NBER WORKING PAPER SERIES

ESTIMATING DYNAMIC EQUILIBRIUM MODELS WITH STOCHASTIC VOLATILITY

Jesus Fernandez-Villaverde
Pablo A. Guerrón-Quintana
Juan Rubio-Ramírez

Working Paper 18399
http://www.nber.org/papers/w18399

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2012

We thank the NSF for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating Dynamic Equilibrium Models with Stochastic Volatility
Jesus Fernandez-Villaverde, Pablo A. Guerrón-Quintana, and Juan Rubio-Ramírez
NBER Working Paper No. 18399
September 2012
JEL No. C1,E30

ABSTRACT

We propose a novel method to estimate dynamic equilibrium models with stochastic volatility. First, we characterize the properties of the solution to this class of models. Second, we take advantage of the results about the structure of the solution to build a sequential Monte Carlo algorithm to evaluate the likelihood function of the model. The approach, which exploits the profusion of shocks in stochastic volatility models, is versatile and computationally tractable even in large-scale models, such as those often employed by policy-making institutions. As an application, we use our algorithm and Bayesian methods to estimate a business cycle model of the U.S. economy with both stochastic volatility and parameter drifting in monetary policy. Our application shows the importance of stochastic volatility in accounting for the dynamics of the data.

Jesus Fernandez-Villaverde
University of Pennsylvania
160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104
and NBER
jesusfv@econ.upenn.edu

Pablo A. Guerrón-Quintana
Federal Reserve Bank of Philadelphia
pablo.guerron@phil.frb.org

Juan Rubio-Ramírez
Duke University
P.O. Box 90097
Durham, NC 27708
juan.rubio-ramirez@duke.edu

1. Introduction

Over the last few years, economists have uncovered the importance of time-varying volatility in accounting for the dynamics of U.S. aggregate data. Stock and Watson (2002) and Sims and Zha (2006) are some prominent examples of this line of research. As a response, a growing literature has developed to investigate the origins and consequences of time-varying volatility. A particularly successful branch of this literature uses dynamic equilibrium models to emphasize the role of volatility shocks (also known as uncertainty shocks, but in any case specified as stochastic volatility as in Shephard, 2008) in business cycle fluctuations. Among those we can highlight are Bloom *et al.* (2007), Fernández-Villaverde and Rubio-Ramírez (2007), Justiniano and Primiceri (2008), Bloom (2009), Fernández-Villaverde *et al.* (2010c), and Bloom *et al.* (2012). In these models, there are two types of shocks: structural shocks (shock to productivity, to preferences, to policy, etc.) and volatility shocks (shocks to the standard deviation of the innovations to the structural shocks).

To fulfill the promise in this literature, it is crucial to have a set of tools to take these dynamic equilibrium models with stochastic volatility to the data. Namely, we want to estimate, evaluate, and compare these models using a likelihood-based approach. However, the task is complicated by the inherent non-linearity that stochastic volatility generates. By construction, linearization is ill-equipped to handle time-varying volatility models because it yields certainty-equivalent policy functions. That is, volatility influences neither the agents' decision rules nor the laws of motion of the aggregate variables. Hence, to consider how volatility affects those, it becomes imperative to employ a non-linear solution method for the dynamics of the economy. Once we consider non-linear solution methods we are forced to use simulation-based estimators of the likelihood.

In principle, and beyond the computational burden, this could seem a mild challenge. For instance, one could be tempted to get around this problem by solving the model non-linearly with a fast algorithm and employing, for estimation, a particle filter to get an unbiased simulation-based estimator of the likelihood. The particle filter presented in Fernández-Villaverde and Rubio-Ramírez (2007) would be a natural candidate for such an undertaking. Unfortunately, that version of the particle filter relies, when estimating models with stochastic volatility, on the presence of linear measurement errors in observables. Otherwise, we would be forced to solve a large quadratic system of equations with multiple solutions, an endeavor for which there are no suitable algorithms. Although measurement errors are both plausible and empirically relevant, their presence complicates the interpretation of the results. Thus, it is desirable to consider a particle filter that departs from requiring measurement errors.

To accomplish this goal, we show how to write a particle filter that allows us to evaluate the likelihood of a dynamic equilibrium model with stochastic volatility by exploiting the structure of the second-order approximation to the decision rules that characterize the equilibrium of the economy without the need of measurement errors. Second-order approximations capture the implications of stochastic volatility and are convenient because they are accurate but not computationally expensive.

We achieve the objective in two steps. First, we characterize the second-order approximation to the decision rules of a rather general dynamic equilibrium model with stochastic volatility.[1] Second, we demonstrate how we can use this characterization to handily simulate the likelihood function of the model with a particle filter without measurement errors. The key is to show how the decision rules characterization reduces the quadratic problem associated with the evaluation of the approximated measurement density to a much simpler matrix inversion problem. This is a novel extension of sequential Monte Carlo methods that can be useful for other applications. After we have evaluated the likelihood, we can combine it with a prior and an Markov chain Monte Carlo (McMc) algorithm to draw from the posterior distribution (Flury and Shephard, 2011).

Our characterization of the second-order approximation to the decision rules should be of interest in itself even for researchers who are not evaluating the likelihood of their model but who are keen to understand how the model works. Among others, the results are useful to analyze the equilibrium of the model, to explore the shape of its impulse-response functions, or to calibrate it.

More concretely, we prove that:

1. The first-order approximation to the decision rules of the agents (or any other equilibrium object of interest) does not depend on volatility shocks and they are certainty equivalent.

2. The second-order approximation to the decision rules of the agents only depends on volatility shocks on terms where volatility is multiplied by the innovation to its own structural shock. For instance, if we have two structural shocks (a productivity and a preference shock) and

[1] These theorems generalize previous results derived by Schmitt-Grohé and Uribe (2004), for the homoscedastic shocks case to the time-varying case.

a volatility shock to each of them, the only non-zero term where the volatility shock to productivity appears is the one where the volatility shock to productivity multiplies the innovation to the productivity shock. Thus, of the terms in the second-order approximation that complicate the evaluation of the likelihood function, only a few are non-zero.

3. The perturbation parameter will only appear in a non-zero term where it is raised to a square. This term is a constant that corrects for precautionary behavior induced by risk.

As an application, we estimate a dynamic equilibrium model of the U.S. economy with stochastic volatility using the particle filter and Bayesian methods. The model, an otherwise standard business cycle model with nominal rigidities, incorporates not only stochastic volatility in the shocks that drive its dynamics but also parameter drifting in the parameters that control monetary policy. In that way, we include two of the main mechanisms that researchers have highlighted to account for the time-varying volatility of U.S. data: heteroscedastic shocks and parameter drifting. We include these together in the model because a "one-at-a-time" investigation is fraught with peril. If we allow only 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 monetary policy, an estimated model without drift but with stochastic volatility in one of the structural shocks may interpret the "true" drift as time-varying volatility in that shock. If, instead, we had time-varying volatility in technological shocks in the data, an estimated model without stochastic volatility and only parameter drifting may conclude, erroneously, that the parameters of monetary policy are changing. Lastly, we have a model that is as rich as the models employed by policy-making institutions in real life. While solving and estimating such a large model is a computational challenge, we wanted to demonstrate that our procedure is of practical use and to make our application a blueprint for the estimation of other dynamic equilibrium models.

Our main empirical findings are as follows. First, the posterior distribution of parameters puts most of its mass in areas that denote a fair amount of stochastic volatility, yet another proof of the importance of time-varying volatility in applied models. Second, a model comparison exercise indicates that, even after controlling for stochastic volatility, the data clearly prefer a specification where monetary policy changes over time. This finding should not be interpreted, though, as implying that volatility shocks did not play a role of their own. It means, instead, that a successful model of the U.S. economy requires the presence of both stochastic volatility and parameter drifting, a result that challenges the results of Sims and Zha (2006). Finally, we document the evolution of the structural shocks, of stochastic volatility, and the parameters of monetary policy. We emphasize the confluence, during the 1970s, of times of high volatility and weak responses to inflation, and, during the 1990s, of positive structural shocks and low volatility even if monetary policy was weaker than often argued. In the appendix, we thoroughly explore the model, including the construction of counterfactual histories of the U.S. data by varying some aspect of the model such as shutting down time-varying volatility or imposing alternative monetary policies.

An alternative to our stochastic volatility framework would be to work with Markov regime-switching models such as those of Bianchi (2009) or Farmer *et al.* (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 in our empirical section suggest that discrete jumps could 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 stylized representations of an economy.

Finally, even if the motivation for our approach and the application belong to macroeconomics, there is nothing specific about that field in the tools we present. One can think about the importance of estimating dynamic equilibrium models with stochastic volatility in many other fields, from finance (Bansal and Yaron, 2004, would be a prominent example) to industrial organization (for instance, an equilibrium model of industry dynamics where the demand and supply shocks have time-varying volatility), international trade (where innovations to the real exchange rates or to the country spreads are well known to have time-varying volatility, see Fernández-Villaverde *et al.*, 2010c), or many others.

The rest of the paper is organized as follows. Section 2 presents a generic dynamic equilibrium model with stochastic volatility to fix notation and discuss how to solve it. Section 3 explains the evaluation of the likelihood of the model. Section 4 introduces a business cycle model as an application of our procedure, takes it to the U.S. data, and comments on the main findings. Section 5 concludes. An extensive technical appendix includes details about the proofs of the main results in the paper, the model, the computation, and additional empirical exercises.

2. Dynamic Equilibrium Models with Stochastic Volatility

2.1. The Model

The set of equilibrium conditions of a wide class of dynamic equilibrium models can be written as

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

where $\mathbb{E}_t$ is the conditional expectation operator at time t, $\mathcal{Y}_t = (\mathcal{Y}_{1t}, \mathcal{Y}_{2t}, \ldots, \mathcal{Y}_{kt})'$ is the $k \times 1$ vector of observables at time t, $\mathcal{S}_t = (\mathcal{S}_{1t}, \mathcal{S}_{2t}, \ldots, \mathcal{S}_{nt})'$ is the $n \times 1$ vector of endogenous states at time t, $\mathcal{Z}_t = (\mathcal{Z}_{1t}, \mathcal{Z}_{2t}, \ldots, \mathcal{Z}_{mt})'$ is the $m \times 1$ vector of structural shocks at time t, f maps $\mathbb{R}^{2(k+n+m)}$ into $\mathbb{R}^{k+n+m}$, and γ is the $n_\gamma \times 1$ vector of parameters that describe preferences and technology. In the context of this paper, γ is also the vector of parameters to be estimated.

We will consider models where the structural shocks, $\mathcal{Z}_{it+1}$, follow a stochastic volatility process of the form

$$\mathcal{Z}_{it+1} = \rho_i \mathcal{Z}_{it} + \Lambda \sigma_i \sigma_{it+1} \varepsilon_{it+1} \tag{2}$$

for all $i \in \{1, \ldots, m\}$, where Λ is a perturbation parameter, σ_i is the mean volatility, and $\log \sigma_{it+1}$, the percentage deviation of the standard deviation of the innovations to the structural shocks with respect to its mean, evolves as

$$\log \sigma_{it+1} = \vartheta_i \log \sigma_{it} + \Lambda \left(1 - \vartheta_i^2\right)^{\frac{1}{2}} \eta_i u_{it+1} \tag{3}$$

for all $i \in \{1, \ldots, m\}$. The combination of levels in (2) and logs in (3) ensures a positive σ_{it+1}. We multiply the innovation in (3) by $\left(1 - \vartheta_i^2\right)^{\frac{1}{2}}$ to normalize its size by the persistence of σ_{it}. It will be clear momentarily why we specify (2) and (3) in terms of the perturbation parameter Λ. It is also convenient to write, for all $i \in \{1, \ldots, m\}$, the laws of motions for $\mathcal{Z}_{it}$ and $\log \sigma_{it}$

$$\mathcal{Z}_{it} = \rho_i \mathcal{Z}_{it-1} + \sigma_i \sigma_{it} \varepsilon_{it} \tag{4}$$

and

$$\log \sigma_{it} = \vartheta_i \log \sigma_{it-1} + \left(1 - \vartheta_i^2\right)^{\frac{1}{2}} \eta_i u_{it} \tag{5}$$

Note that the perturbation parameter Λ appears only in equations (2) and (3) but not in equations (4) and (5). This is because this parameter is used to eliminate, when we later determine the point around which to perform a higher-order approximation to the equilibrium dynamics of the model, any uncertainty about the future. Given the information set in equation (1), there is uncertainty about both $\mathcal{Z}_{it+1}$ and $\log \sigma_{it+1}$ for all $i \in \{1, \ldots, m\}$, but there is no uncertainty about either $\mathcal{Z}_{it}$ or $\log \sigma_{it}$ for any $i \in \{1, \ldots, m\}$.

Let $\Sigma_t = (\log \sigma_{1t}, \ldots, \log \sigma_{mt})'$ be the $m \times 1$ vector of volatility shocks, $\mathcal{E}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{mt})'$ the $m \times 1$ vector of innovations to the structural shocks, and $\mathcal{U}_t = (u_{1t}, \ldots, u_{mt})'$ the $m \times 1$ vector of innovations to the volatility shocks. We assume that $\mathcal{E}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathcal{U}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $\mathbf{0}$ is an $m \times 1$ vector of zeros and $\mathbf{I}$ is an $m \times m$ identity matrix. To ease notation, we assume that:

1. All structural shocks face volatility shocks and that the volatility shocks are uncorrelated. It is straightforward, yet cumbersome, to generalize the notation to other cases.

2. $\mathcal{E}_t$ and $\mathcal{U}_t$ are normally distributed. As we will see below, to implement our particle filter, we only need to be able to simulate $\mathcal{E}_t$ and evaluate the density of $\mathcal{U}_t$.

2.2. The Solution

Given equations (1)-(5), the solution to the model -if one exists- that embodies equilibrium dynamics (that is, agent's optimization and market clearing conditions) is characterized by a policy function describing the evolution of the endogenous state variables

$$\mathcal{S}_{t+1} = h\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda; \gamma\right), \tag{6}$$

and two policy functions describing the law of motion of the observables

$$\mathcal{Y}_t = g\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda; \gamma\right) \tag{7}$$

and

$$\mathcal{Y}_{t+1} = g\left(\mathcal{S}_{t+1}, \mathcal{Z}_t, \Sigma_t, \Lambda\mathcal{E}_{t+1}, \Lambda\mathcal{U}_{t+1}, \Lambda; \gamma\right), \tag{8}$$

together with equations (4) and (5) describing the laws of motion of the structural and volatility shocks. The policy functions h and g map $\mathbb{R}^{n+4m+1}$ into $\mathbb{R}^n$ and $\mathbb{R}^k$, respectively, and are indexed by the vector of parameters γ.

For our purposes, it is important to define the steady state of the model. Our assumptions about the stochastic processes imply that, in the steady state, $\mathcal{Z} = \mathbf{0}$ and $\Sigma = \mathbf{0}$. Given a vector of parameters γ and equations (1)-(8), the steady state of the model is a $k \times 1$ vector of observables $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_k)'$ and an $n \times 1$ vector of endogenous states $\mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_n)'$ such that

$$f\left(g\left(h\left(\mathcal{S}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, 0; \gamma\right), \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, 0; \gamma\right), g\left(\mathcal{S}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, 0; \gamma\right), h\left(\mathcal{S}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, 0; \gamma\right), \mathcal{S}, \mathbf{0}, \mathbf{0}; \gamma\right) = 0. \tag{9}$$

In addition, in the steady state, the following two relationships hold

$$\mathcal{S} = h\left(\mathcal{S}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, 0; \gamma\right), \tag{10}$$

and

$$\mathcal{Y} = g\left(\mathcal{S}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, 0; \gamma\right). \tag{11}$$

The perturbation parameter Λ plays a key role in defining our steady state. If $\Lambda = 0$, the model is in the steady state, since we eliminate any uncertainty about the future. If $\Lambda \neq 0$, the model is not in the steady state and conditions (9)-(11) do not hold. For instance, in general $\mathcal{S} \neq h\left(\mathcal{S}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \Lambda; \gamma\right)$ because of the precautionary behavior of agents.

2.3. The State-Space Representation

Given the solution to the model -the policy functions (6)-(8) together with equations (4) and (5)- we can concisely characterize the equilibrium dynamics of the model by its state-space representation written in terms of the transition and the observation equations.

The transition equation uses the policy function (6), together with equations (4) and (5), to describe the evolution of the states (endogenous states, structural shocks, volatility shocks, and their innovations) as a function of lag states, the perturbation parameter, and the vector of parameters

$$\mathbb{S}_{t+1} = \widetilde{h}\left(\mathbb{S}_t, \Lambda; \gamma\right) + \Xi \mathbb{W}_{t+1} \tag{12}$$

where $\mathbb{S}_t = \left(\mathcal{S}_t', \mathcal{Z}_{t-1}', \Sigma_{t-1}', \mathcal{E}_t', \mathcal{U}_t'\right)'$ is the $(n+4m) \times 1$ vector of the states and $\widetilde{h}$ maps $\mathbb{R}^{n+4m+1}$ into $\mathbb{R}^{n+4m}$. Also, $\mathbb{W}_{t+1} = \left(\mathcal{W}_{1t+1}', \mathcal{W}_{2t+1}'\right)'$ is a $2m \times 1$ vector of random variables, $\mathcal{W}_{1t+1}$ and $\mathcal{W}_{2t+1}$ are $m \times 1$ vectors with distributions $\mathcal{W}_{1t+1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$ and $\mathcal{W}_{2t+1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$, and Ξ is a $(n+4m) \times 2m$ matrix with the the top $n + 2m$ rows equal to zero and the bottom $2m \times 2m$ matrix equal to the identity matrix. Note that $\mathcal{W}_{1t+1}$ and $\mathcal{W}_{2t+1}$ share distributions with $\mathcal{E}_{t+1}$ and $\mathcal{U}_{t+1}$ respectively. If we were to change distributions for either $\mathcal{E}_{t+1}$ or $\mathcal{U}_{t+1}$ we would need to do the same with $\mathcal{W}_{1t+1}$ and $\mathcal{W}_{2t+1}$. Let us define $n_s = n + 4m$. The linearity of the second term of the right-hand side is a consequence of the autoregressive specification of the structural shocks and the evolution of their volatilities. However, through the function $\widetilde{h}$, we let those shocks and their volatilities affect $\mathcal{S}_{t+1}$ nonlinearly.

The measurement equation uses the policy function (7) to describe the relationship of the observables with the states, the perturbation parameter, and the vector of parameters

$$\mathcal{Y}_t = g\left(\mathbb{S}_t, \Lambda; \gamma\right). \tag{13}$$

2.4. Approximating the State-Space Representation

In general, when we deal with the class of dynamic equilibrium models with stochastic volatility, we do not have access to a closed-form solution, that is, the policy functions h and g cannot be found explicitly. Thus, we cannot build the state-space representation described by (12) and (13). Instead, we will approximate numerically the solution of model and use the result to generate an approximated state-space representation.

There are many different solution algorithms for dynamic equilibrium models. Among those, the perturbation method has emerged as a popular way (see Judd and Guu, 1997, Schmitt-Grohé and Uribe, 2004, and Aruoba *et al.*, 2006 among others) to obtain higher-order Taylor series approximations to the policy functions (6)-(7) together with equations (4) and (5) around the steady state. It is key to note that we also get the Taylor expansion of (4) because it is a non-linear function. Beyond being extremely fast for systems with a large number of state variables, perturbation offers high levels of accuracy even relatively far away from the perturbation point (Aruoba, *et al.*, 2006, and, for cases with stochastic volatility, Caldara *et al.*, 2012).

The perturbation method allows the researcher to approximate the solution to the model up to any order. Since we want to analyze models with volatility shocks, we must go beyond a first-order approximation. First-order approximations are certainty equivalent and hence, they are silent about stochastic volatility. As we well see in section 3.1, volatility shocks only appear starting in the second-order approximation. But since we want to evaluate the likelihood function, we need to stop at a second-order approximation. Because of dimensionality issues, higher-order approximation would make the evaluation of the likelihood function exceedingly challenging for current computers for models with a reasonable number of state variables.

Given a second-order approximation to the policy functions (6)-(7) together with equations (4) and (5) around the steady state, the approximated state-space representation can be written in terms of two equations: the approximated transition equation and the approximated measurement equation. The approximated transition equation is

$$\widehat{\mathbb{S}}_{t+1} = \begin{pmatrix} \Psi^1_{s,1}\widehat{\mathbb{S}}_t \\ \vdots \\ \Psi^1_{s,n_s}\widehat{\mathbb{S}}_t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \widehat{\mathbb{S}}'_t\Psi^2_{s,1}\widehat{\mathbb{S}}_t \\ \vdots \\ \widehat{\mathbb{S}}'_t\Psi^2_{s,n_s}\widehat{\mathbb{S}}_t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \Psi^{\Lambda}_{s,1} \\ \vdots \\ \Psi^{\Lambda}_{s,n_s} \end{pmatrix} + \Xi\mathbb{W}_{t+1} \tag{14}$$

where $\widehat{\mathbb{S}}_t = \mathbb{S}_t - \mathbb{S}$ is the $n_s \times 1$ vector of deviations of the states from their steady-state value and $\mathbb{S} = (\mathcal{S}', \mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}')'$. The approximated measurement equation is

$$\mathcal{Y}_t - \mathcal{Y} = \begin{pmatrix} \Psi^1_{y,1}\widehat{\mathbb{S}}_t \\ \vdots \\ \Psi^1_{y,k}\widehat{\mathbb{S}}_t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \widehat{\mathbb{S}}'_t\Psi^2_{y,1}\widehat{\mathbb{S}}_t \\ \vdots \\ \widehat{\mathbb{S}}'_t\Psi^2_{y,k}\widehat{\mathbb{S}}_t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \Psi^{\Lambda}_{y,1} \\ \vdots \\ \Psi^{\Lambda}_{y,k} \end{pmatrix}. \tag{15}$$

where $\mathcal{Y}$ is the steady-state value of $\mathcal{Y}_t$.

In these equations, $\Psi^1_{s,i}$ is a $1 \times n_s$ vector and $\Psi^2_{s,i}$ is an $n_s \times n_s$ matrix for $i = 1, \ldots, n_s$. The first term is the linear approximation to transition equation for the states, while the second term is the quadratic component of the second-order approximation. Similarly, $\Psi^1_{y,i}$ is a $1 \times n_s$ vector and $\Psi^2_{y,i}$ an $n_s \times n_s$ matrix for $i = 1, \ldots, k$. 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 approximation to the law of motion for the measurement equations. The term $\Psi^{\Lambda}_{s,i}$ is the constant that appears in the second-order perturbation and that is often interpreted as the correction for risk in the evolution of state $i = 1, \ldots, n_s$. Similarly, the term $\Psi^{\Lambda}_{y,i}$ is the constant correction for risk of observable $i = 1, \ldots, k$. All these vectors and matrices are non-linear functions of the vector of parameters γ. It is important to emphasize that we are not assuming the presence of any measurement error. We will return to this point in a few paragraphs.

3. Stochastic Volatility and Evaluation of the Likelihood

In this section, we explain how to evaluate the likelihood function of our class of dynamic equilibrium models with volatility shocks in equation (1) using the approximated transition and the measurement equations (14) and (15). If we allow $\mathbb{Y}_t$ to be the data counterpart of our observables $\mathcal{Y}_t$, $\mathbb{Y}^t = (\mathbb{Y}_1, \ldots, \mathbb{Y}_t)$ (with $\mathbb{Y}^0 = \{\varnothing\}$) to be their history up to time t, given a vector of parameters γ, we can write the likelihood of $\mathbb{Y}^T$ as

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

where

$$p\left(\mathcal{Y}_t = \mathbb{Y}_t | \mathbb{Y}^{t-1}; \gamma\right) = \int\int\int\int p\left(\mathcal{Y}_t = \mathbb{Y}_t | \mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t; \gamma\right) p\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t | \mathbb{Y}^{t-1}; \gamma\right) d\mathcal{S}_t d\mathcal{Z}_{t-1} d\Sigma_{t-1} d\mathcal{E}_t \tag{16}$$

for all $t \in \{2, \ldots, T\}$ and

$$p\left(\mathcal{Y}_1 = \mathbb{Y}_1; \gamma\right) = \int\int\int\int p\left(\mathcal{Y}_1 = \mathbb{Y}_1 | \mathcal{S}_1, \mathcal{Z}_0, \Sigma_0, \mathcal{E}_1; \gamma\right) p\left(\mathcal{S}_1, \mathcal{Z}_0, \Sigma_0, \mathcal{E}_1; \gamma\right) d\mathcal{S}_1 d\mathcal{Z}_0 d\Sigma_0 d\mathcal{E}_1. \tag{17}$$

Note that the $\mathcal{U}_t$'s do not show up in these two expressions. It will be momentarily clear why this comes directly from our procedure below to evaluate the likelihood.

Computing this likelihood is a difficult problem. Since we do not have analytic forms for the terms inside the integral, it cannot be evaluated exactly and deterministic numerical integration algorithms are too slow for practical use (we have four integrals per period over large dimensions). As a feasible alternative, we will show how to use a simple particle filter to obtain a simulation-based estimate of (16) for the class of dynamic equilibrium models with volatility shocks we described above. Künsch (2005) proves, under weak conditions, that the particle filter delivers a consistent estimator of the likelihood function and that a central limit theorem applies. A particle filter is a sequential simulation device for filtering of nonlinear and/or non-Gaussian space models (Pitt and Shephard, 1999, and Doucet *et al.*, 2001). In economics the particle filter has been used, among others, by Fernández-Villaverde and Rubio-Ramírez (2007).

As mentioned before, the particle filter has minimal requirements: the ability to evaluate the approximated measurement density associated with the approximated measurement equation, to simulate from the approximated dynamics of the state using the approximated transition equation, and to draw from the unconditional density of the states implied by the approximated transition equation. Usually, the first requirement is the hardest. These three requirements are formally described in the following assumption.

Assumption 1. *To implement the particle filter, we assume that:*

1. *We can evaluate the approximated measurement density*

$$p\left(\mathcal{Y}_t = \mathbb{Y}_t | \mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t; \gamma\right).$$

2. *We can simulate from the approximated transition equation*

$$\left(\mathcal{S}'_{t+1}, \mathcal{Z}'_t, \Sigma'_t, \mathcal{E}'_{t+1}\right)' | \left(\mathcal{S}'_t, \mathcal{Z}'_{t-1}, \Sigma'_{t-1}, \mathcal{E}'_t\right)', \mathcal{F}\left(\mathbb{Y}^t\right); \gamma$$

for all $t \in \{1, \ldots, T\}$, where $\mathcal{F}\left(\mathbb{Y}^t\right)$ is the filtration of $\mathbb{Y}^t$.

3. *We can draw from the unconditional distribution implied by the approximated transition equation*

$$p\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t; \gamma\right).$$

The second requirement asks for the filtration of $\mathbb{Y}^t$, $\mathcal{F}\left(\mathbb{Y}^t\right)$, because we will need to evaluate the volatility shocks (this will be clearer below). The last requirement can be easily implemented using the results in Santos and Peralta-Alva (2005). Given our assumption about $\mathcal{E}_{t+1}$ and the quadratic form of the approximated transition equation (14), the second requirement easily holds. A key novelty of this paper is that we show how the class of dynamic equilibrium models with volatility shocks considered here also satisfies the first requirement.

For our class of models, conditional on having N draws of $\left\{s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i\right\}_{i=1}^N$ from the density $p\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t | \mathbb{Y}^{t-1}; \gamma\right)$, each integral (16) can be consistently approximated by

$$p\left(\mathcal{Y}_t = \mathbb{Y}_t | \mathbb{Y}^{t-1}; \gamma\right) \simeq \frac{1}{N} \sum_{i=1}^{N} p\left(\mathcal{Y}_t = \mathbb{Y}_t | s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i; \gamma\right) \quad (18)$$

for all $t \in \{2, \ldots, T\}$. For $t = 1$, we need N draws from the density $p(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t; \gamma)$, so that the integral (17) can be consistently approximated by

$$p(\mathcal{Y}_1 = \mathbb{Y}_1; \gamma) \simeq \frac{1}{N} \sum_{i=1}^{N} p\left(\mathcal{Y}_1 = \mathbb{Y}_1 | s_1^i, z_0^i, \sigma_0^i, \varepsilon_1^i; \gamma\right). \tag{19}$$

We know that we can make the draws because requirement 3 in assumption 1 holds.

In our framework, checking whether requirement 1 in assumption 1 holds means checking whether we can evaluate the approximated measurement density

$$p\left(\mathcal{Y}_t = \mathbb{Y}_t | s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i; \gamma\right) \tag{20}$$

for each draw in $\left\{s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i\right\}_{i=1}^{N}$ for all $t \in \{1, \ldots, T\}$. This evaluation step is crucial, not only because it is required to compute (18) and (19) but also because, as we will explain momentarily, our particle filter needs to evaluate (20) to resample from $p(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t | \mathbb{Y}^{t-1}; \gamma)$ and get draws from $p(\mathcal{S}_{t+1}, \mathcal{Z}_t, \Sigma_t, \mathcal{E}_{t+1} | \mathbb{Y}^t; \gamma)$ in order to obtain $\left\{s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i\right\}_{i=1}^{N}$ for all $t \in \{2, \ldots, T\}$ in a recursive way.

To check how we can evaluate the approximated measurement density (20), we rewrite (15) in terms of draws $\left(s_t^{i\prime}, z_{t-1}^{i\prime}, \sigma_{t-1}^{i\prime}, \varepsilon_t^{i\prime}\right)'$ and $\mathbb{Y}_t$, instead of $\left(\mathcal{S}_t', \mathcal{Z}_{t-1}', \Sigma_{t-1}', \mathcal{E}_t'\right)'$ and $\mathcal{Y}_t$. Thus, the approximated measurement equation (15) can be rewritten as

$$\mathbb{Y}_t - \mathcal{Y} = \begin{pmatrix} \Psi_{y,1}^1 \widehat{\mathbb{S}}_t^i \\ \vdots \\ \Psi_{y,k}^1 \widehat{\mathbb{S}}_t^i \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \widehat{\mathbb{S}}_t^{i\prime} \Psi_{y,1}^2 \widehat{\mathbb{S}}_t^i \\ \vdots \\ \widehat{\mathbb{S}}_t^{i\prime} \Psi_{y,k}^2 \widehat{\mathbb{S}}_t^i \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \Psi_{y,1}^{\Lambda} \\ \vdots \\ \Psi_{y,k}^{\Lambda} \end{pmatrix} \tag{21}$$

where $\widehat{\mathbb{S}}_t^i = \mathbb{S}_t^i - \mathbb{S}$ and $\mathbb{S}_t^i = \left(s_t^{i\prime}, z_{t-1}^{i\prime}, \sigma_{t-1}^{i\prime}, \varepsilon_t^{i\prime}, \mathcal{U}_t'\right)'$. The new approximated measurement equation (21) implies that evaluating the approximated measurement density (20) involves solving the system of quadratic equations

$$\mathbb{Y}_t - \mathcal{Y} - \begin{pmatrix} \Psi_{y,1}^1 \widehat{\mathbb{S}}_t^i \\ \vdots \\ \Psi_{y,k}^1 \widehat{\mathbb{S}}_t^i \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \widehat{\mathbb{S}}_t^{i\prime} \Psi_{y,1}^2 \widehat{\mathbb{S}}_t^i \\ \vdots \\ \widehat{\mathbb{S}}_t^{i\prime} \Psi_{y,k}^2 \widehat{\mathbb{S}}_t^i \end{pmatrix} - \frac{1}{2} \begin{pmatrix} \Psi_{y,1}^{\Lambda} \\ \vdots \\ \Psi_{y,k}^{\Lambda} \end{pmatrix} = 0 \tag{22}$$

for $\mathcal{U}_t$ given $\mathbb{Y}_t, s_t^i, z_{t-1}^i, \sigma_{t-1}^i$, and ε_t^i.

Solving this system is non-trivial. Since the system is quadratic, we may have either none or several different solutions. In fact, it is even hard to know how many solutions there are. But even if we knew the number of solutions, we are not aware of any accurate and efficient method to solve quadratic problems that finds all the solutions. This difficulty seemingly prevents us from achieving our goal of evaluating the likelihood function.

A solution would be to introduce, as Fernández-Villaverde and Rubio-Ramírez (2007) did, a $k \times 1$ vector of linear measurement errors and solve for those instead of $\mathcal{U}_t$. In this case the system would have a unique, easy to find solution. However, there are three reasons, in increasing order of importance, why this solution is not satisfactory:

1. Although measurement errors are both plausible and empirically relevant, their presence complicates the interpretation of any empirical results. In particular, we are interested in measuring how heteroscedastic structural shocks help in accounting for the data and measurement errors can confound us.

2. The absence of measurement errors will help us to illustrate below how dynamic equilibrium models with volatility shocks have a profusion of shocks that we can exploit in our estimation.

3. As we will also show below, volatility shocks would enter linearly (conditional on the draw) in the system equations. Since, by definition, linear measurement errors would also enter in the same fashion, it would be hard to identify one apart from the others.

Our alternative in this paper is to realize that considering stochastic volatility converts the above-described quadratic system into a linear one. Hence, if a rank condition is satisfied, the system (22) has only one solution and the solution can be found by simply inverting a matrix. Thus, requirement 1 in assumption 1 holds and we can use the particle filter. The core of the argument is to note that, when volatility shocks are considered, the policy functions share a peculiar pattern that we can exploit.

3.1. Structure of the Solution

Our first step will characterize the first- and second-order derivatives of the policy functions h and g evaluated at the steady state. Then, we will describe an interesting pattern in these derivatives. The second step will take advantage of the pattern to show that, when the number of volatility shocks equals the number of observables, our quadratic system becomes a linear one. This characterization is important both for estimation and, more generally, for the analysis of perturbation solutions to dynamic equilibrium models with stochastic volatility.

3.1.1. First- and Second-order Derivatives of the Policy Functions

Let us begin with the characterization of the first- and second-order derivatives of the policy functions. The following theorem shows that the first-order derivatives of h and g with respect to any component of $\mathcal{U}_t$ and Σ_{t-1} evaluated at the steady state are zero; that is, volatility shocks and their innovations do not affect the linear component of the optimal decision rule of the agents. The same occurs with the perturbation parameter Λ. A similar result has been established by Schmitt-Grohé and Uribe (2004) for the homoscedastic shocks case.

Theorem 2. *Let us denote $[\Upsilon_\omega]^i_j$ as the derivative of the $i-th$ element of generic function Υ with respect to the $j-th$ element of generic variable ω evaluated at the non-stochastic steady state (where we drop this index if ω is unidimensional). Then, for the dynamic equilibrium model specified in equation (1), we have that $\left[h_{\Sigma_{t-1}}\right]^{i_1}_j = \left[g_{\Sigma_{t-1}}\right]^{i_2}_j = [h_{\mathcal{U}_t}]^{i_1}_j = [g_{\mathcal{U}_t}]^{i_2}_j = [h_\Lambda]^{i_1} = [g_\Lambda]^{i_2} = 0$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,m\}$.*

Proof. See appendix 6.1.1. ■

The second theorem shows, among other things, that the second partial derivatives of h and g with respect to either $\log \sigma_{it}$ or $u_{i,t}$ and any other variable but $\varepsilon_{i,t}$ is also zero for any $i \in \{1,\ldots,m\}$.

Theorem 3. *Furthermore, if we denote $[\Upsilon_{\omega\xi}]^i_{j_1,j_2}$ as the derivative of the $i-th$ element of generic function Υ with respect to the j_1-th element of generic variable ω and the j_2-th element of generic variable ξ evaluated at the non-stochastic steady state (where again we drop the index for unidimensional variables), we have that $[h_{\Lambda,\mathcal{S}_t}]^{i_1}_j = [g_{\Lambda,\mathcal{S}_t}]^{i_2}_j = 0$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,n\}$,*

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

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

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

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

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

and

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

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

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

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

Proof. See appendix 6.1.2. ■

Since the statement of theorem 3 is long and involved, we clarify it with a table in which we characterize the second derivatives of the functions h and g with respect to the different variables $(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t, \Lambda)$. This pattern is both interesting and useful for our purposes.

Table 3.1: Second Derivatives

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

The way to read table 3.1 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 and g with respect to $\mathcal{S}_t$ and $\mathcal{Z}_{t-1}$ are different from zero (the table is upper triangular because, given the symmetry of second derivatives, we do not need to report those entries). Similarly, entry (3,5), $\Sigma_{t-1}\mathcal{U}_t = 0$ tells us that the cross-derivatives of h and g with respect to Σ_{t-1} and $\mathcal{U}_t$ are all zero.[2] Entries (3,4) and (4,5)

[2]For ease of exposition, in table 3.1 we are not being explicit about the dimensions of the matrices: it is a purely qualititive description of the relevant derivatives.

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 3.1 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, Λ, 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, Σ_{t-1}, appear with coefficients different from zero only in the term $\Sigma_{t-1}\mathcal{E}_t$ where they are multiplied by the innovation to its own structural shock $\mathcal{E}_t$. 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 $\mathcal{E}_t$. Hence, of the terms that complicate the evaluation of the approximated measurement density, only the ones associated with $[h_{\mathcal{E}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_1}$ and $[g_{\mathcal{E}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_1}$ are non-zero.

3.1.2. Evaluating the Likelihood Using the Particle Filter

The second step is to use theorems 2 and 3 to show that the system (22) is linear and it has only one solution. Corollary 4 shows that the pattern described in table 3.1. has an important, yet rather direct implication for the structure of the approximated measurement equation (21).

Corollary 4. *The approximated measurement equation (21) can be written as*

$$\mathbb{Y}_t - \mathcal{Y} = \begin{pmatrix} \widetilde{\Psi}^1_{y,1}\widehat{\widetilde{\mathbb{S}}}^i_t \\ \vdots \\ \widetilde{\Psi}^1_{y,k}\widehat{\widetilde{\mathbb{S}}}^i_t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \widehat{\widetilde{\mathbb{S}}}^{i\prime}_t \widetilde{\Psi}^{2,1}_{y,1}\widehat{\widetilde{\mathbb{S}}}^i_t \\ \vdots \\ \widehat{\widetilde{\mathbb{S}}}^{i\prime}_t \widetilde{\Psi}^{2,1}_{y,k}\widehat{\widetilde{\mathbb{S}}}^i_t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \Psi^{\Lambda}_{y,1} \\ \vdots \\ \Psi^{\Lambda}_{y,k} \end{pmatrix} + \begin{pmatrix} \varepsilon^{i\prime}_t \widetilde{\Psi}^{2,2}_{y,1} \\ \vdots \\ \varepsilon^{i\prime}_t \widetilde{\Psi}^{2,2}_{y,k} \end{pmatrix}\sigma^i_{t-1} + \begin{pmatrix} \varepsilon^{i\prime}_t \widetilde{\Psi}^{2,3}_{y,1} \\ \vdots \\ \varepsilon^{i\prime}_t \widetilde{\Psi}^{2,3}_{y,k} \end{pmatrix}\mathcal{U}_t$$

where $\widetilde{\mathbb{S}}^i_t = \left(s^{i\prime}_t, z^{i\prime}_{t-1}, \varepsilon^{i\prime}_t\right)'$ *is the* $(n+2m)\times 1$ *vector of the simulated states without the stochastic volatility components (that is, endogenous states, structural shocks, and their innovations),* $\widetilde{\mathbb{S}} = (\mathcal{S}', \mathbf{0}', \mathbf{0}')'$ *is its steady state, and* $\widehat{\widetilde{\mathbb{S}}}^i_t = \widetilde{\mathbb{S}}^i_t - \widetilde{\mathbb{S}}$ *is its deviation from the steady state. Let us define* $\widetilde{n}_s = n + 2m$*. The matrix* $\widetilde{\Psi}^1_{y,j}$ *is a* $1\times\widetilde{n}_s$ *vector that represents the first-order component of the second-order approximation to the law of motion for the measurement equation as a function of*

$\widehat{\mathbb{S}}_t$, where $\widetilde{\mathbb{S}}_t = \left(\mathcal{S}'_t, \mathcal{Z}'_{t-1}, \mathcal{E}'_t\right)'$ and $\widehat{\widetilde{\mathbb{S}}}_t = \widetilde{\mathbb{S}}_t - \widetilde{\mathbb{S}}$, for $j = 1, \ldots, k$. The matrix $\widetilde{\Psi}^{2,1}_{y,j}$ is an $\tilde{n}_s \times \tilde{n}_s$ matrix that represents the second-order component of the second-order approximation to the law of motion for the measurement equation as a function of $\widehat{\mathbb{S}}_t$ for $j = 1, \ldots, k$. The matrix $\widetilde{\Psi}^{2,2}_{y,j}$ is an $m \times m$ matrix that represents the second-order component of the second-order approximation to the law of motion for the measurement equation as a function of $\mathcal{E}_t$ and Σ_{t-1} for $j = 1, \ldots, k$. The matrix $\widetilde{\Psi}^{2,3}_{y,j}$ is an $m \times m$ matrix that represents the second-order term of the approximated law of motion for the measurement equation as a function of $\mathcal{E}_t$ and $\mathcal{U}_t$ for $j = 1, \ldots, k$.*

We are now ready to show that the system (22) is linear, and if a rank condition is satisfied, it has only one solution. Define

$$\mathbb{A}\left(\mathbb{Y}_t, s^i_t, z^i_{t-1}, \sigma^i_{t-1}, \varepsilon^i_t; \gamma\right)$$

$$\equiv \mathbb{Y}_t - \mathcal{Y} - \begin{pmatrix} \widetilde{\Psi}^1_{y,1}\widehat{\widetilde{\mathbb{S}}}^i_t \\ \vdots \\ \widetilde{\Psi}^1_{y,k}\widehat{\widetilde{\mathbb{S}}}^i_t \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \widehat{\widetilde{\mathbb{S}}}^{i\prime}_t \widetilde{\Psi}^{2,1}_{y,1}\widehat{\widetilde{\mathbb{S}}}^i_t \\ \vdots \\ \widehat{\widetilde{\mathbb{S}}}^{i\prime}_t \widetilde{\Psi}^{2,1}_{y,k}\widehat{\widetilde{\mathbb{S}}}^i_t \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \Psi^{\Lambda}_{y,1} \\ \vdots \\ \Psi^{\Lambda}_{y,k} \end{pmatrix} - \begin{pmatrix} \varepsilon^{i\prime}_t \widetilde{\Psi}^{2,2}_{y,1} \\ \vdots \\ \varepsilon^{i\prime}_t \widetilde{\Psi}^{2,2}_{y,k} \end{pmatrix}\sigma^i_{t-1}$$

and

$$\mathbb{B}\left(\varepsilon^i_t; \gamma\right) \equiv \left(\left(\varepsilon^{i\prime}_t \widetilde{\Psi}^{2,3}_{y,1}\right)' \ \ldots \ \left(\varepsilon^{i\prime}_t \widetilde{\Psi}^{2,3}_{y,k}\right)'\right)'.$$

Let $k = m$, then $\mathbb{B}\left(\varepsilon^i_t; \gamma\right)$ is a $k \times k$ matrix. If $\mathbb{B}\left(\varepsilon^i_t; \gamma\right)$ is full rank, the solution to the system (22) can be written as

$$\mathcal{U}_t = \mathbb{B}^{-1}\left(\varepsilon^i_t; \gamma\right)\mathbb{A}\left(\mathbb{Y}_t, s^i_t, z^i_{t-1}, \sigma^i_{t-1}, \varepsilon^i_t; \gamma\right).$$

The next theorem shows how to use this solution to evaluate the approximated measurement density.

Theorem 5. *Let $k = m$, then $\mathbb{B}\left(\varepsilon^i_t; \gamma\right)$ is a $k \times k$ matrix. If $\mathbb{B}\left(\varepsilon^i_t; \gamma\right)$ is full rank, the approximated measurement density can be written as*

$$p\left(\mathcal{Y}_t = \mathbb{Y}_t | s^i_t, z^i_{t-1}, \sigma^i_{t-1}, \varepsilon^i_t; \gamma\right) = \det \mathbb{B}^{-1}\left(\varepsilon^i_t; \gamma\right) p\left(\mathcal{U}_t = \mathbb{B}^{-1}\left(\varepsilon^i_t; \gamma\right)\mathbb{A}\left(\mathbb{Y}_t, s^i_t, z^i_{t-1}, \sigma^i_{t-1}, \varepsilon^i_t; \gamma\right)\right) \tag{23}$$

for each draw in $\left\{s^i_t, z^i_{t-1}, \sigma^i_{t-1}, \varepsilon^i_t\right\}_{i=1}^N$ for all $t \in \{1, \ldots, T\}$, which can be evaluated given that we know $\mathbb{B}^{-1}\left(\varepsilon^i_t; \gamma\right)$, $\mathbb{A}\left(\mathbb{Y}_t, s^i_t, z^i_{t-1}, \sigma^i_{t-1}, \varepsilon^i_t; \gamma\right)$, and the distribution of $\mathcal{U}_t$.

Proof. The theorem is a straightforward application of the change of variables theorem. ■

Theorem 5 shows that requirement 1 in assumption 1 holds. Consequently, we can use a variation of the particle filter adapted for our class of dynamic equilibrium models with volatility shocks. We are requiring $B(\varepsilon_t^i;\gamma)$ to be full rank. When can $B(\varepsilon_t^i;\gamma)$ not be full rank? $B(\varepsilon_t^i;\gamma)$ would fail to have full rank when the impact of volatility innovations are identical across several elements of $\mathbb{Y}_t$. This would mean that the observables lack enough information to tell volatility shocks apart. If this is the case, a new set of observables would have to be chosen to estimate the model.

Note also that theorem 5 assumes $k = m$. Given the notation in section 2.1, this means that the number of structural shocks equals to the number of observables. This is not always necessary. What the theorem needs is that the number of volatility shocks equals the number of observables. Since, to simplify notation, we have assumed that all structural shocks face volatility shocks, the number of structural shocks equals the number of observables. As mentioned in section 2.1, we could have structural shocks that do not face volatility shocks (this will be the case in our application to follow). In that case, we could have more structural shocks than observables. From the theorem, it is also clear that if we used $\mathcal{E}_t$ rather than $\mathcal{U}_t$ to compute the approximated measurement equation, we would have to solve a quadratic system, a challenging task.

An outline of the algorithm in pseudo-code is:

Step 0: `Set` $t \rightsquigarrow 1$. `Sample` N `values` $\left\{s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i\right\}_{i=1}^N$ `from` $p\left(\mathcal{S}_t, \mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t;\gamma\right)$.

Step 1: `Compute`

$$p\left(\mathcal{Y}_t = \mathbb{Y}_t | \mathbb{Y}^{t-1};\gamma\right) \simeq \frac{1}{N}\sum_{i=1}^{N} \det \mathbb{B}^{-1}\left(\varepsilon_t^i;\gamma\right) p\left(\mathcal{U}_t = \mathbb{B}^{-1}\left(\varepsilon_t^i;\gamma\right)\mathbb{A}\left(\mathbb{Y}_t, s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i;\gamma\right)\right)$$

`using expression (23) and the importance weights for each draw`

$$q_t^i = \frac{\det \mathbb{B}^{-1}\left(\varepsilon_t^i;\gamma\right) p\left(\mathcal{U}_t = \mathbb{B}^{-1}\left(\varepsilon_t^i;\gamma\right)\mathbb{A}\left(\mathbb{Y}_t, s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i;\gamma\right)\right)}{\sum_{i=1}^{N}\det \mathbb{B}^{-1}\left(\varepsilon_t^i;\gamma\right) p\left(\mathcal{U}_t = \mathbb{B}^{-1}\left(\varepsilon_t^i;\gamma\right)\mathbb{A}\left(\mathbb{Y}_t, s_t^i, z_{t-1}^i, \sigma_{t-1}^i, \varepsilon_t^i;\gamma\right)\right)}$$

Step 2: `Sample` N `times with replacement from` $\left\{s_{t|t-1}^i, z_{t-1|t-1}^i, \sigma_{t-1|t-1}^i, \varepsilon_{t|t-1}^i\right\}_{i=1}^N$ `and probability` $\{q_t^i\}_{i=1}^N$. `This delivers` $\left\{s_{t|t}^i, z_{t-1|t}^i, \sigma_{t-1|t}^i, \varepsilon_{t|t}^i\right\}_{i=1}^N$.

Step 3: `Simulate` $\{s^i_{t+1}, z^i_t, \sigma^i_t, \varepsilon^i_{t+1}\}_{i=1}^N$ `from the approximated transition equation`

$$\left(\mathcal{S}'_{t+1}, \mathcal{Z}'_t, \Sigma'_t, \mathcal{E}'_{t+1}\right)' \mid \left(s^{i\prime}_{t|t}, z^{i\prime}_{t-1|t}, \sigma^{i\prime}_{t-1|t}, \varepsilon^{i\prime}_{t|t}\right)', \mathcal{F}\left(\mathbb{Y}^t\right); \gamma.$$

Step 4: `If` $t < T$, `set` $t \rightsquigarrow t+1$ `and go to step 1. Otherwise stop.`

Once we have evaluated the likelihood, we can nest this algorithm either with a McMc to perform Bayesian inference (as done in our application; see Flury and Shephard, 2011, for technical details) or with some optimization algorithm to undertake maximum likelihood estimation (as done, in a model without volatility shocks, by Van Binsbergen *et al.*, 2012). In this last case, care must be taken to keep the random numbers used in the particle filter constant among iterations to achieve equi-continuity and to use an optimization algorithm that does not rely on derivatives, as the particle filter implies an evaluation of the likelihood function that is not differentiable.

4. An Application: A Business Cycle Model with Stochastic Volatility

As an illustration of our procedure, we estimate a business cycle model of the U.S. economy with nominal rigidities and stochastic volatility. We will show: 1) how we can characterize posterior distributions of the parameters of interest and 2) how we can recover and analyze the smoothed structural and volatility shocks. Those are two of the most relevant exercises in terms of the estimation of dynamic equilibrium models. Once the estimation has been undertaken, there are many other exercises that can be done with the empirical results. For instance, in the appendix, we include several of those such as: 1) finding the impulse response functions (IRFs) of the model; 2) evaluating the fit of the model in comparison with some alternatives; and 3) undertaking some counterfactuals and run alternative histories of the evolution of the U.S. economy.

Before doing so, though, we need to motivate our choice of application. The model that we present has many strengths but also important weaknesses. Suffice it to say here that since this model is the base of much applied policy analysis by academics and policy-making institutions, it is a natural example for this paper. Thus, given that the model is well known, our presentation will be brief. We will depart only along two dimensions from the standard model: we will have stochastic volatility in the shocks that drive the dynamics of the economy (the object of interest in this paper) and parameter drifting in the monetary policy rule. In that way, the likelihood has the chance of picking between two of the alternatives that the literature has highlighted to account for the well-documented time-varying volatility of the U.S. aggregate time series, either a reduced volatility of the shocks that hit the economy (as in Sims and Zha, 2006) or a different monetary policy (Clarida *et al.*, 2000), which makes the application of interest in itself.

4.1. Households

The economy is populated by a continuum of households indexed by j and preferences:

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

which are 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 expectations operator, β is the discount factor, h controls habit persistence, ϑ is the inverse of the Frisch labor supply elasticity, d_t is a shifter to intertemporal preference that follows $\log d_t = \rho_d \log d_{t-1} + \sigma_d \sigma_{dt} \varepsilon_{dt}$ where $\varepsilon_{dt} \sim \mathcal{N}(0, 1)$, and φ_t is a labor supply shifter that evolves as $\log \varphi_t = \rho_\varphi \log \varphi_{t-1} + \sigma_\varphi \sigma_{\varphi t} \varepsilon_{\varphi t}$ where $\varepsilon_{\varphi t} \sim \mathcal{N}(0, 1)$. The two preference shocks are common to all households.

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

$$\log \sigma_{dt} = \rho_{\sigma_d} \log \sigma_{dt-1} + \left(1 - \rho_{\sigma_d}^2\right)^{\frac{1}{2}} \eta_d u_{dt} \text{ where } u_{dt} \sim \mathcal{N}(0, 1)$$

and

$$\log \sigma_{\varphi t} = \rho_{\sigma_\varphi} \log \sigma_{\varphi t-1} + \left(1 - \rho_{\sigma_\varphi}^2\right)^{\frac{1}{2}} \eta_\varphi u_{\varphi t} \text{ where } u_{\varphi t} \sim \mathcal{N}(0, 1).$$

This parsimonious specification introduces only four new parameters, ρ_{σ_d}, ρ_{σ_φ}, η_d, and η_φ, while being surprisingly powerful in capturing important features of the data (Shephard, 2008). All the shocks and innovations throughout the model are observed by the agents when they are realized. Agents have, as well, rational expectations about how they evolve over time.

We can interpret the shocks to preferences and to their volatility as reflecting the random evolution of more complicated phenomena. For example, stochastic volatility may appear as the consequence of changing demographics. 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 $\sigma_{\varphi t}$).

Although we assume complete financial markets, to ease notation we drop the Arrow securities implied by that assumption from the budget constraints (they are in net zero supply at the aggregate level anyway). 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:

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

where x_t is investment, w_{jt} is the real wage, r_t is the real rental price of capital, $u_{jt} > 0$ is the rate of use of capital, $\mu_t^{-1}\Phi\left[u_{jt}\right]$ is the cost of using capital at rate u_{jt} in terms of the final good, μ_t is an investment-specific technological level, T_t is a lump-sum transfer, and F_t is firms' profits. We specify that $\Phi[u] = \Phi_1(u-1) + \frac{\Phi_2}{2}(u-1)^2$, a form that satisfies that $\Phi[1] = 0$, $\Phi'[\cdot] = 0$, and $\Phi''[\cdot] > 0$. 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 $\Phi_1 = \Phi'[1] = \tilde{r}$ where $\tilde{r}$ is the (rescaled) steady-state rental price of capital (determined by the other parameters in the model). This leaves us with only one free parameter, Φ_2.

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

$$k_{jt} = (1-\delta)\,k_{jt-1} + \mu_t\left(1 - \frac{\kappa}{2}\left(\frac{x_{jt}}{x_{jt-1}} - \Lambda_x\right)^2\right)x_{jt}$$

where δ is the depreciation rate and κ is an investment adjustment cost parameter. This function is written in deviations with respect to the balanced growth rate of investment, Λ_x. The investment-specific technology level μ_t, follows a random walk in logs, $\log\mu_t = \Lambda_\mu + \log\mu_{t-1} + \sigma_\mu\sigma_{\mu t}\varepsilon_{\mu t}$ with $\varepsilon_{\mu t} \sim \mathcal{N}(0,1)$ and where Λ_μ is the drift of the process and $\varepsilon_{\mu t}$ is the innovation to its growth rate. The standard deviation $\sigma_{\mu t}$ also evolves as:

$$\log\sigma_{\mu t} = \rho_{\sigma_\mu}\log\sigma_{\mu t-1} + \left(1-\rho_{\sigma_\mu}^2\right)^{\frac{1}{2}}\eta_\mu u_{\mu t} \text{ where } u_{\mu t} \sim \mathcal{N}(0,1).$$

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

Each household j supplies a 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 wage w_t. The labor packer is perfectly competitive and it takes all wages as given. Households set their wages with a Calvo pricing mechanism. At the start of every period, a randomly selected fraction $1-\theta_w$ of households can reoptimize their wages (where, by a law of large numbers, individual probabilities and aggregate fractions are equal). All other households index their wages given past inflation with an indexation parameter $\chi_w \in [0,1]$.

4.2. Firms

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

$$y_t = \left(\int_0^1 y_{it}^{\frac{\varepsilon-1}{\varepsilon}} di\right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{24}$$

where ε is the elasticity of substitution. The final good producer is perfectly competitive and minimizes its costs subject to the production function (24) and taking as given all intermediate goods prices p_{ti} and the final good price p_t.

Each intermediate good is produced by a monopolistic competitor with technology $y_{it} = A_t k_{it-1}^{\alpha} \left(l_{it}^d\right)^{1-\alpha}$, 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, and A_t is neutral productivity. Productivity evolves as $\log A_t = \Lambda_A + \log A_{t-1} + \sigma_A \sigma_{At} \varepsilon_{At}$ where Λ_A is the drift of the process and $\varepsilon_{At} \sim \mathcal{N}(0,1)$ is the innovation to its growth rate. The time-varying standard deviation of this innovation follows:

$$\log \sigma_{At} = \rho_{\sigma_A} \log \sigma_{At-1} + \left(1-\rho_{\sigma_A}^2\right)^{\frac{1}{2}} \eta_A u_{At} \text{ where } u_{At} \sim \mathcal{N}(0,1).$$

Intermediate good producers meet the quantity demanded by the final good producer by renting l_{it}^d and k_{it-1} at prices w_t and r_t. Given their demand function, these producers set prices to maximize profits. However, when they do so, they follow a Calvo pricing scheme. 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 $\chi \in [0,1]$.

4.3. The Monetary Authority

The model is closed with a monetary authority that sets the nominal interest rates by following a modified Taylor rule:

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R}\right)^{\gamma_R} \left(\left(\frac{\Pi_t}{\Pi}\right)^{\gamma_\Pi \gamma_{\Pi,t}} \left(\frac{y_t^d}{y_{t-1}^d} / \exp\left(\Lambda_{y^d}\right)\right)^{\gamma_y \gamma_{y,t}}\right)^{1-\gamma_R} \xi_t. \tag{25}$$

The first term on the right-hand side represents a desire for interest rate smoothing, expressed in terms of R, the balanced growth path nominal interest rate. The second term, an "inflation gap," responds to the deviation of inflation from its balanced growth path level Π. The third term is a "growth gap": the ratio between the growth rate of the economy and Λ_{y^d}, the balanced path gross growth rate of y_t^d, where y_t^d is aggregate demand (defined precisely in appendix 6.2). The last term is the monetary policy shock, where $\log \xi_t = \sigma_\xi \sigma_{\xi,t} \varepsilon_{\xi t}$ with an innovation $\varepsilon_{\xi t} \sim \mathcal{N}(0,1)$ and a time-varying standard deviation, $\sigma_{\xi,t}$, that follows an autoregressive process

$$\log \sigma_{\xi t} = \rho_{\sigma_\xi} \log \sigma_{\xi t-1} + \left(1 - \rho_{\sigma_\xi}^2\right)^{\frac{1}{2}} \eta_\xi u_{\xi,t} \text{ where } u_{\xi,t} \sim \mathcal{N}(0,1).$$

In this policy rule, we have two drifting parameters: the responses of the monetary authority to the inflation gap and the growth gap. The parameters drift over time as:

$$\log \gamma_{\Pi t} = \rho_{\gamma_\Pi} \log \gamma_{\Pi t-1} + \sigma_\pi \varepsilon_{\pi t} \text{ and } \log \gamma_{yt} = \rho_{\gamma_y} \log \gamma_{yt-1} + \sigma_y \varepsilon_{yt} \text{ where } \varepsilon_{\pi t}, \varepsilon_{yt} \sim \mathcal{N}(0,1).$$

For simplicity, the volatility of the innovation to this processes is fixed over time. The agents perfectly observe the changes in monetary policy parameters.

4.4. Equilibrium and Solution

We can characterize the equilibrium of the model in appendix 6.2. This equilibrium conditions are non-stationary because we have two unit roots in the processes for technology. However, we circumvent this problem by rescaling the model using the variable $z_t = A_t^{\frac{1}{1-\alpha}} \mu_t^{\frac{\alpha}{1-\alpha}}$ in the form: $\widetilde{k}_t = \frac{k_t}{z_t \mu_t}$, $\widetilde{c}_t = \frac{c_t}{z_t}$, $\widetilde{x}_t = \frac{x_t}{z_t}$, $\widetilde{y}_t = \frac{y_t}{z_t}$, $\widetilde{w}_t = \frac{w_t}{z_t}$, $\widetilde{r}_t = \mu_t r_t$, $\widetilde{A}_t = \frac{A_t}{\exp(\Lambda_A) A_{t-1}}$, and $\widetilde{\mu}_t = \frac{\mu_t}{\exp(\Lambda_\mu) \mu_{t-1}}$. In the notation of section 2.1 we have that:

1. The states of the (rescaled) economy are

$$\mathcal{S}_t = \left(\log \widetilde{k}_{t-1}, \log \widetilde{c}_{t-1}, \log \widetilde{x}_{t-1}, \log \widetilde{y}_{t-1}, \log v^p_{t-1}, \log v^w_{t-1}, \log \widetilde{w}_{t-1}, \log R_{t-1}, \log \Pi_{t-1}\right)'.$$

2. The structural shocks are $\mathcal{Z}_t = \left(\log d_t, \log \varphi_t, \log \mu_t, \log \widetilde{A}_t, \log \xi_t, \log \gamma_{\Pi t}, \log \gamma_{yt}\right)'$. The parameter drifts are handled as structural shocks in the state-space representation.

3. The volatility shocks are $\Sigma_t = (\log \sigma_{dt}, \log \sigma_{\varphi t}, \log \sigma_{\mu t}, \log \sigma_{At}, \log \sigma_{\xi t})'$, where the last two zeros correspond to the processes for parameter drifting, which have constant volatilities (see also in vector $\mathcal{U}_t$ below). Here, we can see that we do not need as many volatility shocks (5 of them) as structural shocks (7 of them).

4. The innovations to the structural and the volatility shocks are $\mathcal{E}_t = (\varepsilon_{dt}, \varepsilon_{\varphi t}, \varepsilon_{\mu t}, \varepsilon_{At}, \varepsilon_{\xi t}, \varepsilon_{\pi t}, \varepsilon_{yt})'$ and $\mathcal{U}_t = (u_{dt}, u_{\varphi t}, u_{\mu t}, u_{At}, u_{\xi t})'$, respectively.

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, in our notation $\mathcal{Y}_t = (-\triangle \log \mu_t, \log R_t, \log \Pi_t, \triangle \log y_t, \triangle \log w_t)'$. 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. Note that we have the same number of observables as we do of volatility shocks, as required by theorem 5.

We follow the steps in section 2.2 and employ a second-order perturbation around the rescaled steady-state to approximate this equilibrium and build the associated state-space representation (appendix 6.4 gives more details). The second-order approximation is even more relevant in our application because, to complicate matters, a linearization would also imply that the parameter drift in the Taylor rule would disappear as well from the equilibrium dynamics (see appendix 6.3).

4.5. Data and Estimation

We estimate our model using the five time series for the U.S. economy described above. Our sample covers 1959.Q1 to 2007.Q1, with 192 observations. Because of space considerations, we stop at 2007 to avoid having to deal with the recent financial crisis, which would make it difficult to appreciate the points we want to illustrate about how to econometrically deal with stochastic volatility. This could be fixed at the cost of a lengthier discussion. Appendix 6.5 explains how we construct the series.

Once we have evaluated the likelihood, we combine it with a prior. 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 exercises below.

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.* (2010c), 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 to reduce the dimensionality of the problem.

Table 4.1: Fixed Parameters

β	h	ψ	ϑ	δ	α	κ	ε	η	Φ_2	ρ_{γ_Π}	ρ_{γ_y}	σ_y
0.99	0.9	8	1.17	0.025	0.21	9.5	10	10	0.001	0.95	0	0

Table 4.1 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 cost of capital utilization, Φ_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 *et al.* (2009). The first is the inverse of the Frisch labor elasticity, $\vartheta = 1.17$. As argued by Chetty *et al.* (2011), this aggregate elasticity is compatible with micro 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 appear as capital income in the National Income and Product Accounts. Finally, the adjustment cost, $\kappa = 9.5$, is in line with other estimates from similar models (κ 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, ρ_{γ_Π}, is set to 0.95. In preliminary estimations, we discovered that the likelihood pushed this parameter to 1. When this happened, the simulations became numerically unstable: after a series of positive innovations to $\log \gamma_{\Pi t}$, the reaction of nominal interest rates to inflation could be too tepid for too long. The 0.95 value seems to be the highest possible value of ρ_{γ_Π} such that the problem does not appear. The last two parameters, ρ_{γ_y} and σ_y, are equal to zero because, also in exploratory estimations, the likelihood favored values of $\sigma_y \approx 0$. Thus, we decided to forget about them and make $\gamma_{y,t} = 1$.

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 use 10,000 draws from the chain to compute posterior moments.

4.6. Results I: Parameter Estimates

Our first empirical result is the 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 4.2: Posterior, Parameters of Nominal Rigidities and Structural Shocks

θ_p	χ	θ_w	χ_w	ρ_d	ρ_φ	Λ_μ	Λ_A
0.8139 (0.0143)	0.6186 (0.024)	0.6869 (0.0432)	0.6340 (0.0074)	0.1182 (0.0049)	0.9331 (0.0425)	0.0034 (6.6e−5)	0.0028 (4.1e−5)

Table 4.2 presents the results for the nominal rigidities and the stochastic processes for the structural shocks 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 in the data about this result. At the same time, there is a fair amount of indexation, between 0.62-0.63, which brings a strong persistence of inflation. While it is tempting to compare our estimates with the micro evidence on the individual duration of prices, 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.

We estimate a low persistence of the intertemporal preference shock and a high persistence of the intratemporal one. The low estimate of ρ_d produces 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 4.3: Posterior, Parameters of the Stochastic Processes for Volatility Shocks

$\log \sigma_d$	$\log \sigma_\varphi$	$\log \sigma_\mu$	$\log \sigma_A$	$\log \sigma_\xi$
-1.9834 (0.0726)	-2.4983 (0.0917)	-6.0283 (0.1278)	-3.9013 (0.0745)	-6.000 (0.1471)
ρ_{σ_d}	ρ_{σ_φ}	ρ_{σ_μ}	ρ_{σ_a}	ρ_{σ_ξ}
0.9506 (0.0298)	0.1275 (0.0032)	0.7508 (0.035)	0.2411 (0.005)	0.8550 (0.0231)
η_d	η_φ	η_μ	η_a	η_ξ
0.1007 (0.0083)	2.8316 (0.0669)	0.3115 (0.006)	0.7720 (0.013)	0.5723 (0.0185)

The results for the parameters of the stochastic volatility processes appear in table 4.3. In all cases, the ρ's and the η'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, $\eta_\varphi = 2.8316$, is large: the model asks for fast changes in the size of movements in marginal utilities of leisure to reproduce the hours data.

Table 4.4: Posterior, Policy Parameters

γ_R	$\log \gamma_y$	Π	$\log \gamma_\Pi$	σ_π
0.7855 (0.0162)	-1.4034 (0.0498)	1.0005 (0.0043)	0.0441 (0.0005)	0.145 (0.002)

In table 4.4., 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 γ_y (0.24 in levels) is similar to other results in the literature and shows that the likelihood clearly likes parameter drifting, although with mild persistence. The estimated value of Π plus the correction on equilibrium inflation implied by second-order effects of the solution match the average inflation in the data.[3] Finally, the estimated value of γ_Π (1.045 in levels) guarantees local determinacy of the equilibrium even if $\gamma_{\Pi,t}$ is temporarily below 1 (see appendix 6.6 for details).

In appendix 6.7 we plot the impulse response functions of the model implied by our estimates. This exercise allows us to check that the estimates are sensible and that the behavior of the model is consistent with the behavior of related models in the literature. In appendix 6.8 we compare our model against an alternative version without parameter drifting but still with stochastic volatility. That is, we ask whether, once we have included stochastic volatility, it is still important to allow for changes in the monetary policy rule to account for the time-varying volatility of U.S. aggregate data over the last several decades. The results show that, even after controlling for stochastic volatility, the data strongly prefer a specification where the monetary policy rule has changed over time. In appendix 6.9 we show, though, that this finding does not imply that volatility shocks did not play an important role in the time-varying volatilities of U.S. aggregate time series. In other words, both stochastic volatility and parameter drifting are key parts of a successful dynamic equilibrium model of the U.S. economy.

[3] Also, these second-order effects enormously complicate the introduction of time-variation in Π. The likelihood wants to match the moments of the ergodic distribution of inflation, not the level of Π, which is inflation along the balanced growth path. When we have non-linearities, the mean of that ergodic distribution may be far from Π. Thus, learning about Π is hard. Learning about a time-varying Π is even harder.

4.7. Results II: Smoothed Shocks

Figure 4.1 reports the log-deviations with respect to their means for the smoothed intertemporal, intratemporal, and monetary shocks and deviations of the growth rate of the investment and technological shocks with respect to their means (± 2 standard deviations). 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).

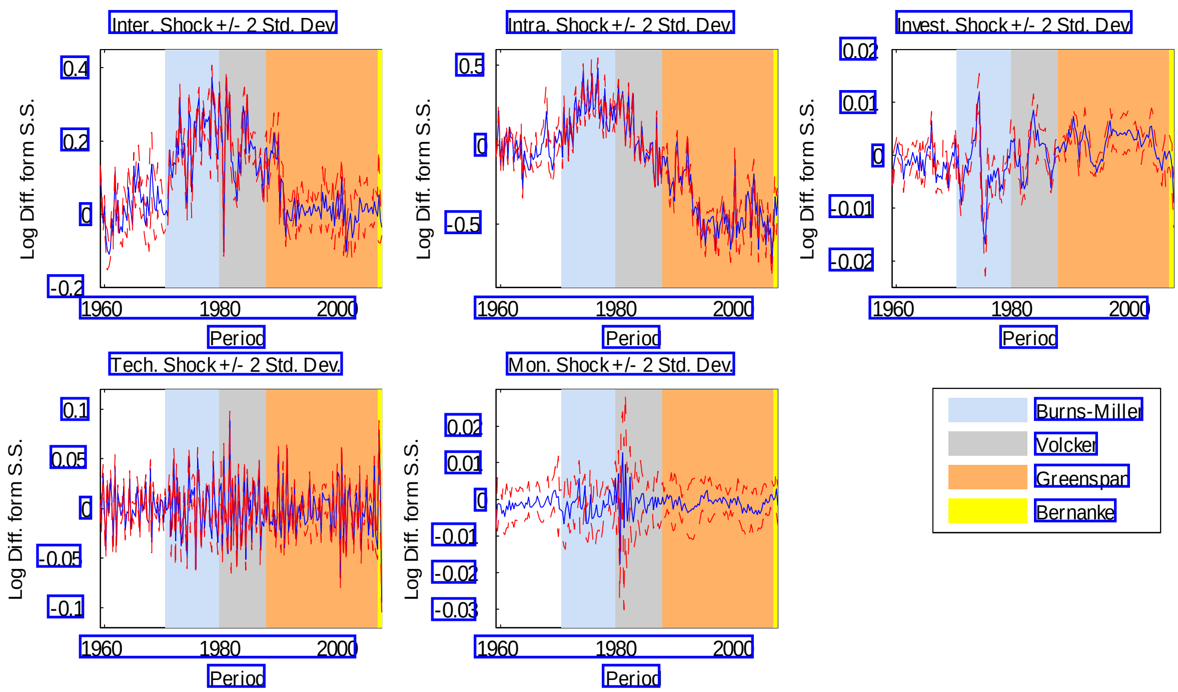

Figure 4.1: Smoothed intertemporal ($\log d_t$) shock, intratemporal ($\log \varphi_t$) shock, investment-specific ($\log \widetilde{\mu}_t$) shock, technology ($\log \widetilde{A}_t$) shock, and monetary policy ($\log \xi_t$) shock +/- 2 s.d.

We see in the top left panel of figure 4.1 that the intertemporal shock, $\log d_t$, 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. The shock has a dramatic drop in the second quarter of 1980. This is precisely the quarter in which the Carter administration invoked the Credit Control Act (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 $\log d_t$. The low values of $\log d_t$ in the 1990s with respect to the 1970s and 1980s eased the inflationary pressures in the economy.

The shock to the utility of leisure, $\log \varphi_t$, 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 technology level). The reduction in marginal costs also helped to reduce inflation during Greenspan's tenure.

The evolution of the investment-specific technology, $\log \mu_t$, shows a sharp 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 large 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 the 1990s. In comparison, the neutral-technology shocks, $\log \widetilde{A}_t$, have been stable since 1959, with only a few big shocks at the end of the sample.

The evolution of the monetary policy shock, $\log \xi_t$, reveals large innovations in the early 1980s. This is due both to the fast change in policy brought about by Volcker 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 also plot in figure 4.1 the same shock (± 2 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.

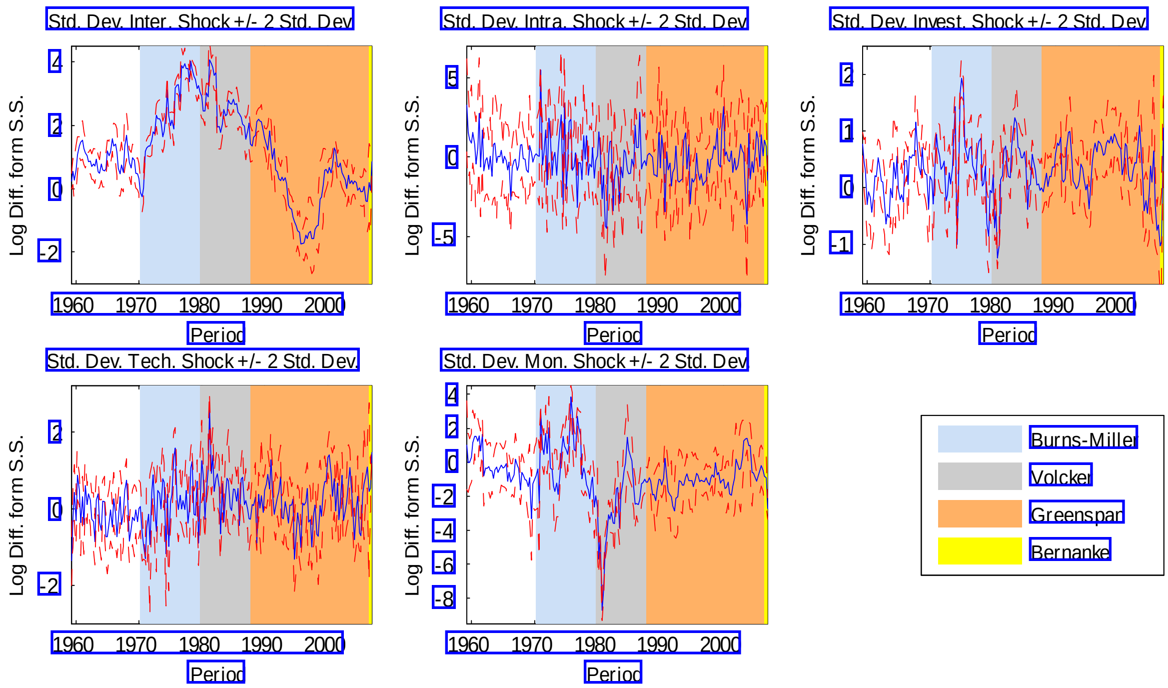

Figure 4.2: Smoothed standard deviation shocks to the intertemporal ($\log \sigma_{dt}$) shock, the intratemporal ($\log \sigma_{\phi t}$) shock, the investment-specific ($\log \sigma_{\mu t}$) shock, the technology ($\log \sigma_{At}$) shock, and the monetary policy ($\log \sigma_{\xi t}$) shock +/- 2 s.d.

We move now, in figure 4.2, to plot the evolution of the volatility shocks, all of them in log-deviations with respect to their estimated means (plus/minus two standard deviations). 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 a large drop in the standard deviation of the monetary policy shock in the early 1980s as well as large 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) and Nason and Smith's (2008) observation that volatility had a downward trend in the 20th century with an abrupt and temporal increase in the 1970s. Also, from the size of the plus/minus two standard deviations, we conclude that the big movements in the different series that we report can be ascertained with a reasonable degree of confidence.

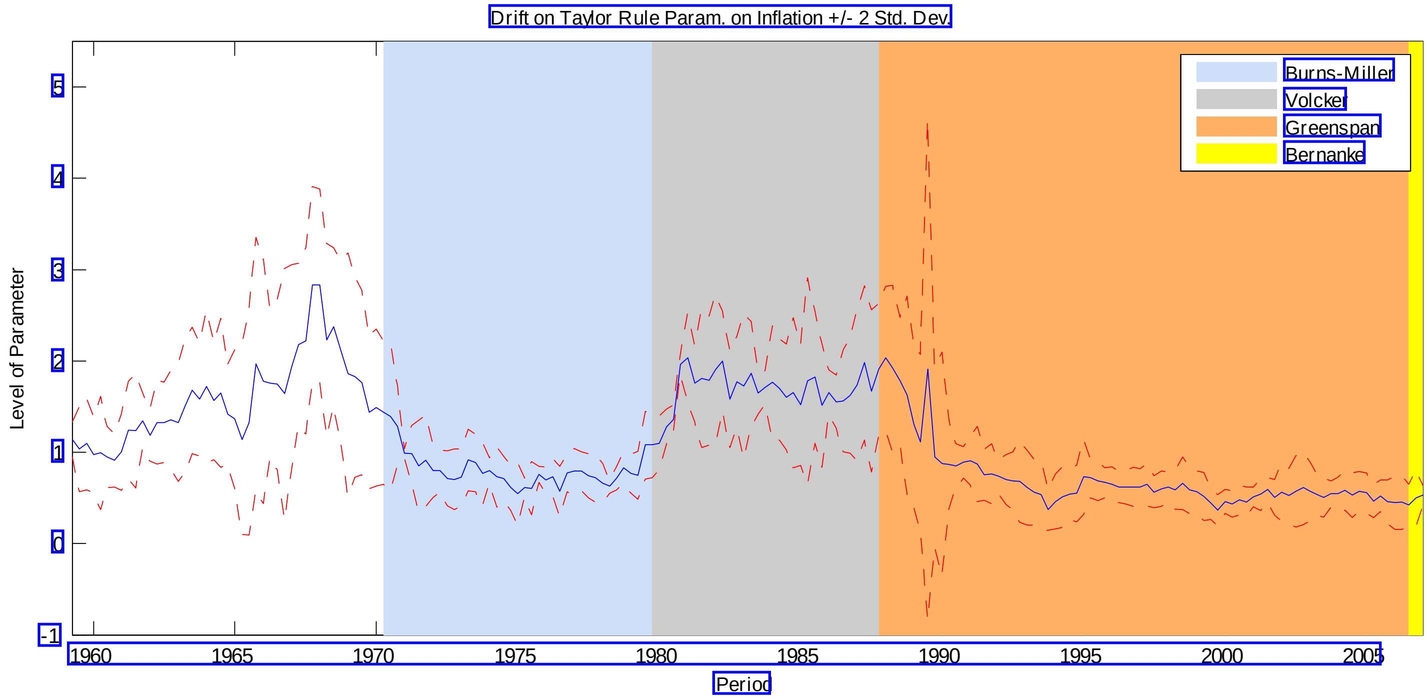

Figure 4.3: Smoothed path for the Taylor rule parameter on inflation +/- 2 standard deviations.

Finally, in figure 4.3, we plot the evolution of the response of monetary policy to inflation plus/minus a two-standard-deviation interval. In particular, we graph $\gamma_{\Pi}\gamma_{\Pi t}$. This graph shows us an intriguing narrative. The parameter 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, it 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. The arrival of Volcker is quickly picked up by our smoothed estimates: it 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. Its level stayed roughly constant at this high during the remainder of Volcker's tenure. But as quickly as it 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. Moreover, our estimates are relatively tight. Fernández-Villaverde *et al.* (2010a) discuss how the results of the estimation relate to historical evidence.

5. Conclusion

In this paper, we have shown how to estimate dynamic equilibrium models with stochastic volatility. The key of the procedure is to realize that a second-order perturbation to the solution of this class of models has a very particular structure that can be easily exploited to build an efficient particle filter. The recent boom in the literature on dynamic equilibrium models with stochastic volatility suggests that this procedure may have many uses. Our characterization of the solution might also be, on many occasions, of interest in itself to understand the dynamic properties of the equilibrium even if the researcher does not want to estimate the model.

As an application to illustrate how the procedure works we have estimated a business cycle model with both stochastic volatility in the structural shocks that drive the economy and parameter drifting in the monetary policy rule. Such a model is motivated by the need to have an empirical framework where we can account for the time-varying volatility of U.S. aggregate time series. In particular, we have explained how you get point estimates in such a model and how to recover and analyze the smoothed structural and volatility shocks. Finally, through different comments -even if brief and not exhaustive- we have discussed the different empirical lessons that one can get from all those steps.

References

[1] 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.

[2] Bansal, R. and A. Yaron (2004). "Risks for the Long Run: a Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59, 1481-1509.

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

[4] Binsbergen, J.V., J. Fernández-Villaverde, R.J. Koijen, and J. Rubio-Ramirez (2012) "The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences." *Journal of Monetary Economics*, forthcoming.

[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., S. Bond, and J. Van Reenen (2007). "Uncertainty and Investment Dynamics." *Review of Economic Studies* 74, 391-415.

[7] Bloom, N. (2009). "The Impact of Uncertainty Shocks." *Econometrica* 77, 623–685.

[8] Bloom, N., N. Jaimovich, N., and M. Floetotto (2012). "Really Uncertain Business Cycles." *NBER* Working Paper 18245.

[9] Caldara, D., J. Fernández-Villaverde, J. Rubio-Ramirez, and W. Yao (2012). "Computing DSGE Models with Recursive Preferences and Stochastic Volatility." *Review of Economic Dynamics* 15, 188-206.

[10] Chetty, R., A. Guren, D. Manoli, and A. Weber (2011)." Micro versus Macro Labor Supply Elasticities." *American Economic Review: Papers and Proceedings*, 101, 472-475.

[11] 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.

[12] Doucet A., N. de Freitas and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. Springer Verlag.

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

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

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

[16] 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.

[17] Fernández-Villaverde, J., P. Guerrón-Quintana, and J. Rubio-Ramírez (2010a). "Reading the Recent Monetary History of the U.S., 1959-2007" *Federal Reserve Bank of St. Louis Review* 92(4), 311-38.

[18] Fernández-Villaverde, J., P. Guerrón-Quintana, and J. Rubio-Ramírez (2010b). "Fortune or Virtue: Time-Variant Volatilities Versus Parameter Drifting in U.S. Data." *NBER Working Papers* 15928.

[19] Fernández-Villaverde, J., P. Guerrón-Quintana, J. Rubio-Ramírez, and M. Uribe (2010c). "Risk Matters: The Real Effects of Volatility Shocks." *American Economic Review* 101, 2530-61.

[20] Flury, T. and N. Shephard (2011). "Bayesian Inference Based Only on Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models." *Econometric Theory* 27, 933-956.

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

[22] Judd, K. and S. Guu (1997). "Asymptotic Methods for Aggregate Growth Models." *Journal of Economic Dynamics and Control* 21, 1025-1042.

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

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

[25] Nason, James M. and Gregor W. Smith (2008). "Great Moderation(s) and US Interest Rates: Unconditional Evidence." The B.E. Journal of Macroeconomics (Contributions) 8, Art. 30.

[26] Pitt, M.K. and N. Shephard (1999). "Filtering via Simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association* 94, 590-599.

[27] Santos, M.S. and A. Peralta-Alva (2005). "Accuracy of Simulations for Stochastic Dynamic Models". *Econometrica* 73, 1939-1976.

[28] 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.

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

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

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

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

6. Technical Appendix (Not for Publication)

This technical appendix is organized as follows. First, it presents the proofs of theorems 2 and 3. Second, it describes the equilibrium of the model in more detail. Third, it shows how parameter drifting in the monetary policy rule does not appear in the first-order approximation. Fourth, it focuses on describing how we efficiently compute the model. Fifth, it offers details on how we elaborated the data. Sixth, it discusses the determinacy of the model. Finally, it includes some additional empirical results not reported in the main text regarding the IRF of the model, the fit of the model in comparison with some alternatives, and some counterfactuals.

6.1. Proofs

6.1.1. Theorem 2

We start by proving theorem 2. In this theorem, we characterize the first-order derivatives of the policy functions h and g evaluated at the steady state. We first show that the first partial derivatives of h and g with respect to any component of Σ_{t-1}, $\mathcal{U}_t$, or Λ evaluated at the steady state are zero (in other words, that the first-order approximation of the policy functions do not depend on volatility shocks nor their innovations nor on the perturbation parameter). Before proceeding, note that using (2), we can write $\mathcal{Z}_{t+1}$, in a compact manner, as a function of $\mathcal{Z}_t, \Sigma_t, \mathcal{E}_{t+1}, \mathcal{U}_{t+1}$, and Λ

$$\mathcal{Z}_{t+1} = \varsigma\left(\mathcal{Z}_t, \Sigma_t, \Lambda\mathcal{E}_{t+1}, \Lambda\mathcal{U}_{t+1}; \gamma\right), \tag{26}$$

that using (3) Σ_{t+1} can be expressed as a function of Σ_t, $\mathcal{U}_{t+1}$, and Λ

$$\Sigma_{t+1} = \vartheta\Sigma_t + \eta\Lambda\mathcal{U}_{t+1}, \tag{27}$$

that using (4) we can write $\mathcal{Z}_t$ as a function of $\mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t$, and $\mathcal{U}_t$

$$\mathcal{Z}_t = \varsigma\left(\mathcal{Z}_{t-1}, \Sigma_{t-1}, \mathcal{E}_t, \mathcal{U}_t; \gamma\right), \tag{28}$$

and that using (5) Σ_t can be expressed as

$$\Sigma_t = \vartheta\Sigma_{t-1} + \eta\mathcal{U}_t \tag{29}$$

where ϑ and η are both $m \times m$ diagonal matrices with diagonal elements equal to ϑ_i and $\left(1-\vartheta_i^2\right)^{\frac{1}{2}} \eta_i$ respectively. If we substitute the policy functions (6)-(8) and (26)-(29) into the set of equilibrium conditions (1), we get that

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

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

where, to ease notation, we do not explicitly write that the functions above depend on γ.

Proof. 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 the homogeneity of a system of linear equations:

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

$$\left[F_{\mathcal{U}_t}\right]_j^i \equiv \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left(\left[g_{\mathcal{S}_t}\right]_{i_1}^{i_2} \left[h_{\mathcal{U}_t}\right]_j^{i_1} + \left[g_{\Sigma_{t-1}}\right]_j^{i_2} \left(1-\vartheta_j^2\right)^{\frac{1}{2}} \eta_j\right) + \left[f_{\mathcal{Y}_t}\right]_{i_2}^i \left[g_{\mathcal{U}_t}\right]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i \left[h_{\mathcal{U}_t}\right]_j^{i_1} = 0$$

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 $[h_{\mathcal{U}_t}]_j^{i_1}$ and $[g_{\mathcal{U}_t}]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$. Thus

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

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

3. Finally, we write the derivative of the $i - th$ element of F with respect to Λ as

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

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

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

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

■

6.1.2. Theorem 3

Let us now prove theorem 3. We show, among other things, that the second partial derivatives of h and g with respect to either $\log \sigma_{it}$ or $u_{i,t}$ and any other variable but $\varepsilon_{i,t}$ are also zero for any $i \in \{1, \ldots, m\}$. We divide the proof into three parts.

Proof, part 1. The first part of the proof deals with the cross-derivatives of the policy functions h and g with respect to Λ and any of $\mathcal{S}_t$, $\mathcal{Z}_{t-1}$, Σ_{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. We again exploit the homogeneity of a system of linear equations.

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

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

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

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

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 Λ and the $j-th$ element of $\mathcal{Z}_{t-1}$

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

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

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

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 Λ and the $j-th$

element of Σ_{t-1}

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

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 Λ and the $j-th$ element of $\mathcal{E}_t$

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

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 $[g_{\Lambda,\mathcal{S}_t}]_j^{i_2} = 0$ for $i_2 \in \{1,\ldots,k\}$ and $j \in \{1,\ldots,n\}$, this is a homogeneous system on $[h_{\Lambda,\mathcal{E}_t}]_j^{i_1}$ and $[g_{\Lambda,\mathcal{E}_t}]_j^{i_2}$ for $i_1 \in \{1,\ldots,n\}$, $i_2 \in \{1,\ldots,k\}$, and $j \in \{1,\ldots,m\}$. Thus

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

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 Λ and the $j-th$ element of $\mathcal{U}_t$

$$[F_{\Lambda,\mathcal{U}_t}]_j^i = \left[f_{\mathcal{Y}_{t+1}}\right]_{i_2}^i \left([g_{\mathcal{S}_t}]_{i_1}^{i_2} [h_{\Lambda,\mathcal{U}_t}]_j^{i_1} + \left[g_{\Lambda,\Sigma_{t-1}}\right]_j^{i_2} \left(1-\vartheta_j^2\right)^{\frac{1}{2}} \eta_j\right)$$
$$+ [f_{\mathcal{Y}_t}]_{i_2}^i [g_{\Lambda,\mathcal{U}_t}]_j^{i_2} + \left[f_{\mathcal{S}_{t+1}}\right]_{i_1}^i [h_{\Lambda,\mathcal{U}_t}]_j^{i_1} = 0$$

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 $[h_{\Lambda,\mathcal{U}_t}]_j^{i_1}$ and $[g_{\Lambda,\mathcal{U}_t}]_j^{i_2}$ for $i_1 \in \{1, \ldots, n\}$, $i_2 \in \{1, \ldots, k\}$, and $j \in \{1, \ldots, m\}$. Then

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

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

$\blacksquare$

Proof, part 2. The second part of the proof deals with the cross-derivatives of the policy functions h and g with respect to Σ_{t-1} and any of $\mathcal{S}_t$, $\mathcal{Z}_{t-1}$, Σ_{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

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

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

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

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

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

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 to be equal to zero from previous derivations).

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 Σ_{t-1}

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

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

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

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

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 Σ_{t-1}

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

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

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

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

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

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

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 Σ_{t-1} if $j_1 \neq j_2$

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

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

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

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

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

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

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 $i \in \{1,\ldots,k+n+m\}$ and $j_1 \in \{1,\ldots,m\}$, we have that this system is not homogeneous and

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

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

■

Proof, part 3. The final part of the proof deals with the cross-derivatives of the policy functions h and g with respect to $\mathcal{U}_t$ and any of $\mathcal{S}_t$, $\mathcal{Z}_{t-1}$, Σ_{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

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

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

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

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

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

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

$$\begin{aligned}[F_{\mathcal{S}_t,\mathcal{U}_t}]^{i}_{j_1,j_2} = \left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2} \left([g_{\mathcal{S}_t}]^{i_2}_{i_1} [h_{\mathcal{S}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_2} + \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2} [h_{\mathcal{S}_t}]^{i_1}_{j_1} \left(1-\vartheta^2_{j_2}\right)^{\frac{1}{2}} \eta_{j_2}\right) \\ + [f_{\mathcal{Y}_t}]^{i}_{i_2} [g_{\mathcal{S}_t,\mathcal{U}_t}]^{i_2}_{j_1,j_2} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1} [h_{\mathcal{S}_t,\mathcal{U}_t}]^{i_1}_{j_1,j_2} = 0\end{aligned}$$

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

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

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

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 $\mathcal{U}_t$

$$
\left[F_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i}_{j_1,j_2} =
$$

$$
\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2} \left(\begin{array}{c} [g_{\mathcal{S}_t}]^{i_2}_{i_1} \left[h_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} [h_{\mathcal{Z}_t}]^{i_1}_{j_1} \left(1-\vartheta^2_{j_2}\right)^{\frac{1}{2}} \eta_{j_2} \\ + \left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} \rho_{j_1} \left(1-\vartheta^2_{j_2}\right)^{\frac{1}{2}} \eta_{j_2} \end{array} \right)
$$

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

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 $\left[g_{\mathcal{Z}_{t-1},\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}$ for $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, and $j_1, j_2 \in \{1,\dots,m\}$. 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 Σ_{t-1} and the $j_2 - th$ element of $\mathcal{U}_t$

$$
\left[F_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i}_{j_1,j_2} =
$$

$$
\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2} \left([g_{\mathcal{S}_t}]^{i_2}_{i_1} \left[h_{\Sigma_{t-1},\mathcal{U}_t}\right]^{i_1}_{j_1,j_2} + \left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} \vartheta_{j_1} \left(1-\vartheta^2_{j_2}\right)^{\frac{1}{2}} \eta_{j_2} \right)
$$

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

for $i \in \{1,\dots,k+n+m\}$, and $j_1, j_2 \in \{1,\dots,m\}$. Since $\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2} = 0$ for $i_2 \in \{1,\dots,k\}$, and $j_1, j_2 \in \{1,\dots,m\}$, 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 $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, and $j_1, j_2 \in \{1,\dots,m\}$. 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 $i_1 \in \{1,\dots,n\}$, $i_2 \in \{1,\dots,k\}$, $j_1, j_2 \in \{1,\dots,m\}$.

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 $\mathcal{U}_t$

$$
\begin{gathered}
\left[F_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i}_{j_1,j_2} = \\
\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\left[g_{\Sigma_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\left(1-\vartheta_{j_1}^2\right)^{\frac{1}{2}}\eta_{j_1}\left(1-\vartheta_{j_2}^2\right)^{\frac{1}{2}}\eta_{j_2}+\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}\right) \\
+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}+\left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{U}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}=0
\end{gathered}
$$

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 $j_1\neq j_2$

$$
\begin{gathered}
\left[F_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i}_{j_1,j_2} = \\
\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\begin{array}{c}\left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}+\left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_2}\left[h_{\mathcal{E}_t}\right]^{i_1}_{j_1}\left(1-\vartheta_{j_2}^2\right)^{\frac{1}{2}}\eta_{j_2} \\ +\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_2}\sigma_{j_1}\exp^{\vartheta_{j_1}\log\sigma_{j_1,t-1}}\left(1-\vartheta_{j_2}^2\right)^{\frac{1}{2}}\eta_{j_2}\end{array}\right) \\
+\left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}+\left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}=0
\end{gathered}
$$

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 $\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_2}$ and $\left[g_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_2}$ for $i_1\in\{1,\ldots,n\}$, $i_2\in\{1,\ldots,k\}$, and $j_1,j_2\in\{1,\ldots,m\}$ if $j_1\neq j_2$. Therefore

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

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

$$
\begin{aligned}
&\left[F_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i}_{j_1,j_1} = \\
&\left[f_{\mathcal{Y}_{t+1}}\right]^{i}_{i_2}\left(\begin{array}{c}\left(\left[g_{\mathcal{Z}_{t-1},\Sigma_{t-1}}\right]^{i_2}_{j_1,j_1} + \left[g_{\mathcal{Z}_{t-1}}\right]^{i_2}_{j_1}\right)\sigma_{j_1}\exp^{\vartheta_{j_1}\log\sigma_{j_1,t-1}}\left(1-\vartheta^2_{j_1}\right)^{\frac{1}{2}}\eta_{j_1} \\ + \left[g_{\mathcal{S}_t,\Sigma_{t-1}}\right]^{i_2}_{i_1,j_1}\left[h_{\mathcal{E}_t}\right]^{i_1}_{j_1}\left(1-\vartheta^2_{j_1}\right)^{\frac{1}{2}}\eta_{j_1} + \left[g_{\mathcal{S}_t}\right]^{i_2}_{i_1}\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_1}\end{array}\right) \\
&+ \left[f_{\mathcal{Y}_t}\right]^{i}_{i_2}\left[g_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_2}_{j_1,j_1} + \left[f_{\mathcal{S}_{t+1}}\right]^{i}_{i_1}\left[h_{\mathcal{E}_t,\mathcal{U}_t}\right]^{i_1}_{j_1,j_1} \\
&+ \left(\left[f_{\mathcal{Z}_t}\right]^{i}_{j_1} + \rho_{j_1}\left[f_{\mathcal{Z}_{t+1}}\right]^{i}_{j_1}\right)\sigma_{j_1}\exp^{\vartheta_{j_1}\log\sigma_{j_1,t-1}}\left(1-\vartheta^2_{j_1}\right)^{\frac{1}{2}}\eta_{j_1} = 0
\end{aligned}
$$

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 $i \in \{1, \ldots, k+n+m\}$ and $j_1 \in \{1, \ldots, m\}$, we have that this system is not homogeneous and hence

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

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

6.2. Equilibrium

In this section we describe the equilibrium conditions of the model. First, we introduce the ones related to the household, then the ones related to the firm and the monetary authority, and finally we present the market clearing and aggregation conditions.

6.2.1. Households

We can define two Lagrangian multipliers, λ_{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 λ_{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:

$$
d_t\left(c_{jt} - hc_{jt-1}\right)^{-1} - b\beta\mathbb{E}_t d_{t+1}\left(c_{jt+1} - hc_{jt}\right)^{-1} = \lambda_{jt}, \tag{30}
$$

$$
\lambda_{jt} = \beta\mathbb{E}_t\{\lambda_{jt+1}\frac{R_t}{\Pi_{t+1}}\}, \tag{31}
$$

$$
r_t = \mu_t^{-1}\Phi'\left[u_{jt}\right], \tag{32}
$$

$$q_{jt} = \beta \mathbb{E}_t \left\{ \frac{\lambda_{jt+1}}{\lambda_{jt}} \left((1-\delta) q_{jt+1} + r_{t+1} u_{jt+1} - \mu_{t+1}^{-1} \Phi \left[u_{jt+1} \right] \right) \right\}, \tag{33}$$

and

$$1 = q_{jt} \mu_t \left(1 - V \left[\frac{x_{jt}}{x_{jt-1}} \right] - V' \left[\frac{x_{jt}}{x_{jt-1}} \right] \frac{x_{jt}}{x_{jt-1}} \right) + \beta \mathbb{E} q_{jt+1} \mu_{t+1} \frac{\lambda_{jt+1}}{\lambda_{jt}} V' \left[\frac{x_{jt+1}}{x_{jt}} \right] \left(\frac{x_{jt+1}}{x_{jt}} \right)^2 . \tag{34}$$

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

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

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. All other households simply index their nominal wages given past inflation with an indexation parameter $\chi_w \in [0,1]$.

Since we postulated in the main text 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:

$$f_t = \frac{\eta - 1}{\eta} \left(w_t^* \right)^{1-\eta} \lambda_t w_t^\eta l_t^d + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{1-\eta} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta-1} f_{t+1} \tag{35}$$

and

$$f_t = \psi d_t \varphi_t \left(\frac{w_t}{w_t^*} \right)^{\eta(1+\vartheta)} \left(l_t^d \right)^{1+\vartheta} + \beta \theta_w \mathbb{E}_t \left(\frac{\Pi_t^{\chi_w}}{\Pi_{t+1}} \right)^{\eta(1+\vartheta)} \left(\frac{w_{t+1}^*}{w_t^*} \right)^{\eta(1+\vartheta)} f_{t+1} \tag{36}$$

on the auxiliary variable f_t.

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

$$w_t^{1-\eta} = \theta_w \left(\frac{\Pi_{t-1}^{\chi_w}}{\Pi_t}\right)^{1-\eta} w_{t-1}^{1-\eta} + (1-\theta_w)\, w_t^{*1-\eta}. \tag{37}$$

6.2.2. Firms

The final good producer is perfectly competitive and minimizes its costs subject to the production function (24) 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:

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

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

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

Each of the intermediate goods is produced by a monopolistic competitor. 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:

$$mc_t = \left(\frac{1}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{\alpha}\right)^{\alpha} \frac{w_t^{1-\alpha} r_t^{\alpha}}{A_t} \tag{38}$$

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

Given the 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 $\chi \in [0,1]$.

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:

$$g_t^1 = \lambda_t mc_t y_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^{\chi}}{\Pi_{t+1}}\right)^{-\varepsilon} g_{t+1}^1, \tag{39}$$

$$g_t^2 = \lambda_t \Pi_t^* y_t^d + \beta\theta_p \mathbb{E}_t \left(\frac{\Pi_t^\chi}{\Pi_{t+1}}\right)^{1-\varepsilon} \left(\frac{\Pi_t^*}{\Pi_{t+1}^*}\right) g_{t+1}^2, \tag{40}$$

and

$$\varepsilon g_t^1 = (\varepsilon - 1) g_t^2 \tag{41}$$

where

$$\Pi_t^* \equiv \frac{p_t^*}{p_t} \tag{42}$$

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 $p_t^{1-\varepsilon}$:

$$1 = \theta_p \left(\frac{\Pi_{t-1}^\chi}{\Pi_t}\right)^{1-\varepsilon} + (1-\theta_p) \Pi_t^{*1-\varepsilon}. \tag{43}$$

6.2.3. The Monetary Authority

The model is closed with a monetary authority that sets the nominal interest rates by a modified Taylor rule described in (25).

6.2.4. Market Clearing and Aggregation

Aggregate demand is given by:

$$y_t^d = c_t + x_t + \mu_t^{-1} \Phi\left[u_t\right] k_{t-1}. \tag{44}$$

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

$$y_t^s = \frac{A_t \left(u_t k_{t-1}\right)^\alpha \left(l_t^d\right)^{1-\alpha} - \phi z_t}{v_t^p} \tag{45}$$

where:

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

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

Market clearing requires that

$$y_t = y_t^d = y_t^s. \tag{46}$$

By the properties of Calvo's pricing:

$$v_t^p = \theta_p \left(\frac{\Pi_{t-1}^{\chi}}{\Pi_t} \right)^{-\varepsilon} v_{t-1}^p + (1 - \theta_p) \Pi_t^{*\,-\varepsilon}. \tag{47}$$

Finally, demanded labor is given by:

$$l_t^d = \frac{1}{v_t^w} \int_0^1 l_{jt} dj = l_t \tag{48}$$

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:

$$v_t^w = \theta_w \left(\frac{w_{t-1}}{w_t} \frac{\Pi_{t-1}^{\chi_w}}{\Pi_t} \right)^{-\eta} v_{t-1}^w + (1 - \theta_w) \left(\Pi_t^{w*} \right)^{-\eta}. \tag{49}$$

Thus an equilibrium is characterized by equations (30)-(49), the Taylor rule (25), the law of motion for the structural shocks (augmented with the parameter drifts), and the law of motion for the volatility shocks.

6.3. Non-linearities in Parameter Drifting

We argued in the main text that, since we wanted to consider the effects of stochastic volatility on our model, it was of the essence to deal with higher-order approximations. In this section we argue that higher-order approximations are also key to dealing with parameter drifting in the Taylor rule. The reason is that parameter drifting disappears from a linear solution. To see this, take the Taylor rule defined in (25) (assuming only in this paragraph and to simplify notation that $\gamma_y = 0$ and $\log \sigma_{\xi t} = 0$) and let us rewrite it:

$$f\left(\widehat{R}_t, \widehat{R}_{t-1}, \widehat{\Pi}_t, \widehat{\gamma}_{\Pi,t}, \varepsilon_{\xi t} \right) = \exp^{\widehat{R}_t} - \exp^{\gamma_R \widehat{R}_{t-1} + (1-\gamma_R)\gamma_\Pi \exp^{\widehat{\gamma}_{\Pi,t}} \widehat{\Pi}_t + \sigma_\xi \varepsilon_{\xi t}} = 0.$$

where we have expressed each variable var_t in terms of log deviation with respect to the steady state, $\widehat{var}_t = \log var_t - \log var$. The log-linear approximation of f around the steady state is

$$f\left(\widehat{R}_t, \widehat{R}_{t-1}, \widehat{\Pi}_t, \widehat{\gamma}_{\Pi,t}, \varepsilon_{\xi t}\right) \simeq f_1(\mathbf{0})\,\widehat{R}_t + f_2(\mathbf{0})\,\widehat{R}_{t-1} + f_3(\mathbf{0})\,\widehat{\Pi}_t + f_4(\mathbf{0})\,\widehat{\gamma}_{\Pi,t} + f_5(\mathbf{0})\,\varepsilon_{\xi t}.$$

where $f_i(\mathbf{0})$ is the first derivative of the function f evaluated at $(0,0,0,0,0)$ with respect to the variable i. Note that:

$$f_4\left(\widehat{R}_t, \widehat{R}_{t-1}, \widehat{\Pi}_t, \widehat{\gamma}_{\Pi,t}, \varepsilon_{\xi t}\right) = -\left(1-\gamma_R\right)\gamma_\Pi \widehat{\Pi}_t \exp^{\widehat{\gamma}_{\Pi,t}} \exp^{\gamma_R \widehat{R}_{t-1} + (1-\gamma_R)\gamma_\Pi \exp^{\widehat{\gamma}_{\Pi,t}} \widehat{\Pi}_t + \sigma_\xi \varepsilon_{\xi t}}.$$

Clearly, $f_4(\mathbf{0}) = 0$ and $\widehat{\gamma}_{\Pi,t}$ does not play any role in this first-order approximation. This is the consequence of one variable (Π_t) being raised to another variable ($\gamma_{\Pi,t}$). Thus, the log-linear approximation of the Taylor rule is:

$$f\left(\widehat{R}_t, \widehat{R}_{t-1}, \widehat{\Pi}_t, \widehat{\gamma}_{\Pi,t}, \varepsilon_{mt}\right) \simeq \widehat{R}_t - \gamma_R \widehat{R}_{t-1} - \left(1-\gamma_R\right)\gamma_\Pi \widehat{\Pi}_t + \sigma_\xi \varepsilon_{\xi t} \tag{50}$$

which does not depend on $\gamma_{\Pi,t}$, but only on the steady state γ_Π and it is exactly the same expression as the one without parameter drifting. Hence, in order to capture parameter drifting in the Taylor rule, we *need*, at least, to perform a second-order approximation.

6.4. Computation

In this section, 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 parse 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.

6.5. Construction of Data

When we estimate the model, we make the series provided by the National Income and Product Accounts (NIPA) consistent with the definition of variables in the theory. The main adjustment we undertake is to express both real output and real gross investment in consumption units. Our 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 2000.Q4, 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.

6.6. Determinacy of Equilibrium

We mentioned in the main text that the estimated value of γ_Π (1.045 in levels) guarantees local determinacy of the equilibrium. To see this note that, for local determinacy, the relevant part of the solution of the model is *only* the linear, first-order component. This component depends on γ_Π, 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 γ_{Π} above 0.98 ensure uniqueness. Since the likelihood assigns zero probability to values of γ_{Π} lower than 1.01, well inside the determinacy region, multiplicity of local equilibria is not an issue in our application.

6.7. Impulse Response Functions

As a check of our estimates, we plot the IRFs generated by the model to a monetary policy shock. This exercise is an important test. 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.

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). The left panel of figure 6.1 plots the IRFs to a monetary shock of 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.

The right panel of figure 6.1 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 stance of monetary policy under his tenure reduces the volatility of output. Inflation responds moderately as well since the agents have the expectation that future shocks will be smoothed out by the monetary authority. 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 in the main body of the text.

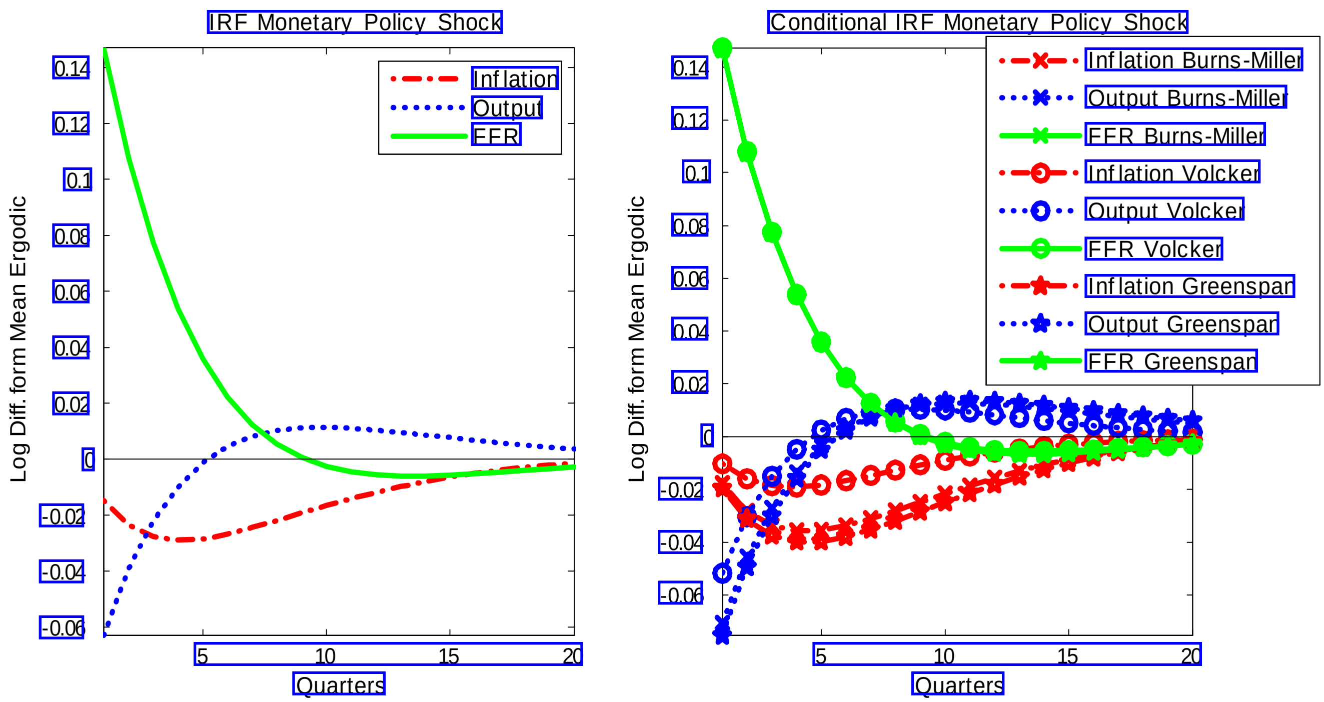

Figure 6.1: IRFs to a Monetary Policy Shock, Unconditional and Conditional

For completeness, we also plot, in figure 6.2, 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. All of those IRFs show that the behavior of the model is standard.

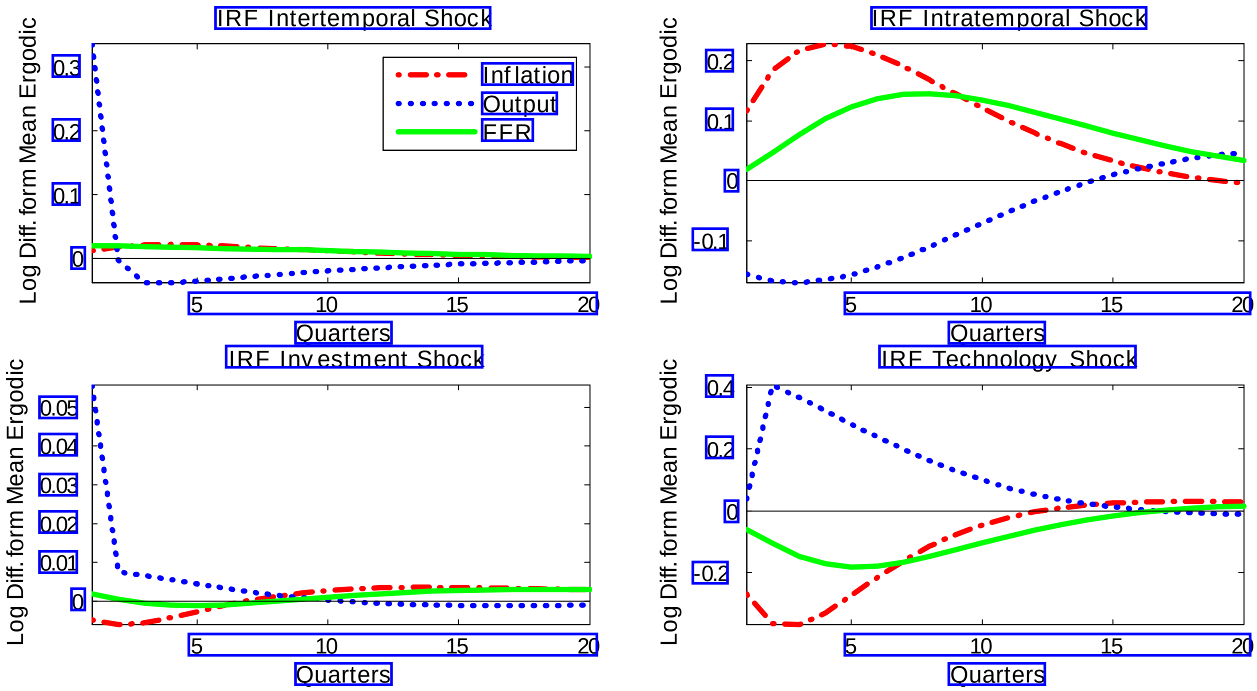

Figure 6.2: IRFs of inflation, output growth, and the federal funds rate to an intertemporal demand (ε_{dt}) shock, an intratemporal demand ($\varepsilon_{\varphi t}$) shock, an investment-specific ($\varepsilon_{\mu t}$) shock, and a neutral technology (ε_{At}) shock. The responses are measured as log differences with respect to the mean of the ergodic distribution.

6.8. Model Comparison

Another use of our procedure to evaluate the likelihood function is to compare our model against alternative models -or alternative versions of the same model. For instance, a natural question is to compare our benchmark model with stochastic volatility and parameter drifting with a version without parameter drifting but with stochastic volatility. That is: once we have included stochastic volatility, is it still important to allow for changes in monetary policy to account for the time-varying volatility of aggregate data in the U.S.?

Given our Bayesian framework, a natural approach for model comparison is the computation of log marginal data densities (log MDD) and log Bayes factors. Let us focus on the proposed example of comparing 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 fewer parameters, ρ_{γ_Π} and σ_π (but we still have γ_Π). To ease notation, we partition the parameter vector γ as $\gamma = \left(\tilde{\gamma}, \rho_{\gamma_{\Pi}}, \sigma_{\pi}\right)$, where $\tilde{\gamma}$ is the vector of all the other parameters, common to the two versions of the model.

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

$$\log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; drift\right) = \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, drift\right) d\gamma + \log p\left(\tilde{\gamma}\right) + \log p\left(\rho_{\gamma_{\Pi}}\right) + \log p\left(\sigma_{\pi}\right),$$

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

$$\log p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; no\ drift\right) = \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \tilde{\gamma}, no\ drift\right) d\tilde{\gamma} + \log p\left(\tilde{\gamma}\right).$$

Thus

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

The difference

$$\log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \gamma, drift\right) d\gamma - \log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \tilde{\gamma}, no\ drift\right) d\tilde{\gamma}$$

tells us how much better the version with parameter drift fits the data in comparison with the version with no drift. The last two terms, $\log p\left(\rho_{\gamma_{\Pi}}\right) + \log p\left(\sigma_{\pi}\right)$, penalize for the presence of two extra parameters.

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 $\log \int p\left(\mathbb{Y}^T = \mathbb{Y}^{data,T}; \tilde{\gamma}, no\ drift\right) d\tilde{\gamma}$. After doing so, we find that

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

This expression shows a potential problem of Bayes factors: by picking uniform priors for

ρ_{γ_Π} and σ_π spread out over a sufficiently large interval, we could overcome any difference in fit. But the prior for ρ_{γ_Π} 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 $\log p\left(\sigma_\pi\right)$. Any sensible prior for σ_π 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 $\log p\left(\sigma_\pi\right) > -1$ ($\log p\left(\sigma_\pi\right) = -1$ would imply a uniform prior between 0 and 2.7183, a considerably wider support than any evidence in the data), and conclude that $\log B_{drift,\ no\ drift} > 124.4400$. 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). Thus, even after controlling for stochastic volatility, the data strongly prefer a specification where monetary policy has changed over time. This finding, however, does not imply that volatility shocks did not play an important role in the time-varying volatilities of U.S. aggregate time series. In fact, as we will see in the next section of this appendix, they were a key mechanism in accounting for it.[4]

It has been noted that the estimation of log MDDs is dangerous because of numerical instabilities in the evaluation of the integral log marginal data density (log MDD). This concern is particularly relevant in our case, since we have a large model saddled with burdensome computation. Thus, 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 the log MDD, can be understood as an asymptotic approximation of the Bayes factor that also automatically penalizes for extra 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, 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

[4] A formal 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.

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 ρ_{γ_Π} and σ_π 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.

6.9. Historical Counterfactuals

One important additional exercise is to quantify how much of the observed changes in the volatility of aggregate U.S. 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. In these exercises, 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 4 of the main text: 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 4.

6.9.1. Counterfactual I: Switching Chairmen

In our first 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 Fed would have followed the policy rule dictated by the average $\gamma_{\Pi t}$ 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 such as the one postulated in section 3. 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. In our model agents have a probability distribution regarding possible changes in monetary policy in the next periods, but those changes are uncorrelated with current conditions. Therefore, our model cannot capture the endogeneity of policy selection.

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.

Moments In table 6.1, we report the mean and the standard deviation of inflation, output growth, and the federal funds rate in the observed and in the sets of counterfactual data. Inflation was high with Burns-Miller, fell with Volcker, and stayed low with Greenspan. Output growth went down during the Volcker years to recover with Greenspan. The federal funds rate reached its peak with Volcker. The standard deviation of output growth fell from 4.7 in Burns-Miller's time to 2.45 with Greenspan, a cut in half. Similarly, inflation volatility fell nearly 54 percent and the federal funds rate volatility 5 percent.

But table 6.1 also tells us one important result: time-varying monetary policy significantly affected average inflation. In particular, Volcker's response to inflation was strong and switching him to either Burns-Miller's or Greenspan's time would have reduced average inflation dramatically. But it also tells us other things: contrary to the conventional wisdom, our estimates suggest that the stance of monetary policy against inflation under Greenspan was not 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 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 higher nominal interest rates would have meant much lower real rates. The counterfactual of Burns-Miller in Greenspan's and Volcker's time casts doubt 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.

Table 6.1: 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

This is an important empirical finding: according to our model, time-varying monetary policy significantly affected average inflation. While Volcker's response to inflation was strong, Greenspan's response was milder and he seems to have behaved quite similarly to how Burns-Miller would have behaved.

Counterfactual Paths An alternative way to analyze our results is to plot the whole counterfactual histories summarized in table 6.1. 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.

In figure 6.3, we move to Burns-Miller being reappointed in Greenspan's time. This plot suggests that the differences in monetary policy under Greenspan and Burns-Miller may have been overstated by the literature.

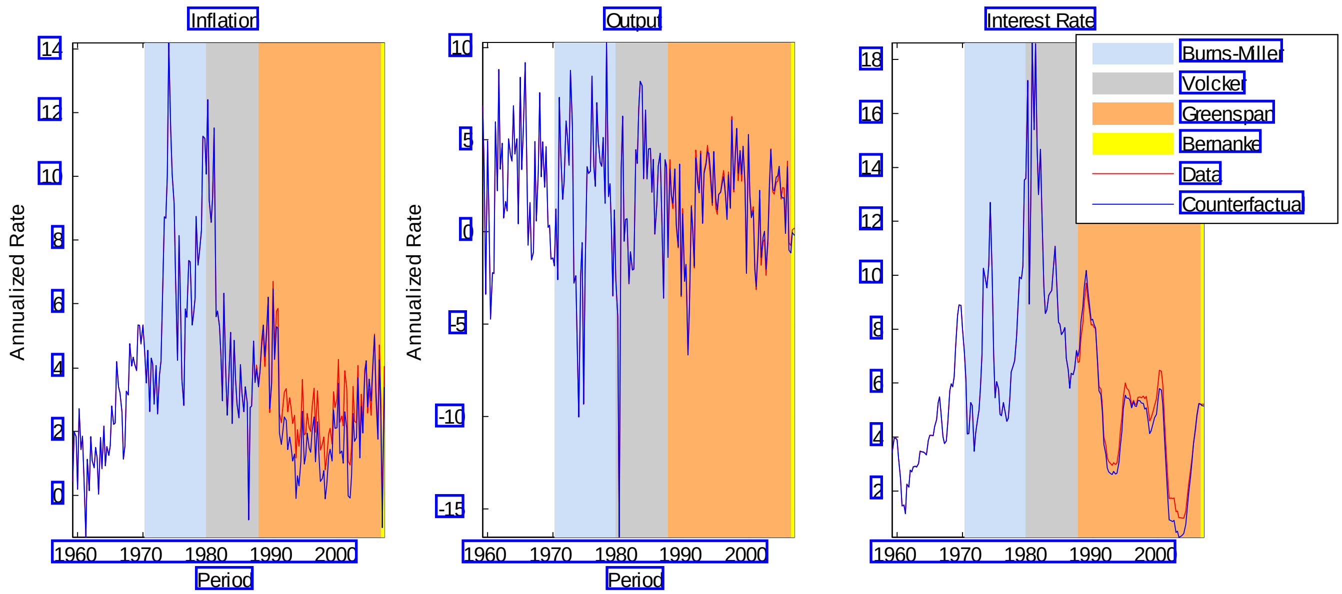

Figure 6.3: Burns-Miller during the Greenspan years

In figure 6.4, 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.

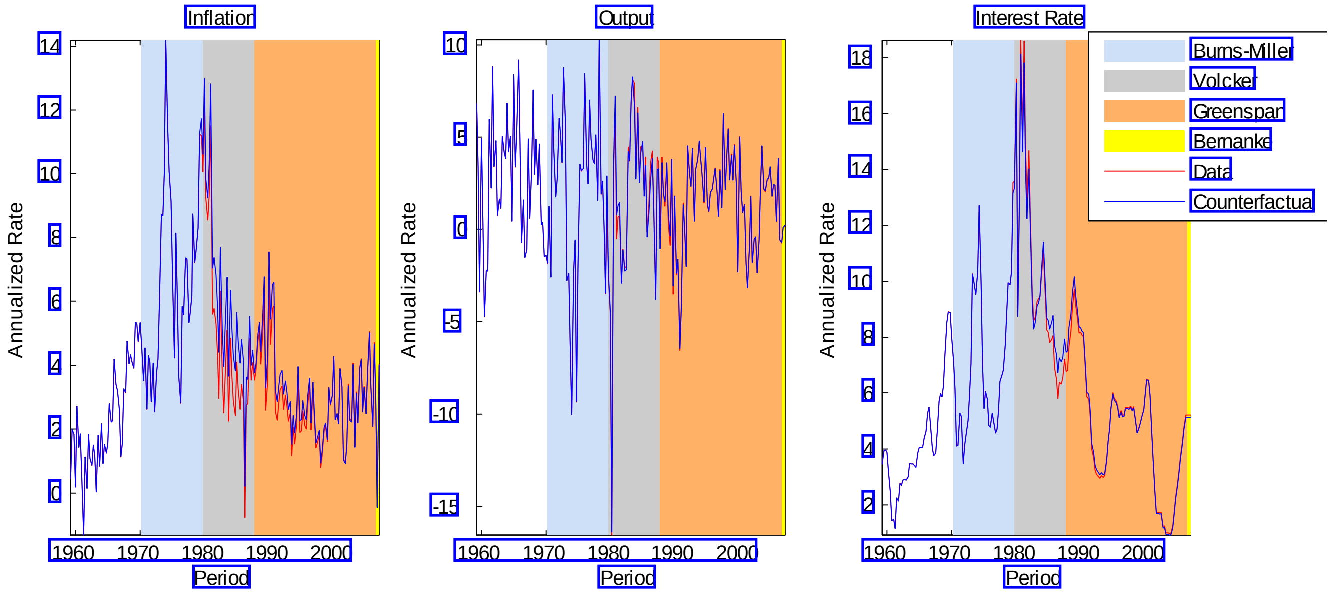

Figure 6.4: Burns-Miller during the Volcker years

A particularly interesting exercise is to check what would have happened if Reagan had decided to reappoint Volcker and not appoint Greenspan. We plot these results in figure 6.5. 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.

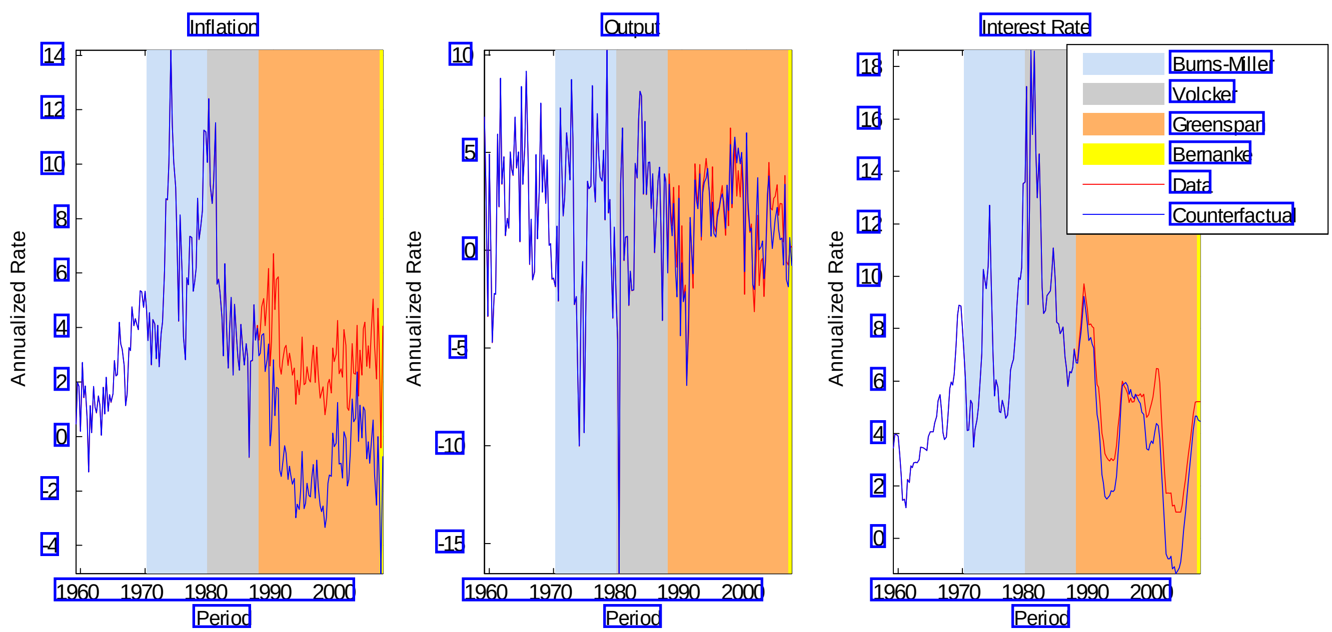

Figure 6.5: Volcker during the Greenspan years

Our final exercise is to plot, in figure 6.6, 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.

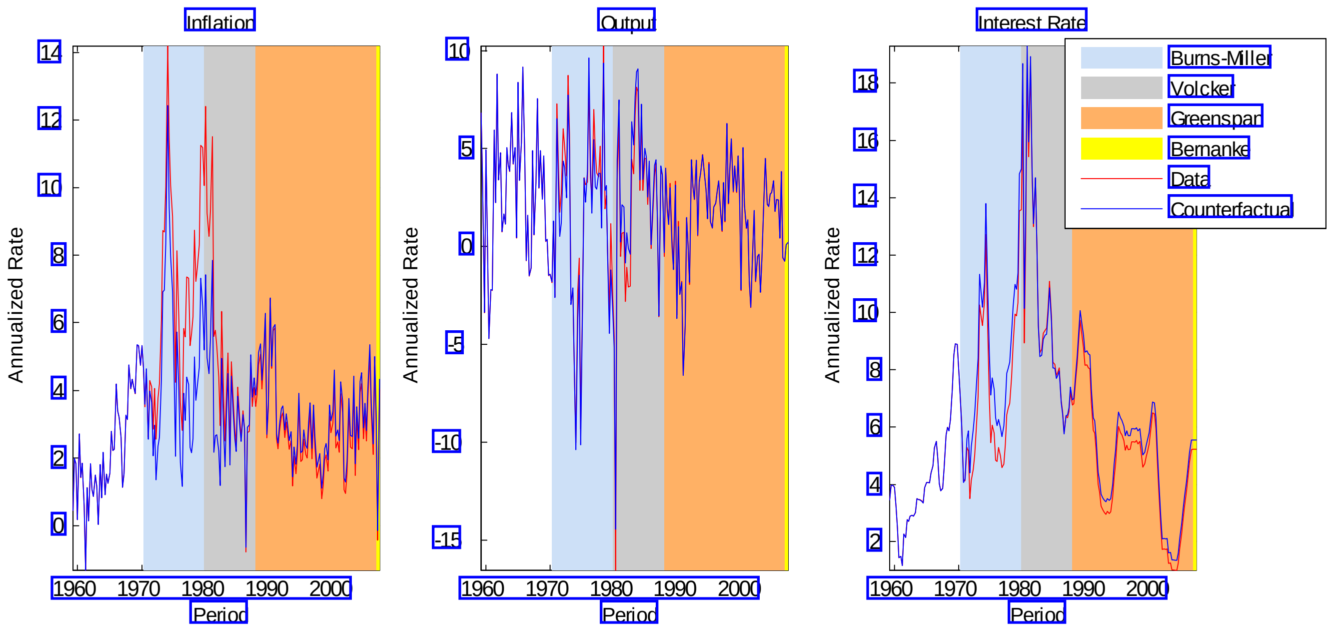

Figure 6.6: Volcker during the Burns-Miller years

6.9.2. Counterfactual II: No Volatility Changes

In our second historical counterfactual, we compute how the economy would have performed 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. To do so, we back out the smoothed structural shocks as we did in section 4.7 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.

Moments Table 6.2 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. The benchmark model with stochastic volatility plus parameter drifting replicates the data exactly.

Some of the numbers in table 6.2 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 the 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.

Table 6.2: 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

The table also reflects the fact that without volatility shocks, the reduction in volatility observed after 1984 would have been noticeably smaller. The standard deviation 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. We must resist here the temptation to undertake 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.

Table 6.2 documents that, while time-varying policy is reflected in changes of average inflation, volatility shocks affect the standard deviation of the observed series. Without time-varying volatility the decrease in observed volatility would not have been nearly as big as we observed in the data. Hence, while changes in the systematic component of monetary policy account for changes in average inflation, volatility shocks account for changes in the standard deviation of inflation, output growth and interest rates observed after 1984. Also, without stochastic volatility, output growth would have been quite lower on average.

Counterfactual Paths Figure 6.7 compares the whole path of the counterfactual history (blue line) with the observed one (red line). Figure 6.7 tells us that volatility shocks mattered throughout the sample. The run-up in 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 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 4 of the paper, this is another manifestation of how placid the 1990s were for policy makers.

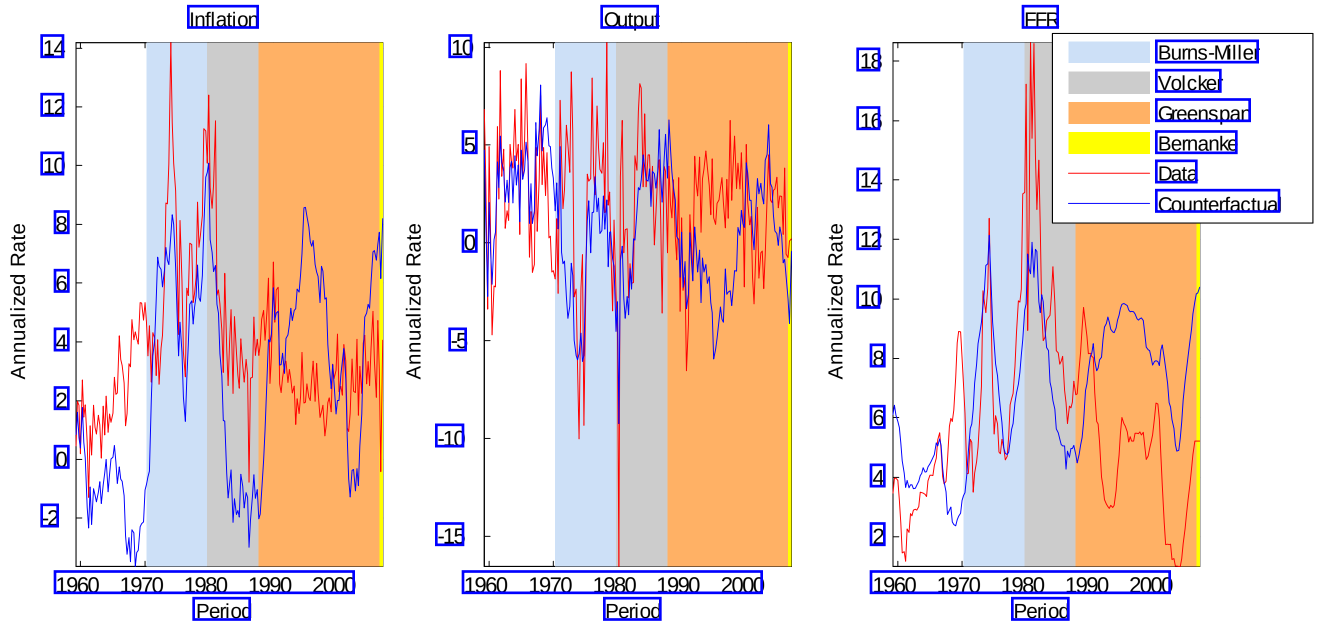

Figure 6.7: Counterfactual “No Changes in Volatility”.

References

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

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

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

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

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

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

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

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

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

[10] 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.

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

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

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

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