firms adjust their prices. That is, all firms are homogeneous, when this is not necessarily a valid assumption. Furthermore, menu-cost models assume that when firms do adjust their prices, they do so optimally. That is, firms have perfect information. In these lines, it would be interesting to evaluate the response of output to monetary shocks of different size when firms are allowed to be heterogeneous in their responses.

References

Ball, Laurence and David Romer (1990). “Are Prices Too Sticky?”. *Quarterly Journal of Economics* 104(3), pp. 507-524.

Ball, Laurence and N. Gregory Mankiw (1994). “Asymmetric Price-Adjustment and Economic Fluctuations”. *The Economic Journal* 423, pp. 247-261.

Bernanke, Ben S. and Mark Gertler (1995). “Inside the Black Box: The Credit Channel of Monetary Policy Transmission”. *Journal of Economic Perspectives* 9, pp. 27-48.

Clark, Peter K. (1987). “The Cyclical Component of U.S. Economic Activity”. *Quarterly Journal of Economics* 102, pp. 797-814.

Cover, James P. (1992). “Asymmetric Effects of Positive and Negative Money-Supply Shocks”. *Quarterly Journal of Economics* 107, pp. 1261-1282.

DeLong, J. Bradford and Lawrence H. Summers (1988). “How does Macroeconomic Policy Affect Output?” In *Brookings Papers on Economic Activity*, edited by William C. Brainard and George L. Perry, pp. 433-494. Washington, DC: The Brookings Institution.

Eo, Yunjong and James Morley (2008). “Likelihood-based Confidence Sets for the Timing of Structural Breaks”, MPRA Paper 10372. Munich: University Library of Munich.

Evans, Paul (1986). “Does the Potency of Monetary Policy Vary with Capacity Utilization?” *Carnegie Rochester Conference Series on Public Policy* 24, pp. 303-332.

García, René (1998). “Asymptotic Null Distribution of the Likelihood Ratio Test in Markov-Switching Models”. *International Economic Review* 39, pp. 763-788.

García, René and Huntley Schaller (2002). “Are the Effects of Interest Rates Changes Asymmetric?” *Economic Inquiry* 40, pp. 801-812.

Hamilton, James D. (1999). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

Hansen, Bruce E. (1992). “The Likelihood Ratio Test under Non-standard Conditions: Testing the Markov-Switching Model of GNP”. *Journal of Applied Econometrics 7*, pp. S61-S82.

Hansen, Bruce E. (1996). “Inference when a Nuisance Parameter is not Identified under the Null Hypothesis”. *Econometrica 64*, pp. 413-430.

Hansen, Bruce E. (1997). “Inference in TAR Models”. *Studies in Nonlinear Dynamics and Econometrics* 2(1), pp. 1-14.

Kaufmann, Sylvia (2001). “Is there an Asymmetric Effect of Monetary Policy over Time?: A Bayesian Analysis using Austrian Data”. *Working Papers* 45, Öesterreichische Nationalbank.

Kim, Chang-Jin and Charles R. Nelson (1999). *State Space Models with Regime-Switching. Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA: The MIT Press.

Kim, Chang-Jin and Charles R. Nelson (1999). "Friedman's Plucking Model of Business Cycles Fluctuations: Tests and Estimates of Permanent and Transitory Components". *Journal of Money, Credit and Banking* 31, pp. 317-334.

Koop, Gary and Simon Potter (2001). "Are Apparent Findings of Nonlinearity due to Structural Instability of Economic Time Series?". *Econometrics Journal*, Vol. 4. pp. 37-55.

Koop, Gary, M. Hashem Pesaran and Simon M. Potter (1996). "Impulse Response Analysis in Nonlinear Multivariate Models". *Journal of Econometrics* 74, pp. 119-147.

Lo, Ming Chien and Jeremy Pager (2005). "Is the Response of Output to Monetary Policy Asymmetric? Evidence from a Regime-Switching Coefficients Model". *Journal of Money, Credit and Banking* 37, pp. 865-887.

MacKinnon, James (2006). "Bootstrap Methods in Econometrics". Queen's Economics Department Working Paper No. 1028. Ontario: Queen's University.

Morley, James C. (2007). "Nonlinear Time Series in Macroeconomics". *Encyclopedia of Complexity and System Science,* forthcoming.

Morley, James C., Charles R. Nelson and Eric Zivot (2003). "Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP so Different?". *The Review of Economics and Statistics* 85(2), pp. 235-243.

Peersman, Gert and Frank Smets (2002). "Are the Effects of Monetary Policy in the Euro Area Greater in Recessions than in Booms?". In *Monetary Transmission in Diverse Economies*, edited by Lavan Mahadeva and Peter Sinclair. Cambridge, UK: Cambridge University Press.

Ravn, Morten and Martin Sola (2004). "Asymmetric Effects of Monetary Policy in the U.S.: Positive versus Negative or Big versus Small?". Federal Reserve Bank of St. Louis Review 86(5), pp. 41-60.

Sensier, Marianne (1996). "The Asymmetric Effect of Monetary Policy in the U.K.". Unpublished Working Paper. University of Oxford.

Sinclair, Tara M. (2007). "Asymmetry in the Business Cycle: Revisiting the Friedman Plucking Model". Working Paper. The George Washington University.

Stock, James H. and Mark Watson (2002). "Has the Business Cycle Changed and Why?". *NBER Macroeconomics Annual* 17, pp. 159-218.

Teräsvirta, Timo (1994). "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models". *Journal of the American Statistical Association* 89, pp. 208-218.

Thoma, Mark (1994). "Subsample Instability and Asymmetries in Money-Income Causality". *Journal of Econometrics* 64, pp. 279-306.

Tong, Howell (1983). *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics. New York-Berlin: Springer-Verlag.

Watson, Mark W. (1986). “Univariate Detrending Methods with Stochastic Trends”. *Journal of Monetary Economics* 18, pp. 49-75.

Weise, Charles (1999). “The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach”. *Journal of Money, Credit and Banking* 31, pp. 85-108.

Appendix 1: State-space representation

The state-space representation for the general *P=p* and *J=j* system given in equations (3) – (5) is provided here. The observation equation is given by:

$$y_t = \begin{bmatrix} 1 & 0 & \ldots & 0 & 1 \end{bmatrix}_{1x(p+1)} \begin{bmatrix} y_t^T \\ y_{t-1}^T \\ \vdots \\ y_{t-p-1}^T \\ y_t^P \end{bmatrix}_{(p+1)x1}$$

The state equation is given by:

$$\begin{bmatrix} y_t^T \\ y_{t-1}^T \\ \vdots \\ y_{t-p-1}^T \\ y_t^P \end{bmatrix}_{(p+1)x1} = F_{(p+1)x(p+1)} \begin{bmatrix} y_{t-1}^T \\ y_{t-2}^T \\ \vdots \\ y_{t-p}^T \\ y_{t-1}^P \end{bmatrix}_{(p+1)x1} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mu \end{bmatrix}_{(p+1)x1} + G^i_{(p+1)xj} \begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-j} \end{bmatrix}_{jx1} + \begin{bmatrix} \nu_t \\ 0 \\ \vdots \\ 0 \\ \varepsilon_t \end{bmatrix}_{(p+1)x1} \quad \text{for } i=1,2$$

where:

$F = \begin{bmatrix} \Phi_{1xp} & 0 \\ I_{p-1} & 0_{px2} \\ 0_{1xp} & 1 \end{bmatrix}$, with $\Phi_{1xp} = \begin{bmatrix} \phi_1 & \ldots & \phi_p \end{bmatrix}$, I_{p-1} being the identity matrix of order (*p-1*) and 0_{ixj} being a matrix with *i* rows and *j* columns of zeros;

$$G^1 = \begin{bmatrix} \alpha_1 & \dots & \alpha_j \\ 0 & & 0 \\ & \ddots & \\ 0 & & 0 \end{bmatrix}; \text{ and } \quad G^2 = \begin{bmatrix} \beta_1 & \dots & \beta_j \\ 0 & & 0 \\ & \ddots & \\ 0 & & 0 \end{bmatrix}$$

The variance-covariance matrix of the transitory component is given by:

$$Q = E\left\{ \begin{bmatrix} v_t \\ 0 \\ \vdots \\ 0 \\ \varepsilon_t \end{bmatrix} \begin{bmatrix} v_t & 0 & \dots & 0 & \varepsilon_t \end{bmatrix} \right\} = \begin{bmatrix} \sigma_v^2 & 0 & \dots & 0 \\ 0 & 0 & & 0 \\ & & \ddots & \\ 0 & \dots & 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

Appendix 2: Computation of Generalized Impulse Response Functions

The procedure to compute the generalized impulse-response functions (GIRFs) follows the one described in Koop *et al.* (1996). The reader is referred there for formal statistical background.

A GIRF can be defined as the effect of a one-time shock on the forecast of variables in a particular model, given a specific history. The response constructed must then be compared to a benchmark "no shock" scenario. In this way, the GIRF can be expressed as follows:

$$GI_Y(q, \nu_t, \omega_{t-1}) = E\left[Y_{t+q} \mid \nu_t, \omega_{t-1}\right] - E\left[Y_{t+q} \mid \omega_{t-1}\right] \qquad \text{for } q = 0, 1, \ldots$$

where GI_Y is the generalized impulse-response function of a variable Y for period q, given the specific history ω_{t-1} and initial shock ν_t, and $E[.]$ is the expectations operator.

To compute the GIRF, we simulate the conditional expectations in the equation above. The nonlinear model is assumed to be known (i.e., sample variability is ignored). The shock to Y, ν_0, occurs in period 0, and responses are computed for q periods ahead. Thus, the GI_Y function is generated according to the following steps:

1. Pick a history $\omega_{i,t-1}$. The history is the actual value of the lagged endogenous variables at a particular date, or for a particular episode (e.g., those values of the endogenous variables that fall under regime 1).
2. Pick a sequence of (2-dimensional) shocks $\nu_{j,t+q}$, $q = 0, 1, \ldots, n$. This vector of shocks includes both monetary and idiosyncratic shocks. They are drawn with replacement from the vector of monetary shocks –the residuals from the identified VAR- and from the estimated residuals of the transitory component of the model (that is, from the vector of residuals from equation (6)).
3. Using $\omega_{i,t-1}$ and $\nu_{j,t+q}$, simulate the path for y_{t+q} over n periods according to equation (6). This benchmark path is denoted as $Y_{t+q}(\omega_{i,t-1}, \nu_{j,t+q})$ for $q = 1, \ldots, n$.
4. Using the same $\omega_{i,t-1}$ and $\nu_{j,t+q}$, plus an additional initial shock ν_0, simulate the path for y_{t+q} over $n+1$ periods according to equation (6). This profile path is denoted $Y_{t+q}(\nu_0, \omega_{i,t-1}, \nu_{j,t+q})$ for $q = 0, 1, \ldots, n$.
5. Repeat steps 2 to 4 B times.
6. Repeat steps 1 to 5 R times and compute the quantiles of the difference between the profile and benchmark paths $Y_{t+q}(\nu_0, \omega_{i,t-1}, \nu_{j,t+q})$ - $Y_{t+q}(\omega_{i,t-1}, \nu_{j,t+q})$.