STAFF REPORTS

NO. 1084
JANUARY 2024

On the Validity of Classical and Bayesian DSGE-Based Inference

Katerina Petrova

FEDERAL RESERVE BANK of NEW YORK

On the Validity of Classical and Bayesian DSGE-Based Inference
Katerina Petrova
Federal Reserve Bank of New York Staff Reports, no. 1084
January 2024
https://doi.org/10.59576/sr.1084

Abstract

This paper studies large sample classical and Bayesian inference in a prototypical linear DSGE model and demonstrates that inference on the *structural* parameters based on a Gaussian likelihood is unaffected by departures from Gaussianity of the *structural* shocks. This surprising result is due to a cancellation in the asymptotic variance resulting into a generalized information equality for the block corresponding to the structural parameters. The underlying reason for the cancellation is the certainty equivalence property of the linear rational expectation model.

The main implication of this result is that classical and Bayesian Gaussian inference achieve a semi-parametric efficiency bound and there is no need for a "sandwich-form" correction of the asymptotic variance of the structural parameters. Consequently, MLE-based confidence intervals and Bayesian credible sets of the deep parameters based on a Gaussian likelihood have correct asymptotic coverage even when the structural shocks are non-Gaussian. On the other hand, inference on the reduced-form parameters characterizing the volatility of the shocks is invalid whenever the structural shocks have a non-Gaussian density and the paper proposes a simple Metropolis-within-Gibbs algorithm that achieves correct large sample inference for the volatility parameters.

JEL classification: C11, C12, C22
Keywords: DSGE models, generalized information equality, sandwich form covariance

---

Petrova: Federal Reserve Bank of New York, Universitat Pompeu Fabra, and Barcelona School of Economics (email: katerina.petrova@ny.frb.org). The author thanks Haroon Mumtaz, Christian Matthes, Geert Mesters, Tassos Magdalinos, Barbara Rossi, Jordi Gali, Domenico Giannone and Argia Sbordone for valuable comments and discussions. She acknowledges support from the General Directorate for Research in the Government of Catalonia through the Beatriu de Pinós grant 2019/BP/00239 and the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S) for the period prior to September 2023.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are routinely used for macroeconomic analysis by both academics and policy-makers. Their success is due to their resulting model-based analysis being consistent with economic theory, since their microfoundations are derived from optimisation of rational agents, in contrast to reduced-form models which typically lack a theory-consistent story for their output.

Two essential ingredients of DSGE models are: (i) the structural 'deep' parameters that define the agents' preferences and economic environment and (ii) the structural shocks that characterise the stochastic component of the model. Given macroeconomic data, the aim of econometric procedures is to infer about the former given assumptions on the latter.

Applied work on DSGE models originally employed data-informed calibration of the parameters (e.g. Kydland and Prescott (1996)), and later full-information estimation procedures: classical MLE (e.g. Altug (1989), Ireland (2004)) or Bayesian methods (e.g. Schorfheide (2000), Fernández-Villaverde and Rubio-Ramírez (2004), Smets and Wouters (2007)). Bayesian methods have become the preferred estimation procedure for DSGE models in the literature since they provide greater control over the parameter space through the use of prior information, which is often available because of the microfoundation of the structural parameters. As such, Bayesian methods amount to a flexible combination between data-based econometric estimation and earlier calibration methods with the tightness of the imposed priors controlling the relative importance of the two.

The aim of this paper is to investigate the asymptotic validity of classical and Bayesian inference on the structural parameters in a DSGE model whenever the distributional assumption on the model's structural shocks is misspecified. Since both classical and Bayesian methods are full-information, such distributional assumptions are required, and the standard assumption made in the literature is Gaussianity, which is convenient since it permits the use of the Kalman filter to evaluate the likelihood function. Even a small degree of distributional misspecification can, in general, invalidate MLE and Bayesian inference whenever the generalised information equality for the quasi-likelihood function is violated, resulting in a 'sandwich-form' large sample variance for the model's parameters; e.g. see White (1982), Gourieroux et al. (1984), Bollerslev and Wooldridge (1992) for MLE and Chernozhukov and Hong (2003) and Müller (2013) for Bayesian estimation under distributional misspecification.

This paper studies large sample inference in the prototypical linear DSGE model and demonstrates that classical and Bayesian inference on the structural parameters based on Gaussian likelihood is unaffected by departures from Gaussianity of the structural shocks. This surprising result is at odds with previous results on Bayesian inference with linear DSGE models (e.g. Müller (2013)) and it is due to a cancellation in the asymptotic variance resulting into information equality for the block corresponding to the structural parameters. The underlying reason for the cancellation is the certainty equivalence property of the linear rational expectation model, which implies that the solution matrices, and hence the model-implied conditional mean of the data, do not depend on the second moment of the structural shocks. The implication of this result is that incorrectly imposing Gaussianity assumption on the structural shocks has no large-sample effect on the validity of classical and Bayesian inference (confidence intervals and credible sets) on the structural parameters of the model, and hence there is no need for any 'sandwich-form' correction for the variance, previously recommended in the literature.

On the other hand, inference on the reduced-form parameters characterising the volatility of the shocks is invalidated whenever the true structural shocks come from a distribution with skewness and kurtosis different from that of the normal density. To this end, the paper proposes a simple Metropolis-within-Gibbs algorithm that achieves valid large sample inference of the volatility parameters and practical implementation of the procedure only requires consistent estimator of the kurtosis of the structural shocks.

The rest of the paper is organised as follows. Section 2 presents the model, assumptions and the main result of the paper. Section 3 proposes a Metropolis-within-Gibbs estimation procedure that achieves valid inference on the reduced-form volatility parameters. Section 4 presents a Monte Carlo exercise demonstrating the validity of the theoretical results of Section 2 as well as the proposed procedure of Section 3, Section 5 applies the proposed algorithm of Section 3 to a DSGE model with financial frictions and Section 6 concludes. The supplementary Appendix contains some auxiliary mathematical results, the proof of Theorem 1 of the paper, as well as some additional results.

2 Econometric Framework

We consider a linearised rational expectation model of the form

$$\Gamma_0(\theta_1)\mathbb{E}_{\mathcal{F}_t}\mathbf{x}_{t+1} = \Gamma_1(\theta_1)\mathbf{x}_t + \Gamma_2(\theta_1)\mathbf{x}_{t-1} + \Gamma_3(\theta_1)\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim (0, \Sigma(\theta_2)), \tag{1}$$

where Γ_0, Γ_1, Γ_2 and Γ_3 are matrix-valued functions of the $k_1 \times 1$ structural parameter vector θ_1 of the model, $\mathbf{x}_t$ is an $s \times 1$ vector of the model's variables, $\boldsymbol{\varepsilon}_t$ is an $m \times 1$ vector of structural shocks with covariance matrix[1] $\Sigma(\theta_2)$, a function of the $k_2 \times 1$ reduced-form parameter vector θ_2, $\mathcal{F}_t$ denotes the natural filtration of the structural shock sequence $\mathcal{F}_t = \sigma(\boldsymbol{\varepsilon}_t, ..., \boldsymbol{\varepsilon}_1)$ and $\mathbb{E}_{\mathcal{F}_{t-1}}$ denotes the conditional expectation operator. The structural parameters θ_1 and reduced-form parameters θ_2 are collected in a $k-$dimensional vector $\theta = [\theta_1', \theta_2']'$. The setup in (1) constitutes the prototypical DSGE model estimated in the literature and used by policy makers and central banks. Under regularity conditions, a solution of the dynamic rational expectation model in (1) exists and is unique, the solution can be obtained numerically (see, for instance, Blanchard and Kahn (1980) or Sims (2002)) and takes the form:

$$\mathbf{x}_t = F(\theta_1)\mathbf{x}_{t-1} + G(\theta_1)\boldsymbol{\varepsilon}_t, \tag{2}$$

[1] In applied work $\Sigma(\theta_2)$ is typically assumed to be diagonal in order to impose orthogonality across shocks and θ_2 contains the volatilities of the shocks. We leave the structure of $\Sigma(\theta_2)$ unrestricted; in Section 3, we provide a discussion on the difference in terms of inference between imposing orthogonality and independence when the shocks are non-Gaussian.

where the solution matrices $F(\theta_1)$ and $G(\theta_1)$ are functions of θ_1, for most models only available numerically. Crucially, linearisation of the underlying nonlinear rational expectation model around the deterministic steady state (i.e. setting $\Sigma(\theta_2) = 0$) implies certainty equivalence: Γ_0, Γ_1, Γ_2 and Γ_3 in (1) and, hence, the solution matrices in (2) do not depend on the second moments of the shocks θ_2. This is the crucial ingredient behind the result of Theorem 1 below: linearity of the solution equation (2) alone would not deliver the generalised information equality if, for example, the solution matrix $F(\cdot)$ depended on θ_2 (see Remark 4 after Theorem 1 for further discussion). To take the linear DSGE model to the data, the solution in (2) is typically augmented by a measurement equation of the form

$$\mathbf{y}_t = C(\theta_1) + H(\theta_1)\mathbf{x}_t \tag{3}$$

where $\mathbf{y}_t$ is an $r \times 1$ vector of observables with $r \leq m$ ($r > m$ results in singularity of the variance for $\mathbf{y}_t$), $C(\theta_1)$ is a vector typically containing model-specific steady-state values and $H(\theta_1)$ selects and, if necessary, transforms the observables of the vector $\mathbf{x}_t$. Additive martingale difference measurement error in (3) can be included without changing the main result of the paper; for brevity we omit such an extension here.

In this paper, we derive the asymptotic variance of classical ML and Bayesian estimators for θ in the DSGE model in (2) and (3) whenever the shocks $\boldsymbol{\varepsilon}_t$ are incorrectly modelled as Gaussian: $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma(\theta_2))$. Before we proceed to laying down the formal assumptions and analysing the DSGE-specific large-sample inference, we provide a brief discussion of the problem that distributional misspecification can cause in the general case in order to give an insight of why the problem is absent for the deep parameters θ_1 of the linear DSGE model considered.

It is well-known[2] that under distributional misspecification and mild regularity conditions, the quasi-ML estimator for θ is consistent and has a 'sandwich-form' asymptotic covariance matrix of the form $\mathcal{C}_0 = \mathcal{A}_0^{-1}\mathcal{B}_0\mathcal{A}_0^{-1}$, with

$$\begin{aligned}\mathcal{A}_0 &= -\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\mathcal{F}_{t-1}}\left[\frac{\partial\ell^2(\mathbf{y}_t;\theta)}{\partial\theta\partial\theta'}\right]_{\theta=\theta^0} \\ \mathcal{B}_0 &= \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\mathcal{F}_{t-1}}\left[\frac{\partial\ell(\mathbf{y}_t;\theta)}{\partial\theta}\frac{\partial\ell(\mathbf{y}_t;\theta)}{\partial\theta'}\right]_{\theta=\theta^0}\end{aligned} \tag{4}$$

where $\ell(.)$ denotes the quasi log-likelihood function of θ and θ^0 denotes the (pseudo) true value of θ. The reason for the different asymptotic variance (relative to the case of correct distributional specification where the asymptotic variance is given by the inverse information matrix $\mathcal{A}_0^{-1}$) is that, in general, the information equality breaks down since expectations are taken with respect to the true (rather than the quasi) distribution. When $\mathcal{A}_0 \neq \mathcal{B}_0$, there is no asymptotic cancellation in the expression for the variance and $\mathcal{C}_0 \neq \mathcal{A}_0^{-1}$. In this case, Bayesian inference is also invalidated even for large samples: Chernozhukov and Hong (2003) show that given a strictly positive and continuous prior density $\pi(\theta)$, the posterior $p(.)$ of θ satisfies

$$p\left(\sqrt{n}\left(\theta - \theta^0\right)\right) \to_d \mathcal{N}\left(0, \mathcal{A}_0^{-1}\right) \text{ as } n \to \infty \tag{5}$$

which has the incorrect asymptotic variance $\mathcal{A}_0^{-1}$ instead of $\mathcal{C}_0$; as a result, posterior credible sets

[2] See, for example, White (1982), Gourieroux et al. (1984), Bollerslev and Wooldridge (1992).

do not contain the (pseudo) true parameter with correct coverage even when the sample size is large. A Bayesian decision-theoretic justification of this result is provided by Müller (2013), who shows that Bayesian inference is of lower asymptotic risk whenever the posterior based on the misspecified likelihood is substituted by an artificial posterior with 'sandwich-form' covariance matrix.

Due to distributional misspecification, classical and Bayesian estimators no longer achieve the parametric Cramér-Rao lower bound and hence the resulting estimators are no longer parametrically efficient. However, there are instances when a generalised version of the information equality $\mathcal{A}_0^{-1} = \mathcal{C}_0$ continues to hold in the presence of distributional misspecification for some or all elements of θ, in which case a semi-parametric lower bound can be achieved and, crucially, no erroneous inference decisions occur if distributional misspecification is ignored by the practitioner. A well-known example of this is the linear regression model under correctly specified first two conditional moments, where Gaussian inference on the conditional mean parameters is semi-parametrically efficient and robust to distributional misspecification. The reason for this result is the separability between the conditional mean and conditional variance parameters, invaliding Gaussian inference only for the variance parameters.

While the DSGE model in (2) and (3) is linear, such separability between the conditional mean and variance parameters is not present. The observables satisfy

$$\mathbf{y}_t|\mathcal{F}_{t-1} \sim \left(\mu_t\left(\theta_1, \mathbf{x}_{t-1}\right), \Omega\left(\theta_1, \theta_2\right)\right)$$

where the conditional moments are given by

$$\begin{aligned} \mu_t\left(\theta_1, \mathbf{x}_{t-1}\right) &= C\left(\theta_1\right) + H\left(\theta_1\right) F\left(\theta_1\right) \mathbf{x}_{t-1} \\ \Omega\left(\theta_1, \theta_2\right) &= H\left(\theta_1\right) G\left(\theta_1\right) \Sigma\left(\theta_2\right) G\left(\theta_1\right)' H\left(\theta_1\right)' \end{aligned} \tag{6}$$

and the conditional variance $\Omega\left(\theta_1, \theta_2\right)$ of $\mathbf{y}_t$ always depends on the structural parameters θ_1 through the matrices $G\left(\theta_1\right)$ and $H\left(\theta_1\right)$. It is perhaps due to this lack of separability that there has been a consensus in the literature that distributional misspecification of $\boldsymbol{\varepsilon}_t$ (and hence $\mathbf{y}_t$) would invalidate classical and Bayesian asymptotic inference on *all* DSGE parameters θ and that a 'sandwich-form' covariance is needed to robustify the posteriors in the non-Gaussian case. For example, Müller (2013) applies his 'sandwich' correction to a linear new-Keynesian DSGE model. Canova and Matthes (2021a) and (2021b) use Müller (2013)'s 'sandwich' posterior correction in order to apply composite likelihood estimation for the parameters of linearised DSGE models. Qu and Tkachenko (2012) propose a frequency domain quasi-maximum likelihood estimator for the parameters of a linearised DSGE model and their asymptotic variance is of a 'sandwich-form'. Guerron-Quintana, Inoue and Kilian (2017) also correct the variance of the (quasi) posteriors of the linear DSGE model for a 'sandwich-form'.

In this paper, we establish a surprising result with important implications for DSGE-based inference: as far as the structural parameters θ_1 are concerned, there is no need for such 'sandwich-form' corrections. In particular, by partitioning the parameter vector in structural and reduced-form parameters $[\theta_1', \theta_2']'$ and looking deeper in the partitioned (quasi-) score vector and Hessian matrix and their moments, we demonstrate that while the variance of the quasi-score vector when Gaussianity is imposed depends on higher (multivariate third and fourth) moments of the shocks, classical and Bayesian objective functions based on Gaussian likelihood for the structural parameters θ_1 continue to satisfy the generalised information equality for large samples, so that the asymptotic variance of the resulting estimators for θ_1 is valid whether or not the true underlying shocks were Gaussian. The advantage of having such an information equality in place for θ_1 is that classical and Bayesian inference can be applied on θ_1 as if Gaussianity holds without the need for any corrections involving consistent estimator for the sandwich-form variance $\mathcal{C}_0$. This is particularly useful in the DSGE setup since it permits the use of the Kalman filter (suited for linear Gaussian state space models) even for models with non-Gaussian shocks, thus considerably simplifying the evaluation of the likelihood function. While the use of sandwich-form covariance is asymptotically valid in theory, in practice, obtaining good quality first and second derivatives of the log-likelihood in order to compute the 'sandwich-form' is difficult and computationally expensive[3] since the model's solution matrices in (2) are not available analytically for most models. Consequently, researchers have resorted to the use of numerical derivatives, which are often of very poor quality particularly when θ is of larger dimension and can add unnecessary noise to already poorly identified models routinely estimated with very small samples of macroeconomic data. The main result of this paper makes 'sandwich-form' MLE and Bayesian posterior corrections obsolete, thus significantly streamlining DSGE inference.

The asymptotic variance for the estimator of the reduced-form volatility parameters θ_2 is affected by the misspecification and depends both on the skewness and kurtosis of the shocks. However, conditional on θ_1, inference on θ_2 is straightforward and only requires a consistent estimator of the kurtosis of the structural shocks. To this end, we design a Metropolis-within-Gibbs algorithm in Section 3 that achieves valid large sample joint posterior inference on θ_1 and θ_2.

We now proceed to the formal analysis of the asymptotic variance of the parameters of the DSGE model in (2) and (3). We make the following assumptions.

Assumptions:

1. Specification. The DSGE model's equations in (1) are correctly specified and the data are generated by (2) and (3) with a true vector $\theta^0 \in \text{int}\Theta$ for a parameter space $\Theta \subseteq \mathbb{R}^k$ with $G(\theta_1)$ full column rank at θ_1^0.

2. Determinacy. The solution (2) of the DSGE model in (1) is uniquely determined: for any $\theta_1 \in \Theta_1 \subseteq \mathbb{R}^{k_1}$ the solution matrices $F(\theta_1)$ and $G(\theta_1)$ are unique.

3. Identification. The DSGE model is globally identified: for any $\tilde{\theta} = [\tilde{\theta}_1', \tilde{\theta}_2']' \in \Theta$, the conditional first two moments in (6) satisfy

$$\mu\left(\theta_1^0, \mathbf{x}_{t-1}\right) = \mu\left(\tilde{\theta}_1, \mathbf{x}_{t-1}\right) \text{ and } \Omega\left(\theta_1^0, \theta_2^0\right) = \Omega\left(\tilde{\theta}_1, \tilde{\theta}_2\right)$$

if and only if $\theta^0 = \tilde{\theta}$.

[3] It may be possible to use the Kalman filter equations in first derivatives to obtain the first derivative of the log-likelihood (and its variance) analytically; however, such an extension increases the dimension of the state vector from s to sk, which would make inference even with small DSGE models prohibitively expensive computationally.

4. Stationarity. The law of motion for $\mathbf{x}_t$ in (2) satisfies $\rho\left(F\left(\theta_1^0\right)\right) < 1$, where $\rho(\cdot)$ denotes the spectral radius of a matrix and the initial condition in (2) satisfies $x_0 = O_p(1)$.

5. Moments. The process for the structural shocks $(\boldsymbol{\varepsilon}_t, \mathcal{F}_t)_{t\geq 1}$ satisfies:

a. $(\boldsymbol{\varepsilon}_t, \mathcal{F}_t)_{t\geq 1}$ is a martingale difference satisfying $\mathbb{E}_{\mathcal{F}_{t-1}}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'] = \Sigma\left(\theta_2^0\right) > 0$ for all t, with $\theta_2 = vech\Sigma(\theta_2)$.

b. $(\boldsymbol{\varepsilon}_t, \mathcal{F}_t)_{t\geq 1}$ has time invariant third and fourth conditional moments:

$$\mathbb{E}_{\mathcal{F}_{t-1}}\left\{\boldsymbol{\varepsilon}_t\left(vech\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right]\right)'\right\} = \mathcal{S} \text{ and } \mathbb{E}_{\mathcal{F}_{t-1}}\left[vech\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\left(vech\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right)'\right] = \mathcal{K} \tag{7}$$

for all t and $\mathcal{K} > 0$.

c. The sequence $\left(\|\boldsymbol{\varepsilon}_t\|^4\right)_{t\geq 1}$ is uniformly integrable:

$$\sup_{t\geq 1}\mathbb{E}\left[\|\boldsymbol{\varepsilon}_t\|^4 1\left\{\|\boldsymbol{\varepsilon}_t\|^4 > \lambda\right\}\right] \to 0 \text{ as } \lambda\to\infty.$$

6. Smoothness. The functions $C(\theta_1)$, $H(\theta_1)$, $F(\theta_1)$, and $G(\theta_1)$ are continuously differentiable over Θ and twice continuously differentiable with Lipschitz continuous second derivatives in a neighbourhood $N\left(\theta^0, \delta\right) = \left\{\theta\in\Theta : \left\|\theta - \theta^0\right\| < \delta\right\}$ for some $\delta > 0$.

7. Rank. The matrices HG and $\dot{HF} := \frac{\partial vec[H(\theta_1)F(\theta_1)]'}{\partial\theta_1}$ satisfy the following rank conditions: $rk(HG) = r$ and $rk\left(\dot{HF}\right) = \dim\theta_1 \leq rm$ at θ^0.

Remarks.

1. Since the focus of this paper is on inference for a well-behaved model, Assumption 1 assumes away model misspecification, Assumption 2 abstracts from indeterminacy issues due to existence of multiple solutions (sunspot equilibria) and Assumption 3 assumes away identification issues and implies that the first two moments globally identify θ (see Iskrev (2010)). Similarly, nonstationarity that gives rise to nonstandard classical inference is assumed away in Assumption 4; $\rho\left(F\left(\theta_1^0\right)\right) < 1$ implies stability (asymptotic covariance stationarity) of $\mathbf{x}_t$ and hence the observables $\mathbf{y}_t$. Weak stationarity implies that the unconditional covariance matrix of the state vector $\mathbf{x}_t$ is given by V_X with $\|V_X\| < \infty$, where V_X satisfies

$$vec(V_X) = \left(I_{s^2} - \left(F\left(\theta^0\right)\otimes F\left(\theta^0\right)\right)\right)^{-1}\left(G\left(\theta^0\right)\otimes G\left(\theta^0\right)\right)vec\Sigma\left(\theta^0\right). \tag{8}$$

2. In the typical DSGE model, $\Sigma(\theta_2)$ is diagonal, since the structural shocks $\boldsymbol{\varepsilon}_t$ are assumed to be mutually uncorrelated, but diagonality is not required for the main result of the paper: we allow the covariance matrix $\Sigma(\theta_2)$ to be full and, for simplicity, model all its reduced-form elements as $\theta_2 = vech\Sigma(\theta_2)$. In the diagonal case, $\theta_2 = Pvec\Sigma$, where P is an $n\times n^2$ semi-orthogonal selector matrix with $[P]_{i,(i-1)n+i} = 1$ for $i = 1, ..., n$ and zeros elsewhere and the asymptotic distribution of $\hat{\theta}_2$ can be obtained from that of $vec\widehat{\Sigma}$ by selecting the relevant elements $\hat{\theta}_2 = Pvec\widehat{\Sigma}$ and their variance $V(\hat{\theta}_2) = PV(vec\widehat{\Sigma})P'$. Similarly, if $\theta_2 = f(vec\Sigma)$ for a smooth function f, a delta method can be used to obtained the asymptotic distribution of $\hat{\theta}_2$ from that of $vec\widehat{\Sigma}$.

3. Assumption 5a imposes conditional homoskedasticity of the structural shocks. The classic GLS result applies here: if heteroskedasticity in Σ is ignored, the generalised information equality of Theorem 1 below breaks down, even if the shocks were Gaussian, and the asymptotic variance of θ_1 is of 'sandwich-form' instead. However, the main result of the paper can be shown to hold under heteroskedasticity in Σ, as long as the heteroskedasticity is correctly modelled explicitly in the quasi-likelihood (for example, as in Justiniano and Primiceri (2008) or Petrova (2019)); that is, conditional on $\Sigma_{1:n}$, the generalised information equality for θ_1 continues to hold.

4. Assumption 5b imposes constant conditional third and fourth moments[4] $\mathcal{S}$ and $\mathcal{K}$ and requires that $\mathcal{K}$ is positive definite. This assumption can be relaxed, allowing the conditional moments $\mathcal{S}_t$ and $\mathcal{K}_t$ to change over time and requiring that $\mathcal{S}^\infty := \text{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^n \mathcal{S}_t$ and $\mathcal{K}^\infty := \text{plim}_{n\to\infty} \frac{1}{n}\sum_{t=1}^n \mathcal{K}_t$ exist; the main result of the paper continues to hold with $\mathcal{S}$ and $\mathcal{K}$ in the limiting quantities replaced by $\mathcal{S}^\infty$ and $\mathcal{K}^\infty$.

5. Assumption 5c requires that the sequence $\left(\|\boldsymbol{\varepsilon}_t\|^4\right)_{t\geq 1}$ is uniformly integrable, this implies existence of four finite moments $\|\mathcal{K}\| < \infty$ but it is weaker than $4+\delta$ finite moments for any $\delta > 0$. Whenever the shocks $\boldsymbol{\varepsilon}_t$ are identically distributed, Assumption 5c is equivalent to $\|\mathcal{K}\| < \infty$.

6. Assumption 6 imposes smoothness on the model's coefficients with respect to θ; while the resulting asymptotic variance only depends on first derivatives, Lipschitz continuity of second derivatives is required in a neighbourhood of θ^0 in order to deal with the intermediate point arising from the linearisation of the score vector and ensure a LLN for the Hessian matrix.

7. Finally Assumption 7 imposes several rank conditions. Requiring that the number of observables is smaller or equal to the number of shocks ($r \leq m$) is not sufficient for Ω to be positive definite, since $rk\Omega \leq \min\left(rk\left(H\right), rk\left(G\right)\right) = r$ when G is full-rank. Requiring instead that the product HG has rank r is necessary and sufficient for Ω to be positive definite, since $\Sigma > 0$ and Ω is a quadratic form $\Omega = HG\Sigma G'H'$. Moreover, $rk(\dot{HF}) = \dim\theta_1 \leq rm$ is a sufficient[5] condition for the asymptotic variance of the (Q)MLE for θ_1 ($[\mathcal{C}_0]_{11}$ in Theorem 1 below) to be nonsingular. This condition is required even under correct distributional specification, to ensure that the asymptotic variance of the MLE for the structural parameters θ_1 is nonsingular.

Under Assumptions 1-7, it follows that, conditional on the information set $\mathcal{F}_{t-1}$,

$$\mathbf{y}_t|\mathcal{F}_{t-1} \sim \left(\mu_t\left(\theta_1, \mathbf{x}_{t-1}\right), \Omega\left(\theta_1, \theta_2\right)\right)$$

with $\mu_t\left(\theta_1, \mathbf{x}_{t-1}\right)$ and $\Omega\left(\theta_1, \theta_2\right)$ defined in (6). When the structural shocks are (incorrectly) assumed to be Gaussian: $\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim \mathcal{N}\left(0, \Sigma\left(\theta_2\right)\right)$, the conditional (quasi-) log-likelihood (except constants) is given by

$$\ell\left(\mathbf{y}_t, \theta|\mathcal{F}_{t-1}\right) = -\frac{1}{2}\log|\Omega| - \frac{1}{2}tr\left[\Omega^{-1}\mathbf{u}_t\mathbf{u}_t'\right] \tag{9}$$

where, for brevity, we suppress dependence of Ω and μ_t on θ and $\mathbf{x}_{t-1}$ and we define the residual function $\mathbf{u}_t = \mathbf{u}_t\left(\theta_1, \mathbf{x}_{t-1}\right) = \mathbf{y}_t - \mu_t$. We denote by $\hat{\theta}$ the QMLE which maximises the Gaussian quasi log-likelihood $\frac{1}{n}\sum_{t=1}^n \ell\left(\mathbf{y}_t, \theta|\mathcal{F}_{t-1}\right)$. The resulting conditional quasi-score vector is given by

$$s_t\left(\theta\right) = \frac{\partial \ell\left(\mathbf{y}_t, \theta|\mathcal{F}_{t-1}\right)}{\partial\theta} = \begin{bmatrix} \frac{1}{2}\dot{\Omega}_1 D_r'\left(\Omega^{-1}\otimes\Omega^{-1}\right)D_r\mathbf{z}_t + \dot{\mu}_t\Omega^{-1}\mathbf{u}_t \\ \frac{1}{2}\dot{\Omega}_2 D_r'\left(\Omega^{-1}\otimes\Omega^{-1}\right)D_r\mathbf{z}_t \end{bmatrix}, \tag{10}$$

[4] Note that $\mathcal{S}$ and $\mathcal{K}$ contain all third and fourth conditional cross moments $\mathbb{E}_{\mathcal{F}_{t-1}}\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}$ and $\mathbb{E}_{\mathcal{F}_{t-1}}\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}$ for $i, j, k, l = 1, ..., m$ respectively. Another way to characterise $\mathcal{S}$ and $\mathcal{K}$ is to compute the covariance matrix of the vector $\left[\boldsymbol{\varepsilon}_t', vech\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)'\right]'$ which is given by $\begin{bmatrix} \Omega & \mathcal{S} \\ \mathcal{S}' & \mathcal{K} \end{bmatrix}$.

[5] Whenever $rk\dot{C} < k_1$ for $\dot{C} = \frac{\partial C(\theta_1)'}{\partial\theta_1}$, which is usually the case since $\dot{C}$ is $k_1 \times r$ and the number of structural parameters k_1 typically exceeds the number of observables r, $rk(\dot{HF}) = \dim\theta_1 \leq rm$ is not only sufficient but also necessary for nonsingularity of the asymptotic variance of θ_1.

where $\dot{\Omega}_1 = \frac{(\partial vech\Omega)'}{\partial\theta_1}$, $\dot{\Omega}_2 = \frac{(\partial vech\Omega)'}{\partial\theta_2}$, $\dot{\mu}_t = \frac{\partial\mu_t'}{\partial\theta_1}$, $\mathbf{z}_t = \mathbf{z}_t(\theta_1, \theta_2, \mathbf{x}_{t-1}) = vech(\mathbf{u}_t\mathbf{u}_t' - \Omega)$ and D_r is the $r^2 \times r(r+1)/2$ duplication matrix. Both $\mathbf{u}_t$ and $\mathbf{z}_t$ are explicit functions of θ and whenever Assumption 1-2 hold (i.e. first two conditional moments are correctly specified), it follows immediately that $\mathbf{z}_t$ and $\mathbf{u}_t$ are martingale difference sequences at the true value θ^0, i.e. $\mathbb{E}_{\mathcal{F}_{t-1}}\left(\mathbf{z}_t\left(\theta^0\right)\right) = 0$ and $\mathbb{E}_{F_{t-1}}\left(\mathbf{u}_t\left(\theta^0\right)\right) = 0$, which together with Assumption 3 is sufficient for consistency[6] of the QMLE $\hat{\theta}$. Standard martingale CLT, established in Lemma 1 of the Appendix, implies that at θ^0

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n s_t\left(\theta^0\right) \to_d \mathcal{N}\left(0, \mathcal{B}_0\right), \quad \mathcal{B}_0 = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{t-1}}\left[s_t\left(\theta^0\right)s_t\left(\theta^0\right)'\right]$$

where the explicit expression for the asymptotic covariance $\mathcal{B}_0$ for the linear DSGE model is given in the Appendix in (A.1). In particular, $\mathcal{B}_0$ is of nonstandard form (e.g. unlike the fully separable case when $\dot{\Omega}_1 = 0$, the higher third and fourth moments $\mathcal{S}$ and $\mathcal{K}$ enter in all elements of $\mathcal{B}_0$). Despite dependence of $\mathcal{B}_0$ on higher moments, computation of the sandwich form covariance $\mathcal{C}_0 = \mathcal{A}_0^{-1}\mathcal{B}_0\mathcal{A}_0^{-1}$ and partitioning it conformably in blocks corresponding to the structural parameters θ_1 and reduced-form parameters θ_2 reveals a cancellation in the block corresponding to the deep parameters θ_1 of the DSGE model. The result in summarised in Theorem 1 below.

Theorem 1 *In the linear DSGE model under Assumptions 1-7,* $\sqrt{n}\left(\hat{\theta}_1 - \theta_1^0\right) \to_d \mathcal{N}\left(0, \left[\mathcal{A}_0^{-1}\right]_{11}\right)$ *where* $\left[\mathcal{A}_0^{-1}\right]_{11}$ *denotes the upper* $k_1 \times k_1$ *principal submatrix of* A_0^{-1} *defined in (4).*

Remarks

1. The key result of Theorem 1 is that the generalised information equality continues to hold for the upper block corresponding to the asymptotic variance of the structural parameters of the DSGE model θ_1 due to a cancellation, that is $[\mathcal{C}_0]_{11} = \left[\mathcal{A}_0^{-1}\right]_{11}$. Relative to the fully separable case when $\dot{\Omega}_1 = 0$ and $\mathcal{A}_0^{-1}$ is block-diagonal, in the DSGE setup[7], $\dot{\Omega}_1 \neq 0$ and the cancellation happens through the off-diagonal blocks in $\mathcal{A}_0^{-1}$ and $\mathcal{B}_0$. Arguably, inference on the deep parameters θ_1 is of utmost importance for applied researchers, frequentist and Bayesian, while inference on the reduced-form volatility parameters θ_2 is of secondary interest. The key implication of Theorem 1 is that ignoring the distributional misspecification and imposing Gaussian assumptions does not affect large sample MLE or Bayesian inference[8] (frequentist confidence intervals and Bayesian credible sets) on θ_1 and robust 'sandwich-form' corrections are unnecessary.

2. Semi-parametric efficiency for θ_1 follows since: (i) the conditional mean μ_t is correctly specified, (ii) the quasi log-likelihood is Gaussian which belongs to the linear exponential family of distributions, and (iii) the generalised information equality holds for the variance corresponding to θ_1; see, for example, Proposition 4 in Gourieroux and Monfort (1993).

3. The exact expression for $[\mathcal{C}_0]_{11}$ is given by $[\mathcal{C}_0]_{11} = \dot{C}\Omega^{-1}\dot{C}' + \dot{HF}\left(V_X \otimes \Omega^{-1}\right)(\dot{HF})'$, where $\dot{C} = \frac{\partial C(\theta_1)'}{\partial\theta_1}$, $\dot{HF} = \frac{\partial vec[H(\theta_1)F(\theta_1)]'}{\partial\theta_1} = \dot{H}\left(F(\theta_1) \otimes I_r\right) + \dot{F}\left(I_s \otimes H(\theta_1)'\right)$, $\dot{H} = \frac{\partial(vecH(\theta_1))'}{\partial\theta_1}$,

[6] This follows by the WLLN $\frac{1}{n}\sum_{t=1}^n s_t(\theta_0) \to_{L_1} 0$ since $\mathbb{E}_{\mathcal{F}_{t-1}}s_t(\theta_0) = 0$ and $\|s_t\|$ is uniformly integrable sequence (e.g. see Hall and Heyde Theorem 2.19).

[7] Note that the result of Theorem 1 is general and applies to any model where $\frac{\partial\mu_t}{\partial\theta_2} = 0$, the specific DSGE model structure is not required for the main result.

[8] For Bayesian estimation, this follows directly from (5) by the result of Chernozhukov and Hong (2003).

$\dot{F} = \frac{\partial(vecF(\theta_1))'}{\partial\theta_1}$ and V_X is the variance of the state vector defined in (8). The complete expressions for $\mathcal{A}_0^{-1}$, $\mathcal{B}_0$ and $\mathcal{C}_0$ can be found in the Appendix in (A.17), (A.1) and (A.18) respectively; we leave them out of the main text for brevity and provide some intuition instead.

4. Since we take the linearised model in (1) as a starting point for our analysis, we provide a brief clarification on the effect of the linearisation on our main result. While the quality of the linear approximation to the underlying nonlinear rational expectation model depends on how close the model is to its deterministic steady state (i.e. how close $\Sigma(\theta_2)$ is to zero), it is not directly affected by departures from Gaussianity of the shocks. Moreover, the result in Theorem 1 is not driven by the linearity of the solution alone; for example, if the solution matrix $F(\cdot)$ and hence the conditional mean μ_t were dependent on θ_2, the result will no longer hold. This is the case, for example, not only with: (i) nonlinear solution methods where the underlying nonlinear rational expectation model is solved via higher-order perturbation or projection methods (e.g. see Auroba et al. (2005)), and so the law of motion for $\mathbf{x}_t$ (and hence the conditional mean μ_t) in general, depend on θ_2; but also (ii) linear risk-adjusted solution methods where the nonlinear rational expectation model is linearlised around a risk-adjusted steady state that depends on the volatility parameters θ_2 (as in Coeurdacier, Rey and Winant (2011)) and the resulting solution matrix $F(\cdot)$ (and hence μ_t) depend on θ_2. Therefore, the result in Theorem 1 above is a *direct consequence* of the certainty equivalence of the linearised rational expectation model. Since the prevailing approximation method in the literature is (log)linearisation around a deterministic steady state, due to its simplicity and computational convenience, the result of Theorem 1 has wide-ranging implications for applied researchers, providing a formal justification for the use of Gaussian assumptions on the structural shocks. A more cautionary spin of the result in Theorem 1 suggests that linearisation around a deterministic steady state not only removes any uncertainty effects from the shocks on the solution of the model, it also eliminates any effects from the non-Gaussian features of the shocks on econometric procedures. This may be particularly undesirable when the objective is to study higher order effects of the stochastic component of the model.

5. $[\mathcal{C}_0]_{11}$ does not depend on the higher moments $\mathcal{S}$ or $\mathcal{K}$ of the structural shocks, so if inference is only needed for the structural parameters θ_1, Assumptions 5b and 5c are not necessary and uniform integrability of $\left(\|\varepsilon_t\|^2\right)_{t\geq 0}$ is sufficient; hence fat tailed shocks (e.g. infinite skewness/kurtosis shocks) do not invalidate inference on θ_1 (see DGP V in Section 4 for simulation results with infinite skewness shocks). For inference on θ_2, existence of higher order moments through the uniform integrability of $\left(\|\boldsymbol{\varepsilon}_t\|^4\right)_{t\geq 1}$ (Assumptions 5c) is necessary.

6. For the $[\mathcal{C}_0]_{22}$ block, there is no generalised information equality, $[\mathcal{C}_0]_{22} \neq \left[\mathcal{A}_0^{-1}\right]_{22}$ unless the standardised shocks $\Sigma^{-1/2}\boldsymbol{\varepsilon}_t$ have skewness $\mathcal{S}_{\Sigma^{-1/2}\boldsymbol{\varepsilon}_t}$=0 and kurtosis $\mathcal{K}_{\Sigma^{-1/2}\boldsymbol{\varepsilon}_t}$=$I_{m^2} + K_m$ where K_m is the $m^2 \times m^2$ commutation matrix, as is the case, for example, under correct specification when the shocks are Gaussian: $\Sigma^{-1/2}\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, I_m)$. Hence, both classical and Bayesian inference on θ_2 would be invalid even for large samples when Gaussianity is incorrectly imposed; the degree to which inference is invalidated will depend on how far the skewness and kurtosis of the true shocks are from those of the Gaussian distribution.

7. The $[\mathcal{C}_0]_{22}$ block corresponding to the asymptotic variance of the reduced-form volatility parameters θ_2 is nonstandard and depends not only on the kurtosis of the structural shocks (as is the case in the fully separable case when $\dot{\Omega}_1 = 0$ and $[\mathcal{C}_0]_{22} = \dot{\Omega}_2'^{-1}\mathcal{K}_u\dot{\Omega}_2^{-1}$) but also on: (i) the skewness of the shocks (whenever there is an intercept included in (3)), and (ii) various derivatives of the solution matrices which cannot be evaluated analytically for a typical DSGE model, rendering a good quality estimator for $[\mathcal{C}_0]_{22}$ computationally difficult and costly[9]. However, the asymptotic variance of θ_2 conditional on θ_1 has a much simpler form that only depends on the fourth moment of the shocks, and in Section 3, we exploit this conditioning argument and demonstrate that robust distribution-free inference on θ_2 can be achieved through the use of a simple Metropolis-within-Gibbs algorithm, at no additional computational cost relative to standard Bayesian estimation based on the Metropolis-Hastings algorithm.

8. Gaussian-based MLE confidence intervals and Bayesian credible sets on quantities that depend on θ_2 are also incorrect whenever the true shocks are non-Gaussian; for example, inference on one-standard-deviation impulse response functions to structural shocks can be invalid, while one-unit impulse response functions are not, since the latter only depend on θ_1 and not on θ_2.

3 Robust conditional inference on θ_2

We now turn attention to inference on θ_2 in the absence of Gaussianity. While the result of Theorem 1 is general and allows $\Sigma(\theta_2)$ to be full $\theta_2 = vech\Sigma$, in applied work, the structural nature of the shocks requires that they are at least mutually uncorrelated. Moreover, whenever the structural shocks $\boldsymbol{\varepsilon}_t$ are non-Gaussian, one needs to take a stance on both their contemporaneous linear and nonlinear dependence. In particular, in this case, mutual orthogonality does not rule out dependence and a stance is needed on what constitutes a non-Gaussian fundamental shock: independence or orthogonality from other shocks[10]. In the case of independence, the resulting asymptotic variance of the volatility parameters is a diagonal matrix with simpler form; however, the procedure we propose in this section works for both orthogonal and independent shocks and we leave this choice to the practitioner.

Since the structural shocks are at least orthogonal, we let $\Sigma(\theta_2)$ be a diagonal covariance matrix and $\theta_2 = Pvec\Sigma(\theta_2)$, where P is an $n \times n^2$ selector matrix with $[P]_{i,(i-1)n+i} = 1$ for $i = 1, ..., n$ and zeros elsewhere and θ_2 contains the diagonal elements $\theta_2 = [\sigma_1^2, ..., \sigma_m^2]'$ of $\Sigma(\theta_2)$: $\sigma_i^2 = [\Sigma(\theta_2)]_{ii}$. If the shocks $(\boldsymbol{\varepsilon}_t)_{t=1}^n$ were observed, it follows that (e.g. see Petrova (2022)),

$$\sqrt{n}\left(\hat{\theta}_2 - \theta_2\right) \mid (\boldsymbol{\varepsilon}_t)_{t=1}^n \to_d \mathcal{N}(0, V_{\theta_2}) \text{ as } n \to \infty$$

where $V_{\theta_2} = PD_m^+\left(\mathcal{K} - vech\Sigma\,(vech\Sigma)'\right)D_m'^+P'$, $\mathcal{K} = \mathbb{E}_{\mathcal{F}_{t-1}}\left[vech\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\left(vech\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right)'\right]$, D_m^+ is the Moore-Penrose inverse of the duplication matrix D_m, and hence V_{θ_2} is a full matrix with typical element $[V_{\theta_2}]_{ij} = \mathbb{E}_{\mathcal{F}_{t-1}}\left[\varepsilon_{it}^2\varepsilon_{jt}^2\right] - \sigma_i^2\sigma_j^2$. If, in addition, the shocks $\boldsymbol{\varepsilon}_t$ are mutually independent, we have $\mathbb{E}_{\mathcal{F}_{t-1}}\left[\varepsilon_{it}^2\varepsilon_{jt}^2\right] = \sigma_i^2\sigma_j^2$ for $i \neq j$ and so V_{θ_2} simplifies to a diagonal matrix with elements $[V_{\theta_2}]_{ii} = \mathbb{E}_{\mathcal{F}_{t-1}}\left[\varepsilon_{it}^4\right] - \sigma_i^4$. Bayesian treatment through an informative prior distribution can easily

[9] Numerical derivatives can be used to obtain an estimator for $[\mathcal{C}_0]_{22}$, but these are typically of poor quality.

[10] Lanne et al. (2017) argue that the correct assumption in a non-Gaussian setup is mutual independence.

be added for θ_2; this is not pursued here since not much prior information is typically available for the reduced-form parameters θ_2 which determine the 'size' of the shocks; instead, we proceed by imposing a flat noninformative prior on θ_2.

The shocks $(\boldsymbol{\varepsilon}_t)_{t=1}^n$ are not observed, but conditional on a draw for θ_1, a draw from the history of structural shocks $(\hat{\boldsymbol{\varepsilon}}_t)_{t=1}^n$ can be obtained through a disturbance smoother (e.g. Carter and Kohn (1994) or Durbin and Koopman (2002)). This allows to exploit the conditional large sample distribution above. To make a draw from it, we need a consistent estimator for V_{θ_2}, for example,

$$\left[\hat{V}_{\theta_2}\right]_{ij} = \frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{it}^2\hat{\varepsilon}_{jt}^2 - \left(\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{it}^2\right)\left(\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{jt}^2\right), \tag{11}$$

which further simplifies to a diagonal matrix with elements

$$\left[\hat{V}_{\theta_2}\right]_{ii} = \frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{it}^4 - \left(\frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_{it}^2\right)^2 \tag{12}$$

if mutual independence is assumed on the shocks.

On the other hand, given θ_2, the DSGE model in (2) and (3) is a standard linear state space with known covariance matrix $\Sigma(\theta_2)$, and so a standard Metropolis-Hastings step can be used to make a draw from the conditional posterior of θ_1. This conditioning argument gives rise to the following Metropolis-within-Gibbs algorithm, designed to approximate the joint posterior of $[\theta_1', \theta_2']'$ by recursively making draws from the conditional posteriors of θ_1 and θ_2 respectively.

——— **Algorithm 1**———

Step 1. Initialise the algorithm at a starting value θ^0; for example, the posterior mode obtained through numerical optimisation can be used: $\theta^0 = \arg\max_\theta p(\theta|y_{1:T})$.

For $i = 1, ..., N^{sim}$, iterate between the following steps:

Step 2. (Disturbance Smoother Step) Given θ_1^{i-1} draw the history of structural shocks $[(\hat{\boldsymbol{\varepsilon}}_t)_{t=1}^n]^i$ from the state space model (2) and (3), e.g. using Carter and Kohn (1994) or Durbin and Koopman (2002) algorithms.

Step 3. (Gibbs Step) Conditional on the fitted shocks $[(\hat{\boldsymbol{\varepsilon}}_t)_{t=1}^n]^i$, draw θ_2^i from $\mathcal{N}\left(0, \frac{1}{n}\hat{V}_{\theta_2}\right)$ with $\hat{V}_{\theta_2}$ defined in (11) (or (12) if mutual independence of the shocks is imposed).

Step 4 (Metropolis Step). Conditional on the draw θ_2^i, draw ϑ from the proposal distribution $\mathcal{N}(\theta_1^{i-1}, c^2\Lambda)$, where Λ is a positive definite symmetric matrix[11], and c^2 is a scaling parameter, controlling the step size and hence the rejection probability. Compute

$$r = \frac{\exp\left(\sum_{t=1}^n \ell(\mathbf{y}_t|\mathcal{F}_{t-1}, \vartheta, \theta_2^i)\right) p(\vartheta)}{\exp\left(\sum_{t=1}^n \ell(\mathbf{y}_t|\mathcal{F}_{t-1}, \theta_1^{i-1}, \theta_2^i)\right) p(\theta_1^{i-1})},$$

accept the proposal (setting $\theta_1^i = \vartheta$) with probability $\tau = min\{1, r\}$ and reject (setting $\theta_1^i = \theta_1^{i-1}$) with probability $1 - \tau$.

4 Monte Carlo

We design a small Monte Carlo exercise to confirm the result in Theorem 1 as well as to study the finite sample properties of the estimator obtained through the Metropolis-within-Gibbs

[11]For example, the Hessian evaluated at the posterior mode might be used for Λ; the theoretical properties of the Metropolis algorithm are unaffected by the choice for Λ as long as it is symmetric p.d. and fixed across draws.

algorithm of Section 3, designed to correct the posterior of the volatility parameters θ_2, and assess how it compares to the standard estimator based on Gaussian likelihood.

We consider a standard small three-equation closed economy New-Keyensian model (e.g. see Lubik and Schorfheide (2004) or Del Negro and Schorfheide (2013)). The linearised model takes the form of a Taylor rule, Phillips curve and an Euler equation respectively:

$$\begin{aligned} r_t &= \rho_r r_{t-1} + (1-\rho_r)\left(\psi_1 \pi_t + \psi_2 (y_t - z_t)\right) + \sigma_r \varepsilon_{rt} \\ \pi_t &= \beta \mathbb{E}_t \pi_{t+1} + \kappa (y_t - z_t) \\ y_t &= \mathbb{E}_t y_{t+1} - \tau (r_t - \mathbb{E}_t \pi_{t+1}) + g_t \end{aligned} \tag{13}$$

where r_t, π_t and y_t denote the nominal interest rate, inflation and output respectively (expressed in deviations from steady states), ψ_1 and ψ_2 are Taylor rule parameters defining the policy maker's inflation and output targeting rule, ε_{rt} is a policy shock, β is the discount factor $\beta = (1 + 0.01r^*)^{1/4}$ where r^* is the steady state interest rate, κ is the slope of the Phillips curve and τ is the intertemporal substitution elasticity. The demand and technology exogenous processes, g_t and z_t, are assumed to follow AR(1) specifications:

$$\begin{aligned} g_t &= \rho_g g_{t-1} + \sigma_g \varepsilon_{gt} \\ z_t &= \rho_z z_{t-1} + \sigma_z \varepsilon_{zt} \end{aligned}$$

where ε_{gt} and ε_{zt} are demand and technology structural shocks respectively. The structural parameters are given by $\theta_1 = \left[\pi^*, r^*, \kappa, \psi_1, \psi_2, \tau^{-1}, \rho_r, \rho_g, \rho_z\right]$ and the volatility parameters are given by $\theta_2 = \left[\sigma_r^2, \sigma_g^2, \sigma_r^2\right]$.

We simulate artificial data from the solution of the model in (13) and generate 5,000 artificial samples for five data generating processes (DGPs) with different shock distributions for the shocks ε_{rt}, ε_{gt} and ε_{zt}. We estimate the model with: (i) a standard Bayesian random walk Metropolis algorithm based on Gaussian likelihood (G-DSGE), and (ii) the Metropolis-within-Gibbs algorithm (MHG-DSGE) proposed in Section 3. Details on the prior distributions, true values as well as point estimate (bias and RMSE) comparison[12] between G-DSGE and MHG-DSGE algorithms can be found in Section 7.4 of the Appendix. Here, we focus on the coverage rates of the resulting posterior distributions based on 5,000 posterior draws, measured by the percentage of times the true parameter value is contained in the 68%, 90%, 95% and 99% credible set respectively.

We begin with Gaussian structural shocks in DGP I:

$$\varepsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \; i \in \{r, g, z\}.$$

DGP I serves as a benchmark to verify that both procedures yield correct coverage rates even when there is no distributional misspecification. Table 1 reports the resulting posterior coverage rates for the G-DSGE and MHG-DSGE estimation procedures respectively for sample sizes $n \in \{200, 500, 1000\}$. Both procedures perform well and the coverage rates get close to the nominal rates as the sample size increases. Next, in DGP II, we consider standardised t-distributed shocks with degrees of freedom $\nu = 5$:

$$\varepsilon_{it} \overset{i.i.d.}{\sim} \frac{1}{\sqrt{\nu/(\nu-2)}} t_5, \; i \in \{r, g, z\}.$$

[12] Both specifications exhibit similar point estimate performance.

Table 1: Posterior Coverage DGP I

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_z
n=200 G-DSGE	68%	69.3%	83.3%	72.5%	62.8%	65.0%	75.7%	87.1%	87.7%	89.5%	66.9%	73.2%	73.7%
	90%	90.7%	97.7%	92.5%	83.1%	90.1%	94.4%	96.4%	95.4%	99.0%	89.4%	93.4%	93.3%
	95%	95.3%	99.4%	96.2%	91.9%	94.6%	97.6%	98.4%	97.5%	99.7%	94.3%	96.7%	97.1%
	99%	98.7%	99.9%	99.3%	96.9%	98.5%	99.6%	99.6%	99.5%	100.0%	98.8%	99.5%	99.7%
n=200 MHG-DSGE	68%	62.1%	76.0%	70.3%	63.9%	65.4%	72.4%	76.8%	75.0%	85.5%	60.6%	71.7%	72.6%
	90%	85.3%	94.1%	90.0%	86.9%	88.0%	92.3%	92.9%	90.4%	98.2%	85.4%	92.3%	92.0%
	95%	91.3%	97.4%	94.3%	91.2%	93.6%	95.9%	96.1%	94.7%	99.4%	91.6%	96.4%	96.2%
	99%	97.5%	99.6%	98.7%	97.4%	98.2%	99.0%	99.1%	98.4%	100.0%	97.8%	99.4%	99.2%
n=500 G-DSGE	68%	71.4%	78.2%	68.4%	63.6%	64.6%	74.1%	94.2%	95.8%	81.4%	70.8%	70.8%	69.2%
	90%	92.0%	95.7%	91.4%	85.7%	87.0%	93.3%	98.7%	98.7%	95.6%	91.8%	91.3%	90.7%
	95%	96.4%	98.0%	95.5%	91.7%	92.4%	96.7%	99.3%	99.3%	98.1%	95.9%	95.6%	95.6%
	99%	99.3%	99.6%	98.8%	96.8%	97.2%	99.1%	99.8%	99.8%	99.5%	99.0%	99.1%	99.1%
n=500 MHG-DSGE	68%	68.2%	72.7%	68.3%	63.7%	64.5%	70.7%	91.4%	91.1%	77.3%	68.8%	69.3%	68.9%
	90%	90.6%	92.5%	90.5%	84.8%	84.9%	90.4%	97.2%	96.9%	94.0%	90.1%	91.4%	90.4%
	95%	95.0%	96.3%	95.2%	90.9%	91.5%	94.9%	98.3%	97.9%	96.7%	94.5%	95.3%	95.0%
	99%	98.8%	99.1%	98.8%	96.4%	96.7%	98.7%	99.1%	99.2%	99.4%	98.7%	98.9%	99.1%
n=1000 G-DSGE	68%	72.3%	71.2%	67.8%	64.5%	66.6%	71.0%	95.1%	96.6%	73.7%	72.2%	68.6%	68.2%
	90%	92.5%	91.5%	90.0%	86.2%	88.0%	91.3%	99.6%	99.5%	92.1%	92.3%	90.3%	90.3%
	95%	96.3%	95.8%	94.9%	89.0%	91.0%	95.5%	99.7%	99.7%	96.4%	96.5%	95.3%	95.2%
	99%	99.4%	99.2%	99.1%	97.6%	97.6%	99.1%	99.9%	99.8%	98.9%	99.4%	99.0%	99.0%
n=1000 MHG-DSGE	68%	71.1%	67.7%	67.9%	64.9%	66.8%	66.4%	95.0%	96.7%	68.8%	71.0%	68.0%	67.7%
	90%	92.2%	88.2%	88.9%	87.1%	87.2%	88.3%	99.3%	99.3%	89.8%	92.2%	89.8%	90.3%
	95%	96.0%	93.6%	94.4%	92.0%	93.0%	93.2%	99.7%	99.6%	94.1%	96.6%	95.0%	94.9%
	99%	99.3%	98.0%	98.6%	97.4%	97.9%	98.8%	99.9%	99.9%	98.4%	99.3%	98.7%	98.7%

Table 2: Posterior Coverage DGP II

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_z
n=200 G-DSGE	68%	50.5%	79.3%	58.6%	63.2%	69.2%	77.3%	85.7%	86.8%	88.9%	66.1%	73.8%	74.5%
	90%	73.6%	95.4%	82.2%	85.5%	90.0%	94.9%	96.0%	95.0%	98.8%	89.1%	93.1%	93.4%
	95%	81.6%	97.7%	88.5%	91.0%	94.3%	97.6%	97.9%	96.9%	99.7%	93.6%	96.8%	96.7%
	99%	92.4%	99.6%	95.8%	97.2%	98.4%	99.8%	99.5%	98.7%	99.9%	98.7%	99.6%	99.4%
n=200 MHG-DSGE	68%	59.3%	73.1%	69.1%	64.9%	69.2%	74.1%	71.6%	68.2%	82.9%	61.1%	74.1%	75.3%
	90%	80.3%	92.0%	89.8%	86.7%	89.5%	92.8%	90.3%	86.8%	96.9%	85.4%	93.1%	94.2%
	95%	87.3%	95.9%	93.9%	91.9%	94.8%	96.0%	94.5%	92.8%	98.8%	91.8%	96.5%	96.9%
	99%	94.5%	99.0%	98.1%	97.4%	98.5%	98.9%	99.2%	98.1%	99.8%	97.9%	99.2%	99.3%
n=500 G-DSGE	68%	49.4%	72.4%	55.5%	63.7%	65.2%	74.8%	93.8%	96.1%	81.4%	70.2%	69.9%	69.4%
	90%	73.4%	92.4%	78.6%	86.4%	87.0%	94.3%	99.1%	98.9%	96.6%	91.5%	91.8%	90.8%
	95%	81.4%	96.0%	85.7%	91.6%	92.9%	97.6%	99.4%	99.3%	98.3%	95.7%	96.2%	95.7%
	99%	91.9%	99.2%	94.1%	97.2%	97.8%	99.5%	99.9%	99.7%	99.5%	99.1%	99.2%	99.1%
n=500 MHG-DSGE	68%	64.2%	71.5%	68.3%	64.4%	65.9%	73.2%	86.4%	86.4%	77.0%	68.1%	72.2%	72.3%
	90%	85.2%	91.7%	89.6%	86.2%	86.9%	93.1%	96.2%	95.1%	93.2%	90.3%	92.5%	92.3%
	95%	90.9%	95.9%	94.2%	91.8%	92.3%	96.8%	97.7%	97.2%	96.4%	95.2%	96.8%	96.3%
	99%	96.5%	98.8%	98.1%	97.0%	97.3%	99.1%	99.1%	98.9%	99.1%	98.7%	99.6%	99.3%
n=1000 G-DSGE	68%	52.0%	67.6%	53.1%	64.7%	65.5%	69.8%	94.9%	96.0%	71.6%	71.4%	69.3%	69.3%
	90%	76.6%	88.4%	76.5%	83.3%	88.5%	91.0%	99.7%	99.6%	90.3%	92.0%	90.3%	90.7%
	95%	83.8%	93.2%	84.5%	89.8%	95.3%	94.9%	99.9%	99.9%	94.7%	96.2%	94.8%	95.1%
	99%	92.6%	98.0%	93.6%	96.4%	97.4%	98.7%	100.0%	99.9%	98.0%	99.4%	99.2%	99.0%
n=1000 MHG-DSGE	68%	69.0%	69.3%	69.9%	65.7%	67.0%	69.3%	93.9%	95.2%	70.0%	74.3%	69.0%	72.0%
	90%	88.1%	89.7%	90.3%	87.5%	87.6%	90.6%	98.5%	98.2%	88.9%	93.6%	92.4%	92.0%
	95%	93.1%	94.4%	95.0%	93.2%	90.4%	95.0%	99.1%	99.0%	93.9%	96.7%	96.1%	96.5%
	99%	98.3%	98.5%	98.4%	97.2%	96.8%	98.6%	99.7%	99.6%	98.2%	99.1%	99.1%	99.3%

In Table 2, we report the resulting posterior coverage rates for the G-DSGE and MHG-DSGE estimation procedures respectively for different sample sizes. From Table 2, it is clear that the coverage rates of the standard G-DSGE procedure for the structural parameters θ_1 are not distorted by the distributional misspecification, as implied by Theorem 1. Moreover, it is evident that the associated coverage rates for the volatility parameters θ_2 of the G-DSGE procedure are only slightly distorted but crucially do not improve with the sample size, as expected. On the other hand, the proposed procedure MHG-DSGE, based on a consistent estimator for the sample kurtosis of the structural shocks and designed to provide valid inference on θ_2, delivers satisfactory coverage for θ_2, as well as for the structural parameters θ_1.

In DGP III, we consider a mixture distribution between Gaussian and t-distribution for the shocks:

$$\tilde{\varepsilon}_{it} \overset{i.i.d.}{\sim} \begin{cases} \zeta_{it} \text{ w.p. } \lambda \ , \ \zeta_{it} \overset{i.i.d.}{\sim} \mathcal{N}\left(\alpha_1, 1\right), \ \alpha_1 = 1, \ \lambda = 0.95 \\ w_{it} \text{ w.p. } \left(1-\lambda\right), \ w_{it} \overset{i.i.d.}{\sim} \alpha_2 + \frac{1}{\sqrt{\nu/(\nu-2)}} t_5, \ \nu = 5, \ \alpha_2 = -20 \end{cases} \quad i \in \{r, g, z\},$$

and ε_{it} are standardised to have unit variance: $\varepsilon_{it} = \frac{1}{\sqrt{v_i}}\tilde{\varepsilon}_{it}$ where $\sqrt{v_i} \approx 2.17$. We report the associated posterior coverage rates for DGP III in Table 3. Once again, it is clear that the coverage rates for the volatility parameters θ_2 of the standard G-DSGE procedure are distorted (in this DGP more severely since the departure from Gaussianity is more serious); for example the volatility of the monetary policy shock σ_r is contained around 33% of the time in the 68% credible set and this does not improve even for $n = 1000$. On the other hand, the proposed procedure MHG-DSGE delivers good coverage for θ_2 converging to the nominal rates as the sample increases.

Table 3: Posterior Coverage DGP III

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_z
n=200 G-DSGE	68%	33.4%	62.0%	38.3%	63.5%	67.7%	74.4%	84.7%	86.7%	85.4%	65.9%	74.6%	75.4%
	90%	52.3%	84.9%	60.1%	84.1%	89.5%	93.3%	95.7%	95.5%	97.3%	88.3%	93.7%	94.4%
	95%	59.9%	91.2%	68.7%	90.4%	94.7%	96.7%	97.7%	97.1%	98.9%	93.6%	97.2%	97.2%
	99%	74.8%	97.3%	83.4%	96.3%	98.3%	99.3%	99.2%	98.9%	99.7%	98.8%	99.6%	99.3%
n=200 MHG-DSGE	68%	62.5%	61.1%	68.6%	64.6%	67.8%	71.4%	48.1%	40.5%	65.3%	55.8%	79.8%	78.7%
	90%	83.3%	83.3%	89.1%	86.5%	89.3%	91.3%	80.3%	74.0%	88.8%	82.9%	95.8%	94.0%
	95%	88.9%	88.9%	94.1%	91.5%	93.8%	95.6%	88.8%	84.6%	93.8%	91.1%	98.1%	96.7%
	99%	94.2%	95.2%	97.4%	97.3%	97.8%	98.7%	96.4%	94.9%	98.3%	98.0%	99.4%	99.2%
n=500 G-DSGE	68%	33.8%	63.5%	38.5%	63.8%	65.0%	73.1%	93.8%	95.6%	78.5%	70.5%	72.6%	72.6%
	90%	52.1%	85.0%	59.8%	85.1%	87.2%	93.5%	98.5%	98.6%	95.6%	91.8%	92.6%	93.7%
	95%	59.3%	90.9%	67.9%	91.0%	92.9%	96.8%	99.3%	99.2%	98.1%	96.1%	96.4%	96.9%
	99%	74.3%	97.3%	82.3%	97.1%	97.9%	99.5%	99.6%	99.5%	99.4%	99.1%	99.1%	99.3%
n=500 MHG-DSGE	68%	63.6%	65.3%	68.7%	66.4%	66.7%	71.5%	71.4%	68.0%	66.6%	67.0%	81.0%	80.9%
	90%	84.4%	86.6%	90.8%	87.3%	88.2%	91.2%	89.4%	87.4%	88.6%	89.1%	96.3%	95.8%
	95%	90.6%	91.7%	95.2%	92.3%	93.8%	95.4%	94.3%	93.0%	93.4%	94.3%	98.5%	98.3%
	99%	96.6%	97.3%	98.6%	97.1%	97.9%	98.7%	98.2%	97.5%	98.4%	98.3%	99.6%	99.6%
n=1000 G-DSGE	68%	33.6%	63.0%	38.9%	64.4%	66.3%	72.5%	94.6%	95.8%	75.5%	72.7%	71.3%	69.2%
	90%	55.6%	85.0%	58.6%	86.3%	87.4%	92.6%	98.7%	98.8%	93.0%	92.3%	91.7%	90.9%
	95%	63.9%	91.0%	68.0%	92.6%	93.8%	96.4%	99.3%	99.4%	96.5%	96.3%	95.8%	95.5%
	99%	77.3%	96.6%	82.2%	97.1%	98.0%	99.2%	99.7%	99.6%	99.1%	99.4%	99.3%	99.0%
n=1000 MHG-DSGE	68%	62.8%	68.2%	70.7%	69.0%	69.5%	75.2%	71.4%	68.9%	72.3%	67.4%	82.1%	81.1%
	90%	85.9%	89.4%	91.6%	89.4%	89.6%	93.0%	88.4%	87.2%	91.3%	90.7%	97.1%	96.9%
	95%	91.7%	94.1%	95.6%	94.0%	93.8%	96.3%	93.1%	92.8%	95.0%	95.8%	99.0%	98.8%
	99%	97.6%	98.3%	99.0%	97.9%	98.0%	99.1%	98.7%	97.5%	98.3%	99.0%	99.8%	99.9%

In DGP IV, we consider another mixture distribution between Gaussian and inverse-Gaussian distributions for the shocks of the form:

$$\tilde{\varepsilon}_{it} \overset{i.i.d.}{\sim} \begin{cases} \zeta_{it} \text{ w.p. } \lambda \ , \ \zeta_{it} \overset{i.i.d.}{\sim} \mathcal{N}\left(\alpha_1, 1\right), \ \alpha_1 = 1, \ \lambda = 0.71 \\ w_{it} \text{ w.p. } \left(1-\lambda\right), \ w_{it} \overset{i.i.d.}{\sim} IG\left(\alpha_2, 3\right), \ \alpha_1 = -2.5 \end{cases} \quad i \in \{r, g, z\}$$

and ε_{it} are standardised: $\varepsilon_{it} = \frac{1}{\sqrt{v_i}}\tilde{\varepsilon}_{it}$ where $\sqrt{v_i} \approx 4.58$. We report the associated posterior coverage rates for DGP IV in Table 4 below. The results reported in Table 4 confirm the conclusions from the other DGPs; namely, (i) the standard Gaussian G-DSGE model works well and delivers valid coverage for the structural parameters θ_1 supporting the main result of Theorem 1; (ii) Gaussian inference is distorted for the volatility parameters θ_2 and distortions do not disappear with the sample size increasing, and (iii) the proposed Metropolis-within-Gibbs algorithm corrects the coverage of the credible sets for θ_2 while delivering satisfactory performance for θ_1.

Table 4: Posterior Coverage DGP IV

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_z
n=200 G-DSGE	68%	39.9%	69.0%	44.6%	65.0%	68.0%	76.3%	86.0%	86.7%	89.2%	65.2%	74.4%	74.5%
	90%	61.1%	90.5%	69.1%	86.2%	89.7%	94.6%	96.2%	95.6%	98.5%	88.6%	92.9%	93.9%
	95%	68.4%	95.2%	77.2%	91.4%	94.1%	98.0%	98.0%	97.2%	99.5%	93.5%	96.8%	97.2%
	99%	82.5%	99.0%	89.8%	96.9%	97.9%	99.5%	99.1%	99.0%	100.0%	98.7%	99.5%	99.4%
n=200 MHG-DSGE	68%	57.9%	69.2%	65.1%	64.4%	67.4%	72.2%	62.6%	57.7%	79.3%	58.1%	76.0%	76.8%
	90%	76.5%	87.9%	86.3%	86.4%	89.8%	92.2%	87.0%	83.3%	95.3%	84.3%	94.1%	94.9%
	95%	82.1%	92.7%	91.2%	91.7%	93.6%	96.1%	93.1%	90.3%	97.6%	91.3%	97.4%	97.5%
	99%	88.9%	97.3%	96.3%	97.0%	98.3%	99.1%	98.2%	97.2%	99.6%	97.4%	99.7%	99.4%
n=500 G-DSGE	68%	39.6%	68.3%	43.5%	61.9%	63.9%	74.1%	93.6%	95.8%	82.2%	70.1%	70.0%	70.9%
	90%	61.1%	89.0%	67.6%	85.3%	86.1%	93.6%	98.8%	98.6%	96.3%	91.6%	91.4%	91.6%
	95%	68.7%	93.8%	76.4%	91.4%	91.6%	97.1%	99.4%	99.2%	98.2%	96.0%	96.3%	96.1%
	99%	82.7%	98.6%	88.1%	97.3%	97.2%	99.1%	99.8%	99.8%	99.4%	99.1%	99.3%	99.1%
n=500 MHG-DSGE	68%	63.2%	70.3%	67.8%	65.3%	67.0%	72.3%	80.3%	78.0%	73.3%	68.3%	75.3%	76.2%
	90%	82.6%	90.3%	87.9%	86.3%	88.3%	91.8%	93.6%	92.2%	91.8%	90.0%	94.5%	94.1%
	95%	88.0%	94.3%	91.6%	91.4%	93.1%	95.8%	96.1%	95.1%	95.3%	94.5%	97.6%	97.3%
	99%	94.2%	98.5%	96.8%	97.3%	97.8%	98.9%	98.8%	98.4%	98.4%	98.5%	99.6%	99.3%
n=1000 G-DSGE	68%	39.4%	66.6%	43.6%	64.5%	66.2%	72.5%	94.9%	96.5%	75.3%	71.9%	69.2%	70.0%
	90%	61.5%	87.3%	66.2%	86.9%	87.5%	92.8%	98.9%	98.8%	93.9%	92.4%	90.9%	91.3%
	95%	70.4%	93.5%	74.6%	91.3%	92.0%	96.8%	99.3%	99.2%	97.0%	96.6%	95.9%	95.9%
	99%	83.8%	98.1%	88.3%	97.2%	97.7%	99.3%	99.9%	99.7%	99.2%	99.4%	99.5%	99.1%
n=1000 MHG-DSGE	68%	63.9%	71.1%	69.4%	68.3%	67.2%	73.9%	79.0%	77.9%	72.5%	69.6%	77.9%	79.4%
	90%	83.3%	90.4%	90.7%	89.3%	89.5%	92.7%	92.1%	91.7%	90.9%	90.3%	95.8%	96.2%
	95%	89.0%	95.2%	94.8%	94.3%	93.9%	96.6%	95.6%	95.0%	95.1%	95.3%	98.3%	98.4%
	99%	95.6%	98.4%	98.0%	98.1%	98.1%	99.2%	98.4%	97.8%	98.5%	98.9%	99.8%	99.8%

Figure 1 displays the shock distributions for DGP III and IV respectively, which clearly exhibit very non-Gaussian features. Finally, in DGP V, we consider standardised t-distributed shocks with degrees of freedom $\nu = 3$:

$$\varepsilon_{it} \overset{i.i.d.}{\sim} \frac{1}{\sqrt{\nu/(\nu-2)}} t_3, \; i \in \{r, g, z\}.$$

The resulting shocks are fat-tailed with infinite skewness, which intentionally violates Assumptions 5b and 5c to investigate if Gaussian inference on the structural parameters θ_1 is affected.

In Table 5, we report the resulting posterior coverage rates for the G-DSGE for different sample sizes (the MHG-DSGE procedure is infeasible since it uses the sample kurtosis which in this case blows up since the corresponding population moment does not exist). From Table 5, it is clear that fat-tailed shocks do not distort the validity of Gaussian inference on the structural parameters θ_1, whose coverage is converging to the nominal rate, as the sample increases.

Figure 1: Error Distributions for DGP III and IV

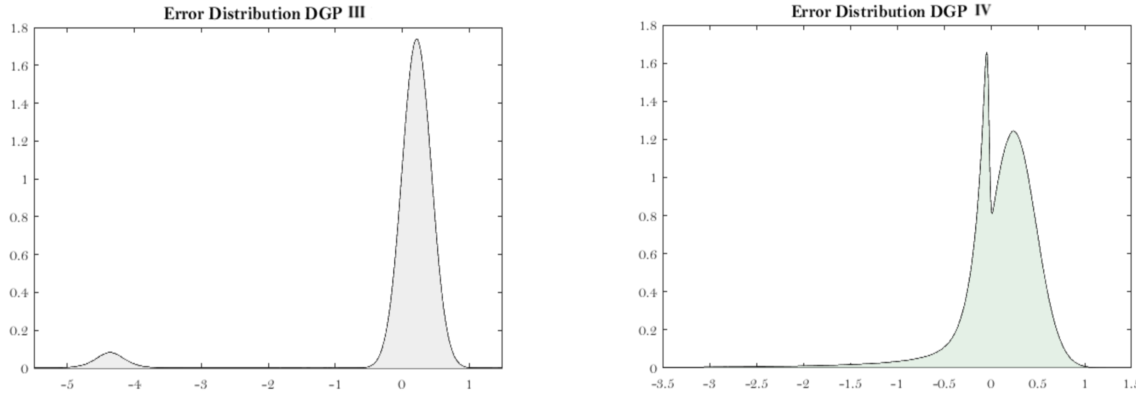


Table 5: Posterior Coverage DGP V

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_z
n=200 G-DSGE	68%	27.7%	56.1%	32.7%	57.2%	59.3%	68.1%	91.2%	94.6%	72.9%	67.8%	70.2%	72.9%
	90%	46.0%	80.4%	54.9%	81.0%	81.9%	90.0%	99.1%	99.3%	90.0%	90.9%	91.0%	93.8%
	95%	55.4%	87.6%	64.2%	87.6%	88.6%	93.9%	99.6%	99.7%	94.1%	95.2%	94.8%	96.8%
	99%	73.1%	95.4%	80.4%	94.9%	96.0%	98.3%	99.9%	99.8%	97.9%	98.8%	98.7%	99.3%
n=500 G-DSGE	68%	23.0%	55.2%	27.7%	59.4%	60.4%	68.9%	92.5%	95.4%	71.8%	69.8%	71.2%	70.1%
	90%	37.8%	78.5%	46.7%	82.4%	83.7%	91.0%	99.2%	99.4%	90.7%	91.1%	91.9%	91.3%
	95%	46.2%	86.2%	56.7%	89.1%	89.6%	95.0%	99.7%	99.7%	94.8%	95.5%	95.6%	95.2%
	99%	64.5%	95.4%	75.3%	96.1%	96.4%	98.8%	100.0%	99.9%	98.3%	99.1%	99.2%	99.0%
n=1000 G-DSGE	68%	22.1%	54.0%	24.4%	61.5%	62.8%	69.1%	91.4%	95.0%	70.9%	69.5%	69.3%	69.3%
	90%	36.3%	76.6%	41.3%	83.8%	83.9%	90.2%	97.3%	96.2%	90.5%	92.6%	90.9%	91.5%
	95%	43.7%	83.9%	50.4%	89.8%	90.5%	94.9%	99.7%	99.6%	94.3%	96.5%	95.6%	95.4%
	99%	60.5%	94.3%	67.7%	96.3%	96.0%	98.5%	99.9%	99.9%	98.0%	99.3%	99.1%	99.3%

5 Non-Gaussian shocks in financial friction DSGE model

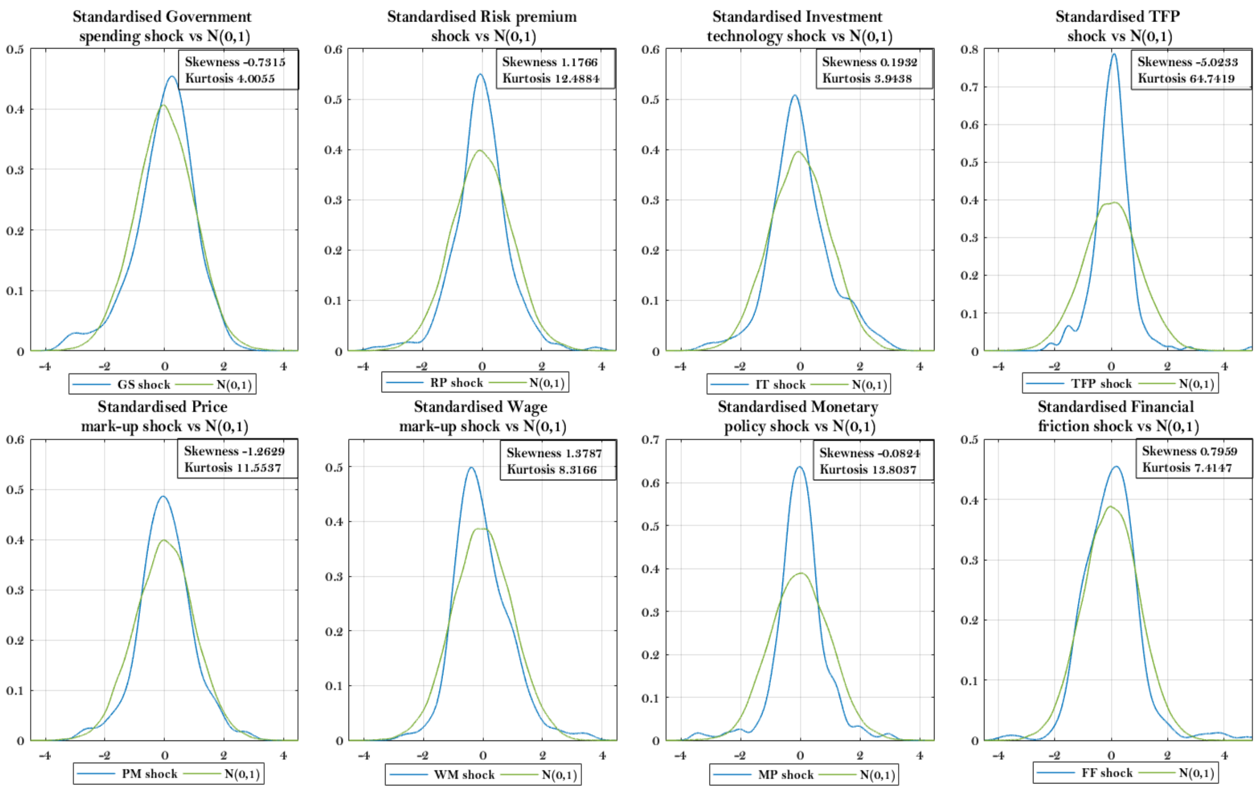
In this section, we estimate a Smets and Wouters (2007) model with an added financial sector as in Bernanke, Gertler and Gilchrist (1999). The choice of model is motivated by the possible non-Gaussian features of the observables and we make use of the model to highlight the large differences in the posterior distributions for the volatility parameters if the MHG algorithm from Section 3 is used, which is robust to non-Gaussianity of the shocks.

The financial sector in the model consists of entrepreneurs, subject to aggregate and idiosyncratic shocks, who borrow funds from banks at a premium. The financial friction is designed to 'accelerate' the impact of negative shocks increasing the default risk during recessions. The model follows the financial friction specification of Del Negro and Schorfheide (2013), with the only difference that we do not impose stochastic trend on productivity and estimate the autoregressive parameter on the productivity process instead, as in Smets and Wouters (2007). The complete log-linearised specification of the model, the measurement equations, prior distributions and data description can be found in Section 7.5 of the Appendix.

In Figure 2, we display the kernel estimated density of the standardised fitted shocks on the US data for the sample 1962Q1-2022Q4 and compare that to the standard normal density, as a simple diagnostic on the degree of non-Gaussianity of the fitted shocks. It is clear from the figure that some shocks (particularly TFP, monetary policy, risk premium and price shocks) display very non-Gaussian features over the sample with kurtosis much larger than the kurtosis of the standard normal. Consequently, we expect that imposing Gaussianity will have a considerable effect on Bayesian inference on the volatility of shocks in the model and may deliver posteriors with invalid coverage, as the theoretical results established in paper suggest.

We estimate the model on the US sample, using the random-walk Metropolis algorithm with Gaussian likelihood, as well as the Metropolis-within-Gibbs algorithm described in Section 3, designed to provide valid inference on the volatility of shocks, in the presence of shocks displaying non-Gaussian features. In particular, we estimate two specifications for our MHG algorithm: (i) imposing mutual independence on the shocks, and (ii) imposing mutual orthogonality but allowing for nonlinear dependence. The estimated posteriors for the volatilities of all eight shocks are displayed in Figure 3 below. The Appendix contains the posterior distributions for the structural parameters of the model for the different specifications[13].

Figure 2: Density of fitted Shocks

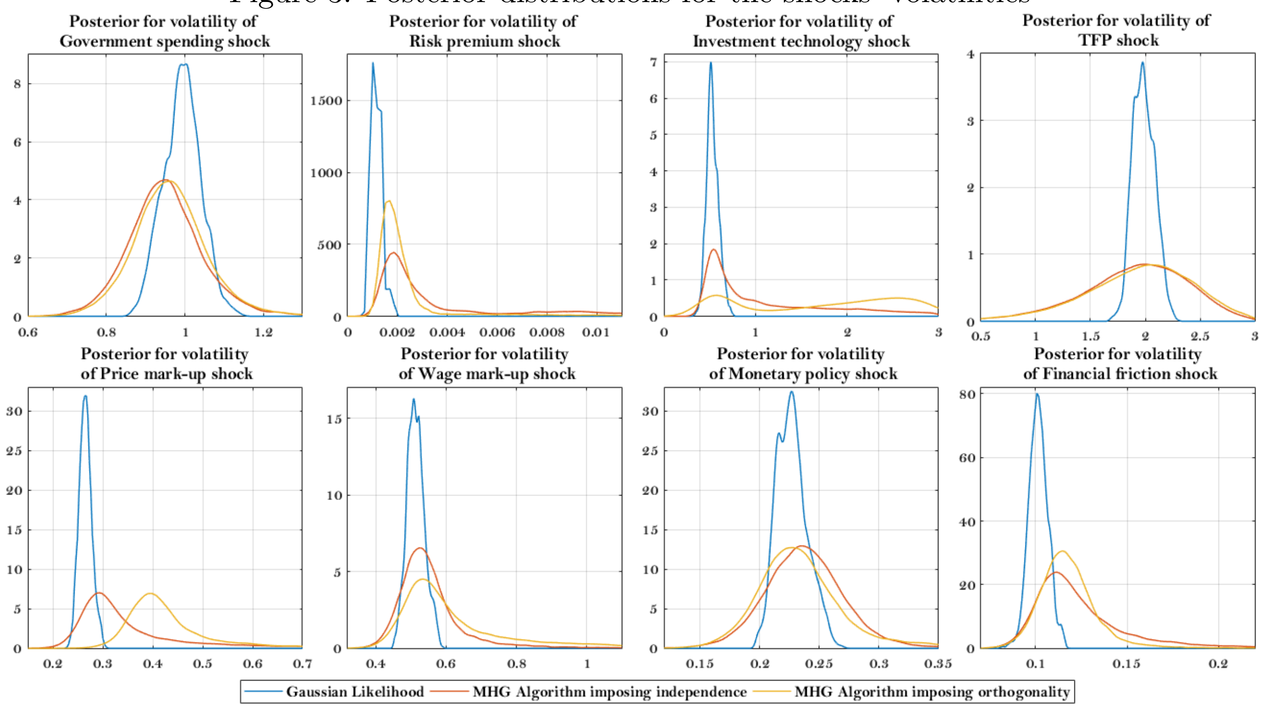


As anticipated, allowing for the possibility of non-Gaussian shocks delivers very different posterior distributions for the volatility parameters in Figure 3 and this can have serious implications when computing credible sets for quantities which depend on the volatility of shocks; for example, one standard deviation impulse responses, routinely reported in the literature, will have invalid credible sets if Gaussianity was incorrectly imposed. Another conclusion from Figure 3 is that

[13] As illustrated in the Monte Carlo exercise, any differences between the Gaussian and non-Gaussian specifications for θ_1 are due to small samples and are expected vanish as the sample size increases.

allowing the structural shocks to be mutually dependent can lead to slightly different posteriors for the volatility. Under dependence, the correct asymptotic sampling distribution for θ_2 (conditional on θ_1) has a non-diagonal covariance matrix as established in Section 3. Differences relative to the case when independence is imposed can arise: (i) due to small sample noise arising from the estimation of the additional $m(m-1)/2$ off-diagonal elements (28 in the context of the model considered), or (ii) whenever the true off-diagonal elements involving higher moment cross terms are not zero. In other words, the two specifications can give rise to different posterior distributions whenever some of the estimated volatility parameters θ_2 co-vary, which happens when the underlying structural shocks exhibit some nonlinear dependence.

Figure 3: Posterior distributions for the shocks' volatilities



6 Conclusion

While non-Gaussianity is an undeniable feature of many macroeconomic and financial time series, incorrectly imposing Gaussian assumptions on the structural shocks of a linear DSGE model is shown to have no asymptotic effect on classical and Bayesian inference on the structural parameters of the model. Consequently, the resulting MLE confidence intervals and Bayesian credible sets for the deep parameters have correct asymptotic coverage and no 'sandwich-form' corrections for the posterior variance are required. This surprising result is due to a cancellation in the asymptotic variance of the structural parameters leading to a generalised information equality for the corresponding block. The underlying reason for the cancellation is the certainty equivalence property of the linear rational expectation model, which ensures that the conditional first moment of the model's variables does not depend on the second moment of the structural shocks.

The main positive implication of the result is that DSGE-based inference is surprisingly robust: imposing a Gaussianity assumption on the structural shocks, which is convenient since it permits the use of the Kalman filter for likelihood evaluation, has no effect on the validity of classical and Bayesian inference for the structural parameters even when the true underlying structural shocks are non-Gaussian. On a more cautionary note, the result implies that linearisation of the DSGE model around a deterministic steady state not only washes away any uncertainty effects from the volatility of the shocks on the solutions matrices, but also any effects from the non-Gaussian features of the shocks on econometric procedures for the structural parameters.

References

Abadir, K. and Magnus, J. (2010) *Matrix Algebra*, Cambridge University Press

Altug, S. (1989) Time-to-build and aggregate fluctuations: some new evidence, *International Economic Review* **30**(4): 889-920

Auroba, B., Fernandez-Villaverde, J., Rubio-Ramirez, J. F. (2005). Comparing solution methods for dynamic equilibrium economies, *Journal of Economic Dynamics and Control* **30**: 2477-2508

Bernanke, B., Gertler, M., Gilchrist, S., 1999. The financial accelerator in a quantitative business cycle framework. In: Taylor, J.B., Woodford, M. (Eds.), *Handbook of Macroeconomics*. vol. 1C. Elsevier, pp. 1341–1393

Blanchard, O. J. and Kahn, C. M. (1980) The solution of linear difference models under rational expectations, *Econometrica* **48**(5): 1305-1312

Bollerslev, T. and Wooldridge, J. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews* **11**(2): 143-172

Canova, F. and Matthes, C. (2021a) A composite likelihood approach for dynamic structural models, *The Economic Journal* **131**: 2447–2477

Canova, F. and Matthes, C. (2021b) Dealing with misspecification in structural macroeconometric models, *Quantitative Economics* **12**: 313–350

Carter, C. K. and Kohn, R. (1994) On Gibbs sampling for state space models, *Biometrika* **81**: 541-553

Chernozhukov, V. and Hong, H. (2003) An MCMC approach to classical estimation, *Journal of Econometrics* **115**: 239-346

Christiano, L., Eichenbaum, M. and Evans, C. (2005) Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1-45

Coeurdacier, N., Rey, H., Winant, P. (2011) The risky steady state, *The American Economic Review* **101**(3), 398–401

Del Negro, M. and Schorfheide, F. (2013) DSGE Model-Based Forecasting in: G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2, Part A, pp. 57–140, Elsevier, NY

Durbin, J. and Koopman, S. (2002) A simple and efficient simulation smoother for state space time series analysis, *Biometrika* **89**(3): 603-616

Gourieroux, C., Monfort, A. and Trognon, A. (1984) Pseudo maximum likelihood methods: Theory, *Econometrica* **52**(3): 681-700

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2004) Comparing dynamic equilibrium models to data: a Bayesian approach, *Journal of Econometrics* **123**(1): 153-187

Gourieroux, C. and Monfort, A. (1993) Pseudo likelihood methods in: G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.), *Handbook of Statistics*, Volume 11, pp. 335-362, Elsevier

Guerron-Quintana, P., Inoue, A. and Kilian, L. (2017) Impulse response matching estimators for DSGE models. *Journal of Econometrics* **196**: 144-155

Hall, P. and Heyde, C. C. (1980) *Martingale Limit Theory and its Applications*, Academic Press

Ireland, P. (2004) Technology shocks in the new Keynesian model, *The Review of Economics and Statistics* **86**(4): 923-936

Iskrev, N. (2010) Local identification in DSGE models. *Journal of Monetary Economics* **57**: 189–202

Justiniano, A. and Primiceri, G. E. (2008) The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3): 604-641

Kydland, F. and Prescott, E. (1996) The computational experiment: an econometric tool, *Journal of Economic Perspectives* **10**: 69-85

Lanne, M., Meitz, M. and Saikkonen, P. (2017) Identification and estimation of non-Gaussian structural vector autoregressions, *Journal of Econometrics* **196**(2): 288-304

Lubik, T. and Schorfheide, F. (2004) Testing for Indeterminacy: An Application to U.S. Monetary Policy, *American Economic Review* **94**: 190–217

Magnus and Neudecker (2007) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley Series in Probability and Statistics

Müller, U. (2013) Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix, *Econometica* **81**(5): 1805-1849

Petrova, K. (2019) Quasi-Bayesian estimation of time varying volatility in DSGE models, *Journal of Time Series Analysis* **40**: 151-157

Petrova, K. (2022) Asymptotically valid Bayesian inference in the presence of distributional misspecification in VAR models, *Journal of Econometrics* **230** (1): 154-182

Rotemberg, J. and Woodford, M. (1997). An optimization-based econometric framework for the evaluation of monetary policy, in B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 297-346

Schorfheide, F. (2000) Loss function-based evaluation of DSGE models, *Journal of Applied Econometrics* **15**(6): 645-670

Sims, C. (2002) Solving linear rational expectations models, *Computational Economics* **20**(1-2): 1-20

Smets, F. and Wouters, R. (2007) Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review* **97**(3): 586-606

Qu, Z. and Tkachenko, D.(2012) Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models, *Quantitative Economics* **3**:95-132

White, H. (1982) Maximum likelihood estimation of misspecified models, *Econometica* **50**(1): 1-25

7 Appendix

This Appendix contains: (i) two auxiliary results (Lemmata 1 and 2) and their proofs in Section 7.2, (ii) the proof of Theorem 1 in Section 7.3, (iii) details on the Monte Carlo design and some additional simulation results in Section 7.4, and (vi) addition details and results on the empirical application in Section 7.5 below.

7.1 Notation

We make use of the following notation throughout. For an $m \times n$ matrix-variate function $X(\theta)$ of θ, for notational convenience, we suppress dependence on θ and write X instead of $X(\theta)$. We denote by $\dot{X}$ the $k_1 \times mn$ Jacobian matrix of first derivatives with respect to θ_1 : $\dot{X} = \frac{\partial (vecX)'}{\partial \theta_1}$, and for quantities that depend on both θ_1 and θ_2, we use $\dot{X}_1$ and $\dot{X}_2$ to denote the $k_1 \times mn$ and $k_2 \times mn$ Jacobian matrices of derivatives with respect to θ_1 and θ_2 respectively. If X is symmetric, we denote the $k_1 \times n(n+1)/2$ Jacobian of first derivatives by $\dot{X} = \frac{(\partial vechX)'}{\partial \theta_1}$ instead, where $vech(.)$ is the half-vec operator, satisfying $vecX = D_n vechX$, where D_n is the $n^2 \times n(n+1)/2$ duplication matrix with D_n^+ denoting its Moore Penrose inverse, such that $D_n^+ D_n = I_{n(n+1)/2}$ and $vechX = D_n^+ vecX$ (see Abadir and Magnus (2010)).

7.2 Auxiliary Results

Lemma 1. The score vector $s_t(\theta)$ in (10) satisfies $n^{-1/2}\sum_{t=1}^n s_t(\theta^0) \rightarrow_d \mathcal{N}(0, \mathcal{B}_0)$, with

$$\mathcal{B}_0 = \begin{bmatrix} \frac{1}{4}\dot{\Omega}_1\Phi\mathcal{K}_u\Phi\dot{\Omega}_1' + \frac{1}{2}\dot{\Omega}_1\Phi L + \frac{1}{2}L'\Phi\dot{\Omega}_1' + V & \frac{1}{4}\dot{\Omega}_1\Phi\mathcal{K}_u\Phi\dot{\Omega}_2' + \frac{1}{2}L'\Phi\dot{\Omega}_2' \\ \frac{1}{4}\dot{\Omega}_2\Phi\mathcal{K}_u\Phi\dot{\Omega}_1' + \frac{1}{2}\dot{\Omega}_2\Phi L & \frac{1}{4}\dot{\Omega}_2\Phi\mathcal{K}_u\Phi\dot{\Omega}_2' \end{bmatrix} \tag{A.1}$$

where $L = \mathcal{S}_u'\Omega^{-1}\dot{C}'$, $V = \dot{C}\Omega^{-1}\dot{C}' + \dot{HF}(V_X \otimes \Omega^{-1})(\dot{HF})'$ with $\dot{HF} = \frac{\partial vec(H(\theta_1)F(\theta_1))'}{\partial \theta_1}$, V_X defined in (8), $\Phi = D_r'(\Omega^{-1}\otimes\Omega^{-1})D_r$ where D_r denotes the $r^2 \times r(r+1)/2$ duplication matrix, $\mathcal{S}_u$ and $\mathcal{K}_u$ denote the multivariate skewness and kurtosis of $\mathbf{u}_t$ respectively given by $\mathcal{S}_u = \mathbb{E}_{\mathcal{F}_{t-1}}\left\{\mathbf{u}_t(vech[\mathbf{u}_t\mathbf{u}_t'])'\right\}$and $\mathcal{K}_u = \mathbb{E}_{\mathcal{F}_{t-1}}\left[vech(\mathbf{u}_t\mathbf{u}_t')(vech(\mathbf{u}_t\mathbf{u}_t'))'\right]$, and all quantities in (A.1) are evaluated at $\theta = \theta^0$.

Lemma 2. The Hessian matrix of second derivatives $\mathcal{H}_t(\theta) = \frac{\partial^2 \ell(\mathbf{y}_t,\theta|\mathcal{F}_{t-1})}{\partial\theta\partial\theta'}$ of $\ell(\mathbf{y}_t,\theta|\mathcal{F}_{t-1})$ satisfies

$$\frac{1}{n}\sum_{t=1}^n \left(\mathcal{H}_t(\theta_n^*) - \mathbb{E}_{\mathcal{F}_{t-1}}\mathcal{H}_t(\theta^0)\right) \rightarrow_p 0 \text{ as } n \rightarrow \infty \tag{A.2}$$

for all θ_n^* satisfying $\left\|\theta_n^* - \theta^0\right\| \leq \left\|\hat{\theta}_n - \theta^0\right\|$.

Proof of Lemma 1. The Gaussian (quasi) conditional log-likelihood is

$$\ell(\mathbf{y}_t, \theta|\mathcal{F}_{t-1}) = -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}tr\Omega^{-1}\mathbf{u}_t\mathbf{u}_t'$$

with $\mathbf{u}_t(\theta_1, \mathbf{x}_{t-1}) = \mathbf{y}_t - \mu_t$. To obtain the conditional quasi-score vector $s_t(\theta) = \frac{\partial \ell(\mathbf{y}_t,\theta|\mathcal{F}_{t-1})}{\partial\theta}$, first compute the first differential of ℓ_t

$$\begin{aligned} d\ell_t &= -\frac{1}{2}tr\left[(\Omega^{-1})d\Omega\right] + \frac{1}{2}tr\left[\Omega^{-1}(d\Omega)\Omega^{-1}\mathbf{u}_t\mathbf{u}_t'\right] + tr\left[\Omega^{-1}\mathbf{u}_t d\mu_t'\right] \\ &= \frac{1}{2}(dvech\Omega)' D_r'(\Omega^{-1}\otimes\Omega^{-1})D_r vech(\mathbf{u}_t\mathbf{u}_t' - \Omega) + d\mu_t'\Omega^{-1}\mathbf{u}_t \end{aligned}$$

where we have used the identities $tr(AB) = (vecA')' vecB$, $vec\Omega = D_r vech\Omega$ and $vec(ABC) = (C'\otimes A)vecB$. Taking derivatives with respect to θ_1 and θ_2, we obtain the expression for conditional (quasi) score vector as in (10):

$$s_t(\theta) = \begin{bmatrix} \frac{\partial \ell(\mathbf{y}_t,\theta|\mathcal{F}_{t-1})}{\partial\theta_1} \\ \frac{\partial \ell(\mathbf{y}_t,\theta|\mathcal{F}_{t-1})}{\partial\theta_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\dot{\Omega}_1 D_r'(\Omega^{-1}\otimes\Omega^{-1})D_r\mathbf{z}_t + \dot{\mu}_t\Omega^{-1}\mathbf{u}_t \\ \frac{1}{2}\dot{\Omega}_2 D_r'(\Omega^{-1}\otimes\Omega^{-1})D_r\mathbf{z}_t \end{bmatrix}, \tag{A.3}$$

where $\mathbf{z}_t(\theta_1, \theta_2, \mathbf{x}_{t-1}) = vech(\mathbf{u}_t\mathbf{u}_t' - \Omega)$. At $\theta = \theta^0$, $\mathbf{u}_t(\theta^0) = H(\theta_1^0) G(\theta_1^0) \boldsymbol{\varepsilon}_t$ and $\mathbf{z}_t(\theta^0) = vech(\mathbf{u}_t(\theta^0) \mathbf{u}_t'(\theta^0) - \Omega(\theta^0))$ are $\mathcal{F}_t$-martingale difference sequences since under Assumptions 1-2, the first two conditional moments of the observables are correctly specified; the functional form for the score vector in (A.3) then implies that $s_t(\theta^0)$ is an $\mathcal{F}_t$-martingale difference sequence at θ^0: $\mathbb{E}_{\mathcal{F}_{t-1}}(s_t(\theta^0)) = 0$. This, together with global identification imposed by Assumption 3 and uniform integrability (UI) of $(\|s_t(\theta^0)\|)_{t\geq 1}$ (implied by square UI established below) are sufficient for the consistency of QML estimator: $\hat{\theta}_n \to_p \theta^0$.

We now show that the sequences $\left(\|\mathbf{u}_t(\theta^0)\|^2\right)_{t\geq 1}$ and $\left(\|\mathbf{z}_t(\theta^0)\|^2\right)_{t\geq 1}$ are uniformly integrable. Firstly, since $\|H(\theta_1^0)\| \|G(\theta_1^0)\|$ is a non-random constant, UI of $\left(\|\mathbf{u}_t(\theta^0)\|^2\right)_{t\geq 1}$ follows from the UI of $(\|\varepsilon_t\|^2)_{t\geq 1}$. Since

$$
\begin{aligned}
\|\mathbf{z}_t(\theta^0)\|^2 &\leq \left(\left\|\mathbf{u}_t(\theta^0)\mathbf{u}_t(\theta^0)'\right\| + \|\Omega(\theta^0)\|\right)^2 \leq 2\|\mathbf{u}_t(\theta^0)\|^4 + 2\|\Omega(\theta^0)\|^2 \\
&\leq 2\|H(\theta_1^0)\|^4 \|G(\theta_1^0)\|^4 \|\varepsilon_t\|^4 + 2\|\Omega(\theta^0)\|^2,
\end{aligned}
$$

the UI of $\left(\|\mathbf{z}_t(\theta^0)\|^2\right)_{t\geq 1}$ follows from that of $(\|\varepsilon_t\|^4)_{t\geq 1}$. In view of (A.3), the UI of the sequences $\left(\|\mathbf{u}_t(\theta^0)\|^2\right)_{t\geq 1}$ and $\left(\|\mathbf{z}_t(\theta^0)\|^2\right)_{t\geq 1}$ implies the UI of the sequence $\left(\|s_t(\theta^0)\|^2\right)_{t\geq 1}$.

UI of $\left(\|s_t(\theta^0)\|^2\right)_{t\geq 1}$ may be used to establish the Lindeberg condition for the $\mathcal{F}_t$-martingale difference array $\xi_{nt} = n^{-1/2}s_t(\theta^0)$, namely $L_n(\delta) = \sum_{t=1}^n \mathbb{E}\left(\|\xi_{nt}\|^2 \mathbf{1}\{\|\xi_{nt}\| > \delta\}\right) \to 0$ for any $\delta > 0$. Substituting $\xi_{nt} = n^{-1/2}s_t(\theta^0)$, we obtain $L_n(\delta) \leq \max_{1\leq t\leq n} \mathbb{E}\left(\|s_t(\theta^0)\|^2 \mathbf{1}\{\|s_t(\theta^0)\|^2 > n\delta\}\right)$ is $o(1)$ by uniform integrability of $\left(\|s_t(\theta^0)\|^2\right)_{t\geq 1}$. Hence, as long as

$$
\mathcal{B}_0 = \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \mathbb{E}_{\mathcal{F}_{t-1}}\left[s_t(\theta^0) s_t(\theta^0)'\right] \qquad \text{(A.4)}
$$

exists and is positive definite, a martingale CLT on $\xi_{nt} = n^{-1/2}s_t(\theta^0)$ (e.g. Corollary 3.1 of Hall and Heyde (1980)) implies that $n^{-1/2}\sum_{t=1}^n s_t(\theta^0) \to_d \mathcal{N}(0, \mathcal{B}_0)$. In what follows, we show that $\mathcal{B}_0$ in (A.4) exists and coincides with the expression for $\mathcal{B}_0$ in (A.1).

We start by computing the second conditional moments of $\mathbf{z}_t$ and $\mathbf{u}_t$ at θ^0:

$$\mathbb{E}_{\mathcal{F}_{t-1}}[\mathbf{u}_t(\theta^0)\mathbf{u}_t(\theta^0)'] = \mathbb{E}_{\mathcal{F}_{t-1}}(\mathbf{y}_t - \mu_t^0)(\mathbf{y}_t - \mu_t^0)' = \left[HG\mathbb{E}_{\mathcal{F}_{t-1}}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'G'H'\right]_{\theta=\theta^0} = [\Omega]_{\theta=\theta^0}$$

$$\mathbb{E}_{\mathcal{F}_{t-1}}[\mathbf{z}_t(\theta^0)\mathbf{z}_t(\theta^0)'] = \mathbb{E}_{\mathcal{F}_{t-1}}[vech(\mathbf{u}_t\mathbf{u}_t' - \Omega)(vech(\mathbf{u}_t\mathbf{u}_t' - \Omega))']_{\theta=\theta^0} =: \left[\mathcal{K}_u - vech\Omega(vech\Omega)'\right]_{\theta=\theta^0}$$

$$\mathbb{E}_{\mathcal{F}_{t-1}}[\mathbf{z}_t(\theta^0)\mathbf{u}_t(\theta^0)'] = \left[\mathbb{E}_{\mathcal{F}_{t-1}}(vech\mathbf{u}_t\mathbf{u}_t')\mathbf{u}_t' - vech\Omega\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{u}_t'\right]_{\theta=\theta^0} =: \mathcal{S}_u'$$

since $\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{u}_t = 0$ at θ^0, and $\mathcal{S}_u$ and $\mathcal{K}_u$ denote the multivariate skewness and kurtosis of $\mathbf{u}_t$ and are related to the skewness and kurtosis of the structural shocks $\mathcal{S}_\varepsilon$ and $\mathcal{K}_\varepsilon$ in the following way:

$$
\begin{aligned}
\mathcal{S}_u &= \mathbb{E}_{\mathcal{F}_{t-1}}\left\{\mathbf{u}_t(vech[\mathbf{u}_t\mathbf{u}_t'])'\right\} = \mathbb{E}_{\mathcal{F}_{t-1}}\left[HG\boldsymbol{\varepsilon}_t\left(D_r^+(HG\otimes HG)vec(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\right)'\right] \\
&= \left[HG\mathcal{S}_\varepsilon\left(D_m'(HG\otimes HG)'D_r^{+'}\right)\right] \qquad \text{(A.5)}
\end{aligned}
$$

$$\mathcal{K}_u = \mathbb{E}_{\mathcal{F}_{t-1}}\left[vech(\mathbf{u}_t\mathbf{u}_t')(vech(\mathbf{u}_t\mathbf{u}_t'))'\right] = \left[D_r^+(HG\otimes HG)D_m\mathcal{K}_\varepsilon D_m'(HG\otimes HG)'D_r^{+\prime}\right]$$

where $\mathcal{S}_\varepsilon = \mathbb{E}_{\mathcal{F}_{t-1}}\boldsymbol{\varepsilon}_t(vech[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'])'$ and $\mathcal{K}_\varepsilon = \mathbb{E}_{\mathcal{F}_{t-1}}vech(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')(vech(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'))'$ are the conditional third and fourth moment of the structural shocks $\boldsymbol{\varepsilon}_t$ defined in Assumption 5b. Next, we compute the conditional variance $\mathcal{B}_0$ of the score vector at θ^0 in (A.4). We define $\Phi = D_r'(\Omega^{-1}\otimes\Omega^{-1})D_r$ and

$C_1 = \frac{1}{2}\dot{\Omega}_1\Phi$, $C_{2,t} = \dot{\mu}_t\Omega^{-1}$ and $C_3 = \frac{1}{2}\dot{\Omega}_2\Phi$. We have

$$\begin{aligned}[\mathcal{B}_0]_{11} &= \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left[C_1\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{z}_t\mathbf{z}_t'C_1' + C_1\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{z}_t\mathbf{u}_t'C_{2,t}' + C_{2,t}\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{u}_t\mathbf{z}_t'C_1' + C_{2,t}\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{u}_t\mathbf{u}_t'C_{2,t}'\right] \\ &= C_1\mathcal{K}_uC_1' + C_1\mathcal{S}_u'\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}C_{2,t}' + \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}C_{2,t}\mathcal{S}_uC_1' + \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}C_{2,t}\Omega C_{2,t}'.\end{aligned}$$

Since $\dot{\mu}_t = \frac{\partial\mu(\theta_1,\mathbf{x}_{t-1})'}{\partial\theta_1} = \dot{C} + \dot{HF}\left(\mathbf{x}_{t-1}\otimes I_r\right)$ where $\dot{HF} = \frac{\partial vec(H(\theta_1)F(\theta_1))'}{\partial\theta_1}$ and $\dot{C} = \frac{\partial C(\theta_1)'}{\partial\theta_1}$, $\dot{\mu}_t$ is $\mathcal{F}_{t-1}$-measurable and we have $\frac{1}{n}\sum_{t=1}^{n}C_{2,t} = \frac{1}{n}\sum_{t=1}^{n}\dot{\mu}_t\Omega^{-1} \to_p \dot{C}\Omega^{-1}$ since $\frac{1}{n}\sum_{t=1}^{n}x_{t-1} \to_p 0$ by Assumption 4. Moreover,

$$\frac{1}{n}\sum_{t=1}^{n}\dot{\mu}_t\dot{\mu}_t' = \dot{C}\dot{C}' + \dot{HF}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\otimes I_r\right)(\dot{HF})' + o_p(1) \to_p \dot{C}\dot{C}' + \dot{HF}\left(V_X\otimes I_r\right)(\dot{HF})' =: V_\mu$$

where $V_X = \mathbb{E}\left[\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right]$ satisfying $vec\left(V_X\right) = \left(I_{s^2} - F\left(\theta^0\right)\otimes F\left(\theta^0\right)\right)^{-1}\left(G\left(\theta^0\right)\otimes G\left(\theta^0\right)\right)vec\Sigma\left(\theta^0\right)$.

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}C_{2,t}\Omega C_{2,t}' &= \frac{1}{n}\sum_{t=1}^{n}\dot{\mu}_t\Omega^{-1}\dot{\mu}_t' = \dot{C}\Omega^{-1}\dot{C}' + \left[\dot{HF}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\otimes\Omega^{-1}\right)(\dot{HF})'\right] + o_p(1) \\ &\to_p \dot{C}\Omega^{-1}\dot{C}' + \dot{HF}\left(V_X\otimes\Omega^{-1}\right)(\dot{HF})' =: V.\end{aligned}$$

Defining $L = \mathcal{S}_u'\Omega^{-1}\dot{C}'$, we obtain the 11 block of (A.4):

$$[\mathcal{B}_0]_{11} = C_1\mathcal{K}_uC_1' + C_1L + L'C_1' + V.$$

Likewise, the 12 block of (A.4) is given by

$$\begin{aligned}[\mathcal{B}_0]_{12} &= \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left[C_1\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{z}_t\mathbf{z}_t'C_3' + C_{2,t}\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{u}_t\mathbf{z}_t'C_3'\right] \\ &= C_1\mathcal{K}_uC_3' + \left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}C_{2,t}\right)\mathcal{S}_uC_3' = C_1\mathcal{K}_uC_3' + L'C_3'\end{aligned}$$

and, hence, $[\mathcal{B}_0]_{21} = C_3\mathcal{K}_uC_1' + C_3L$. Finally, the 22 block of (A.4) is given by

$$[\mathcal{B}_0]_{22} = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}C_3\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{z}_t\mathbf{z}_t'C_3' = C_3\mathcal{K}_uC_3'.$$

Letting $\mathcal{B}_0 = \left\{[\mathcal{B}_0]_{ij}, i,j\in\{1,2\}\right\}$ with blocks given by the above expressions establishes that the probability limit in (A.4) exists and is equal to the expression for $\mathcal{B}_0$ in (A.1).

It remains to show that $\mathcal{B}_0$ is positive definite. It is easy to see that $\mathcal{B}_0^{22} := \dot{\Omega}_2\Phi\mathcal{K}_u\Phi\dot{\Omega}_2' > 0$. To see this, recall that $\dot{\Omega}_2 = \frac{\partial(vech\Omega)'}{\partial\theta_2} = D_m'\left(HG\otimes HG\right)'D_r^+$ since $\frac{\partial(vech\Sigma)'}{\partial\theta_2} = I_{k_2}$ as $\theta_2 = vech\Sigma$, and so $rk(\dot{\Omega}_2) = k_2$ since $rk\left(HG\otimes HG\right) = r^2$ since $rk\left(HG\right) = r$ by Assumption 7 and hence $rk\dot{\Omega}_2 = r\left(r+1\right)/2 = rk\left(D_r^+\right)$ since $r\leq m$.

Since $\mathcal{B}_0^{22} > 0$, $\mathcal{B}_0$ will be positive definite if and only if the Schur complement $\mathcal{B}_0^{11}\big|\,\mathcal{B}_0^{22} > 0$:

$$\begin{aligned}\mathcal{B}_0^{11}\big|\,\mathcal{B}_0^{22} &= \mathcal{B}_0^{11} - \mathcal{B}_0^{12}\left(\mathcal{B}_0^{22}\right)^{-1}\mathcal{B}_0^{21} = \dot{\Omega}_1\Phi\mathcal{K}_u\Phi\dot{\Omega}_1'/4 + \dot{\Omega}_1\Phi L/2 + L'\Phi\dot{\Omega}_1'/2 + V \\ &\quad -(\dot{\Omega}_1\Phi\mathcal{K}_u\Phi\dot{\Omega}_2'/2 + L'\Phi\dot{\Omega}_2')\left(\dot{\Omega}_2\Phi\mathcal{K}_u\Phi\dot{\Omega}_2'\right)^{-1}(\dot{\Omega}_2\Phi\mathcal{K}_u\Phi\dot{\Omega}_1'/2 + \dot{\Omega}_2\Phi L) \\ &= V - L'\mathcal{K}_u^{-1}L = \dot{C}\Omega^{-1}\left(\Omega - \mathcal{S}_u\mathcal{K}_u^{-1}\mathcal{S}_u'\right)\Omega^{-1}\dot{C}' + \dot{HF}\left(V_X\otimes\Omega^{-1}\right)(\dot{HF})'.\end{aligned}$$

Both terms above are matrix quadratic with $(\Omega - \mathcal{S}_u\mathcal{K}_u^{-1}\mathcal{S}_u') > 0$ (by p.d. of $\mathbb{V}\left(\left[u_t', vech\left(u_tu_t'\right)'\right]'\right)$) and $(V_X\otimes\Omega^{-1}) \geq 0$ and hence both are p.s.d. It suffice to show that the second term[14] is p.d. The second term is a quadratic form, with $rk\left(V_X\otimes\Omega^{-1}\right) = mr$ since $rk\left(V_X\right) = m < \dim\left(x_t\right)$ due to G being in general not-square full column rank. We have that $rk(\dot{HF}) = k_1$ and hence $rk(\dot{HF}\left(V_X\otimes\Omega^{-1}\right)(\dot{HF})') = k_1$ since $k_1 < mr$ by Assumption 7. It follows that $\mathcal{B}_0$ is p.d.

[14] The first can be rank-deficient: it contains $\dot{C} = \frac{\partial C'}{\partial\theta_1}$ which in a typical model containing more structural parameters than observables $k_1 > r$ will have rank r.

Proof of Lemma 2. The second differential of $\ell_t(\theta)$ takes the form $d^2\ell_t(\theta) =$

$$-d\mu_t'\Omega^{-1}d\mu_t - \frac{1}{2}tr\left[\Omega^{-1}(d\Omega)\Omega^{-1}d\Omega\right] - \frac{1}{2}tr\left[\Omega^{-1}(d\Omega)\Omega^{-1}d\mu_t\mathbf{u}_t'\right]$$
$$-\frac{1}{2}tr\left[\Omega^{-1}(d\Omega)\Omega^{-1}\mathbf{u}_t d\mu_t'\right] + \frac{1}{2}tr\left[d\left(\Omega^{-1}\right)(d\Omega)\Omega^{-1}\mathbf{Z}_t\right] + \frac{1}{2}tr\left[\Omega^{-1}(d\Omega)d\left(\Omega^{-1}\right)\mathbf{Z}_t\right]$$
$$+\left[d\mu_t'd\left(\Omega^{-1}\right)\right]\mathbf{u}_t + \frac{1}{2}tr\left[\Omega^{-1}\left(d^2\Omega\right)\Omega^{-1}\mathbf{Z}_t\right] + \left[d^2\left(\mu_t'\right)\Omega^{-1}\right]\mathbf{u}_t. \tag{A.6}$$

Recall that, by (6), for $k \in \{0,1,2\}$

$$d^k\mu_t(\theta_1, \mathbf{x}_{t-1}) = d^kC(\theta_1) + d^k\left[H(\theta_1)F(\theta_1)\right]\mathbf{x}_{t-1} \tag{A.7}$$

with the convention $d^0 f_t = f_t$. The functions C, H and F are continuously differentiable over Θ and twice continuously differentiable with Lipschitz continuous second derivatives in a neighbourhood $N\left(\theta^0, \delta\right) = \left\{\theta \in \Theta : \left\|\theta - \theta^0\right\| < \delta\right\}$ for some $\delta > 0$. For each $\theta_1 \in N_\delta\left(\theta_1^0\right)$, $\left\|\mu_t(\theta_1) - \mu_t\left(\theta_1^0\right)\right\| \leq$
$\left\|C(\theta_1) - C\left(\theta_1^0\right)\right\| + \left\|F(\theta_1) - F\left(\theta_1^0\right)\right\|\left\|H(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\| + \left\|H(\theta_1) - H\left(\theta_1^0\right)\right\|\left\|F\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\|,$
and $\left\|d\mu_t(\theta_1) - d\mu_t\left(\theta_1^0\right)\right\| \leq$

$$\left\|dC(\theta_1) - dC\left(\theta_1^0\right)\right\| + \left\|dF(\theta_1) - dF\left(\theta_1^0\right)\right\|\left\|H(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\| + \left\|dH(\theta_1) - dH\left(\theta_1^0\right)\right\|\left\|F\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\|$$
$$+\left\|H(\theta_1) - H\left(\theta_1^0\right)\right\|\left\|dF\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\| + \left\|F(\theta_1) - F\left(\theta_1^0\right)\right\|\left\|dH(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\|$$

and, hence, $\left\|d^2\mu_t(\theta_1) - d^2\mu_t\left(\theta_1^0\right)\right\|$

$$\begin{aligned}
&\leq \left\|d^2C(\theta_1) - d^2C\left(\theta_1^0\right)\right\| + \left\|d^2H(\theta_1)F(\theta_1) - d^2H\left(\theta_1^0\right)F\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\| \\
&\quad +2\left\|dH(\theta_1)dF(\theta_1) - dH\left(\theta_1^0\right)dF\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\| + \left\|H(\theta_1)d^2F(\theta_1) - H\left(\theta_1^0\right)d^2F\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\| \\
&\leq \left\|d^2C(\theta_1) - d^2C\left(\theta_1^0\right)\right\| + \left\|d^2H(\theta_1) - d^2H\left(\theta_1^0\right)\right\|\left\|F\left(\theta_1^0\right)\right\|\left\|\mathbf{x}_{t-1}\right\| \\
&\quad +\left\|F(\theta_1) - F\left(\theta_1^0\right)\right\|d^2H(\theta_1)\left\|\mathbf{x}_{t-1}\right\| + 2\left\|dH(\theta_1) - dH\left(\theta_1^0\right)\right\|\left\|dF(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\| \\
&\quad +2\left\|dF(\theta_1) - dF\left(\theta_1^0\right)\right\|\left\|dH(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\| + \left\|d^2F(\theta_1) - d^2F\left(\theta_1^0\right)\right\|\left\|H(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\| \\
&\quad +\left\|H(\theta_1) - H\left(\theta_1^0\right)\right\|\left\|d^2F(\theta_1)\right\|\left\|\mathbf{x}_{t-1}\right\|.
\end{aligned}$$

Since the functions d^kC, d^kF and d^kH are Lipschitz continuous on $N_\delta\left(\theta_1^0\right)$ for $k \in \{0,1,2\}$, we conclude that there exists $c \in (0,\infty)$ such that

$$\left\|d^k\mu_t(\theta_1) - d^k\mu_t\left(\theta_1^0\right)\right\| \leq c\left\|\theta_1 - \theta_1^0\right\|\left(1 + \left\|\mathbf{x}_{t-1}\right\|\right), \quad k \in \{0,1,2\} \tag{A.8}$$

for all $\theta_1 \in N_\delta\left(\theta_1^0\right)$. Similarly, since $\Omega(\theta_1, \theta_2) = H(\theta_1)G(\theta_1)\Sigma(\theta_2)G(\theta_1)'H(\theta_1)'$ and finite products of bounded Lipschitz continuous functions are Lipschitz continuous, Assumption 6 on H, G implies that $d^k\Omega$, Ω^{-1} and $d\Omega^{-1}$ for $k \in \{0,1,2\}$ are Lipschitz continuous on $N_\delta\left(\theta_1^0\right)$. Also, for each $\theta \in N_\delta\left(\theta^0\right)$,

$$\left\|\mathbf{u}_t(\theta) - \mathbf{u}_t\left(\theta^0\right)\right\| \leq \left\|\mu_t(\theta_1) - \mu_t\left(\theta_1^0\right)\right\| \leq c\left\|\theta_1 - \theta_1^0\right\|\left(1 + \left\|\mathbf{x}_{t-1}\right\|\right) \tag{A.9}$$

by (A.8) and

$$\left\|\mathbf{Z}_t(\theta) - \mathbf{Z}_t\left(\theta^0\right)\right\| \leq c\left\|\theta - \theta^0\right\|\left(1 + \left\|\mathbf{x}_{t-1}\right\| + \left\|\mathbf{x}_{t-1}\right\|^2 + \left\|\mathbf{x}_{t-1}\right\|\left\|y_t\right\|\right) \tag{A.10}$$

because (A.7) and (A.8) imply $\left\|\mathbf{Z}_t(\theta) - \mathbf{Z}_t\left(\theta^0\right)\right\|$

$$\begin{aligned}
&\leq \left\|\mathbf{u}_t(\theta)\mathbf{u}_t(\theta)' - \mathbf{u}_t\left(\theta^0\right)\mathbf{u}_t\left(\theta^0\right)'\right\| + \left\|\Omega(\theta) - \Omega\left(\theta^0\right)\right\| \\
&\leq 2\left\|\mu_t(\theta_1) - \mu_t\left(\theta_1^0\right)\right\|\left\|y_t\right\| + \left\|\mu_t(\theta_1) - \mu_t\left(\theta_1^0\right)\right\|\left(\left\|\mu_t(\theta_1)\right\| + \left\|\mu_t\left(\theta_1^0\right)\right\|\right) + \left\|\Omega(\theta) - \Omega\left(\theta^0\right)\right\| \\
&\leq c\left\|\theta - \theta^0\right\|\left(1 + \left\|\mathbf{x}_{t-1}\right\| + \left\|\mathbf{x}_{t-1}\right\|^2 + \left\|\mathbf{x}_{t-1}\right\|\left\|y_t\right\|\right).
\end{aligned}$$

Since $\theta_n^* \in N_\delta\left(\theta^0\right)$ for all but finitely many n, applying the Lipschitz continuity properties in (A.8), (A.9), (A.10) and the Lipschitz continuity of $d^k\Omega$, Ω^{-1} and $d\Omega^{-1}$ to the expression for $d^2\ell_t(\theta)$ in (A.6), we obtain that $\frac{1}{n}\sum_{t=1}^n\left|d^2\ell_t\left(\theta_n^*\right) - d^2\ell_t\left(\theta^0\right)\right| \leq$

$$c\left\|\theta_n^* - \theta^0\right\|\left\{1 + \frac{1}{n}\sum_{t=1}^n\left(\left\|\mathbf{Z}_t\left(\theta_n^*\right)\right\| + \left\|d\mu_t\left(\theta_n^*\right)\right\|^2 + \left\|d^k\mu_t\left(\theta_n^*\right)\right\|\left\|\mathbf{u}_t\left(\theta_n^*\right)\right\|\right)\right\} \tag{A.11}$$

for $k \in \{1,2\}$. Since $\left\|\theta_n^* - \theta^0\right\| \to_p 0$,

$$\frac{1}{n}\sum_{t=1}^n\left|d^2\ell_t\left(\theta_n^*\right) - d^2\ell_t\left(\theta^0\right)\right| \to_p 0 \tag{A.12}$$

follows by showing that each of the sample means in (A.11) is $O_p(1)$. For the first, (A.10) gives

$$\frac{1}{n}\sum_{t=1}^{n}\|\mathbf{Z}_t(\theta_n^*)\| \leq \frac{1}{n}\sum_{t=1}^{n}\left\|\mathbf{Z}_t\left(\theta^0\right)\right\| + \frac{1}{n}\sum_{t=1}^{n}\left\|\mathbf{Z}_t(\theta_n^*) - \mathbf{Z}_t\left(\theta^0\right)\right\|$$

$$\leq \frac{1}{n}\sum_{t=1}^{n}\left\|\mathbf{Z}_t\left(\theta^0\right)\right\| + c\left\|\theta-\theta^0\right\|\frac{1}{n}\sum_{t=1}^{n}\left(1+\|\mathbf{x}_{t-1}\|+\|\mathbf{x}_{t-1}\|^2+\|\mathbf{x}_{t-1}\|\,\|y_t\|\right) = \frac{1}{n}\sum_{t=1}^{n}\left\|\mathbf{Z}_t\left(\theta^0\right)\right\| + o_p(1)$$

since $\max_{t\leq n}\|\mathbf{x}_t\|_{L_2}$ and $\max_{t\leq n}\|y_t\|_{L_2}$ are $O(1)$ implying that $n^{-1}\sum_{t=1}^{n}\|\mathbf{x}_{t-1}\|$, $n^{-1}\sum_{t=1}^{n}\|\mathbf{x}_{t-1}\|^2$ and $n^{-1}\sum_{t=1}^{n}\|\mathbf{x}_{t-1}\|\,\|y_t\|$ are all bounded in L_1 norm. Since $\left\|\mathbf{Z}_t\left(\theta^0\right)\right\| \leq \|\mathbf{u}_t\|^2 + \|\Omega\|$ and $\max_{t\leq n}\mathbb{E}\|\mathbf{u}_t\|^2 = O(1)$, $n^{-1}\sum_{t=1}^{n}\left\|\mathbf{Z}_t\left(\theta^0\right)\right\| = O_p(1)$ and $n^{-1}\sum_{t=1}^{n}\|\mathbf{Z}_t(\theta_n^*)\| = O_p(1)$ as required. For the third sample mean in (A.11), $\mathbf{u}_t(\theta) = y_t - \mu_t(\theta)$ implies that

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\|d^k\mu_t(\theta_n^*)\right\|\|\mathbf{u}_t(\theta_n^*)\| \leq \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\|d^k\mu_t(\theta_n^*)\right\|\|y_t\| + \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\|d^k\mu_t(\theta_n^*)\right\|\|\mu_t(\theta_n^*)\|$$

$$\leq \max_{t\leq n}\left\|d^k\mu_t(\theta_n^*)\right\|_{L_2}(\max_{t\leq n}\|y_t\|_{L_2} + \max_{t\leq n}\|\mu_t(\theta_n^*)\|_{L_2}) \leq C\max_{t\leq n}\|x_t\|_{L_2}(\max_{t\leq n}\|y_t\|_{L_2} + \max_{t\leq n}\|x_t\|_{L_2})$$

which is $O(1)$, where the second line uses CS inequality and the third uses the functional form of (6) and its derivatives which implies that $\left\|d^k\mu_t(\theta_n^*)\right\|_{L_2} \leq C\|x_t\|_{L_2}$ for all $\theta \in N_\delta\left(\theta^0\right)$. The last inequality for $k=1$ shows that the second sample mean in (A.11) is $O_p(1)$, showing (A.12).

Further, UI of $(\left\|\mathbf{u}_t\left(\theta^0\right)\right\|^2)_{t\geq 1}$, $(\left\|\mathbf{z}_t\left(\theta^0\right)\right\|^2)_{t\geq 1}$ and $(\|x_t\|^2)_{t\geq 1}$ (the first two proven in Lemma 1 and the last implied by the short memory linear process representation of x_t with innovations ε_t and UI of $(\|\varepsilon_t\|^2)_{t\geq 1}$) ensure the uniform integrability (UI) of the sequences $(\left\|d^k\mu_t\left(\theta^0\right)\right\|^2)_{t\geq 1}$ for $k \in \{1,2\}$ (since $\left\|d^k\mu_t\left(\theta^0\right)\right\| \leq C\|\mathbf{x}_{t-1}\|$ and $(\|x_t\|^2)_{t\geq 1}$ is UI) and of $(\left\|d^k\mu_t\left(\theta^0\right)\right\|\left\|\mathbf{u}_t\left(\theta^0\right)\right\|)_{t\geq 1}$ for $k\in\{1,2\}$ (since $\left\|d^k\mu_t\left(\theta^0\right)\right\|\left\|\mathbf{u}_t\left(\theta^0\right)\right\| \leq \left\|d^k\mu_t\left(\theta^0\right)\right\|^2 \vee \left\|\mathbf{u}_t\left(\theta^0\right)\right\|^2$ and $(\left\|d^k\mu_t\left(\theta^0\right)\right\|^2)_{t\geq 1}$ and $(\left\|\mathbf{u}_t\left(\theta^0\right)\right\|^2)_{t\geq 1}$ are UI sequences). The functional form of $d^2\ell_t(\theta)$ in (A.6) then implies the UI of the sequence $\left(d^2\ell_t\left(\theta^0\right)\right)_{t\geq 1}$ which, in turn, implies the LLN

$$n^{-1}\sum_{t=1}^{n}\left(d^2\ell_t\left(\theta^0\right) - \mathbb{E}_{\mathcal{F}_{t-1}}d^2\ell_t\left(\theta^0\right)\right) \to_{L_1} 0. \tag{A.13}$$

Combining (A.12) and (A.13) implies that

$$n^{-1}\sum_{t=1}^{n}\left(d^2\ell_t(\theta_n^*) - \mathbb{E}_{\mathcal{F}_{t-1}}d^2\ell_t\left(\theta^0\right)\right) \to_{L_1} 0 \tag{A.14}$$

for any θ_n^* satisfying $\left\|\theta_n^* - \theta^0\right\| \leq \left\|\hat{\theta}_n - \theta^0\right\|$. The approximation (A.2) of the Hessian sequence $(\mathcal{H}_t\left(\theta^0\right))_{t\geq 1}$ follows from (A.14) and the identification theorem between the second differential and the Hessian (Theorem 7 in Magnus and Neudecker (2007)).

7.3 Proof of Theorem 1

The QML estimator $\hat{\theta}_n$ solves $\frac{1}{n}\sum_{t=1}^{n}s_t(\hat{\theta}_n) = 0$ and the mean value theorem on $\frac{1}{n}\sum_{t=1}^{n}s_t(\theta^0)$ around $\hat{\theta}_n$ yields

$$\frac{1}{n}\sum_{t=1}^{n}s_t\left(\theta^0\right) = -\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_t(\theta_n^*)(\hat{\theta}_n - \theta^0) \tag{A.15}$$

where $\mathcal{H}_t(\theta) = \frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'}$ and θ_n^* is an intermediate point satisfying $\left\|\theta_n^* - \theta^0\right\| \leq \left\|\hat{\theta}_n - \theta^0\right\|$. Lemma 2 implies that $\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_t(\theta_n^*) \to_p \text{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\mathcal{F}_{t-1}}\mathcal{H}_t\left(\theta^0\right)$, provided that the probability limit exists. In what follows, we show that the probability limit exists and is given by

$$\mathcal{A}_0 := \text{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}_{\mathcal{F}_{t-1}}\mathcal{H}_t\left(\theta^0\right) = \begin{bmatrix} \frac{1}{2}\dot{\Omega}_1\Phi\dot{\Omega}_1' + V & \frac{1}{2}\dot{\Omega}_1\Phi\dot{\Omega}_2' \\ \frac{1}{2}\dot{\Omega}_2\Phi\dot{\Omega}_1' & \frac{1}{2}\dot{\Omega}_2\Phi\dot{\Omega}_2' \end{bmatrix}_{\theta=\theta^0} \tag{A.16}$$

where V and Φ in (A.16)) are defined in Lemma 1. Using the fact that $\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{z}_t\left(\theta^0\right)=0$ and $\mathbb{E}_{\mathcal{F}_{t-1}}\mathbf{u}_t\left(\theta^0\right)=0$, we obtain $\frac{1}{n}\sum_{t=1}^n\mathbb{E}_{\mathcal{F}_{t-1}}d^2\ell_t\left(\theta^0\right)$

$$= \frac{1}{n}\sum_{t=1}^n\mathbb{E}_{\mathcal{F}_{t-1}}\left[\frac{1}{2}tr\left[\Omega^{-1}\left(d\Omega\right)\Omega^{-1}d\mathbf{Z}_t\right]+d\mu_t'\Omega^{-1}d\mathbf{u}_t\right]$$

$$= -\frac{1}{2}\left(dvech\Omega\right)'\Phi dvech\Omega-\frac{1}{n}\sum_{t=1}^n\mathbb{E}_{\mathcal{F}_{t-1}}\left(d\mu_t'\Omega^{-1}d\mu_t\right)$$

where $\Phi=D_r'\left(\Omega^{-1}\otimes\Omega^{-1}\right)D_r$ and all terms are evaluated at θ^0. Hence, $\mathcal{A}_0$ is given by (A.16) since

$$\left[\frac{1}{n}\sum_{t=1}^n\dot{\mu}_t\Omega^{-1}\dot{\mu}_t'\right]_{\theta=\theta^0}\to_p\left[\dot{C}\Omega^{-1}\dot{C}'\right]_{\theta=\theta^0}+\left[\dot{H}F\left(V_X\otimes\Omega^{-1}\right)(\dot{H}F)'\right]_{\theta=\theta^0}=V.$$

This completes the proof of (A.16). We first show that $\mathcal{A}_0$ is p.d. under the maintained Assumptions and then compute its inverse. We have that $\mathcal{A}_0^{22}>0$. Therefore, $\mathcal{A}_0$ will be positive definite if and only if the Schur complement $\mathcal{A}_0^{11}|\,\mathcal{A}_0^{22}$ is p.d. We have that $\mathcal{A}_0^{11}|\,\mathcal{A}_0^{22}=\mathcal{A}_0^{11}-\mathcal{A}_0^{12}\left(\mathcal{A}_0^{22}\right)^{-1}\mathcal{A}_0^{21}$

$$=\frac{1}{2}\dot{\Omega}_1\Phi\dot{\Omega}_1'+V-\frac{1}{2}\dot{\Omega}_1\Phi\dot{\Omega}_2'\left(\dot{\Omega}_2\Phi\dot{\Omega}_2'\right)^{-1}\dot{\Omega}_2\Phi\dot{\Omega}_1'=\frac{1}{2}\dot{\Omega}_1\Phi\dot{\Omega}_1'+V-\frac{1}{2}\dot{\Omega}_1\Phi\dot{\Omega}_1'=V>0$$

as we already established that $V>0$ at θ^0 in Lemma 1. Computing the inverse of $\mathcal{A}_0$, we have

$$\mathcal{A}_0^{-1}=\begin{bmatrix} V^{-1} & -V^{-1}X_1'X_2\left(X_2'X_2\right)^{-1} \\ -\left(X_2'X_2\right)^{-1}X_2'X_1V^{-1} & 2\left(X_2'X_2\right)^{-1}+\left(X_2'X_2\right)^{-1}X_2'X_1V^{-1}X_1'X_2\left(X_2'X_2\right)^{-1}\end{bmatrix} \quad \text{(A.17)}$$

where $X_1=\Phi^{1/2}\dot{\Omega}_1'$ and $X_2=\Phi^{1/2}\dot{\Omega}_2'$, and the second line uses that $P_{X_2}=X_2\left(X_2'X_2\right)^{-1}X_2'=I_{r(r+1)/2}$, since $\dim(\dot{\Omega}_2)=\dim\Phi=r\left(r+1\right)/2$ and so X_2 is square and $M_{X_2}=I_{r(r+1)/2}-P_{X_2}=0$.

Applying Lemmata 1 and 2 to (A.15), we obtain that $\sqrt{n}(\hat{\theta}_n-\theta^0)\to_d\mathcal{N}\left(0,\mathcal{C}_0\right)$, where $\mathcal{C}_0=\mathcal{A}_0^{-1}\mathcal{B}_0\mathcal{A}_0^{-1}$ and the explicit formulae for $\mathcal{B}_0$, $\mathcal{A}_0$ and $\mathcal{A}_0^{-1}$ are given in (A.1), (A.16) and (A.17). It remains to derive an expression for the asymptotic sandwich-form variance, by combining the expressions for $\mathcal{B}_0$ and $\mathcal{A}_0^{-1}$ in (A.17) and (A.1):

$$\begin{aligned}\left[\mathcal{C}_0\right]_{11} &= V^{-1}+\frac{1}{4}V^{-1}X_1'M_{X_2}\Phi^{1/2}\mathcal{K}_u\Phi^{1/2}X_1V^{-1}-\frac{1}{4}V^{-1}X_1'M_{X_2}\Phi^{1/2}\mathcal{K}_u\Phi^{1/2}P_{X_2}X_1V^{-1}\\ &\quad+\frac{1}{2}V^{-1}X_1'M_{X_2}\Phi^{1/2}LV^{-1}+\frac{1}{2}V^{-1}L'\Phi^{1/2}M_{X_2}X_1V^{-1}=V^{-1}\end{aligned}$$

since $M_{X_2}=0$. Similarly,

$$\left[\mathcal{C}_0\right]_{12}=-V^{-1}X_1'\left(X_2'\right)^{-1}+V^{-1}L'\Phi^{1/2}\left(X_2'\right)^{-1}.$$

$$\left[\mathcal{C}_0\right]_{22}=\dot{\Omega}_2'^{-1}\left(\mathcal{K}_u-LV^{-1}\dot{\Omega}_1-\dot{\Omega}_1'V^{-1}L'+\dot{\Omega}_1'V^{-1}\dot{\Omega}_1\right)\dot{\Omega}_2^{-1}.$$

We conclude that the asymptotic variance $\mathcal{C}_0$ is given by

$$\mathcal{C}_0=\begin{bmatrix} V^{-1} & V^{-1}(\dot{C}\Omega^{-1}\mathcal{S}_u-\dot{\Omega}_1)\dot{\Omega}_2^{-1} \\ \dot{\Omega}_2'^{-1}(\mathcal{S}_u'\Omega^{-1}\dot{C}'-\dot{\Omega}_1')V^{-1} & \dot{\Omega}_2'^{-1}(\mathcal{K}_u-\mathcal{S}_u'\Omega^{-1}\dot{C}'V^{-1}\dot{\Omega}_1-\dot{\Omega}_1'V^{-1}\dot{C}\Omega^{-1}\mathcal{S}_u+\dot{\Omega}_1'V^{-1}\dot{\Omega}_1)\dot{\Omega}_2^{-1}\end{bmatrix}.$$

(A.18)

7.4 Additional Monte Carlo Results

Table 6: Prior distributions and DGP values

	σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
DGP θ^0	1.00	1.00	1.00	4.00	2.00	0.50	1.50	0.50	2.00	0.50	0.70	0.70
Prior pdf	$\mathcal{IG}$	$\mathcal{IG}$	$\mathcal{IG}$	$\mathcal{G}$	$\mathcal{G}$	$\mathcal{G}$	$\mathcal{G}$	$\mathcal{G}$	$\mathcal{G}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{B}$
Prior mean	1.00	1.00	1.00	4.00	2.00	0.50	1.50	0.50	2.00	0.50	0.70	0.70
Prior std	2.00	2.00	2.00	2.00	1.00	0.30	0.25	0.25	0.50	0.20	0.10	0.10

Table 7: Bias DGP I

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	-0.01	0.07	0.02	-0.13	-0.12	0.01	0.11	-0.14	0.01	-0.01	0.00	0.00
	MHG-DSGE	-0.04	0.03	0.01	-0.12	-0.11	0.00	0.19	-0.26	0.07	-0.02	0.00	0.00
n=500	G-DSGE	0.00	0.05	0.01	-0.10	-0.10	0.01	0.06	-0.06	0.02	0.00	0.00	0.00
	MHG-DSGE	-0.02	0.04	0.00	-0.11	-0.11	0.00	0.09	-0.11	0.05	-0.01	0.00	0.00
n=1000	G-DSGE	0.00	0.03	0.01	-0.05	-0.06	0.00	0.03	-0.04	0.03	0.00	0.00	0.00
	MHG-DSGE	-0.01	0.02	0.00	-0.05	-0.06	0.00	0.04	-0.05	0.04	0.00	0.00	0.00

RMSE DGP I

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	0.11	0.20	0.13	0.58	0.49	0.05	0.20	0.28	0.22	0.03	0.03	0.04
	MHG-DSGE	0.13	0.22	0.15	0.59	0.51	0.05	0.28	0.40	0.27	0.04	0.04	0.04
n=500	G-DSGE	0.07	0.17	0.09	0.43	0.42	0.04	0.13	0.16	0.24	0.02	0.02	0.03
	MHG-DSGE	0.08	0.19	0.09	0.44	0.44	0.04	0.18	0.23	0.28	0.02	0.02	0.03
n=1000	G-DSGE	0.05	0.14	0.07	0.33	0.34	0.03	0.08	0.10	0.23	0.01	0.02	0.02
	MHG-DSGE	0.05	0.16	0.07	0.33	0.34	0.03	0.10	0.12	0.26	0.01	0.02	0.02

Table 8: Bias DGP II

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	0.00	0.06	0.03	-0.12	-0.12	0.01	0.11	-0.14	0.00	-0.01	0.00	0.00
	MHG-DSGE	-0.05	-0.02	0.02	-0.11	-0.10	-0.01	0.23	-0.32	0.14	-0.02	0.00	0.00
n=500	G-DSGE	0.00	0.05	0.01	-0.10	-0.11	0.01	0.05	-0.06	0.01	0.00	0.00	0.00
	MHG-DSGE	-0.03	0.01	0.01	-0.08	-0.09	0.00	0.12	-0.16	0.10	-0.01	0.00	0.00
n=1000	G-DSGE	0.00	0.03	0.00	-0.06	-0.06	0.00	0.03	-0.03	0.01	0.00	0.00	0.00
	MHG-DSGE	-0.01	0.01	0.00	-0.06	-0.06	0.00	0.05	-0.07	0.07	0.00	0.00	0.00

RMSE DGP II

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	0.19	0.23	0.21	0.59	0.49	0.05	0.21	0.29	0.22	0.03	0.03	0.04
	MHG-DSGE	0.22	0.25	0.25	0.60	0.50	0.05	0.31	0.45	0.32	0.04	0.04	0.04
n=500	G-DSGE	0.12	0.19	0.13	0.42	0.41	0.04	0.13	0.15	0.24	0.02	0.02	0.03
	MHG-DSGE	0.12	0.25	0.14	0.44	0.43	0.04	0.22	0.29	0.31	0.02	0.02	0.03
n=1000	G-DSGE	0.09	0.16	0.09	0.33	0.34	0.03	0.08	0.09	0.23	0.01	0.02	0.02
	MHG-DSGE	0.09	0.17	0.09	0.34	0.35	0.03	0.13	0.16	0.28	0.01	0.02	0.02

Table 9: Bias DGP III

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	0.00	0.08	0.03	-0.13	-0.09	0.01	0.12	-0.14	-0.01	-0.01	0.00	0.00
	MHG-DSGE	-0.10	-0.11	0.03	-0.12	-0.05	-0.03	0.35	-0.53	0.37	-0.03	0.01	0.02
n=500	G-DSGE	0.00	0.05	0.01	-0.09	-0.08	0.01	0.06	-0.06	0.01	0.00	0.00	0.00
	MHG-DSGE	-0.06	-0.06	0.01	-0.08	-0.06	-0.02	0.23	-0.31	0.26	-0.01	0.00	0.01
n=1000	G-DSGE	0.00	0.04	0.01	-0.06	-0.06	0.00	0.05	-0.06	0.01	0.00	0.00	0.00
	MHG-DSGE	-0.05	-0.03	0.01	-0.05	-0.05	-0.01	0.14	-0.11	0.14	-0.01	0.00	0.00

RMSE DGP III

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	0.27	0.29	0.28	0.61	0.51	0.05	0.21	0.29	0.24	0.03	0.03	0.04
	MHG-DSGE	0.28	0.36	0.32	0.65	0.55	0.06	0.40	0.60	0.51	0.05	0.04	0.04
n=500	G-DSGE	0.18	0.22	0.19	0.43	0.42	0.04	0.13	0.16	0.25	0.02	0.02	0.03
	MHG-DSGE	0.19	0.26	0.20	0.47	0.45	0.04	0.32	0.43	0.43	0.03	0.02	0.03
n=1000	G-DSGE	0.13	0.19	0.14	0.34	0.35	0.03	0.12	0.15	0.23	0.01	0.02	0.02
	MHG-DSGE	0.14	0.21	0.14	0.37	0.37	0.04	0.22	0.23	0.35	0.02	0.02	0.02

Table 10: Bias DGP IV

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	-0.01	0.07	0.02	-0.13	-0.11	0.01	0.11	-0.14	-0.01	-0.01	0.00	0.00
	MHG-DSGE	-0.08	-0.05	0.01	-0.13	-0.09	-0.01	0.27	-0.39	0.21	-0.02	0.00	0.01
n=500	G-DSGE	0.00	0.05	0.01	-0.08	-0.08	0.01	0.05	-0.06	0.01	0.00	0.00	0.00
	MHG-DSGE	-0.04	-0.03	0.01	-0.07	-0.06	-0.01	0.17	-0.23	0.17	-0.01	0.00	0.00
n=1000	G-DSGE	-0.01	0.04	0.01	-0.06	-0.06	0.00	0.05	-0.05	0.01	0.00	0.00	0.00
	MHG-DSGE	-0.04	-0.01	0.01	-0.06	-0.06	-0.01	0.11	-0.22	0.14	-0.01	0.00	0.00

RMSE DGP IV

n=200	G-DSGE	0.23	0.26	0.24	0.58	0.50	0.05	0.21	0.29	0.22	0.03	0.03	0.04
	MHG-DSGE	0.23	0.29	0.26	0.61	0.52	0.06	0.35	0.51	0.37	0.04	0.04	0.04
n=500	G-DSGE	0.15	0.21	0.17	0.43	0.43	0.04	0.13	0.16	0.24	0.02	0.02	0.03
	MHG-DSGE	0.16	0.22	0.17	0.46	0.44	0.04	0.27	0.36	0.36	0.03	0.02	0.03
n=1000	G-DSGE	0.11	0.17	0.12	0.33	0.34	0.03	0.12	0.14	0.23	0.01	0.02	0.02
	MHG-DSGE	0.12	0.19	0.12	0.35	0.36	0.03	0.24	0.29	0.31	0.02	0.02	0.02

Table 11: Bias DGP V

		σ_r	σ_g	σ_a	π^*	r^*	κ	ψ_1	ψ_2	τ^{-1}	ρ_r	ρ_g	ρ_a
n=200	G-DSGE	-0.02	0.01	-0.01	-0.08	-0.10	0.01	0.03	-0.03	-0.01	-0.01	0.00	0.00
n=500	G-DSGE	-0.02	0.00	-0.02	-0.06	-0.07	0.01	0.03	-0.03	0.00	0.00	0.00	0.00
n=1000	G-DSGE	-0.01	0.01	-0.01	-0.05	-0.05	0.00	0.02	-0.03	0.01	0.00	0.00	0.00

RMSE DGP V

n=200	G-DSGE	0.37	0.29	0.39	0.49	0.46	0.05	0.10	0.12	0.29	0.03	0.04	0.04
n=500	G-DSGE	0.32	0.25	0.34	0.38	0.39	0.04	0.09	0.10	0.26	0.02	0.02	0.03
n=1000	G-DSGE	0.28	0.24	0.29	0.32	0.33	0.03	0.08	0.09	0.24	0.01	0.02	0.02

7.5 Additional results

7.5.1 Linearised Model

For completeness, we list below the linearised equations and refer the reader to the original Smets and Wouters (2007) paper for full derivation of the model's equations and steady states. The derivations and steady state expressions of the financial friction block can be found in Del Negro and Schorfheide (2013). The main difference with the specification of Del Negro and Schorfheide (2013) is that we do not impose a stochastic (unit root) trend in productivity, as in the original Smets and Wouters (2007) specification.

The resource constraint in the model is given by

$$y_t = (1 - g_y - i_y)c_t + ((\gamma - 1 - \delta)k_y)i_t + (R^k_* k_y)z_t + \varepsilon^g_t.$$

The consumption Euler equation is

$$c_t = \frac{(\lambda/\gamma)}{(1+\lambda/\gamma)}c_{t-1} + \frac{1}{(1+\lambda/\gamma)}\mathbb{E}_t c_{t+1} + \frac{(\sigma_c - 1)W^h_* L_*/C_*}{\sigma_c(1+\lambda/\gamma)}\mathbb{E}_t(l_t - l_{t+1}) - \frac{(1-\lambda/\gamma)}{(1+\lambda/\gamma)\sigma_c}(r_t - \mathbb{E}_t\pi_{t+1} + \varepsilon^b_t).$$

The investment Euler equation,

$$i_t = \frac{1}{1+\beta\gamma^{1-\sigma_c}}i_{t-1} + (1 - \frac{1}{1+\beta\gamma^{1-\sigma_c}})\mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi}q_t + \varepsilon^i_t.$$

The aggregate production function is $y_t = \phi(\alpha k^s_t + (1-\alpha)l_t + \varepsilon^a_t)$.

The relation between effectively rented capital and capital follows $k^s_t = k_{t-1} + z_t$, with degree of capital utilization given by $z_t = \frac{1-\psi}{\psi} r^k_t$.

The capital accumulation equation follows $k_t = \frac{1-\delta}{\gamma}k_{t-1} + (1 - \frac{1-\delta}{\gamma})i_t + (1 - \frac{1-\delta}{\gamma})((1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi)\varepsilon^i_t$.

The price mark-up is $\mu^p_t = \alpha(k^s_t - l_t) + \varepsilon^a_t - w_t$; the resulting new Keynesian Phillips curve is

$$\pi_t = \frac{\iota_p}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\pi_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\mathbb{E}_t\pi_{t+1} - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\left\{\frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_p)(1-\xi_p)}{\xi_p((\phi-1)\varepsilon_p+1)}\right\}\mu^p_t + \varepsilon^p_t.$$

The rental rate of capital is $r^k_t = -(k_t - l_t) + w_t$.

The wage block is characterised by: (i) a wage mark-up equation: $\mu_t^w = w_t - (\sigma_l l_t + \frac{1}{1-\lambda/\gamma}(c_t - \lambda/\gamma c_{t-1}))$ and (ii) a wage equation

$$\begin{aligned} w_t &= \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} w_{t-1} + (1 - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}})(\mathbb{E}_t w_{t+1} + \mathbb{E}_t \pi_{t+1}) - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}\pi_t \\ &+ \frac{\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}\pi_{t-1} - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}\left\{\frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{\xi_w((\phi_w - 1)\varepsilon_w + 1)}\right\}\mu_t^w + \varepsilon_t^w. \end{aligned}$$

The Taylor Rule is given by

$$r_t = \rho r_{t-1} + (1-\rho)\left\{r_\pi \pi_t + r_y(y_t - y_t^p)\right\} + r_{\Delta y}\left((y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)\right) + \varepsilon_t^r.$$

The financial friction block is characterised by three equations: (i) a corporate spread equation:

$$\mathbb{E}_t\left[\tilde{R}_{t+1}^k - r_t\right] = \frac{(1-\lambda/\gamma)}{(1+\lambda/\gamma)\sigma_c}\varepsilon_t^b + \varsigma_{sp,b}(q_t + \bar{k}_t - n_t) + \varepsilon_t^\omega,$$

(ii) an arbitrage condition $\tilde{R}_t^k - \pi_t = \frac{r_*^k}{r_*^k + (1-\delta)} r_t^k + \frac{(1-\delta)}{r_*^k + (1-\delta)} q_t - q_{t-1}$, and (iii) an entrepreneurs' net worth which follows

$$n_t = \varsigma_{n,R^K}(\tilde{R}_t^k - \pi_t) - \varsigma_{n,R}(r_{t-1} - \pi_t) + \varsigma_{n,q}(q_{t-1} + \bar{k}_{t-1}) + \varsigma_{n,n} n_{t-1} - \frac{\varsigma_{n,\omega}}{\varsigma_{sp,\omega}}\varepsilon_{t-1}^\omega.$$

The eight stochastic processes in the model are: (i) government spending: $\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \sigma_g \eta_t^g + \rho_{ga}\sigma_z \eta_t^a$, (ii) TFP: $\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \sigma_a \eta_t^a$, (iii) risk premium process: $\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \sigma_b \eta_t^b$, (iv) an investment-technology process: $\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \sigma_i \eta_t^i$, (v) monetary policy process: $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \sigma_r \eta_t^r$, (vi) a price mark-up process: $\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \sigma_p \eta_t^p + \mu_p \sigma_p \eta_{t-1}^p$, (vii) wage mark-up process: $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \sigma_w \eta_t^w + \mu_w \sigma_w \eta_{t-1}^w$, (viii) financial friction process: $\varepsilon_t^\omega = \rho_\omega \varepsilon_{t-1}^\omega + \sigma_\omega \eta_t^\omega$. The structural shocks given by η_t^j with volatility parameter σ_j for $j \in \{g, a, b, i, r, p, w, \omega\}$.

7.5.2 Measurement Equation and Data

The measurement equation is given and transformed as

$$Y_t = \begin{bmatrix} Y_t \\ C_t \\ I_t \\ W_t \\ H_t \\ \Pi_t \\ R_t \\ S_t \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \\ SP^* \end{bmatrix} + \begin{bmatrix} y_t - y_{t-1} \\ c_t - c_{t-1} \\ i_t - i_{t-1} \\ w_t - w_{t-1} \\ l_t \\ \pi_t \\ r_t \\ 100 * \mathbb{E}_t(\tilde{R}_{t+1}^k - r_t) \end{bmatrix}, Y_t = \begin{bmatrix} Y_t = 100 * \Delta\ln(GDP_t/POP_t) \\ C_t = 100 * \Delta\ln(CON_t/POP_t) \\ I_t = 100 * \Delta\ln((INV_t/CPI_t)/POP_t) \\ W_t = 100 * \Delta\ln(WAGE_t) \\ H_t = HOURS_t/POP_t - \bar{H} \\ \Pi_t = 100 * \Delta\ln(CPI_t) \\ R_t = 1/4 * FFR_t \\ S_t = 1/4 * (BAA_t - TR_t) \end{bmatrix}$$

The series used for the estimation are described in Table 10. Table 11 presents the first four sample moments of the observables Y_t. The prior distributions as well as posterior mean and 95% credible sets are described in Table 13. The number of draws for all models is 30,000, from which we drop the first 10,000. The scaling parameter for the MH has been adjusted in order to obtain rejection rates of 20%-30%.

Table 10: Data Description

Variable	Description	Source
GDP_t	GDP, Total, Constant Prices, AR, SA, USD, 2012 chnd prices	U.S. Bureau of Economic Analysis
CON_t	PCE, Total, Constant Prices, AR, SA, USD, 2012 chnd prices	U.S. Bureau of Economic Analysis
INV_t	Private Fixed Investment, Total, Current Prices, AR, SA, USD	U.S. Bureau of Economic Analysis
CPI_t	Consumer price index, AR, SA, Index 1984=100	U.S. Bureau of Economic Analysis
$WAGE_t$	Real hourly compensation, nonfarm business, index, SA, Index 2012=100	U.S. Bureau of Labor Statistics
$HOURS_t$	Hours worked all workers, AR, SA Index Q3 1969=100	U.S. Bureau of Labor Statistics
POP_t	Popultaion Total, all ages	U.S. Bureau of Economic Analysis
FFR_t	Federal Funds Effective Rate	Federal Reserve, U.S.
BAA_t	Moody's Baa-Rated Corporate Bond Yield	Reuters
TR_t	Constant Maturity Yields, 10 Year	Federal Reserve, U.S.

Table 11: Sample moments of observables

	Y_t	C_t	I_t	W_t	H_t	Π_t	R_t	S_t
Mean	0.48	1.37	0.29	0.29	0.00	0.94	1.22	0.51
Variance	1.25	1.64	22.34	0.85	4.11	0.58	0.87	0.04
Skewness	-2.03	-2.41	-0.58	1.26	-1.32	0.70	0.84	0.79
Kurtosis	34.49	42.63	5.34	11.90	24.98	5.68	3.81	5.52

Table 12: Fixed Parameters

	Parameter	Fixed at value
λ^w	Steady state mark up in labour market	1.50
ϵ^w	Curvature Kimball aggregator labour market	10.00
δ	Capital Depreciation rate	0.025
ϵ^p	Curvature Kimball aggregator goods market	10.00
g_y	Exogenous spending GDP ratio	0.18
$\overline{F}^*$	Steady state default probability	0.03
γ^*	Survival rate of entrepreneurs	0.99

Table 13: Prior and posterior distributions for the parameters.

Parameter		Prior Distribution			Posterior Distribution					
					G		MHG Ind		MHG Orth	
		pdf	Mean	St.Dev.	Mean	95% set	Mean	95% set	Mean	95% set
φ	Elasticity Capital Adj Cost	Normal	4	1.5	8.68	[7.67,9.69]	9.02	[4.68,12.66]	10.69	[5.82,13.58]
σ_c	Elasticity Int Substitution	Normal	1.5	0.3	1.23	[0.93,1.38]	1.41	[1.03,1.85]	1.41	[1.13,1.88]
λ	Habit Formation	Beta	0.7	0.1	0.78	[0.73,0.82]	0.72	[0.61,0.80]	0.72	[0.48,0.81]
ξ_w	Calvo Probability Labour	Beta	0.5	0.1	0.98	[0.97,0.99]	0.97	[0.95,0.99]	0.98	[0.97,0.99]
σ_l	Elasticity Labour Supply	Normal	2	0.75	3.03	[1.86,4.38]	2.69	[1.37,4.28]	2.23	[0.97,3.96]
ξ_p	Calvo Probability Goods	Beta	0.5	0.1	0.77	[0.87,0.64]	0.72	[0.64,0.92]	0.87	[0.81,0.93]
ι_w	Wage Indexation	Beta	0.5	0.15	0.98	[0.97,0.99]	0.93	[0.89,0.98]	0.80	[0.69,0.97]
ι_p	Price Indexation	Beta	0.5	0.15	0.90	[0.80,0.96]	0.90	[0.71,0.98]	0.38	[0.19,0.94]
ψ	Elasticity of Capital	Beta	0.5	0.2	0.80	[0.70,0.89]	0.83	[0.72,0.92]	0.86	[0.74,0.94]
Φ	Fixed Costs Producers	Normal	1.25	0.12	1.46	[1.15,1.63]	1.46	[1.09,1.55]	1.59	[1.10,1.92]
r_π	Inflation Coefficient	Normal	1.5	0.25	1.01	[1.00,1.02]	1.01	[1.00,1.05]	1.02	[1.00,1.09]
ρ	Interest Rate Smoothing	Beta	0.75	0.1	0.94	[0.92,0.96]	0.93	[0.90,0.96]	0.87	[0.83,0.93]
r_y	Output Gap Coefficient	Normal	0.12	0.05	0.02	[0.02,0.03]	0.03	[0.01,0.05]	0.01	[-0.01,0.02]
$r_{\Delta y}$	Coefficient ΔOutput Gap	Normal	0.12	0.05	0.06	[0.04,0.09]	0.06	[0.03,0.09]	0.04	[0.01,0.06]
$100(\beta^{-1}-1)$	Households' Discount Factor	Gamma	0.25	0.1	0.14	[0.05,0.29]	0.26	[0.08,0.43]	0.18	[0.06,0.31]
π^*	Steady State Inflation	Gamma	0.62	0.1	0.61	[0.43,0.81]	0.59	[0.39,0.80]	0.63	[0.43,0.84]
l^*	Steady State Hours	Normal	0	2	0.16	[-3.83,4.13]	-0.97	[-4.59,2.87]	-0.10	[-3.94,3.38]
γ^*	SS Quarterly Growth	Normal	0.4	0.1	0.26	[0.24,0.28]	0.27	[0.24,0.29]	0.26	[0.25,0.28]
α	Capital Share	Normal	0.3	0.05	0.07	[0.05,0.10]	0.09	[0.06,0.13]	0.10	[0.07,0.15]
SP^*	Steady State Spread	Gamma	2	0.3	0.33	[0.26,0.41]	0.36	[0.28,0.45]	0.43	[0.31,0.54]
$\zeta_{sp,b}$	Effect of spread on Tobin's Q	Beta	0.05	0.015	0.01	[0.01,0.01]	0.01	[0.01,0.01]	0.01	[0.01,0.02]
σ_a	St. Dev. TFP Shock	Uniform	0	5	1.99	[1.80,2.12]	1.90	[0.78,2.72]	1.93	[0.78,2.78]
σ_b	St. Dev. Risk Premium Shock	Uniform	0	5	0.001	[0.000,0.002]	0.006	[0.001,0.038]	0.003	[0.001,0.012]
σ_g	St. Dev. Spending Shock	Uniform	0	5	0.99	[0.90,1.09]	0.95	[0.77,1.16]	0.96	[0.78,1.18]
σ_i	St. Dev. Investment Shock	Uniform	0	5	0.53	[0.41,0.67]	1.09	[0.44,2.76]	1.71	[0.47,2.94]
σ_r	St. Dev. Monetary Policy Shock	Uniform	0	5	0.23	[0.21,0.26]	0.24	[0.17,0.31]	0.24	[0.17,0.37]
σ_p	St. Dev. Price Mark-Up Shock	Uniform	0	5	0.26	[0.24,0.29]	0.44	[0.23,1.70]	0.50	[0.32,1.48]
σ_w	St. Dev. Wage Mark-Up Shock	Uniform	0	5	0.51	[0.47,0.57]	0.55	[0.43,0.85]	0.67	[0.44,1.56]
σ_ω	St. Dev. Financial Friction Shock	Uniform	0	5	0.10	[0.09,0.11]	0.13	[0.08,0.24]	0.12	[0.08,0.16]
ρ_a	Persistence of TFP	Beta	0.5	0.2	0.99	[0.98,1.00]	0.98	[0.96,1.00]	0.98	[0.97,1.00]
ρ_b	Persistence of Risk Premium	Beta	0.5	0.2	0.99	[0.99,1.00]	0.99	[0.99,1.00]	0.99	[0.99,1.00]
ρ_g	Persistence of Spending	Beta	0.5	0.2	0.99	[0.99,1.00]	0.99	[0.99,1.00]	0.99	[0.99,1.00]
ρ_i	Persistence of Investment	Beta	0.3	0.2	0.85	[0.78,0.91]	0.68	[0.24,0.91]	0.39	[0.08,0.89]
ρ_r	Persistence of Monetary Policy	Beta	0.3	0.2	0.15	[0.06,0.24]	0.14	[0.02,0.27]	0.19	[0.07,0.34]
ρ_p	Persistence of Price Mark Up	Beta	0.3	0.2	0.99	[0.99,1.00]	0.99	[0.98,1.00]	0.91	[0.85,1.00]
ρ_w	Persistence of Wage Mark Up	Beta	0.5	0.2	0.85	[0.84,0.87]	0.83	[0.81,0.85]	0.84	[0.82,0.85]
ρ_ω	Persistence of Financial Friction	Beta	0.5	0.2	0.98	[0.96,0.99]	0.98	[0.95,1.00]	0.98	[0.95,1.00]
μ_p	MA Coefficient Price Mark Up	Beta	0.5	0.2	0.97	[0.96,0.99]	0.95	[0.92,0.99]	0.94	[0.90,0.99]
μ_w	MA Coefficient Wage Mark Up	Beta	0.5	0.2	0.89	[0.89,0.90]	0.90	[0.89,0.91]	0.89	[0.88,0.90]
ρ_{ga}	TFP Coefficient Spending	Normal	0.5	0.2	0.06	[0.02,0.11]	0.11	[0.06,0.21]	0.8	[0.04,0.14]