# DISCUSSION PAPER SERIES

No. 7813

## FORTUNE OR VIRTUE: TIME-VARIANT VOLATILITIES VERSUS PARAMETER DRIFTING IN U.S. DATA

Jesús Fernández-Villaverde,
Pablo A. Guerron-Quintana and
Juan Francisco Rubio-Ramirez

***INTERNATIONAL MACROECONOMICS***

Centre for Economic Policy Research

**www.cepr.org**

**Available online at:** **www.cepr.org/pubs/dps/DP7813.asp**

*ISSN 0265-8003*

# FORTUNE OR VIRTUE: TIME-VARIANT VOLATILITIES VERSUS PARAMETER DRIFTING IN U.S. DATA

**Jesús Fernández-Villaverde,** University of Pennsylvania and CEPR
**Pablo A. Guerron-Quintana,** Federal Reserve Bank of Philadelphia
**Juan Francisco Rubio-Ramirez,** Duke University, Federal Reserve Bank of Atlanta andFEDEA

Discussion Paper No. 7813
May 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 7813

May 2010

# ABSTRACT

## Fortune or Virtue: Time-Variant Volatilities Versus Parameter Drifting in U.S. Data*

This paper compares the role of stochastic volatility versus changes in monetary policy rules in accounting for the time-varying volatility of U.S. aggregate data. Of special interest to us is understanding the sources of the great moderation of business cycle fluctuations that the U.S. economy experienced between 1984 and 2007. To explore this issue, we build a medium-scale dynamic stochastic general equilibrium (DSGE) model with both stochastic volatility and parameter drifting in the Taylor rule and we estimate it non-linearly using U.S. data and Bayesian methods. Methodologically, we show how to confront such a rich model with the data by exploiting the structure of the high-order approximation to the decision rules that characterize the equilibrium of the economy. Our main empirical findings are: 1) even after controlling for stochastic volatility (and there is a fair amount of it), there is overwhelming evidence of changes in monetary policy during the analyzed period; 2) however, these changes in monetary policy mattered little for the great moderation; 3) most of the great performance of the U.S. economy during the 1990s was a result of good shocks; and 4) the response of monetary policy to inflation under Burns, Miller, and Greenspan was similar, while it was much higher under Volcker.

JEL Classification: C11, E10 and E30
Keywords: Bayesian methods, DSGE models, parameter drifting and stochastic volatility

Jesús Fernández-Villaverde
Department of Economics
University of Pennsylvania
160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104
USA
Email: jesusfv@econ.upenn.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=162452

Pablo A. Guerron-Quintana
Federal Reserve Bank of Philadelphia
Ten Independence Mall
Philadelphia, PA 19106
USA
Email: pablo.guerron@phl.frb.org

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=169712

Juan Francisco Rubio-Ramirez
Economics Department
Duke University
P.O. Box 90097
Durham, NC 27708
USA
Email: juan.rubio-ramirez@duke.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=156639

* We thank André Kurmann, Jim Nason, Serena Ng, Ricardo Reis, Frank Schorfheide, Tao Zha, and participants at several seminars for useful comments, and Béla Személy for invaluable research assistance. Beyond the usual disclaimer, we must note that any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System. Finally, we also thank the NSF for financial support.

Submitted 15 April 2010

# 1. Introduction

This paper addresses one of the main open questions in empirical macroeconomics: what is the role of time-varying variances versus changes in monetary policy rules in accounting for the evolving volatility of U.S. aggregate data? This discussion is particularly relevant for understanding the sources of the great moderation of business cycle fluctuations that the U.S. economy experienced between 1984 and 2007 and to forecast whether low volatility will return after the turbulence of 2008-2009.[1] To answer this question, we build a medium-scale dynamic stochastic general equilibrium (DSGE) model with both stochastic volatility in the structural shocks that drive the economy, parameter drifting in the Taylor rule followed by the monetary authority, and rational expectations of agents regarding these changes. Then, we estimate the model non-linearly using U.S. data and Bayesian methods, assess its fit, and compute its impulse response functions (IRFs). We use our results to run a battery of counterfactual exercises in which we build artificial histories of economies in which some source of variation has been eliminated or modified in an illustrative manner.

The motivation for this investigation is transparent. Time-varying volatility tells a history built around the changing size of the variance of structural shocks that hit the economy. The great moderation is, then, a tale of fortune: for two and a half decades we were favored by fate in the form of small variance of shocks. It is also a pessimistic perspective: we dwell in joy during periods of low volatility and we struggle through times of high volatility, but there is disappointingly little scope for the policy maker to battle the elements. Therefore, our current turbulence may be the opening stages of an era of large business cycle swings.

Parameter drifting constructs a radically divergent account of the cause of higher stability. It argues that some other changes in the economy, besides heteroscedastic disturbances, explain the evolution of aggregate volatility. Some versions of the parameter drifting narrative emphasize technological change. Two commonly cited factors are better inventory control (McConnell and Pérez-Quirós, 2000, Ramey and Vine, 2006) or financial innovation (Dynan, Elmendorf, and Sichel, 2006, or Guerrón-Quintana, 2009a). Other versions of the parameter drift history, the most prominent of which is Clarida, Galí, and Gertler (2000), single out monetary policy as the key to the reduced size of business cycle fluctuations. Thus, parameter drifting is a tale of virtue: thanks to either better technologies or changed policies, the

[1] Kim and Nelson (1998), McConnell and Pérez-Quirós (2000), and Blanchard and Simon (2001) were the first papers to point out that time-varying volatility was an important component of U.S. aggregate fluctuations. While Kim and Nelson and McConnell and Pérez-Quirós emphasized a big change in volatility around 1984, Blanchard and Simon saw the great moderation as part of a long-run trend toward lower volatility only momentarily interrupted during the 1970s. Stock and Watson (2002) undertake a thorough review of the evidence.

economy is more stable than before. It is also an optimistic view. As long as we do not abandon new technologies or unlearn the lessons of monetary economics, we should expect the great moderation to continue, current maladies notwithstanding.

There is evidence in favor of parameter drifting, particularly in monetary policy. Besides the classic work by Clarida, Galí, and Gertler (2000), basic references include Cogley and Sargent (2002), Lubick and Schorfheide (2004), Boivin and Giannoni (2006), and Canova (2009), among others. More recently, and most directly related to our investigation, Fernández-Villaverde and Rubio-Ramírez (2008) report compelling evidence of parameter drifting in the parameters that control the Taylor rule and in the degree of nominal rigidities in a standard DSGE model.

Another branch of the literature, which appeared largely as a response to Clarida, Galí, and Gertler (2000) and the other follow-up papers, has presented a strong case in favor of heteroscedastic structural shocks. Perhaps the most influential work in this tradition is Sims and Zha (2006). Relying on a structural vector autoregression (SVAR) with regime switching, Sims and Zha find that the model that best fits the data only has changes over time in the variances of structural disturbances and no variation in the monetary rule or in the private sector of the model. But even when they allow for policy regime changes, Sims and Zha find that the estimated changes cannot account for the evolution of observed volatility.[2] Using similar approaches, other papers corroborate this view. Among others, we can cite Cogley and Sargent (2005), Primiceri (2005), and Canova and Gambetti (2009). In general, once time-varying volatility is allowed, SVARs find little support for the tale of virtue; fortune seems to be the preferred option.

But using an SVAR approach presents challenges of its own. Benati and Surico (2009) use data generated from a simple New Keynesian DSGE model to show how regime-switching SVARs may misinterpret changes in policy as changes in variances because changes in policy also have implications for the volatility of endogenous variables (we can think about this argument as one instance of the Lucas critique). They read their results as suggesting that existing SVAR evidence may be uninformative for the question at hand.

To avoid these problems we follow a perspective more firmly grounded in explicit equilibrium models. First attempts along this direction are Fernández-Villaverde and Rubio-Ramírez (2007) and, in an important contribution, Justiniano and Primiceri (2008). These papers estimate DSGE economies that incorporate stochastic volatility on the structural shocks and show that such models fit the data considerably better than economies with homoscedastic structural shocks. However, neither of them allows for policy changes.

[2] Furthermore, Sims and Zha (2006) reject single-equation approaches because they require the use of instruments, which the authors argue rely on implausible restriction assumptions and fragile identification.

The natural next step is, thus, to estimate a DSGE model that can measure how much of the volatility change observed in the U.S. aggregate data can be attributed to either fortune, through heteroscedastic shocks, or virtue, in our case through changes in monetary policy. The project is challenging because, to get an econometrically satisfying answer, we need to simultaneously allow for both stochastic volatility and parameter drifting. A "one-at-a-time" approach is fraught with peril. If we only allow one source of variation in the model, the likelihood may want to take advantage of this extra degree of flexibility to fit the data better. For example, if the "true" model is one with parameter drifting in nominal rigidities, an estimated DSGE model with stochastic volatility may interpret this drift as time-varying volatility in mark-up shocks. If, instead, we had time-varying volatility in technological shocks in the data, an estimated model with only parameter drifting may conclude, erroneously, that the parameters of the Taylor rule are changing. Finally, it is important to have a model where agents have rational expectations over changes in monetary policy and incorporate the distributions over these changes into their decision rules (although, at a first pass and because of computational constraints, we will assume that the agents recognize right away when these changes occur, a restrictive hypothesis, since changes in policy are often difficult to detect even with hindsight).

Our contributions are both methodological and substantive. Methodologically, we show how to confront a rich non-linear DSGE model with stochastic volatility with the data by exploiting the structure of the second-order approximation to the decision rules that characterize the equilibrium of the economy. We prove a theorem, for a general class of DSGE models, that characterizes the structure of these decision rules. This theorem allows us to handily evaluate the likelihood function of the model. As an added bonus, this approach allows us to estimate the model without measurement errors in observables. One of the advantages of having stochastic volatility is that we multiply the number of random shocks in the model by two: for each exogenous stochastic process, we have a shock to level and a shock to volatility.

Our substantive findings are as follows. First, and challenging the SVAR evidence, we show that there is overwhelming evidence of changes in monetary policy during the analyzed period even after controlling for the fair amount of stochastic volatility present in the data. Second, we estimate that most of the reduction in aggregate volatility was caused by a reduction in the volatility of the innovation to the structural shocks in the economy, whereas the changes in monetary policy mattered less for the great moderation. Structural shocks were large and volatility under Burns, Miller, and Volcker and smaller and favorable under Greenspan. Third, our model suggests that the response of monetary policy to inflation during Volcker's tenure was stronger than under Burns, Miller, or Greenspan. According to our econometric estimates, had monetary policy been conducted during the 1970s as it was during the Volcker years, inflation would have been lower, in particular during the second oil shock. Interestingly, the clear change in monetary policy during Volcker's tenure is consistent with one of the policy regimes identified in Sims and Zha (2006). Fourth, in comparison, our estimated model indicates that Greenspan's response to inflation was mild (indeed, it was slightly lower than under Burns and Miller), but that he ruled over a period of extraordinarily positive shocks.

We establish these facts with a number of exercises. First, we compare the fit of different versions of the model with and without parameter drifting. Second, we inspect the smoothed series of the estimated parameter that controls the response of the monetary policy rule to inflation. The series shows a steep increase at the arrival of Volcker and a fast drop after the coming of Greenspan. In fact, the response of monetary policy to inflation is back to the levels of Burns-Miller times by the early 1990s. Third, we construct two sets of counterfactual histories. In the first set, we eliminate stochastic volatility by fixing variances at their historical means. In the second set, we feed alternative policy rules to different periods of time. In particular, we compute how the economy would have behaved during the tenure of one chairman if the monetary authority had followed the policy rule dictated by the average estimated response to inflation of some other chairman. In other words, we use our model to evaluate how the "average" monetary policy rule under Greenspan would have done during Volcker's times.

An alternative to our stochastic volatility framework would be to work with Markov regime-switching models (either in policy rules or in variances of shocks) with discrete jumps such as those of Bianchi (2009) or Farmer, Waggoner, and Zha (2009). This class of models provides an extra degree of flexibility in modelling aggregate dynamics that is highly promising. In fact, some of the fast changes in policy parameters that we document suggest that discrete jumps may be a good representation of the data. However, current technical limitations regarding the computation of the equilibria induced by regime switches force researchers to focus on small models that are only very stylized representations of an economy.

The rest of the paper is organized as follows. Section 2 presents the benchmark model that we use for our exercise, while section 3 defines its equilibrium and how we approximate it. Section 4, the core of the methodological contribution, explains how we evaluate the likelihood of the model. This is achieved by stating a theorem that characterizes the structure of the solution of second-order approximations to DSGE models with stochastic volatility. After describing the data and the estimation approach in section 5, the empirical results appear in sections 6 to 9. Section 10 constructs counterfactual histories. Section 11 concludes and three appendices provide further technical details.

# 2. A Benchmark Model

We adopt as the benchmark economy for our empirical investigation what has become the standard New Keynesian DSGE model in the literature (Woodford, 2003). The model is based on Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2007) and we have used it, without stochastic volatility, in Fernández-Villaverde and Rubio-Ramírez (2008) and in Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2009). The model has many strengths but also important weaknesses. Suffice it to say here that since this model has been the base of much applied policy analysis by central banks,[3] it is the natural laboratory for this paper.

Since the model is well known, our presentation will be brief.[4] A continuum of households consume, save, hold real money balances, supply labor, and set wages subject to a demand curve and nominal rigidities in the form of Calvo's pricing with partial indexation. The final good is produced by a representative firm that aggregates a continuum of intermediate goods produced by monopolistic competitive firms. These firms manufacture the intermediate good by renting the labor supplied and the capital accumulated by the households. Intermediate good producers set their prices subject to a demand curve and nominal rigidities in the form of Calvo's pricing with partial indexation. The model is closed with a monetary authority that fixes the one-period nominal interest rate according to a Taylor policy rule. In our specification, we introduce long-run growth through two unit roots: one in the neutral technology and one in the investment-specific technology. Stochastic volatility appears in the form of changing standard deviations of the five structural shocks to the model (two shocks to preferences, two shocks to technology, and one shock to monetary policy). Parameter drifting appears in the form of changing values of the parameters in the Taylor policy rule.

## 2.1. Households

The economy is populated by a continuum of households indexed by $j$. Household $j$'s preferences are representable by a lifetime utility function:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t d_t \left\{ \log\left(c_{jt} - hc_{jt-1}\right) + \upsilon \log\left(\frac{m_{jt}}{p_t}\right) - \varphi_t \psi \frac{l_{jt}^{1+\vartheta}}{1+\vartheta} \right\},$$

[3] Closely related models are used by the Federal Reserve Board (Edge, Kiley, and Laforte, 2007), the European Central Bank (Christoffel, Coenen, and Warne, 2008) and the Bank of Sweden (Adolfson *et al.*, 2007).

[4] The interested reader can find the web document, `www.econ.upenn.edu/~jesusfv/benchmark_DSGE.pdf`, in which we present the model without stochastic volatility or parameter drifting in careful detail.

which is separable in consumption, $c_{jt}$, real money balances, $m_{jt}/p_t$, and hours worked, $l_{jt}$. In our notation, $\mathbb{E}_0$ is the conditional expectation operator, $\beta$ is the discount factor, $h$ controls habit persistence, $\vartheta$ is the inverse of the Frisch labor supply elasticity, $d_t$ is a shifter to intertemporal preference that follows:

$$\log d_t = \rho_d \log d_{t-1} + \sigma_{dt}\varepsilon_{dt} \text{ where } \varepsilon_{dt} \sim \mathcal{N}(0,1),$$

and $\varphi_t$ is a labor supply shifter that evolves as:

$$\log \varphi_t = \rho_\varphi \log \varphi_{t-1} + \sigma_{\varphi t}\varepsilon_{\varphi t} \text{ where } \varepsilon_{\varphi t} \sim \mathcal{N}(0,1).$$

These two preference shocks are common to all households and provide flexibility for the equilibrium dynamics of the model to capture fluctuations in interest rates and changes in hours worked not accounted for by variations in consumption and wages.

The principal novelty of these preferences is that, for both shifters $d_t$ and $\varphi_t$, the standard deviations, $\sigma_{dt}$ and $\sigma_{\varphi t}$, of their innovations, $\varepsilon_{dt}$ and $\varepsilon_{\varphi t}$, are indexed by time; that is, they stochastically move period by period according to the autoregressive processes:

$$\log \sigma_{dt} = \left(1-\rho_{\sigma_d}\right)\log \sigma_d + \rho_{\sigma_d}\log \sigma_{dt-1} + \eta_d u_{dt} \text{ where } u_{dt} \sim \mathcal{N}(0,1)$$

and

$$\log \sigma_{\varphi t} = \left(1-\rho_{\sigma_\varphi}\right)\log \sigma_\varphi + \rho_{\sigma_\varphi}\log \sigma_{\varphi t-1} + \eta_\varphi u_{\varphi t} \text{ where } u_{\varphi t} \sim \mathcal{N}(0,1).$$

Our specification for the law of motion of the standard deviations of the innovations is parsimonious and it introduces only four new parameters, $\rho_{\sigma_d}$, $\rho_{\sigma_\varphi}$, $\eta_d$, and $\eta_\varphi$. At the same time, it is surprisingly powerful in capturing some important features of the data (Shephard, 2008). All the shocks and innovations (here and later in the paper) are perfectly observed by the agents when they are realized. Agents have, as well, rational expectations about how they evolve over time.

We can think about the shocks to preferences and to their stochastic volatility as reflecting the random evolution of more complicated phenomena. For example, stochastic volatility may appear as the consequence of changing demographic structures. An economy with older agents might be both less patient because of higher mortality risk (in our notation, a lower $d_t$) and less prone to reallocations in the labor force because of longer attachments to particular jobs (in our notation, a lower $\sigma_{\varphi t}$).

We assume complete financial markets: households can trade a whole set of securities contingent on idiosyncratic and aggregate events. An amount of those securities, $a_{jt+1}$, which pay one unit of consumption in event $\omega_{jt+1,t}$, is traded at time $t$ at unitary price $q_{jt+1,t}$ in terms of the consumption good. We drop the dependence on the event to ease the notational burden. In addition, households also hold $b_{jt}$ government bonds that pay a nominal gross interest rate of $R_{t-1}$. Therefore, the $j-th$ household's budget constraint is given by:

$$c_{jt} + x_{jt} + \frac{m_{jt}}{p_t} + \frac{b_{jt+1}}{p_t} + \int q_{jt+1,t} a_{jt+1} d\omega_{jt+1,t}$$
$$= w_{jt} l_{jt} + \left(r_t u_{jt} - \mu_t^{-1} \Phi\left[u_{jt}\right]\right) k_{jt-1} + \frac{m_{jt-1}}{p_t} + R_{t-1} \frac{b_{jt}}{p_t} + a_{jt} + T_t + \digamma_t$$

where $x_t$ is investment, $w_{jt}$ is the real wage, $r_t$ the real rental price of capital, $u_{jt} > 0$ the rate of use of capital, $\mu_t^{-1} \Phi\left[u_{jt}\right]$ is the cost of utilizing capital at rate $u_{jt}$ in terms of the final good, $\mu_t$ is an investment-specific technological level, $T_t$ is a lump-sum transfer, and $\digamma_t$ is the profits of the firms in the economy. We specify that

$$\Phi\left[u\right] = \Phi_1 \left(u - 1\right) + \frac{\Phi_2}{2} (u-1)^2$$

a form that satisfies the standard conditions that $\Phi\left[1\right] = 0$, $\Phi'\left[\cdot\right] = 0$, and $\Phi''\left[\cdot\right] > 0$. This function carries the normalization that $u = 1$ in the balanced growth path of the economy. Using the relevant first-order conditions, we can find $\Phi_1 = \Phi'\left[1\right] = \widetilde{r}$ where $\widetilde{r}$ is the (rescaled) balanced growth path rental price of capital (determined by all the other parameters in the model). This will leave us with only one free parameter, $\Phi_2$.

The capital accumulated by household $j$ at the end of period $t$ is given by:

$$k_{jt} = \left(1 - \delta\right) k_{jt-1} + \mu_t \left(1 - V\left[\frac{x_{jt}}{x_{jt-1}}\right]\right) x_{jt}$$

where $\delta$ is the depreciation rate and $V\left[\cdot\right]$ is a quadratic adjustment cost function:

$$V\left[\frac{x_t}{x_{t-1}}\right] = \frac{\kappa}{2} \left(\frac{x_t}{x_{t-1}} - \Lambda_x\right)^2$$

with adjustment parameter $\kappa$. This function is written in deviations with respect to the balanced growth rate of investment, $\Lambda_x$. Therefore, along the balanced growth path, $V\left[\Lambda_x\right] = V'\left[\Lambda_x\right] = 0$.

Our third structural shock, the investment-specific technology level $\mu_t$, follows a random walk in logs:

$$\log \mu_t = \Lambda_\mu + \log \mu_{t-1} + \sigma_{\mu t} \varepsilon_{\mu t} \text{ where } \varepsilon_{\mu t} \sim \mathcal{N}(0,1)$$

where $\Lambda_\mu$ is the drift of the process and $\varepsilon_{\mu t}$ is the innovation to its growth rate (see Greenwood, Herkowitz, and Krusell, 1997, for the classic motivation for this shock). In a similar way to the standard deviation of the innovations to the preference shocks, the standard deviation $\sigma_{\mu t}$ evolves as an autoregressive process:

$$\log \sigma_{\mu t} = \left(1 - \rho_{\sigma_\mu}\right) \log \sigma_\mu + \rho_{\sigma_\mu} \log \sigma_{\mu t-1} + \eta_\mu u_{\mu t} \text{ where } u_{\mu t} \sim \mathcal{N}(0,1).$$

Again, we can think about stochastic volatility as a stand-in for a more detailed explanation of technological progress in capital production that we do not model explicitly.

We can define two Lagrangian multipliers, $\lambda_{jt}$, the multiplier associated with the budget constraint, and $q_{jt}$ (the marginal Tobin's Q), the multiplier associated with the investment adjustment constraint normalized by $\lambda_{jt}$. Thus, the first order conditions of the household problem with respect to $c_{jt}$, $b_{jt}$, $u_{jt}$, $k_{jt}$, and $x_{jt}$ can be written as:

$$d_t \left(c_{jt} - hc_{jt-1}\right)^{-1} - b\beta \mathbb{E}_t d_{t+1} \left(c_{jt+1} - hc_{jt}\right)^{-1} = \lambda_{jt},$$

$$\lambda_{jt} = \beta \mathbb{E}_t \{\lambda_{jt+1} \frac{R_t}{\Pi_{t+1}}\},$$

$$r_t = \mu_t^{-1} \Phi' \left[u_{jt}\right],$$

$$q_{jt} = \beta \mathbb{E}_t \left\{ \frac{\lambda_{jt+1}}{\lambda_{jt}} \left( (1-\delta) q_{jt+1} + r_{t+1} u_{jt+1} - \mu_{t+1}^{-1} \Phi \left[u_{jt+1}\right]\right) \right\},$$

and

$$1 = q_{jt}\mu_t \left(1 - V\left[\frac{x_{jt}}{x_{jt-1}}\right] - V'\left[\frac{x_{jt}}{x_{jt-1}}\right]\frac{x_{jt}}{x_{jt-1}}\right) + \beta \mathbb{E} q_{jt+1}\mu_{t+1}\frac{\lambda_{jt+1}}{\lambda_{jt}} V'\left[\frac{x_{jt+1}}{x_{jt}}\right]\left(\frac{x_{jt+1}}{x_{jt}}\right)^2.$$

We need more work to find the optimality condition with respect to labor and wages because of the presence of monopolistic competition and nominal rigidities. Each household $j$ supplies a slightly different type of labor services $l_{jt}$ that are aggregated by a "labor packer" into homogeneous labor $l_t^d$ with the production function:

$$l_t^d = \left(\int_0^1 l_{jt}^{\frac{\eta-1}{\eta}} dj\right)^{\frac{\eta}{\eta-1}}$$

that is rented to intermediate good producers at the wage $w_t$. The "labor packer" is perfectly competitive and it takes all differentiated labor wages $w_{jt}$ and the wage $w_t$ as given.

The first-order conditions of the "labor packer" imply a demand function for labor:

$$l_{jt} = \left(\frac{w_{jt}}{w_t}\right)^{-\eta} l_t^d \qquad \forall j$$

and, together with a zero profit condition $w_t l_t^d = \int_0^1 w_{jt} l_{jt} dj$, an expression for the wage:

$$w_t = \left( \int_0^1 w_{jt}^{1-\eta} dj \right)^{\frac{1}{1-\eta}}.$$

Households follow a Calvo pricing mechanism when they set their wages. At the start of every period, a randomly selected fraction $1 - \theta_w$ of households can reoptimize their wages (where, by an appropriate law of large numbers, individual probabilities and aggregate fractions are equal). All other households index their wages given past inflation with an indexation parameter $\chi_w \in [0, 1]$. Therefore, the real wage of a household $j$ that has not changed wages for $\tau$ periods is:

$$\prod_{s=1}^{\tau} \frac{\Pi_{t+s-1}^{\chi_w}}{\Pi_{t+s}} w_{jt}.$$

Since we postulated above both complete financial markets for the households and separable utility in consumption, the marginal utilities of consumption are the same for all households. Thus, in equilibrium, $c_{jt} = c_t$, $u_{jt} = u_t$, $k_{jt-1} = k_t$, $x_{jt} = x_t$, $q_{jt} = q_t$, $\lambda_{jt} = \lambda_t$, and $w_{jt}^* = w_t^*$.

The last two equalities are the most relevant to simplify our analysis: they tell us that the shadow cost of consumption is equated across households and that all households that can reset their wages optimally will do it at the same level $w_t^*$. With these two results, and after several steps of algebra, we find that the evolution of wages is described by two recursive equations:

$$f_t = \frac{\eta - 1}{\eta} \left( w_t^* \right)^{1-\eta} \lambda_t w_t^{\eta} l_t^d + \beta \theta_w \mathbb{E}_t \left( \frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{1-\eta} \left( \frac{w_{t+1}^*}{w_t^*} \right)^{\eta-1} f_{t+1}$$

and

$$f_t = \psi d_t \varphi_t \left( \frac{w_t}{w_t^*} \right)^{\eta(1+\vartheta)} \left( l_t^d \right)^{1+\vartheta} + \beta \theta_w \mathbb{E}_t \left( \frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{-\eta(1+\vartheta)} \left( \frac{w_{t+1}^*}{w_t^*} \right)^{\eta(1+\vartheta)} f_{t+1}$$

on the auxiliary variable $f_t$.

Taking advantage that, in every period, a fraction $1 - \theta_w$ of households set $w_t^*$ as their wage and the remaining fraction $\theta_w$ partially index their price by past inflation, we can write the law of motion of real wage as:

$$w_t^{1-\eta} = \theta_w \left( \frac{\Pi_{t-1}^{\chi_w}}{\Pi_t} \right)^{1-\eta} w_{t-1}^{1-\eta} + (1 - \theta_w) w_t^{*1-\eta}.$$

## 2.2. The Final Good Producer

There is one final good producer that aggregates a continuum of intermediate goods according to the production function:

$$y_t^d = \left( \int_0^1 y_{it}^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{1}$$

where $\varepsilon$ is the elasticity of substitution.

The final good producer is perfectly competitive and minimizes its costs subject to the production function (1) and taking as given all intermediate goods prices $p_{ti}$ and the final good price $p_t$. The optimality conditions of this problem result in a demand function for each intermediate good with the classic form:

$$y_{it} = \left( \frac{p_{it}}{p_t} \right)^{-\varepsilon} y_t^d \qquad \forall i$$

where $y_t^d$ is the aggregate demand and a price for the final good:

$$p_t = \left( \int_0^1 p_{it}^{1-\varepsilon} di \right)^{\frac{1}{1-\varepsilon}}.$$

## 2.3. Intermediate Good Producers

Each of the intermediate goods is produced by a monopolistic competitor whose technology is given by a Cobb-Douglas production function with a fixed cost:

$$y_{it} = A_t k_{it-1}^{\alpha} \left( l_{it}^d \right)^{1-\alpha} - \phi z_t$$

where $k_{it-1}$ is the capital rented by the firm, $l_{it}^d$ is the amount of the "packed" labor input rented by the firm, the parameter $\phi$ corresponds to the fixed cost of production, and $A_t$ (our fourth structural shock) is neutral productivity that follows:

$$\log A_t = \Lambda_A + \log A_{t-1} + \sigma_{At} \varepsilon_{At} \text{ where } \varepsilon_{At} \sim \mathcal{N}(0,1).$$

In this specification, $\Lambda_A$ is the drift of the neutral technological level and $\varepsilon_{At}$ is the innovation to its growth rate.

The time-varying standard deviation of this innovation evolves stochastically following our, by now already familiar, specification:

$$\log \sigma_{At} = \left(1 - \rho_{\sigma_A}\right) \log \sigma_A + \rho_{\sigma_A} \log \sigma_{At-1} + \eta_A u_{At} \text{ where } u_{At} \sim \mathcal{N}(0,1).$$

The technology is translated by a fixed cost parameter $\phi$ and a scale variable $z_t = A_t^{\frac{1}{1-\alpha}} \mu_t^{\frac{\alpha}{1-\alpha}}$. Given our definitions of neutral productivity, $A_t$, and investment-specific productivity, $\mu_t$, we have that:

$$\log z_t = \Lambda_z + \log z_{t-1} + z_{zt}$$

where $\Lambda_z = \frac{\Lambda_A + \alpha \Lambda_\mu}{1-\alpha}$, $z_{zt} = \frac{z_{At} + \alpha z_{\mu t}}{1-\alpha}$, $z_{At} = \sigma_{At} \varepsilon_{At}$, and $z_{\mu t} = \sigma_{\mu t} \varepsilon_{\mu t}$. We can think about $z_t$ as the weighted level of technology in the economy, where the weight is given by the elasticity of output with respect to capital. The constant $\Lambda_z$ is the average growth rate of the economy. The role of $\phi$ is to make economic profits roughly equal to zero. We scale it by $z_t$ to keep the fixed costs constant in relative terms to the technology level. Finally, note that $z_{zt}$ will also have a stochastic volatility structure that is the product of the mixture of two processes with stochastic volatility themselves.

Intermediate good producers produce the quantity demanded of the good by renting $l_{it}^d$ and $k_{it-1}$ at prices $w_t$ and $r_t$. Then, by minimization, we have a marginal cost of:

$$mc_t = \left( \frac{1}{1-\alpha} \right)^{1-\alpha} \left( \frac{1}{\alpha} \right)^{\alpha} \frac{w_t^{1-\alpha} r_t^{\alpha}}{A_t}$$

The marginal cost is constant for all firms and all production levels given $A_t$, $w_t$, and $r_t$.

The quantity sold of the good is determined by the demand function derived above. Given this demand function, the intermediate good producers set prices to maximize profits. However, when they do so, they follow the same Calvo pricing scheme as households. In each period, a fraction $1 - \theta_p$ of intermediate good producers reoptimize their prices. All other firms partially index their prices by past inflation with an indexation parameter $\chi \in [0, 1]$.

Therefore, prices are set to solve the problem:

$$\max_{p_{it}} \mathbb{E}_t \sum_{\tau=0}^{\infty} (\beta \theta_p)^{\tau} \frac{\lambda_{t+\tau}}{\lambda_t} \left\{ \left( \prod_{s=1}^{\tau} \Pi_{t+s-1}^{\chi} \frac{p_{it}}{p_{t+\tau}} - mc_{t+\tau} \right) y_{it+\tau} \right\}$$

subject to

$$y_{it+\tau} = \left( \prod_{s=1}^{\tau} \Pi_{t+s-1}^{\chi} \frac{p_{it}}{p_{t+\tau}} \right)^{-\varepsilon} y_{t+\tau}^d.$$

In this problem, future profits are discounted using the pricing kernel of the economy, $\beta^{\tau} \lambda_{t+\tau} / \lambda_t$ (which is the right valuation criteria from the perspective of the households), and the probability of the event "only indexation for $\tau$ periods," $\theta_p^{\tau}$.

The solution for the firm's pricing problem has a recursive structure in two new auxiliary variables $g_t^1$ and $g_t^2$ that take the form:

$$g_t^1 = \lambda_t mc_t y_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}}\right)^{-\varepsilon} g_{t+1}^1$$
$$g_t^2 = \lambda_t \Pi_t^* y_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}}\right)^{1-\varepsilon} \left(\frac{\Pi_t^*}{\Pi_{t+1}^*}\right) g_{t+1}^2$$

and

$$\varepsilon g_t^1 = (\varepsilon - 1) g_t^2$$

where

$$\Pi_t^* = \frac{p_t^*}{p_t}$$

is the ratio between the optimal new price (common across all firms that can reset their prices) and the price of the final good. With this structure, the price index follows:

$$p_t^{1-\varepsilon} = \theta_p \left(\Pi_{t-1}^\chi\right)^{1-\varepsilon} p_{t-1}^{1-\varepsilon} + (1-\theta_p)\, p_t^{*1-\varepsilon}$$

or, normalizing by $p_t^{1-\varepsilon}$:

$$1 = \theta_p \left(\frac{\Pi_{t-1}^\chi}{\Pi_t}\right)^{1-\varepsilon} + (1-\theta_p)\, \Pi_t^{*1-\varepsilon}.$$

### 2.4. The Monetary Authority

The model is closed by the presence of a monetary authority that sets the nominal interest rates through open market operations financed with lump-sum transfers $T_t$ and a balanced budget. The monetary authority follows a modified Taylor rule:

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R}\right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi}\right)^{\gamma_{\Pi,t}} \left(\frac{\frac{y_t^d}{y_{t-1}^d}}{\exp\left(\Lambda_{y^d}\right)}\right)^{\gamma_{y,t}}\right)^{1-\gamma_R} \xi_t.$$

The first term on the right-hand side, $\frac{R_{t-1}}{R}$, represents a desire for interest rate smoothing, expressed in terms of $R$, the balanced growth path nominal return of capital. The second term, $\frac{\Pi_t}{\Pi}$, an "inflation gap," responds to the deviation of inflation from its balanced growth path level $\Pi$. The third term,

$$\frac{\frac{y_t^d}{y_{t-1}^d}}{\exp\left(\Lambda_{y^d}\right)}$$

is a "growth gap": the ratio between the growth rate of the economy and $\Lambda_{y^d}$, the balanced path gross growth rate of $y_t^d$. The term, $\log \xi_t = \sigma_{m,t}\varepsilon_{mt}$, is the monetary policy shock. The innovation $\varepsilon_{mt}$ to the monetary policy shock follows a $\mathcal{N}(0,1)$ process with a time-varying standard deviation, $\sigma_{m,t}$, that follows an autoregressive process:

$$\log \sigma_{mt} = \left(1 - \rho_{\sigma_m}\right) \log \sigma_m + \rho_{\sigma_m} \log \sigma_{mt-1} + \eta_m u_{m,t}.$$

In this policy rule, we have two drifting parameters: the responses of the monetary authority, $\gamma_{\Pi,t}$ and $\gamma_{y,t}$, to the inflation and growth gap. The parameters drift over time in an autoregressive fashion:

$$\log \gamma_{\Pi t} = \left(1 - \rho_{\gamma_\Pi}\right) \log \gamma_\Pi + \rho_{\gamma_\Pi} \log \gamma_{\Pi t-1} + \eta_\pi \varepsilon_{\pi t} \text{ where } \varepsilon_{\pi t} \sim \mathcal{N}(0,1)$$

and

$$\log \gamma_{yt} = \left(1 - \rho_{\gamma_y}\right) \log \gamma_y + \rho_{\gamma_y} \log \gamma_{yt-1} + \eta_y \varepsilon_{yt} \text{ where } \varepsilon_{yt} \sim N(0,1).$$

We assume here that the agents perfectly observe the changes in monetary policy parameters. A more plausible scenario would involve some filtering in real time by the agents who need to learn the stand of the monetary authority from observed decisions. A similar argument can be made for the values of the standard deviations of all the other shocks in the economy. But since this learning would further complicate what is already a large model, we leave this extension for future work.

### 2.5. Aggregation

Aggregate demand is given by:

$$y_t^d = c_t + x_t + \mu_t^{-1} \Phi \left[u_t\right] k_{t-1}.$$

By relying on the observation that the capital-labor ratio is constant across firms, we can derive that aggregate supply is:

$$y_t^s = \frac{A_t \left(u_t k_{t-1}\right)^\alpha \left(l_t^d\right)^{1-\alpha} - \phi z_t}{v_t^p}$$

where:

$$v_t^p = \int_0^1 \left(\frac{p_{it}}{p_t}\right)^{-\varepsilon} di$$

is the aggregate loss of efficiency induced by price dispersion of the intermediate goods.

Market clearing requires that

$$y_t = y_t^d = y_t^s.$$

By the properties of Calvo's pricing:

$$v_t^p = \theta_p \left( \frac{\Pi_{t-1}^{\chi}}{\Pi_t} \right)^{-\varepsilon} v_{t-1}^p + (1 - \theta_p) \Pi_t^{*-\varepsilon}.$$

Finally, demanded labor is given by:

$$l_t^d = \frac{1}{v_t^w} \int_0^1 l_{jt} dj$$

where:

$$v_t^w = \int_0^1 \left( \frac{w_{jt}}{w_t} \right)^{-\eta} dj.$$

is the aggregate loss of labor input induced by wage dispersion among differentiated types of labor. Again, by Calvo's pricing, this inefficiency evolves as:

$$v_t^w = \theta_w \left( \frac{w_{t-1}}{w_t} \frac{\Pi_{t-1}^{\chi_w}}{\Pi_t} \right)^{-\eta} v_{t-1}^w + (1 - \theta_w) \left( \Pi_t^{w*} \right)^{-\eta}.$$

## 3. Equilibrium

We can characterize an equilibrium in our economy by compiling all the first-order conditions of the household and firms, the Taylor rule of the monetary authority, and market clearing. This equilibrium is not stationary because we have two unit roots in the processes for technology. However, we circumvent this problem by rescaling the model $\widetilde{c}_t = \frac{c_t}{z_t}$, $\widetilde{\lambda}_t = \lambda_t z_t$, $\widetilde{r}_t = r_t \mu_t$, $\widetilde{q}_t = q_t \mu_t$, $\widetilde{x}_t = \frac{x_t}{z_t}$, $\widetilde{w}_t = \frac{w_t}{z_t}$, $\widetilde{w}_t^* = \frac{w_t^*}{z_t}$, $\widetilde{k}_t = \frac{k_t}{z_t \mu_t}$, and $\widetilde{y}_t^d = \frac{y_t^d}{z_t}$. The model is stationary in these transformed variables and, therefore, along the balanced growth path:

$$\Lambda_c = \Lambda_x = \Lambda_w = \Lambda_{w^*} = \Lambda_{y^d} = \Lambda_z.$$

With these rescaled variables, the states of the economy are stacked in a vector $S_t$ with 19 components:

$$S_t = \begin{pmatrix} \Lambda, \widehat{R}_{t-1}, \widehat{\widetilde{k}}_{t-1}, \widehat{v}_{t-1}^p, \widehat{v}_{t-1}^w, \widehat{\widetilde{w}}_{t-1}, \widehat{\widetilde{c}}_{t-1}, \widehat{\Pi}_{t-1}, \widehat{\widetilde{x}}_{t-1}, \widehat{\widetilde{y}}_{t-1}, \\ \widehat{d}_{t-1}, \widehat{\varphi}_{t-1}, \widehat{\gamma}_{\Pi t-1}, \widehat{\gamma}_{yt-1}, \widehat{\sigma}_{dt-1}, \widehat{\sigma}_{\varphi t-1}, \widehat{\sigma}_{\mu t-1}, \widehat{\sigma}_{At-1}, \widehat{\sigma}_{mt-1} \end{pmatrix}'$$

where we have expressed each variable $var_t$ in terms of log deviation with respect to the steady state, $\widehat{var}_t = \log var_t - \log var$, and $\Lambda$ is the perturbation parameter to be described below.

At this moment, it is also useful to clarify our wording with respect to the sources of randomness in the model. Thus, $(d_t, \varphi_t, A_t, \mu_t, \xi_t)$ is the vector of structural shocks (two to preferences, two to technology, and one to monetary policy), $(\varepsilon_{dt}, \varepsilon_{\varphi t}, \varepsilon_{At}, \varepsilon_{\mu t}, \varepsilon_{mt})$ is the vector of innovations to the structural shocks, and $(\varepsilon_{\pi t}, \varepsilon_{yt})$ is the vector of innovations to the parameter drifts (one to the response to inflation and one to the response to output). We stack these two vectors of innovations in the vector $W_{1t} = (\varepsilon_{dt}, \varepsilon_{\varphi t}, \varepsilon_{\mu t}, \varepsilon_{At}, \varepsilon_{mt}, \varepsilon_{\pi t}, \varepsilon_{yt})'$. Finally, $(\sigma_{dt}, \sigma_{\varphi t}, \sigma_{At}, \sigma_{\mu t}, \sigma_{mt})$ is the vector of volatility shocks and $W_{2t} = (u_{dt}, u_{\varphi t}, u_{\mu t}, u_{At}, u_{mt})'$ is the vector of innovations to the volatility shocks.

The equilibrium does not have a closed-form solution and we need to resort to a numerical approximation to compute it. This computation presents three challenges. The first challenge is that the dynamics of the state variables depend on $\mathbb{S}_t = (S_t', W_{1t}', W_{2t}')$, a vector that stacks states and innovations and has 31 components, an extremely demanding structure to keep track of. The second challenge is that since we have stochastic volatility, an inherently nonlinear structure, standard linearization techniques cannot be applied. More pointedly, if we linearized the model, stochastic volatility would disappear from the scene because the solution of the model would be certainty equivalent. The third challenge is that since we will need to compute the model for a large number of different parameter values in our estimation process, speed is of the utmost importance.

Perturbation methods provide a nice solution to the computation of the model that addresses these challenges. Beyond being extremely fast, perturbation offers high levels of accuracy even relatively far away from the perturbation point (Aruoba, Fernández-Villaverde, and Rubio-Ramírez, 2006). Therefore, we perform a second order perturbation around the deterministic steady state of the model and with respect to the perturbation parameter $\Lambda$ that we introduced earlier. The quadratic terms of this approximation allow us to capture, to a large extent, the effects of volatility shocks and parameter drift while keeping computational complexity at a reasonable level.

The solution of the model is then given by a transition equation for states:

$$S_{t+1} = \begin{pmatrix} \Psi_{s1}^1 \mathbb{S}_t' \\ \vdots \\ \Psi_{s19}^1 \mathbb{S}_t' \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \mathbb{S}_t \Psi_{s1}^2 \mathbb{S}_t' \\ \vdots \\ \mathbb{S}_t \Psi_{s19}^2 \mathbb{S}_t' \end{pmatrix} \tag{2}$$

and an observation equation:

$$\mathbb{Y}_t = \mathbb{C} + \begin{pmatrix} \Psi_{o1}^1 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \\ \vdots \\ \Psi_{o5}^1 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right) \Psi_{o1}^2 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \\ \vdots \\ \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right) \Psi_{o5}^2 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \end{pmatrix} \tag{3}$$

where $\mathbb{Y}_t$ is a vector of observables for the econometrician (five in our case) and $\mathbb{C}$ is a vector of means of these observables. The lagged vector $\mathbb{S}_{t-1}$ appears in the equation because, as we will see momentarily, our observables have components in first differences.

In these equations, $\Psi_{si}^1$ is a $1\times31$ vector and $\Psi_{si}^2$ is a $31\times31$ matrix for $i = 1, \ldots, 19$. The first term is the linear solution of the model while the second term is the quadratic component of the solution. Similarly, $\Psi_{oi}^1$ is a $1\times62$ vector and $\Psi_{oi}^2$ a $62\times62$ matrix for $i = 1, \ldots, 5$ and the interpretation of each term is the same as before: the first term is the linear component and the second one the quadratic component of the solution.

It is important to emphasize that we are *not* assuming the presence of any measurement error. Although we consider measurement errors both plausible and empirically relevant, in our exercise we want to measure how heteroscedastic structural shocks and parameter drifting help in accounting for the data. Consequently, we eliminate measurement errors to sharpen our analysis. This decision also helps to illustrate how DSGE models with stochastic volatility have a profusion of shocks that we can exploit in our estimation. We will elaborate on this last point below.

The transition equation (2) is unique (up to an equivalence class of representations) but the observable equation (3) is not, because it depends on what we assume the econometrician can observe. Guerrón-Quintana (2009b) discusses the consequences for inference of selecting different observables.

We pick as observables the first difference of the log of the relative price of investment, the log federal funds rate, log inflation, the first difference of log output, and the first difference of log real wages, or in our notation:

$$\mathbb{Y}_t = (-\triangle \log \mu_t, \log R_t, \log \Pi_t, \triangle \log y_t, \triangle \log w_t)'.$$

This implies that $\mathbb{C} = \left(-\Lambda_\mu, \log R, \log \Pi, \Lambda_{y^d}, \Lambda_w\right)'$. Later, when we take the model to the data, we will let the likelihood pick these means, since $\Lambda_\mu$, $\log R$, $\log \Pi$, $\Lambda_{y^d}$, and $\Lambda_w$ depend on the structural parameters. We select these variables because they bring us information about aggregate behavior (output), the stand of monetary policy (the interest rate and inflation), and the different shocks (the relative price of investment about investment-specific technological change, the other four variables about technology and preference shocks) that we are concerned about.

The state space representation generated by the transition equation (2) and the measurement equation (3) has an interesting structure that we exploit to evaluate the likelihood of the model. In the next section, we present a general description of that structure and how it applies to our economy.

# 4. Stochastic Volatility and Evaluation of the Likelihood

In this section we explain how to evaluate the likelihood function of our model. If we allow $\mathbb{Y}_t^{data}$ to be the data counterpart of our observables, $\mathbb{Y}^{data,t} = \left(\mathbb{Y}_1^{data}, \ldots, \mathbb{Y}_t^{data}\right)$ to be the history up to time $t$ of our observables, we can write the likelihood as:

$$p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma\right) = \prod_{t=1}^{T} p\left(\mathbb{Y}_t = \mathbb{Y}_t^{data} | \mathbb{Y}^{data,t-1}; \gamma\right),$$

where

$$p\left(\mathbb{Y}_t = \mathbb{Y}_t^{data} | \mathbb{Y}^{data,t-1}; \gamma\right)$$
$$= \int\int\int p\left(\mathbb{Y}_t = \mathbb{Y}_t^{data} | S_t, W_{1t}, \mathbb{S}_{t-1}; \gamma\right) p\left(S_t, W_{1t}, \mathbb{S}_{t-1} | \mathbb{Y}^{data,t-1}; \gamma\right) dS_t dW_{1t} d\mathbb{S}_{t-1} \tag{4}$$

and $\mathbb{Y}^{data,0} = \{\varnothing\}$.

Computing this likelihood is a difficult problem. It cannot be evaluated exactly and deterministic integration problems are too slow for practical use (we have three integrals period per period over large dimensions). Instead, we use a sequential Monte Carlo method to obtain a numerical estimate of (4).[5] As shown in Fernández-Villaverde and Rubio-Ramírez (2007), conditional on having $N$ draws of $\left\{s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i\right\}_{i=1}^N$ from the densities $p\left(S_t, W_{1t}, \mathbb{S}_{t-1} | \mathbb{Y}^{data,t-1}; \gamma\right)$ (we will explain later how we generate them), a law of large numbers implies that the integral (4) can be approximated by:

$$p\left(\mathbb{Y}_t = \mathbb{Y}_t^{data} | \mathbb{Y}^{data,t-1}; \gamma\right) \simeq \frac{1}{N} \sum_{i=1}^{N} p\left(\mathbb{Y}_t = \mathbb{Y}_t^{data} | s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i; \gamma\right) \tag{5}$$

Hence, we need to evaluate:

$$p\left(\mathbb{Y}_t = \mathbb{Y}_t^{data} | s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i; \gamma\right) \tag{6}$$

for each draw. This evaluation step is crucial not only because it is a term in (5), but also because, in the sequential Monte Carlo that we will implement, we need (6) to resample from the draws from $p\left(S_t, W_{1t}, \mathbb{S}_{t-1} | \mathbb{Y}^{data,t-1}; \gamma\right)$ and, in that way, get draws from $p\left(S_{t+1}, W_{1t+1}, \mathbb{S}_t | \mathbb{Y}^{data,t}; \gamma\right)$.

[5] This is not the only possible algorithm to do so, although it is a procedure that we have found useful in previous work. Alternatives include DeJong *et al.* (2007), Kim, Shephard, and Chib (1998), Fiorentini, Sentana, and Shepard (2004), and Fermanian and Salanié (2004).

The measurement equation (3) implies that evaluating (6) involves solving the equation:

$$\mathbb{Y}_t^{data} = \mathbb{C} + \begin{pmatrix} \Psi_{o1}^1 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \\ \vdots \\ \Psi_{o5}^1 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right) \Psi_{o1}^2 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \\ \vdots \\ \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right) \Psi_{o5}^2 \left(\mathbb{S}_t, \mathbb{S}_{t-1}\right)' \end{pmatrix} \tag{7}$$

for $W_{2,t}$ given $\mathbb{Y}_t^{data}$, $s_t^i$, $w_{1t}^i$, and $\boldsymbol{s}_{t-1}^i$. Since (7) is quadratic, we will have $2^5$ different solutions to this equation. We are not aware of any accurate and efficient way to find these $2^5$ different solutions. This problem would seem to prevent us from achieving our goal of evaluating the likelihood function of this model.

But considering stochastic volatility allows us to convert the above-described quadratic problem into a linear and simpler one. In particular, we illustrate how, when stochastic volatility is present in the problem, equation (7) has only one solution. Moreover, that solution can be found by simply inverting a matrix. Thanks to this insight, the evaluation of the likelihood function becomes possible.[6]

The key to our approach is to note that, when stochastic volatility is considered, the optimal policies functions of many economies share a particular pattern that we can exploit. To make this point more generally, we switch in the next few paragraphs to a more abstract notation.

The set of equilibrium conditions of a wide variety of DSGE models (including the one described in the paper) can be written as:

$$\mathbb{E}_t f\left(\mathcal{Y}_{t+1}, \mathcal{Y}_t, \mathcal{S}_{t+1}, \mathcal{S}_t, \mathcal{Z}_{t+1}, \mathcal{Z}_t\right) = 0 \tag{8}$$

where $\mathbb{E}_t$ is the expectation operator conditional on information available at time $t$, $\mathcal{Y}_t = \left(\mathcal{Y}_{1t}, \mathcal{Y}_{2t}, \ldots, \mathcal{Y}_{kt}\right)$ is the vector of non-predetermined variables of size $k$, $\mathcal{S}_t = \left(\mathcal{S}_{1t}, \mathcal{S}_{2t}, \ldots, \mathcal{S}_{nt}\right)$ is the vector of endogenous predetermined variables of size $n$, $\mathcal{Z}_t = \left(\mathcal{Z}_{1t}, \mathcal{Z}_{2t}, \ldots, \mathcal{Z}_{mt}\right)$ is the vector of exogenous predetermined variables of size $m$ (which we call structural shocks), and $f$ maps $\mathbb{R}^{2\times k+2\times n+2\times m}$ into $\mathbb{R}^{k+n+m}$.

We want to consider the case where the structural shocks follow a stochastic volatility

[6] Stochastic volatility may also help to circumvent a problem of some DSGE models: stochastic singularity. In general, we need at least as many shocks as observables for the likelihood function to be well defined. This requirement forces researchers to add extra shocks or measurement errors in situations where they might not desire to do so. Stochastic volatility, by introducing an additional volatility shock for each structural shock, doubles the number of shocks in the model. Even if in our model this is not necessary, on some other occasions, the researcher might want to take advantage of this extra flexibility and either augment the number of observables or reduce the number of shocks.

process of the form:

$$\mathcal{Z}_{it+1} = \rho_i \mathcal{Z}_{it} + \Lambda \sigma_{it+1} \varepsilon_{it+1}$$

where the standard deviation of the innovations evolves as:

$$\log \sigma_{it+1} = \vartheta_i \log \sigma_{it} + \Lambda \eta_i u_{it+1}$$

for all $i = \{1, \ldots, m\}$.

To ease notation, we are assuming that all structural shocks face volatility shocks. It is straightforward yet cumbersome to generalize the notation to other cases (in fact, in our model, some of the shocks, the ones in the parameter drifting, do not follow a stochastic volatility process).

The solution to the model given in equation (8) can be summarized by the following two equations, one describing the evolution of predetermined variables:

$$\mathcal{S}_{t+1} = h\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda\right) \tag{9}$$

and one describing the evolution of non-predetermined ones:

$$\mathcal{Y}_t = g\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda\right), \tag{10}$$

where $\Sigma_t = (\log \sigma_{1t}, \log \sigma_{2t}, \ldots, \log \sigma_{mt})$, $\mathcal{E}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{mt})$, and $\mathcal{U}_t = (u_{1t}, u_{2t}, \ldots, u_{mt})$ (this assumes that the volatility shocks are uncorrelated, a restriction that could be relaxed by the appropriate extension of the state space). To clarify notation, we think of $\Sigma_t$ as the volatility shocks, $\mathcal{E}_t$ are the innovations to the structural shocks, and $\mathcal{U}_t$ are innovations to volatility shocks.

We wish to find a second-order approximation of the functions $h(\cdot) : \mathbb{R}^{n+(4\times m)+1} \to \mathbb{R}^n$ and $g(\cdot) : \mathbb{R}^{n+(4\times m)+1} \to \mathbb{R}^k$ around the steady state, $\mathcal{S}_t = \mathcal{S}$ and $\Lambda = 0$. Therefore, we need to characterize the first- and second-order derivatives of the functions $h(\cdot)$ and $g(\cdot)$ evaluated at the steady state. The following theorem shows that the first partial derivatives of $h(\cdot)$ and $g(\cdot)$ with respect to any component of $\mathcal{U}_t$ and $\Sigma_{t-1}$ evaluated at the steady state is zero; that is, volatility shocks and their innovations do not affect the linear component of the optimal decision rule of the agents for any $i = \{1, \ldots, m\}$ (the same occurs with the perturbation parameter $\Lambda$). A similar result has been already established by Schmitt-Grohé and Uribe (2004) for the homoscedastic shocks case. More important, the theorem also shows that the second partial derivative of $h(\cdot)$ and $g(\cdot)$ with respect to $u_{i,t}$ and any other variable but $\varepsilon_{i,t}$ is also zero for any $i = \{1, \ldots, m\}$.

**Theorem 1.** *Let us denote* $[\Upsilon_\omega]_j^i$ *as the derivative of the* $i-th$ *element of generic function* $\Upsilon$ *with respect to the* $j-th$ *element generic variable* $\omega$ *evaluated at the non-stochastic steady state (where we drop this index if* $\omega$ *is unidimensional). Then, for the dynamic equilibrium model specified in equation (8), we have that:*

$$\left[h_{\Sigma_{t-1}}\right]_j^{i_1} = \left[g_{\Sigma_{t-1}}\right]_j^{i_2} = \left[h_{\mathcal{U}_t}\right]_j^{i_1} = \left[g_{\mathcal{U}_t}\right]_j^{i_2} = \left[h_\Lambda\right]^{i_1} = \left[g_\Lambda\right]^{i_2} = 0$$

*for* $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, *and* $j \in \{1, \ldots, m\}$.

*Furthermore, if we denote* $[\Upsilon_{\omega\xi}]_{j_1,j_2}^i$ *as the derivative of the* $i-th$ *element of generic function* $\Upsilon$ *with respect to the* $j_1-th$ *element generic variable* $\omega$ *and the* $j_2-th$ *element generic variable* $\xi$ *evaluated at the non-stochastic steady state (where again we drop the index for unidimensional variables), we have that:*

$$\left[h_{\Lambda,\mathcal{S}_t}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{S}_t}\right]_j^{i_2} = 0$$

*for* $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, *and* $j \in \{1, \ldots, n\}$ ,

$$\left[h_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2} = \left[h_{\Lambda,\Sigma_{t-1}}\right]_j^{i_1} = \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} = \left[h_{\Lambda,\mathcal{E}_t}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{E}_t}\right]_j^{i_2} = \left[h_{\Lambda,\mathcal{U}_t}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{U}_t}\right]_j^{i_2} = 0$$

*for* $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, *and* $j \in \{1, \ldots, m\}$,

$$\left[h_{\mathcal{S}_t,\Sigma_{t-1}}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]_{j_1,j_2}^{i_2} = \left[h_{\mathcal{S}_t,\mathcal{U}_t}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{S}_t,\mathcal{U}_t}\right]_{j_1,j_2}^{i_2} = 0$$

*for* $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, $j_1 \in \{1, \ldots, n\}$, *and* $j_2 \in \{1, \ldots, m\}$ ,

$$\left[h_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]_{j_1,j_2}^{i_2} = \left[h_{\Sigma_{t-1},\Sigma_{t-1}}\right]_{j_1,j_2}^{i_1} = \left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]_{j_1,j_2}^{i_2} = 0$$

*and*

$$\left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]_{j_1,j_2}^{i_2} = \left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]_{j_1,j_2}^{i_1} = \left[g_{\Sigma_{t-1},\mathcal{U}_t}\right]_{j_1,j_2}^{i_2} = \left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{U}_t,\mathcal{U}_t}\right]_{j_1,j_2}^{i_2} = 0$$

*for* $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, *and* $j_1, j_2 \in \{1, \ldots, m\}$, *and*

$$\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]_{j_1,j_2}^{i_2} = \left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]_{j_1,j_2}^{i_1} = \left[g_{\mathcal{E}_t,\mathcal{U}_t}\right]_{j_1,j_2}^{i_2} = 0$$

*for* $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, *and* $j_1, j_2 \in \{1, \ldots, m\}$ *if* $j_1 \neq j_2$.

**Proof.** See Appendix. ■

Since the statement of the theorem is long and involved, we clarify it with a table in which we characterize the second derivatives of $h(\cdot)$ and $g(\cdot)$ with respect to the different variables $(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda)$. The way to read the table is as follows. Take an arbitrary entry, for instance entry (1,2), $\mathcal{S}_t\mathcal{Z}_{t-1} \neq 0$. In this entry, we state that the cross-derivatives of $h(\cdot)$ and $g(\cdot)$ with respect to $\mathcal{S}_t$ and $\mathcal{Z}_{t-1}$ are different from zero. Similarly, entry (3,3), $\Sigma_{t-1}\mathcal{U}_t = 0$, tells us that the cross-derivatives of $h(\cdot)$ and $g(\cdot)$ with respect to $\Sigma_{t-1}$ and $\mathcal{U}_t$ are all zero. Entries (3,2) and (4,2) have a "*" to denote that the only cross-derivatives of those entries that are different from zero are those that correspond to the same index $j$ (that is, the cross derivatives of each innovation to the structural shocks with respect to its own volatility shock and the cross derivatives of the innovation to the structural shocks to the innovation to its own volatility shock). The lower triangular part of the table is empty because of the symmetry of second derivatives.

Table 4.1: Second Derivatives

| | | | | | |
|---|---|---|---|---|---|
| $\mathcal{S}_t\mathcal{S}_t \neq 0$ | $\mathcal{S}_t\mathcal{Z}_{t-1} \neq 0$ | $\mathcal{S}_t\Sigma_{t-1} = 0$ | $\mathcal{S}_t\mathcal{E}_t \neq 0$ | $\mathcal{S}_t\mathcal{U}_t = 0$ | $\mathcal{S}_t\Lambda = 0$ |
| $\mathcal{Z}_{t-1}\mathcal{Z}_{t-1} \neq 0$ | $\mathcal{Z}_{t-1}\Sigma_{t-1} = 0$ | $\mathcal{Z}_{t-1}\mathcal{E}_t \neq 0$ | $\mathcal{Z}_{t-1}\mathcal{U}_t = 0$ | $\mathcal{Z}_{t-1}\Lambda = 0$ | |
| $\Sigma_{t-1}\Sigma_{t-1} = 0$ | $\Sigma_{t-1}\mathcal{E}_t \neq 0^*$ | $\Sigma_{t-1}\mathcal{U}_t = 0$ | $\Sigma_{t-1}\Lambda = 0$ | | |
| $\mathcal{E}_t\mathcal{E}_t \neq 0$ | $\mathcal{E}_t\mathcal{U}_t \neq 0^*$ | $\mathcal{E}_t\Lambda = 0$ | | | |
| $\mathcal{U}_t\mathcal{U}_t = 0$ | $\mathcal{U}_t\Lambda = 0$ | | | | |
| $\Lambda\Lambda \neq 0$ | | | | | |

Table 4.1 tells us that, of the 21 possible sets of second derivatives, 12 are zero and 9 are not. The implications for the decision rules of agents and for the equilibrium function are striking. The perturbation parameter, $\Lambda$, will only have a coefficient different from zero in the term where it appears in a square by itself. This term is a constant that corrects for precautionary behavior induced by risk. Volatility shocks, $\Sigma_{t-1}$ , appear with coefficients different from zero only in the term where they are multiplied by the innovation to its own structural shock. Finally, innovations to the volatility shocks, $\mathcal{U}_t$, also appear with coefficients different from zero when they show up with the innovation to their own structural shock $\mathcal{E}_t$.

The main implication of Theorem 1 for our goal of evaluating the likelihood function is that, of the terms that complicate our work, only the ones associated with $[h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_1}$ and $[g_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_1}$ are different from zero. As we will see in the next corollary, this result has an important yet rather direct implication for the structure of the observation equation.

**Corollary 2.** *The second-order approximation to the measurement equation (3) can be written as:*

$$\mathbb{Y}_t = \mathbb{C} + \begin{pmatrix} \Psi_{o1}^1 (\mathbb{S}_t, \mathbb{S}_{t-1})' \\ \vdots \\ \Psi_{o5}^1 (\mathbb{S}_t, \mathbb{S}_{t-1})' \end{pmatrix} + \frac{1}{2} \begin{pmatrix} (S_t', W_{1t}', \mathbb{S}_{t-1}) \Psi_{o1}^{2,1} (S_t', W_{1t}', \mathbb{S}_{t-1})' \\ \vdots \\ (S_t', W_{1t}', \mathbb{S}_{t-1}) \Psi_{o5}^{2,1} (S_t', W_{1t}', \mathbb{S}_{t-1})' \end{pmatrix}$$
$$+ \begin{pmatrix} W_{1t}' \Psi_{o1}^{2,2} \\ \vdots \\ W_{1t}' \Psi_{o5}^{2,2} \end{pmatrix} W_{2,t}$$

*where $\Psi_{oi}^{2,1}$ denotes the cross-derivative between elements of $(S_t', W_{1t}', \mathbb{S}_{t-1})$ and $\Psi_{oi}^{2,2}$ denotes the cross-derivative between elements of $W_{1t}$ and elements of $W_{2t}$ for $i \in \{1, \ldots, 5\}$.*

We are now ready to evaluate the likelihood function. Using corollary 2 and if we define:

$$\mathbb{A}_t (S_t', W_{1t}', \mathbb{S}_{t-1}) \equiv \mathbb{Y}_t^{data} - \mathbb{C} - \begin{pmatrix} \Psi_{o1}^1 \\ \vdots \\ \Psi_{o5}^1 \end{pmatrix} (\mathbb{S}_t, \mathbb{S}_{t-1})' -$$
$$\frac{1}{2} \begin{pmatrix} (S_t', W_{1t}', \mathbb{S}_{t-1}) \Psi_{o1}^{2,1} (S_t', W_{1t}', \mathbb{S}_{t-1})' \\ \vdots \\ (S_t', W_{1t}', \mathbb{S}_{t-1}) \Psi_{o5}^{2,1} (S_t', W_{1t}', \mathbb{S}_{t-1})' \end{pmatrix}'$$

and

$$\mathbb{B}_t (W_{1t}') \equiv \begin{pmatrix} W_{1t}' \Psi_{o1}^{2,2} \\ \vdots \\ W_{1t}' \Psi_{o5}^{2,2} \end{pmatrix}$$

we have that:

$$p \left( \mathbb{Y}_t = \mathbb{Y}_t^{data} | S_t, W_{1t}, \mathbb{S}_{t-1}; \gamma \right) = \det \mathbb{B}_t^{-1} (W_{1t}') \, p \left( W_{2,t} = \mathbb{B}_t^{-1} (W_{1t}') \, \mathbb{A}_t (S_t', W_{1t}', \mathbb{S}_{t-1}) \right)$$

which is evaluated directly given that we know $\mathbb{B}_t (W_{1t}')$, $\mathbb{A}_t (S_t', W_{1t}', \mathbb{S}_{t-1})$, and the distribution of $W_{2t}$.

With this expression, evaluated at $N$ draws of $\left\{ s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i \right\}_{i=1}^N$ from the conditional densities $p \left( S_t, W_{1t}, \mathbb{S}_{t-1} | \mathbb{Y}^{data,t-1}; \gamma \right)$, the likelihood (4) can be approximated by:

$$p \left( \mathbb{Y}_t = \mathbb{Y}_t^{data} | \mathbb{Y}^{data,t-1} \right) \simeq \frac{1}{N} \sum_{i=1}^{N} \det \left( \mathbb{B}_t^{-1} \left( w_{1t}^{i\prime} \right) \right) p \left( W_{2,t} = \mathbb{B}_t^{-1} \left( w_{1t}^{i\prime} \right) A_t \left( s_t^{i\prime}, w_{1t}^{i\prime}, \boldsymbol{s}_{t-1}^i \right) \right)$$

In addition, we can find the importance weights for each draw:

$$q_t^i = \frac{\det\left(\mathbb{B}_t^{-1}\left(w_{1t}^{i\prime}\right)\right) p\left(W_{2,t} = \mathbb{B}_t^{-1}\left(w_{1t}^{i\prime}\right) A_t\left(s_t^{i\prime}, w_{1t}^{i\prime}, \boldsymbol{s}_{t-1}^i\right)\right)}{\sum_{i=1}^N \det\left(\mathbb{B}_t^{-1}\left(w_{1t}^{i\prime}\right)\right) p\left(W_{2,t} = \mathbb{B}_t^{-1}\left(w_{1t}^{i\prime}\right) A_t\left(s_t^{i\prime}, w_{1t}^{i\prime}, \boldsymbol{s}_{t-1}^i\right)\right)} \tag{11}$$

that we will use momentarily to update our swarm of particles.

To generate the $N$ draws of $\left\{s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i\right\}_{i=1}^N$, we rely on a sequential Monte Carlo that proceeds as follows (see Fernández-Villaverde and Rubio-Ramírez, 2007, for details):

---

**Step 0, Initialization:** Set $t \rightsquigarrow 1$. Sample $N$ values $\left\{s_{t-1|t-1}^i, w_{1t-1|t-1}^i, \boldsymbol{s}_{t-2|t-1}^i\right\}_{i=1}^N$ from $p\left(S_{t-1}, W_{1t-1}, \mathbb{S}_{t-2} | \mathbb{Y}^{data,t-1}; \gamma\right)$.

**Step 1, Prediction:** Sample $N$ values $\left\{s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i\right\}_{i=1}^N$ from $p\left(S_t, W_{1t}, \mathbb{S}_{t-1} | \mathbb{Y}^{data,t-1}; \gamma\right)$ using $\left\{s_{t-1|t-1}^i, w_{1t-1|t-1}^i, \boldsymbol{s}_{t-2|t-1}^i\right\}_{i=1}^N$, the law of motion for states and the distribution of shocks $\{W_{1t}, W_{2t-1}\}$.

**Step 2, Filtering:** Assign to each draw $\left(s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i\right)$ the weight $q_t^i$ in (11).

**Step 3, Sampling:** Sample $N$ times with replacement from $\left\{s_t^i, w_{1t}^i, \boldsymbol{s}_{t-1}^i\right\}_{i=1}^N$ and weights $\{q_t^i\}_{i=1}^N$. Call each draw $\left(s_{t|t}^i, w_{1t|t}^i, \boldsymbol{s}_{t-1|t}^i\right)$. If $t < T$, set $t \rightsquigarrow t+1$ and go to step 1. Otherwise stop.

---

Del Moral and Jacod (2002) and Künsch (2005) prove, under weak conditions, that this sequential Monte Carlo delivers a consistent estimator of the likelihood function and that a central limit theorem applies.

## 5. Data and Estimation

We estimate our model using five time series for the U.S. economy: 1) the relative price of investment goods with respect to the price of consumption goods, 2) the federal funds rate, 3) real output per capita growth, 4) the consumer price index, and 5) real wages per capita. Our sample covers 1959.Q1 to 2007.Q1, with 192 observations. Appendix B explains how we construct the series.

Once we have evaluated the likelihood as outlined earlier, we can either maximize it or combine it with a prior and rely on a Markov chain Monte Carlo algorithm to simulate from the posterior distribution. We follow the second route. However, we pick flat priors on a bounded support for all the parameters. The bounds are either natural economic restrictions (for instance, the Calvo and indexation parameters lie between 0 and 1) or are so wide that the likelihood assigns (numerically) zero probability to values outside them. Bounded flat priors induce a proper posterior, a convenient feature for our exercise below that assesses the fit of the model.

We resort to flat priors for two reasons. First, to reduce the impact of presample information and show that our results arise mainly from the shape of the likelihood and not from the prior (although, of course, flat priors are not invariant to reparameterization). Thus, the reader who wants to interpret our posterior modes as maximum likelihood point estimates can do so. Second, because as we learned in Fernández-Villaverde *et al.* (2009), eliciting priors for stochastic volatility is difficult, since we deal with unfamiliar units, such as the variance of volatility shocks, about which we do not have clear beliefs.

Flat priors come, though, at a price: before proceeding to the estimation, we have to fix several parameters. We are dealing with a large model that suffers from weak identification along some dimensions. Our ability to learn from the data is sharpened if we avoid searching over these dimensions.

Table 5.1: Fixed Parameters

| $\beta$ | $h$ | $\psi$ | $\vartheta$ | $\delta$ | $\alpha$ | $\kappa$ |
|---|---|---|---|---|---|---|
| 0.99 | 0.9 | 8 | 1.17 | 0.025 | 0.21 | 9.5 |
| $\varepsilon$ | $\eta$ | $\phi$ | $\Phi_2$ | $\rho_{\gamma_\Pi}$ | $\rho_{\gamma_y}$ | $\eta_y$ |
| 10 | 10 | 0 | 0.001 | 0.95 | 0 | 0 |

Table 5.1 lists the fixed parameters. Our guiding criterion in selecting them was to pick conventional values in the literature. The discount factor, $\beta = 0.99$, is a default choice, habit persistence, $h = 0.9$, matches the observed sluggish response of consumption to shocks, the parameter controlling the level of labor supply, $\psi = 8$, captures the average amount of hours in the data, and the depreciation rate, $\delta = 0.025$, induces the appropriate capital-output ratio. The elasticities of substitution, $\varepsilon = \eta = 10$, deliver average mark-ups of around 10 percent, a common value in these models. We set the fixed cost of production, $\phi$, to zero, since it is nearly irrelevant for the dynamics of the model, and we set the cost of capital utilization, $\Phi_2$, to a small number to introduce some curvature in this decision.

Three parameter values are borrowed from the point estimates from a similar model without stochastic volatility or parameter drifting presented in Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2009). The first is the inverse of the Frisch labor elasticity, $\vartheta = 1.17$. As argued by Rogerson and Wallenius (2009), this aggregate elasticity is compatible with microeconomic data, once we allow for intensive and extensive margins on labor supply. The second is the coefficient of the intermediate goods production function, $\alpha = 0.21$. This value is lower than the common calibration of Cobb-Douglas production functions in real business cycle models because, in our environment, we have positive profits that also appear as capital income in the National Income and Product Accounts. The third value that we borrow is the adjustment cost, $\kappa = 9.5$, a number in line with other estimates from DSGE models ($\kappa$ would be particularly hard to identify, since investment is not one of our observables).

The autoregressive parameter of the evolution of the response to inflation, $\rho_{\gamma_\Pi}$, is set to 0.95. In preliminary estimations, we discovered that the likelihood pushed this parameter to 1. When this happened, and although the model was still in the determinacy region, the simulations became numerically unstable: after some shocks, the reaction of interest rates to inflation could be too tepid for too long. The 0.95 value seems to be the highest possible value of $\rho_{\gamma_\Pi}$ such that the problem does not appear. The last two parameters, $\rho_{\gamma_y}$ and $\eta_y$, are equal to zero because, also in exploratory estimations, the likelihood favored values of $\eta_y \approx 0$. Thus, we decided to forget about them and make $\gamma_{y,t} = \gamma_y$.

To find the posterior, we proceed as follows. First, we define a grid of parameter values and check for the regions of high posterior density by evaluating the likelihood function in each point of the grid. This is a time-consuming procedure, but it ensures that we are searching in the right zone of the parameter space. Once we have identified the global maximum in the grid, we initialize a random-walk Metropolis-Hastings algorithm from this point. After an extensive fine-tuning of the algorithm, we draw 10,000 times from the chain.

## 6. Parameter Estimates

We now examine our parameter estimates. To ease the discussion, we group them in different tables, one for each set of parameters dealing with related aspects of the model. In all cases, we report the mode of the posterior and the standard deviation in parenthesis below (in the interest of space, we do not include the whole histograms of the posterior).

Table 6.1: Posterior, Parameters of Nominal Rigidities

| $\theta_p$ | $\chi$ | $\theta_w$ | $\chi_w$ |
|---|---|---|---|
| 0.8139<br>(0.0143) | 0.6186<br>(0.024) | 0.6869<br>(0.0432) | 0.6340<br>(0.0074) |

Table 6.1 presents the results for the nominal rigidities parameters. Our estimates indicate an economy with substantial rigidities in prices, which are reoptimized roughly once every five quarters, and in wages, which are reoptimized approximately every three quarters. Moreover, since the standard deviations are small, there is enough information on the data about this result. At the same time, there is a fair degree of indexation, between 0.62-0.63, which brings

a strong persistence of inflation. While it is tempting to compare our estimates with the evidence on the individual duration of prices as reviewed by Klenow and Malin (2009), in our model all prices and wages change every quarter. That is why, to a naive observer, our economy would look like one displaying tremendous price flexibility.

Table 6.2: Posterior, Parameters of the Stochastic Processes for Structural Shocks

| $\rho_d$ | $\rho_\varphi$ | $\Lambda_\mu$ | $\Lambda_A$ |
|---|---|---|---|
| 0.1182 (0.0049) | 0.9331 (0.0425) | 0.0034 ($6.6e-5$) | 0.0028 ($4.1e-5$) |

Table 6.2 reports the findings for the parameters of the stochastic processes for the structural shocks. We estimate a low persistence of the intertemporal preference shock and a high persistence of the intratemporal one. The low estimate of $\rho_d$ gets the quick variations in marginal utilities of consumption that match output growth and inflation fluctuations. The intratemporal shock is persistent to account for long-lived movements in hours worked. We estimate mean growth rates of technology of 0.0034 (neutral) and 0.0028 (investment-specific). Those numbers give us an average growth of the economy of 0.44 percent per quarter, or around 1.77 percent on an annual basis (0.46 and 1.86 percent in the data, respectively). Technology shocks, in our model, are deviations with respect to these drifts. Thus, we estimate that $A_t$ falls in only 8 of the 192 quarters in our sample (which roughly corresponds to the percentage of quarters where measured productivity falls in the data), even if we estimate negative innovations to neutral technology in 103 quarters.

Table 6.3: Posterior, Parameters of the Stochastic Processes for Volatility Shocks

| $\log\sigma_d$ | $\log\sigma_\varphi$ | $\log\sigma_\mu$ | $\log\sigma_A$ | $\log\sigma_m$ |
|---|---|---|---|---|
| −1.9834 (0.0726) | −2.4983 (0.0917) | −6.0283 (0.1278) | −3.9013 (0.0745) | −6.000 (0.1471) |
| $\rho_{\sigma_d}$ | $\rho_{\sigma_\varphi}$ | $\rho_{\sigma_\mu}$ | $\rho_{\sigma_a}$ | $\rho_{\sigma_m}$ |
| 0.9506 (0.0298) | 0.1275 (0.0032) | 0.7508 (0.035) | 0.2411 (0.005) | 0.8550 (0.0231) |
| $\eta_d$ | $\eta_\varphi$ | $\eta_\mu$ | $\eta_a$ | $\eta_m$ |
| 0.3246 (0.0083) | 2.8549 (0.0669) | 0.4716 (0.006) | 0.7955 (0.013) | 1.1034 (0.0185) |

The results for the parameters of the stochastic volatility processes appear in table 6.3. In all cases, the $\rho$'s and the $\eta$'s are far away from zero: the likelihood strongly favors values where stochastic volatility plays an important role. The standard deviations of the innovations of the intertemporal preference shock and of the monetary policy shock are the most persistent, while the standard deviation of the innovation of the intratemporal preference shock is the least persistent. The standard deviation of the innovations of the volatility shock to the intratemporal preference shock, $\eta_\varphi = 2.8549$, is large: the model asks for fast changes in the size of movements in marginal utilities of leisure to reproduce the hours data.

Table 6.4: Posterior, Policy Parameters

| $\gamma_R$ | $\log \gamma_y$ | $\Pi$ | $\log \gamma_\Pi$ | $\eta_\pi$ |
|---|---|---|---|---|
| 0.7855 (0.0162) | −1.4034 (0.0498) | 1.0005 (0.0043) | 0.0441 (0.0005) | 0.1479 (0.002) |

In table 6.4., we have the estimates of the policy parameters. The autoregressive component of the federal funds rate is high, 0.7855, although somewhat smaller than in estimations without parameter drift. The value of $\gamma_y$ (0.24 in levels) is similar to other results in the literature. The inflation along the balanced growth path is estimated to be 0.0005 per quarter. This parameter requires some comment. As for all the other endogenous variables, the law of motion of inflation is not centered around its value along the balanced growth path. Instead, it is moved away by the second-order terms in the solution. In this case, the constant in the second order associated with the squared value of the perturbation parameter is 0.0013. These effects allow the model to capture the level of inflation in the data.[7]

Finally, the estimated value of $\gamma_\Pi$ (1.045 in levels) guarantees local determinacy of the equilibrium. To see this, note that the relevant part of the solution of the model for local determinacy is the linear component. This component depends on $\gamma_\Pi$, the mean policy response, and not on the current value of $\gamma_{\Pi t}$. The economic intuition is that local unicity survives even if $\gamma_{\Pi t}$ temporarily violates the Taylor principle as long as there is reversion to the mean in the policy response and, thus, the agents have the expectation that $\gamma_{\Pi t}$ will satisfy the Taylor principle on average. For a related result in models with Markov-switching regime changes, see Davig and Leeper (2006). While we cannot find an analytical expression for the determinacy region, numerical experiments show that, conditional on the other point estimates, values of $\gamma_\Pi$ above 0.98 ensure uniqueness. Since the likelihood assigns zero probability to values of $\gamma_\Pi$ lower than 1.01, well inside the determinacy region, multiplicity of local equilibria is not an issue here.

[7] These effects also enormously complicate the introduction of time variation in $\Pi$. The likelihood wants to match the moments of the ergodic distribution of inflation, not the level of $\Pi$, which is inflation along the balanced growth path. When we have non-linearities, the mean of that ergodic distribution may be far from $\Pi$. Therefore, learning from that mean about $\Pi$ is hard. Learning from that mean about a time-varying $\Pi$ is even harder.

## 7. Impulse Response Functions

Before continuing the exploration of our results, we plot the impulse response functions (IRFs) generated by the model to a monetary policy shock. This exercise is a powerful reality check. If the IRFs match the shapes and sizes of those gathered by time series methods such as SVARs, it will strengthen our belief in the rest of our results. Otherwise, we should at least understand where the differences come from.

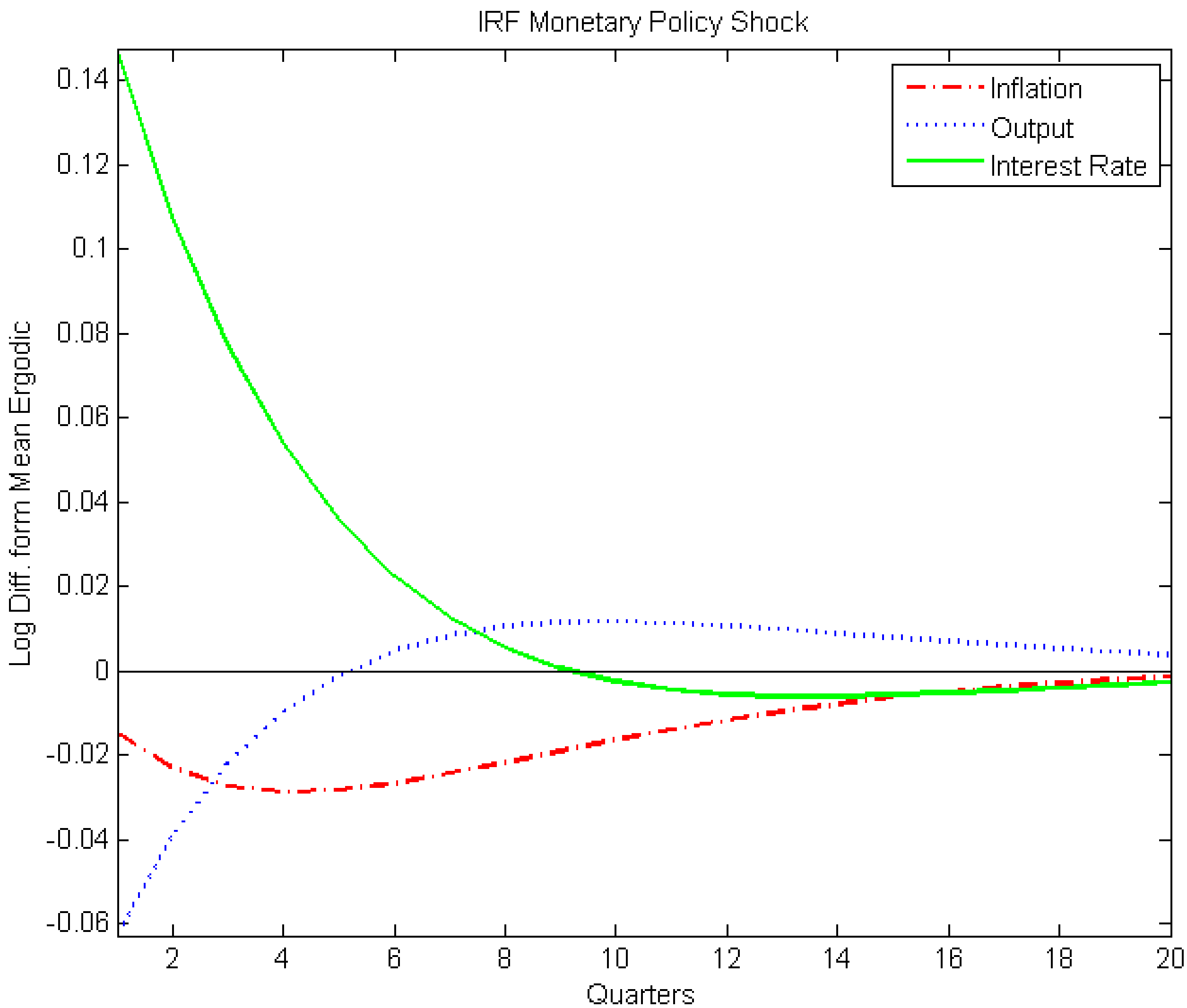


Figure 7.1: IRFs of inflation, output growth, and the federal funds rate to a monetary policy ($\varepsilon_{mt}$) shock. The responses are measured as log differences with respect to the mean of the ergodic distribution.

Auspiciously, the answer is positive: our model generates dynamic responses that are close to the ones from SVARs (see, for instance, Sims and Zha, 2006). Figure 7.1 plots the IRFs to three variables commonly discussed in monetary models: the federal funds rate, output growth, and inflation. Since we have a non-linear model, in all the figures in this section, we report the generalized IRFs starting from the mean of the ergodic distribution (Koop, Pesaran, and Potter, 1996). After a one-standard-deviation shock to the federal funds rate, inflation goes down in a hump-shaped pattern for many quarters and output growth drops.

Figure 7.2 plots the IRFs after a one-standard-deviation innovation to the monetary policy shock computed conditional on fixing $\gamma_{\Pi t}$ to the estimated mean during the tenure of each of three different chairmen of the Board of Governors: the combination Burns-Miller, Volcker, and Greenspan. This exercise tells us how the variation on the systematic component of monetary policy has affected the dynamics of aggregate variables. Furthermore, it allows a comparison with numerous similar exercises done in the literature with SVARs where the IRFs are estimated on different subsamples.

The most interesting difference is that the response of output growth under Volcker was the mildest: the estimated average stand of monetary policy under his tenure reduces the volatility of output. Inflation responds more moderately as well since the agents have the expectation that future shocks will be smoothed out by the monetary authority. In models such as ours, this stabilization of the economy is generally welfare improving since it reduces price and wage dispersion and lowers consumption fluctuations. This finding also explains why the IRFs of the interest rate are nearly on top of each other for all three periods: while we estimate that monetary policy responded more during Volcker's years for any given level of inflation than under Burns-Miller or Greenspan, this policy lowers inflation deviations and hence moderates the actual movement along the equilibrium path of the economy. Moreover, this second set of IRFs already points out one important result of this paper: we estimate that monetary policy under Burns-Miller and Greenspan was similar, while it was different under Volcker. This finding will be reinforced by the results we present below.

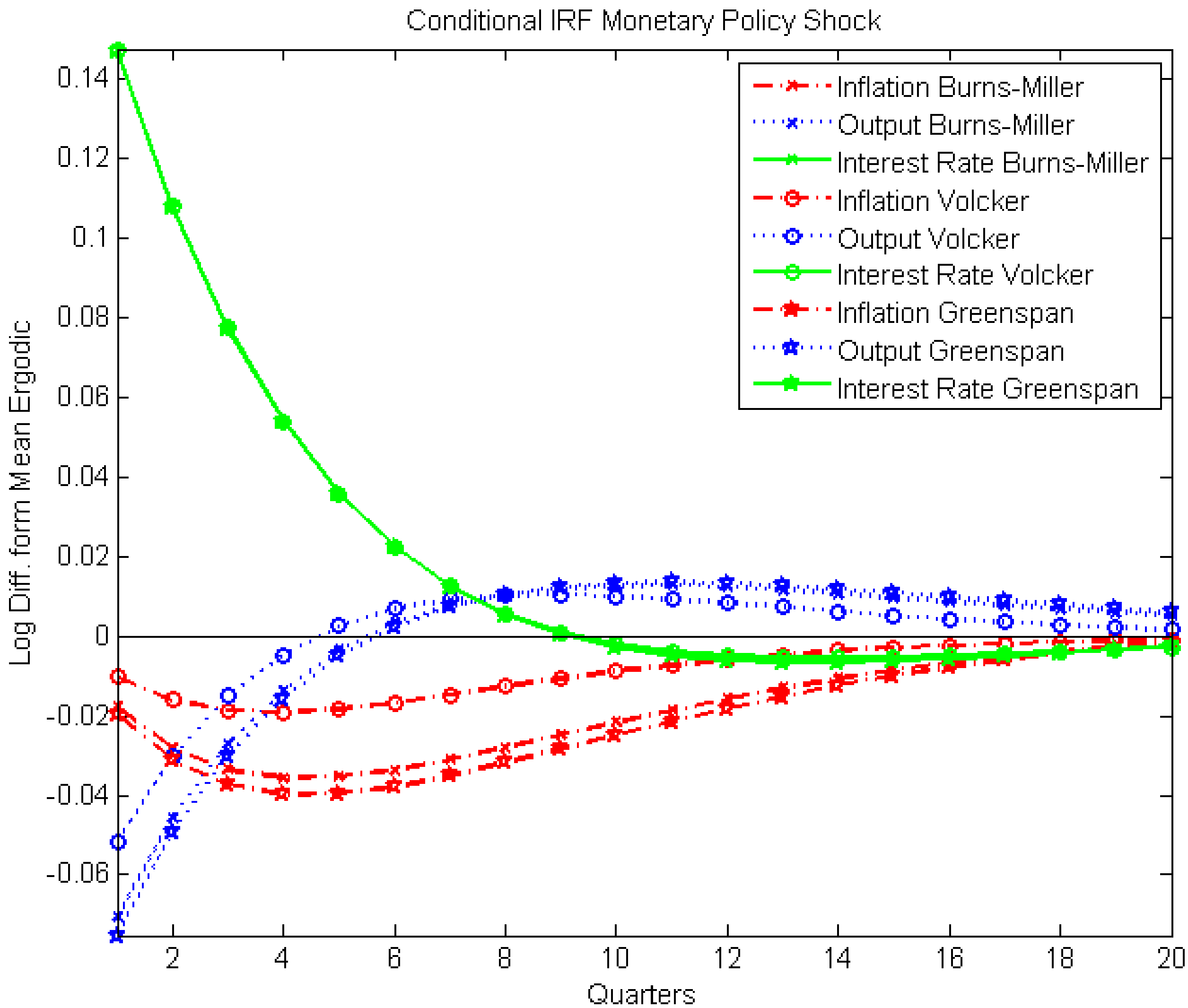


Figure 7.2: IRFs of inflation, output growth, and the federal funds rate to a monetary policy ($\varepsilon_{mt}$) shock conditional on the mean of the Taylor rule parameters of each chairmen. The responses are measured as log differences with respect to the mean of the ergodic distribution.

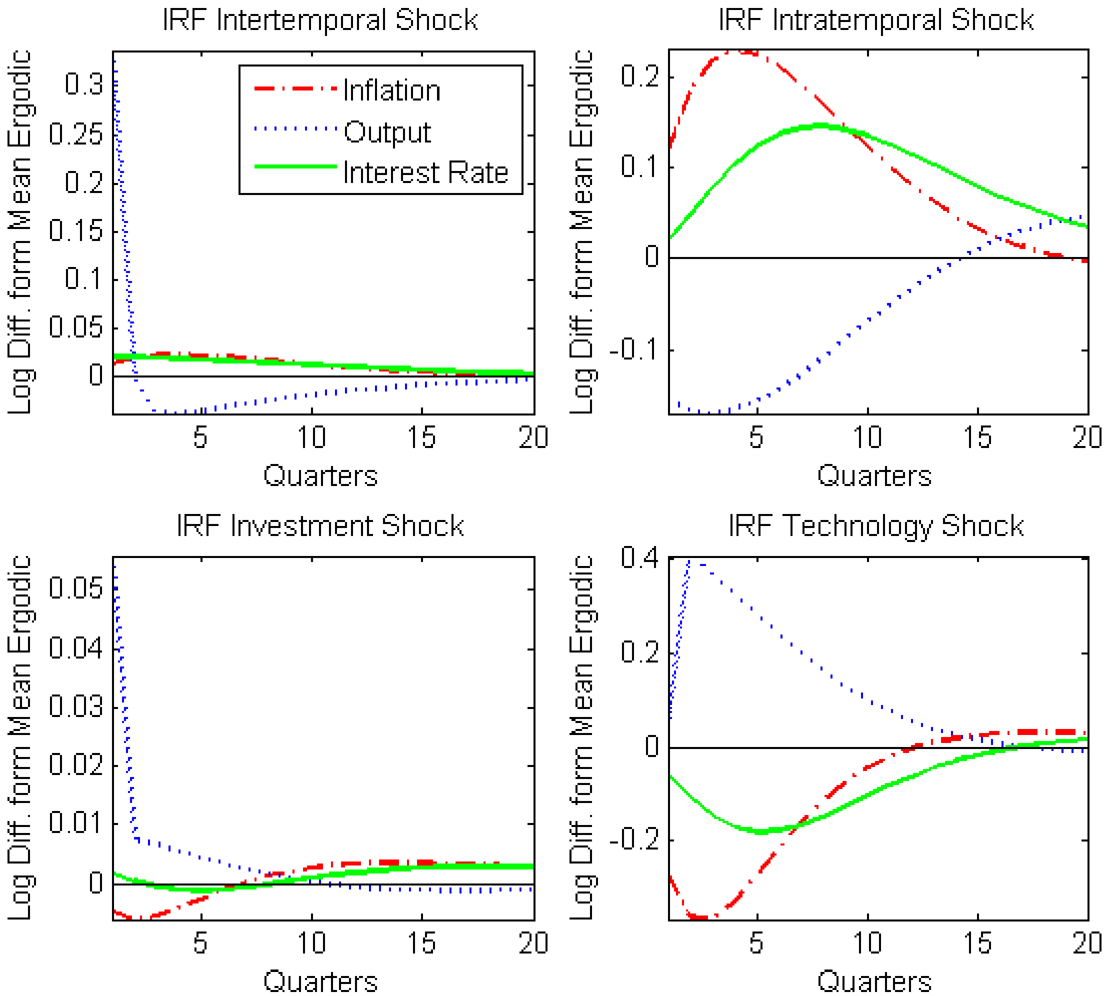


Figure 7.3: IRFs of inflation, output growth, and the federal funds rate to an intertemporal demand ($\varepsilon_{dt}$) shock, an intratemporal demand ($\varepsilon_{\phi t}$) shock, an investment-specific ($\varepsilon_{\mu t}$) shock, and a neutral technology ($\varepsilon_{At}$) shock. The responses are measured as log differences with respect to the mean of the ergodic distribution.

For completeness, we also plot, in figure 7.3, the IRFs to each of the other four shocks in our model: the two preferences shocks (intertemporal and intratemporal) and the two technology shocks (investment-specific and neutral). The behavior of the model is standard. A one-standard-deviation intertemporal preference shock raises output growth and inflation because there is an increase in the desire for consumption in the current period. The intratemporal shock lowers output because labor becomes less attractive, driving up the marginal costs and with it, prices. The two supply shocks raise output growth and lower inflation by increasing productivity.

## 8. Model Fit

After the point estimates and the IRFs, our next step is to examine the fit of the model and how it compares with alternative specifications. In this way, we document the first main finding of the paper: the data strongly support the view that monetary policy has changed over time even after including stochastic volatility. This finding corroborates our interpretation that the differences in IRFs reported in figure 7.2 are empirically relevant.

Given our Bayesian framework, a natural approach for model comparison is the computation of log marginal data densities (log MDD) and log Bayes factors. The log MDD of version $i$ of a model is defined as

$$\log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; i\right) = \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, i\right) p\left(\gamma; i\right) d\gamma, \tag{12}$$

where $p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, i\right)$ is the likelihood and $p\left(\gamma; i\right)$ is the prior for the parameters of version $i$. The log Bayes factor between specifications $i$ and $j$, a measure of the evidence in the data for one specification above the other, is

$$\log B_{i,j} = \log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; i\right) - \log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; j\right).$$

The Bayes factor is attractive because it automatically penalizes specifications with unneeded free parameters.

We compare the full model with stochastic volatility and parameter drifting (*drift*) with a version without parameter drifting (*no drift*) but with stochastic volatility. In this second case, we have two parameters less, $\rho_{\gamma_\Pi}$ and $\eta_\pi$ (but we still have the mean response $\log \gamma_\Pi$ of monetary policy to inflation deviations). To ease notation, we partition the parameter vector $\gamma$ as $\gamma = \left(\widetilde{\gamma}, \rho_{\gamma_\Pi}, \eta_\pi\right)$ where $\widetilde{\gamma}$ is the vector of all the other parameters, common to the two versions of the model.

Given that 1) our priors are uniform, 2) independent of each other, and 3) covers all the area where the likelihood is (numerically) positive, and that 4) the priors on $\widetilde{\gamma}$ are common across the two specifications of the model, we can write

$$\begin{aligned}
& \log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \textit{drift}\right) \\
= \; & \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, \textit{drift}\right) p\left(\widetilde{\gamma}\right) p\left(\rho_{\gamma_\Pi}\right) p\left(\eta_\pi\right) d\gamma = \\
= \; & \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, \textit{drift}\right) d\gamma + \log p\left(\widetilde{\gamma}\right) + \log p\left(\rho_{\gamma_\Pi}\right) + \log p\left(\eta_\pi\right),
\end{aligned}$$

where $\log p(\widetilde{\gamma})$, $\log p\left(\rho_{\gamma_\Pi}\right)$, and $\log p(\eta_\pi)$ are constants and

$$\begin{aligned} & \log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \textit{no drift}\right) \\ = \; & \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \widetilde{\gamma}, \textit{no drift}\right) p(\widetilde{\gamma})\, d\widetilde{\gamma} = \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \widetilde{\gamma}, \textit{no drift}\right) d\widetilde{\gamma} + \log p(\widetilde{\gamma}). \end{aligned}$$

Thus

$$\begin{aligned} & \log B_{\textit{drift, no drift}} = \log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \textit{drift}\right) - \log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \textit{no drift}\right) \\ = & \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, \textit{drift}\right) d\widetilde{\gamma} - \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \widetilde{\gamma}, \textit{no drift}\right) d\widetilde{\gamma} - \log p\left(\rho_{\gamma_\Pi}\right) - \log p(\eta_\pi). \end{aligned}$$

The first two terms, $\log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, \textit{drift}\right) d\gamma - \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \widetilde{\gamma}, \textit{no drift}\right) d\widetilde{\gamma}$, tell us how much better the version with parameter drift fits the data in comparison with the version with no drift. The last two terms, $\log p\left(\rho_{\gamma_\Pi}\right) + \log p(\eta_\pi)$, penalize for the presence of two extra parameters in the version with parameter drift.

We estimate the log MDDs following Geweke's (1998) harmonic mean method. This requires us to generate a new draw of the posterior of the model for the specification with no parameter drift to compute $\log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \widetilde{\gamma}, \textit{no drift}\right) d\widetilde{\gamma}$. After doing so, we find that

$$\log B_{\textit{drift, no drift}} = 126.1331 + \log p\left(\rho_{\gamma_\Pi}\right) + \log p(\eta_\pi).$$

This expression shows a potential problem of Bayes factors: by picking uniform priors for $\rho_{\gamma_\Pi}$ and $\eta_\pi$ spread out over a sufficiently large interval, we could overcome any difference in fit. But the prior for $\rho_{\gamma_\Pi}$ is pinned down by our desire to keep that process stationary, which imposes natural bounds in $[-1, 1]$ and makes $\log p\left(\rho_{\gamma_\Pi}\right) = -0.6931$. Thus, there is only one degree of freedom left: our choice of $\log p(\eta_\pi)$.

Any sensible prior for $\eta_\pi$ will only put mass in a relatively small interval: the point estimate is 0.1479, the standard deviation is 0.002, and the likelihood is numerically zero for values bigger than 0.2. Hence, we can safely impose that $\log p(\eta_\pi) > -1$ ($\log p(\eta_\pi) = -1$ would imply a uniform prior between 0 and 2.7183, a considerably wider support than any evidence in the data) and conclude that $\log B_{\textit{drift, no drift}} > 124.4400$. This is conventionally considered overwhelming evidence in favor of the model with parameter drift (Jeffreys, 1961, for instance, suggests that differences bigger than 5 are decisive).[8] Thus, even after controlling for stochastic volatility, the data strongly prefer a specification of the model where monetary policy has changed over time. This finding, however, does not imply that volatility shocks

[8] An alternative way to see this is that, to overcome the evidence in the data as recorded by the difference in loglikelihoods, $\log p(\eta_\pi)$ should be defined between 0 and $3.0054e + 054$, clearly an absurd proposition.

did not play an important role in the great moderation. In fact, as we will see in section 10, they were a key mechanism to account for it.

It has been noted that the estimation of log MDDs is dangerous because of numerical instabilities in the evaluation of the integral (12). This concern is particularly relevant in our case, since we have a large model saddled with a burdensome computation. As a robustness analysis, we also computed the Bayesian Information Criterion (BIC) (Schwarz, 1978). The BIC, which avoids the need to handle the integral in (12), can be understood as an asymptotic approximation of the Bayes factor that also automatically penalizes for unneeded free parameters. The BIC of model $i$ is defined:

$$BIC_i = -2 \ln p \left( \mathbb{Y}^T = \mathbb{Y}^{data,T}; \widehat{\gamma}, i \right) + k_i \ln n$$

where $\widehat{\gamma}$ is the maximum likelihood estimator (or, in our case given our flat priors, the mode of the posterior), $k_i$ is the number of parameters, and $n$ is the number of observations. Then, the BIC of the model with stochastic volatility and parameter drifting is:

$$BIC_{drift} = -2 * 3885 + 28 * \ln 192 = -7,622.8$$

If we eliminate parameter drifting and the parameters $\rho_{\gamma_\Pi}$ and $\eta_\pi$ associated with it (and, of course, with a new point estimate of the other parameters):

$$BIC_{no\ drift} = -2 * 3810.7 + 26 * \ln 192 = -7,484.7$$

The difference is, therefore, of over 138 log points, which is again overwhelming evidence in favor of the model with parameter drifting.

The comparison with the case without stochastic volatility is more difficult, since we are taking advantage of its presence to evaluate the likelihood. Fortunately, Justiniano and Primiceri (2008) and Fernández-Villaverde and Rubio-Ramírez (2007) estimate models similar to ours with and without stochastic volatility (in the first case, using only a first-order approximation to the decision rules of the agents and in the second with measurement errors). Both papers find that the fit of the model improves substantially when we include stochastic volatility. Finally, Fernández-Villaverde and Rubio-Ramírez (2008) compare a model with parameter drifting and no stochastic volatility with a model without parameter drifting and no stochastic volatility and report that parameter drifting is also strongly preferred by the likelihood. These findings corroborate the evidence of the importance of changes in volatility gathered by Sims and Zha (2006) using an SVAR perspective.

## 9. Smoothed Shocks

We present now the smoothed estimates of the structural shocks, volatility shocks, and drifting parameters of the model. Figure 9.1 reports the log-deviations with respect to their means for the intertemporal, intratemporal, and monetary shocks and deviations of the growth rate of the investment and technological shocks with respect to their means. To ease reading of the results, we color different vertical bars to represent each of the periods at the Federal Reserve: the McChesney Martin years from the start of our sample in 1959 to the appointment of Burns in February 1970 (white), the Burns-Miller era (light blue), the Volcker interlude from August 1979 to August 1987 (grey), the Greenspan times (orange), and Bernanke's tenure from February 2006 (yellow).

We see in the top left panel of figure 9.1 that the intertemporal shock, $\widehat{d}_t$, is particularly high in the 1970s. This increases households' desire for current consumption (for instance, because of the entrance of baby boomers into adulthood). A higher aggregate demand triggers, in the model, the higher inflation observed in the data for those years and generates challenges for the monetary authority that we will discuss below. The shock has a dramatic drop in the second quarter of 1980. This is precisely the quarter where the Carter administration invoked the Credit Control Act (started on March 14, 1980). Schreft (1990) documents that this measure caused turmoil in financial markets and, most likely, distorted intertemporal choices of households, which is reflected in the large negative innovation to $\widehat{d}_t$. The low values of $\widehat{d}_t$ in the 1990s with respect to the 1970s and 1980s eased the inflationary pressures in the economy.

The shock to the utility of leisure, $\widehat{\varphi}_t$, grows in the 1970s and falls in the 1980s to stabilize at a very low value in the 1990s. The likelihood wants to track, in this way, the path of average hours worked: low in the 1970s, increasing in the 1980s, and stabilizing in the 1990s. Higher hours also lower the marginal cost of firms (wages fall relative to the technological level). The reduction in marginal costs also helped to reduce inflation during Greenspan's tenure.

The evolution of the investment-specific technology, $\widehat{\frac{\mu_t}{\mu_{t-1}}}$, shows a clear drop after 1973 (when it is likely that energy-intensive capital goods suffered the consequences of the oil shocks in the form of economic obsolescence) and very positive realizations in the late 1990s (our model interprets the sustained boom of those years as the consequence of strong improvements in investment technology). These positive realizations were an additional help to contain inflation during those years. In comparison, the neutral-technology shocks, $\widehat{\frac{A_t}{A_{t-1}}}$, have been stable since 1959, with only a few big shocks at the end of the sample.

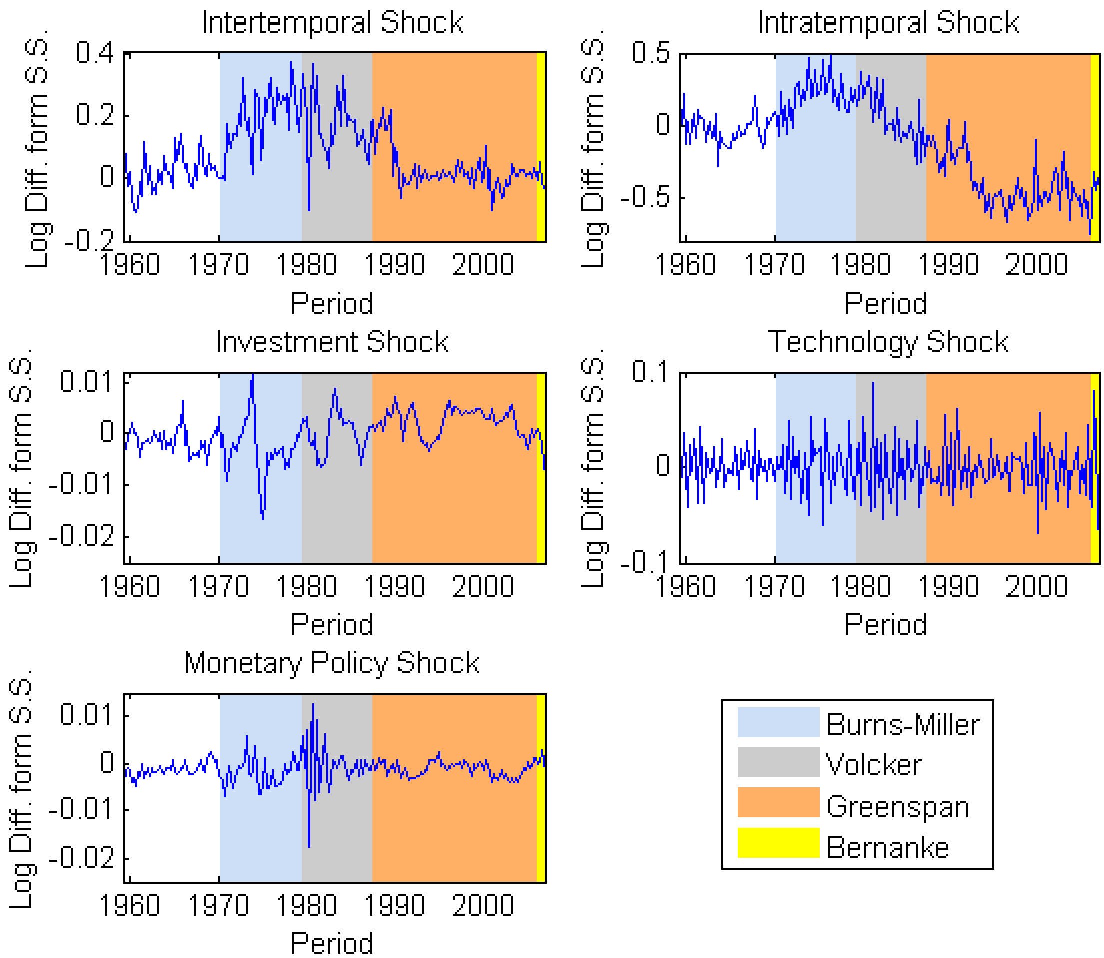


Figure 9.1: Smoothed intertemporal demand ($\widehat{d_t}$) shock, intratemporal demand ($\widehat{\phi_t}$) shock, investment-specific ($\widehat{\frac{\mu_t}{\mu_{t-1}}}$) shock, technology ($\widehat{\frac{A_t}{A_{t-1}}}$) shock, and monetary policy ($\widehat{\xi_t}$) shock.

The evolution of the monetary policy shock, $\widehat{\xi}_t$, reveals large innovations the early 1980s. This is due both to the fast change in policy brought about by Volcker in and to the fact that a Taylor rule might not fully capture the dynamics of monetary policy during a period in which money growth targeting was attempted. Sims and Zha (2006) also find that the Volcker period appears to be one with large disturbances to the policy rule and argue that the Taylor rule formalism can be a misleading perspective from which to view policy during that time. Our evidence from the estimated intertemporal, intratemporal, and investment shocks suggests that monetary authorities faced a more difficult environment in the 1970s and early 1980s than in the 1990s.

As a way to gauge the level of uncertainty of our smoothed estimates, we plot in figure 9.2 the same shock plus/minus two standard deviations. The lesson to take away from this figure is that, in all cases, the data are informative about the history we just narrated.

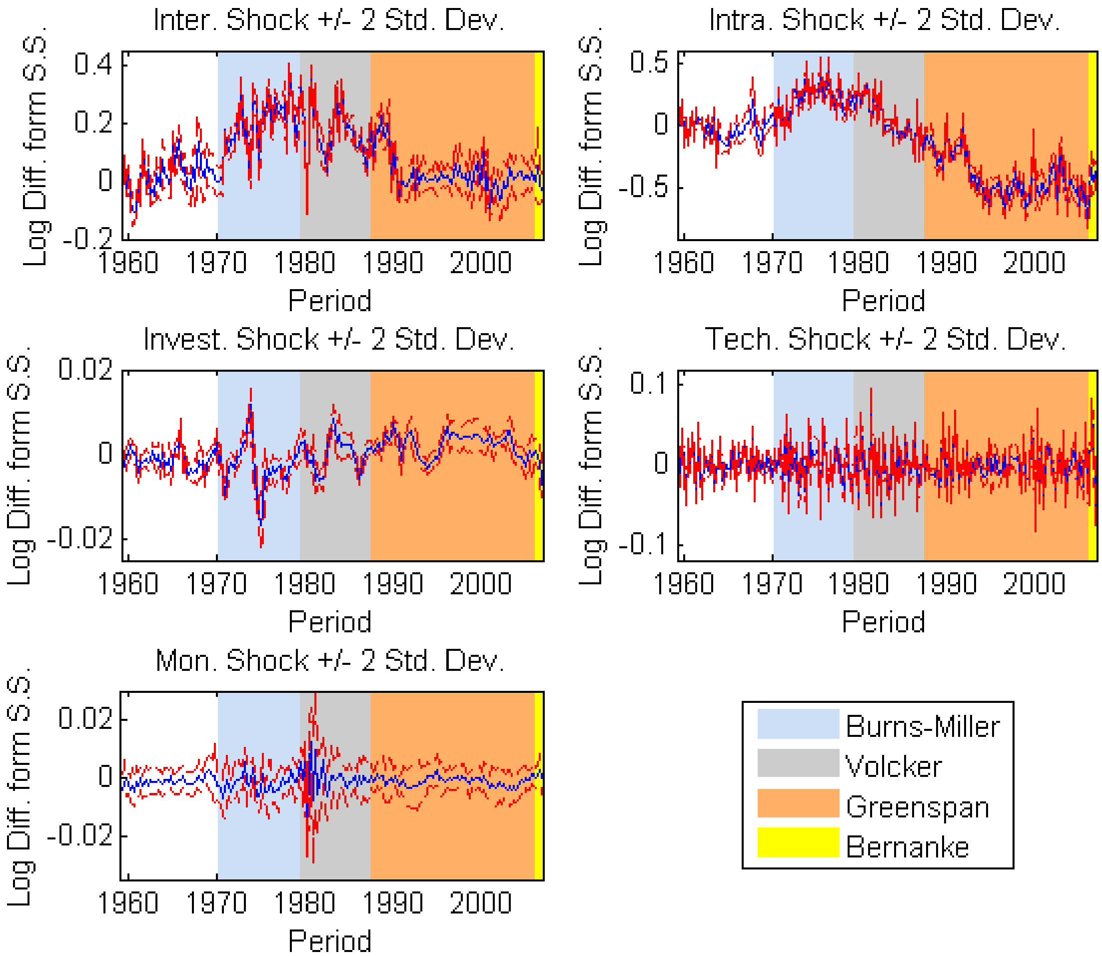


Figure 9.2: Smoothed intertemporal demand ($\widehat{d_t}$) shock, intratemporal demand ($\widehat{\phi_t}$) shock, investment-specific ($\widehat{\frac{\mu_t}{\mu_{t-1}}}$) shock, technology ($\widehat{\frac{A_t}{A_{t-1}}}$) shock, and monetary policy ($\widehat{\xi_t}$) shock +/- 2 Standard Deviations.

We move now, in figure 9.3, to plot the evolution of the standard deviation of the innovation of the structural shocks, all of them in log-deviations with respect to their estimated means. We see in this figure that the standard deviation of the intertemporal shock was particularly high in the 1970s and only slowly went down during the 1980s and early 1990s.

By the end of the sample, the standard deviation of the intertemporal shock was roughly at the level where it started. In comparison, the standard deviation of all the other shocks is relatively stable except, perhaps, for the big drop in the standard deviation of the monetary policy shock in the early 1980s and the big changes in the standard deviation of the investment shock during the period of oil price shocks. Hence, the 1970s and the 1980s were more volatile than the 1960s and the 1990s, creating a tougher environment for monetary policy. This result also confirms Blanchard and Simon's (2001) observation that volatility had a downward trend in the 20th century with an abrupt and temporal increase in the 1970s.

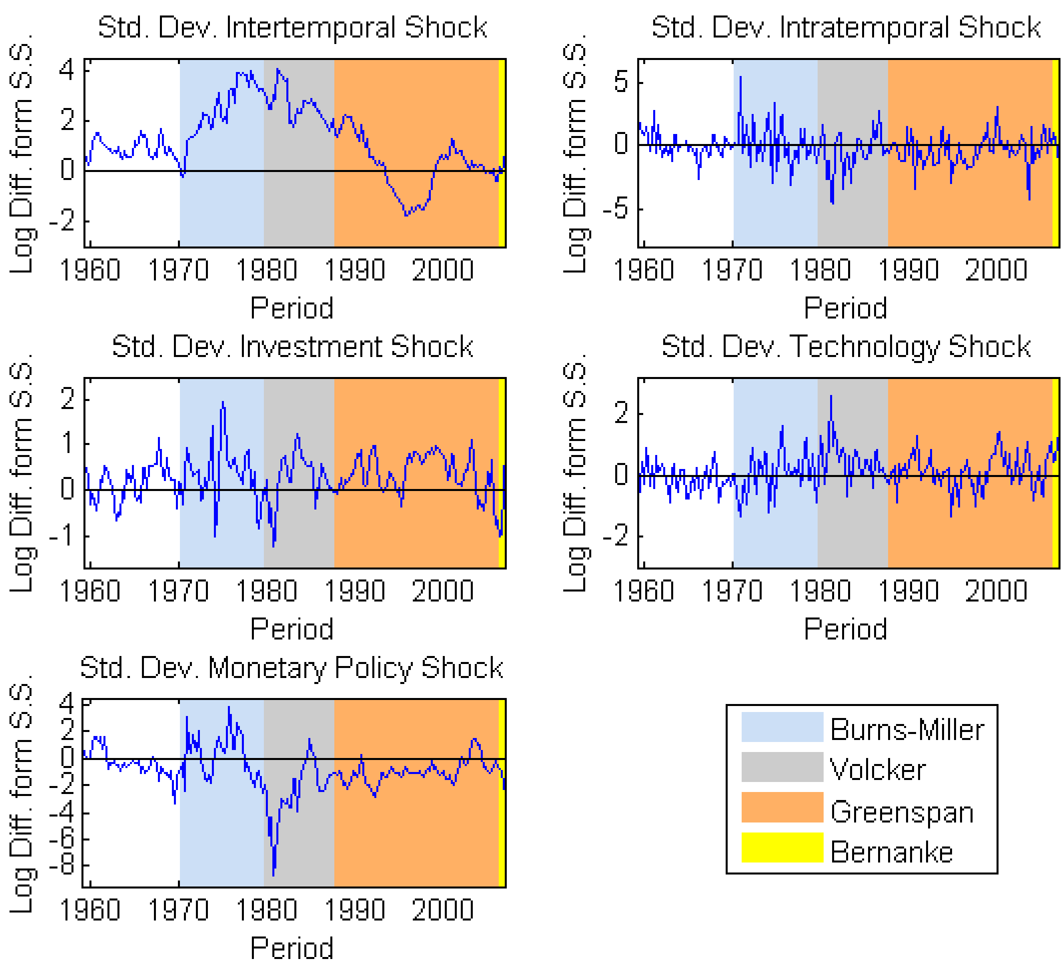


Figure 9.3: Smoothed standard deviation shocks to the intertemporal demand ($\widehat{\sigma}_{dt}$) shock, the intratemporal demand ($\widehat{\sigma}_{\phi t}$) shock, the investment-specific ($\widehat{\sigma}_{\mu t}$) shock, the technology ($\widehat{\sigma}_{At}$) shock, and the monetary policy ($\widehat{\sigma}_{mt}$) shock.

In figure 9.4, we plot the same results except that now we add two standard deviations to assess posterior uncertainty. Again, the lesson from this figure is that the big movements in the different series that we report can be ascertained with a reasonable degree of confidence.

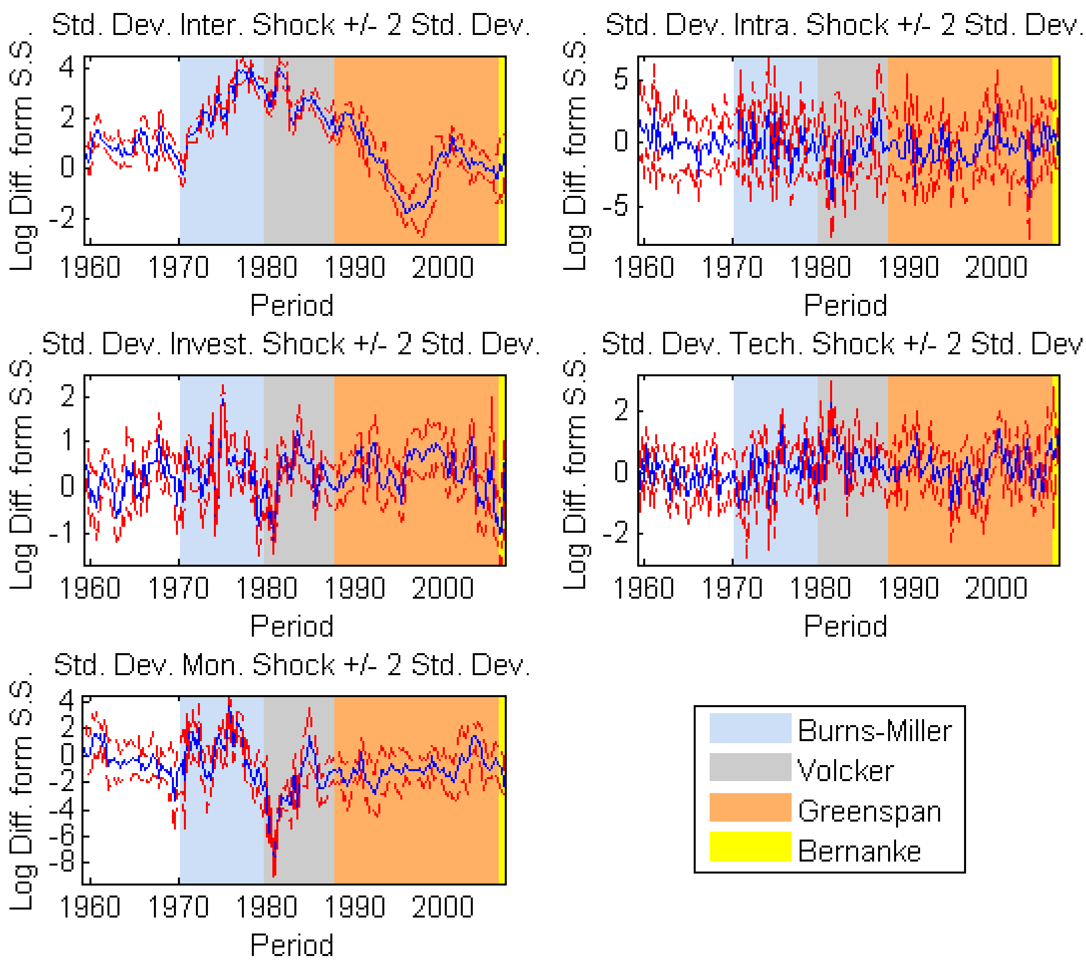


Figure 9.4: Smoothed standard deviation shocks to the intertemporal demand ($\widehat{\sigma}_{dt}$) shock, the intratemporal demand ($\widehat{\sigma}_{\phi t}$) shock, the investment-specific ($\widehat{\sigma}_{\mu t}$) shock, the technology ($\widehat{\sigma}_{At}$) shock, and the monetary policy ($\widehat{\sigma}_{mt}$) shock +/- 2 standard deviations.

Finally, in figure 9.5, we plot the evolution of the response of monetary policy to inflation plus/minus a two-standard-deviation interval. Figure 9.5 shows us an intriguing narrative. The parameter $\gamma_{\Pi t}$ started the sample around its estimated mean, slightly over 1, and it grew more or less steadily during the 1960s until reaching a peak in early 1968. After that year, $\gamma_{\Pi t}$

suffered a fast collapse that took it below 1 in 1971. To put this evolution in perspective, it is useful to remember that Burns was appointed chairman in February 1970. The parameter stayed below 1 for all of the 1970s, showing either that monetary policy did not satisfy the Taylor principle or that our postulated monetary policy rule is not a good description of the behavior of the Fed at the time (for example, because the Fed was using real-time data to make its decisions; see Orphanides, 2002).

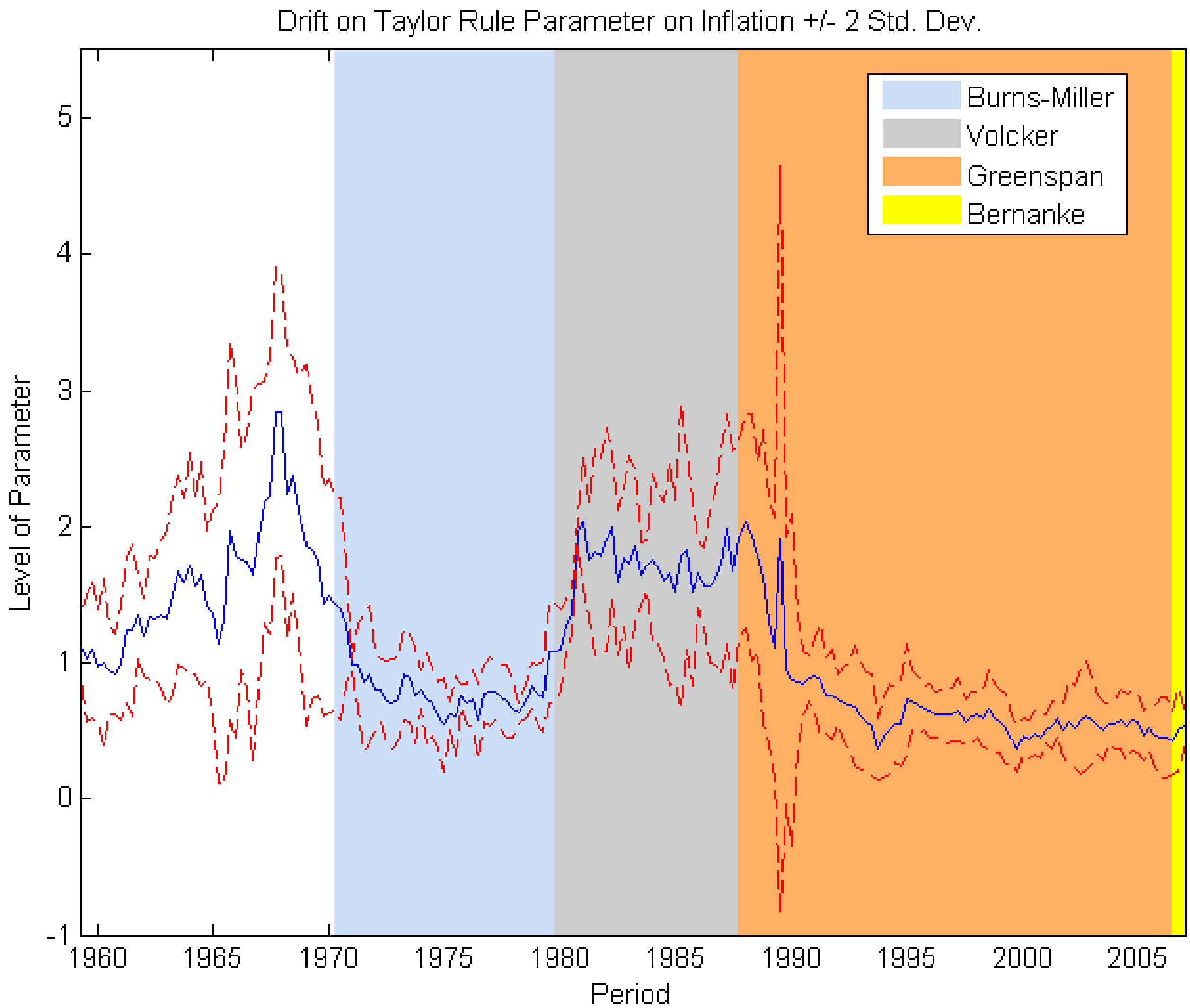


Figure 9.5: Smoothed path for the Taylor rule parameter on inflation +/- 2 standard deviations.

The arrival of Volcker is quickly picked up by our smoothed estimates: $\gamma_{\Pi t}$ increases to over 2 after a few months and stays high during all the years of Volcker's tenure. Interestingly, our estimate captures well the observation by Goodfriend and King (2007) that monetary policy tightened in the spring of 1980 as inflation and long-run inflation expectations continued to grow. The level of $\gamma_{\Pi t}$ stayed roughly constant at this high during the remainder of Volcker's tenure.

But as quickly as $\gamma_{\Pi t}$ rose when Volcker arrived, it went down again when he departed. Greenspan's tenure at the Fed meant that, by 1990, the response of the monetary authority to inflation was again below 1. During all the following years, $\gamma_{\Pi t}$ was low, probably even below the values that it took during Burns-Miller's time. Moreover, our estimates of $\gamma_{\Pi t}$ are relatively tight, suggesting that posterior uncertainty may not be the full explanation behind these movements. But if monetary policy in the 1970s was similar to policy in the 1990s, how can our model account for the different economic performance between the two decades? While our smoothed shocks already hint at the reason (the 1990s were characterized by favorable shocks and low volatility, compared with either Burns-Miller's or Volcker's time), to satisfactorily answer this question, in the next section we build counterfactual histories.

## 10. Historical Counterfactuals

One important goal of our research is to quantify how much of the observed changes in the volatility of aggregate variables can be accounted for by changes in the standard deviations of shocks and how much by changes in policy. To accomplish this, we build a number of historical counterfactuals. These are internally coherent exercises in which we remove one source of variation at a time and we measure how aggregate variables would have behaved when hit only by the remaining shocks. Since our model is structural in the sense of Hurwicz (1962) (it is invariant to interventions, including shocks by nature such as the ones we are postulating), we will obtain an answer that is robust to the Lucas critique.

In the next two subsections, we will always plot the same three basic variables that we used in Section 7: inflation, output growth, and the federal funds rate. Counterfactual histories of other variables could be built analogously. Also, we will have vertical bars for the tenure of each chairman, following the same color scheme as in Section 9.

### 10.1. Counterfactual I: No Volatility Changes

In our first historical counterfactual, we compute how the economy would have behaved in the absence of changes in the volatility of the shocks, that is, if the volatility of the innovation of the structural shocks had been fixed at its historical mean. We can think about this exercise as measuring the effect of virtue by eliminating fortune. To do so, we back up the smoothed structural shocks as we did in section 9 and we feed them to the model, given our parameter point estimates and the historical mean of volatility, to generate series for inflation, output, and the federal funds rate. We present our findings in two steps. First, we discuss several moments of the data and the counterfactuals. Second, we compare the paths of our counterfactual histories over time with the observed data.

### 10.1.1. Moments

Table 10.1 reports the moments of the data (in annualized terms) and the moments from the counterfactual history (*no s.v.* in the table stands for "no stochastic volatility"). In both cases, we include the moments for the whole sample and for the sample divided before and after 1984.Q1, a conventional date for the start of the great moderation (McConnell and Pérez-Quirós, 2000). In the last two rows of the table, we compute the ratio of the moments after 1984.Q1 over the moments before 1984.Q1.[9]

Table 10.1: No Volatility Changes, Data versus Counterfactual History

| | Means | | | Standard Deviations | | |
|---|---|---|---|---|---|---|
| | Inflation | Output Growth | FFR | Inflation | Output Growth | FFR |
| Data | 3.8170 | 1.8475 | 6.0021 | 2.6181 | 3.5879 | 3.3004 |
| Data, pre 1984.1 | 4.6180 | 1.9943 | 6.7179 | 3.2260 | 4.3995 | 3.8665 |
| Data, after 1984.1 | 2.9644 | 1.6911 | 5.2401 | 1.3113 | 2.4616 | 2.3560 |
| No s.v. | 2.5995 | 0.7169 | 6.9388 | 3.5534 | 3.1735 | 2.4128 |
| No s.v., pre-1984.1 | 2.0515 | 0.9539 | 6.3076 | 3.7365 | 3.4120 | 2.7538 |
| No s.v., after-1984.1 | 3.1828 | 0.4647 | 7.6106 | 3.2672 | 2.8954 | 1.7673 |
| Data, post-1984.1/pre-1984.1 | 0.6419 | 0.8480 | 0.7800 | 0.4065 | 0.5595 | 0.6093 |
| No s.v., post-1984.1/pre-1984.1 | 1.5515 | 0.4871 | 1.2066 | 0.8744 | 0.8486 | 0.6418 |

Some of the numbers in table 10.1 are well known. For instance, after 1984, the standard deviation of inflation falls by nearly 60 percent, the standard deviation of output growth falls by 44 percent, and the standard deviation of federal funds rate falls by 39 percent. In terms of means, after 1984, there is less inflation and the federal funds rate is lower, but output growth is also 15 percent lower.

More novel are the numbers that come from our counterfactual. Without changes in volatility, the great moderation would have been noticeably smaller. The standard deviation

[9] The benchmark model with stochastic volatility plus parameter drifting replicates the data exactly. Hence, there is no need to report the volatility results from the model in table 10.1.

of inflation would have fallen by only 13 percent, the standard deviation of output growth would have fallen by 16 percent, and the standard deviation of the federal funds rate would have fallen by 35 percent, that is, only 33, 20, and 87 percent, respectively, of how much they would have fallen otherwise.[10]

Table 10.1 documents the second main finding of the paper: without changes in the standard deviations of the innovation of the structural shocks, the great moderation would not have been nearly as big as we observed in the data. How do we reconcile these findings with the findings in section 8? While there is strong evidence of changes in the systematic component of monetary policy, they can account for only a fraction of the great moderation observed after 1984. In other words: according to our estimated model, monetary policy has a relatively small role in affecting aggregate outcomes besides inflation. Also, without stochastic volatility, output growth would have been quite lower on average. As we will see in the next section, this would have had important consequences for the performance of the economy during the last 15 years of our sample.

### 10.1.2. Counterfactual Paths

To further illustrate the previous results, figure 10.1 compares the whole path of the counterfactual history (blue line) with the observed one (red line). Figure 10.1 tells us that volatility shocks mattered all across the sample. The run-up for inflation would have been much slower in the late 1960s (inflation would have actually been negative during the last years of Martin's tenure) with small effects on output growth or the federal funds rate (except at the very end of the sample). Inflation would not have picked up as nearly as much during the first oil shock, but output growth would have suffered. During Volcker's time, inflation would also have fallen faster with little cost to output growth. These are indications that both Burns-Miller and Volcker suffered from large and volatile shocks to the economy.

In comparison, during the 1990s, inflation would have been more volatile, with a big increase in the middle of the decade. Similarly, during those years, output growth would have been much lower, with a long recession between 1994 and 1998, and the federal funds rate would have been prominently higher. Confirming the results presented in section 9, this is yet another manifestation of how placid the 1990s were for policymakers.

[10] We must resist here the temptation of undertaking a standard variance decomposition exercise. Since we have a second-order approximation to the policy function and its associated cross-product terms, we cannot neatly divide total variance among the different shocks as we could do in the linear case.

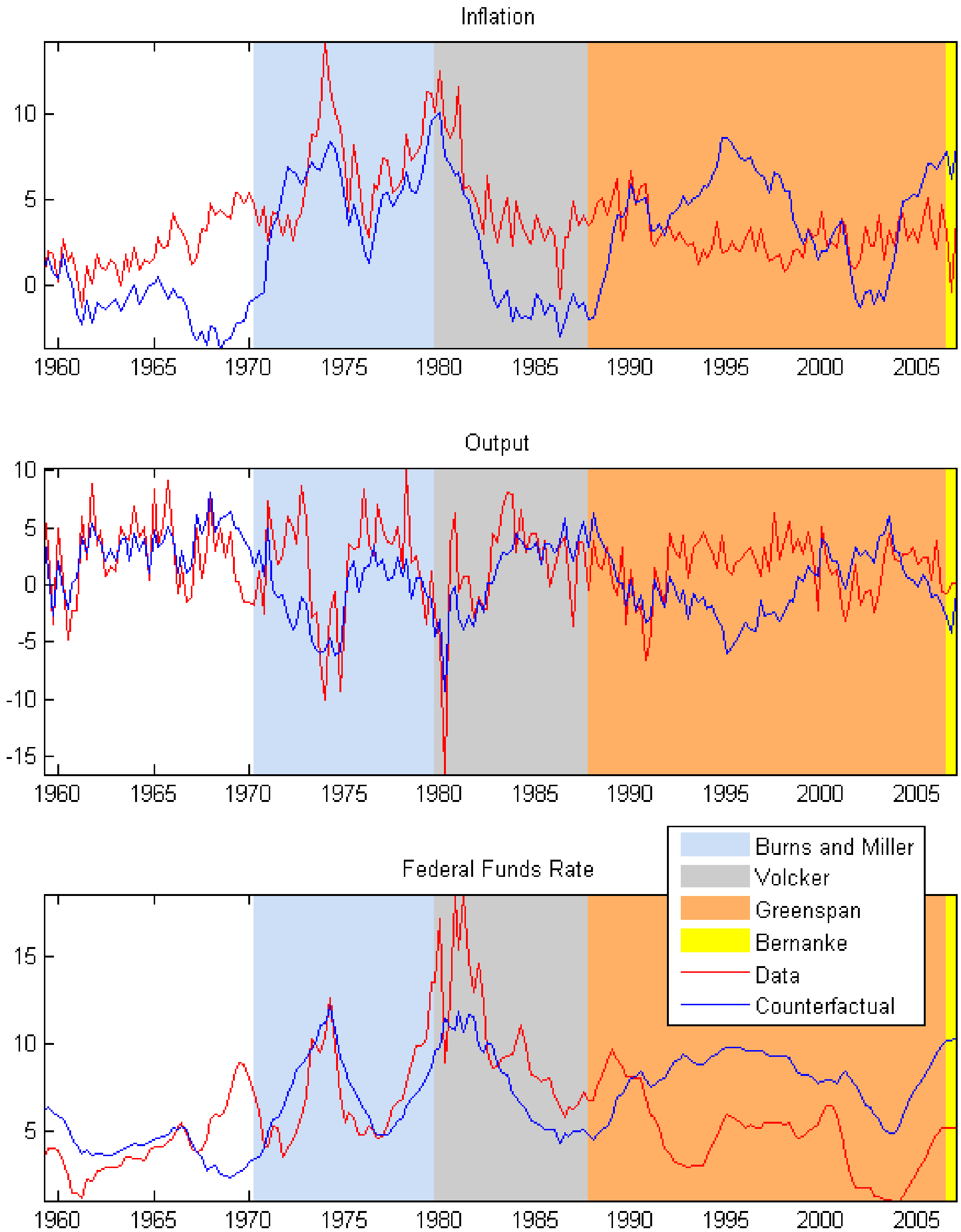


Figure 10.1: Counterfactual history with no changes in volatility.

### 10.2. Counterfactual II: Switching Chairmen

In our second counterfactual we move one chairman from his mandate to an alternative time period. For example, we appoint Greenspan as chairman during the Burns-Miller years. By that, we mean that the monetary authority would have followed the policy rule dictated by the average $\gamma_{\Pi t}$ that we estimated during Greenspan's time while starting from the same states as Burns-Miller and suffering the same shocks (both structural and of volatility). We repeat this exercise with all the other possible combinations: Volcker in the Burns-Miller decade, Burns-Miller in Volcker's mandate, Greenspan in Volcker's time, Burns-Miller in the Greenspan years, and, finally, Volcker in Greenspan's time.

It is important to be careful in interpreting this exercise. By appointing Greenspan at Volcker's time, we do not literally mean Greenspan as a person, but Greenspan as a convenient label for a particular monetary policy response to shocks that according to our model were observed during his tenure. The real Greenspan could have behaved in a different way, for example, as a result of some non-linearities in monetary policy that are not properly captured by a simple rule as the one we postulated in section 2. In fact, the argument could be pushed one step further and we could think about the appointment of Volcker as an endogenous response of the political-economic equilibrium to high inflation.[11]

Another issue that we sidelined is the evolution of expectations. In our model, agents have rational expectations and observe the changes in monetary policy parameters. This hypothesis may be a poor approximation of the agents' behavior in real life. It could be the case that $\gamma_{\Pi t}$ was high in 1984, even though inflation was already low by that time, because of the high inflationary expectations that economic agents held during most of the 1980s (this point is also linked to issues of commitment and credibility that our model does not address). While we see all these arguments as interesting lines of research, we find it important to focus first on our basic counterfactual conditional. As before, in the next two subsections, we explore first the moments of the counterfactual and later we plot the whole paths.

#### 10.2.1. Moments

In table 10.2, we report the mean and the standard deviation of inflation, output growth, and the federal funds rate in the observed data and in the six counterfactual ones. The means of the observed data present us with a history similar to that in table 10.1. Inflation was high with Burns-Miller, fell with Volcker, and stayed low with Greenspan. Output growth went

[11] In our model agents have a probability distribution regarding possible changes of monetary policy in the next periods, but those changes are uncorrelated with current conditions. Therefore, our model cannot capture the endogeneity of policy selection.

down during the Volcker years to recover with Greenspan. The federal funds rate reached its peak with Volcker. The standard deviation of the observed data tells us yet one more time about the great moderation: from a standard deviation of output growth of 4.7 in Burns-Miller's time, we went to a standard deviation of 2.45 with Greenspan, a cut in half. Similarly, inflation volatility fell nearly 54 percent and the federal funds rate volatility 5 percent.

Table 10.2: Switching Chairmen, Data versus Counterfactual Histories

| | Means | | | Standard Deviations | | |
|---|---|---|---|---|---|---|
| | Inflation | Output Gr. | FFR | Inflation | Output Gr. | FFR |
| BM (data) | 6.2333 | 2.0322 | 6.5764 | 2.7347 | 4.7010 | 2.2720 |
| Greenspan to BM | 6.8269 | 1.8881 | 6.5046 | 3.3732 | 4.6781 | 2.0103 |
| Volcker to BM | 4.3604 | 1.5010 | 7.6479 | 2.4620 | 4.6219 | 2.3470 |
| Volcker (data) | 5.3584 | 1.3846 | 10.3338 | 3.1811 | 4.4811 | 3.4995 |
| BM to Volcker | 6.4132 | 1.3560 | 10.4126 | 2.9728 | 4.4220 | 3.0648 |
| Greenspan to Volcker | 6.7284 | 1.3423 | 10.4235 | 2.9824 | 4.3730 | 2.8734 |
| Greenspan (data) | 2.9583 | 1.5177 | 4.7352 | 1.2675 | 2.4567 | 2.1887 |
| BM to Greenspan | 2.3355 | 1.5277 | 4.4529 | 1.5625 | 2.4684 | 2.4652 |
| Volcker to Greenspan | -0.4947 | 1.3751 | 3.6560 | 1.7700 | 2.4705 | 2.7619 |

But table 10.2 also tells us other things. Contrary to the conventional wisdom, our estimates suggest that the stand of monetary policy against inflation under Greenspan was not particularly strong. In Burns-Miller's time, the monetary policy under Greenspan would have delivered slightly higher average inflation, 6.83 versus the observed 6.23, accompanied by a slightly lower federal funds rate and lower output growth, 1.89 versus the observed 2.03. The difference is even bigger in Volcker's time, during which average inflation would have been nearly 1.4 percent higher, while output growth would have been virtually identical (1.34 versus 1.38). The key for this finding is in the behavior of the federal funds rate, which would have increased only by 9 basis points, on average, if Greenspan had been in charge of the Fed instead of Volcker. Given the higher inflation in the counterfactual, the slightly higher nominal interest rates would have implied much lower real rates.

The counterfactual of Burns-Miller in Greenspan's and Volcker's time casts doubts on the malignant reputations of these two short-lived chairmen, at least when compared with Greenspan. Burns-Miller would have brought even slightly lower inflation than Greenspan, thanks to a higher real federal funds rate and a bit higher output growth. However, Burns-Miller would have delivered higher inflation than Volcker.

This is the third main empirical finding of the paper: according to our estimates, Volcker's response to inflation was high. Greenspan's response was milder. Instead, he was favored by a long sequence of good shocks and low volatility. In fact, Greenspan seems to have behaved quite similarly to how Burns-Miller would have behaved. This result will be confirmed by the counterfactual paths in the next subsection.

### 10.2.2. Counterfactual Paths

Our next exercise is to plot the whole counterfactual histories summarized in table 10.2. We find it interesting to plot the whole history because changes in the economy's behavior in one period will propagate over time and we want to understand, for example, how Greenspan's legacy would have molded Volcker's tenure. Also, plotting the whole history allows us to track the counterfactual response of monetary policy to large economic events such as the oil shocks.

Our first plot is figure 10.2, where we put Greenspan in Burns-Miller's time. The pick up in inflation after the first oil shock of 1973 would have been even higher. This would have been caused by a lower federal funds rate, which, coupled with faster price increases, would have turned out to yield a much lower real interest rate. Greenspan would have left Volcker with approximately the same legacy as Burns-Miller. We can see this from how close the counterfactual history was during Volcker's mandate (where, remember, Volcker is responding to inflation as our estimates indicate he did in the data but where he is facing the states of the economy left by Greenspan, not the ones left by Burns-Miller).

In figure 10.3, we repeat the same exercise for Greenspan during Volcker's time. The main difference is that, in our counterfactual history, Greenspan would have disinflated more slowly than Volcker did. Moreover, the disinflation would have lasted longer, well into the 1990s. As we saw before, this slower disinflation would not have had increased output growth. In terms of nominal interest rates, those would have been lower at the beginning of the 1980s, reflecting a softer stand against inflation, but higher in the late 1980s because of the higher inflation during those years.

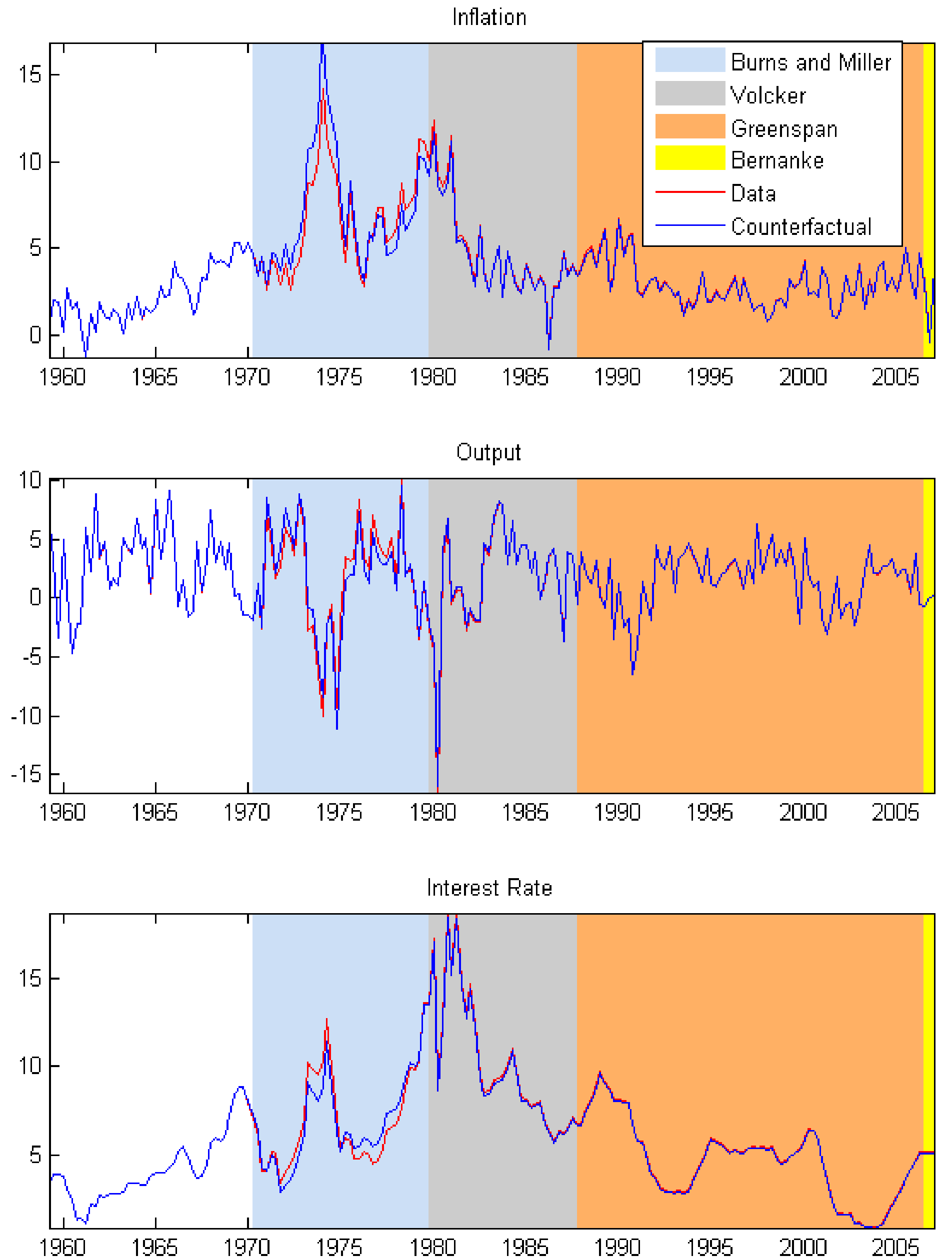


Figure 10.2: Greenspan during the Burns-Miller years

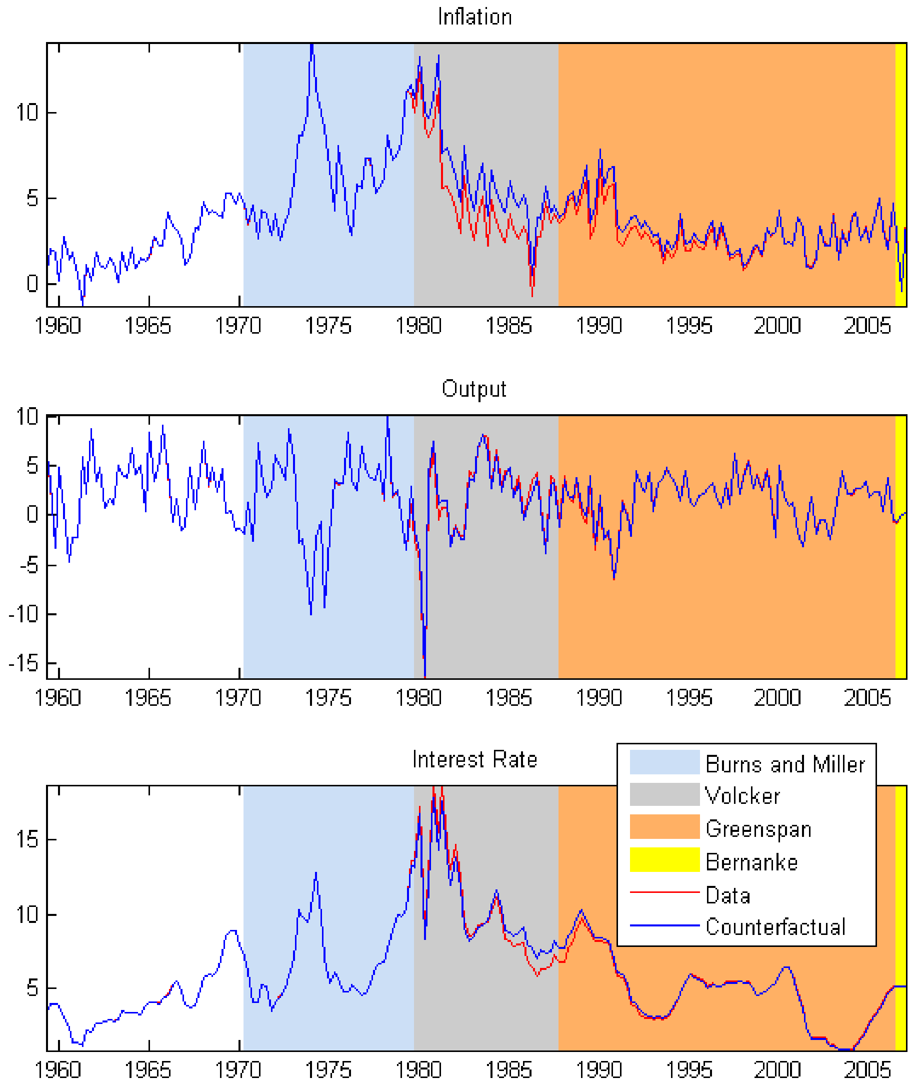


Figure 10.3: Greenspan during the Volcker years

In figure 10.4, we move to Burns-Miller being reappointed in Greenspan's time. The first thing we see is that inflation would have been lower with Burns-Miller than it was in the data, while output growth would have been nearly identical. Also, the federal funds rate would have been a bit lower due to lower inflation. This is yet another piece of evidence that the differences in monetary policy under Greenspan's and Burns-Miller may have been overstated by the literature.

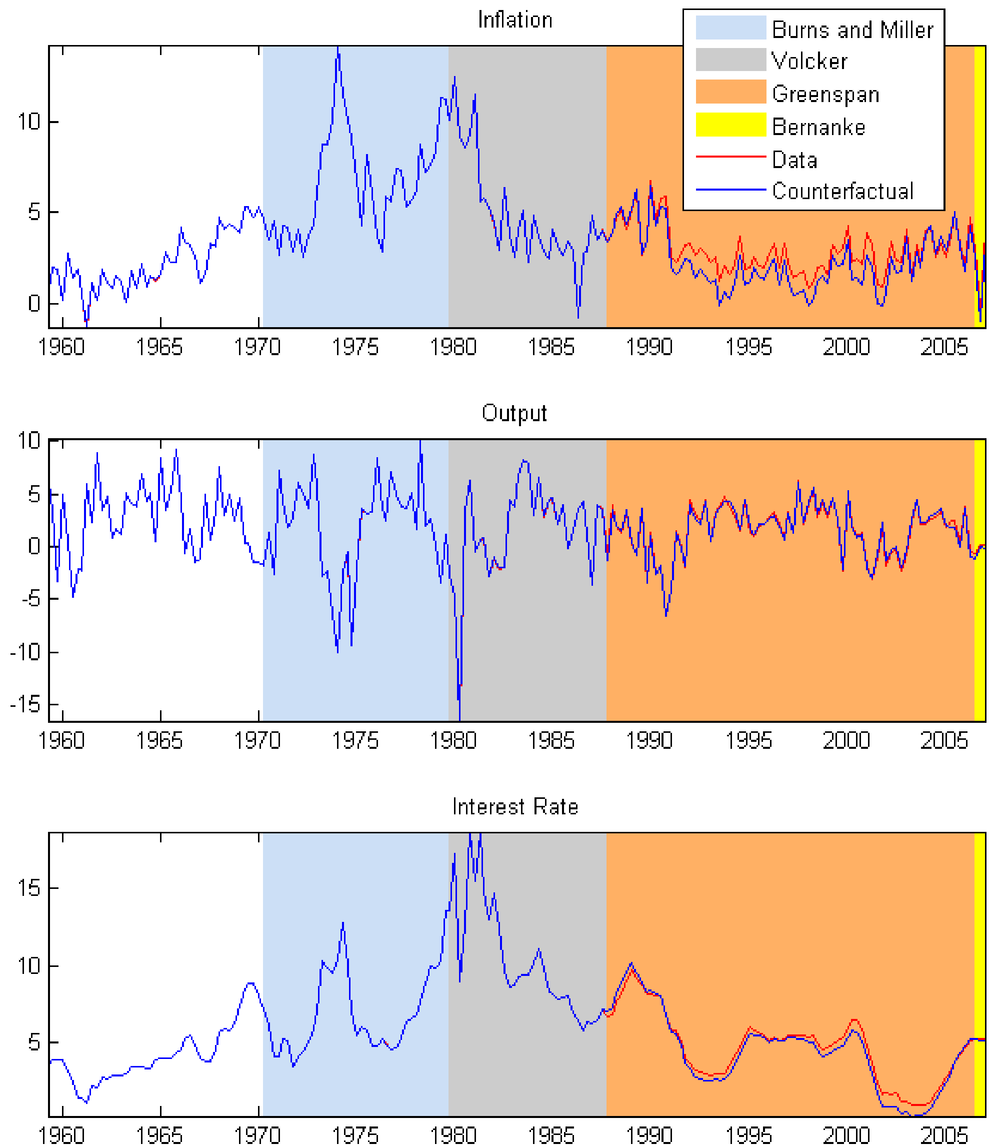


Figure 10.4: Burns-Miller during the Greenspan years

In figure 10.5, we plot the counterfactual of Burns-Miller extending their tenure to 1987. The results are very similar to the case in which we move Greenspan to the same period: slower disinflation and no improvement in output growth.

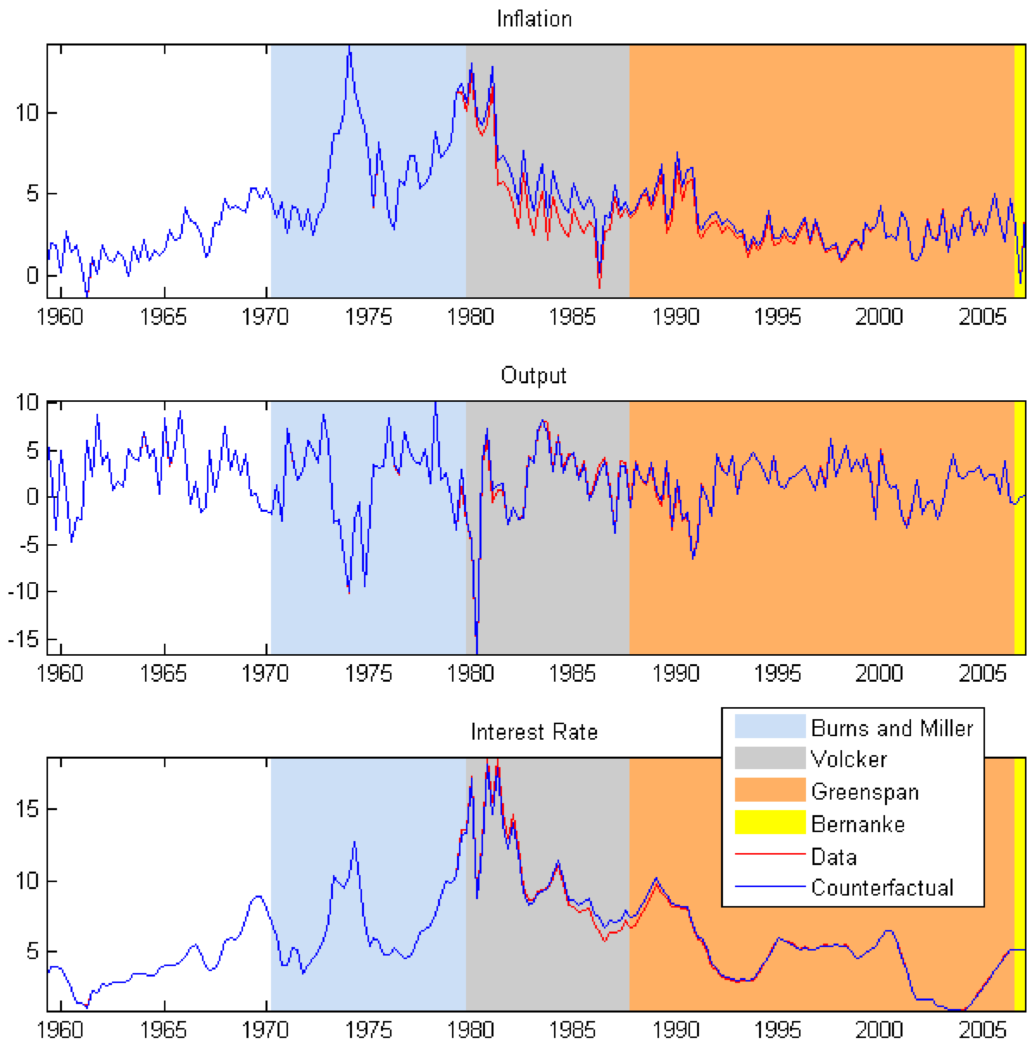


Figure 10.5: Burns-Miller during the Volcker years

A particularly interesting exercise is to check what would have happened if Reagan had decided to reappoint Volcker and not Greenspan. We plot these results in figure 10.6. The quick answer is: lower inflation and interest rates. Our estimates also suggest that Volcker would have reduced price increases with little cost to output. In fact, the difference in output growth is mainly caused by the period 2004-2007. Without those years, output growth under Volcker would have been higher than under Greenspan.

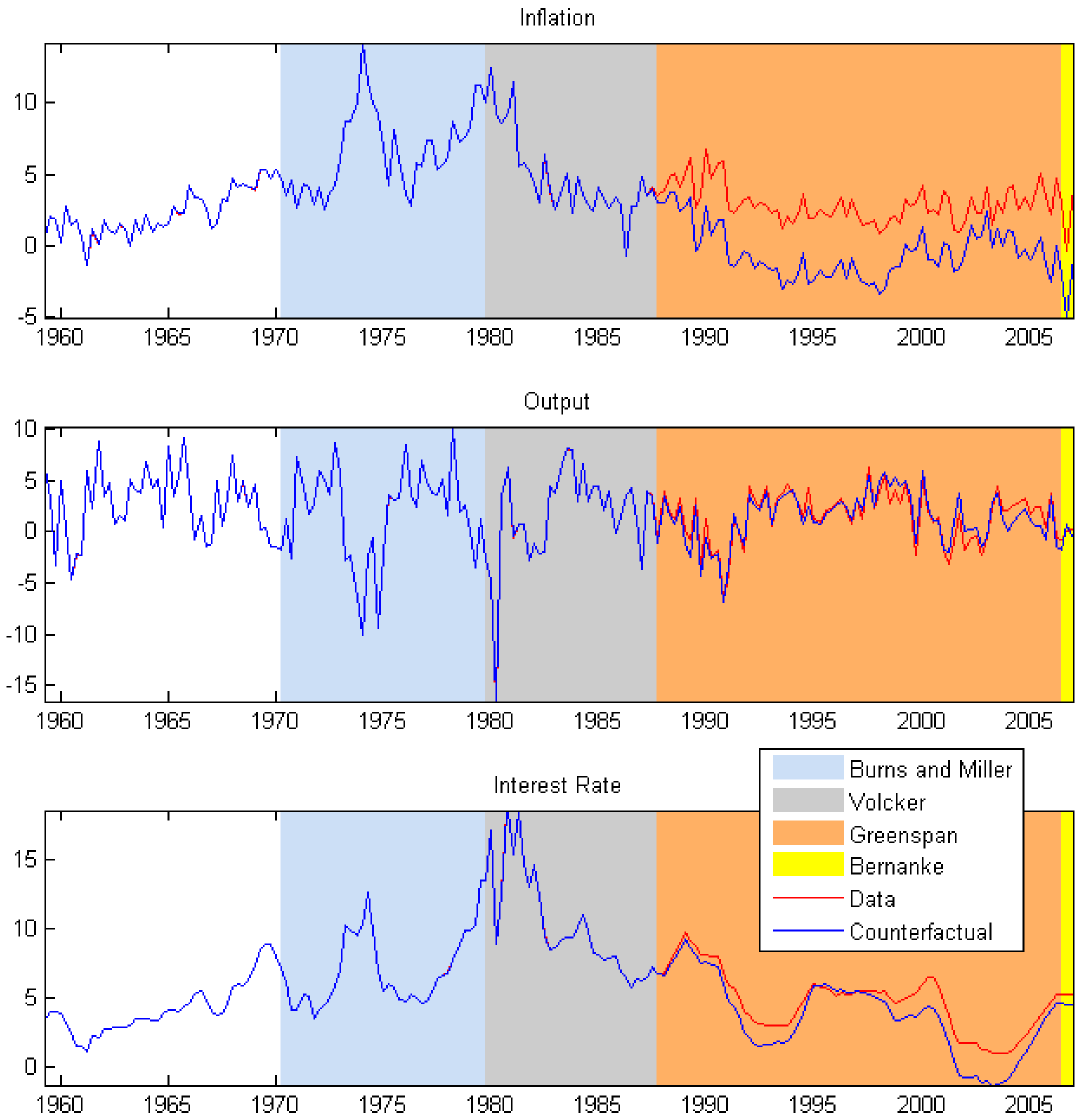


Figure 10.6: Volcker during the Greenspan years

This exercise has the problem that, according to the estimated policy rule, during 2004-2005, the Federal Reserve should have implemented a negative federal funds rate. Forgetting for a moment the practicality of schemes to push the nominal interest rate below zero, this evidence points out that an extension of the model should take the zero-lower bound on interest rates seriously. This, however, would preclude us from using a perturbation method and, given the current computational frontier, make it impossible to estimate such a model.

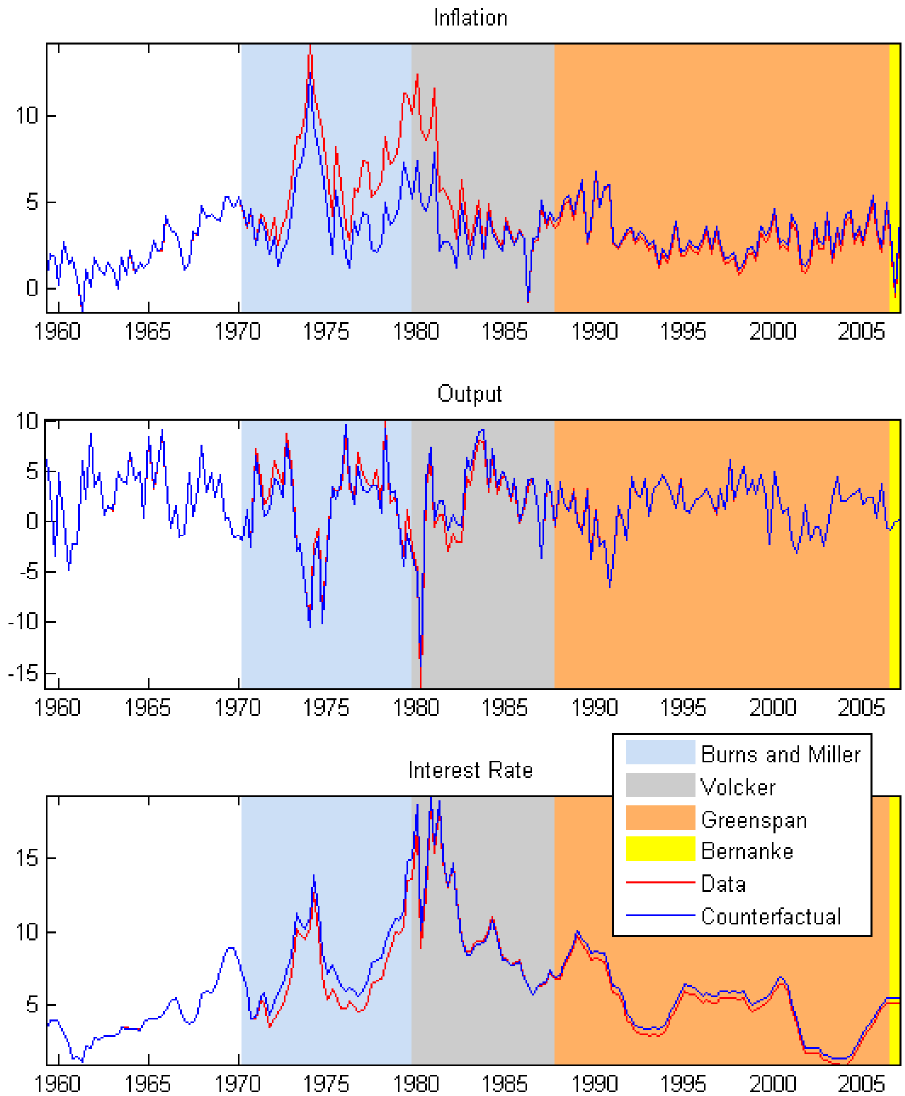


Figure 10.7: Volcker during the Burns-Miller years

Our final exercise is to plot, in figure 10.7, the counterfactual in which we move Volcker to the time of Burns-Miller. The main finding is that inflation would have been rather lower especially because the effects of the second oil shock would have been much more muted. This counterfactual is plausible: other countries, such as Germany, Switzerland, and Japan, that undertook a more aggressive monetary policy during the late 1970s were able to keep inflation under control at levels below 5 percent at an annual rate, while the U.S. had peaks of price increases over 10 percent.

## 11. Conclusion

In this paper, we have built and estimated a non-linear DSGE model with both stochastic volatility in the structural shocks that drive the economy and parameter drifting in the monetary policy rule. We have shown how such a rich model can be successfully taken to the data by using our characterization of a second-order approximation of the decision rules of a large class of DSGE models. We can see many future applications for these tools in all those situations in which being explicit about heteroscedastic shocks is crucial (see, among others, Bloom, Jaimovich, and Floetotto, 2008, or Fernández-Villaverde *et al.*, 2009).

With respect to our main empirical findings, a simple way to summarize them is to think about the recent monetary history of the U.S. as being characterized by three eras:

1. An era of small responses to inflation and large and volatile structural shocks: Burns and Miller, 1970-1979.

2. An era of large responses to inflation and large and volatile structural shocks: Volcker, 1979-1987.

3. An era of small responses to inflation and small and positive structural shocks: Greenspan, 1987-2006.

Like all empirical work, our approach suffers from several shortcomings, many of which we have discussed in the main body of the paper. The most important, in our opinion, is the limitation of how much we can learn from the data given our relatively short sample. In fact, we are reluctant to estimate even more complicated versions of our model precisely because of these limits of aggregate time series. A source of information that can complement our quantitative investigation is a historical narrative based on the Federal Reserve statements and documents. We have in mind the type of work pioneered by Romer and Romer (2004) or Hetzel (2008). If the changes in policy uncovered by our estimates did, in fact, occur, we should find telltale signs of them in the written record. We undertake this exercise in Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2010), where we use our estimates to read the recent monetary history of the U.S.

## 12. Appendix A: Theorem 1

Let us now prove theorem 1. In this theorem, we characterize the first- and second-order derivatives of the functions $h(\cdot)$ and $g(\cdot)$ evaluated at the non-stochastic steady state. We first show that the first partial derivatives of $h(\cdot)$ and $g(\cdot)$ with respect to any component of $\Sigma_{t-1}$, $\mathcal{U}_t$, or $\Lambda$ evaluated at the non-stochastic steady state is zero (or, in other words, that the the first order of the solution does not depend on volatility levels or shocks nor on the perturbation parameter). Second, it shows that, among many other results, the second partial derivative of $h(\cdot)$ and $g(\cdot)$ with respect to $u_{j,t}$ and any other state variable but $\varepsilon_{j,t}$ is also zero for any $j = \{1, \ldots, m\}$.

Before proceeding, note that we can write $Z_t$ as a function of $Z_{t-1}, \Sigma_{t-1}, \mathcal{E}_t$, and $\mathcal{U}_t$:

$$\mathcal{Z}_t = \varsigma\left(\mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t\right)$$

and that $\Sigma_t$ can be expressed as:

$$\Sigma_t = \vartheta \Sigma_{t-1} + \eta \mathcal{U}_t$$

where $\vartheta$ and $\eta$ are both $m \times m$ diagonal matrices with diagonal elements equal to $\vartheta_i$ and $\eta_i$ respectively. If we substitute the two functions (9) and (10) into (8) we get that:

$$F\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda\right) \equiv$$

$$\mathbb{E}_t f \left( \begin{array}{c} g\left(h\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda\right), \varsigma\left(\mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t\right), \vartheta\Sigma_{t-1} + \eta\mathcal{U}_t, \Lambda\mathcal{E}_{t+1}, \Lambda\mathcal{U}_{t+1}, \Lambda\right), \\ g\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda\right), h\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda\right), \mathcal{S}_t, \\ \varsigma\left(\varsigma\left(\mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t\right), \vartheta\Sigma_{t-1} + \eta\mathcal{U}_t, \Lambda\mathcal{E}_{t+1}, \Lambda\mathcal{U}_{t+1}\right), \varsigma\left(\mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t\right) \end{array} \right) = 0.$$

To ease reading, we divide the proof into four parts, the first dealing with the first derivatives and the next three dealing with the second derivatives.

**Proof, part 1.** The first part of the proof deals with the first derivatives of (9) and (10) that are equal to zero. In particular, we want to show that:

$$\left[h_{\Sigma_{t-1}}\right]_j^{i_1} = \left[g_{\Sigma_{t-1}}\right]_j^{i_2} = \left[h_{\mathcal{U}_t}\right]_j^{i_1} = \left[g_{\mathcal{U}_t}\right]_j^{i_2} = \left[h_\Lambda\right]^{i_1} = \left[g_\Lambda\right]^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$.

We show this result in three steps that basically repeat the same argument based on homogeneity of a system of linear equations:

1. We can write the derivative of the $i-th$ element of $F$ with respect to the $j-th$ element

of $\Sigma_{t-1}$ as:

$$\left[F_{\Sigma_{t-1}}\right]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\Sigma_{t-1}}\right]_j^{i_1} + \left[g_{\Sigma_{t-1}}\right]_j^{i_2} \vartheta_j\right) + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\Sigma_{t-1}}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\Sigma_{t-1}}\right]_j^{i_1} = 0$$

for $i \in \{1, \ldots, k+n+m\}$ and $j \in \{1, \ldots, m\}$. This is a homogeneous system on $\left[h_{\Sigma_{t-1}}\right]_j^{i_1}$ and $\left[g_{\Sigma_{t-1}}\right]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$. Thus:

$$\left[h_{\Sigma_{t-1}}\right]_j^{i_1} = \left[g_{\Sigma_{t-1}}\right]_j^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$.

2. We can write the derivative of the $i-th$ element of $F$ with respect to the $j-th$ element of $\mathcal{U}_t$ as:

$$\left[F_{\mathcal{U}_t}\right]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\mathcal{U}_t}\right]_j^{i_1} + \left[g_{\Sigma_{t-1}}\right]_j^{i_2} \eta_j\right) + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\mathcal{U}_t}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\mathcal{U}_t}\right]_j^{i_1} = 0$$

for $i \in \{1, \ldots, k+n+m\}$ and $j \in \{1, \ldots, m\}$. Since we have already shown that $\left[g_{\Sigma_{t-1}}\right]_j^{i_2} = 0$ for $i_2 \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, m\}$, this is a homogeneous system on $\left[h_{\mathcal{U}_t}\right]_j^{i_1}$ and $\left[g_{\mathcal{U}_t}\right]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$, Thus:

$$\left[h_{\mathcal{U}_t}\right]_j^{i_1} = \left[g_{\mathcal{U}_t}\right]_j^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$.

3. Finally, we can write the derivative of the $i-th$ element of $F$ with respect to $\Lambda$ as:

$$\left[F_{\Lambda}\right]^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\Lambda}\right]^{i_1} + \left[g_{\Lambda}\right]^{i_2}\right) + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\Lambda}\right]^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\Lambda}\right]^{i_1} = 0$$

for $i \in \{1, \ldots, k+n+m\}$. Since this is a homogeneous system on $\left[h_{\Lambda}\right]^{i_1}$ and $\left[g_{\Lambda}\right]^{i_2}$ for $i_1 \in \{1, \ldots, n\}$ and $i_2 \in \{1, \ldots, k\}$, we have that:

$$\left[h_{\Lambda}\right]^{i_1} = \left[g_{\Lambda}\right]^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$ and $i_2 \in \{1, \ldots, k\}$.

■

**Proof, part 2.** The second part of the proof deals with the cross-derivatives of (9) and (10) with respect to $\Lambda$ and any of $\mathcal{S}_t$, $\mathcal{Z}_{t-1}$, $\Sigma_{t-1}$, $\mathcal{E}_t$, or $\mathcal{U}_t$ and it shows that all of them are equal to zero. In particular, we want to show that:

$$[h_{\Lambda,\mathcal{S}_t}]_j^{i_1} = [g_{\Lambda,\mathcal{S}_t}]_j^{i_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,n\}$ and:

$$\left[h_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2} = \left[h_{\Lambda,\Sigma_{t-1}}\right]_j^{i_1} = \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} = [h_{\Lambda,\mathcal{E}_t}]_j^{i_1} = [g_{\Lambda,\mathcal{E}_t}]_j^{i_2} = [h_{\Lambda,\mathcal{U}_t}]_j^{i_1} = [g_{\Lambda,\mathcal{U}_t}]_j^{i_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,m\}$.

We show this result in five steps that, as in part 1 of the proof, exploit the homogeneity of a system of linear equations (and where we have already taken advantage of the terms that we know from part 1 of the proof that they are equal to zero and eliminate them from our expressions):

1. We consider the cross-derivative of the $i-th$ element of $F$ with respect to $\Lambda$ and the $j-th$ element of $\mathcal{S}_t$ :

   $$[F_{\Lambda,\mathcal{S}_t}]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left([g_{\mathcal{S}_t}]_{i_1}^{i_2} [h_{\Lambda,\mathcal{S}_t}]_j^{i_1} + [g_{\Lambda,\mathcal{S}_t}]_{i_1}^{i_2} [h_{\mathcal{S}_t}]_j^{i_1}\right) + [f_{\mathcal{Y}_t}]_{i_2}^i [g_{\Lambda,\mathcal{S}_t}]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i [h_{\Lambda,\mathcal{S}_t}]_j^{i_1} = 0$$

   for $i \in \{1,\ldots,k+n+m\}$ and $j \in \{1,\ldots,n\}$. This is a homogeneous system on $[h_{\Lambda,\mathcal{S}_t}]_j^{i_1}$ and $[g_{\Lambda,\mathcal{S}_t}]_j^{i_2}$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,n\}$. Thus:

   $$[h_{\Lambda,\mathcal{S}_t}]_j^{i_1} = [g_{\Lambda,\mathcal{S}_t}]_j^{i_2} = 0$$

   for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,n\}$.

2. We consider the cross-derivative of the $i-th$ element of $F$ with respect to $\Lambda$ and the $j-th$ element of $\mathcal{Z}_{t-1}$:

   $$\begin{aligned}\left[F_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i &\left([g_{\mathcal{S}_t}]_{i_1}^{i_2} \left[h_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_1} + [g_{\Lambda,\mathcal{S}_t}]_{i_1}^{i_2} \left[h_{\mathcal{Z}_{t-1}}\right]_j^{i_1} + \rho_j \left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2}\right) \\ &+ [f_{\mathcal{Y}_t}]_{i_2}^i \left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_1} = 0\end{aligned}$$

   for $i \in \{1,\ldots,k+n+m\}$ and $j \in \{1,\ldots,m\}$. Since $[g_{\Lambda,\mathcal{S}_t}]_j^{i_2} = 0$ for $i_2 \in \{1,\ldots,k\}$ and $j \in \{1,\ldots,n\}$, this is a homogeneous system on $\left[h_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_1}$ and $\left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2}$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,m\}$. Hence:

   $$\left[h_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2} = 0$$

   for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,m\}$.

3. We consider the cross-derivative of the $i-th$ element of $F$ with respect to $\Lambda$ and the $j-th$ element of $\Sigma_{t-1}$:

$$\begin{aligned}\left[F_{\Lambda,\Sigma_{t-1}}\right]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\Lambda,\Sigma_{t-1}}\right]_j^{i_1} + \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} \vartheta_j\right) \\ + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\Lambda,\Sigma_{t-1}}\right]_j^{i_1} = 0\end{aligned}$$

for $i \in \{1, \ldots, k+n+m\}$ and $j \in \{1, \ldots, m\}$. This is a homogeneous system on $\left[h_{\Lambda,\Sigma_{t-1}}\right]_j^{i_1}$ and $\left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$. Hence:

$$\left[h_{\Lambda,\Sigma_{t-1}}\right]_j^{i_1} = \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$.

4. We consider the cross-derivative of the $i-th$ element of $F$ with respect to $\Lambda$ and the $j-th$ element of $\mathcal{E}_t$:

$$\begin{aligned}\left[F_{\Lambda,\mathcal{E}_t}\right]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\Lambda,\mathcal{E}_t}\right]_j^{i_1} + \left[g_{\Lambda,\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\mathcal{E}_t}\right]_j^{i_1} + \exp^{\vartheta_j \sigma_{j,t-1}} \left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2}\right) \\ + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\Lambda,\mathcal{E}_t}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\Lambda,\mathcal{E}_t}\right]_j^{i_1} = 0\end{aligned}$$

for $i \in \{1, \ldots, k+n+m\}$ and $j \in \{1, \ldots, m\}$. Since $\left[g_{\Lambda,\mathcal{Z}_{t-1}}\right]_j^{i_2} = 0$ for $i_2 \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, m\}$ and $\left[g_{\Lambda,\mathcal{S}_t}\right]_j^{i_2} = 0$ for $i_2 \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, n\}$, this is a homogeneous system on $\left[h_{\Lambda,\mathcal{E}_t}\right]_j^{i_1}$ and $\left[g_{\Lambda,\mathcal{E}_t}\right]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$. Thus:

$$\left[h_{\Lambda,\mathcal{E}_t}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{E}_t}\right]_j^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$.

5. We consider the cross-derivative of the $i-th$ element of $F$ with respect to $\Lambda$ and the $j-th$ element of $\mathcal{U}_t$:

$$\left[F_{\Lambda,\mathcal{U}_t}\right]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\Lambda,\mathcal{U}_t}\right]_j^{i_1} + \eta_j \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2}\right) + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\Lambda,\mathcal{U}_t}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\Lambda,\mathcal{U}_t}\right]_j^{i_1} = 0$$

for $i \in \{1, \ldots, k+n+m\}$ and $j \in \{1, \ldots, m\}$. Since we have shown that $\left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} = 0$ for $i_2 \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, m\}$ , we have that the above system is a homogeneous system on $\left[h_{\Lambda,\mathcal{U}_t}\right]_j^{i_1}$ and $\left[g_{\Lambda,\mathcal{U}_t}\right]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$. Then:

$$\left[h_{\Lambda,\mathcal{U}_t}\right]_j^{i_1} = \left[g_{\Lambda,\mathcal{U}_t}\right]_j^{i_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$.

■

**Proof, part 3.** The third part of the proof deals with the cross-derivatives of (9) and (10) with respect to $\Sigma_{t-1}$ and any of $\mathcal{S}_t$, $\mathcal{Z}_{t-1}$, $\Sigma_{t-1}$, or $\mathcal{E}_t$ and it shows that all of them are equal to zero with one exception. In particular, we want to show that:

$$\left[h_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, $j_1 \in \{1,\ldots,n\}$, and $j_2 \in \{1,\ldots,m\}$,

$$\left[h_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = \left[h_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = \left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j_1, j_2 \in \{1,\ldots,m\}$, and

$$\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j_1, j_2 \in \{1,\ldots,m\}$ if $j_1 \neq j_2$.

We show this result in four steps (and where we have already taken advantage of the terms that we know from part 1 of the proof are equal to zero and eliminate them from our expressions):

1. We consider the cross-derivative of the $i-th$ element of $F$ with respect to the $j_1-th$ element of $\mathcal{S}_t$ and the $j_2-th$ element of $\Sigma_{t-1}$:

$$\begin{aligned}\left[F_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i}_{j_1,j_2} = \left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} + \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{S}_t}\right]^{i_1}_{j_1}\vartheta_{j_2}\right)\\ + \left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = 0\end{aligned}$$

   for $i \in \{1,\ldots,k+n+m\}$, $j_1 \in \{1,\ldots,n\}$, and $j_2 \in \{1,\ldots,m\}$. This is a homogeneous system on $\left[h_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, $j_1 \in \{1,\ldots,n\}$, and $j_2 \in \{1,\ldots,m\}$. Therefore:

$$\left[h_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_2,j_2} = 0$$

   for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, $j_1 \in \{1,\ldots,n\}$, and $j_2 \in \{1,\ldots,m\}$.

2. We consider the cross-derivative of the $i-th$ element of $F$ with respect to the $j_1-th$

element of $\mathcal{Z}_{t-1}$ and the $j_2-th$ element of $\Sigma_{t-1}$:

$$\left[F_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i}_{j_1,j_2}$$
$$= \left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}+\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{Z}_{t-1}}\right]^{i_1}_{j_1}\vartheta_{j_2}+\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\vartheta_{j_2}\rho_{j_1}\right)$$
$$+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}+\left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}=0$$

for $i\in\{1,\ldots,k+n+m\}$, and $j_1,j_2\in\{1,\ldots,m\}$. Since we just found that $\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}=0$ for $i_2\in\{1,\ldots,k\}$, $j_1\in\{1,\ldots,n\}$, and $j_2\in\{1,\ldots,m\}$, this is a homogeneous system on $\left[h_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{i_2,j_2}$ for $i_1\in\{1,\ldots,n\}$, $i_2\in\{1,\ldots,k\}$, and $j_1,j_2\in\{1,\ldots,m\}$. Therefore:

$$\left[h_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}=\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{i_2,j_2}=0$$

for $i_1\in\{1,\ldots,n\}$, $i_2\in\{1,\ldots,k\}$, and $j_1,j_2\in\{1,\ldots,m\}$.

3. We consider the cross-derivative of the $i-th$ element of $F$ with respect to the $j_1-th$ element of $\Sigma_{t-1}$ and the $j_2-th$ element of $\Sigma_{t-1}$:

$$\left[F_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i}_{j_1,j_2}=$$
$$\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}+\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\vartheta_{j_1}\vartheta_{j_2}\right)$$
$$+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}+\left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}=0$$

for $i\in\{1,\ldots,k+n+m\}$ and $j_1,j_2\in\{1,\ldots,m\}$. This is a homogeneous system on $\left[h_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}$ for $i_1\in\{1,\ldots,n\}$, $i_2\in\{1,\ldots,k\}$, and $j_1,j_2\in\{1,\ldots,m\}$, therefore:

$$\left[h_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}=\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}=0$$

for $i_1\in\{1,\ldots,n\}$, $i_2\in\{1,\ldots,k\}$, and $j_1,j_2\in\{1,\ldots,m\}$.

4. We consider the cross-derivative of the $i-th$ element of $F$ with respect to the $j_1-th$ element of $\mathcal{E}_t$ and the $j_2-th$ element of $\Sigma_{t-1}$ if $j_1\neq j_2$:

$$\left[F_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i}_{j_1,j_2}=$$
$$\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}+\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{E}_t}\right]^{i_1}_{j_1}\vartheta_{j_2}+\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\exp^{\vartheta_{j_1}\sigma_{j_1,t-1}}\vartheta_{j_2}\right)$$
$$+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}+\left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}=0$$

for $i \in \{1, \dots, k+n+m\}$ and $j_1, j_2 \in \{1, \dots, m\}$. Since we know that $\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j,j_2} = 0$ for $i_2 \in \{1, \dots, k\}$, $j \in \{1, \dots, n\}$, and $j_1, j_2 \in \{1, \dots, m\}$, this is a homogeneous system on $\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}$ for $i_1 \in \{1, \dots, n\}$, $i_2 \in \{1, \dots, k\}$, and $j_1, j_2 \in \{1, \dots, m\}$ if $j_1 \neq j_2$. Therefore:

$$\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = \left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_2} = 0$$

for $i_1 \in \{1, \dots, n\}$, $i_2 \in \{1, \dots, k\}$, and $j_1, j_2 \in \{1, \dots, m\}$ if $j_1 \neq j_2$.

Note that if $j_1 = j_2$, we have that:

$$\begin{gathered}\left[F_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i}_{j_1,j_1} = \left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2} * \\ * \left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_1} + \vartheta_{j_1}\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_1}\left[h_{\mathcal{E}_t}\right]^{i_1}_{j_1} + \left(\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_1} + \left[g_{\mathcal{Z}_{t-1}}\right]^{i_2}_{j_1}\right)\exp^{\vartheta_{j_1}\sigma_{j_1,t-1}}\vartheta_{j_1}\right) \\ + \left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_1} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_1} \\ + \left(\left[f_{\mathcal{Z}_t}\right]^{i}_{j_1} + \rho_{j_1}\left[f_{\mathcal{Z}_{t+1}}\right]^{i}_{j_1}\right)\exp^{\vartheta_{j_1}\sigma_{j_1,t-1}}\vartheta_{j_1} = 0\end{gathered}$$

and since $\left[f_{\mathcal{Z}_t}\right]^{i}_{j_1}$ and $\left[f_{\mathcal{Z}_{t+1}}\right]^{i}_{j_1}$ are different from zero in general for $i \in \{1, \dots, k+n+m\}$ and $j_1 \in \{1, \dots, m\}$ , we have that this system is not homogeneous and

$$\left[h_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_1} = \left[g_{\mathcal{E}_t,\Sigma_{t-1}}\right]^{i_1}_{j_1,j_1} \neq 0$$

for $i_1 \in \{1, \dots, n\}$, $i_2 \in \{1, \dots, k\}$, and $j_1 \in \{1, \dots, m\}$.

■

**Proof, part 4.** The fourth, and final, part of the proof deals with the cross-derivatives of (9) and (10) with respect to $\mathcal{U}_t$ and any of $\mathcal{S}_t$, $\mathcal{Z}_{t-1}$, $\Sigma_{t-1}$, $\mathcal{E}_t$, or $\mathcal{U}_t$ and it shows that all of them are equal to zero with one exception. In particular, we want to show that:

$$\left[h_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1, \dots, n\}$, $i_2 \in \{1, \dots, k\}$, $j_1 \in \{1, \dots, n\}$, and $j_2 \in \{1, \dots, m\}$,

$$\left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = \left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = \left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1, \dots, n\}$, $i_2 \in \{1, \dots, k\}$, and $j_1, j_2 \in \{1, \dots, m\}$, and

$$\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, $j_1, j_2 \in \{1,\dots,m\}$, and $j_1 \neq j_2$.

Again, we follow the same steps for each part of the result as before and use our previous findings regarding which terms are zero.

1. We consider the cross derivative of the $i-th$ element of $F$ with respect to the $j_1-th$ element of $S_t$ and the $j_2-th$ element of $\mathcal{U}_t$:

$$\begin{aligned}\left[F_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i}_{j_1,j_2} = \left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{S}_t}\right]^{i_1}_{j_1}\eta_{j_2}\right)\\ + \left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = 0\end{aligned}$$

for $i \in \{1,\dots,k+n+m\}$, $j_1 \in \{1,\dots,n\}$, and $j_2 \in \{1,\dots,m\}$. Since $\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$ for $i_2 \in \{1,\dots,k\}$, $j_1 \in \{1,\dots,n\}$, and $j_2 \in \{1,\dots,m\}$, this is a homogeneous system on $\left[h_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}$.Therefore:

$$\left[h_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, $j_1 \in \{1,\dots,n\}$, and $j_2 \in \{1,\dots,m\}$.

2. We consider the cross-derivative of the $i-th$ element of $F$ with respect to the $j_1-th$ element of $\mathcal{Z}_{t-1}$ and the $j_2-th$ element of $\mathcal{U}_t$:

$$\begin{aligned}\left[F_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i}_{j_1,j_2} =\\ \left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \eta_{j_2}\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{Z}_t}\right]^{i_1}_{j_1} + \rho_{j_1}\eta_{j_2}\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\right)\\ + \left[f_{\mathcal{Y}_t}\right]^{i}_{i_1}\left[g_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = 0\end{aligned}$$

for $i \in \{1,\dots,k+n+m\}$, and $j_1, j_2 \in \{1,\dots,m\}$. Since $\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j,j_2} = 0$ for $i_2 \in \{1,\dots,k\}$, $j \in \{1,\dots,n\}$, and $j_1, j_2 \in \{1,\dots,m\}$, this is a homogeneous system on $\left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}$ for $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, and $j_1, j_2 \in \{1,\dots,m\}$. Therefore:

$$\left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, and $j_1, j_2 \in \{1,\dots,m\}$.

3. We consider the cross-derivative of the $i-th$ element of $F$ with respect to the $j_1-th$

element of $\Sigma_{t-1}$ and the $j_2 - th$ element of $\mathcal{U}_t$:

$$\left[F_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i}_{j_1,j_2} =$$
$$\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\vartheta_{j_1}\eta_{j_2}\right)$$
$$+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_1}\left[g_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = 0$$

for $i \in \{1,\ldots,k+n+m\}$, and $j_1, j_2 \in \{1,\ldots,m\}$. Since $\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$ for $i_2 \in \{1,\ldots,k\}$, and $j_1, j_2 \in \{1,\ldots,m\}$, this is a homogeneous system on $\left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j_1, j_2 \in \{1,\ldots,m\}$. Therefore:

$$\left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, $j_1, j_2 \in \{1,\ldots,m\}$.

4. We consider the cross-derivative of the $i - th$ element of $F$ with respect to the $j_1 - th$ element of $\mathcal{U}_t$ and the $j_2 - th$ element of $\mathcal{U}_t$:

$$\left[F_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i}_{j_1,j_2} =$$
$$\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\eta_{j_1}\eta_{j_2}\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} + \left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}\right)$$
$$+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = 0$$

for $i \in \{1,\ldots,k+n+m\}$ and $j_1, j_2 \in \{1,\ldots,m\}$. Since $\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$ for $i_1 \in \{1,\ldots,k\}$ and $j_1, j_2 \in \{1,\ldots,m\}$, this is a homogeneous system on $\left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j_1, j_2 \in \{1,\ldots,m\}$. Therefore:

$$\left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = \left[g_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} = 0$$

for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j_1, j_2 \in \{1,\ldots,m\}$.

5. Finally, consider the cross-derivative of the $i - th$ element of $F$ with respect to the $j_1 - th$ element of $\mathcal{E}_t$ and the $j_2 - th$ element of $\mathcal{U}_t$ if $j_1 \neq j_2$:

$$\left[F_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i}_{j_1,j_2} =$$
$$\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \eta_{j_2}\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{E}_t}\right]^{i_1}_{j_1} + \left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\exp^{\vartheta_{j_1}\sigma_{j_1,t-1}}\eta_{j_2}\right)$$
$$+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} = 0$$

for $i \in \{1, \ldots, k+n+m\}$ and $j_1, j_2 \in \{1, \ldots, m\}$. Since $\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j,j_2} = 0$ for $i_2 \in \{1, \ldots, k\}$, $j \in \{1, \ldots, n\}$, and $j_1, j_2 \in \{1, \ldots, m\}$, this is a homogeneous system on $[h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_2}$ and $[g_{\mathcal{E}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j_1, j_2 \in \{1, \ldots, m\}$ if $j_1 \neq j_2$. Therefore:

$$[h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_2} = [g_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_2} = 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j_1, j_2 \in \{1, \ldots, m\}$ if $j_1 \neq j_2$.

Note that if $j_1 = j_2$, we have that:

$$\begin{gathered}
[F_{\mathcal{E}_t,\mathcal{U}_t}]^{i}_{j_1,j_1} = \\
\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2} \left(\exp^{\vartheta_{j_1}\sigma_{j_1,t-1}} \eta_{j_1} \left(\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_1} + \left[g_{\mathcal{Z}_{t-1}}\right]^{i_2}_{j_1}\right) + \eta_{j_1} \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_1} [h_{\mathcal{E}_t}]^{i_1}_{j_1} + [g_{\mathcal{S}_t}]^{i_2}_{i_1} [h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_1}\right) \\
+ [f_{\mathcal{Y}_t}]^{i}_{i_2} [g_{\mathcal{E}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_1} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1} [h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_1} \\
+ \exp^{\vartheta_{j_1}\sigma_{j_1,t-1}} \eta_{j_1} \left([f_{\mathcal{Z}_t}]^{i}_{j_1} + \rho_{j_1} \left[f_{\mathcal{Z}_{t+1}}\right]^{i}_{j_1}\right) = 0
\end{gathered}$$

and since $[f_{\mathcal{Z}_t}]^{i}_{j_1}$ and $\left[f_{\mathcal{Z}_{t+1}}\right]^{i}_{j_1}$ are different from zero in general for $i \in \{1, \ldots, k+n+m\}$ and $j_1 \in \{1, \ldots, m\}$ , we have that this system is not homogeneous and hence:

$$[h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_1} = [g_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_1} \neq 0$$

for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j_1 \in \{1, \ldots, m\}$. ■

## 13. Appendix B: Computation

In this appendix we provide some more details regarding the computation of the paper. We generate all the derivatives required by our second-order perturbation with `Mathematica` 6.0. In that way, we do not need to recompute the derivatives, the most time-intensive step, for each set of parameter values in our estimation. Once we have all the relevant derivatives, we export them automatically into Fortran files. This whole process takes about 3 hours.

Then, we compile the resulting files with the `Intel Fortran Compiler` version 10.1.025 with `IMSL`. Previous versions failed to compile our project because of the length of some of the expressions. Compilation takes about 18 hours. The project has 1798 files and occupies 2.33 Gbytes of memory.

The next step is, for given parameter values, to compute the first- and second-order approximation to the decision rules around the deterministic steady state using the analytic derivatives we found before. For this task, Fortran takes around 5 seconds. Once we have the solution, we approximate the likelihood using the particle filter with $10,000$ particles. This number delivered a good compromise between accuracy and time to compute the likelihood. The evaluation of one likelihood requires 22 seconds on a Dell server with 8 processors. Once we have the likelihood evaluation, we guess new parameter values and we start again. This means that drawing 5,000 times from the posterior (even forgetting about the initial search over a grid of parameter values) takes around 38 hours.

It is important to emphasize that the `Mathematica` and `Fortran` code were highly optimized in order to 1) keep the size of the project within reasonable dimensions (otherwise, the compiler cannot sparse the files and, even when it can, it delivers code that is too inefficient) and 2) provide a fast computation of the likelihood.

Perhaps the most important task in that optimization was the parallelization of the `Fortran` code using `OPENMP` as well as the compilation options: `OG` (global optimizations) and `Loop Unroll`. In addition, we tailored specialized code to perform the matrix multiplications required in the first- and second-order terms of our model solution.

Implementing corollary 1 requires the solution of a linear system of equations and the computation of a Jacobian. For our particular application, we found that the following sequence of `LAPACK` operations delivered the fastest solution:

1. `DGESV` (computes the solution to a real system of linear equations $\mathtt{A} * \mathtt{X} = \mathtt{B}$).

2. `DGETRI` (computes the inverse of a matrix using the LU factorization from the previous line).

3. `DGETRF` (helps to compute the determinant of the inverse from the previous line).

Without the parallelization and our optimized code, the solution of the model and evaluation of its likelihood take about 70 seconds.

With respect to the random-walk Metropolis-Hastings, we performed an intensive process of fine-tuning of the chain, both in terms of initial conditions as well as in terms of getting the right acceptance level. The only other important remark is to remember that as pointed out by McFadden (1989) and Pakes and Pollard (1989), we must keep the random numbers used for resampling in the particle filter constant across draws of the Markov chain. This is required to achieve stochastic equi-continuity, and even if this condition is not strictly necessary in a Bayesian framework, it reduces the numerical variance of the procedure, which was a serious concern for us given the complexity of our problem.

## 14. Appendix C: Construction of Data

When we estimate the model, we need to make the series provided by the National Income and Product Accounts (NIPA) consistent with the definition of variables in the theory. The main adjustment that we undertake is to express both real output and real gross investment in consumption units. Our DSGE model implies that there is a numeraire in terms of which all the other prices need to be quoted. We pick consumption as the numeraire. The NIPA, in comparison, uses an index of all prices to transform nominal GDP and investment into real values. In the presence of changing relative prices, such as the ones we have seen in the U.S. over the last several decades with the fall in the relative price of capital, NIPA's procedure biases the valuation of different series in real terms.

We map theory into the data by computing our own series of real output and real investment. To do so, we use the relative price of investment, defined as the ratio of an investment deflator and a deflator for consumption. The denominator is easily derived from the deflators of non-durable goods and services reported in the NIPA. It is more complicated to obtain the numerator because, historically, NIPA investment deflators were poorly constructed. Instead, we rely on the investment deflator computed by Fisher (2006). Since the series ends early in 2000Q4, we have extended it to 2007.Q1 by following Fisher's methodology.

For the real output per capita series, we first define nominal output as nominal consumption plus nominal gross investment. We define nominal consumption as the sum of personal consumption expenditures on non-durable goods and services. We define nominal gross investment as the sum of personal consumption expenditures on durable goods, private residential investment, and non-residential fixed investment. Per capita nominal output is equal to the ratio between our nominal output series and the civilian non-institutional population between 16 and 65. To obtain per capita values, we divide the previous series by the civilian non-institutional population between 16 and 65. Finally, real wages are defined as compensation per hour in the non-farm business sector divided by the CPI deflator.

## References


[1] Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2007). "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through." *Journal of International Economics* 72, 481-511.

[2] Aruoba, S.B., J. Fernández-Villaverde, and J. Rubio-Ramírez (2006). "Comparing Solution Methods for Dynamic Equilibrium Economies." *Journal of Economic Dynamics and Control* 30, 2477-2508.

[3] Benati, L. and P. Surico (2009). "VAR Analysis and the Great Moderation." *American Economic Review* 99, 1636-1652.

[4] Bianchi, F. (2009). "Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics." *Mimeo*, Duke University.

[5] Blanchard, O.J. and J. Simon (2001). "The Long and Large Decline in U.S. Output Volatility." *Brookings Papers on Economic Activity* 2001-1, 187-207.

[6] Bloom, N., N. Jaimovich, N., and M. Floetotto (2008). "Really Uncertain Business Cycles." *Mimeo*, Stanford University.

[7] Boivin, J. and M.P. Giannoni (2006). "Has Monetary Policy Become More Effective?" *Review of Economics and Statistics* 88, 445-462.

[8] Canova, F. (2009). "What Explains the Great Moderation in the U.S.? A Strutural Analysis." *Journal of the European Economic Association* 7, 697-721.

[9] Canova, F. and L. Gambetti (2009). "Structural Changes in the US Economy: Is There a Role for Monetary Policy?" *Journal of Economic Dynamics and Control* 33, 477-490.

[10] Christiano, L., M. Eichenbaum, and C.L. Evans (2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113, 1-45.

[11] Christoffel, K., G. Coenen, and A. Warne (2008) "The New Area-Wide Model of the Euro Area: A Micro-Founded Open-Economy Model for Forecasting and Policy Analysis." ECB Working Paper No. 944.

[12] Clarida, R., J. Galí, and M. Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory." *Quarterly Journal of Economics* 115, 147-180.

[13] Cogley, T. and T.J. Sargent (2002). "Evolving Post-World War II U.S. Inflation Dynamics." *NBER Macroeconomics Annual* 2001, 331-388.

[14] Cogley, T. and T.J. Sargent (2005). "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S." *Review of Economic Dynamics* 8, 262-302.

[15] Davig, T. and E.M. Leeper (2006). "Generalizing the Taylor Principle." *American Economic Review* 97, 607-635.

[16] DeJong, D.N., H. Dharmarajan, R. Liesenfeld, and J.F. Richard (2007). "An Efficient Filtering Approach to Likelihood Approximation for State-Space Representations." *Mimeo*, University of Pittsburgh.

[17] Del Moral P. and J. Jacod (2002). "The Monte-Carlo Method for Filtering with Discrete Time Observations. Central Limit Theorems." in T. J. Lyons and T. S. Salisbury (eds.), *Numerical Methods and Stochastics*. The Fields Institute Communications, American Mathematical Society.

[18] Dynan, K.E., D.W. Elmendorf, and D.E. Sichel (2006). "Can Financial Innovation Help to Explain the Reduced Volatility of Economic Activity?" *Journal of Monetary Economics* 53, 123-150.

[19] Edge, R., M.T. Kiley, and J.-P. Laforte (2007). "Documentation of the Research and Statistics Division's Estimated DSGE Model of the U.S. Economy: 2006 Version." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2007-53.

[20] Farmer, R.E., D.F. Waggoner, and T. Zha (2009). "Understanding Markov-Switching Rational Expectations Models." *Journal of Economic Theory* 144, 1849-1867.

[21] Fermanian, J.D. and B. Salanié (2004). "A Nonparametric Simulated Maximum Likelihood Estimation Method." *Econometric Theory* 20, 701-734.

[22] Fernández-Villaverde, J. and J. Rubio-Ramírez (2007). "Estimating Macroeconomic Models: A Likelihood Approach." *Review of Economic Studies* 74, 1059-1087.

[23] Fernandez-Villaverde, J. and J. Rubio-Ramirez (2008). "How Structural Are Structural Parameters?" *NBER Macroeconomics Annual 2007*, 83-137.

[24] Fernández-Villaverde, J., P. Guerrón-Quintana, and J. Rubio-Ramírez (2009). "The New Macroeconometrics: A Bayesian Approach." in A. O'Hagan and M. West (eds.) *Handbook of Applied Bayesian Analysis*, Oxford University Press.

[25] Fernández-Villaverde, J., P. Guerrón-Quintana, and J. Rubio-Ramírez (2010). "Reading the Recent Monetary History of the U.S., 1959-2007" *Mimeo*, University of Pennsylvania.

[26] Fernández-Villaverde, J., P. Guerrón-Quintana, J. Rubio-Ramírez, and M. Uribe (2009). "Risk Matters: The Real Effects of Volatility Shocks." *NBER Working Paper* 14875.

[27] Fiorentini, G., E. Sentana, and N. Shephard (2004). "Likelihood Estimation of Latent Generalised ARCH Structures." *Econometrica* 72, 1481-1517.

[28] Fisher, J., (2006). "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." *Journal of Political Economy* 114, 413-52.

[29] Geweke, J. (1998). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication." *Staff Report* 249, Federal Reserve Bank of Minneapolis.

[30] Goodfriend, M. and R. King (2007). "The Incredible Volcker Disinflation." *Journal of Monetary Economics* 52, 981-1015.

[31] Greenwood, J., Z. Herkowitz, and P. Krusell (1997). "Long-Run Implications of Investment-Specific Technological Change." *American Economic Review* 87, 342-362.

[32] Guerrón-Quintana, P. (2009a). "Money Demand Heterogeneity and the Great Moderation External Link." *Journal of Monetary Economics* 56, 255-266.

[33] Guerrón-Quintana, P. (2009b). "What You Match Does Matter: The Effects of Data on DSGE Estimation." *Journal of Applied Econometrics*, forthcoming.

[34] Hetzel, R.L. (2008). *The Monetary Policy of the Federal Reserve: a History*. Cambridge University Press.

[35] Hurwicz, L. (1962). "On the Structural Form of Interdependent Systems." Logic, Methodology and Philosophy of Science, Proceedings of the 1960 International Congress, 232-39.

[36] Jeffreys H. (1961). *The Theory of Probability* (3rd ed). Oxford University Press.

[37] Justiniano A. and G.E. Primiceri (2008). "The Time Varying Volatility of Macroeconomic Fluctuations." *American Economic Review* 98, 604-641.

[38] Kim, C. and C.R. Nelson (1998) "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle." *Review of Economics and Statistics* 81, 608-616.

[39] Kim, S., N. Shephard, and S. Chib (1998). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies* 65, 361-93.

[40] Klenow, P.J. and B.A. Malin (2009). "Microeconomic Evidence on Price-Setting." Chapter prepared for the *Handbook of Monetary Economic*s.

[41] Koop G., M. Hashem Pesaran, and S. Potter (1996). "Impulse Response Analysis in Nonlinear Multivariate Models." *Journal of Econometrics* 74, 199-147.

[42] Künsch, H.R. (2005). "Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis." *Annals of Statistics* 33, 1983-2021.

[43] Lubick, T. and F. Schorfheide (2004). "Testing for Indeterminacy: An Application to U.S. Monetary Policy." *American Economic Review* 94, 190-217.

[44] McConnell, M.M. and G. Pérez-Quirós (2000). "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review* 90, 1464-1476.

[45] McFadden, D.L. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57, 995-1026.

[46] Orphanides, A. (2002). "Monetary Policy Rules and the Great Inflation." *American Economic Review* 92, 115-120.

[47] Pakes, A. and D. Pollard (1989). "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57, 1027-1057.

[48] Primiceri, G.E. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy." *Review of Economic Studies* 72, 821-852.

[49] Ramey, V.A. and D.J. Vine (2006). "Declining Volatility in the U.S. Automobile Industry." *American Economic Review* 96, 1876-1889.

[50] Rogerson, R. and J. Wallenius (2009). "Micro and Macro Elasticities in a Life Cycle Model with Taxes." *Journal of Economic Theory*, forthcoming.

[51] Romer, C.D. and D. H. Romer (2004). "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 1055-1084.

[52] Schmitt-Grohé, S., and M. Uribe (2004). "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function." *Journal of Economic Dynamics and Control* 28, 755-775.

[53] Schreft, S. (1990). "Credit Controls: 1980." Federal Reserve Bank of Richmond *Economic Review* 1990, 6, 25-55.

[54] Schwarz, G.E. (1978). "Estimating the Dimension of a Model." *Annals of Statistics* 6, 461–464.

[55] Shephard, N. (2008). "Stochastic Volatility." *The New Palgrave Dictionary of Economics*. Palgrave MacMillan.

[56] Sims, C.A. and T. Zha (2006). "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96, 54-81.

[57] Smets, F. and R. Wouters (2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97, 586-606.

[58] Stock, J.H. and M.W. Watson (2002). "Has the Business Cycle Changed, and Why?" *NBER Macroeconomics Annual* 17, 159-218.

[59] Woodford, M.D. (2003). *Interest and Prices*. Princeton University Press.