

Rare Shocks, Great Recessions

Vasco Cúrdia , Marco Del Negro, Daniel L. Greenwald*

*Federal Reserve Bank of San Francisco, Federal Reserve Bank of New York
and New York University*

October 9, 2012; First Draft: March 2011

WORK IN PROGRESS

Abstract

We estimate a DSGE model where rare large shocks can occur, by replacing the commonly used Gaussian assumption with a Student- t distribution. We show that the Student- t specification is strongly favored by the data using the [Smets and Wouters \(2007\)](#) model estimated on the usual set of macroeconomic time series over the 1964-2011 period, even when we allow for low frequency variation in the volatility of the shocks, and that the estimated degrees of freedom are quite low for several shocks that drive U.S. business cycles, implying an important role for rare large shocks. This result holds even if we exclude the Great Recession period from the sample. We also show that inference about low frequency changes in volatility—and in particular, inference about the magnitude of Great Moderation—is different once we allow for fat tails.

JEL CLASSIFICATION: C32, E3

KEY WORDS: Bayesian Analysis, DSGE Models, Fat tails, Stochastic volatility, Great Recession.

*Vasco Cúrdia: Federal Reserve Bank of San Francisco, vasco.curdia@sf.frb.org. Marco Del Negro: Research Department, Federal Reserve Bank of New York, 33 Liberty Street, New York NY 10045, marco.delnegro@ny.frb.org. Daniel L. Greenwald: New York University, dlg340@nyu.edu. We thank Marek Jarocinski, Ulrich Mueller, Andriy Norets, as well as seminar participants at the Bank of France, 2011 EEA, ECB, 2011 ESOBE, 2012 ESSIM, FRB Chicago, FRB St Louis Econometrics Workshop, RCEF 2012, Seoul National University (Conference in Honor of Chris Sims), 2011 SCE, for useful comments and suggestions. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of New York, the Federal Reserve Bank of San Francisco or the Federal Reserve System.

1 Introduction

Great Recessions do not happen every decade — this is why they are dubbed “great” in the first place. Therefore, to the extent that DSGE models rely on shocks in order to generate macroeconomic fluctuations, they may need to account for the occurrence of rare but extremely large shocks to generate an event like the Great Recession. We therefore estimate a linearized DSGE model assuming that shocks are generated from a Student- t distribution, which is designed to capture fat tails, in place of a Gaussian distribution, which is the standard assumption in the DSGE literature. The number of degrees of freedom in the Student- t distribution, which determines the likelihood of observing rare large shocks and which we allow to vary across shocks, is estimated from the data. We show that estimating DSGE models with Student- t distributed shocks is a fairly straightforward extension of current methods (described, for instance, in [An and Schorfheide \(2007\)](#)). In fact, the Gibbs sampler is a simple extension of [Geweke \(1993\)](#)’s Gibbs sampler for a linear model to the DSGE framework. The paper is closely related to [Chib and Ramamurthy \(2011\)](#) who in independent and contemporaneous work also propose a similar approach to the one developed here for estimating DSGE models with Student- t distributed shocks. One difference between this work and [Chib and Ramamurthy \(2011\)](#)’s is that we also allow for low frequency changes in the volatility of the shocks, in light of the evidence provided by several papers in the DSGE literature ([Justiniano and Primiceri \(2008\)](#), [Fernández-Villaverde and Rubio-Ramírez \(2007\)](#), [Liu et al. \(2011\)](#), among others). We show that ignoring low frequency movements in volatility will bias the results toward finding evidence in favor of fat tails.¹

We apply our methodology to the [Smets and Wouters \(2007\)](#) model (henceforth, SW), estimated on the same seven macroeconomic time series used in SW. Our

¹Specifically, we follow the approach in [Justiniano and Primiceri \(2008\)](#), who postulate a random walk as the law of motion of the volatilities. Another difference between this paper and [Chib and Ramamurthy \(2011\)](#) is that Chib and Ramamurthy use a simple three equation New Keynesian model, while we use the full [Smets and Wouters \(2007\)](#) model. As discussed below, this difference may matter in the assessment of the importance of fat tails for the macroeconomy.

baseline dataset starts in 1964Q4 and ends in 2011Q1, but we also consider a subsample ending in 2004Q4 to analyze the extent to which our findings depend on the inclusion of the Great Recession in our sample. We use the SW model both because it is a prototypical medium-scale DSGE model, and because its empirical success has been widely documented.² Models that fit the data poorly will necessarily have large shocks. We therefore chose a DSGE model that is at the frontier in terms of empirical performance to assess the extent to which macro variables have fat tails.

The motivation for our work arises from evidence like that displayed in Figure 1. The top panel of Figure 1 shows the time series of the smoothed “discount rate” and “marginal efficiency of investment” shocks (in absolute value) from the SW model estimated under Gaussianity. The shocks are normalized, so that they are expressed in standard deviations units. The solid line is the median, and the dashed lines are the posterior 90% bands. The Figure shows that the size of the shocks is between 3.5 and 4 standard deviations on several occasions, one of which is the recent recession. The probability of observing such large shocks under Gaussianity is very low.³

While the visual evidence against Gaussianity is strong, there are several reasons to perform a more careful quantitative analysis. First, these shocks are obtained under the assumption of Gaussianity, and are not informative about the distribution of estimated shocks once that assumption is removed. Second, a quantitative estimate of the fatness of the tails is an obvious object of interest. Third, it is important to disentangle the relative contribution of fat tails from that of (slow moving) time-varying volatility. The bottom panel of Figure 1, which shows the evolution of the smoothed monetary policy shocks estimated under Gaussianity (again, normalized, and in absolute value), provides a case in point, as the clustering of large shocks

² The forecasting performance of the SW model was found to be competitive in terms of accuracy relative to private forecasters and reduced-form models not only during the Great Moderation period (see Smets and Wouters (2007) and Edge and Gürkaynak (2010)), but also including data for the Great Recession (Del Negro and Schorfheide (2012)).

³In addition to this DSGE model-based evidence, there is work showing that the unconditional distribution of macro variables is not Gaussian (see Christiano (2007) for pre-Great Recession evidence, and Ascari et al. (2012) for more recent work).

in the late 70s and 80s is quite evident. In general, studying the importance of fat tails by looking only at the kurtosis in the unconditional distribution of either macro variables (as in [Ascari et al. \(2012\)](#)) or of estimated shocks can be misleading, because the evidence against Gaussianity can be due to low frequency changes in volatility.

Conversely, the presence of large shocks can potentially distort the assessment of low frequency movements in volatility. To see this, imagine estimating a model that allows only for slow moving time variation in volatility, but no fat tails, in presence of shocks that fit the pattern shown in the top panel of [Figure 1](#). As the stochastic volatility will try to fit the squared residuals, such a model will produce a time series of volatilities peaking around 1980, and then again during the Great Recession. Put differently, very large shocks may be interpreted as persistent changes in volatility, when they may in fact be rare realizations from a process with a time invariant distribution. For instance, the extent to which the Great Recession can be interpreted as a permanent rise in macroeconomic volatility may depend on whether we allow for rare large shocks.

Finally, we expect that the evidence provided in this paper will be further motivation for the study of non-linear models. First, if shocks have fat tails, linearization may simply produce a poor approximation of the full model. Second, non-linearities may explain away the fat tails: what we capture as large rare shocks may in fact be Gaussian shocks whose effect is amplified through a non-linear propagation mechanism. Assessing whether this is the case will be an important line of research. In fact, the extent to which non-linearities can alleviate the need for fat-tailed shocks to explain business cycles could possibly become an additional metric for evaluating their usefulness.

Our findings are the following. We provide strong evidence that the Gaussianity assumption in DSGE models is counterfactual, even after allowing for low frequency changes in the volatility of shocks. Such strong evidence is remarkable considering that our sample consists of macro variables only. This finding is robust to excluding the Great Recession from the sample. We follow two approaches in our analysis:

comparing the fit of different specifications using Bayesian marginal likelihoods, and inference on the posterior estimates of the degrees of freedom of the Student- t distribution. We demonstrate that the model fit improves considerably if we allow for Student- t shocks in addition to stochastic volatility. Further, we show that the posterior estimates of the Student- t distribution's degrees of freedom for some shocks are quite low, indicating a substantial amount of fat tails. From our results, we can cluster the shocks in the model into three broad categories. Shocks to productivity, to the household's discount rate, to the marginal efficiency of investment, and to the wage markup all have fat tails, even in the case with stochastic volatility. Conversely, shocks to government expenditures and to price markups have posterior means for the degrees of freedom that are somewhat high, indicating that their distribution is not far from Gaussian, regardless of whether we allow for stochastic volatility. Finally, the degrees of freedom for monetary policy shocks are estimated to be extremely low in the case with constant volatility, but shift dramatically toward higher values when we allow for stochastic volatility.

In order to evaluate the importance of fat tails in the study of the business cycle, we consider an experiment in which we shut down the fat tails, and recreate a counterfactual path of the economy in their absence. We show that in this case almost all recessions in the sample would then have roughly the same magnitudes, and the Great Recession would have been essentially a "run-of-the-mill" recession. Finally, we show that allowing for fat tails will change the inference about slow moving stochastic volatility. Specifically, we reevaluate the evidence in favor of the Great Moderation hypothesis discussed for example in [Justiniano and Primiceri \(2008\)](#) and find that when we consider Student- t shocks, the magnitude of the reduction in the volatility of output and other variables is smaller. Similarly, we show that the evidence in favor of a permanent increase in volatility following the Great Recession is weaker when we consider the possibility that shocks have a Student- t distribution.

The rest of the paper proceeds as follows. [Section 2](#) discusses Bayesian inference. [Section 3](#) describes the model, as well as our set of observables. [Section 4](#) describes

the results. We conclude in section 5.

2 Bayesian Inference

This section begins with a description of the estimation procedure for a DSGE model with both Student- t distributed shocks and time-varying volatilities. The Gibbs sampler combines the algorithm proposed by Geweke (1993)'s for a linear model with Student- t distributed shocks (see also Geweke (1994), and Geweke (2005) for a textbook exposition) with the approach for sampling the parameters of DSGE models with time-varying volatilities discussed in Justiniano and Primiceri (2008). Section A.2 discusses the computation of the marginal likelihood.

The model consists of the standard measurement and transition equations:

$$y_t = Z(\theta)s_t, \quad (1)$$

$$s_{t+1} = T(\theta)s_t + R(\theta)\varepsilon_t, \quad (2)$$

for $t = 1, \dots, T$, where y_t , s_t , and ε_t are $n \times 1$, $k \times 1$, and $\bar{q} \times 1$ vector of observables, states, and shocks, respectively. Call $p(\theta)$ the prior on the vector of DSGE model parameters θ . We assume that:

$$\varepsilon_{q,t} = \sigma_{q,t} \tilde{h}_{q,t}^{-1/2} \eta_{q,t}, \quad \text{all } q, t, \quad (3)$$

where

$$\eta_{q,t} \sim \mathcal{N}(0, 1), \quad \text{i.i.d. across } q, t, \quad (4)$$

$$\lambda_q \tilde{h}_{q,t} \sim \chi^2(\lambda_q), \quad \text{i.i.d. across } q, t. \quad (5)$$

For the prior on the parameters λ_q we assume a gamma distributions with parameters $\underline{\lambda}/\underline{\nu}$ and $\underline{\nu}$:

$$p(\lambda_q | \underline{\lambda}, \underline{\nu}) = \frac{(\underline{\lambda}/\underline{\nu})^{-\underline{\nu}}}{\Gamma(\underline{\nu})} \lambda_q^{\underline{\nu}-1} \exp(-\underline{\nu} \frac{\lambda_q}{\underline{\lambda}}), \quad \text{i.i.d. across } q. \quad (6)$$

where $\underline{\lambda}$ is the mean and $\underline{\nu}$ is the number of degrees of freedom (Geweke (2005) assumes a Gamma with one degree of freedom).

Define

$$\tilde{\sigma}_{q,t} = \log(\sigma_{q,t}/\sigma_q), \quad (7)$$

where the parameters $\sigma_{1:\bar{q}}$ (the non-time varying component of the shock variances) are included in the vector of DSGE parameters θ . We assume that the $\tilde{\sigma}_{q,t}$ follows an autoregressive process:

$$\tilde{\sigma}_{q,t} = \rho_q \tilde{\sigma}_{q,t-1} + \zeta_{q,t}, \quad \zeta_{q,t} \sim \mathcal{N}(0, \omega_q^2), \text{ i.i.d. across } q, t. \quad (8)$$

The prior distribution for ω_q^2 is an inverse gamma $\mathcal{IG}(\nu_\omega/2, \nu_\omega \underline{\omega}^2/2)$, that is:

$$p(\omega_q^2 | \nu_\omega, \underline{\omega}^2) = \frac{(\nu_\omega \underline{\omega}^2/2)^{\frac{\nu_\omega}{2}}}{\Gamma(\nu_\omega/2)} (\omega_q^2)^{-\frac{\nu_\omega}{2}-1} \exp\left[-\frac{\nu_\omega \underline{\omega}^2}{2\omega_q^2}\right], \text{ i.i.d. across } q. \quad (9)$$

We consider two types of priors for ρ_q :

$$p(\rho_q | \omega_q^2) = \begin{cases} 1 & \text{SV-UR} \\ \mathcal{N}(\bar{\rho}, \omega_q^2 \bar{v}_\rho) \mathcal{I}(\rho_q), \text{ i.i.d. across } q, & \mathcal{I}(\rho_q) = \begin{cases} 1 \text{ if } |\rho_q| < 1 & \text{SV-S} \\ 0 \text{ otherwise,} & \end{cases} \end{cases} \quad (10)$$

In the SV-UR case $\tilde{\sigma}_{q,t}$ follows a random walk as in [Justiniano and Primiceri \(2008\)](#), while in the SV-S it follows a stationary process as in [Fernández-Villaverde and Rubio-Ramírez \(2007\)](#). In both cases the $\sigma_{q,t}$ process is very persistent: in the SV-UR case the persistence is wired into the assumed law of motion for $\tilde{\sigma}_{q,t}$, while in the SV-AR case it is enforced by choosing the hyperparameters $\bar{\rho}$ and $\bar{\sigma}_\rho$ in such a way that the prior for ρ_q puts most mass on high values of ρ_q . As a consequence, $\sigma_{q,t}$ and $\tilde{h}_{q,t}$ play very different roles in (3): $\sigma_{q,t}$ allows for slow-moving trends in volatility, while $\tilde{h}_{q,t}$ allows for large shocks. Finally, to close the model we make the following distributional assumptions on the initial conditions $\tilde{\sigma}_{q,0}$, $q = 1, \dots, \bar{q}$:

$$p(\tilde{\sigma}_{q,0} | \rho_q, \omega_q^2) = \begin{cases} 0 & \text{SV-UR} \\ \mathcal{N}(0, \omega_q^2/(1 - \rho_q^2)), \text{ i.i.d. across } q & \text{SV-S} \end{cases} \quad (11)$$

where the restriction under the SV-UR case is needed to obtain identification. In the stationary case we have assumed that $\tilde{\sigma}_{q,0}$ is drawn from the ergodic distribution.

2.1 The Gibbs-Sampler

The joint distribution of data and unobservables (parameters and latent variables) is given by:

$$p(y_{1:T}|s_{1:T}, \theta)p(s_{1:T}|\varepsilon_{1:T}, \theta)p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) \\ p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)p(\lambda_{1:\bar{q}})p(\rho_{1:\bar{q}}|\omega_{1:\bar{q}}^2)p(\omega_{1:\bar{q}}^2)p(\theta), \quad (12)$$

where $p(y_{1:T}|s_{1:T}, \theta)$ and $p(s_{1:T}|\varepsilon_{1:T}, \theta)$ come from the measurement and transition equation, respectively, $p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)$ obtains from (3) and (4):

$$p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) \propto \prod_{q=1}^{\bar{q}} \left(\prod_{t=1}^T \tilde{h}_{q,t}^{-1/2} \sigma_{q,t} \right) \exp \left[- \sum_{t=1}^T \tilde{h}_{q,t} \varepsilon_{q,t}^2 / 2\sigma_{q,t}^2 \right], \quad (13)$$

$p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})$ obtains from (5)

$$p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) = \prod_{q=1}^{\bar{q}} \prod_{t=1}^T \left(2^{\lambda_q/2} \Gamma(\lambda_q/2) \right)^{-1} \lambda_q^{\lambda_q/2} \tilde{h}_{q,t}^{(\lambda_q-2)/2} \exp(-\lambda_q \tilde{h}_{q,t}/2), \quad (14)$$

and $p(\tilde{\sigma}_{1:T}|\omega_{1:\bar{q}}^2)$ obtains from expressions (8) and (11):

$$p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2) \propto \prod_{q=1}^{\bar{q}} (\omega_q^2)^{-(T-1)/2} \exp \left[- \sum_{t=2}^T (\tilde{\sigma}_{q,t} - \rho_q \tilde{\sigma}_{q,t-1})^2 / 2\omega_q^2 \right] p(\tilde{\sigma}_{q,1}|\rho_q, \omega_q^2), \quad (15)$$

where

$$p(\tilde{\sigma}_{q,1}|\rho_q, \omega_q^2) \propto \begin{cases} (\omega_q^2)^{-1/2} \exp \left(-\frac{\tilde{\sigma}_{q,1}^2}{2\omega_q^2} \right), & \text{SV-UR} \\ (\omega_q^2(1-\rho_q^2))^{-1/2} \exp \left(-\frac{\tilde{\sigma}_{q,1}^2}{2\omega_q^2(1-\rho_q^2)} \right). & \text{SV-S} \end{cases} \quad (16)$$

Finally, $p(\lambda_{1:\bar{q}}) = \prod_{q=1}^{\bar{q}} p(\lambda_q|\underline{\lambda})$, $p(\omega_{1:\bar{q}}^2) = \prod_{q=1}^{\bar{q}} p(\omega_q^2|\nu, \underline{\omega}^2)$.

The sampler consists of six blocks.

- (1) Draw from $p(\theta, s_{1:T}, \varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$. This is accomplished in two steps:

(1.1) Draw from the marginal $p(\theta|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$, where

$$\begin{aligned} & p(\theta|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T}) \\ & \propto \left[\int p(y_{1:T}|s_{1:T}, \theta) p(s_{1:T}|\varepsilon_{1:T}, \theta) p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) \cdot d(s_{1:T}, \varepsilon_{1:T}) \right] p(\theta) \\ & = p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) p(\theta) \end{aligned} \quad (17)$$

where

$$p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) = \int p(y_{1:T}|s_{1:T}, \theta) p(s_{1:T}|\varepsilon_{1:T}, \theta) p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) \cdot d(s_{1:T}, \varepsilon_{1:T})$$

is computed using the Kalman filter with (1) as the measurement equation and (2) as transition equation, with

$$\varepsilon_t|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T} \sim \mathcal{N}(0, \Delta_t), \quad (18)$$

where Δ_t are $\bar{q} \times \bar{q}$ diagonal matrices with $\sigma_{q,t}^2 \cdot \tilde{h}_{q,t}^{-1}$ on the diagonal. The draw is obtained from a Metropolis-Hastings step.

(1.2) Draw from the conditional $p(s_{1:T}, \varepsilon_{1:T}|\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$.

This is accomplished using the simulation smoother of [Durbin and Koopman \(2002\)](#).

(2) Draw from $p(\tilde{h}_{1:T}|\theta, s_{1:T}, \varepsilon_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$. This is accomplished by drawing from

$$\begin{aligned} & p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) \\ & \propto \prod_{q=1}^{\bar{q}} \prod_{t=1}^T \tilde{h}_{q,t}^{(\lambda_q-1)/2} \exp(-[\lambda_q + \varepsilon_{q,t}^2/\sigma_{q,t}^2] \tilde{h}_{q,t}/2), \end{aligned} \quad (19)$$

which implies

$$[\lambda_q + \varepsilon_{q,t}^2/\sigma_{q,t}^2] \tilde{h}_{q,t}|\theta, \varepsilon_{1:T}, \tilde{\sigma}_{1:T}, \lambda_q \sim \chi^2(\lambda_q + 1). \quad (20)$$

(3) Draw from $p(\lambda_{1:\bar{q}}|\tilde{h}_{1:T}, \theta, s_{1:T}, \varepsilon_{1:T}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$. This is accomplished by drawing from

$$\begin{aligned} & p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) p(\lambda_{1:\bar{q}}) \propto \prod_{q=1}^{\bar{q}} ((\underline{\lambda}/\underline{\nu})^{\underline{\nu}} \Gamma(\underline{\nu}))^{-1} [2^{\lambda_q/2} \Gamma(\lambda_q/2)]^{-T} \lambda_q^{T\lambda_q/2 + \underline{\nu} - 1} \\ & \left(\prod_{t=1}^T \tilde{h}_{q,t}^{(\lambda_q-2)/2} \right) \exp \left[- \left(\frac{\underline{\nu}}{\underline{\lambda}} + \frac{1}{2} \sum_{t=1}^T \tilde{h}_{q,t} \right) \lambda_q \right]. \end{aligned} \quad (21)$$

This is a non-standard distribution, hence the draw is obtained from a Metropolis-Hastings step.

- (4) Draw from $p(\tilde{\sigma}_{1:T}|\theta, s_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$. This is accomplished by drawing from

$$p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2) \quad (22)$$

using the algorithm developed by Kim et al. (1998), which we briefly describe in Appendix A.3.

- (5) Draw from $p(\omega_{1:\bar{q}}^2, \rho_{1:\bar{q}}|\tilde{\sigma}_{1:T}, \theta, s_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, \lambda_{1:\bar{q}}, y_{1:T})$ using

$$p(\tilde{\sigma}_{1:T}|\rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)p(\omega_{1:\bar{q}}^2)p(\rho_{1:\bar{q}}|\omega_{1:\bar{q}}^2) \propto \prod_{q=1}^{\bar{q}} (\omega_q^2)^{-\frac{\nu+T-1}{2}-1} \exp\left[-\frac{\nu\omega^2 + \sum_{t=2}^T (\tilde{\sigma}_{q,t} - \rho_q \tilde{\sigma}_{q,t-1})^2}{2\omega_q^2}\right] p(\tilde{\sigma}_{q,1}|\rho_q, \omega_q^2)p(\rho_q|\omega_q^2), \quad (23)$$

where $p(\tilde{\sigma}_{q,1}|\rho_q, \omega_q^2)$ is given by equation (16). In the SV-UR case ρ_q is fixed to 1, and we can draw ω_q^2 (i.i.d. across q) from:

$$\omega_q^2|\tilde{\sigma}_{1:T}, \dots \sim \mathcal{IG}\left(\frac{\nu+T}{2}, \frac{1}{2}\left(\nu\omega^2 + \sum_{t=2}^T (\tilde{\sigma}_{q,t} - \tilde{\sigma}_{q,t-1})^2 + \tilde{\sigma}_{q,1}^2\right)\right). \quad (24)$$

In the SV-S case the joint posterior of ρ_q, ω_q^2 is non-standard because of the likelihood of the first observation $p(\tilde{\sigma}_1|\rho_q, \omega_q^2)$. We therefore use the Metropolis-Hastings step proposed by Chib and Greenberg (1994). Specifically, we use as proposal density the standard Normal-Inverted Gamma distribution, that is,

$$\omega_q^2|\tilde{\sigma}_{1:T}, \dots \sim \mathcal{IG}\left(\frac{\nu+T-1}{2}, \frac{1}{2}\left(\nu\omega^2 + \sum_{t=2}^T \tilde{\sigma}_{q,t}^2 + \bar{v}_\rho^{-1}\bar{\rho}^2 - \hat{V}_q^{-1}\hat{\rho}_q^2\right)\right), \quad (25)$$

$$\rho_q|\omega_q^2, \tilde{\sigma}_{1:T}, \dots \sim \mathcal{N}\left(\hat{\rho}_q, \omega_q^2\hat{V}_q\right), \text{ i.i.d. across } q,$$

where $\hat{\rho}_q = \hat{V}_q\left(\bar{v}_\rho^{-1}\bar{\rho} + \sum_{t=2}^T \tilde{\sigma}_{q,t}\tilde{\sigma}_{q,t-1}\right)$, $\hat{V}_q = (\bar{v}_\rho^{-1} + \sum_{t=2}^T \tilde{\sigma}_{q,t}^2)^{-1}$. We then accept/reject this draw using the proposal density and the acceptance ratio $\frac{p(\tilde{\sigma}_1, \rho_q^{(*)}, \omega_q^{2(*)})\mathcal{I}(\rho_q^{(*)})}{p(\tilde{\sigma}_1, \rho_q^{(j-1)}, \omega_q^{2(j-1)})\mathcal{I}(\rho_q^{(j-1)})}$, with $(\rho^{(j-1)}, \omega_q^{2(j-1)})$ and $(\rho^{(*)}, \omega_q^{2(*)})$ being the draw at the $(j-1)^{\text{th}}$ iteration and the proposed draw, respectively.

3 The DSGE Model

The model considered is the one used in [Smets and Wouters \(2007\)](#), which is based on earlier work by [Christiano et al. \(2005\)](#) and [Smets and Wouters \(2003\)](#). It is a medium-scale DSGE model, which augments the standard neoclassical stochastic growth model with nominal price and wage rigidities as well as habit formation in consumption and investment adjustment costs.

3.1 The Smets-Wouters Model

We begin by briefly describing the log-linearized equilibrium conditions of the [Smets and Wouters \(2007\)](#) model. We follow [Del Negro and Schorfheide \(2012\)](#) and detrend the non-stationary model variables by a stochastic rather than a deterministic trend. This approach makes it possible to express almost all equilibrium conditions in a way that encompasses both the trend-stationary total factor productivity process in [Smets and Wouters \(2007\)](#), as well as the case where technology follows a unit root process. We refer to the model presented in this section as SW model. Let \tilde{z}_t be the linearly detrended log productivity process which follows the autoregressive law of motion

$$\tilde{z}_t = \rho_z \tilde{z}_{t-1} + \sigma_z \varepsilon_{z,t}. \quad (26)$$

We detrend all non stationary variables by $Z_t = e^{\gamma t + \frac{1}{1-\alpha} \tilde{z}_t}$, where γ is the steady state growth rate of the economy. The growth rate of Z_t in deviations from γ , denoted by z_t , follows the process:

$$z_t = \ln(Z_t/Z_{t-1}) - \gamma = \frac{1}{1-\alpha}(\rho_z - 1)\tilde{z}_{t-1} + \frac{1}{1-\alpha}\sigma_z\varepsilon_{z,t}. \quad (27)$$

All variables in the subsequent equations are expressed in log deviations from their non-stochastic steady state. Steady state values are denoted by *-subscripts and steady state formulas are provided in a Technical Appendix (available upon

request). The consumption Euler equation takes the form:

$$c_t = -\frac{(1 - he^{-\gamma})}{\sigma_c(1 + he^{-\gamma})} (R_t - \mathbb{E}_t[\pi_{t+1}] + b_t) + \frac{he^{-\gamma}}{(1 + he^{-\gamma})} (c_{t-1} - z_t) \\ + \frac{1}{(1 + he^{-\gamma})} \mathbb{E}_t [c_{t+1} + z_{t+1}] + \frac{(\sigma_c - 1)}{\sigma_c(1 + he^{-\gamma})} \frac{w_* L_*}{c_*} (L_t - \mathbb{E}_t[L_{t+1}]), \quad (28)$$

where c_t is consumption, L_t is labor supply, R_t is the nominal interest rate, and π_t is inflation. The exogenous process b_t drives a wedge between the intertemporal ratio of the marginal utility of consumption and the riskless real return $R_t - \mathbb{E}_t[\pi_{t+1}]$, and follows an AR(1) process with parameters ρ_b and σ_b . The parameters σ_c and h capture the relative degree of risk aversion and the degree of habit persistence in the utility function, respectively. From the optimality condition for the capital producers, we next obtain the following condition, which expresses the relationship between the value of capital in terms of consumption q_t^k and the level of investment i_t measured in terms of consumption goods:

$$q_t^k = S'' e^{2\gamma} (1 + \beta e^{(1-\sigma_c)\gamma}) \left(i_t - \frac{1}{1 + \beta e^{(1-\sigma_c)\gamma}} (i_{t-1} - z_t) \right. \\ \left. - \frac{\beta e^{(1-\sigma_c)\gamma}}{1 + \beta e^{(1-\sigma_c)\gamma}} \mathbb{E}_t [i_{t+1} + z_{t+1}] - \mu_t \right), \quad (29)$$

which is affected by both investment adjustment cost (S'' is the second derivative of the adjustment cost function) and by μ_t , an exogenous process called the ‘‘marginal efficiency of investment’’ that affects the rate of transformation between consumption and installed capital (see Greenwood et al. (1998)). The capital stock, \bar{k}_t , evolves as

$$\bar{k}_t = \left(1 - \frac{i_*}{\bar{k}_*} \right) (\bar{k}_{t-1} - z_t) + \frac{i_*}{\bar{k}_*} i_t + \frac{i_*}{\bar{k}_*} S'' e^{2\gamma} (1 + \beta e^{(1-\sigma_c)\gamma}) \mu_t, \quad (30)$$

where i_*/\bar{k}_* is the steady state ratio of investment to capital. μ_t follows an AR(1) process with parameters ρ_μ and σ_μ . The parameter β captures the intertemporal discount rate in the utility function of the households. The arbitrage condition between the return to capital and the riskless rate is:

$$\frac{r_*^k}{r_*^k + (1 - \delta)} \mathbb{E}_t [r_{t+1}^k] + \frac{1 - \delta}{r_*^k + (1 - \delta)} \mathbb{E}_t [q_{t+1}^k] - q_t^k = R_t + b_t - \mathbb{E}_t [\pi_{t+1}], \quad (31)$$

where r_t^k is the rental rate of capital, r_*^k its steady state value, and δ the depreciation rate. Capital is subject to variable capacity utilization u_t . The relationship between \bar{k}_t and the amount of capital effectively rented out to firms k_t is

$$k_t = u_t - z_t + \bar{k}_{t-1}. \quad (32)$$

The optimality condition determining the rate of utilization is given by

$$\frac{1 - \psi}{\psi} r_t^k = u_t, \quad (33)$$

where ψ captures the utilization costs in terms of foregone consumption. From the optimality conditions of goods producers it follows that all firms have the same capital-labor ratio:

$$k_t = w_t - r_t^k + L_t. \quad (34)$$

Real marginal costs for firms are given by

$$mc_t = w_t + \alpha L_t - \alpha k_t, \quad (35)$$

where α is the income share of capital (after paying markups and fixed costs) in the production function.

All of the equations so far maintain the same form whether technology has a unit root or is trend stationary. However, a few small differences arise for the following two equilibrium conditions. The production function is:

$$y_t = \Phi_p (\alpha k_t + (1 - \alpha)L_t) + \mathcal{I}\{\rho_z < 1\} (\Phi_p - 1) \frac{1}{1 - \alpha} \tilde{z}_t, \quad (36)$$

under trend stationarity. The last term $(\Phi_p - 1) \frac{1}{1 - \alpha} \tilde{z}_t$ drops out if technology has a stochastic trend, because in this case one has to assume that the fixed costs are proportional to the trend. Similarly, the resource constraint is:

$$y_t = g_t + \frac{c_*}{y_*} c_t + \frac{i_*}{y_*} i_t + \frac{r_*^k k_*}{y_*} u_t - \mathcal{I}\{\rho_z < 1\} \frac{1}{1 - \alpha} \tilde{z}_t, \quad (37)$$

where again the term $-\frac{1}{1 - \alpha} \tilde{z}_t$ disappears if technology follows a unit root process. Government spending g_t is assumed to follow the exogenous process:

$$g_t = \rho_g g_{t-1} + \sigma_g \varepsilon_{g,t} + \eta_{gz} \sigma_z \varepsilon_{z,t}.$$

Finally, the price and wage Phillips curves are, respectively:

$$\begin{aligned} \pi_t = & \frac{(1 - \zeta_p \beta e^{(1-\sigma_c)\gamma})(1 - \zeta_p)}{(1 + \iota_p \beta e^{(1-\sigma_c)\gamma}) \zeta_p ((\Phi_p - 1)\epsilon_p + 1)} m c_t \\ & + \frac{\iota_p}{1 + \iota_p \beta e^{(1-\sigma_c)\gamma}} \pi_{t-1} + \frac{\beta e^{(1-\sigma_c)\gamma}}{1 + \iota_p \beta e^{(1-\sigma_c)\gamma}} \mathbb{E}_t[\pi_{t+1}] + \lambda_{f,t}, \end{aligned} \quad (38)$$

and

$$\begin{aligned} w_t = & \frac{(1 - \zeta_w \beta e^{(1-\sigma_c)\gamma})(1 - \zeta_w)}{(1 + \beta e^{(1-\sigma_c)\gamma}) \zeta_w ((\lambda_w - 1)\epsilon_w + 1)} (w_t^h - w_t) \\ & - \frac{1 + \iota_w \beta e^{(1-\sigma_c)\gamma}}{1 + \beta e^{(1-\sigma_c)\gamma}} \pi_t + \frac{1}{1 + \beta e^{(1-\sigma_c)\gamma}} (w_{t-1} - z_t - \iota_w \pi_{t-1}) \\ & + \frac{\beta e^{(1-\sigma_c)\gamma}}{1 + \beta e^{(1-\sigma_c)\gamma}} \mathbb{E}_t [w_{t+1} + z_{t+1} + \pi_{t+1}] + \lambda_{w,t}, \end{aligned} \quad (39)$$

where ζ_p , ι_p , and ϵ_p are the Calvo parameter, the degree of indexation, and the curvature parameters in the Kimball aggregator for prices, and ζ_w , ι_w , and ϵ_w are the corresponding parameters for wages. The variable w_t^h corresponds to the household's marginal rate of substitution between consumption and labor, and is given by:

$$w_t^h = \frac{1}{1 - h e^{-\gamma}} (c_t - h e^{-\gamma} c_{t-1} + h e^{-\gamma} z_t) + \nu_l L_t, \quad (40)$$

where ν_l characterizes the curvature of the disutility of labor (and would equal the inverse of the Frisch elasticity in absence of wage rigidities). The mark-ups $\lambda_{f,t}$ and $\lambda_{w,t}$ follow exogenous ARMA(1,1) processes

$$\lambda_{f,t} = \rho_{\lambda_f} \lambda_{f,t-1} + \sigma_{\lambda_f} \varepsilon_{\lambda_f,t} + \eta_{\lambda_f} \sigma_{\lambda_f} \varepsilon_{\lambda_f,t-1}, \text{ and}$$

$$\lambda_{w,t} = \rho_{\lambda_w} \lambda_{w,t-1} + \sigma_{\lambda_w} \varepsilon_{\lambda_w,t} + \eta_{\lambda_w} \sigma_{\lambda_w} \varepsilon_{\lambda_w,t-1},$$

respectively. Last, the monetary authority follows a generalized feedback rule:

$$\begin{aligned} R_t = & \rho_R R_{t-1} + (1 - \rho_R) \left(\psi_1 \pi_t + \psi_2 (y_t - y_t^f) \right) \\ & + \psi_3 \left((y_t - y_t^f) - (y_{t-1} - y_{t-1}^f) \right) + r_t^m, \end{aligned} \quad (41)$$

where the flexible price/wage output y_t^f is obtained from solving the version of the model without nominal rigidities (that is, Equations (28) through (37) and (40)), and the residual r_t^m follows an AR(1) process with parameters ρ_r^m and σ_r^m .

3.2 Observation Equation, Data, and Priors

We use the method in Sims (2002) to solve the log-linear approximation of the DSGE model. We collect all the DSGE model parameters in the vector θ , stack the structural shocks in the vector ϵ_t , and derive a state-space representation for our vector of observables y_t , which is composed of the transition equation:

$$s_t = \mathcal{T}(\theta)s_{t-1} + \mathcal{R}(\theta)\epsilon_t, \quad (42)$$

which summarizes the evolution of the states s_t , and the measurement equation:

$$y_t = \mathcal{Z}(\theta)s_t + \mathcal{D}(\theta), \quad (43)$$

which maps the states onto the vector of observables y_t , where $\mathcal{D}(\theta)$ represents the vector of steady state values for these observables. Specifically, the SW model is estimated based on seven quarterly macroeconomic time series. The measurement equations for real output, consumption, investment, and real wage growth, hours, inflation, interest rates and long-run inflation expectations are given by:

$$\begin{aligned} \text{Output growth} &= \gamma + 100(y_t - y_{t-1} + z_t) \\ \text{Consumption growth} &= \gamma + 100(c_t - c_{t-1} + z_t) \\ \text{Investment growth} &= \gamma + 100(i_t - i_{t-1} + z_t) \\ \text{Real Wage growth} &= \gamma + 100(w_t - w_{t-1} + z_t), \\ \text{Hours} &= \bar{l} + 100l_t \\ \text{Inflation} &= \pi_* + 100\pi_t \\ \text{FFR} &= R_* + 100R_t \end{aligned} \quad (44)$$

where all variables are measured in percent, where π_* and R_* measure the steady state level of net inflation and short term nominal interest rates, respectively and where \bar{l} captures the mean of hours (this variable is measured as an index).

Appendix A.1 provides further details on the data. In our benchmark specification we use data from 1964Q4 to 2011Q1, but we also consider a shorter sample in which end the sample in 2004Q4, so that we exclude the Great Recession. Table 1 shows the priors for the DSGE model parameters, which coincide with those used in Smets and Wouters (2007).

4 Results

4.1 Evidence of Fat Tails

In the introduction we showed that shocks extracted from standard Gaussian estimation are sometimes quite large, up to four standard deviations in size. In this section we consider more formal evidence against Gaussianity. Specifically, this section addresses two questions. First, do we still need fat tailed shocks once we allow for low frequency movements in volatility? Second, if so, which shocks are fat tailed, and how fat are the tails? We address these questions by: i) assessing the improvement in fit obtained by allowing for Student- t distributed shocks relative to both the standard model as well as models with low frequency movements in the volatility; ii) presenting the posterior distribution of the degrees of freedom for each shock.

Before we delve into the results, we provide an intuitive description of the relationship between the degrees of freedom of the Student- t distribution and the likelihood of observing large shocks. Recall from equation (3) that a Student- t distributed shock ε_t can be decomposed as

$$\varepsilon_t = \sigma_t \tilde{h}_t^{-1/2} \eta_t,$$

where η_t is drawn from a standard Gaussian distribution, and \tilde{h}_t is . Therefore, given σ_t , the chances of observing a very large ε_t depend on the chances of \tilde{h}_t being small. The prior for \tilde{h}_t is given by (5):

$$\lambda \tilde{h}_t \sim \chi^2(\lambda).$$

If λ is high, this prior concentrates around one, and the likelihood of observing large shocks is slim (the $\lambda \rightarrow \infty$ limit represents Gaussianity). As λ drops, the distribution of \tilde{h}_t spreads out and the chances of observing a low \tilde{h}_t increase. The following table provides a quantitative feel for what different λ s imply in terms of the model's ability to generate fat tailed shocks. Specifically, the table shows the number of shocks larger (in absolute value) than x standard deviations per 200 periods, which is the size of our sample. The table shows that even with 9 degrees of

freedom the chances of seeing even a single shock as large as those shown in Figure 1 are not high (.28 for shocks larger than 4 standard deviations), and become negligible for 15 or more degrees of freedom.

$\lambda, x:$	3	4	5
∞	.54	.012	$1e^{-4}$
15	1.14	.13	.02
9	1.57	.28	.06
6	2.08	.54	.17

In what follows, we consider three different Gamma priors of the form (6) for the degrees of freedom parameter λ , which capture different *a priori* views on the importance of fat tails. The first prior, $\underline{\lambda} = 15$, captures the view that the world is not quite Gaussian, but not too far from Gaussianity either. The second prior, $\underline{\lambda} = 9$, embodies the idea that the world is quite far from Gaussian, yet not too extreme. The last prior, $\underline{\lambda} = 6$, implies prior belief in a model with somewhat heavy tails.

The tightness of these priors depends on the degrees of freedom parameter $\underline{\nu}$ in equation (6), which we set equal to 4.⁴ Figure 2 shows the three priors for $\underline{\lambda} = 6, 9, \text{ and } 15$. Since the variance of the prior is $\underline{\lambda}^2/\underline{\nu}$, lower values of $\underline{\lambda}$ correspond to tighter prior. Note however that the prior with higher variance ($\underline{\lambda} = 15$) puts most of the mass on high values of the degrees of freedom. For instance, the prior probability put on the regions $\{\lambda < 6\}$ and $\{\lambda < 4\}$ by the $\underline{\lambda} = 15$ prior is less than 8 and 2.5%, respectively.

With the description of the prior in hand, we are now ready to discuss our evidence on the importance of fat tails.⁵ Table 2 shows the log marginal likelihood—the standard measure of fit in a Bayesian framework—for models with different assumptions on the shocks distribution.⁶ We consider four different combinations: i) Gaussian shocks with constant volatility (baseline), ii) Gaussian shocks with time-varying volatility, iii) Student-*t* distributed shocks with constant volatility, and

⁴In an appendix not for publication we also report the results for $\underline{\nu} = 1$, which are very similar.

⁵Our results are based on 4 chains, each beginning from a different starting point, of 220K draws each, of which we discard the first 20K draws. We checked for convergence across chains.

⁶For details on the computation of these marginal likelihoods, see the appendix.

iv) both Student- t shocks and time-variation in volatility.⁷ The Gaussian-constant volatility model is clearly rejected by the data. This is not surprising in light of the evidence contained in Justiniano and Primiceri (2008), Fernández-Villaverde and Rubio-Ramírez (2007), and Liu et al. (2011). What is perhaps more surprising is that the marginal likelihood indicates that if one had to choose between the two features, a random walk in volatility (SV) or Student- t shocks, the data unequivocally point to the latter feature as the more important: the difference in marginal likelihood between the SV and the best fitting model with student- t shocks is very large, about 71 log points. Of course one does not have to choose between the two features, and the marginal likelihoods indicate that even after accounting for fat-tailed shocks, allowing for time-variation in the volatility improves fit: for any row in Table 2, the marginal likelihood increases moving from the left (no SV) to the right (SV) column. From the perspective of this paper, however, the main finding is provided by the fact that the numbers increase moving down the rows of Table 2, regardless of the column. In summary, the data strongly favor Student- t distributed shocks with a non-negligible degree of fat tails, whether or not we allow for low frequency movements in volatility.

Our second piece of evidence comes from the posterior distribution of the degrees of freedom λ . Table 4 shows the posterior mean and 90% bands for the degrees of freedom for each shock, in the specifications with and without stochastic volatility. Two results emerge from Table 4. First, for quite a few shocks the estimated degrees of freedom are small, even when we allow for low frequency movements in volatility. Second, allowing for low frequency movements in volatility changes substantially the inference about the degrees of freedom, implying that this feature is necessary for a proper assessment of how fat-tailed macroeconomic shocks are.

⁷For the stochastic volatility model we use an \mathcal{IG} prior on ω_q^2 (see equation (9)) with parameters $\underline{\omega}^2 = (.01)^2$ (following Justiniano and Primiceri (2008)) and $\nu_\omega = .1$. It is well known that when $\underline{\omega}^2$ is very low, as is the case here, designing a prior that is not too informative is challenging. In Montecarlo experiments we found that using very low degrees of freedom ($\nu_\omega = .1$) led to a good performance whenever the true value of ω^2 was 10^{-4} , 10^{-3} , or 10^{-2} , and therefore settled on this value.

As for the first result, Table 4 shows that four shocks have a mean below 6 for the best fitting prior on the degrees of freedom according to the marginal likelihood criterion ($\underline{\lambda} = 6$). For priors with higher $\underline{\lambda}$ the posterior mean degrees of freedom for these four shocks increases, but this is mostly because the posterior distribution becomes more skewed to the right (that is, it places some mass on higher values for λ). Still, for many shocks the posterior distribution puts sizable mass on the $\{\lambda < 6\}$ region even for $\underline{\lambda} = 15$. It is also worth noting that in four cases the 5th quintile of the posterior distribution barely changes as a function of $\underline{\lambda}$. The shocks with the fattest tails (lowest posterior degrees of freedom) are those affecting the discount rate (b), TFP productivity (z), the marginal efficiency of investment (μ), and the wage markup (λ_w). Not surprisingly, these shocks are the usual suspects as key drivers of business cycles (see Smets and Wouters (2007), Justiniano et al. (2009)).

As for the result that allowing for low frequency movements in volatility substantially changes the inference about the degrees of freedom, the monetary policy shock r^m is a case in point. Its estimated degrees of freedom are very low when one ignores time variations in volatility that seem to be apparent in figure 1. Ignoring time-variation in volatility, the model interprets the large shocks of the late seventies/early eighties as evidence of fat tails. Once these secular changes in volatility are taken into account, the posterior estimates of the degrees of freedom increases substantially.⁸ What is the intuition for this finding? The posterior distribution of \tilde{h}_t , which determines how fat the tails are, is given by (20):

$$[\lambda + \varepsilon_t^2/\sigma_t^2] \tilde{h}_t | \theta, \varepsilon_{1:T}, \tilde{\sigma}_{1:T}, \lambda \sim \chi^2(\lambda + 1).$$

⁸In an appendix available upon request we show the smoothed shocks and the “tamed” version of these shocks (that is, the counterfactual shocks after shutting down the Student- t component – see below) for both the discount rate and policy shocks for the estimation without stochastic volatility. Consistent with the above analysis, the “tamed” plots look much more consistent with a Gaussian distribution than the original estimates, confirming the role of fat tails. However, for the policy shock, we find that there is still a cluster of higher variance innovations in the late 70s and early 80s, suggesting that stochastic volatility still has an important role in the evolution of this shock.

For a given value of σ_t , the a larger estimated shock ε_t implies a smaller posterior value of \tilde{h}_t . Not surprisingly, large shocks are interpreted by the model as evidence for fat tails. However, the shock is standardized by σ_t : if a large shock occurs during a period where all shocks tend to be large, it is discounted, and the posterior value of \tilde{h}_t may not be particularly small.

A question of obvious interest is whether this evidence in favor of fat-tailed shocks depends on the Great Recession being part of the estimation sample. To address this, we estimate the model for the sub-sample ending in the fourth quarter of 2004—the same sample used in [Justiniano and Primiceri \(2008\)](#). Table 3 shows the marginal likelihood for all the specifications considered above but estimated on the shorter sub-sample. The results for this specification, displayed in Table 3, are in line the results for the full sample: having Student- t distributed shocks improves fit, regardless of whether we also consider stochastic volatility, and the lower the prior mean for the degrees of freedom the higher the marginal likelihood we obtain. The posterior means of the degrees of freedom of the Student- t distribution (not shown) are also in line with those presented in Table 4.

4.2 Large Shocks and Macroeconomic Fluctuations

We have shown that quite a few important shocks in the SW model have fat tails (i.e., their estimated degrees of freedom are low). But what does this mean in terms of business cycle fluctuations? This section tries to provide a quantitative answer to this question by performing a counterfactual experiment. Recall again from equation (3) that

$$\varepsilon_{q,t} = \sigma_{q,t} \tilde{h}_{q,t}^{-1/2} \eta_{q,t}.$$

Therefore, once we compute the posterior distribution of $\varepsilon_{q,t}$ (the smoothed shocks) and $\tilde{h}_{q,t}$, we can purge the Student- t component from $\varepsilon_{q,t}$ using

$$\tilde{\varepsilon}_{q,t} = \sigma_{q,t} \eta_{q,t}.$$

We can then compute counterfactual histories that would have occurred had the shocks been $\tilde{\varepsilon}_{q,t}$ instead of $\varepsilon_{q,t}$. All these counterfactuals are computed for the best

fitting model – that with stochastic volatility and Student- t shocks, with the prior for λ centered at 6.

The left panel of Figure 3 shows these counterfactual histories for output, consumption growth, and hours. For all plots the pink solid lines are the median counterfactual paths, the pink dashed lines represent the 90% bands, and the solid black lines represent the actual data. The right panel uses actual and counterfactual histories to compute a rolling window standard deviation, where each window contains the prior 20 quarters as well as the following 20 quarters, for a total of 41 quarters. These rolling window standard deviations are commonly used measures of time-variation in the volatility of the series. The difference between actual and counterfactual standard deviations measures the extent to which the change in volatility is accounted for by fat-tailed shocks.⁹

The left panels show that fat-tailed shocks seem to account for a non negligible part of fluctuations in two of the variables of interest, output and consumption growth. For output growth, the Student- t component accounted for a sizable fraction of the contraction in output growth during the Great Recession. In particular, if the fat tail component were absent the Great Recession would be of about the same size as more mild recessions, such as the 1990-91 recession. In general, without the fat-tailed component of the shocks all recessions (with the exception of the 2001 recession) would be of roughly the same magnitude. Further, the rolling window standard deviation shown in the right panel shows that the Student- t component explains a non-negligible part of changes in the realized volatility in the data. One can interpret this evidence as saying that the 70s and early 80s were more volatile than the Great Moderation period at least in part because rare shocks took place. For example at the peak of the volatility in 1978, the data standard deviation of output is about 1.25, but once we shut down the Student- t component it drops to 1.05, which is a reduction of about 16% in volatility. Similar conclusions apply to the decomposition of consumption and investment growth (middle and bottom

⁹The distribution of $\tilde{h}_{q,t}$ is non-time varying. However, since large shocks rarely occur, they may account for changes in the rolling window volatility.

panels, respectively). It is notable that in the case of investment, and to a lesser extent of consumption, the rolling-window volatility computed from the data has spiked up recently to levels near those prior to the Great Moderation. When we take the Student- t component into account, however, the recent increase in the rolling-window volatility appears much milder.

4.3 Student- t Shocks and Inference about Time-Variation in Volatility

This section discusses the extent to which accounting for fat tails makes us reevaluate the magnitude of low frequency changes in volatility. Why might this be the case? Inference about the stochastic volatility is conducted using state-space methods, where equation (48) in the appendix, which we repeat here for convenience,

$$\log(\sigma^{-2}\tilde{h}_t\varepsilon_t^2 + c) = 2\tilde{\sigma}_t + \eta_t^*,$$

is the measurement equation (with c a small constant), and equation (8) is the transition equation. Intuitively, the estimated time-varying volatilities will try to fit the time series $\log(\sigma^{-2}\tilde{h}_t\varepsilon_t^2 + c)$. Since this quantity depends on ε_t^2 , the model will interpret changes over time in the size of the squared shocks ε_t^2 as evidence of time variation in the volatilities $\tilde{\sigma}_t$. In a world with fat tails, ε_t^2 will vary over time simply because \tilde{h}_t changes. In order to take this into account, in the expression above ε_t^2 is pre-multiplied by \tilde{h}_t . If one ignores variations in \tilde{h}_t by assuming Gaussian shocks, one may obtain the wrong inference about the time variation in the σ_t s. For instance, one may conclude that the Great Recession signals a permanent change in the level of macroeconomic volatility, when in fact it may (at least in part) be the result of a particularly large realization of the shocks.

Does all of this matter in practice? An implication of stochastic volatility is that the model-implied variance of the endogenous variables changes over time (see Figure 5 of [Justiniano and Primiceri \(2008\)](#)). Therefore, rather than looking at the posterior estimates of the stochastic volatility component for individual shocks

(which we show in an appendix available upon request) we focus here on the time-variation of the standard deviation of output and consumption—two objects that are directly relevant to macroeconomists. Specifically, Figure 4 shows the model-implied volatility of output and consumption growth, as measured by the unconditional standard deviation of the series computed at each point in time t assuming that the standard deviations of the shocks is going to remain equal to the estimated value of $\sigma_q \sigma_{q,t}$ forever after (that is, abstracting from the fact that future $\sigma_{q,t}$ s will evolve according to equation (8); this is the same objects computed in Figure 5 of Justiniano and Primiceri (2008)). In the top panel, the red line shows this measure for the estimation with stochastic volatility but Gaussian shocks, while the black lines show this volatility for the estimation with both stochastic volatility and Student- t components (with $\lambda = 6$). As in the other plots, the solid line is the posterior median and the dashed lines correspond to the 90% bands around the median. For both variables the model-implied volatility is generally higher when we account for fat tails. This may be intuitive: in a model that allows for fat-tailed shocks the implied volatility of output and consumption can be higher. However, the difference between the models is not constant over time. At the peak of the high-volatility period (late 70s and early 80s), the two models agree. However, during the Great Moderation the model that does not allow for fat tailed shocks seems to overestimate the decline in macroeconomic volatility.

The middle panel of Figure 4 hones in on this finding. This panel shows the posterior distribution of the ratio of the volatility in 1981 (roughly, the peak of the volatility series) relative to the volatility in 1994 (roughly, the bottom) for output and consumption growth, respectively. The red bars are for the model with stochastic volatility only, while the black bars stand for the model with stochastic volatility and Student- t shocks. Numbers greater than one indicate that volatility was higher in 1981 relative to 1994. There is no doubt that this is the case, as both histograms are well to the right of one. However, the magnitude of the decrease in volatility depends on whether or not we allow for fat-tailed shocks. The posterior distribution for the ratio of the output growth volatility in 1981 relative to the

volatility in 1994 in the estimation with Student- t shocks is smaller relative to the case with Gaussian shocks. The median is 1.8 in the former, compared to 2.6 in the latter. For the red bars, most of the mass is to the right of two, implying that volatility dropped by more than half between 1981 and 1994. The converse holds for the black bars, which show a much smaller decline in volatility according to the model with Student- t shocks. This same pattern is also evident for the consumption growth.

As a result of the Great Recession there has been an increase in volatility in many macroeconomic variables since 2008, as measured by the rolling window standard deviations shown in Figure 3. To what extent does this increase reflect a permanent increase in the volatility of the underlying shocks, and, potentially, the end of the Great Moderation? The bottom panels of Figure 4 show the ratio of the volatility in 2011 (end of the sample) relative to the volatility in 2005 (pre-Great Recession) for output and consumption growth, respectively, with numbers greater than one indicating a permanent rise in volatility. Under the model with time-varying volatility and Gaussian shocks, the probability that volatility in both output and consumption has increased after the Great Recession is quite high. The probability of the ratio being below one is 4.6% and 8.2% for output and consumption growth, respectively. The model that has, Student- t shocks in addition to time-varying volatility is less sure: the probability of the ratio being below one increases to 12% in the case of output, and 21% in the case of consumption growth. Moreover, this model implies that if such an increase took place, it was fairly modest, with most of the mass below 1.25.

5 Conclusions

We provide strong evidence that the Gaussianity assumption in DSGE models is counterfactual, even after allowing for low frequency changes in the volatility of shocks. However, it is important to point out a number of caveats regarding our analysis. First, in the current draft we allow for excess kurtosis but not for skewness.

The shocks plots in in Figure 1 make it plain that most large shocks occur during recessions, implying that skewness may also be an salient feature of the shocks distribution. A recent paper by Müller (2011) describes some of the dangers associated with departures from Gaussianity when the alternative shock distribution is also misspecified. Importantly for our analysis, not allowing for skewness may lead to an underestimation of the importance of fat tails during recessions, as we only estimate the “average” amount of kurtosis. Second, we allow for permanent (random walk) and i.i.d (Student- t distribution) changes in the variance of the shocks. These assumptions are convenient, but also extreme. Our main point is that together with low frequency changes in the standard deviation of shocks, there are also short run spikes in volatility. So far, the literature for the U.S. has mainly focused on the former phenomenon; in this paper we emphasize the latter. Still, in future research it may be important to relax the assumption that these short run spikes are identically distributed over time. Finally, in order to study the full implications of fat tailed shocks on the macroeconomy we need to use non linear models, as we discuss in the introduction. We leave these important extensions for future research.

References

- AN, S. AND F. SCHORFHEIDE (2007): “Bayesian Analysis of DSGE Models—Rejoinder,” *Econometric Reviews*, 26, 211 – 219.
- ASCARI, G., G. FAGIOLO, AND A. ROVENTINI (2012): “Fat Tails Distributions and Business Cycle Models,” *Manuscript, OFCE-Science-Po*.
- CHIB, S. AND E. GREENBERG (1994): “Bayes inferences in regression models with ARMA(p,q) errors,” *Journal of Econometrics*, 64, 183–206.
- CHIB, S. AND S. RAMAMURTHY (2011): “DSGE Models with Student-t errors,” *mimeo, Washington University in St. Louis*.
- CHRISTIANO, L., M. EICHENBAUM, AND C. L. EVANS (2005): “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, 113, 1–45.
- CHRISTIANO, L. J. (2007): “Comment on ‘On the Fit of New Keynesian Models’ by Del Negro ,Schorfheide, Smets ,Wouters,” *Journal of Business and Economic Statistics*, 25, 143–151.
- DEL NEGRO, M. AND F. SCHORFHEIDE (2012): “DSGE Model-Based Forecasting,” *FRBNY Working Paper*.
- DURBIN, J. AND S. J. KOOPMAN (2002): “A Simple and Efficient Simulation Smoother for State Space Time Series Analysis,” *Biometrika*, 89, 603–616.
- EDGE, R. AND R. GÜRKAYNAK (2010): “How Useful Are Estimated DSGE Model Forecasts for Central Bankers,” *Brookings Papers of Economic Activity*, forthcoming.
- FERNÁNDEZ-VILLAYERDE, J. AND J. F. RUBIO-RAMÍREZ (2007): “Estimating Macroeconomic Models: A Likelihood Approach,” *Review of Economic Studies*, 74, 1059–1087.

- GEWEKE, J. (1993): “Bayesian Treatment of the Independent Student-t Linear Model,” *Journal of Applied Econometrics*, 8, S19–S40.
- (1994): “Priors for Macroeconomic Time Series and Their Application,” *Econometric Theory*, 10, 609–632.
- (1999): “Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication,” *Econometric Reviews*, 18, 1–126.
- (2005): *Contemporary Bayesian Econometrics and Statistics*, Wiley.
- GREENWOOD, J., Z. HERCOVITZ, AND P. KRUSELL (1998): “Long-Run Implications of Investment-Specific Technological Change,” *American Economic Review*, 87, 342–36.
- JUSTINIANO, A. AND G. PRIMICERI (2008): “The Time-Varying Volatility of Macroeconomic Fluctuations,” *American Economic Review*, 98, 604 – 641.
- JUSTINIANO, A., G. E. PRIMICERI, AND A. TAMBALOTTI (2009): “Investment Shocks and Business Cycles,” *NBER Working Paper*, 15570.
- KIM, S., N. SHEPHARD, AND S. CHIB (1998): “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *Review of Economic Studies*, 65, 361–393.
- LIU, Z., D. F. WAGGONER, AND T. ZHA (2011): “Sources of macroeconomic fluctuations: A regime-switching DSGE approach,” *Quantitative Economics*, 2, 251–301.
- MÜLLER, U. K. (2011): “Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix,” *Mimeo, Princeton University*.
- SIMS, C. A. (2002): “Solving Linear Rational Expectations Models,” *Computational Economics*, 20(1-2), 1–20.
- SMETS, F. AND R. WOUTERS (2003): “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,” *Journal of the European Economic Association*, 1, 1123 – 1175.

——— (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 97, 586 – 606.

A Appendix

A.1 Data

The data set is obtained from Haver Analytics (Haver mnemonics are in italics). We compile observations for the variables that appear in the measurement equation (44). Real GDP (*GDPC*), the GDP price deflator (*GDPDEF*), nominal personal consumption expenditures (*PCEC*), and nominal fixed private investment (*FPI*) are constructed at a quarterly frequency by the Bureau of Economic Analysis (BEA), and are included in the National Income and Product Accounts (NIPA).

Average weekly hours of production and nonsupervisory employees for total private industries (*PRS85006023*), civilian employment (*CE16OV*), and civilian noninstitutional population (*LNSINDEX*) are produced by the Bureau of Labor Statistics (BLS) at the monthly frequency. The first of these series is obtained from the Establishment Survey, and the remaining from the Household Survey. Both surveys are released in the BLS Employment Situation Summary (ESS). Since our models are estimated on quarterly data, we take averages of the monthly data. Compensation per hour for the nonfarm business sector (*PRS85006103*) is obtained from the Labor Productivity and Costs (LPC) release, and produced by the BLS at the quarterly frequency. Last, the federal funds rate is obtained from the Federal Reserve Board's H.15 release at the business day frequency, and is not revised. We take quarterly averages of the annualized daily data.

All data are transformed following [Smets and Wouters \(2007\)](#). Specifically:

$$\begin{aligned}
 \textit{Output growth} &= \textit{LN}((\textit{GDPC})/\textit{LNSINDEX}) * 100 \\
 \textit{Consumption growth} &= \textit{LN}((\textit{PCEC}/\textit{GDPDEF})/\textit{LNSINDEX}) * 100 \\
 \textit{Investment growth} &= \textit{LN}((\textit{FPI}/\textit{GDPDEF})/\textit{LNSINDEX}) * 100 \\
 \textit{Real Wage growth} &= \textit{LN}(\textit{PRS85006103}/\textit{GDPDEF}) * 100 \\
 \textit{Hours} &= \textit{LN}((\textit{PRS85006023} * \textit{CE16OV}/100)/\textit{LNSINDEX}) * 100 \\
 \textit{Inflation} &= \textit{LN}(\textit{GDPDEF}/\textit{GDPDEF}(-1)) * 100 \\
 \textit{FFR} &= \textit{FEDERAL FUNDS RATE}/4
 \end{aligned}$$

A.2 Marginal likelihood

The marginal likelihood is the marginal probability of the observed data, and is computed as the integral of (12) with respect to the unobserved parameters and latent variables:

$$\begin{aligned}
p(y_{1:T}) &= \int p(y_{1:T}|s_{1:T}, \theta) p(s_{1:T}|\varepsilon_{1:T}, \theta) p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) \\
&\quad p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) p(\tilde{\sigma}_{1:T}|\omega_{1:\bar{q}}^2) p(\lambda_{1:\bar{q}}) p(\omega_{1:\bar{q}}^2) p(\theta) \\
&\quad d(s_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, \theta), \\
&= \int p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) p(\tilde{\sigma}_{1:T}|\omega_{1:\bar{q}}^2) \\
&\quad p(\lambda_{1:\bar{q}}) p(\omega_{1:\bar{q}}^2) p(\theta) d(\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, \theta)
\end{aligned} \tag{45}$$

where the quantity

$$\begin{aligned}
p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) &= \int p(y_{1:T}|s_{1:T}, \theta) p(s_{1:T}|\varepsilon_{1:T}, \theta) \\
&\quad p(\varepsilon_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) \cdot d(s_{1:T}, \varepsilon_{1:T})
\end{aligned}$$

is computed at step 1a of the Gibb-sampler described above.

We obtain the marginal likelihood using Geweke (1999)'s modified harmonic mean method. If $f(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)$ is any distribution with support contained in the support of the posterior density such that

$$\int f(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2) \cdot d(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2) = 1,$$

it follows from the definition of the posterior density that:

$$\begin{aligned}
\frac{1}{p(y_{1:T})} &= \int \frac{f(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)}{p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta) p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) p(\tilde{\sigma}_{1:T}|\omega_{1:\bar{q}}^2) p(\lambda_{1:\bar{q}}) p(\omega_{1:\bar{q}}^2) p(\theta)} \\
&\quad p(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2 | y_{1:T}) \cdot d(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)
\end{aligned}$$

We follow Justiniano and Primiceri (2008) in choosing

$$f(\theta, \tilde{h}_{1:T}) = f(\theta) \cdot p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}}) p(\tilde{\sigma}_{1:T}|\omega_{1:\bar{q}}^2) p(\lambda_{1:\bar{q}}) p(\omega_{1:\bar{q}}^2), \tag{46}$$

where $f(\theta)$ is a truncate multivariate distribution as proposed by Geweke (1999).

Hence we approximate the marginal likelihood as:

$$\hat{p}(y_{1:T}) = \left[\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \frac{f(\theta^j)}{p(y_{1:T}|\tilde{h}_{1:T}^j, \tilde{\sigma}_{1:T}^j, \theta^j) p(\theta^j)} \right]^{-1} \tag{47}$$

where θ^j , $\tilde{h}_{1:T}^j$, and $\tilde{\sigma}_{1:T}^j$ are draws from the posterior distribution, and n_{sim} is the total number of draws. We are aware of the problems with (46), namely that it does not ensure that the random variable

$$\frac{f(\theta, \tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2)}{p(y_{1:T}|\tilde{h}_{1:T}, \tilde{\sigma}_{1:T}, \theta)p(\tilde{h}_{1:T}|\lambda_{1:\bar{q}})p(\tilde{\sigma}_{1:T}|\omega_{1:\bar{q}}^2)p(\lambda_{1:\bar{q}})p(\omega_{1:\bar{q}}^2)p(\theta)}$$

has finite variance. Nonetheless, like Justiniano and Primiceri (2008) we found that this method delivers very similar results across different chains.

A.3 Drawing the stochastic volatilities

We draw the stochastic volatilities using the procedure in Kim et al. (1998), which we briefly describe. Taking squares and then logs of (3) one obtains:

$$\varepsilon_{q,t}^* = 2\tilde{\sigma}_{q,t} + \eta_{q,t}^* \tag{48}$$

where

$$\varepsilon_{q,t}^* = \log(\sigma_q^{-2}\tilde{h}_{q,t}\varepsilon_{q,t}^2 + c), \tag{49}$$

$c = .001$ being an offset constant, and $\eta_{q,t}^* = \log(\eta_{q,t}^2)$. If $\eta_{q,t}^*$ were normally distributed, $\sigma_{q,1:T}$ could be drawn using standard methods for state-space systems. In fact, $\eta_{q,t}^*$ is distributed as a $\log(\chi_1^2)$. Kim et al. (1998) address this problem by approximating the $\log(\chi_1^2)$ with a mixture of normals, that is, expressing the distribution of $\eta_{q,t}^*$ as:

$$p(\eta_{q,t}^*) = \sum_{k=1}^K \pi_k^* \mathcal{N}(m_k^* - 1.2704, \nu_k^{*2}) \tag{50}$$

The parameters that optimize this approximation, namely $\{\pi_k^*, m_k^*, \nu_k^*\}_{k=1}^K$ and K , are given in Kim et al. (1998). Note that these parameters are independent of the specific application. The mixture of normals can be equivalently expressed as:

$$\eta_{q,t}^* | s_{q,t} = k \sim \mathcal{N}(m_k^* - 1.2704, \nu_k^{*2}), \Pr(s_{i,t} = k) = \pi_k^*. \tag{51}$$

Hence step (4) of the Gibbs sampler actually consists in two steps:

- (4.1) Draw from $p(\varsigma_{1:T}|\tilde{\sigma}_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, s_{1:T}\lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$ using (50) for each q . Specifically:

$$Pr\{\varsigma_{q,t} = k|\tilde{\sigma}_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T} \dots\} \propto \pi_k^* \nu_k^{-1} \exp\left[-\frac{1}{2\nu_k^*}(\eta_{q,t}^* - m_k^* + 1.2704)^2\right]. \quad (52)$$

where from (48) $\eta_{q,t}^* = \varepsilon_{q,t}^* - 2\tilde{\sigma}_{q,t}$.

- (4.2) Draw from $p(\tilde{\sigma}_{1:T}|\varsigma_{1:T}, \varepsilon_{1:T}, \tilde{h}_{1:T}, s_{1:T}\lambda_{1:\bar{q}}, \rho_{1:\bar{q}}, \omega_{1:\bar{q}}^2, y_{1:T})$ using Durbin and Koopman (2002), where (48) is the measurement equation and (8) is the transition equation.

Note that in principle we should make it explicit that we condition on $\varsigma_{1:T}$ in the other steps of the Gibbs sampler as well. In practice, all other conditional distributions do not depend on $\varsigma_{1:T}$, hence we omit the term for simplicity.

Table 1: Priors for the Medium-Scale Model

	Density	Mean	St. Dev.		Density	Mean	St. Dev.
<i>Policy Parameters</i>							
ψ_1	Normal	1.50	0.25	ρ_R	Beta	0.75	0.10
ψ_2	Normal	0.12	0.05	ρ_{r^m}	Beta	0.50	0.20
ψ_3	Normal	0.12	0.05	σ_{r^m}	InvG	0.10	2.00
<i>Nominal Rigidities Parameters</i>							
ζ_p	Beta	0.50	0.10	ζ_w	Beta	0.50	0.10
<i>Other “Endogenous Propagation and Steady State” Parameters</i>							
α	Normal	0.30	0.05	π^*	Gamma	0.75	0.40
Φ	Normal	1.25	0.12	γ	Normal	0.40	0.10
h	Beta	0.70	0.10	S''	Normal	4.00	1.50
ν_l	Normal	2.00	0.75	σ_c	Normal	1.50	0.37
ι_p	Beta	0.50	0.15	ι_w	Beta	0.50	0.15
r_*	Gamma	0.25	0.10	ψ	Beta	0.50	0.15
<i>$\rho_s, \sigma_s, \text{ and } \eta_s$</i>							
ρ_z	Beta	0.50	0.20	σ_z	InvG	0.10	2.00
ρ_b	Beta	0.50	0.20	σ_b	InvG	0.10	2.00
ρ_{λ_f}	Beta	0.50	0.20	σ_{λ_f}	InvG	0.10	2.00
ρ_{λ_w}	Beta	0.50	0.20	σ_{λ_w}	InvG	0.10	2.00
ρ_μ	Beta	0.50	0.20	σ_μ	InvG	0.10	2.00
ρ_g	Beta	0.50	0.20	σ_g	InvG	0.10	2.00
η_{λ_f}	Beta	0.50	0.20	η_{λ_w}	Beta	0.50	0.20
η_{gz}	Beta	0.50	0.20				

Notes: Note that $\beta = (1/(1 + r_*/100))$. The following parameters are fixed in [Smets and Wouters \(2007\)](#): $\delta = 0.025$, $g_* = 0.18$, $\lambda_w = 1.50$, $\varepsilon_w = 10.0$, and $\varepsilon_p = 10$. The columns “Mean” and “St. Dev.” list the means and the standard deviations for Beta, Gamma, and Normal distributions, and the values s and ν for the Inverse Gamma (InvG) distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region. The prior for \bar{l} is $\mathcal{N}(-45, 5^2)$.

Table 2: Marginal Likelihoods

	Without Stochastic Volatility	With Stochastic Volatility
<i>Gaussian shocks</i>		
	-1117.9	-1050.9
<i>Student-t distributed shocks</i>		
$\underline{\lambda} = 15$	-999.0	-989.9
$\underline{\lambda} = 9$	-988.1	-986.0
$\underline{\lambda} = 6$	-980.0	-966.3

Notes: The parameter $\underline{\lambda}$ represents the prior mean for the degrees of freedom in the Student- t distribution λ .

Table 3: Marginal Likelihoods, Sample Ending in 2004Q4

	Constant Volatility	Stochastic Volatility
<i>Gaussian shocks</i>		
	-962.8	-926.7
<i>Student-t distributed shocks</i>		
$\underline{\lambda} = 15$	-878.4	-847.9
$\underline{\lambda} = 9$	-866.8	-842.2
$\underline{\lambda} = 6$	-853.9	-835.0

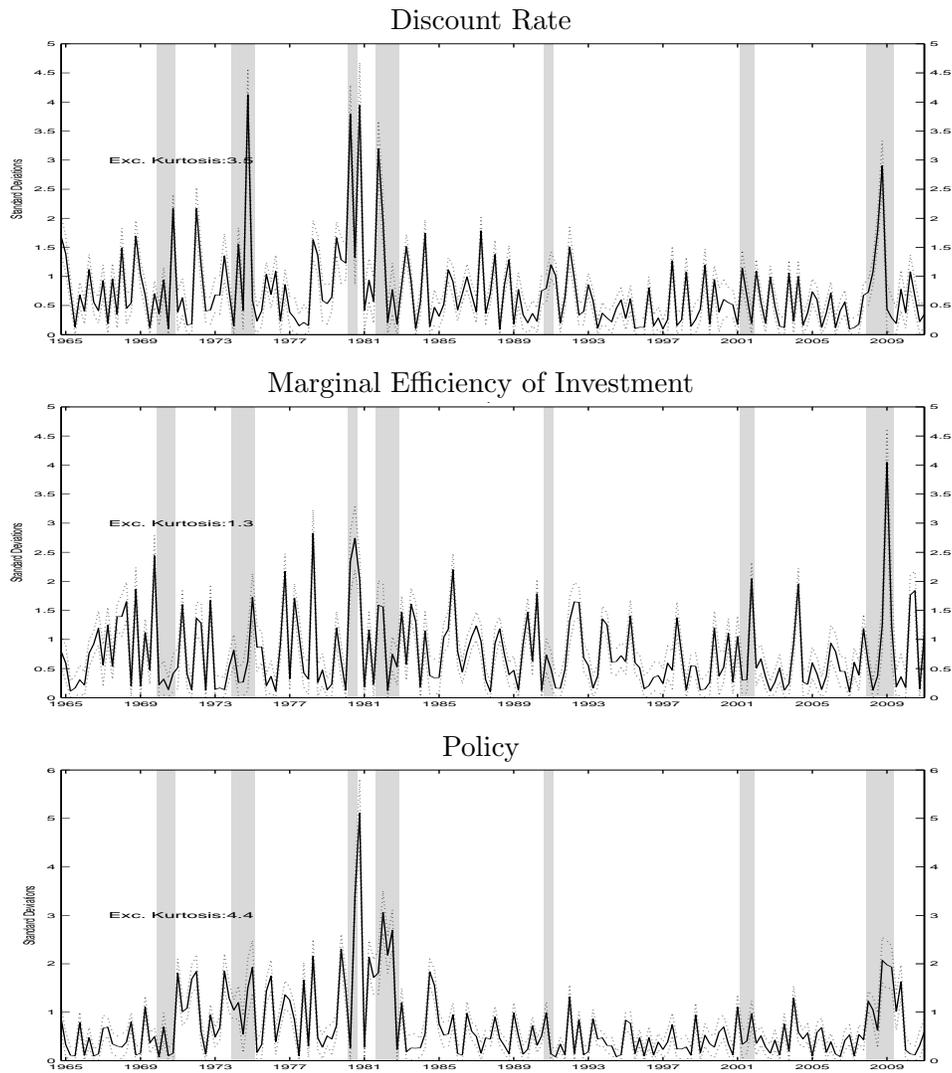
Notes: The parameter $\underline{\lambda}$ represents the prior mean for the degrees of freedom in the Student- t distribution.

Table 4: Posterior of the Student's t Degrees of Freedom

	<i>Without Stochastic Volatility</i>			<i>With Stochastic Volatility</i>		
	$\underline{\lambda} = 15$	$\underline{\lambda} = 9$	$\underline{\lambda} = 6$	$\underline{\lambda} = 15$	$\underline{\lambda} = 9$	$\underline{\lambda} = 6$
g	9.8 (3.3,16.6)	7.2 (3.1,11.4)	5.8 (2.8,8.8)	14.8 (5.5,24.0)	10.5 (4.7,16.1)	8.2 (4.1,12.2)
b	4.4 (2.3,6.4)	4.4 (2.4,6.4)	4.1 (2.3,5.8)	9.1 (3.0,15.4)	6.9 (3.0,10.9)	5.8 (2.8,8.6)
μ	9.9 (3.5,16.5)	7.1 (3.0,11.2)	5.8 (2.9,8.8)	10.2 (3.3,17.2)	7.4 (3.0,11.8)	5.9 (2.8,9.0)
z	5.9 (2.1,9.9)	4.9 (2.2,7.6)	4.2 (2.1,6.2)	7.5 (2.4,13.1)	5.7 (2.3,9.2)	4.8 (2.2,7.3)
λ_f	11.1 (3.6,18.7)	8.4 (3.5,13.2)	6.6 (3.2,10.1)	16.4 (6.3,26.1)	11.5 (5.2,17.6)	9.0 (4.6,13.4)
λ_w	9.5 (3.5,15.5)	7.2 (3.3,11.1)	6.0 (3.0,8.9)	7.5 (3.1,12.1)	6.2 (3.0,9.3)	5.3 (2.9,7.8)
r^m	3.4 (1.9,5.0)	3.2 (1.8,4.5)	3.0 (1.8,4.2)	15.1 (5.6,24.5)	10.7 (4.7,16.6)	8.4 (4.1,12.4)

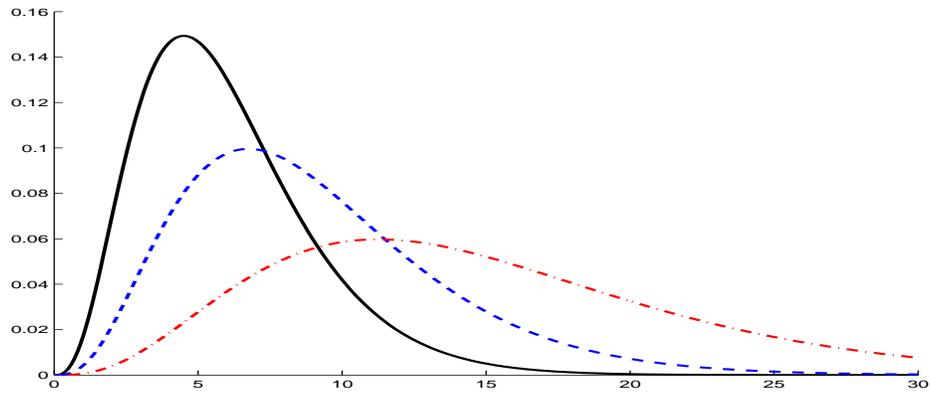
Notes: Numbers shown for the posterior mean and the 90% intervals of the degrees of freedom parameter.

Figure 1: Smoothed Shocks under Gaussianity (Absolute Value, Standardized)



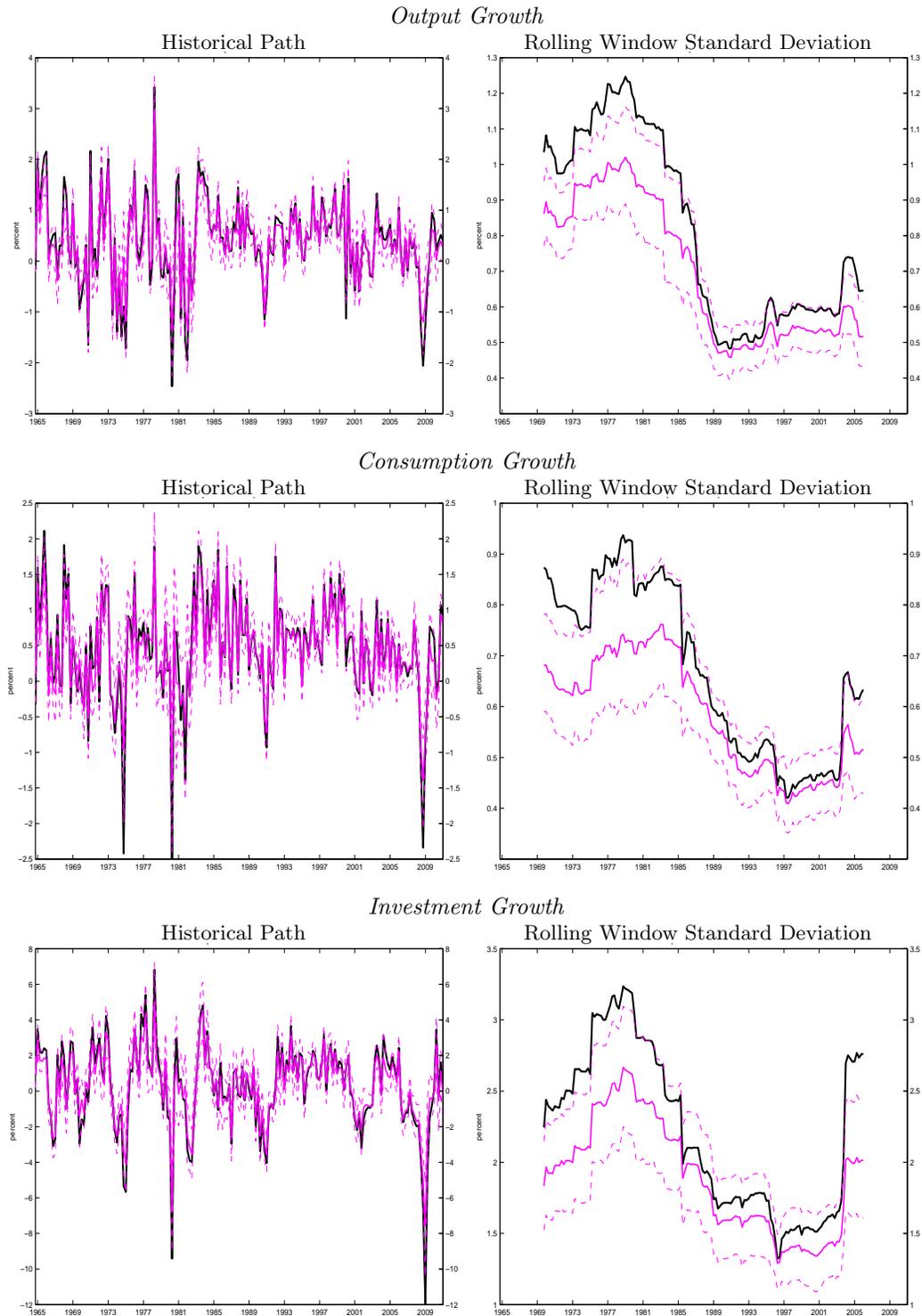
Notes: The solid line is the median, and the dashed lines are the posterior 90% bands. The vertical shaded regions identify NBER recession dates.

Figure 2: Priors on degrees of freedom of Student- t distribution



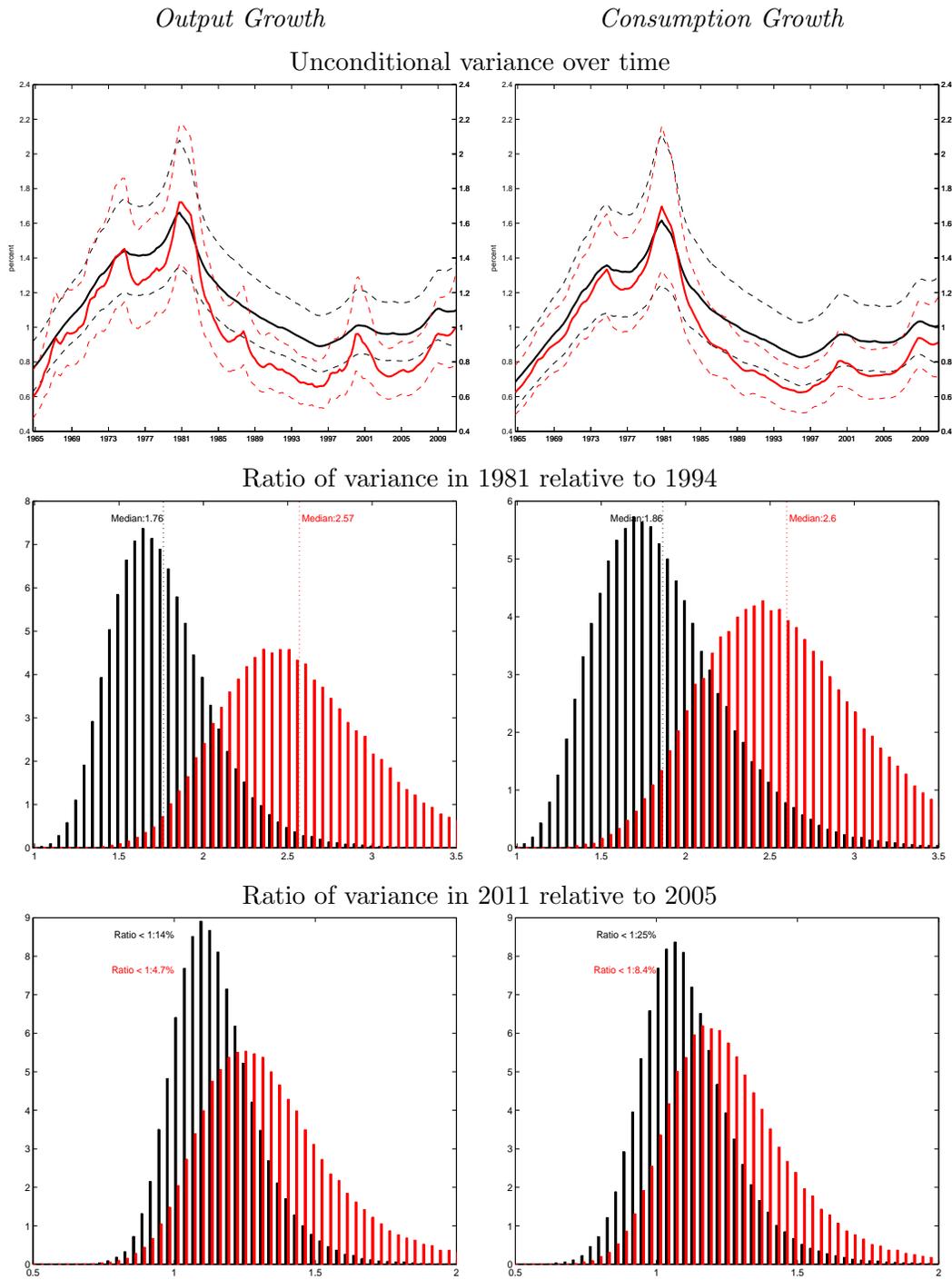
Notes: Prior density for $\lambda = 6$ (solid), 9 (dashed), and 15 (dash-and-dotted). All priors have $\nu = 4$ degrees of freedom.

Figure 3: Counterfactual evolution of output, consumption and hours worked when the Student- t distributed component is turned off, estimation with Student- t distributed shocks and stochastic volatility.



Notes: Black lines are the historical evolution of the variable, and pink lines are the median counterfactual evolution of the same variable if we shut down the Student- t distributed component of all shocks. The rolling window standard deviation uses 20 quarters before and 20 quarters after a given quarter.

Figure 4: Time-Variation in the unconditional variance of output and consumption; models estimated with and without the Student- t distributed component.



Notes: Black line in the top panel is the unconditional variance in the estimation with both stochastic volatility and Student- t components, while the red line is the unconditional variance in the estimation with stochastic volatility component only. On the middle panel the black bars correspond to the posterior histogram of the ratio of volatility in 1981 over the variance in 1994 for the estimation with both stochastic volatility and Student- t components, while the red bars are for the estimation with with stochastic volatility component only. The lower panel replicates the same analysis as in the middle panel but for the ratio of volatility in 2011 over the variance in 2005.

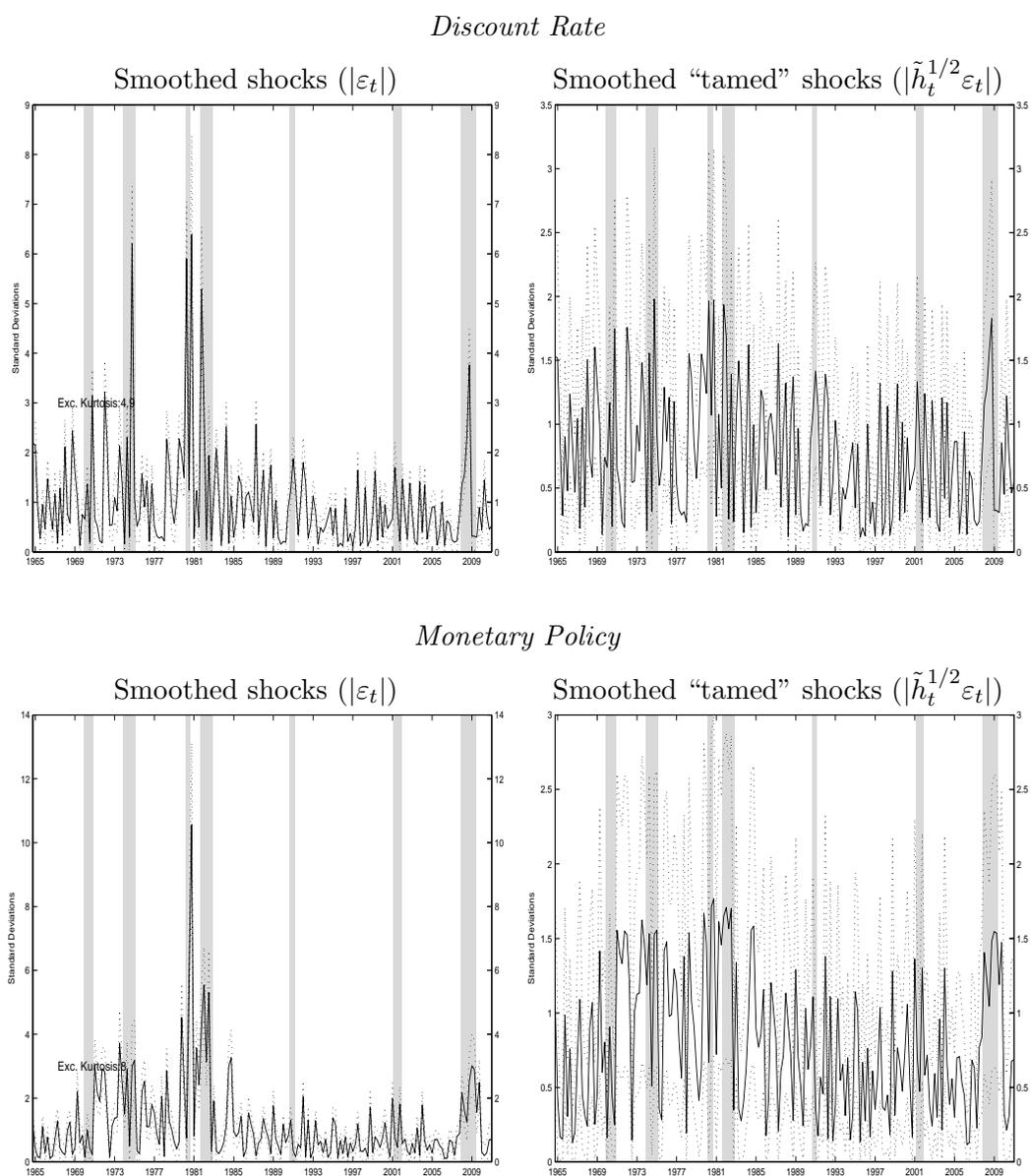
B Appendix – Not intended for publication

Table 5: Marginal Likelihoods, prior with 1 degree of freedom

	Without Stochastic Volatility	With Stochastic Volatility
<i>Student-t distributed shocks</i>		
$\underline{\lambda} = 15$	-994.2	-987.5
$\underline{\lambda} = 9$	-982.4	-974.0
$\underline{\lambda} = 6$	-975.6	-976.0

Notes: The parameter $\underline{\lambda}$ represents the prior mean for the degrees of freedom in the Student- t distribution.

Figure 5: Shocks and “Tamed” Shocks (Absolute Value, Standardized)

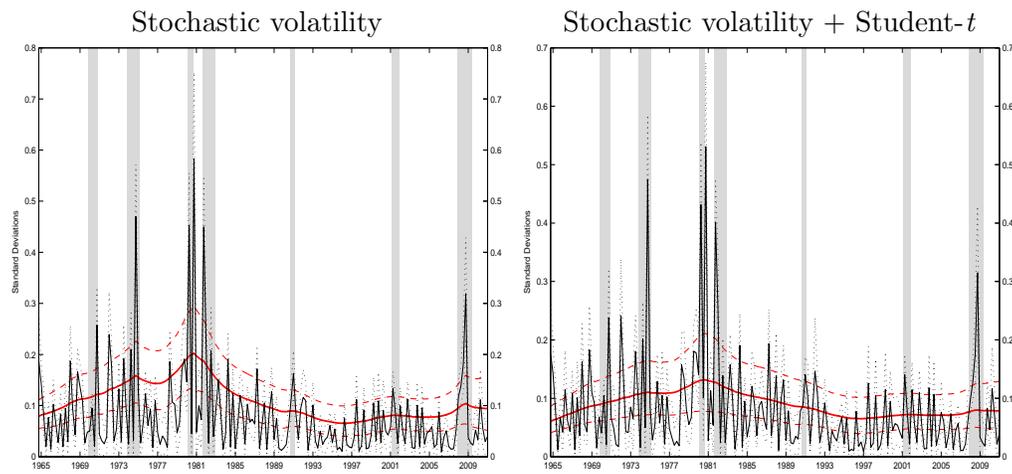


Notes: Estimation with Student- t distribution with $\lambda = 6$. The solid line is the median, and the dashed lines are the posterior 90% bands. Shocks are expressed in units of the standard deviation σ_q . The vertical shaded regions identify NBER recession dates.

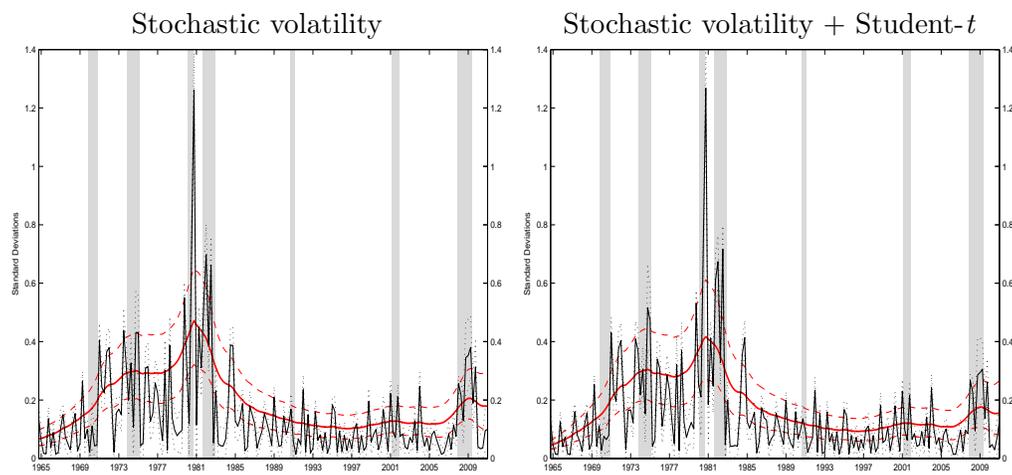
Figure 6 shows the stochastic volatility component for the discount rate and policy shocks. On the left panel we show these for the estimation with stochastic volatility and otherwise Gaussian shocks, while on the right panels we consider both stochastic volatility and Student- t shocks. The black lines correspond to the absolute value of the shocks, as in Figure 1, and the red lines correspond to the evolution of the stochastic volatility component, $\sigma_q\sigma_{q,t}$. Solid lines correspond to the median and dot/dashed lines to the 90% bands around the median. The Figure shows that when we ignore fat tails fluctuations in volatility for the discount rate shock are sizable: volatility drops by half between the early 1980s and the mid 1990s, and the decline is significant. Once we account for the fat tails movements in the volatility are more subdued, and less significant. Recall that the degrees of freedom for the discount rate shock are still quite low even after allowing for stochastic volatility. For the policy shock, in the lower panels, there is hardly any difference in the stochastic volatility component whether or not we allow for Student- t shocks – consistent with the fact that the estimated degrees of freedom of the Student- t component for this shock is high once we allow for stochastic volatility (shown in Table 4).

Figure 6: Shocks (absolute values) and smoothed stochastic volatility component, $\sigma_q \sigma_{q,t}$

Discount Rate



Monetary Policy



Notes: Estimation with Student- t distribution with $\underline{\lambda} = 15$. The solid line is the median, and the dashed lines are the posterior 90% bands. Black line is the absolute value of the shock, and the red line is the stochastic volatility component.

Table 6: Posterior Means of the DSGE Model Parameters

	Baseline	SV	St- t	St- t +SV
α	0.150	0.127	0.148	0.143
ζ_p	0.734	0.786	0.771	0.717
ι_p	0.315	0.295	0.391	0.387
Φ	1.580	1.539	1.587	1.575
S''	4.686	5.491	5.631	5.541
h	0.611	0.567	0.606	0.592
ψ	0.714	0.868	0.696	0.703
ν_l	2.088	2.327	2.438	2.535
ζ_w	0.803	0.820	0.821	0.767
ι_w	0.541	0.561	0.487	0.472
β	0.206	0.181	0.210	0.176
ψ_1	1.953	1.966	1.911	2.041
ψ_2	0.083	0.083	0.100	0.064
ψ_3	0.245	0.207	0.203	0.179
π^*	0.683	0.729	0.711	0.859
σ_c	1.236	1.097	1.251	1.288
ρ	0.835	0.856	0.869	0.856
γ	0.306	0.350	0.324	0.298
\bar{l}	-44.17	-43.29	-43.17	-49.01
ρ_g	0.977	0.993	0.982	0.978
ρ_b	0.758	0.829	0.820	0.800
ρ_μ	0.748	0.782	0.775	0.794
ρ_z	0.994	0.966	0.993	0.995
ρ_{λ_f}	0.791	0.830	0.758	0.811
ρ_{λ_w}	0.981	0.924	0.967	0.951
ρ_{rm}	0.154	0.210	0.216	0.250
σ_g	2.892	0.128	2.377	2.318
σ_b	0.125	0.075	0.077	0.058
σ_μ	0.430	0.181	0.324	0.098
σ_z	0.493	0.506	0.355	0.342
σ_{λ_f}	0.164	0.071	0.141	0.049
σ_{λ_w}	0.281	0.095	0.213	0.046
σ_{rm}	0.228	0.061	0.127	0.040
η_{gz}	0.787	0.786	0.780	0.757
η_{λ_f}	0.670	0.711	0.661	0.694
η_{λ_w}	0.948	0.867	0.924	0.868

Notes: We use a prior mean of 6 degrees of freedom for the Student- t distributed component. The stochastic volatility component assumes a prior mean for the size of the shocks to volatility of $(0.01)^2$.