Macroeconomic Dynamics, **19**, 2015, 58–78. Printed in the United States of America.
doi:10.1017/S1365100513000217

THE QUANTITY THEORY REVISITED: A NEW STRUCTURAL APPROACH

MAKRAM EL-SHAGI AND SEBASTIAN GIESEN
Halle Institute for Economic Research

LOGAN J. KELLY
University of Wisconsin, River Falls

We propose a unified identification scheme to identify monetary shocks and track their propagation through the economy. We combine three approaches dealing with the consequences of monetary shocks. First, we adjust a state space version of the P-star type model employing money overhang as the driving force of inflation. Second, we identify the contemporaneous impact of monetary policy shocks by applying a sign restriction identification scheme to the reduced form given by the state space signal equations. Third, to ensure that our results are not distorted by the measurement error exhibited by the official monetary data, we employ the Divisia M4 monetary aggregate provided by the Center for Financial Stability. Our approach overcomes one of the major difficulties of previous models by using a data-driven identification of equilibrium velocity. Thus, we are able to show that a P-star model can fit U.S. data and money did indeed matter in the United States.

Keywords: Divisia Money, State Space Decomposition, Sign Restrictions

1. INTRODUCTION

Ever since Sims (1980) cast doubt on the informational content of monetary aggregates, the question of how monetary policy shocks propagate throughout the economy has been of interest to the profession. But although the relationship between money and prices as suggested by the quantity theory has often been validated using cross-country data [see, e.g., McCandless and Weber, 1995; Teles and Uhlig (2010)], tracking the transmission of monetary shocks within a given economy remains a hotly debated topic. Most of the recent literature concentrates on the role of interest rates in explaining the short-run dynamics of inflation as suggested by the New Keynesian paradigm, and thereby ignores the long-run relationship of money growth and inflation. In this paper, we provide a structural

The authors would like to thank William A. Barnett, Jane Binner, Walter Hyll, John Keating, Gunnar Pippel, and Hans-Eggert Reimers for valuable comments and suggestions. We also benefited from seminar participants at the University of Kansas 2011, the Annual Conference of the Scottish Economic Society 2011, the Annual Conference of the Royal Economic Society 2011, the Midwest Macroeconomics Meetings 2011, and the 15th International Conference on Macroeconomic Analysis and International Finance (Crete). Address correspondence to: Makram El-Shagi, Halle Institute for Economic Research, Kleine Märkerstraße 8, D-06108 Halle (Saale), Germany; e-mail: Makram.El-Shagi@iwh-halle.de.

© 2013 Cambridge University Press 1365-1005/13

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

econometric approach that allows both monetarist and New Keynesian transmission of monetary policy shocks by combining the three strands of literature that have dealt with these issues: the P-star literature, the structural VAR literature, and the monetary aggregation literature.

The P-star literature, beginning with Hallman et al. (1991), has attempted to reestablish the role of the money supply in explaining price movements. The most successful of these models follows the two-pillar Phillips curve approach proposed by Gerlach and Svensson (2003), where a standard Phillips curve framework is extended by including a measure of excess liquidity.[1] Movements of equilibrium velocity have proven to be a key obstacle to identifying excess liquidity. Most approaches that try to include nonlinear movements of velocity rely on explaining money demand either through the opportunity costs of holding money [see, e.g., Orphanides and Porter, 2000)] or by accounting for the role of wealth and asset prices [see, e.g., Dreger and Wolters, 2009]. In this paper, we follow an alternative P-star approach proposed by El-Shagi and Giesen (2013) and El-Shagi (2011), who consider any persistent movement to reflect a change in the equilibrium velocity. To do so, El-Shagi and Giesen replace the standard cointegration techniques commonly used to estimate P-star models with a state space framework, where equilibrium velocity and potential output are modeled as latent variables. Although this approach does not allow us to analyze the driving forces of velocity, it does ensure that equilibrium changes caused by institutional changes, technological progress, and other unobservable factors are captured. To allow the identification of monetary policy shocks, we augment the original framework by adding interest rates and enhancing the estimation framework to cope with monthly data, and the resulting higher lag order, in a computationally feasible way.

Whereas the P-star literature deals mostly with whether excess liquidity causes inflation, the structural VAR literature deals with the identification of monetary shocks and the transmission of those shocks throughout the economy.[2] In the last decade, sign restrictions have become a popular method of identifying structural shocks [see, e.g., Faust, 1998, Canova and De Nicoló, 2002, and Uhlig, 2005]. This technique also allows us to embed a more convincing economic explanation into structural econometric models than does the standard Cholesky decomposition. We apply this approach to the signal equations of the state space model of money, prices, production, and an interest rate. Although the state space model allows us to identify the long-run shocks (i.e., shocks changing the equilibrium path of the economy), the addition of sign restrictions permits a structural interpretation of the transitory shocks.

Finally, the monetary aggregation literature attributes the lack of informational content of monetary aggregates to the theoretical shortcomings of the simple sum monetary aggregates frequently used in empirical work [see, e.g., Barnett 1980, Barnett and Serletis 2000, and Barnett and Chauvet 2011]. A number of studies have shown that simple sum aggregates significantly overstate the money stock [Barnett et al. 2005; Barnett et al. 2008] and mask the dynamics of the relationship between money and interest rates [Kelly 2009; Kelly et al. 2011]. Therefore, in line

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

with recent approaches to incorporating money or liquidity effects into structural macro-models [see, e.g., Belongia and Ireland, 2012, p. 1], we employ a Divisia M4 monetary aggregate provided by the Center for Financial Stability [Barnett et al. (2013)] rather than a simple sum monetary aggregate. This solves a number of problems that could potentially obfuscate the macroeconomic impact of money supply, which in the recent literature has been referred to as the Barnett critique [see Chrystal and MacDonald (1994, p. 76)].

By combining the aforementioned literature, we are able to create a parsimonious model that allows a detailed look into the dynamics of a monetary policy shock by identifying persistent and nonpersistent money demand and supply shocks. This is crucial because previous research has found that there exists a time inconsistency problem in monetary policy, because monetary policy has short-run benefit, i.e., current monetary policy affects GDP in the short run, and long-run costs, i.e., money overhang affects prices in the long run [see, e.g., Nicolini, 1998 and Taylor, 1993].

The key contribution of this approach is (i) to show that the quantity theory of money is still relevant to price movements in the United States, even considering the Great Moderation as part of the sample that ranges from 1967M1 to 2006M12, and (ii) to provide common ground for the fully structural DSGE models and the purely data-driven VAR approaches. By explicitly modeling the latent variables driving the economy and thereby making it possible to correctly identify transitory disequilibria and their adjustment, we are able to capture the dynamics of the economy not unlike DSGE models. Actually, the estimation of DSGE models usually relies on reduced-form state space models similar to our own [see Ireland (2004)]. However, we impose fewer restrictions on the data, because we do not have to assume specific functional forms of the underlying concepts, such as utility and production functions. Thus, similarly to the VAR literature, we can truly allow the data to speak, providing impulse responses closer to what we observe in the real world.

The remainder of the paper is structured as follows. Section 2 outlines the elementary theoretical considerations that underlie our econometric specification as well as our basic empirical strategy. Section 3 describes the state space framework and the identification scheme for the remaining structural shocks. Section 4 describes the data used. Section 5 presents the results, and Section 6 concludes.

2. THEORETICAL CONSIDERATIONS AND EMPIRICAL STRATEGY

2.1. Empirical Strategy

We propose a two-step identification scheme for structural shocks that combines two established approaches to structural shock identification: (i) state space decomposition (using a Kalman filter) and (ii) structural VAR (SVAR).

Step 1. We identify the persistent developments in the data, i.e., potential output and equilibrium velocity, using a state space model. Our approach roughly

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

follows that of El-Shagi and Giesen (2013) and El-Shagi (2011), who estimate a P-Star type model [Hallman et al. (1991)] in its two-pillar Phillips curve version [Gerlach and Svensson (2003)] using a state space framework. We augment this model by including an interest rate and embedding it in an estimation procedure that allows the use of monthly data. Shocks to potential output and equilibrium velocity are modeled as latent variables and estimated using a Kalman filter. Because only supply shocks can plausibly affect output persistently, the shocks to potential output are interpreted as supply shocks. Assuming that the central bank is able to determine the quantity of money in the long run, by counteracting endogenous developments if necessary, shocks in the persistent component velocity can be read as changes in equilibrium money demand.

Step 2. We treat the residuals from the signal equations of the state space model as if they were residuals from a reduced-form (first difference) VAR. Because shocks to supply and persistent shocks to money demand have already been removed from the data, the remaining shocks reflect the consequences of structural shocks to aggregate demand, money supply, and costs, i.e., transitory supply-side shocks, as well as transitory shocks to money demand. The latter might be interpreted as financial market shocks, that change the need for liquidity in the short run, but do not change the behavior of market participants as institutional changes (such as changing regulation) or financial innovations do. We employ a sign restriction approach for the structural identification, as proposed by Canova and De Nicoló (2002) and Uhlig (2005), with the extensions proposed by Fry and Pagan (2007).

2.2. Developing the P-Star Model with Business-Cycle-Adjusted Monetary Overhang

To derive the excess liquidity, we start from a standard log-linearized quantity equation,

$$m + v = y + p,$$

where m, v, y, and p are the natural logarithms of money, velocity, production, and prices, respectively. Suppose that, to be in general equilibrium, $v = v^*$, where the star marks equilibrium values. Then an increase in the money supply constitutes the creation of excess liquidity by construction. Clearly, this increase has to be reflected by a corresponding decrease in velocity if both prices and production exhibit some degree of rigidity. More precisely, the negative deviation of velocity from equilibrium corresponds to excess liquidity. Accordingly, changes in the persistent component of velocity can be regarded as persistent changes in money demand. If v^* increases, only a corresponding increase in real money can reestablish equilibrium; i.e., either the central bank has to increase the money supply or the price level will adjust.[3]

Decomposing total velocity to obtain an estimate of excess liquidity may create an undesirable statistical artifact, as shown by El-Shagi (2011). To see this

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

statistical artifact, consider that in a recession, production, y, falls below potential output, y^*. Because neither money nor prices can plausibly compensate for this contemporaneously, the increased output gap is once again mirrored by a corresponding deviation of velocity from equilibrium. Strictly speaking, this constitutes excess liquidity, because less money is needed during the recession. However, the return of velocity to equilibrium takes place in the course of output returning to potential output, which may create a spurious correlation between "excess liquidity" and output growth. Thus, we define excess liquidity such that it does not rely on the decomposition of total velocity but on the decomposition of an adjusted velocity index, v^{a},

$$m - p - y^* = -v^{\mathrm{a}} = -(\tilde{v}^{\mathrm{a}} + v^{a*}), \tag{1}$$

where the tilde marks the cyclical component. For simplicity, we drop the index "a" in the components of velocity. The variables $\tilde{v}$ and v^* refer to the components of adjusted velocity in the remainder of the paper, unless noted otherwise.

3. IDENTIFICATION

3.1. The State Space Model

The model. Our model aims at explaining the joint dynamics of money, prices, output, and interest rates. Although we allow all four variables and the derived disequilibria to affect the others in a way similar to a standard VAR setup, the individual equations closely mirror widespread theoretical concepts, such as a monetary reaction function and a Philips curve. But although it is often possible to determine the implicit structural equation that most strongly drives the coefficient estimates of a single equation, the equations we use are reduced forms rather than structural equations.

This is most important when the money growth equation is considered. Although the central bank can control money growth in the long run, if necessary by counteracting changes in the behavior of banks, the short-run money growth rate is the outcome of the interaction of policy makers and market participants. However, because of the importance of central bank behavior, the variables that enter the typical monetary policy reaction function, i.e., interest rates, inflation, and real economic indicators, should play a major role.

We also account for the momentum of money growth. Although this momentum might reflect a smoothing preference of the central bank, it more likely results from the time lag of monetary policy actions affecting the monetary aggregate. Finally, we add excess liquidity, to allow the central bank to compensate for its past mistakes, and banks reducing their credit supply because of limited demand.

Production movements are expected to rely mostly on their own momentum, and more importantly, the attraction to potential output. Including monetary indicators allows the central bank to actually affect the business cycle. Prices are necessary to account for the joint dynamics created by supply and demand shocks.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

The inflation equation resembles a two-pillar Phillips curve. That is, inflation is driven by real indicators, i.e., output gap and production growth, and monetary indicators, i.e., excess liquidity, money growth, and the interest rate. We further account for the momentum of inflation dynamics by including an autoregressive process that has been established by the previous literature.

The interest rate equation roughly resembles an extended Taylor rule. Although there are many minor differences, such as including several real indicators, growth and output gap, and a larger variety of lags of the considered variables, the most important differences are that we allow interest rates to be affected by money growth and money overhang. Including money growth and money overhang in the interest rate equation accounts for the possibility that the central bank considers monetary indicators in policy making, but also allows for the possibility that money market conditions could affect the interest rate through financial markets. However, the latter interpretation seems unlikely, given that the federal funds rate rarely strays far from the policy target.

Thus, the model then takes the form

$$\begin{bmatrix} \Delta m \\ \Delta p \\ \Delta y \\ i \end{bmatrix}_t = A(m_t - p_t - y_t^* - v_t^*) + B(y_t - y_t^*) + \psi \Delta y_t^* + \phi \Delta v_t^*$$

$$+ C(L) \begin{bmatrix} \Delta m \\ \Delta p \\ \Delta y \\ i \end{bmatrix}_t + u_t, \tag{2}$$

$$\begin{bmatrix} y_t^* \\ v_t^* \end{bmatrix} = \begin{bmatrix} y_{t-1}^* \\ v_{t-1}^* \end{bmatrix} + \varepsilon_t, \tag{3}$$

where m, p, and y are the logged levels of money, prices, and production, i is the monthly average of the effective federal funds rate, v is adjusted velocity, A, B, $C(L)$, ψ, and ϕ are coefficient vectors or matrices—$C(L)$ being a lag polynomial, u and ε are the error term vectors of the signal and state equations, respectively, and t is the time index. The star denotes the equilibrium levels of velocity and output.

We use a fairly flexible approach, which follows the convention of the business cycle literature, where the trend components y^* and v^* are modeled as a random walk. This implies that the original time series follows a random walk with drift.[4] The drift term is omitted in this specification because the linear component of the trend has been removed. The term equilibrium is used in its traditional sense in this paper, i.e. a situation that the economy gravitates to and that could be sustained permanently, rather than in the New Keynesian sense of market clearing. Because persistent changes in the time series effectively reflect changes in the point of attraction of the economy, this is in line with the employed equilibrium concept.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

The latent variables are identified by a Kalman filter. As spelled out in the equations, we do only include the trend as latent variable rather than trend and cycle. Otherwise the shock to production would not be uniquely identified because there were two independent production predictions. Therefore to ensure economically meaningful results, we restrict the solution space to cases where the implicit cycles are stationary with a mean that is statistically indistinguishable from zero. Estimation is done using restricted maximum likelihood.

Reducing the number of parameters. State space models of this size quickly become unfeasible to estimate when the lag order increases, because of the huge number of parameters. Using monthly data, we need to use a lag order of at least 12 to capture the key dynamics that are ongoing. Because this would result in more than 200 parameters, we propose to restrict the matrix $C(L)$ to zero, where the correlation coefficient is unlikely to matter.

We obtain a first approximate estimate of the model by replacing the Kalman-filtered cyclical components of production and velocity by their HP-filtered counterparts. We then identify the least significant parameter in $C(L)$, restrict it to zero, and reestimate until all remaining parameters are significant at the 10% level, or only 50 entries of $C(L)$ remain unconstrained. The state space model is then estimated using these zero restrictions. Because the entries of $C(L)$ are particularly robust to minor misspecification of the cyclical components, we can reduce the number of parameters substantially without strongly affecting the model.[5] Although this might seem restrictive, our approach still is substantially less restrictive than fully structural DSGE models, where most lags are omitted entirely and thus implicitly restricted. Therefore, our approach is still able to capture the empirically founded dynamics of the system in greater detail.

3.2. Sign Restrictions

To identify our structural innovations, we adopt the sign restriction approach introduced by Canova and De Nicoló (2002), Uhlig (2005), and Fry and Pagan (2010). Fundamental to this method of shock identification is the requirement that the shocks should be uncorrelated. Recalling our model from equation (2), we obtain the structural shocks by treating the residuals from the signal equations of the state space model as if they were residuals from the following reduced-form VAR:

$$\begin{bmatrix} \Delta m \\ \Delta p \\ \Delta y \\ i \end{bmatrix}_t = A(m_t - p_t - y_t^* - v_t^*) + B(y_t - y_t^*) + \psi \Delta y_t^* + \phi \Delta v_t^*$$

$$+ C(L) \begin{bmatrix} \Delta m \\ \Delta p \\ \Delta y \\ i \end{bmatrix}_t + \Gamma^{\{-1\}} \xi_t,$$

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

such that $\hat{u}_t = \hat{\Gamma}^{-1}\hat{\xi}_t$, where the structural (i.e., orthogonal) residuals are denoted by $\hat{\xi}_t$. We define a normalized residual vector, $\eta_t = S^{-1}\hat{\xi}_t$, such that the diagonal elements S are the standard deviations of $\hat{\xi}$. Then we are able to write these residuals as

$$\hat{u}_t = T\eta, \tag{4}$$

where $T = \widehat{\Gamma}^{-1}S$.

To identify a set of estimated structural shocks, Fry and Pagan (2007) define a matrix Q such that $Q'Q = QQ' = I$, where I is an identity matrix, and

$$\hat{u}_t = TQ'Q\eta_t = T^*\eta_t^*. \tag{5}$$

Valid choices for Q include any Givens matrix or product of Givens matrices. Because there are an infinite number of Givens matrices, there is an infinite set of possible T^*, which we shall denote as $\mathfrak{T}$. We can generate all possible series of structural shocks that fulfill the required assumptions from a single estimate of Γ that can be obtained by a Cholesky decomposition of the estimated covariance matrix $\hat{u}'\hat{u}$. The covariance matrix of all η^* generated by these Givens rotations[6] of the original η is equal to the identity matrix, I, because $E(\eta_t^*\eta_t^{*'}) = QE(\eta_t\eta_t)Q' = I$.

We consider only those responses where the following sign restrictions hold:

$$\begin{bmatrix} + & + & + & + \\ * & + & + & - \\ + & - & + & - \\ - & + & + & + \end{bmatrix}. \tag{6}$$

We do not, however, prevent the response of any variable to any shock from being zero. Thus, we never impose that a particular response to a shock must happen, but rather allow for the possibility that they may happen. Because the columns of our sign restriction matrix do not allow linear dependent columns of TQ, the shocks are still orthogonal by construction.[7]

Monetary policy shock. An expansionary monetary policy shock, which is a negative shock to interest rates, has a positive contemporaneous effect on the money supply and production. The contemporaneous effect on price level is unconstrained. By remaining agnostic about the contemporaneous response of price level, we allow the possibility of a price puzzle. Restricting the contemporaneous response of real output to be positive allows the possibility that money is nonneutral in the short run, though our model does impose long-run money neutrality. Finally, restricting the contemporaneous response of money to be positive simply imposes a downward-sloping demand for money, eliminating a liquidity puzzle.

Cost shock. A cost shock has a positive contemporaneous effect on the money supply, price level, and interest rates and a negative contemporaneous effect on real output. Costs have a positive impact on prices and a negative impact on

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

real production, as in Leeper et al. (1996). To justify a positive contemporaneous effect on the money supply and the interest rate, consider that because the money supply process is at least partially endogenous, when price level increases, interest rates increase, inducing lenders to temporarily lend more. In other words, because neither lenders nor the central bank can react to the change in price level instantly, the new price level causes demand for money to shift to the right before money supply can change, causing a temporary increase in m and i. Even if the central bank reacts immediately to the unintended movement in prices and money and thereby adds to the rise in interest rates by moving policy in the same direction, it seems unlikely that the immediate response overcompensates for the original shock (and its implications) contemporaneously when monthly data are considered. Thus, although it is possible that the reaction of money might be mitigated, it is almost impossible to observe a contemporaneous reduction in money supply as a result of a cost shock.

Demand shock. An increase in aggregate demand causes an increase in both output and price level. Both higher price level and increased volume of trade cause demand for money change before money supply can respond, leading to a temporary increase in m and i. Thus, a demand shock has a positive contemporaneous effect on money, prices, production, and interest rates.

Money demand shock. An increase in money demand leads to an increase in interest rates, which in turn lowers aggregate demand, causing prices and output to fall. Thus, a money demand shock has a positive contemporaneous effect on money and interest rates and a negative contemporaneous effect on prices and production.

3.3. Computation of Impulse Response Functions

To compute our impulse responses and the corresponding confidence bounds, we combine a moving blocks bootstrap of the coefficients of the state space model with randomly generated Givens rotations. This allows us to account for the uncertainty from the estimation of the state space model and the uncertainty about T^* simultaneously.

We bootstrap 1,000 estimates for the coefficient matrices A and B and the corresponding covariance matrices $\hat{u}'\hat{u}$. We then draw from these coefficient estimates and randomly generate Givens rotations until we generate a matrix T^*, where the sign restrictions (6) hold. If this is not achieved within 200,000 draws, the corresponding coefficient matrices are discarded. This procedure is repeated until we obtain 2,000 sets $\{A, B, T^*\}$.

The confidence bounds for the impulse responses are obtained by a moving block bootstrap,with a block length of 7, which has been shown to be robust to autocorrelation [Li and Maddala (1996); Davidson and MacKinnon (2006)].[8]

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

To compute the point estimates of the impulse responses, we follow Fry and Pagan (2007). The impulse responses we report are selected from our bootstraps by minimizing the squared differences between the impulse response and the median of the confidence bounds. Fry and Pagan (2007) argue that this is a more appropriate approach than the commonly used pointwise median of the confidence bounds, because the median of the confidence bounds is not necessarily a possible response that could be generated by a single model.

4. DATA AND DESCRIPTIVE ANALYSIS

For our analysis, we use monthly U.S. data that include a monetary aggregate, price level, output, and interest rate from 1967M1 to 2006M12. All data series, excluding the interest rate, are seasonally adjusted. Variables used are as follows. The price aggregate, p, is the log of the Consumer Price Index (CPI), which we obtained from the Bureau of Labor Statistics. Output, y, is proxied by the log of industrial production (IP) and is obtained from the Board of Governors of the Federal Reserve System. The monetary aggregate, m, is the log of Divisia M4.[9] Divisia M4 is the Tornqvist–Theil discrete-time approximation of the the Divisia monetary aggregate first developed by Barnett (1978) and Barnett (1980), adjusted for retail sweeps [see Anderson and Rasche, 2000 and Dutkowsky and Cynamon, 2003 for further explanation of the issues of retail sweeps], which is published by the Center for Financial Stability.[10] See Barnett et al. (2013) for a detailed description of the calculation of the Divisia M4 aggregate. The interest rate, i, is the effective federal funds rate. Using the DF-GLS test we find (log) money, prices and, production to be stationary in first differences and interest rates to be stationary in levels.

We choose to use a Divisia monetary aggregate because the proper measure of the liquidity provided by the monetary portfolio is fundamental to our model. In a world with many monetary assets, which are clearly not perfect substitutes, we must use some weighting scheme to calculate the aggregate liquidity available for transactions, which affects prices. The main advantage of this approach is that it allows us to include a broad array of monetary assets with minimal risk of including assets that are too far removed from money, as less liquid assets have lower weights. Conventional wisdom suggests that using a sufficiently narrow monetary aggregate is necessary to avoid the many empirical puzzles plaguing monetary economists; however, Kelly et al. (2011) find that when money is measured using a reputable index number, a broad monetary aggregate exhibits empirical results more consistent with theoretical expectations than narrow aggregates.

At this point, an objection might be raised. Money in the traditional equation of exchange is a stock variable, but the Monetary Service Index measures the service flow provided by the monetary portfolio. However, Barnett (1991) showed that under the assumption of stationary expectations—i.e., the monetary service flow is expected to follow a Martingale process—the monetary stock is proportional to the monetary service flow. Therefore, given the complexity of correctly measuring

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

the money stock, which includes dealing with the equity premium puzzle and the fact that the assumptions needed to use the monetary service flow in this case are reasonably nonrestrictive, we chose to use Divisia M4.

This system of money, prices, production, and an interest rate has a long history in the analysis of monetary policy shocks. Leeper and Gordon (1992), for example, examine a very similar system. The sample period was chosen because calculation of the monetary aggregate begins in 1967M1, and we choose to exclude the financial crisis because its magnitude and position in the sample cause it to have an exaggerated influence on the parameter estimates. We use monthly data rather than quarterly, as are commonly used in this type of analysis, to simplify shock identification. By increasing the frequency of observations, we reduce the probability that reactions to the initial economic shock that reverse the original direction of the true contemporaneous effect occur within one observation.

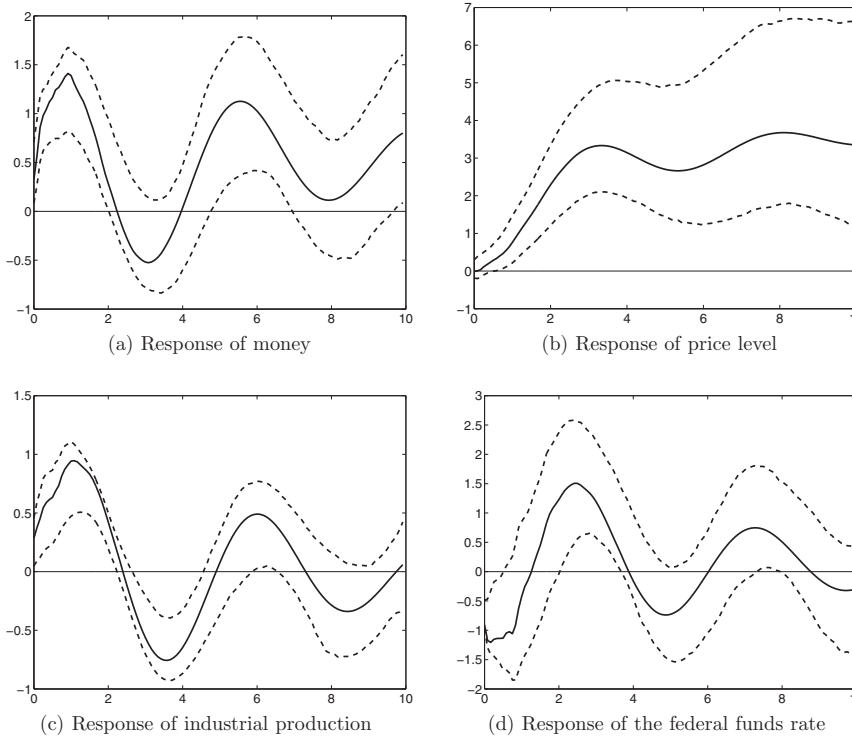
5. RESULTS

Figures 1–4 show the impulse response obtained from our analysis. The structural shocks are normalized to a standard deviation of one and each shock in the figures is a one-unit structural shock. The impulse responses are scaled to the standard deviation of the residuals of the corresponding equation. All figures cover impulse responses over a horizon of 10 years. Although this is too long for forecasting, it illustrates both short-run dynamics and long-run effects. The impulse responses are reported in levels; i.e., the impulse responses of money, prices, and production are aggregated over time. Although the first differences of money and prices return to their equilibrium in the long run, there can be persistent effects on levels. This is not true for production, because IP returns to the potential output.

5.1. Monetary Policy Shock

Figure 1 presents the impulse responses to a positive monetary policy shock. A positive monetary policy shock induces a strongly significant and persistent positive response in money and a strongly negative response in the federal funds rate that lasts about 20 months on average and becomes insignificant after about nine months. In the short run, we see a strongly positive response in IP, lasting for roughly two years. However, we see that monetary policy starts a strongly cyclical movement, with the first trough reaching magnitudes in both length and strength similar to those for the first peak. This cyclical movement is mirrored in all other macroeconomic variables except prices, which change rather smoothly because of the high persistence of inflation we find. There is no evidence for a price puzzle in the initial response of prices, which is very small and insignificant. After that prices begin growing quickly, leading to a significantly positive price response after about half a year. The momentum of inflation causes the price movement to surpass the monetary expansion in magnitude eventually, thereby causing excess liquidity to become negative.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press



(a) Response of money

(b) Response of price level

(c) Response of industrial production

(d) Response of the federal funds rate

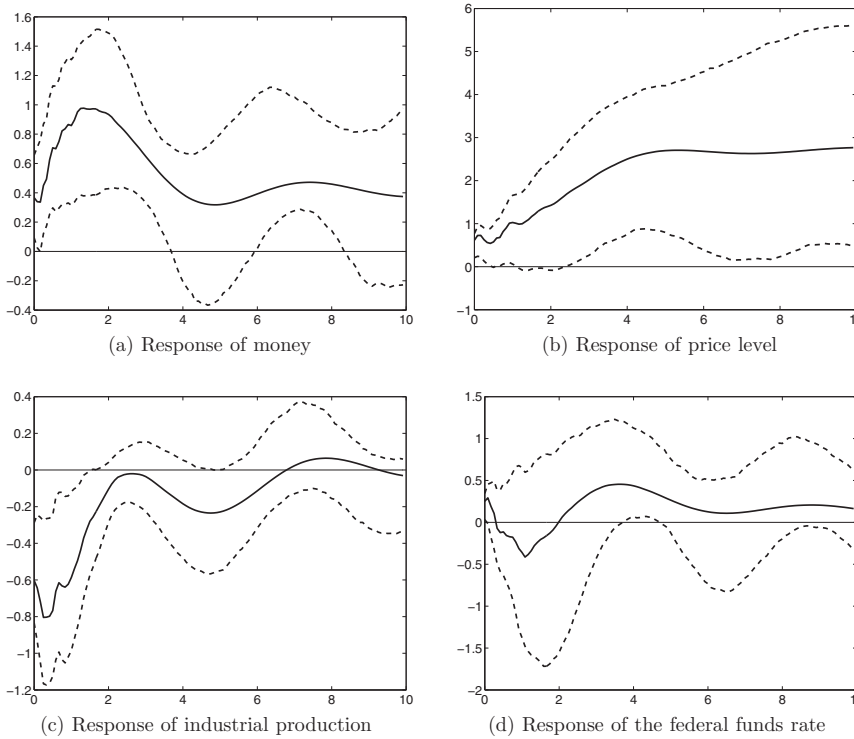
FIGURE 1. Impulse responses to a positive monetary policy shock.

Although money is subject to strong fluctuations after the initial shock, partly because of endogenous responses of money supply to the macroeconomic environment, the central bank ends up accommodating the higher price level, to an extent where the total response of prices is usually stronger than the initial expansionary monetary policy. That is, a monetary policy shock (1) causes strong business cycle fluctuation, (2) causes increased inflation, and (3) forces further monetary expansion to deal with the dynamics that is initiated.

5.2. Cost Shock

Initially, a positive cost shock induces a significant increase in price level and a decrease in real production, which is consistent with theoretical expectations; see Figure 2. In the first period, which is likely before the central bank can react, we see an increase in the federal funds rate because of increasing money demand. Once sufficent time has passed for the central bank to implement a policy response, there are very broad confidence bounds, indicating strong heterogeneity in the central bank reaction. Because both strongly increasing and decreasing interest rate responses are within the confidence bounds, we cannot determine whether

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press



(a) Response of money

(b) Response of price level

(c) Response of industrial production

(d) Response of the federal funds rate

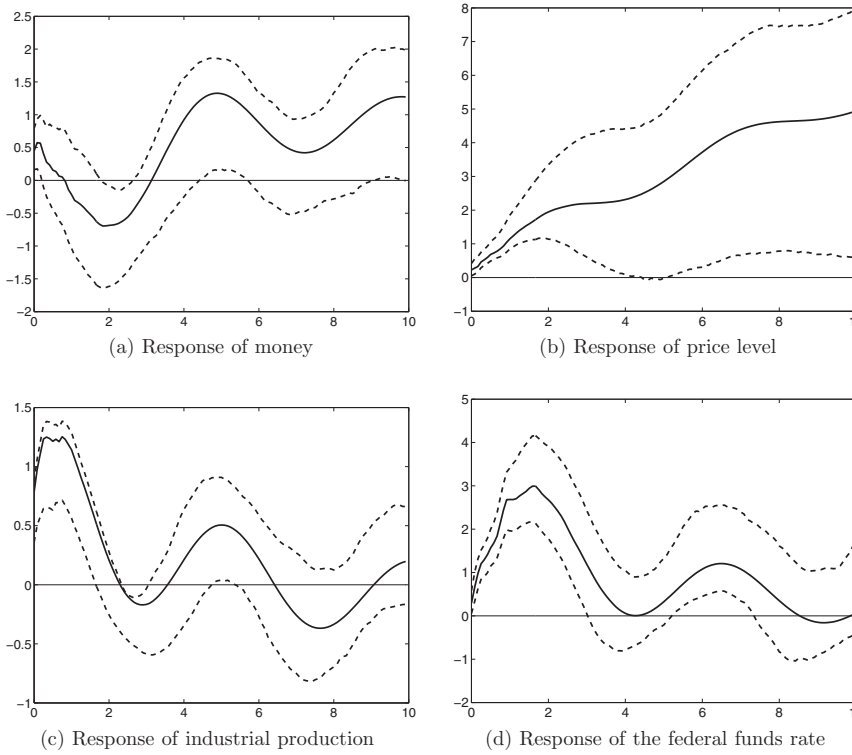
FIGURE 2. Impulse responses to a positive cost shock.

the central bank mistakes the cost shock for a reaction to its own policy and compensates, or recognizes the cost shock as exogenous and quickly accommodates. We see that the central bank accommodates the shock in the long run, at least partially, in the sense that the cost shock remains persistent in the price level and correspondingly in money supply. However, there is evidence that the central bank reacts by contracting the money supply about two years after the shock, thereby stopping the momentum of inflation. After about three years, interest rates rise slightly, but significantly, over their long-run average and inflation returns to its preshock levels. Although the negative response of IP is only significant for less than a year, IP seems to remain slightly depressed for much longer, on average. This may explain the delayed reaction of the central bank.

5.3. Demand Shock

A positive demand shock—see Figure 3—induces a significant positive response in IP, price level, and the federal funds rate. The interest rate continues to rise. Although the small contemporaneous reaction of the federal funds rate most likely is market-driven, the magnitude of the subsequent changes indicates a policy

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press



(a) Response of money

(b) Response of price level

(c) Response of industrial production

(d) Response of the federal funds rate

FIGURE 3. Impulse responses to a positive demand shock.

reaction. The monetary policy reaction to a demand shock causes money to drop quickly, counteracting the initial positive response. However, the response is too late and not sufficiently strong to stop the inflation precipitated by the demand shock. In the end, the central bank has to accommodate the higher prices at least partly, leading to permanently increased money supply and price level. This might be driven by the priority on stabilizing output. Although we do observe a cyclical movement of production after the demand shock, the oscillations dampen quickly.

5.4. Money Demand Shock

The economic response to a positive money demand shock is complex—see Figure 4. Because long-run equilibrium velocity remains unchanged, the change in money demand, as identified through the signal equations, is transitory by definition. One might think of this type of shock as a financial market phenomena, such as an emerging bubble causing higher asset demand that in turn drives liquidity demand. In response to this shock, we first observe an immediate increase in the federal funds rate and a decrease in IP, which is perfectly consistent with a leftward shift in the LM curve triggered by the money demand shock.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

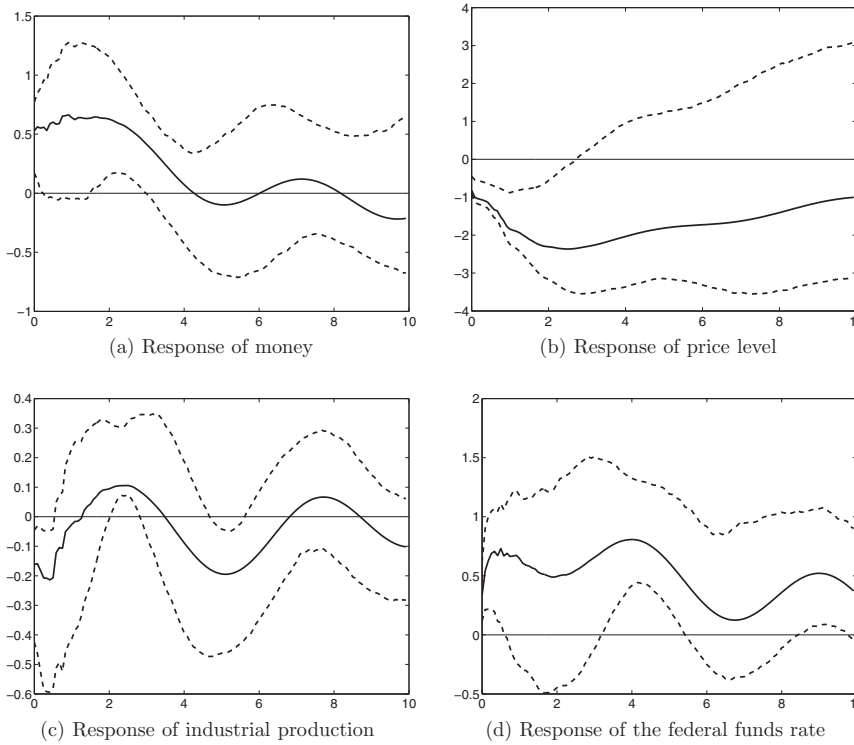


FIGURE 4. Impulse responses to a positive money demand shock.

There is a positive, though insignificant, response in money for the first three years after the initial money demand shock. This may be due either to the central bank attempting to accommodate the increase in money demand or to an increase in financial market participants' willingness to supply liquidity. Financial markets are willing to supply more liquidity because the real interest rate has increased. However, the increase in money supply is insufficient to compensate for the increased money demand, which explains why we observe both a significant positive response in the federal funds rate and a significant negative response in price level.

IP recovers from its initial decline within about 18 months, starting a period of damped oscillation around zero. To explain the mildness of the output response, consider that, because of the increased scarcity of money, we see an immediate decrease in price level that remains significant for about three years. This price level decrease creates a Pigou wealth effect that softens the impact of money demand shock [see Pigou, 1912 and Pigou, 1917].

Last, we observe that the federal funds rate remains above its preshock level over the entire period covered in the figures, though the rate does eventually return to its long-run average. To explain this persistence, consider that as long as money

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

demand is increased over its long-run average, the higher relative scarcity of money exerts downward pressure on prices. Thus, prices fall and, partly because of the momentum of inflation, keep falling or, more precisely, increase unusually slowly for some time. By the time liquidity has returned to the benchmark level, as it eventually does, prices are below benchmark. Assuming that the transitory change in money demand has expired, this implies a monetary overhang that creates inflationary pressure. Therefore, despite a contractory monetary policy, as indicated by the high interest rate, inflation is slightly above average after that.

6. CONCLUSION

In this paper, we have created a parsimonious model that allows a detailed look into the dynamics of a monetary policy shock. By applying shock identification to an econometric model that allows us to track equilibria such as potential output and money demand, we are able to combine some of the major advantages of DSGE models, which fully account for the driving forces of economic development, and VAR approaches, which allow the data to speak.

The most important of our model's results are as follows. First, contrary to most of the literature, we can show that the quantity of money matters in the United States (even when a sample that spans the Great Moderation is used). This holds even though interest rates are included in the model; that is, the impact of money does not represent a proxy for current monetary policy but indeed reflects a genuine quantity effect. Second, without imposing restrictions beyond the initial effect—as done in the previous literature on this issue—we find plausible short- to long-run impulse responses that are consistent with the New Keynesian short-run dynamics and the neoclassical long-run effects. Third, our model fits the data in a parsimonious specification without producing puzzles such as the price puzzle. In this framework, we find strong evidence for a time inconsistency problem of monetary policy. Inflation remains above the most likely policy target, i.e., the long-run average inflation rate, for about four years. This is due to both the creation of monetary overhang and the momentum of inflation itself. After that four-year period, the initially positive response of production has not only dissipated but reached a new trough. The resulting disequilibria enforce further policy action that creates a high cyclicality. Therefore, although monetary policy is able to dampen the business cycle fluctuations created by other shocks such as demand shocks, an "activist" monetary policy that aims to provide additional stimulus beyond what is needed to dampen business cycles is problematic.

NOTES

1. Although there have been difficulties in applying these models to the U.S. samples that include the Great Moderation, P-star models have been particularly successful in explaining inflation in Europe [see, e.g., Svensson, 2000, Tödter, 2002, and Kaufmann and Kugler, 2008].

2. This includes the seminal contributions of Gordon and Leeper (1994), Christiano et al. (1996), Leeper et al. (1996), Bernanke and Mihov (1998), and Christiano et al. (1999).

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

3. Because y is restricted to return to its potential y^* over time, real balances are determined by money demand in the long run. Given y^*, the long-run equilibrium of $m - p$ is exclusively driven by v^*. Therefore, persistent changes in real money demand are mirrored by corresponding changes in v^*. These changes are not necessarily due to changing preferences, but might be caused by institutional development, financial innovations, etc.

4. See, e.g., Gerlach and Smets (1999). The seminal Beveridge and Nelson (1981) decomposition essentially relies on similar assumptions on the trend component.

5. We tested the statistical significance of the difference between coefficients obtained from the constrained regression model using HP-filtered cycles and coefficients obtained from our state space model. Although the improved estimation of latent variables by the state space approach does lead to different estimates of corresponding coefficients, in both models we find no statistically significant parameters in $C(L)$, i.e., in the autoregressive behavior of the vector of endogenous variables, using a t-test.

6. By Givens rotation, we mean multiplication by one or more Givens matrices. For more details see Section 3.2 in Fry and Pagan (2007).

7. The only situation where this is not necessarily true is that where a range of several entries of TQ equal zero simultaneously. However, our sample algorithm does produce any TQ meeting this requirement.

8. Following the convention of the sign restriction literature [see, e.g., Uhlig (2005)], we report the 16th and 84th percentiles of the distribution as confidence bounds in the figures.

9. We also estimate the model using the log of the revised monetary service index at the M2 level of aggregation (MSI2) produced by the Federal Reserve Bank of St. Louis [see Anderson and Jones 2011]. Our results were robust to the choice of monetary aggregate.

10. Divisia M4 data can be obtained for the Center for Financial Stability Web site: www.centerforfinancialstability.org.

REFERENCES

Anderson, R.G. and B.E. Jones (2011) A comprehensive revision of the monetary services (Divisia) indexes for the United States. *Federal Reserve Bank of St. Louis Review* September/October, 325–359.

Anderson, R. and R. Rasche (2000) Retail sweep programs and bank reserves, 1994–1999. *Federal Reserve Bank of St. Louis Review* January/February, 51–72.

Barnett, W., J. Liu, R. Mattson, and J. van den Noort (2013) The new CFS Divisia monetary aggregates: Design, construction, and data sources. *Open Economies Review* 24(1), 101–124.

Barnett, W.A. (1978) The user cost of money. *Economics Letters* 1(2), 145–149.

Barnett, W.A. (1980) Economic monetary aggregates: An application of index number and aggregation theory. *Journal of Econometrics* 14(1), 11–48.

Barnett, W.A. (1991) A reply to Julio J. Rotemberg. In M. T. Belongia (ed.), *Monetary Policy on the 75th Anniversary of the Federal Reserve System*, pp. 189–222. Proceedings of the Fourteenth Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis. New York: Kluwer.

Barnett, W.A., U. Chae, and J.W. Keating (2005) The discounted economic stock of money with VAR forecasting. *Annals of Finance* 2(2), 229–258.

Barnett, W.A. and M. Chauvet (2011) How better monetary statistics could have signaled the financial crisis. *Journal of Econometrics* 161(1), 6–23.

Barnett, W.A., J.W. Keating, and L.J. Kelly (2008) Toward a bias corrected currency equivalent index *Economics Letters* 100(3), 448–451.

Barnett, W.A. and A. Serletis (2000) *The Theory of Monetary Aggregation. Contributions to Economic Analysis*. Amsterdam: Elsevier.

Belongia, M. and P. Ireland (2012) The Barnett Critique after Three Decades: A New Keynesian Analysis. NBER working paper 17885.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

Bernanke, B.S. and I. Mihov (1998) Measuring monetary policy. *Quarterly Journal of Economics* 113(3), 869–902.
Beveridge, S. and C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics* 7(2), 151–174.
Canova, F. and G. De Nicoló (2002) Monetary disturbances matter for business fluctuations in the G-7. *Journal of Monetary Economics* 49(6), 1131–1159.
Christiano, L., M. Eichenbaum, and C.L. Evans (1999) Monetary policy shocks: What have we learned and to what end? In *Handbook of Macroeconomics*, vol. 1. Amsterdam: Elsevier.
Christiano, L.J., M. Eichenbaum, and C. Evans (1996) The effects of monetary policy shocks: Evidence from the flow of funds. *Review of Economics and Statistics* 78(1), 16–34.
Chrystal, K. and R. MacDonald (1994) Empirical evidence on the recent behavior and usefulness of simple-sum and weighted measures of the money stock. *Federal Reserve Bank of Saint Louis Review* March/April, 73–73.
Davidson, R. and J.G. MacKinnon (2006) Bootstrap methods in econometrics. In K. Patterson and T. C. Mills (eds.), *Palgrave Handbooks of Econometrics*, vol. 1 of *Econometric Theory*, chap. 25. Houndmills, UK: Palgrave Macmillan.
Dreger, C. and J. Wolters (2009) Money velocity and asset prices in the euro area. *Empirica* 36(1), 51–63.
Dutkowsky, D. and B. Cynamon (2003) Sweep programs: the fall of M1 and rebirth of the medium of exchange. *Journal of Money, Credit and Banking* 35(2), 263–279.
Elliott, G., T.J. Rothenberg, and J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64(4), 813–836.
El-Shagi, M. (2011) Did the crisis affect potential output? *Applied Economics Letters* 18(8), 735–738.
El-Shagi, M. and S. Giesen (2013) Money and inflation: Consequences of the recent monetary policy. *Journal of Policy Modeling* 35(4), 520–537.
Faust, J. (1998) The robustness of identified VAR conclusions about money. *Carnegie–Rochester Conference Series on Public Policy* 49, 207–244.
Fry, R. and A. Pagan (2007) Some Issues in Using Sign Restrictions for Identifying Structural VARs. Working paper 14, NCER.
Fry, R. and A. Pagan (2010) Sign Restrictions in Structural Vector Autoregressions: A Critical Review. Working paper 22/2010, CAMA.
Gerlach, S. and F. Smets (1999) Output gap and monetary policy in the EMU area. *European Economic Review* 43(4–6), 801–812.
Gerlach, S. and L.E.O. Svensson (2003) Money and inflation in the euro area: A case for monetary indicators? *Journal of Monetary Economics* 50(8), 1649–1672.
Gordon, D.B. and E.M. Leeper (1994) The dynamic impacts of monetary policy: An exercise in tentative identification. *Journal of Political Economy* 102(6), 1228–1247.
Hallman, J.J., R.D. Porter, and D.H. Small (1991) Is the price level tied to the M2 monetary aggregate in the long run? *American Economic Review* 81(4), 841–858.
Ireland, P.N. (2004) A method for taking models to the data. *Journal of Economic Dynamics and Control* 28(6), 1205–1226.
Kaufmann, S. and P. Kugler (2008) Does money matter for inflation in the euro area? *Contemporary Economic Policy* 26(4), 590–606.
Kelly, L.J. (2009) *The Stock of Money and Why You Should Care*. Bingley, UK: Emerald Group Publishing.
Kelly, L.J., W.A. Barnett, and J.W. Keating (2011) Rethinking the liquidity puzzle: Application of a new measure of the economic money stock. *Journal of Banking and Finance* 35(4), 768–774.
Leeper, E. and D. Gordon (1992) In search of the liquidity effect. *Journal of Monetary Economics* 29(3), 341–369.
Leeper, E.M., C.A. Sims, and T. Zha (1996) What does monetary policy do? *Brookings Papers on Economic Activity* 1996(2), 1–78.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

Li, H. and G.S. Maddala (1996) Bootstrapping time series models. *Econometric Reviews* 15(2), 115–158.
McCandless, G.T. and W.E. Weber (1995) Some monetary facts. *Federal Reserve Bank of Minneapolis Quarterly Review* 19, 2–11.
Nicolini, J. (1998) More on the time consistency of monetary policy. *Journal of Monetary Economics* 41(2), 333–350.
Orphanides, A. and R. Porter (2000) P* revisited: Money-based inflation forecasts with a changing equilibrium velocity. *Journal of Economics and Business* 52(1–2), 87–100.
Pigou, A. (1912) *Wealth and Welfare*. London: Macmillan.
Pigou, A.C. (1917) The value of money. *Quarterly Journal of Economics* 32(1), 38–65.
Sims, C.A. (1980) Comparison of interwar and postwar business cycles: Monetarism reconsidered. *American Economic Review* 70(2), 250–257.
Svensson, L.E.O. (2000) Does the P* model provide any rationale for monetary targeting? *German Economic Review* 1(1), 69–81.
Taylor, J. (1993) Discretion versus policy rules in practice. In *Carnegie–Rochester Conference Series on Public Policy*, vol. 39, pp. 195–214. Amsterdam: Elsevier.
Teles, P. and H. Uhlig (2010) Is Quantity Theory Still Alive? NBER working paper 16393.
Tödter, K.H. (2002) Monetary Indicators and Policy Rules in the P-Star Model. Bundesbank discussion paper 18/02.
Uhlig, H. (2005) What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics* 52(2), 381–419.

APPENDIX: STATIONARITY

Figure A.1 presents an overview of the individual data series in logarithms and Figure A.2 shows the corresponding growth rates. All data series except interest rates are identified as upward-trending. Looking at the first differences of money, prices, and production, we see that the series is affected by some persistent behavior, especially for the CPI variable. The log differences of the money, price, and production series and the level of the interest rate series are tested for stationarity using the DF-GLS test (Table A.1), a more powerful version of the augmented Dickey–Fuller test proposed by Elliott et al. (1996). The lag length is selected using the Schwarz criterion with up to 17 lags. For all variables a lag length clearly below our maximum lags order (12) is chosen. The test indicates that all tested variables are stationary at the 5% significance level at least.

TABLE A.1. Dickey–Fuller GLS tests

Dickey–Fuller	CFSM4	CPI	IP	FedFunds
Test statistic	−5.71***	−2.02**	−2.70***	−2.35**
Lags	4	8	3	2

***indicates significance at the 1% level; **indicates significance at the 5% level.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

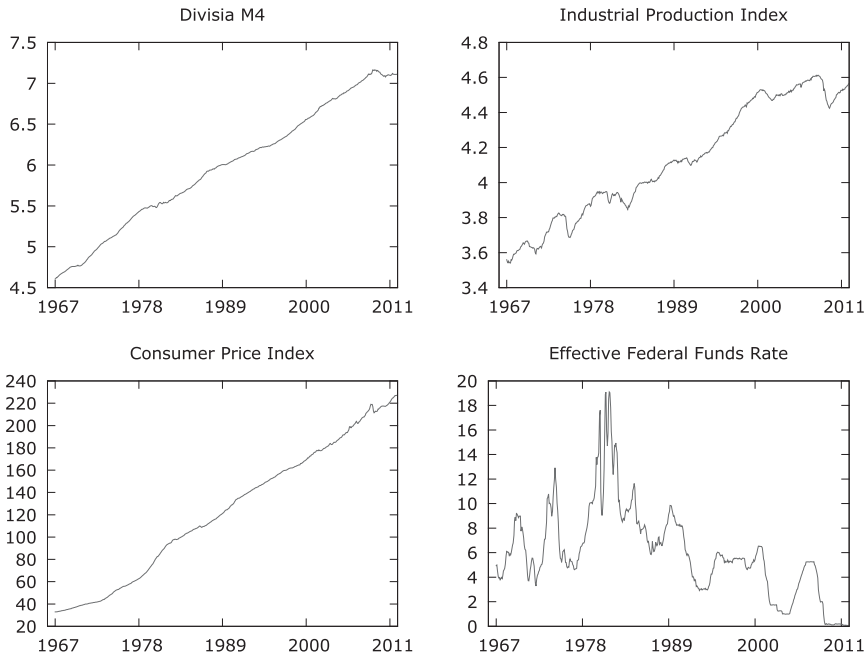


FIGURE A.1. Data series. For Divisia M4, industrial production, and CPI, log levels are plotted, whereas the level of the effective federal funds rate is plotted.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press

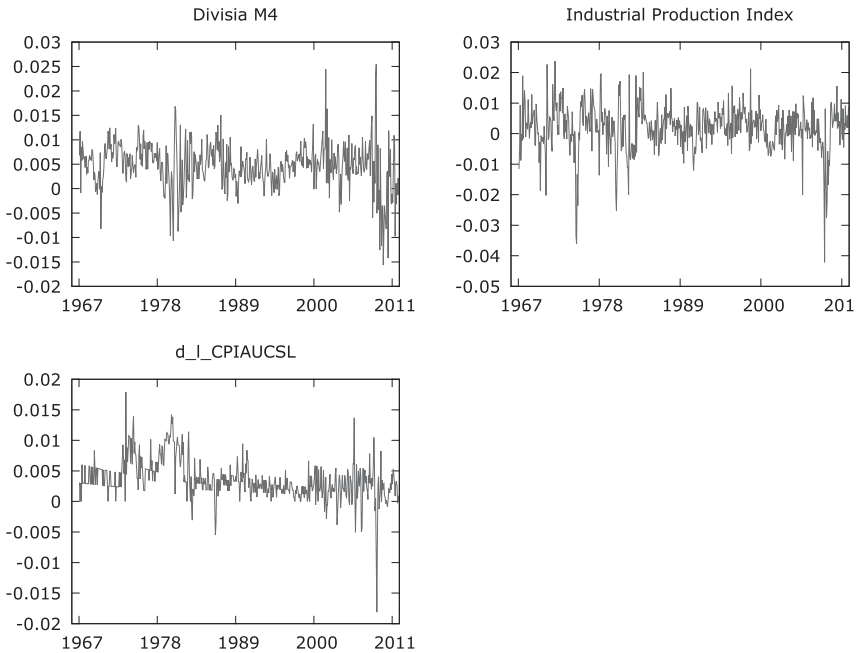


FIGURE A.2. Growth rates of Divisia M4, industrial production, and CPI.

https://doi.org/10.1017/S1365100513000217 Published online by Cambridge University Press