Presently there is growing interest in DSGE models that have more parameters, endogenous variables, exogenous shocks, and observables than the Smets and Wouters (2007) model, and substantial additional complexities from non-Gaussian distributions and the incorporation of time-varying volatility. The popular DYNARE software package, which has proved useful for small and medium-scale models is, however, not capable of handling such models, thus inhibiting the formulation and estimation of more realistic DSGE models. Our goal in this paper is to show that the DSGE models of current interest can be efficiently estimated by the tailored random block Metropolis-Hastings (TaRB-MH) algorithm of Chib and Ramamurthy (2010) and that the marginal likelihood of such high-dimensional DSGE models can be computed reliably by the method of Chib (1995) and Chib and Jeliazkov (2001). These methods are available for wide-spread use in a user-friendly MATLAB software program that simulates the posterior distribution, reports the marginal likelihood and includes various post-estimation tools that are important for policy analysis, for example, functions for conducting impulse response and variance decomposition analyses, and point and density forecasts. Pointers on the fitting these DSGE models are given in an extended version of the new Keynesian model of Leeper, Traum and Walker (2017) that has 51 parameters, 21 endogenous variables, 8 exogenous shocks, 8 observables, as well as 1,494 non-Gaussian and nonlinear latent variables.
**Keywords**: Bayesian inference; MCMC; Metropolis-Hastings; Marginal likelihood; Tailored proposal densities; Random blocks; Student-$t$ shocks; Stochastic volatility.

**JEL Classification**: C11, C15, C32, E37, E63

1 Introduction

Over the past 20 years or so, dynamic stochastic general equilibrium (DSGE) models have become the mainstay of macroeconomic policy analysis and forecasting. Presently there is growing interest in DSGE models that have more parameters, endogenous variables, exogenous shocks, and observables than the Smets and Wouters (2007) model and substantial additional complexities from non-Gaussian distributions, as in Chib and Ramamurthy (2014) and Cúrdia, Del Negro and Greenwald (2014), and the incorporation of time-varying volatility, as in Justiniano and Primiceri (2008).\(^1\) This is because these higher-dimensional DSGE models are more realistic and have the potential to provide better statistical fit to the data. Despite widespread use of Bayesian estimation techniques, based on Markov chain Monte Carlo (MCMC) simulation methods [see Chib and Greenberg (1995) and Herbst and Schorfheide (2016) for further details about these methods], the estimation of high-dimensional DSGE models is challenging. The popular DYNARE software package, which has proved useful for small and medium-scale models is, however, currently not capable of handling the preceding DSGE models, thus inhibiting the formulation, estimation and comparison of such models for policy analysis and prediction.

Our goal in this paper is to show that the DSGE models of current interest can be efficiently estimated by the tailored random block Metropolis-Hastings (TaRB-MH) algorithm of Chib and Ramamurthy (2010). Recent applications of this algorithm to DSGE models include, e.g., Born and Pfeifer (2014), Rathke, Straumann and Woitek (2017), Kulish, Morley and Robinson (2017) and Kapetanios et al. (2019), while applications to other problems in economics include Kim and Kang (2019) and Mele (2020), amongst many others. This algorithm is effective because it breaks up the parameters into randomly grouped blocks, which are updated by Metropolis-Hastings steps based on proposal distributions that are tailored to the conditional distribution of that block. This random-grouping and tailoring is particularly effective in DSGE models where the structural parameters in the likelihood are subject to complex cross-equation restrictions of unknown form arising from the rational expectations solution of the equilibrium conditions.

Like in small-scale and medium sized DSGE models, but perhaps even more so, weak identification of the structural parameters is the norm, and the prior distribution is an important part of the model. We point to

---

\(^1\)See also, e.g., Dave and Malik (2017), Chiu, Mumtaz and Pinter (2017), Franta (2017), and Liu (2019) for macroeconomic implications of fat-tailed shocks and stochastic volatility.
additional steps in the formulation of the prior that can be helpful. For one, starting with recognized prior assumptions about structural parameters, it is important to examine (and modify) these assumptions if the implied distributions of the endogenous variables are not realistic. Second, using a portion of the initial sample to train/adjust the resulting modified prior, and adopting thick-tailed Student-t priors for location-type parameters, instead of the normal distribution family that is typically used, reduces the potential for prior-likelihood conflict and leads to more robust results. Another point is that in implementing the TaRB-MH algorithm there are payoffs to investing in the most efficient way of sampling the different blocks, for example, sampling the non-structural parameters and the latent variables by the integration sampler of Kim, Shephard and Chib (1998). This goes a long way in ensuring that the MCMC sampler traverses the parameter-space effectively.

Another goal of this paper is to introduce a user-friendly MATLAB software program for estimating DSGE models with student-t shocks and stochastic volatility. The TaRB-MH algorithm for DSGE models is thus available for wide-spread use, absent set-up costs. Amongst various summaries of the posterior distribution, the package reports the marginal likelihood by the method of Chib (1995) and Chib and Jeliazkov (2001), as modified for the TaRB-MH algorithm in Chib and Ramamurthy (2010). The package also offers various post-estimation tools that are important and useful for policy analysis, for example, functions for conducting impulse response and variance decomposition analyses, and point and density forecasts. All computations reported in this paper are generated from this package.

We do not discuss alternative methods for estimating and comparing high-dimensional DSGE models. For example, the popular random-walk M-H algorithm does not work reliably or efficiently in such settings. In addition, the Justiniano and Primiceri (2008) marginal likelihood estimation method, based on the method of Gelfand and Dey (1994), which has been applied to medium-scale DSGE model with stochastic volatility (as in Cúrdia, Del Negro and Greenwald (2014)), is subject to upward finite-sample bias, to the extent that reliance on this estimator runs the risk of misleading model comparisons [see Sims, Waggoner and Zha (2008) and Chan and Grant (2015) for such examples]. A dire consequence of this upward finite-sample bias, for instance, is that it can mistakenly favor fatter tails and incorrect time-varying variance dynamics. In contrast, we illustrate that the marginal likelihood computed by our package does not face similar problems using simulated data.

The rest of the paper is organized as follows. The next section specifies a prototypical high-dimensional DSGE model for the subsequent analysis. Section 3 provides pointers on prior formulation, posterior sampling, and model comparison accompanied by both empirical results and simulation evidence. Section 4 conducts an out-of-sample forecast analysis. Section 5 concludes. The appendix contains a detailed
summary of the implied equilibrium and steady state relations (Appendix A), a practical user guide on how to run our MATLAB package called ‘TaRB-t-SV’ (Appendix B), as well as a description of the small-scale DSGE model used in Section 4 (Appendix C).²

2 High-Dimensional DSGE Model

As a template, consider the new Keynesian model of Leeper, Traum and Walker (2017) that fills fiscal details into an otherwise standard medium-scale DSGE model presented in Christiano, Eichenbaum and Evans (2005) and Smets and Wouters (2007). To make this model more realistic, we introduce fat-tailed shocks and time-varying volatility. The resulting model is high-dimensional, consisting of 51 parameters, 21 endogenous variables, 8 exogenous shocks, 8 observables, as well as 1,494 non-Gaussian and nonlinear latent variables. This section outlines the model structure briefly to conserve space. Unless otherwise noted, we let \( \hat{x}_t \equiv \ln x_t - \ln x \) denote the log-deviation of a generic variable \( x_t \) from its steady state \( x \). We also divide a non-stationary variable \( X_t \) by the level of technology \( A_t \) and express the detrended variable as \( x_t = X_t / A_t \).

2.1 Firms

The production sector consists of firms that produce intermediate and final goods. A perfectly competitive final goods producer uses intermediate goods supplied by a continuum of intermediate goods producers indexed by \( i \) on the interval \([0,1]\) to produce the final goods. The production technology \( Y_t \leq \left( \int_0^1 Y_t(i)^{1/(1+\eta_t^p)}di \right)^{1+\eta_t^p} \) is constant-return-to-scale, where \( \eta_t^p \) is an exogenous price markup shock, \( Y_t \) is the aggregate demand of final goods, and \( Y_t(i) \) is the intermediate goods produced by firm \( i \).

Each intermediate goods producer follows a production technology \( Y_t(i) = K_t(i)^\alpha (A_t L_t^d(i))^{1-\alpha} - A_t \Omega, \) where \( K_t(i) \) and \( L_t^d(i) \) are the capital and the amount of ‘packed’ labor input rented by firm \( i \) at time \( t \), and \( 0 < \alpha < 1 \) is the income share of capital. \( A_t \) is the labor-augmenting neutral technology shock and its growth rate \( u_t^a \equiv \ln(A_t/A_{t-1}) \) equals \( \gamma > 0 \) when \( A_t \) evolves along the balanced growth path. The parameter \( \Omega > 0 \) represents the fixed cost of production.

Intermediate goods producers maximize their profits in two stages. First, they take the input prices, i.e., nominal wage \( W_t \) and nominal rental rate of capital \( R_t^k \), as given and rent \( L_t^d(i) \) and \( K_t(i) \) in perfectly competitive factor markets. Second, they choose the prices that maximize their discounted real profits. Here we introduce Calvo-pricing mechanism for nominal price rigidities. Specifically, a fraction \( 0 < \omega_p < 1 \)

²The toolbox is publicly available at https://sites.google.com/a/slu.edu/tanf.
of firms cannot change their prices each period. All other firms can only partially index their prices by the rule \( P_t(i) = P_{t-1}(i) \left( \pi_{t-1} \beta \right)^{\chi_t} \), where \( P_{t-1}(i) \) is indexed by the geometrically weighted average of past inflation \( \pi_{t-1} \) and steady state inflation \( \pi \). The weight \( 0 < \chi_t < 1 \) controls the degree of partial indexation.

The production sector can be summarized by four log-linearized equilibrium equations in terms of six parameters \((\alpha, \Omega, \beta, \omega_p, \chi_p, \eta_p)\), seven endogenous variables \((\hat{y}_t, \hat{k}_t, \hat{L}_t, \hat{r}_t, \hat{w}_t, \hat{m}_c, \hat{\eta}_t)\), and one exogenous shock \( \hat{u}_t \):

\[
\begin{align*}
\text{Production function:} & \quad \hat{y}_t = \frac{y + \Omega}{y} \left[ \alpha \hat{k}_t + (1 - \alpha)\hat{L}_t \right] \quad (2.1) \\
\text{Capital-labor ratio:} & \quad \hat{r}_t = \hat{L}_t - \hat{k}_t \quad (2.2) \\
\text{Marginal cost:} & \quad \hat{m}_c = \alpha \hat{r}_t + (1 - \alpha)\hat{w}_t \quad (2.3) \\
\text{Phillips equation:} & \quad \hat{\pi}_t = \frac{\beta}{1 + \beta \chi_p} \mathbb{E}_t \hat{\pi}_{t+1} + \frac{\chi_p}{1 + \beta \chi_p} \hat{\pi}_{t-1} + \kappa_p \hat{m}_c + \hat{u}_t \quad (2.4)
\end{align*}
\]

where \( \kappa_p = \left[ (1 - \beta \omega_p)/(1 - \omega_p) \right]/\left[ \omega_p (1 + \beta \chi_p) \right] \), \( \hat{\eta}_t = \ln(1 + \eta_t^p) - \ln(1 + \eta_t^z) \), \( \hat{\eta}_t^z \) is normalized to \( \hat{u}_t = \kappa_p \hat{\eta}_t^z \), and \( \mathbb{E}_t \) represents mathematical expectation given information available at time \( t \).

### 2.2 Households

The economy is populated by a continuum of households indexed by \( j \) on the interval \([0, 1]\). Each optimizing household \( j \) derives utility from composite consumption \( C_t(j) \), relative to a habit stock defined in terms of lagged aggregate composite consumption \( hC_{t-1}^* \) where \( 0 < h < 1 \). The composite consumption consists of private \( C_t(j) \) and public \( G_t \) consumption goods, i.e., \( C_t^*(j) = C_t(j) + \alpha_G G_t \), where \( \alpha_G \) governs the degree of substitutability of the consumption goods. Each household \( j \) also supplies a continuum of differentiated labor services \( L_t(j, l) \) where \( l \in [0, 1] \). Households maximize their expected lifetime utility

\[
\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u_t^b \left[ \ln(C_t^*(j) - hC_{t-1}^*) - L_t(j)^{1+\xi}/(1 + \xi) \right],
\]

where \( 0 < \beta < 1 \) is the discount rate, \( \xi > 0 \) is the inverse of Frisch labor supply elasticity, and \( u_t^b \) is an exogenous preference shock.

Households have access to one-period nominal private bonds \( B_{s,t} \) that pay one unit of currency at time \( t + 1 \), sell at price \( R_t^{-1} \) at time \( t \), and are in zero net supply. They also have access to a portfolio of long-term nominal government bonds \( B_t \), which sell at the price \( P_t^{B_t} \) at time \( t \). Maturity of these zero-coupon bonds decays at the constant rate \( 0 < \rho < 1 \) to yield the average duration \( (1 - \rho \beta)^{-1} \). Households receive bond earnings, labor and capital rental income, lump-sum transfers from the government \( Z_t \), and profits from firms \( \Pi_t \). They spend income on consumption, investment \( I_t \), and bonds. The nominal flow budget
constraint for household $j$ is given by

$$(1 + \tau^C)P_tC_t(j) + P_tI_t(j) + P_t^BB_t(j) + R_t^{-1}B_{s,t}(j) = (1 + \rho P_t^B)B_{t-1}(j) + B_{s,t-1}(j)$$

$$+(1 - \tau^L)\int_0^1 W_t(l)L_t(j,l)dl \quad (2.4)$$

where $W_t(l)$ is the nominal wage charged by the household for type $l$ labor service. Consumption and labor income are, in nominal terms, subject to a sales tax $\tau^C > 0$ and a labor income tax $\tau^L > 0$, respectively.

Effective capital $K(j)$, which is subject to a rental income tax $\tau^K > 0$, is related to physical capital $K_t(j)$ via $K_t(j) = v_t(j)K_{t-1}(j)$, where $v_t(j)$ is the utilization rate of capital chosen by households and incurs a nominal cost of $\Psi(v_t)$ per unit of physical capital.\(^3\) Physical capital is accumulated by households according to $K_t(j) = (1 - \delta)K_{t-1}(j) + u_t^i \left(1 - S \left( \frac{I_t(j)}{K_{t-1}(j)} \right) \right)I_t(j)$, where $0 < \delta < 1$ is the depreciation rate, $S(\cdot)$ is an investment adjustment cost and $u_t^i$ is an exogenous investment-specific efficiency shock.\(^4\)

There are perfectly competitive labor packers that hire a continuum of differentiated labor inputs $L_t(l)$, pack them to produce an aggregate labor service and then sell it to intermediate goods producers. The labor packer uses the Dixit-Stiglitz aggregator for labor aggregation $L_t^d = \left( \int_0^1 L_t(l) \frac{1}{1 + \eta_l^w} dl \right)^{1 + \eta_l^w}$, where $L_t^d$ is the aggregate labor service demanded by intermediate goods producers, $L_t(l)$ is the $l$th type labor service supplied by all the households and demanded by the labor packer, and $\eta_l^w$ is an exogenous wage markup shock.

For the optimal wage setting problem, we adopt the Calvo-pricing mechanism for nominal wage rigidities. Specifically, of all the types of labor services within each household, a fraction $0 < \omega_w < 1$ of wages cannot be changed each period. The wages for all other types of labor services follow a partial indexation rule $W_t(l) = W_{t-1}(l) \left( \pi_{t-1}(\omega_w) \right)^{\chi_w(\pi^\gamma(1 - \omega_w))^\chi_w}$, where $W_{t-1}(l)$ is indexed by the geometrically weighted average of the growth rates of nominal wage in the past period and in the steady state, respectively. The weight $0 < \chi_w < 1$ controls the degree of partial indexation.

The household sector can be summarized by ten log-linearized equilibrium equations in terms of seventeen parameters ($h, \gamma, \alpha_G, \rho, \tau^C, \tau^K, \tau^L, \psi, \beta, \gamma, s, \delta, \xi, \omega_w, \chi_w, \rho_a, \eta^w$), fifteen endogenous variables ($\hat{\lambda}_t, \hat{c}_t^*, \hat{c}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\pi}_t, \hat{P}_t^B, \hat{r}_t^k, \hat{v}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\lambda}_t, \hat{\lambda}_t$), and four exogenous shocks ($\hat{u}_t^a, \hat{u}_t^b, \hat{u}_t^c, \hat{u}_t^w$):

Optimal consumption:

$$\hat{\lambda}_t = \hat{u}_t^b - \frac{h}{e^\gamma - \hat{h}} \hat{u}_t^a - \frac{e^\gamma}{e^\gamma - \hat{h}} \hat{c}_t^* + \frac{h}{e^\gamma - \hat{h}} \hat{c}_{t-1}^* - \frac{\tau^C}{1 + \tau^C} \hat{\pi}_t^C$$

Composite consumption:

$$\hat{c}_t^* = \frac{c}{c + \alpha_G \hat{g}_t} \hat{c}_t + \frac{\alpha_G g}{c + \alpha_G \hat{g}_t} \hat{g}_t$$

\(^3\)Define the parameter $0 < \psi < 1$ such that $\frac{\psi^\gamma}{\psi^{1-\gamma}} = \frac{\psi}{1 - \psi}$.\(^4\) $S(\cdot)$ satisfies $S'(e^\gamma) = 0$ and $S''(e^\gamma) \equiv s > 0$.\(^2\)
Consumption Euler: \[ \dot{\lambda}_t = \hat{R}_t + E_t \dot{\lambda}_{t+1} - E_t \hat{\pi}_{t+1} - E_t \hat{u}^a_{t+1} \] (2.7)

Bond pricing: \[ \hat{R}_t + \hat{P}^B_t = \frac{\rho P^B}{1 + \rho P^B} E_t \hat{P}^B_{t+1} = \frac{\rho}{R} E_t \hat{P}^B_{t+1} \] (2.8)

Optimal capital utilization: \[ \dot{r}_t^k - \frac{\tau^K}{1 - \tau^K} \dot{r}_t^k = \frac{\psi}{1 - \psi} \dot{v}_t \] (2.9)

Optimal physical capital: \[ \dot{q}_t = E_t \dot{\lambda}_{t+1} - \dot{\lambda}_t - E_t \hat{u}^a_{t+1} + \beta e^{-\gamma} (1 - \tau^K) r^k E_t \dot{r}_t^k \] \[ - \beta e^{-\gamma} \tau^K r^k E_t \dot{r}_t^k + \beta e^{-\gamma} (1 - \delta) E_t \hat{q}_t + \beta e^{-\gamma} (1 - \delta) E_t \hat{q}_t \] (2.10)

Optimal investment: \[ \dot{\lambda}_t = -\frac{1}{1 + \beta} \dot{u}^a_t + \frac{1}{(1 + \beta) \sigma} \dot{q}_t + \hat{u}^i_t + \frac{\beta}{1 + \beta} E_t \hat{u}_{t+1} \] \[ + \frac{\beta}{1 + \beta} E_t \hat{u}_{t+1} + \frac{1}{1 + \beta} \dot{\lambda}_{t-1} \] (2.11)

Effective capital: \[ \dot{\hat{k}}_t = \dot{\hat{v}}_t + \dot{\hat{k}}_{t-1} - \dot{\hat{u}}^a_t \] (2.12)

Capital law of motion: \[ \dot{\hat{k}}_t = [1 - (1 - \delta)e^{-\gamma}][1 + \beta] \sigma \dot{\hat{u}}^i_t + \dot{\hat{q}}_t \] \[ + (1 - \delta)e^{-\gamma} (\dot{\hat{k}}_{t-1} - \dot{\hat{u}}^a_t) \] (2.13)

Wage equation: \[ \dot{w}_t = \kappa_w \left[ \dot{w}_t - \xi \dot{L}_t - \dot{u}^b_t + \hat{\lambda}_t - \frac{\tau^L}{1 - \tau^L} \dot{\hat{z}}_t \right] + \frac{1}{1 + \beta} \dot{\hat{w}}_{t-1} \] \[ + \frac{\beta}{1 + \beta} E_t \dot{\hat{u}}_{t+1} + \frac{\beta}{1 + \beta} E_t \dot{\hat{u}}_{t+1} - \frac{1 + \beta \chi_w}{1 + \beta} \frac{\dot{\hat{w}}_t}{1 + \beta} \] (2.14)

where \( \kappa_w = [(1 - \beta \omega_w)(1 - \omega_w)]/[\omega_w(1 + \beta)(1 + (1/\eta^w + 1/\xi))] \), \( \dot{\hat{w}}_t = \ln(1 + \eta^w) - \ln(1 + \eta^w), \) \( \dot{\hat{w}}^w_t \) is normalized to \( \dot{\hat{w}}^w_t = \kappa_w \dot{\hat{u}}^w_t, \) \( \dot{\hat{u}}^w_t \) is normalized to \( \dot{\hat{u}}^w_t = \frac{1}{(1 + \beta) \sigma} \dot{\hat{u}}^w_t, \) and \( \hat{\lambda}_t \) is the Lagrange multiplier associated with the household’s budget constraint. We set the capital, labor, and consumption tax rates to their constant steady states so that \( \dot{\hat{r}}^K_t = \dot{\hat{r}}^L_t = \dot{\hat{r}}^C_t = 0. \)

### 2.3 Monetary and Fiscal Policy

The central bank implements monetary policy according to a Taylor-type interest rate rule. The government collects revenues from capital, labor, and consumption taxes, and sells nominal bond portfolio to finance its interest payments and expenditures. The fiscal choices must satisfy the following government budget constraint \( P_t^B B_t + \tau^K R_t^K K_t + \tau^L W_t L_t + \tau^C P_t C_t = (1 + \rho P_t^B) B_{t-1} + P_t G_t + P_t Z_t, \) where we have assumed the lump sum transfers are equal across households, i.e., \( \int_0^1 Z_t(j) dj = Z_t, \) and fiscal instruments follow simple rules specified below.

The government sector can be summarized by seven log-linearized equilibrium equations in terms of thirteen parameters \( (\tau^C, \tau^K, \tau^L, \beta, \gamma, \rho, \rho_r, \rho_g, \rho_z, \phi_\pi, \phi_y, \gamma_g, \gamma_z), \) sixteen endogenous variables \( (\dot{b}_t, \dot{r}_t^K, \dot{k}_t, \dot{w}_t, \dot{L}_t, \)
where \( s_{t-1} = \frac{P_{t-1}^B B_{t-1}}{P_{t-1} Y_{t-1}} \) denotes the market value of the debt-to-GDP ratio, \( s = \tau^K r^k + \tau^L w L + \tau^C c - g - z \), \( 0 < \rho_r, \rho_g, \rho_z < 1 \) measure policy smoothness, \( \phi_\pi, \phi_y > 0 \) and \( \gamma_g, \gamma_z \) are policy parameters, and \((\hat{\omega}_t^m, \hat{\omega}_t^g, \hat{\omega}_t^z)\) are exogenous policy shocks.

Following Leeper, Traum and Walker (2017), we consider two distinct regions of the policy parameter space \((\phi_\pi, \gamma_g, \gamma_z)\) that deliver unique bounded rational expectations equilibria. The conventional active monetary/passive fiscal policy regime, or regime-M, has the monetary authority raise the nominal rate aggressively in response to inflation while the fiscal authority adjust expenditures and tax rates to stabilize debt. The alternative passive monetary/active fiscal policy regime, or regime-F, has monetary policy respond weakly to inflation while fiscal instruments adjust weakly to debt.

### 2.4 Exogenous Processes

All exogenous shocks follow autoregressive processes

\[
\hat{\omega}_t^s = \rho_{es} \hat{\omega}_{t-1}^s + \epsilon_t^s, \quad s \in \{a, b, i, p, w, m, g, z\}
\]

where \( \rho_{es} \in (0, 1) \) and the innovations \( \epsilon_t^s \) are serially uncorrelated and independent of each other at all leads and lags. We complete the model by assuming a multivariate Student-t distribution for the shock
innovations collected in an $8 \times 1$ vector $\epsilon_t$, i.e., $\epsilon_t \sim t_\nu(0, \Sigma_t)$, where $\nu$ denotes the degrees of freedom and $\Sigma_t$ is an $8 \times 8$ diagonal matrix with time-varying volatility $\sigma^2_{s,t}$ of $\epsilon^s_t$ on its main diagonal.\footnote{It is straightforward to introduce an independent Student-$t$ distribution with different degrees of freedom for each shock innovation. For exhibition ease, we do not consider this generalization.} For estimation convenience, it is useful to represent each element of $\epsilon_t$ as a mixture of normals by introducing a Gamma distributed random variable $\lambda_t$

$$\epsilon^s_t = \lambda_t^{-1/2} e^{h^s_t/2} \epsilon^s, \quad \lambda_t \sim \mathcal{G} \left( \frac{\nu}{2}, \frac{\nu}{2} \right), \quad \epsilon^s_t \sim \mathcal{N}(0, 1)$$

(2.23)

where, following the stochastic volatility literature (see, e.g., Kim, Shephard and Chib (1998)), the logarithm of each volatility $h^s_t = \ln \sigma^2_{s,t}$ collected in an $8 \times 1$ vector $h_t$ evolves as a stationary ($|\phi_s| < 1$) process

$$h^s_t = (1 - \phi_s)\mu_s + \phi_s h^s_{t-1} + \eta^s_t, \quad \eta^s_t \sim \mathcal{N}(0, \omega^2_s)$$

(2.24)

\section{2.5 Taking Model to Data}

Define the private sector's one-step-ahead endogenous forecast errors as

$$\eta^x_t \equiv \hat{x}_t - \mathbb{E}_{t-1} \hat{x}_t, \quad x \in \{ \lambda, \pi, i, q, r^k, w, P^B \}$$

(2.25)

The model consists of 36 log-linearized equilibrium equations and can be cast into the rational expectations system

$$
\begin{pmatrix}
\Gamma^e e\ 0 \\
0 \\
[0, I]
\end{pmatrix}_{(36 \times 36)}
\begin{pmatrix}
x^e_t \\
x^e_t
\end{pmatrix}_{(36 \times 1)}
= 
\begin{pmatrix}
\Gamma^e e \ 0 \\
0 \\
[0, I]
\end{pmatrix}_{(36 \times 36)}
\begin{pmatrix}
x^e_{t-1} \\
x^e_{t-1}
\end{pmatrix}_{(36 \times 1)}
+ 
\begin{pmatrix}
0 \\
I
\end{pmatrix}_{(36 \times 1)}
\begin{pmatrix}
\epsilon_t \\
\eta_t
\end{pmatrix}_{(8 \times 1)}
+ 
\begin{pmatrix}
0 \\
I
\end{pmatrix}_{(36 \times 7)}
\begin{pmatrix}
\Psi \\
\Pi
\end{pmatrix}_{(36 \times 8)}
$$

(2.26)

where $I$ denotes the identity matrix, $P = \text{diag} \left( \rho_{ea}, \rho_{eb}, \rho_{ei}, \rho_{ep}, \rho_{ew}, \rho_{em}, \rho_{eg}, \rho_{ez} \right)$, $x^e_t = \text{diag} \left( \tilde{g}_t, \tilde{c}_t, \tilde{k}_t, \tilde{h}_t, \tilde{w}_t, \tilde{R}_t, \tilde{s}_t^b, \tilde{s}_t^x, \tilde{\lambda}_t, \tilde{\pi}_t, \tilde{q}_t, \tilde{r}_t^k, \tilde{w}_t, \tilde{P}_t^B \right)$.
are the endogenous variables,

\[ x_t^z = [\hat{u}_t^a, \hat{u}_t^b, \hat{u}_t^i, \hat{u}_t^p, \hat{u}_t^w, \hat{u}_t^m, \hat{u}_t^q, \hat{u}_t^z] \]

are the exogenous shocks,

\[ x^d_t = [E_t\hat{\lambda}_{t+1}, E_t\hat{\pi}_{t+1}, E_t\hat{\iota}_{t+1}, E_t\hat{q}_{t+1}, E_t\hat{r}_{t+1}, E_t\hat{w}_{t+1}, E_t\hat{P}_B_{t+1}] \]

are the conditional expectations of the last seven elements of \( x_{t+1}^e \),

\[ \epsilon_t = [\epsilon_t^a, \epsilon_t^b, \epsilon_t^i, \epsilon_t^p, \epsilon_t^m, \epsilon_t^q, \epsilon_t^z] \]

are the shock innovations, and

\[ \eta_t = [\eta_t^\lambda, \eta_t^\pi, \eta_t^i, \eta_t^q, \eta_t^{rk}, \eta_t^w, \eta_t^{P_B}] \]

are the forecast errors.

Here the first row of (2.26) stacks the 21 structural equations (2.1)–(2.21), the second row stacks the 8 shock processes (2.22), and the third row stacks the 7 definitional equations (2.25). The unknown parameters \( \theta \) consist of the structural parameters

\[ \theta^S_{27 \times 1} = [100\gamma, \xi, h, \alpha_G, \psi, s, \omega_p, \omega_w, \chi_p, \chi_w, \phi_\pi, \phi_y, \gamma_g, \gamma_z, \rho_r, \rho_g, \rho_z, \rho_{ea}, \rho_{eb}, \rho_{ei}, \rho_{ep}, \rho_{ew}, \rho_{em}, \rho_{eg}, \rho_{ez}, \bar{L}, \bar{\pi}] \]

and the volatility parameters

\[ \theta^V_{24 \times 1} = [\mu_a, \mu_b, \mu_i, \mu_p, \mu_w, \mu_m, \mu_g, \mu_z, \phi_a, \phi_b, \phi_i, \phi_p, \phi_w, \phi_m, \phi_g, \phi_z, \phi_2, \phi_2^a, \phi_2^b, \phi_2^i, \phi_2^p, \phi_2^w, \phi_2^m, \phi_2^g, \phi_2^z, \omega_2, \omega_2^a, \omega_2^b, \omega_2^i, \omega_2^p, \omega_2^w, \omega_2^m, \omega_2^g, \omega_2^z] \]

Conditional on \( \theta^S \) and independent of the volatility processes, the above structural system can be solved by the procedure of Sims (2002) to deliver a linear solution of the form

\[ x_t = G(\theta^S)x_{t-1} + M(\theta^S)\epsilon_t \]

which is then estimated over a vector \( y_t \) of 8 observables stacked in \( y = [y_1, \ldots, y_T] \), including log differences (denoted dl) of consumption, investment, real wage, government spending, and government debt; log (denoted l) hours worked, inflation, and nominal interest rate.\(^6\) The observables are linked to the model

---

\(^6\)See the Online Appendix of Leeper, Traum and Walker (2017) for details on data construction.
variables $x_t$ via the following measurement equations

$$
\begin{bmatrix}
\text{dlCons}_t \\
\text{dlInv}_t \\
\text{dlWage}_t \\
\text{dlGovSpend}_t \\
\text{dlGovDebt}_t \\
\text{lHours}_t \\
\text{lInfl}_t \\
\text{lFedFunds}_t
\end{bmatrix}
= 
\begin{bmatrix}
100\gamma \\
100\gamma \\
100\gamma \\
100\gamma \\
100\gamma \\
L \\
\bar{\pi} \\
\bar{\pi} + 100(\gamma/\beta - 1)
\end{bmatrix}
\begin{bmatrix}
\hat{c}_t - \hat{c}_{t-1} + \hat{u}_t^a \\
\hat{\delta}_t - \hat{\delta}_{t-1} + \hat{u}_t^a \\
\hat{w}_t - \hat{w}_{t-1} + \hat{u}_t^a \\
\hat{g}_t - \hat{g}_{t-1} + \hat{u}_t^a \\
\hat{b}_t - \hat{b}_{t-1} + \hat{u}_t^a \\
\hat{\pi}_t \\
\hat{\pi}_t \\
\hat{\pi}_t + 100(\gamma/\beta - 1)
\end{bmatrix} + 
\begin{bmatrix}
\hat{\delta}_t \\
\hat{\delta}_t \\
\hat{w}_t \\
\hat{g}_t \\
\hat{b}_t \\
\hat{\pi}_t \\
\hat{\pi}_t \\
\hat{\pi}_t
\end{bmatrix}
$$

(2.28)

Let $\lambda = [\lambda_1, \ldots, \lambda_T]'$ contain all non-Gaussian latent states and $h = [h_1', \ldots, h_T']'$ contain all nonlinear latent states. Further collect them in $z = [\lambda', h']'$, which in our empirical application ($T = 166$) has a total of 1,494 elements. In conjunction with the shock volatility specification (2.23) and (2.24), equations (2.27) and (2.28) form a state space representation of the DSGE model whose conditional likelihood function $f(y|\theta, z)$ can be evaluated with the Kalman filter.

### 3 Estimation

The general framework for estimating DSGE models, employing Bayesian tools, is by now quite well established. The state space model (2.27)–(2.28) is supplemented by a prior distribution $\pi(\theta)$ summarizing the researcher’s initial views of the model parameters. This prior information is updated with the sample information via Bayes’ theorem

$$
\pi(\theta|y) \propto f(y|\theta) \cdot \pi(\theta) \cdot 1 \{\theta \in \Theta_D\}
$$

where $f(y|\theta)$ is the observed likelihood function, and the posterior distribution $\pi(\theta|y)$ characterizing the researcher’s updated parameter beliefs is calculated up to the normalization constant (i.e., the marginal likelihood). Moreover, $1 \{\theta \in \Theta_D\}$ is an indicator function that equals one if $\theta$ is in the determinacy region $\Theta_D$ and zero otherwise. Because $\pi(\theta|y)$ is typically intractable, it is customary to summarize this distribution numerically based on a sample of draws $\{\theta^{(g)}\}_{g=1}^G$ from it using an appropriate MCMC procedure.
The first element of our approach is the use of training sample priors beyond a prior predictive analysis of the model, instead of the priors used in the DSGE literature to date (Section 3.1). Although training sample priors are common in the vector autoregression (VAR) literature, they have not previously been used in the DSGE setting. In high-dimensions, these priors are superior to conventional priors because they tend to avoid prior-likelihood conflicts. We further mitigate the possibility of such conflicts by using a Student-\(t\) prior for the location parameters, where the hyperparameters are constructed based on the training sample.

In the current high-dimensional context, there is still little experience on how an MCMC sampling procedure should be implemented to estimate such a model. Our second point is that reliable estimation is possible by the TaRB-MH algorithm of Chib and Ramamurthy (2010) (Section 3.2). Two defining features of this algorithm are worth mentioning. One is the random clustering of the structural parameters \(\theta^S\) at every iteration into an arbitrary number of blocks. Each block is then sequentially updated through an M-H step. Another is the adaptation of the proposal density to the location and curvature of the posterior distribution for a given block using a mix of simulated annealing and a deterministic optimizer. The TaRB-MH algorithm may appear to require work, but random blocking and tailoring are central to generating efficient exploration of the posterior distribution. Our MATLAB package, by eliminating entry costs, makes this powerful algorithm widely available to macroeconomists.

Finally, for model comparisons, we compute the marginal likelihood by the Chib (1995) and Chib and Jeliazkov (2001) method (Section 3.3). This method provides reliable estimates of the marginal likelihood because it is based on the output of the efficient TaRB-MH algorithm.

### 3.1 Prior Distribution

Table A.1 of Appendix A lists the marginal prior distributions for all model parameters.\(^7\) The priors on the structural parameters follow closely Leeper, Traum and Walker (2017), and those on the volatility parameters imply a fairly persistent volatility process for each shock innovation. In the Bayesian estimation of DSGE models, an informative prior distribution (such as those on the policy parameters \(\phi_\pi, \gamma_g, \gamma_z\)) can play an important role in shifting the posterior distribution toward regions of the parameter space that are economically meaningful. It can also introduce curvature into the posterior surface that facilitates numerical optimization and MCMC simulations (such as the tailoring of proposal densities in the TaRB-MH algorithm).

\(^7\)Because some parameters are held fixed under each regime, effectively, \(\theta\) has 49 elements and \(\theta^S\) has 25 elements.
When it comes to high-dimensions, however, developing a sensible prior becomes increasingly difficult due to the complex mapping from the structural parameters to those of the state space form. Consequently, a reasonable prior for the parameters per se may still imply a distribution of the data that is strongly at odds with actual observations. For instance, Figure 1 shows the implied distributions for selected sample moments under the original regime-M prior and model specification of Leeper, Traum and Walker (2017) (red dashed lines). Most notably, this prior places little or no mass in the neighborhood of the actual mean of government spending and the actual standard deviations of investment, government spending, debt, and hours worked (vertical lines). After taking the model to data, we also find that the posterior mass for several parameters (e.g., the habit parameter $h$, the nominal rigidity parameters $\omega_p$ and $\omega_w$, and the government spending persistence $\rho_g$) lies entirely in the far tail of the corresponding prior, thereby introducing fragility to the inferences. To cope with these issues, we suggest two pointers on formulating an appropriate prior that can avoid such prior-sample conflict.

### 3.1.1 Sampling the Prior

The first step follows the sampling the prior approach in, e.g., Geweke (2005) and Chib and Ergashev (2009). In particular, start with the prior for the structural parameters $\theta^S$ adopted in Leeper, Traum and Walker (2017) and an initial prior for the volatility parameters $\theta^V$. Then sample this joint prior a large number of times (say 1,000) and simulate a data set $y^{(g)}$ for each parameter draw $\theta^{(g)}$, $g = 1, \ldots, G$, under which the model has a unique bounded solution. Finally, compute the implied distributions for various functions of the data (such as the sample mean, standard deviation, and autocorrelation) and check whether these are close to corresponding estimates from the real data. If not, adjust some or all marginal components of the prior for $\theta^V$ and repeat the above process.\(^8\)

It is clear from Figure 1 that under the adjusted prior, reported in Table A.1 of Appendix A, the Leeper, Traum and Walker (2017) model extended with Student-$t$ shocks and stochastic volatility implies distributions of the data that capture the corresponding real data quantities in their relatively high density regions (represented by the shaded areas).

\(^8\)In the Leeper, Traum and Walker (2017) setting with Gaussian shocks and constant volatility, this step suggests that the original prior for the standard deviation parameters $100\sigma_s$, $s \in \{a, b, i, p, w, m, g, z\}$, each of which follows an Inverse-Gamma type-2 distribution with mean 0.1 and standard deviation 1, should be adjusted. Alternatively, one could also adjust some or all marginal components of the prior for $\theta^S$.  

13
3.1.2 Training Sample Prior

In the second step, given the adjusted prior from the first step, we use the TaRB-MH algorithm to estimate the DSGE model on the initial 50 observations running from 1955:Q1 to 2008:Q4. The posterior draws from this run are used to form the prior. Specifically, we keep the prior type of each parameter unchanged but set its location (dispersion) to the corresponding mean (twice standard deviation). Note also that we adopt a Student-$t$ prior with 2.1 degrees of freedom for each location parameter $\mu_s$, $s \in \{a, b, i, p, w, m, g, z\}$, of the volatility process. As will be shown below, this training sample prior avoids any stark conflict between the prior and the likelihood.

3.2 Posterior Estimates

Throughout this section, we apply the TaRB-MH algorithm to estimate our high-dimensional DSGE model based on the post-training sample from 1967:Q3 to 2008:Q4, which includes 166 quarterly observations. With the ultimate goal of forecasting in mind, we present the estimation results for the model of best fit among all competing specifications. This specification stands out from an extensive model search based on the marginal likelihood computation, as performed in the next section. It has regime-M in place and features heavy-tailed shocks with $\nu = 5$ degrees of freedom and persistent volatilities.

We first discuss a full MCMC procedure that explores the conditionally linear and Gaussian structure of our high-dimensional DSGE model. This procedure consists of two primary MCMC steps that are executed in the program \texttt{tarb\_full.m}. The first step samples 25 structural parameters in $\theta^S$ from the conditional posterior $\pi(\theta^S|y, \theta^V, \lambda, h)$ by the TaRB-MH algorithm of Chib and Ramamurthy (2010), which avoids the need to evaluate the observed likelihood $f(y|\theta)$ with the computationally costly particle filter. The second step samples the remaining blocks, including 24 volatility parameters in $\theta^V$, 166 non-Gaussian latent variables in $\lambda$, and 1,328 nonlinear latent variables in $h$, from the conditional posterior $\pi(\theta^V, \lambda, h|y, \theta^S)$ by the methods developed in Kim, Shephard and Chib (1998) and Chib and Ramamurthy (2014). Iterating the above cycle until convergence produces a sample from the joint posterior $\pi(\theta, z|y)$, where recall that $\theta = [\theta^S, \theta^V]$ and $z = [\lambda', h']$. We refer interested readers to the original papers above for complete details of these methods.
3.2.1 Sampling Structural Parameters

The first step entails sampling $\theta^S$ from

$$\pi(\theta^S|y, \theta^V, z) \propto f(y|\theta^S, z) \cdot \pi(\theta^S) \cdot 1\{\theta \in \Theta_D\} \quad (3.1)$$

using the TaRB-MH algorithm. To fix ideas, consider the $g$th iteration where a random partition of $B$ blocks from a permuted sequence of $\theta^S$ has been formed, i.e., $\theta^S = (\theta^S_1, \ldots, \theta^S_B)$. Specifically, we initialize $\theta^S_1$ with the first element of this shuffled sequence, and start a new block with every next element with probability $1 - p$. As a result, the average size of a block is given by $(1 - p)^{-1}$. In our benchmark setting, we set $p = 0.7$ so that each block contains three to four parameters on average. This effectively breaks a 25-dimensional sampling problem into about seven smaller ones. The random block feature is also useful as the researcher typically does not have a priori knowledge about the correlation pattern of $\theta^S$. Now suppose the blocks $\theta_{1:b-1}^{S,(g)} = (\theta^S_1, \ldots, \theta^S_{b-1})$ have been updated in the current iteration, whereas the remaining blocks $\theta_{b:B}^{S,(g-1)} = (\theta^S_{b}, \ldots, \theta^S_B)$ and $z^{(g-1)}$ take values in the previous iteration. Then the steps for updating the $b$th block can be summarized as follows:

1. Use the simulated annealing (SA) optimization method (available as a MATLAB built-in function `simulannealbnd`) to obtain an initial solution to

   $$\hat{\theta}^S_b = \arg \min_{\theta^S_b} \ln f(y|\theta^S_{1:b-1}^{S,(g)}, \theta^S_b, \theta_{b+1:B}^{S,(g-1)}, z^{(g-1)}) \cdot \pi(\theta^S_b)$$

   This SA version of the posterior mode is further used to initiate the BFGS quasi-Newton method (available as a MATLAB function `csminwel` written by Chris Sims) that refines the initial solution. `csminwel` also approximates the inverse of the Hessian matrix evaluated at $\hat{\theta}^S_b$, denoted by $\hat{V}_b$, and returns it as a byproduct.\(^9\)

2. Generate a candidate draw from the tailored Student-$t$ proposal density

   $$\theta^S_{b}^{(g)} \sim t_\nu(\hat{\theta}^S_b, \hat{V}_b)$$

   where the degrees of freedom is set to $\nu = 15$. The local tailoring feature allows for sizable moves from the neighborhood of the current parameter draw.

\(^9\)The same optimization procedure, executed in the program `chain_init.m`, is repeated multiple times to obtain a starting value $\theta^{S,(0)}$ for the chain.
3. Accept \( \theta_b^{S,g} \) as the updated value of \( \theta_b^S \) with probability

\[
\alpha = \min \left\{ 1, \frac{f(y|\theta_{b+1:B}^{S,g}, \theta_{b+1:B}^{S,(g-1)}, z^{(g-1)}) \cdot \pi(\theta_b^{S,g})}{f(y|\theta_{b+1:B}^{S,g}, \theta_{b+1:B}^{S,(g-1)}, z^{(g-1)}) \cdot \pi(\theta_b^{S,(g-1)})} \cdot \frac{\nu(\theta_b^{S,g}|\theta_{b+1:B}^*, \tilde{Z}_b)}{\nu(\theta_b^{S,(g-1)}|\theta_{b+1:B}^*, \tilde{Z}_b)} \right\}
\]

We also introduce a new procedure, i.e., tailoring at random frequency, to accelerate the TaRB-MH algorithm. This procedure allows the researcher to efficiently estimate a high-dimensional DSGE model within a reasonable amount of runtime. The idea is similar in essence to grouping the structural parameters into random blocks. Because the tailored proposal density in the current iteration may remain efficient for the next few iterations, there is typically no need to re-tailor the proposal density in every iteration. Nevertheless, there is still a chance that the re-tailored proposal density will be quite different from the recycled one. Therefore, randomizing the number of iterations before new blocking and tailoring ensures that the proposal density remains well-tuned on average. The reciprocal of this average number, which we call the tailoring frequency \( \omega \), as well as a number of optional user inputs (e.g., the blocking probability \( p \)), can be specified flexibly in the program \texttt{tarb_spec.m}. In our benchmark setting, we set \( \omega = 0.5 \) so that each proposal density is tailored every second iterations on average.

We apply the TaRB-MH algorithm to sample a total of 11,000 draws from the posterior distribution and discard the first 1,000 draws as burn-in phase. Figure 2 provides a graphical comparison of the prior and posterior for each structural parameter. For most parameters, the post-training sample carries information which is beyond but by no means conflicts with that contained in the training sample prior. In particular, the data imply quite high habit formation and relatively high degrees of price and wage stickiness. See also Table A.2 of Appendix A for a detailed summary of the posterior parameter estimates.

### 3.2.2 Sampling Latent Variables and Volatility Parameters

The second step involves augmenting the remaining blocks with 1,328 shock innovations \( \epsilon = [\epsilon_1', \ldots, \epsilon_T'] \) and then sampling the joint posterior \( \pi(\theta^V, \epsilon, \lambda, h|y, \theta^S) \). To this end, Gibbs sampling is applied to the following conditional densities

\[
\pi(\epsilon|y, \theta, \lambda, h), \quad \pi(\lambda|y, \theta, \epsilon, h), \quad \pi(\theta^V, h|y, \theta^S, \epsilon, \lambda)
\]

using the steps below:
1. Sample $\epsilon$ from the first density in (3.2)

$$
\pi(\epsilon|y, \theta, \lambda, h) = \pi(\epsilon|y, \theta^S, \lambda, h)
$$

with the disturbance smoother of Durbin and Koopman (2002) applied to the state space form (2.27)–(2.28), where each $\epsilon^s_t \sim \mathcal{N}(0, e^{h^s_t}/\lambda_t)$, $s \in \{a, b, i, p, w, m, g, z\}$, due to the gamma-normal representation (2.23).

2. Sample $\lambda$ from the second density in (3.2)

$$
\pi(\lambda|y, \theta, \epsilon, h) \propto \prod_{t=1}^{T} f(\epsilon_t|\lambda_t, h_t) \cdot \pi(\lambda_t)
$$

by independently sampling each $\lambda_t$ from

$$
\lambda_t \sim \mathcal{G}\left(\frac{\nu + n_\epsilon}{2}, \frac{\nu + \epsilon_t^T \Sigma_t^{-1} \epsilon_t}{2}\right), \quad \Sigma_t = \text{diag}(e^{h^s_t})
$$

as in Chib and Ramamurthy (2014), where $n_\epsilon = 8$ is the dimension of $\epsilon_t$.

3. Following Kim, Shephard and Chib (1998), the nonlinear measurement equation (2.23) can be transformed into a linear one by squaring and taking logarithm. In conjunction with the volatility state equation (2.24), this leads to the state space model

$$
\begin{align*}
    h^s_t &= (1 - \phi_s) \mu_s + \phi_s h^s_{t-1} + \eta^s_t \\
    \tilde{\epsilon}^s_t &= h^s_t + \epsilon^s_t
\end{align*}
$$

for $s \in \{a, b, i, p, w, m, g, z\}$, where $\tilde{\epsilon}^s_t = \ln \lambda_t (\epsilon^s_t)^2$ and $\epsilon^s_t = \ln (\tilde{\epsilon}^s_t)^2$. Practically, we set $\tilde{\epsilon}^s_t = \ln[\lambda_t (\epsilon^s_t)^2 + c]$ with $c = 10^{-5}$ being an offset constant, and accurately approximate the distribution of $\epsilon^s_t$ by the 10-component mixture normal density proposed by Omori et al. (2007)

$$
p(\epsilon^s_t) = \sum_{k=1}^{10} q_k \cdot p_n(\epsilon^s_t|s^s_t = k)
$$

where $s^s_t$ is an indicator variable and $p_n(\cdot|s^s_t = k)$ denotes a normal density function with mean $m_k$, variance $\nu^2_k$, and component probability $q_k$. Now, to sample the last density in (3.2), we further augment the remaining blocks with 1,328 indicator variables $s = [s^a_t, \ldots, s^z_T]$, where $s_t = [s^a_t, s^b_t, s^i_t, s^p_t, s^w_t, s^m_t, s^g_t, s^z_t]$, and sample the joint posterior $\pi(s, \theta^V, h|y, \theta^S, \epsilon, \lambda)$ by sampling the con-
ditional densities

\[ \pi(s|y, \theta, \epsilon, \lambda, h), \quad \pi(\theta^V, h|y, \theta^S, \epsilon, \lambda, s) = \pi(\theta^V|y, \theta^S, \epsilon, \lambda, s) \cdot \pi(h|y, \theta, \epsilon, \lambda, s) \]

using these steps:

(a) Sample each \( s^s = [s^s_1, \ldots, s^s_T]' \), \( s \in \{a, b, i, p, w, m, g, z\} \), independently from

\[ \pi(s^s|y, \theta, \epsilon, \lambda, h) \propto \prod_{t=1}^T f(\tilde{\epsilon}^s_t|h^s_t, s^s_t) \cdot \pi(s^s_t) \]

where \( f(\cdot|h^s_t, s^s_t = k) \) is a normal density function with mean \( h^s_t + m_k \) and variance \( v^2_k \), and \( \pi(s^s_t = k) = q_k, k = 1, \ldots, 10 \).

(b) Sample \( \theta^V \) marginalized over \( h \) (the ‘integration sampler’ in Kim, Shephard and Chib (1998)) by sampling each triplet \( (\mu_s, \phi_s, \omega^2_s) \), \( s \in \{a, b, i, p, w, m, g, z\} \), independently from

\[ \pi(\mu_s, \phi_s, \omega^2_s|y, \theta^S, \epsilon, \lambda, s) \propto f(\tilde{\epsilon}^s|\mu_s, \phi_s, \omega^2_s, s^s) \cdot \pi(\mu_s, \phi_s, \omega^2_s) \]

using a tailored proposal density, where \( \tilde{\epsilon}^s = [\tilde{\epsilon}^s_1, \ldots, \tilde{\epsilon}^s_T]' \) and \( f(\tilde{\epsilon}^s|\mu_s, \phi_s, \omega^2_s, s^s) \) is available from the Kalman filter, followed by the sampling of each \( h^s = [h^s_1, \ldots, h^s_T]' \), \( s \in \{a, b, i, p, w, m, g, z\} \), using the ‘filter-forward-sample-backward’ method of Carter and Kohn (1994).

Figure 3 plots the estimated historical log-volatility series for 1967:Q3 to 2008:Q4 based on the \( h \) draws. Overall, these estimates display clear countercyclical time variation, with pronounced increases in volatility accompanying the recessions. For several shock innovations, volatility becomes lower by historical standards since the 1980s so that the Great Moderation is also evident.

3.2.3 Sampling Efficiency

To see the sampling efficiency of the TaRB-MH algorithm, it is informative to examine the serial correlation among the sampled draws. Figures A.1–A.2 of Appendix A display the autocorrelation function for each element of \( \theta \). As can be observed, the serial correlations for most parameters decay quickly to zero after a few lags. Due to the efficiency gains achieved by the TaRB-MH algorithm, the number of draws is substantially smaller (say 10,000 draws) than the number typically used for the random-walk M-H algorithm (say one million draws).
Another useful measure of the sampling efficiency is the so-called inefficiency factor, which approximates the ratio between the numerical variance of the estimate from the MCMC draws and that from the hypothetical i.i.d. draws. An efficient sampler produces reasonably low serial correlations and hence inefficiency factors. Figure 4 compares the inefficiency factors resulting from two different sampling schemes, each corresponding to a choice of the tailoring frequency $\omega \in \{0.5, 1.0\}$ with the same blocking probability $p = 0.7$. Compared to the more efficient setting that tailors in every iteration ($\omega = 1.0$), our benchmark setting that tailors on average every second iterations ($\omega = 0.5$) leads to very similar inefficiency factors, ranging from $3.37$ ($1.74$) to $103.56$ ($68.97$) with most values below $20$ ($15$) for the structural (volatility) parameters. In conjunction with a rejection rate of approximately $50\%$ in the M-H step, the small inefficiency factors suggest that the chain mixes well. Moreover, choices of $\omega$ have no material effect on the sampling efficiency for most volatility parameters due to the use of the integration sampler that is tailored in every iteration.

While single blocking or infrequent tailoring can greatly reduce the overall runtime, it may also add considerably to the average inefficiency factor. Therefore, in practice, we suggest setting $p \in [0.6, 0.9]$ and $\omega \in [0.2, 1.0]$ to maintain a good balance between runtime and efficiency.

### 3.2.4 Simulation Evidence

We also estimate the same high-dimensional DSGE model based on a simulated data set that is generated under fat-tailed shocks with $\nu = 15$ degrees of freedom and persistent volatilities. We set the sample size to 200, which is meant to be 50 years of quarterly observations, and use the initial 50 observations to construct a training sample prior. Table A.1 of Appendix A lists the parameter values used for the data generating process under regime-M (column ‘DGP’). Figure A.3 provides a graphical comparison of priors and posteriors. For most parameters, the posterior mass concentrates around the corresponding true value. Figure A.4 further reveals that the estimated log-volatility series largely captures the level and all major trends of the true series for each shock innovation. These plots are relegated to Appendix A.

### 3.3 Marginal Likelihood

Given the output of the efficient TaRB-MH algorithm, the method of Chib (1995), as modified for M-H chains in Chib and Jeliazkov (2001) and implemented in the program `tarb_reduce.m`, can be readily

---

10 Formally, the inefficiency factor is computed as $1 + 2 \sum_{j=1}^{K} w(j/K) \rho(j)$, where the autocorrelation function $\rho(\cdot)$ is weighed by the Parzen kernel $w(\cdot)$ and $K$ is the truncation level at which $\rho(\cdot)$ tapers off (see Kim, Shephard and Chib (1998)).

applied to calculate the marginal likelihood of different DSGE models. For instance, one may be interested in comparing regime-M with Student-\(t\) shocks with regime-F with stochastic volatility. Other models could be of interest as we discuss below.

The marginal likelihood is the quantity
\[
m(y|\mathcal{M}) = \frac{1}{c} \int f(y|\mathcal{M}, \theta) \cdot \pi(\theta|\mathcal{M}) \cdot \mathbb{1}\{\theta \in \Theta_D\} \, d\theta
\]
where \(\mathcal{M}\) denotes the model label and \(c = \int_{\theta \in \Theta_D} \pi(\theta|\mathcal{M}) \, d\theta\). In the Chib (1995) approach, this is computed via the identity
\[
m(y|\mathcal{M}) = \frac{1}{c} \frac{f(y|\mathcal{M}, \theta) \cdot \pi(\theta|\mathcal{M}) \cdot \mathbb{1}\{\theta \in \Theta_D\}}{\pi(\theta|\mathcal{M}, y)}
\]
where the right hand side terms are evaluated at a single high density point \(\theta^*\). We obtain the likelihood ordinate by a mixture version of the Kalman filter introduced by Chen and Liu (2000), as facilitated by the conditionally Gaussian and linear structure of the DSGE model solution. In our application, we find that 10,000 particles are sufficient to deliver a robust estimate of \(f(y|\mathcal{M}, \theta^*)\). We obtain the high-dimensional ordinate in the denominator after decomposing it as
\[
\pi(\theta^*|\mathcal{M}, y) = \pi(\theta^*_1|\mathcal{M}, y) \cdot \pi(\theta^*_2|\mathcal{M}, y, \theta^*_1) \cdots \pi(\theta^*_B|\mathcal{M}, y, \theta^*_1, \ldots, \theta^*_{B-1})
\]
where \(B\) refers to the number of blocks (that is under our control), and then estimate each of these reduced ordinates from the MCMC output of reduced runs (see Chib (1995) and Chib and Jeliazkov (2001) for further details).

An interesting point is that these reduced runs are independent of each other and can be done in parallel. Thus, all reduced ordinates can be estimated at the cost of one reduced run, regardless of the size of \(B\). This parallel computation is built into our MATLAB package. In our application, we set the total number of blocks to \(B = 15\), including seven almost equally sized blocks for \(\theta^S\) arranged first, followed by eight blocks \((\mu_s, \phi_s, \omega^2_s), \ s \in \{a, b, i, p, w, m, g, z\}\), for \(\theta^V\). All ordinates are then simultaneously estimated using MATLAB’s multi-core processing capacity via its Parallel Computing Toolbox.

### 3.3.1 Reliability

We recommend the Chib and Jeliazkov (2001) method because it is reliable and because other methods do not generalize to our large-scale DSGE models with non-Gaussian and/or nonlinear latent variables.\(^\text{12}\)

\(^\text{12}\)For instance, the modified harmonic mean (MHM) estimator of Gelfand and Dey (1994), used, for example, in Justiniano and Primiceri (2008) and Cúrdia, Del Negro and Greenwald (2014) in medium-scale DSGE models with Student-\(t\) shocks.
As shown in Chib and Ramamurthy (2010), efficient MCMC estimation automatically delivers an efficient estimate of the conditional posterior ordinate \( \pi(\theta_0, \ldots, \theta_B, z|y, \theta_1^*, \ldots, \theta_{b-1}^*) \) from the output of the reduced MCMC simulation in which \( \theta_0 \) is a fixed block and the remaining structural parameters, if any, form random blocks.\(^{13}\) Figure 5 displays the sequence of posterior ordinate and marginal likelihood estimates from the best fit model, as functions of the number of MCMC draws, for efficient and (relatively) less efficient TaRB-MH implementations, as described in Section 3.2.3. These estimates settle down quickly (after say 1,000 draws are made) and converge to the same limit point, leading to an estimated log marginal likelihood of about \(-1579.65\) with a numerical standard error of about 0.12. This underscores the point that since the Chib (1995) method is underpinned by whatever MCMC algorithm is used in the posterior simulation, the efficiency of the MCMC simulator is germane to the calculation of the marginal likelihood.

### 3.3.2 Regime Comparison

Because regimes M and F of the Leeper, Traum and Walker (2017) model imply completely different mechanisms for price level determination and therefore different policy advice, identifying which monetary-fiscal regime produced the real data is key to making good policy choices. While it is difficult to explore the entire model space, we perform extensive regime comparisons by estimating the marginal likelihood for both regimes with four choices of \( \nu \in \{2.1, 5, 15, 30\} \) and three choices of \( \phi_s \in \{0.1, 0.5, 0.95\} \), \( s \in \{a, b, i, p, w, m, g, z\} \). The resulting model space contains a total of 24 relevant models that are simultaneously confronted with the data over the period from 1967:Q3 to 2008:Q4, similar in spirit to the Bayesian model scan framework proposed by Chib and Zeng (2019).\(^{14}\)

Two aspects of the marginal likelihood estimates reported in Table 1 are worth highlighting. First, the data systematically prefer regime-M over regime-F in all cases, which corroborates the regime ranking found by Leeper, Traum and Walker (2017) with Gaussian shocks and constant volatility.\(^{15}\) The small numerical standard errors point to the numerical accuracy of the marginal likelihood estimates. Second, reading the table by row (column) for each regime suggests that the data exhibit quite strong evidence in favor of heavy-tailed shocks (persistent volatility process). Indeed, each feature is important for improving and stochastic volatility, always favors a model specification with stronger latent features, e.g., shocks with fatter tails or volatilities with more persistence. This extreme result emerges even when the true model exhibits weak evidence of these features, such as those considered in Section 3.3.3.

\(^{13}\)In contrast, Justiniano and Primiceri (2008, p. 636) and Herbst and Schorfheide (2016, p. 97) estimate the posterior ordinate in a single-block, with the random-walk M-H, both detrimental to getting reliable and efficient marginal likelihood estimates, as already documented in Chib and Jeliazkov (2001).

\(^{14}\)All computations performed in this section are executed on the High Performance Computing Cluster maintained by Saint Louis University (https://sites.google.com/a/slu.edu/atg/home).

\(^{15}\)We also find the reversed regime ranking with the inclusion of the recent financial crisis sample, a period of nearly zero policy rate and unprecedented fiscal stimulus that regime-F policy rules embody. See Section 4 for more details.
Table 1: Log marginal likelihood estimates

<table>
<thead>
<tr>
<th>ν</th>
<th>φₙ = 0.1 (weak)</th>
<th>φₙ = 0.5 (moderate)</th>
<th>φₙ = 0.95 (strong)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>30 (light)</td>
<td>−1640.73 (0.15)</td>
<td>−1650.03 (0.15)</td>
<td>−1627.24 (0.14)</td>
</tr>
<tr>
<td>15 (fat)</td>
<td>−1622.26 (0.14)</td>
<td>−1631.66 (0.14)</td>
<td>−1612.62 (0.13)</td>
</tr>
<tr>
<td>5 (heavy)</td>
<td>−1605.77 (0.15)</td>
<td>−1616.95 (0.14)</td>
<td>−1600.18 (0.14)</td>
</tr>
<tr>
<td>2.1 (heavy)</td>
<td>−1622.31 (0.15)</td>
<td>−1629.38 (0.15)</td>
<td>−1618.37 (0.14)</td>
</tr>
</tbody>
</table>

Notes: Numerical standard errors are reported in parentheses. All estimates are obtained using 15 reduced TaRB-MH runs under the benchmark setting (p = 0.7, ω = 0.5), including 7 runs for the structural parameters and 8 runs for the volatility parameters. 10,000 posterior draws are made for each reduced run.

Table 2: Number of picks for each model specification

<table>
<thead>
<tr>
<th>DGP 1: regime-M with ν = 15</th>
<th>DGP 2: regime-F with φ = 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ν</td>
<td>regime-M</td>
</tr>
<tr>
<td>30 (light)</td>
<td>4</td>
</tr>
<tr>
<td>15 (fat)</td>
<td>15</td>
</tr>
<tr>
<td>5 (heavy)</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: The shock innovations have constant volatility under DGP 1 and follow Gaussian distribution under DGP 2. The number of simulations performed for each DGP is 20.

the fit, even after accounting for the other, and the model that fits best is regime-M with ν = 5 and φₙ = 0.95.

3.3.3 Simulation Evidence

This section furnishes additional evidence that demonstrates the reliability of the Chib (1995) method. For each regime, we generate 20 data sets of 100 quarterly observations using the subsample parameter estimates reported in Leeper, Traum and Walker (2017), which are also reproduced in Table A.3 of Appendix A. We then estimate three versions of each regime model that differ in the volatility specification. Based on the marginal likelihood estimates, we count the number of times that each of the six regime-volatility specifications is picked across the 20 simulated data sets. Table 2 summarizes the simulation results.
The first data generating process assumes that regime-M is in place and the shock innovations follow a multivariate Student-t distribution with fat tails, i.e., $\nu = 15$, and constant volatility. For each regime, we fit the model with three degrees of freedom: $\nu = 30$ (light), $\nu = 15$ (fat), and $\nu = 5$ (heavy). As can be seen from the left panel of Table 2, the correct degrees of freedom in 15 times. The correct policy regime is always picked.\footnote{Although not reported here, we have also computed the marginal likelihood by the MHM method as implemented in Justiniano and Primiceri (2008). Using a larger set of grid points for $\nu$, we find that nearly all data sets favor the lowest value of $\nu$. On the other hand, the computation based on the Chib (1995) method continues to find the correct value of $\nu$.}

The second data generating process assumes that regime-F is in place and the shock innovations follow a multivariate Gaussian distribution with moderate time-varying volatility, i.e., $\phi_s = 0.5$ for $s \in \{a, b, i, p, w, m, g, z\}$. For each regime, we fit the model with three degrees of persistence in volatility: $\phi_s = 0.1$ (weak), $\phi_s = 0.5$ (moderate), and $\phi_s = 0.9$ (strong). As shown in the right panel of Table 2, with only one exception, the data overwhelmingly favor weak to moderate degree of persistence in volatility under the true regime, which is preferred by all data sets over the alternative regime.\footnote{Like the case of regime-M with Student-t shocks, the computation based on the MHM method always overestimates the importance of stochastic volatility and selects $\phi_s = 0.9$. This result emerges despite the fact that all data sets are relatively short-lived and generated by a model with ‘close’ to constant volatility process.}

\section{Prediction}

Because a good understanding of the current and future state of the economy is essential to develop and implement sound economic policies, generating a predictive distribution for the future path of the economy constitutes an important part of the policy analysis. To facilitate this goal, our toolbox also produces, as a byproduct of the efficient TaRB-MH algorithm and the marginal likelihood computation by the Chib (1995) method, the joint predictive distribution for all observable variables at any forecasting horizon. For illustration purposes, Section 4.1 presents such a predictive distribution based on the best fitting model that is selected by the marginal likelihood comparison. Using the predictive distribution for wages as an example, Section 4.2 highlights the importance of allowing for non-Gaussian structural shocks with time-varying variances in the context of out-of-sample prediction. Finally, Section 4.3 evaluates the predictive performance by comparing the accuracy of point and density forecasts between a small-scale DSGE model and our high-dimensional DSGE model.
4.1 Sampling the Predictive Distribution

Let \( y_{1:T} \) be the data used to perform estimation, inference, and model selection. In addition, denote \( y_{T+1:T+h} \) the future path of the observables in the model economy. Then the predictive distribution is defined as

\[
p(y_{T+1:T+h}|y_{1:T}) = \int p(y_{T+1:T+h}|y_{1:T}, \theta) \cdot p(\theta|y_{1:T}) d\theta
\]

where the above integration is numerically approximated by first sampling the posterior \( p(\theta|y_{1:T}) \) a large number of times by the TaRB-MH algorithm and then simulating a future path \( y_{T+1:T+h}^{(g)} \) for each parameter draw. This amounts to moving model variables forward with \( \theta \) and \( y_{1:T} \). We call \( p(y_{i,T+h}|y_{1:T}) \) the \( h \)-step-ahead predictive distribution for the \( i \)th variable generated in period \( T \).

Now we generate the one-quarter-ahead predictive distribution for all eight observables based on the best fitting model as measured by the marginal likelihood. Throughout the entire forecasting horizon, this model operates under regime-M model and has Student-\( t \) shocks with stochastic volatilities. The first predictive distribution is generated using observations from the third quarter of 1967 to the fourth quarter of 2008, which is about six months before the Business Cycle Dating Committee of the National Bureau of Economic Research announces the end of the Great Recession. The forecasting horizon starts from the first quarter of 2009 and ends at second quarter of 2014, covering the whole economic recovery period from the Great Recession. Figure 6 displays the median forecasts with 90\% credible bands computed from the predictive distribution of regime-M over the full forecasting horizon. Overall the model performs quite well in tracking the recovery path of most observables.

4.2 Importance of Non-Gaussian Shocks

As the marginal likelihood comparison reveals, one needs a flexible way to model structural shocks in the model economy to explain the U.S. macroeconomic variables. The need of flexible distributional assumptions, such as Student-\( t \) shocks with stochastic volatility, can also be seen from our generated predictive densities as well. The left panel of Figure 7 plots the 90\% credible sets for wages based on two predictive distributions: one under Gaussian shocks with constant variance and another under Student-\( t \) shocks with time-varying variance. It is noticeable that the uncertainty bands are much wider for the model under Student-\( t \) shocks with time-varying variance. To understand this stark difference, the right panel of Figure 7 plots the time series of wages over the full sample. As pointed out by Champagne and Kurmann (2013), wages in the U.S. have become more volatile over the past 20 years. For example, the standard deviation of wages was 0.55 between 1955:Q1 and 1999:Q4, and 1.05 between 2000:Q1 and
2014:Q2. The heightened volatility of wages after 2000 is captured by the model with stochastic volatility, which adaptively widens the predictive distribution for wages. On the other hand, the model with constant variance misses this important change in volatility. In turn, its predictive distribution of wages is too narrow, underestimating the uncertainty in the future path of wages. In general, allowing for time-varying volatility produces similar improvements in the quality of DSGE-based interval and density forecasts (see, e.g., Diebold, Schorfheide and Shin (2017)). Thus, we expect that our toolbox, by making it easy to incorporate non-Gaussian errors and time-varying variances, will be useful for researchers and policymakers interested in better out-of-sample performance of DSGE models.

4.3 Predictive Performance Comparison

Although regime-M yields a higher marginal likelihood relative to regime-F, one may still be interested in knowing how the two policy regimes compare in terms of the quality of point and density forecasts over the forecasting horizon. It is also interesting to compare the forecasts from a medium-scale DSGE model with those from a small-scale one when both models are equipped with Student-$t$ shocks and stochastic volatility. Specifically, we compare the point and density forecasts generated from regimes M and F, and a small-scale DSGE model described in Appendix C. Starting from the first quarter of 2009, we recursively estimate the three models and generate one-quarter-ahead to two-year-ahead point and density forecasts until the second quarter of 2014, which results in 22 quarters of evaluation points for the one-quarter-ahead prediction. Since the small-scale model contains fewer observables, our evaluation exercise only considers the common set of observables: consumption growth, inflation rate, and federal funds rate. The aim of this comparison is to get information about the strengths and weaknesses of DSGE model elaborations.

In each model, for each observable and forecasting horizon, the point prediction is the mean of the corresponding predictive distribution. Let $\hat{y}_{i,t+h|t}$ denote the $h$-step-ahead point prediction for the $i$th variable generated at time $t$. To compare the quality of point forecasts, we report the root mean squared error (RMSE) for the point prediction

$$\text{RMSE}(\hat{y}_{i,t+h|t}, y_{i,t+h}) = \sqrt{\frac{1}{22} \sum_{t=2009:Q1}^{2014:Q2-h} (y_{i,t+h} - \hat{y}_{i,t+h|t})^2}$$

where $2014:Q2-h$ denotes $h$-quarters before 2014:Q4 and $y_{i,t+h}$ is the actual value for the $i$th variable at time $t + h$. The model with a smaller RMSE is preferred as the smaller forecast error is desirable. To compare the precision of predictive densities, we compute the continuous ranked probability score (CRPS),
Table 3: Point forecast comparison, RMSE

<table>
<thead>
<tr>
<th>Model</th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 4Q$</th>
<th>$h = 8Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Consumption growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-scale</td>
<td>0.32</td>
<td>0.28</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>Regime-M</td>
<td>0.44 (0.06)</td>
<td>0.48 (0.20)</td>
<td>0.50 (0.22)</td>
<td>0.48 (0.12)</td>
</tr>
<tr>
<td>Regime-F</td>
<td>0.40 (0.23)</td>
<td>0.39 (0.16)</td>
<td>0.36 (0.38)</td>
<td>0.37 (0.11)</td>
</tr>
<tr>
<td>(b) Inflation rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-scale</td>
<td>0.26</td>
<td>0.32</td>
<td>0.46</td>
<td>0.58</td>
</tr>
<tr>
<td>Regime-M</td>
<td>0.24 (0.40)</td>
<td>0.28 (0.31)</td>
<td>0.37 (0.12)</td>
<td>0.44 (0.04)</td>
</tr>
<tr>
<td>Regime-F</td>
<td>0.34 (0.00)</td>
<td>0.53 (0.08)</td>
<td>0.86 (0.10)</td>
<td>1.14 (0.14)</td>
</tr>
<tr>
<td>(c) Federal funds rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-scale</td>
<td>0.21</td>
<td>0.38</td>
<td>0.64</td>
<td>0.94</td>
</tr>
<tr>
<td>Regime-M</td>
<td>0.06 (0.00)</td>
<td>0.12 (0.01)</td>
<td>0.19 (0.01)</td>
<td>0.42 (0.01)</td>
</tr>
<tr>
<td>Regime-F</td>
<td>0.06 (0.00)</td>
<td>0.12 (0.02)</td>
<td>0.18 (0.02)</td>
<td>0.22 (0.01)</td>
</tr>
</tbody>
</table>

Notes: Each entry reports the RMSE based on the point forecast with the $p$-value of Diebold-Mariano (DM) tests of equal MSE in parentheses, obtained using the fixed-$b$ critical values. The standard errors entering the DM statistics are computed using the equal-weighted Type II discrete cosine transform (EWC) estimator with the truncation rule recommended by Lazarus et al. (2018).

which is defined as

$$CRPS(F_{i,t+h|t}(z), y_{i,t+h}) = \int_{\mathbb{R}} (F_{i,t+h|t}(z) - 1\{y_{i,t+h} \leq z\})^2 dz$$

where $F_{i,t+h|t}(z)$ is the $h$-step-ahead predictive cumulative distribution of the $i$th variable generated at time $t$. The CRPS is one of the proper scoring rules, and the predictive distribution with a smaller CRPS is preferred as this measure can be viewed as the divergence between the given predictive distribution and the unattainable oracle predictive distribution that puts a probability mass only on the realized value. Tables 3 and 4 report the RMSE and average CRPS, respectively, of consumption growth, inflation rate, and federal funds rate based on all three models.

Forecasts from the medium-scale models are significantly more accurate for the federal funds rate at all horizons. On the other hand, forecasts from the small-scale model are more accurate for the consumption growth at all horizons although the difference is only statistically significant at the one-quarter-ahead horizon. The major difference between regimes M and F lies in the inflation forecasts, and the model
Table 4: Density forecast comparison, average CRPS

<table>
<thead>
<tr>
<th>Model</th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 4Q$</th>
<th>$h = 8Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Consumption growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-scale</td>
<td>0.21</td>
<td>0.2</td>
<td>0.19</td>
<td>0.2</td>
</tr>
<tr>
<td>Regime-M</td>
<td>0.26 (0.08)</td>
<td>0.28 (0.28)</td>
<td>0.29 (0.31)</td>
<td>0.28 (0.20)</td>
</tr>
<tr>
<td>Regime-F</td>
<td>0.23 (0.40)</td>
<td>0.22 (0.41)</td>
<td>0.21 (0.72)</td>
<td>0.22 (0.46)</td>
</tr>
<tr>
<td>(b) Inflation rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-scale</td>
<td>0.15</td>
<td>0.18</td>
<td>0.26</td>
<td>0.34</td>
</tr>
<tr>
<td>Regime-M</td>
<td>0.14 (0.48)</td>
<td>0.17 (0.53)</td>
<td>0.23 (0.29)</td>
<td>0.28 (0.11)</td>
</tr>
<tr>
<td>Regime-F</td>
<td>0.20 (0.00)</td>
<td>0.31 (0.03)</td>
<td>0.52 (0.05)</td>
<td>0.69 (0.08)</td>
</tr>
<tr>
<td>(c) Federal funds rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-scale</td>
<td>0.13</td>
<td>0.24</td>
<td>0.43</td>
<td>0.67</td>
</tr>
<tr>
<td>Regime-M</td>
<td>0.04 (0.00)</td>
<td>0.07 (0.02)</td>
<td>0.13 (0.01)</td>
<td>0.27 (0.01)</td>
</tr>
<tr>
<td>Regime-F</td>
<td>0.04 (0.00)</td>
<td>0.07 (0.02)</td>
<td>0.12 (0.02)</td>
<td>0.19 (0.01)</td>
</tr>
</tbody>
</table>

Notes: Each entry reports the average CRPS over the evaluation period with the $p$-value of Diebold-Mariano (DM) tests of equal CRPS in parentheses, obtained using the fixed-$b$ critical values. The standard errors entering the DM statistics are computed using the equal-weighted Type II discrete cosine transform (EWC) estimator with the truncation rule recommended by Lazarus et al. (2018).

under regime-M produces forecasts with lower RMSEs (CRPSs). The RMSE (CRPS) gaps get wider as the forecasting horizon extends, and the RMSE (CRPS) from regime-M becomes more than half of that from regime-F. In contrast, the forecasts from regime-F fare slightly better for the consumption growth at all horizons and are most accurate for the federal funds rate at the two-year-ahead horizon.

In sum, there is no clear winner in this comparison. The small-scale model performs better for forecasting the consumption growth. The medium-scale model, on the other hand, performs the best under regime-M for forecasting the inflation rate but does not generate better forecasts under regime-F except for forecasting the federal funds rate in the long run. Although the evaluation period is too short-lived to draw a definite conclusion, the results from this out-of-sample forecasting exercise indicate that there is still room for improvement even for the more complex models.
5 Concluding Remarks

We have given pointers on the fitting and comparison of high-dimensional DSGE models with latent variables and shown that the TaRB-MH algorithm of Chib and Ramamurthy (2010) allows for the efficient estimation of such models. We emphasize the importance of training sample priors, which is new in the DSGE context, and the use of the Student-t, as opposed to the normal family, as the prior distribution for location type parameters. In addition, we show that the method of Chib (1995) and Chib and Jeliazkov (2001), in conjunction with a parallel implementation of the required reduced MCMC runs, can be used to get reliable and fast estimates of the marginal likelihood. With the help of a user-friendly MATLAB package, these methods can be readily employed in academic and central bank applications to conduct DSGE model comparisons, impulse response and variance decomposition analyses, and to generate point and density forecasts. Finally, in ongoing work, we are applying this toolkit, without modification and any erosion in performance, to open economy DSGE models that contain more than twice as many parameters and latent variables as the model showcased in this paper. Findings from this analysis will be reported elsewhere.

References


Appendix

Siddhartha Chib, Minchul Shin, and Fei Tan

Appendix A  Leeper-Traum-Walker Model

A.1 Equilibrium System

Since the economy features a stochastic trend induced by the permanent technology shock $A_t$, some variables are not stationary. To induce stationarity, we therefore detrend these variables as: $y_t \equiv \frac{y_t}{A_t}$, $c_t^* \equiv \frac{C_t^*}{A_t}$, $k_t \equiv \frac{K_t}{A_t}$, $\bar{k}_t \equiv \frac{\bar{K}_t}{A_t}$, $i_t \equiv \frac{I_t}{A_t}$, $g_t \equiv \frac{G_t}{A_t}$, $z_t \equiv \frac{Z_t}{A_t}$, $b_t \equiv \frac{P_t B_t}{P_t A_t}$, $w_t \equiv \frac{W_t}{P_t A_t}$, $\lambda_t \equiv \Lambda_t A_t$. The model’s equilibrium system in terms of the detrended variables can be summarized as follows.

Production function:

$$y_t \Delta^p_t = k_t^\alpha (L^d_t)^{1-\alpha} - \Omega$$  \hspace{1cm} (A.1)

Capital-labor ratio:

$$\frac{k_t}{L^d_t} = \frac{w_t \alpha}{r_t^k 1 - \alpha}$$  \hspace{1cm} (A.2)

Real marginal cost:

$$mc_t = (1 - \alpha)^{\alpha - 1} \alpha^{-\alpha} (r_t^k)^\alpha w_t^{1-\alpha}$$  \hspace{1cm} (A.3)

Intermediate goods producer’s optimal price:

$$\mathbb{E}_t \left[ \sum_{k=0}^{\infty} (\omega_p^k)^k \lambda_{t+k} \bar{y}_{t+k} \left( \pi_t^p \prod_{s=1}^{k} \left( \frac{\pi_{t+s-1}}{\pi} \right)^{\chi_p} \frac{\pi}{\pi_{t+s}} - (1 + \eta_{t+k}^p) mc_{t+k} \right) \right] = 0$$  \hspace{1cm} (A.4)

Evolution of aggregate price index:

$$1 = (1 - \omega_p)(\pi_t^*)^{-\frac{1}{\eta_p^t}} + \omega_p \left( \frac{\pi_{t-1}^p}{\pi} \frac{\pi}{\pi_t} \right)^{-\frac{1}{\eta_p^t}}$$  \hspace{1cm} (A.5)
Optimal consumption:

$$\lambda_t(1 + \tau_t^C) = \frac{u_t^b}{c_t^* - \theta c_{t-1}^* e^{-u_t^g}}$$  \hspace{1cm} (A.6)$$

Composite consumption:

$$c_t^* = c_t + \alpha G_y t$$  \hspace{1cm} (A.7)$$

Consumption Euler equation:

$$\lambda_t = \beta R_t \mathbb{E}_t \left[ \frac{\lambda_{t+1} e^{-u_{t+1}^g}}{\pi_{t+1}} \right]$$  \hspace{1cm} (A.8)$$

Bond pricing relation:

$$P_t^B = \mathbb{E}_t \left[ 1 + \rho P_{t+1}^B \right]$$  \hspace{1cm} (A.9)$$

Optimal capital utilization:

$$(1 - \tau^K_t) r_t^k = \psi'(v_t)$$  \hspace{1cm} (A.10)$$

Optimal physical capital:

$$q_t = \beta \mathbb{E}_t \left[ \frac{\lambda_{t+1} e^{-u_{t+1}^g}}{\lambda_t} \left( (1 - \tau^K_{t+1}) r_{t+1}^k v_{t+1} - \psi(v_{t+1}) + (1 - \delta) q_{t+1} \right) \right]$$  \hspace{1cm} (A.11)$$

where $q_t$ is the real price of capital in terms of consumption goods (i.e., Tobin’s Q).

Optimal investment:

$$1 = q_t \tilde{u}_t^i \left( 1 - S \left( \frac{i_t e^{u_t^g}}{i_{t-1}} \right) - S' \left( \frac{i_t e^{u_t^g}}{i_{t-1}} \right) \right) + \beta \mathbb{E}_t \left[ q_{t+1} \frac{\lambda_{t+1} e^{-u_{t+1}^g}}{\lambda_t} \tilde{u}_{t+1}^k \left( \frac{i_{t+1} e^{u_{t+1}^g}}{i_t} \right) \left( \frac{i_{t+1} e^{u_{t+1}^g}}{i_t} \right)^2 \right]$$  \hspace{1cm} (A.12)$$

Effective capital:

$$k_t = v_t \tilde{k}_{t-1} e^{-u_t^g}$$  \hspace{1cm} (A.13)$$
Law of motion for capital:

\[ \tilde{k}_t = (1 - \delta)e^{-u_t^a} \tilde{k}_{t-1} + \bar{u}_t \left( 1 - S \left( \frac{i_t e^{u_t^q}}{i_{t-1}} \right) \right) i_t \]  

(A.14)

Optimal wage:

\[ \mathbb{E}_t \left[ \sum_{k=0}^\infty (\beta \omega_w)^k \lambda_{t+k} \bar{L}_{t+k} \left( \frac{w_{t+s}^a}{w_{t+k}^a} \prod_{s=1}^k \left( \frac{\pi_{t+s-1} e^{u_{t+s-1}^q}}{\pi e^{\gamma}} \right)^{x_w} \frac{\pi e^{\gamma}}{\pi_{t+s} e^{u_{t+s}^q}} - \frac{(1 + \eta_{t+k}^w) u_{t+k}^b \bar{L}_{t+k}^c}{(1 - \tau_{t+k} L) \lambda_{t+k}} \right) \right] = 0 \]  

(A.15)

where

\[ \bar{L}_{t+k} = \left[ \frac{w_{t+s}^a}{w_{t+k}^a} \prod_{s=1}^k \left( \frac{\pi_{t+s-1} e^{u_{t+s-1}^q}}{\pi e^{\gamma}} \right)^{x_w} \frac{\pi e^{\gamma}}{\pi_{t+s} e^{u_{t+s}^q}} \right]^{-\frac{1 + \eta_{t+k}^w}{\eta_{t+k}^w}} L_{t+k}^d \]  

(A.16)

Evolution of aggregate wage index:

\[ w_t^{-\frac{1}{\eta_w}} = (1 - \omega_w) (w_t^a)^{-\frac{1}{\eta_w}} + \omega_w \left[ \left( \frac{\pi_{t-1} e^{u_{t-1}^q}}{\pi e^{\gamma}} \right)^{x_w} \left( \frac{\pi e^{\gamma}}{\pi_{t} e^{u_{t}^q}} \right)^{w_{t-1}} \right]^{-\frac{1}{\eta_w}} \]  

(A.17)

Government budget constraint:

\[ b_t + \tau^K r_t^k \tilde{k}_t + \tau^L w_t L_t + \tau^C c_t = \frac{1 + \rho P_t^B b_{t-1}}{P_{t-1}^B} \frac{i_{t-1} e^{u_{t-1}^q}}{\pi e^{u_{t-1}^q}} + g_t + z_t \]  

(A.18)

Aggregate resource constraint:

\[ y_t = c_t + i_t + g_t + \psi(v_t) \tilde{k}_{t-1} e^{-u_t^a} \]  

(A.19)

A.2 Steady States

To solve for the steady states, we calibrate \( \beta = 0.99, \alpha = 0.33, \delta = 0.025, \) the average maturity of government bond portfolio \( AD = 20, \eta^w = \eta^p = 0.14, g/y = 0.11, b/y = 1.47, \tau^C = 0.023, \tau^K = 0.218, \) and \( \tau^L = 0.186. \) By assumption, \( v = 1, \psi(v) = 0, \) and \( S(e^\gamma) = S'(e^\gamma) = 0. \) The remaining steady states can be solved as follows.

From \( AD: \)

\[ \rho = \left( 1 - \frac{1}{AD} \right) \frac{1}{\beta} \]  

(A.20)
From (A.8):

\[ R = \frac{e^{\gamma \pi}}{\beta} \]  \hspace{1cm} (A.21)

From (A.9):

\[ P^B = \frac{\beta}{e^{\gamma \pi} - \rho \beta} \]  \hspace{1cm} (A.22)

From \( \tilde{w}^i = 1 \) and (A.12):

\[ q = 1 \]  \hspace{1cm} (A.23)

From (A.11):

\[ r^k = \frac{e^{\gamma \pi} - (1 - \delta)}{1 - \tau^K} \]  \hspace{1cm} (A.24)

From (A.10):

\[ \psi'(1) = r^k (1 - \tau^K) \]  \hspace{1cm} (A.25)

From (A.4):

\[ mc = \frac{1}{1 + \eta_p} \]  \hspace{1cm} (A.26)

From (A.3):

\[ w = [mc(1 - \alpha)^{1 - \alpha} \alpha \alpha^k - \alpha]^{\frac{1}{1 - \alpha}} \]  \hspace{1cm} (A.27)

From (A.2):

\[ \frac{k}{L} = \frac{\alpha w}{1 - \alpha r^k} \]  \hspace{1cm} (A.28)

From \( \Delta^p = 1 \), the final goods producer’s zero profit condition, and (A.1):

\[ \frac{\Omega}{L} = \left( \frac{k}{L} \right)^{\alpha} - r^k \frac{k}{L} - w \]  \hspace{1cm} (A.29)
From (A.29):

\[
y \frac{L}{L} = \left( \frac{k}{L} \right)^\alpha - \frac{\Omega}{L} \tag{A.30}
\]

From (A.13):

\[
\bar{k} = ke^\gamma \tag{A.31}
\]

From (A.14):

\[
\frac{i}{L} = \left[ 1 - (1 - \delta)e^{-\gamma} \right] e^\gamma \frac{k}{L} \tag{A.32}
\]

From (A.19):

\[
\frac{c}{L} = \frac{y}{L} \left( 1 - \frac{g}{y} \right) - \frac{i}{L} \tag{A.33}
\]

From (A.18):

\[
\frac{z}{L} = \left[ 1 - \left( \frac{R}{\pi e^\gamma} \right) \frac{b}{y} - \frac{g}{y} \right] \frac{y}{L} + \tau^C c \frac{c}{L} + \tau^L \frac{w}{L} + \tau^K \frac{k}{L} \tag{A.34}
\]

From (A.7):

\[
\frac{c^*}{L} = \frac{c}{L} + \alpha G \frac{g}{y} \frac{y}{L} \tag{A.35}
\]

From (A.15) and (A.17):

\[
L = \frac{\frac{w(1 - \tau^L)}{(1 + \tau^C)(1 + \eta^w)(1 - \theta e^{-\gamma})} \frac{1}{L}}{\frac{1}{L}} \tag{A.36}
\]

from which all level variables can be calculated from the steady state ratios given above.

### A.3 Tables and Figures

- Table A.1 lists the marginal prior distributions and the true values for the high-dimensional DSGE model under regime-M.

- Table A.2 summarizes the posterior parameter estimates for the model of best fit.

- Table A.3 reproduces the subsample posterior parameter estimates reported in Leeper, Traum and
Walker (2017), which are used to generate the simulated data sets in Section 3.3.3.

- Figures A.1–A.2 display the autocorrelation function for each model parameter.
- Figures A.3–A.4 compare the model’s estimated parameters and volatilities with their true values, respectively.

**Appendix B  Guide to MATLAB Package**

We provide a library of numerical subroutines called ‘TaRB-t-SV’ that implements the TaRB-MH algorithm discussed in the main text. All of the subroutines are programmed in MATLAB or compiled as executable (MEX) functions. The MEX versions of these subroutines are included in the subfolders ‘mex/mac’ (for Mac OS users) and ‘mex/win’ (for Windows OS users) to improve computational time. The subfolder ‘utils’ contains various supporting packages that will be utilized by the main programs below.

The TaRB-MH algorithm can be broken down into a sequence of easily implementable steps. These steps can be executed with the following set of functions in the main folder ‘TaRB-t-SV’, which are extensively annotated:

- **tarb_demo.m**—main function that estimates one of the DSGE models specified under the subfolder ‘user’.

  - **tarb_spec.m**—admits all user-specified optional settings for the TaRB-MH algorithm. Like MATLAB built-in functions, each setting enters as a string-value pair.

  - **chain_init.m**—finds the posterior mode and its associated inverse Hessian matrix.

  - **tarb_full.m**—implements the full MCMC run for parameter estimation.

  - **tarb_reduce.m**—implements the reduced MCMC runs for marginal likelihood estimation.

  - **readme.m**—displays general information about TaRB-t-SV.

  - **test_bench.m**—preliminary DSGE model analysis.

  The subfolder ‘user/ltw17’ contains the following files for the Leeper-Traum-Walker model, which can be modified as needed for alternative model specifications:

  - **data.txt**—prepared in matrix form where each row corresponds to the observations for a given period.
• user_parvar.m—defines the model parameters, variables, shock innovations, forecast errors, and observables.

• user_mod.m—defines the model and measurement equations.

• user_ssp.m—defines the steady state, implied, and/or fixed parameters.

• user_svp.m—defines the stochastic volatility parameters.

To estimate the Leeper-Traum-Walker model with Student-$t$ shocks and stochastic volatility, for example, simply set the MATLAB current directory to the main folder ‘TaRB-t-SV’ and run the following blocks of code in tarb_demo.m.

1. Specify the model, data, and save directories and generate, if needed, the required MEX files that are compatible with the user machine.

   ```matlab
   \%
   \% Housekeeping
   clear
   close all
   clc
   readme

   \% User search path & mex files
   modpath = ['user ' filesep 'ltw17 '];
   datpath = ['user ' filesep 'ltw17 ' filesep 'data.txt '];
   savepath = ['user ' filesep 'ltw17 '];
   spec = tarb_spec([], 'modpath', modpath, 'datpath', datpath, 'savepath', savepath);
   OneFileToMexThemAll
   \%
   ```

2. Find the posterior mode and evaluate the corresponding inverse of the Hessian matrix.

   ```matlab
   \%
   \% Find posterior mode
   sa_spec = optimoptions(@simulannealbnd,... \% simulated annealing
                             'TemperatureFcn', @temperaturefast,...
                             'InitialTemperature', 2,...
                             'TolFun', 1e-3,...
                             'MaxTime', 10,...
                             'Display', 'iter',...
                             'DisplayInterval', 10);
   dof = 5; \% Shock degrees of freedom
   ```
3. Sample the posterior distribution by the TaRB-MH algorithm of Chib and Ramamurthy (2010). The estimation results will be stored in the MATLAB data file `tarb_full.mat`, which is saved to the subfolder ‘user/ltw17’.

4. Conditional on the estimated latent variables, find the posterior mode again and evaluate the corresponding inverse of the Hessian matrix.
5. Compute the marginal likelihood by the Chib (1995) and Chib and Jeliazkov (2001) method. The estimation results will be stored in the MATLAB data file `tarb_reduce.mat`, which is saved to the subfolder ‘user/ltw17’.

```matlab
%% Marginal likelihood (reduced run)
spec = tarb_spec(spec,’sa’,[]);
B = 7; % number of blocks
tarb_reduce(M,burn,B,spec)
```

### Appendix C  Small-Scale DSGE Model

A log-linear approximation to the model’s equilibrium conditions around the steady state can be summarized as follows:

**Dynamic IS equation:**
\[
\dot{c}_t = \mathbb{E}_t \dot{c}_{t+1} - \frac{1}{\tau} (\dot{R}_t - \mathbb{E}_t \dot{\pi}_{t+1} - \mathbb{E}_t \dot{z}_{t+1})
\]

**New Keynesian Phillips curve:**
\[
\dot{\pi}_t = \beta \mathbb{E}_t \dot{\pi}_{t+1} + \kappa \dot{c}_t
\]

**Monetary policy:**
\[
\dot{R}_t = \rho_R \dot{R}_{t-1} + (1 - \rho_R) [\psi_1 \dot{\pi}_t + \psi_2 (\dot{c}_t + \dot{g}_t)] + \epsilon_{R,t}
\]

**Technology shock:**
\[
\dot{z}_t = \rho_z \dot{z}_{t-1} + \epsilon_{z,t}
\]

**Government spending shock:**
\[
\dot{g}_t = \rho_g \dot{g}_{t-1} + \epsilon_{g,t}
\]

Here $\tau > 0$ is the coefficient of relative risk aversion, $0 < \beta < 1$ is the discount factor, $\kappa > 0$ is the slope of the new Keynesian Phillips curve, $\psi_1 > 0$ and $\psi_2 > 0$ are the policy rate responsive coefficients, and $0 \leq \rho_R, \rho_z, \rho_g < 1$. Moreover, $c_t$ is the detrended consumption, $\pi_t$ is the inflation between periods $t-1$ and $t$, $R_t$ is the nominal interest rate, $z_t$ is an exogenous shock to the labor-augmenting technology that grows on average at the rate $\gamma$, and $g_t$ is an exogenous government spending shock. Finally, the shock innovations $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}, \epsilon_{g,t}]'$ follow a multivariate Student-$t$ distribution, i.e., $\epsilon_t \sim t_\nu(0, \Sigma_t)$, where $\Sigma_t = \text{diag}(e^{h_t})$ and each element of $h_t = [h_t^R, h_t^z, h_t^g]'$ follows a stationary process

\[
h_t^s = (1 - \phi_s)\mu_s + \phi_s h_{t-1}^s + \eta_t^s, \quad \eta_t^s \sim \mathcal{N}(0, \omega_s^2), \quad s \in \{R, z, g\}
\]

The model is estimated over three observables, including log difference of consumption, log inflation, and log nominal interest rate. The observables are linked to the model variables via the following measurement
equations

\[

dlCons_t \\
llInfl_t \\
lFedFunds_t \\
\begin{bmatrix}
\gamma^{(Q)} \\
\pi^{(Q)} \\
\pi^{(Q)} + r^{(Q)} + \gamma^{(Q)}
\end{bmatrix} + \begin{bmatrix}
\hat{c}_t - \hat{c}_{t-1} + \hat{z}_t \\
\hat{\pi}_t \\
\hat{R}_t
\end{bmatrix}
\]

(C.7)

where \((\gamma^{(Q)}, \pi^{(Q)}, r^{(Q)})\) are connected to the model’s steady states via \(\gamma = 1 + \gamma^{(Q)}/100, \beta = 1/(1 + r^{(Q)}/100), \) and \(\pi = 1 + \pi^{(Q)}/100.\) Table C.1 lists the marginal prior distributions for the small-scale DSGE model parameters.

Table C.1: Priors for small-scale DSGE model parameters

<table>
<thead>
<tr>
<th>Structural parameters</th>
<th>Volatility parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Density (1, 2)</td>
</tr>
<tr>
<td>(\tau)</td>
<td>(G (2.00, 0.50))</td>
</tr>
<tr>
<td>(\kappa)</td>
<td>(G (0.20, 0.10))</td>
</tr>
<tr>
<td>(\psi_1)</td>
<td>(N (1.50, 0.20))</td>
</tr>
<tr>
<td>(\psi_2)</td>
<td>(G (0.125, 0.05))</td>
</tr>
<tr>
<td>(\gamma^{(Q)})</td>
<td>(N (0.40, 0.05))</td>
</tr>
<tr>
<td>(\pi^{(Q)})</td>
<td>(G (0.75, 0.25))</td>
</tr>
<tr>
<td>(\rho_R)</td>
<td>(\beta (0.50, 0.20))</td>
</tr>
<tr>
<td>(\rho_z)</td>
<td>(\beta (0.50, 0.20))</td>
</tr>
<tr>
<td>(\rho_g)</td>
<td>(\beta (0.50, 0.20))</td>
</tr>
</tbody>
</table>

**Notes:** See Table A.1.
Figure 1: Distributions of simulated selected quantities obtained by sampling the prior, and then the outcomes given drawings from the prior. Notes: Each panel compares the resulting densities under Gaussian shocks with constant volatility (red dashed line) with that under Student-\(t\) shocks with stochastic volatility (shaded area). Vertical lines denote the real data counterparts.
Figure 2: Marginal prior and posterior distributions of each structural parameter. Notes: Each panel compares the prior (red dashed line) with the posterior (shaded area). Vertical lines denote posterior means. The kernel smoothed posterior densities are estimated using 10,000 TaRB-MH draws.
Figure 3: Stochastic volatility of each shock innovation. Notes: Blue dashed lines denote median estimates, while blue shaded areas delineate 90% highest posterior density bands. Vertical bars indicate recessions as designated by the National Bureau of Economic Research.
Figure 4: Inefficiency factor of each parameter. Notes: The horizontal axis indicates parameter indices. The vertical line separates the structural (indexed by 1–25) and volatility (indexed by 26–49) parameters.
Figure 5: Recursive posterior ordinates and marginal likelihood. Notes: Ordinates 1–7 (8–15) correspond to structural (volatility) parameters. The last panel depicts the estimated marginal likelihood. Black solid lines (with cross marker) correspond to the benchmark setting \((p = 0.7, \omega = 0.5)\). All estimates are in logarithm scale.
Figure 6: DSGE model forecast of each observable. Notes: Each panel compares the one-quarter-ahead posterior forecast of regime M with real data (black solid lines). Blue dashed lines denote median forecasts, while blue shaded areas delineate 90% highest predictive density bands.
Figure 7: Predictive distribution and data for wages. Notes: Predictive distributions are constructed using data up to 2008:Q4. The one-step-ahead prediction corresponds to 2009:Q1. The left panel plots 90% prediction intervals of regime-M under Gaussian shocks with constant variance (labeled ‘CV-N’, thick line) and Student-\(t\) shocks with time-varying variance (labeled ‘SV-t’, thin line). The right panel plots the time series of wages (solid line). Dashed lines delineate two standard deviations from the mean for two sub-samples, i.e., pre- and post-2000.
<table>
<thead>
<tr>
<th>Name</th>
<th>Structural parameters</th>
<th>Volatility parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Density (1, 2)</td>
<td>DGP</td>
</tr>
<tr>
<td>100γ</td>
<td>N (0.40, 0.05)</td>
<td>0.25</td>
</tr>
<tr>
<td>ξ</td>
<td>G (2.00, 0.50)</td>
<td>1.77</td>
</tr>
<tr>
<td>h</td>
<td>B (0.50, 0.20)</td>
<td>0.99</td>
</tr>
<tr>
<td>αG</td>
<td>U (-1.75, 1.75)</td>
<td>-0.25</td>
</tr>
<tr>
<td>ψ</td>
<td>B (0.60, 0.15)</td>
<td>0.16</td>
</tr>
<tr>
<td>s</td>
<td>N (6.00, 1.50)</td>
<td>5.46</td>
</tr>
<tr>
<td>ωp</td>
<td>B (0.50, 0.20)</td>
<td>0.92</td>
</tr>
<tr>
<td>ωw</td>
<td>B (0.50, 0.20)</td>
<td>0.91</td>
</tr>
<tr>
<td>Xp</td>
<td>B (0.50, 0.20)</td>
<td>0.06</td>
</tr>
<tr>
<td>Xw</td>
<td>B (0.50, 0.20)</td>
<td>0.18</td>
</tr>
<tr>
<td>φπ, regime-M</td>
<td>N (1.50, 0.20)</td>
<td>0.90</td>
</tr>
<tr>
<td>φπ, regime-F</td>
<td>B (0.50, 0.15)</td>
<td>n/a</td>
</tr>
<tr>
<td>φg</td>
<td>N (0.125, 0.05)</td>
<td>0.10</td>
</tr>
<tr>
<td>ρr</td>
<td>B (0.50, 0.20)</td>
<td>0.71</td>
</tr>
<tr>
<td>γg, regime-M</td>
<td>N (0.15, 0.10)</td>
<td>0.26</td>
</tr>
<tr>
<td>γz, regime-M</td>
<td>N (0.15, 0.10)</td>
<td>-0.11</td>
</tr>
<tr>
<td>ρg</td>
<td>B (0.50, 0.20)</td>
<td>0.98</td>
</tr>
<tr>
<td>ρz, regime-F</td>
<td>B (0.50, 0.20)</td>
<td>n/a</td>
</tr>
<tr>
<td>ρea</td>
<td>B (0.50, 0.20)</td>
<td>0.23</td>
</tr>
<tr>
<td>ρeb</td>
<td>B (0.50, 0.20)</td>
<td>0.40</td>
</tr>
<tr>
<td>ρei</td>
<td>B (0.50, 0.20)</td>
<td>0.69</td>
</tr>
<tr>
<td>ρep</td>
<td>B (0.50, 0.20)</td>
<td>0.74</td>
</tr>
<tr>
<td>ρew</td>
<td>B (0.50, 0.20)</td>
<td>0.18</td>
</tr>
<tr>
<td>ρem</td>
<td>B (0.50, 0.15)</td>
<td>0.39</td>
</tr>
<tr>
<td>ρeg</td>
<td>B (0.50, 0.15)</td>
<td>0.13</td>
</tr>
<tr>
<td>ρez, regime-F</td>
<td>B (0.50, 0.15)</td>
<td>n/a</td>
</tr>
<tr>
<td>L</td>
<td>N (468, 5.00)</td>
<td>481.12</td>
</tr>
<tr>
<td>π</td>
<td>N (0.75, 0.25)</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Notes: Density (1, 2) refer to Gamma (G), Normal (N), Beta (B), and Student-tν (ν = 2.1 degrees of freedom) distributions with means and standard deviations indicated in parentheses; Uniform (U) distribution with lower and upper bounds; Inverse-Gamma type-1 (IG-1) distribution with parameters ν and s, where \( p(\sigma) \propto \sigma^{-\nu-1} \exp \left(-\frac{\nu s^2}{2\sigma^2}\right) \); Inverse-Gamma type-2 (IG-2) distribution with parameters α and β, where \( p(\omega^2) \propto (\omega^2)^{-\alpha-1} \exp \left(-\frac{\beta}{\omega^2}\right) \). The effective priors are truncated at the boundary of the determinacy region.
<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>90% HPD</th>
<th>Ineff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100\gamma$</td>
<td>0.22</td>
<td>[0.16, 0.28]</td>
<td>4.4</td>
</tr>
<tr>
<td>$\xi$</td>
<td>1.65</td>
<td>[0.77, 2.41]</td>
<td>12.3</td>
</tr>
<tr>
<td>$h$</td>
<td>0.99</td>
<td>[0.98, 0.99]</td>
<td>103.6</td>
</tr>
<tr>
<td>$\alpha_G$</td>
<td>-0.09</td>
<td>[-0.19, -0.00]</td>
<td>5.2</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.20</td>
<td>[0.11, 0.29]</td>
<td>9.0</td>
</tr>
<tr>
<td>$s$</td>
<td>7.35</td>
<td>[5.14, 9.57]</td>
<td>6.2</td>
</tr>
<tr>
<td>$\omega_p$</td>
<td>0.91</td>
<td>[0.89, 0.94]</td>
<td>20.9</td>
</tr>
<tr>
<td>$\omega_w$</td>
<td>0.84</td>
<td>[0.79, 0.89]</td>
<td>13.6</td>
</tr>
<tr>
<td>$\chi_p$</td>
<td>0.09</td>
<td>[0.01, 0.16]</td>
<td>6.3</td>
</tr>
<tr>
<td>$\chi_w$</td>
<td>0.07</td>
<td>[0.02, 0.11]</td>
<td>3.4</td>
</tr>
<tr>
<td>$\phi_x$</td>
<td>1.00</td>
<td>[0.71, 1.31]</td>
<td>20.6</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>0.17</td>
<td>[0.13, 0.22]</td>
<td>7.2</td>
</tr>
<tr>
<td>$\rho_r$</td>
<td>0.76</td>
<td>[0.70, 0.82]</td>
<td>15.7</td>
</tr>
<tr>
<td>$\gamma_g$</td>
<td>0.27</td>
<td>[0.16, 0.37]</td>
<td>17.8</td>
</tr>
<tr>
<td>$\gamma_z$</td>
<td>-0.07</td>
<td>[-0.20, 0.07]</td>
<td>39.9</td>
</tr>
<tr>
<td>$\rho_g$</td>
<td>0.98</td>
<td>[0.97, 0.99]</td>
<td>12.0</td>
</tr>
<tr>
<td>$\rho_ea$</td>
<td>0.24</td>
<td>[0.12, 0.35]</td>
<td>3.5</td>
</tr>
<tr>
<td>$\rho_{eb}$</td>
<td>0.38</td>
<td>[0.26, 0.50]</td>
<td>4.5</td>
</tr>
<tr>
<td>$\rho_{ei}$</td>
<td>0.64</td>
<td>[0.54, 0.73]</td>
<td>15.1</td>
</tr>
<tr>
<td>$\rho_{ep}$</td>
<td>0.74</td>
<td>[0.66, 0.80]</td>
<td>24.0</td>
</tr>
<tr>
<td>$\rho_{ew}$</td>
<td>0.45</td>
<td>[0.35, 0.58]</td>
<td>31.1</td>
</tr>
<tr>
<td>$\rho_{em}$</td>
<td>0.57</td>
<td>[0.46, 0.68]</td>
<td>16.1</td>
</tr>
<tr>
<td>$\rho_{eg}$</td>
<td>0.06</td>
<td>[0.01, 0.11]</td>
<td>3.9</td>
</tr>
<tr>
<td>$\bar{L}$</td>
<td>474.43</td>
<td>[473.31, 475.50]</td>
<td>23.9</td>
</tr>
<tr>
<td>$\bar{\pi}$</td>
<td>0.27</td>
<td>[0.00, 0.47]</td>
<td>3.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>90% HPD</th>
<th>Ineff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_a$</td>
<td>-0.33</td>
<td>[-0.74, 0.08]</td>
<td>4.9</td>
</tr>
<tr>
<td>$\mu_b$</td>
<td>6.41</td>
<td>[5.73, 7.06]</td>
<td>69.0</td>
</tr>
<tr>
<td>$\mu_i$</td>
<td>-0.93</td>
<td>[-1.56, -0.35]</td>
<td>5.8</td>
</tr>
<tr>
<td>$\mu_p$</td>
<td>-5.55</td>
<td>[-6.02, -5.04]</td>
<td>9.9</td>
</tr>
<tr>
<td>$\mu_w$</td>
<td>-3.63</td>
<td>[-4.27, -2.98]</td>
<td>13.7</td>
</tr>
<tr>
<td>$\mu_m$</td>
<td>-4.59</td>
<td>[-5.46, -3.76]</td>
<td>2.7</td>
</tr>
<tr>
<td>$\mu_g$</td>
<td>0.79</td>
<td>[0.45, 1.18]</td>
<td>5.1</td>
</tr>
<tr>
<td>$\mu_z$</td>
<td>1.78</td>
<td>[1.09, 2.46]</td>
<td>43.0</td>
</tr>
<tr>
<td>$\phi_a$</td>
<td>0.95</td>
<td>[0.94, 0.97]</td>
<td>2.0</td>
</tr>
<tr>
<td>$\phi_b$</td>
<td>0.95</td>
<td>[0.93, 0.96]</td>
<td>2.0</td>
</tr>
<tr>
<td>$\phi_i$</td>
<td>0.95</td>
<td>[0.94, 0.97]</td>
<td>1.7</td>
</tr>
<tr>
<td>$\phi_p$</td>
<td>0.95</td>
<td>[0.93, 0.97]</td>
<td>1.9</td>
</tr>
<tr>
<td>$\phi_w$</td>
<td>0.95</td>
<td>[0.93, 0.97]</td>
<td>1.7</td>
</tr>
<tr>
<td>$\phi_m$</td>
<td>0.95</td>
<td>[0.94, 0.97]</td>
<td>2.0</td>
</tr>
<tr>
<td>$\phi_g$</td>
<td>0.95</td>
<td>[0.93, 0.96]</td>
<td>2.3</td>
</tr>
<tr>
<td>$\phi_z$</td>
<td>0.95</td>
<td>[0.93, 0.97]</td>
<td>2.0</td>
</tr>
<tr>
<td>$\omega_a^2$</td>
<td>0.02</td>
<td>[0.01, 0.04]</td>
<td>15.0</td>
</tr>
<tr>
<td>$\omega_b^2$</td>
<td>0.02</td>
<td>[0.01, 0.04]</td>
<td>14.1</td>
</tr>
<tr>
<td>$\omega_i^2$</td>
<td>0.05</td>
<td>[0.01, 0.10]</td>
<td>13.1</td>
</tr>
<tr>
<td>$\omega_p^2$</td>
<td>0.03</td>
<td>[0.01, 0.06]</td>
<td>8.7</td>
</tr>
<tr>
<td>$\omega_w^2$</td>
<td>0.05</td>
<td>[0.01, 0.08]</td>
<td>12.6</td>
</tr>
<tr>
<td>$\omega_m^2$</td>
<td>0.10</td>
<td>[0.03, 0.17]</td>
<td>7.5</td>
</tr>
<tr>
<td>$\omega_g^2$</td>
<td>0.02</td>
<td>[0.01, 0.03]</td>
<td>24.9</td>
</tr>
<tr>
<td>$\omega_z^2$</td>
<td>0.03</td>
<td>[0.01, 0.05]</td>
<td>19.2</td>
</tr>
</tbody>
</table>

**Notes:** The posterior means and 90% highest probability density (HPD) intervals are computed using 10,000 posterior draws from the TaRB-MH algorithm. The estimated model has regime-M in place and Student-t shocks with $\nu = 5$ degrees of freedom.
<table>
<thead>
<tr>
<th>Name</th>
<th>Regime-M</th>
<th>Regime-F</th>
<th>Name</th>
<th>Regime-M</th>
<th>Regime-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$100\gamma$</td>
<td>0.34</td>
<td>0.27</td>
<td>$\rho_{ca}$</td>
<td>0.23</td>
<td>0.30</td>
</tr>
<tr>
<td>$\xi$</td>
<td>1.78</td>
<td>2.25</td>
<td>$\rho_{cb}$</td>
<td>0.47</td>
<td>0.22</td>
</tr>
<tr>
<td>$h$</td>
<td>0.96</td>
<td>0.96</td>
<td>$\rho_{ci}$</td>
<td>0.76</td>
<td>0.47</td>
</tr>
<tr>
<td>$\alpha_G$</td>
<td>−0.24</td>
<td>−0.38</td>
<td>$\rho_{ep}$</td>
<td>0.48</td>
<td>0.61</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.35</td>
<td>0.31</td>
<td>$\rho_{ew}$</td>
<td>0.36</td>
<td>0.20</td>
</tr>
<tr>
<td>$s$</td>
<td>7.08</td>
<td>3.47</td>
<td>$\rho_{em}$</td>
<td>0.52</td>
<td>0.87</td>
</tr>
<tr>
<td>$\omega_p$</td>
<td>0.94</td>
<td>0.95</td>
<td>$\rho_{eg}$</td>
<td>0.16</td>
<td>0.29</td>
</tr>
<tr>
<td>$\omega_w$</td>
<td>0.81</td>
<td>0.74</td>
<td>$\rho_{ez}$</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td>$\chi_p$</td>
<td>0.21</td>
<td>0.11</td>
<td>$100\sigma_a$</td>
<td>0.80</td>
<td>1.18</td>
</tr>
<tr>
<td>$\chi_w$</td>
<td>0.19</td>
<td>0.06</td>
<td>$100\sigma_b$</td>
<td>14.54</td>
<td>16.95</td>
</tr>
<tr>
<td>$\phi_x$</td>
<td>1.34</td>
<td>0.19</td>
<td>$100\sigma_i$</td>
<td>0.44</td>
<td>1.30</td>
</tr>
<tr>
<td>$\phi_y$</td>
<td>0.16</td>
<td>0.21</td>
<td>$100\sigma_p$</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>$\rho_{r}$</td>
<td>0.79</td>
<td>0.37</td>
<td>$100\sigma_w$</td>
<td>0.27</td>
<td>0.23</td>
</tr>
<tr>
<td>$\gamma_g$</td>
<td>0.25</td>
<td>0.0</td>
<td>$100\sigma_m$</td>
<td>0.15</td>
<td>0.21</td>
</tr>
<tr>
<td>$\gamma_z$</td>
<td>−0.06</td>
<td>0.0</td>
<td>$100\sigma_g$</td>
<td>1.63</td>
<td>2.05</td>
</tr>
<tr>
<td>$\rho_{g}$</td>
<td>0.97</td>
<td>0.95</td>
<td>$100\sigma_z$</td>
<td>3.57</td>
<td>0.78</td>
</tr>
<tr>
<td>$\rho_{z}$</td>
<td>0.98</td>
<td>0.97</td>
<td>$\bar{L}$</td>
<td>470.67</td>
<td>468.98</td>
</tr>
<tr>
<td>$\bar{\pi}$</td>
<td>0.70</td>
<td>0.75</td>
<td>$\bar{T}$</td>
<td>0.70</td>
<td>0.75</td>
</tr>
</tbody>
</table>

**Notes:** The DGPs under regimes M and F correspond to their post-Volcker (1955:Q1–1979:Q4) and pre-Volcker (1982:Q1–2007:Q4) estimates reported in Leeper, Traum and Walker (2017), respectively.
Figure A.1: Autocorrelation function of each structural parameter. Notes: Red horizontal lines correspond to an autocorrelation of 0.1.
Figure A.2: Autocorrelation function of each volatility parameter. Notes: See Figure A.1.
Figure A.3: Marginal prior and posterior distributions of each structural parameter. Notes: Each panel compares the prior (red dashed line) with the posterior (shaded area). Vertical lines denote the true parameter values. The kernel smoothed posterior densities are estimated using 10,000 TaRB-MH draws.
Figure A.4: Stochastic volatility of each shock innovation. Notes: Each panel compares the model's estimated log-variances with their true values (red solid line). Blue dashed lines denote median estimates, while shaded areas delineate 90% highest posterior density bands.