HIAS-E-128

Posterior Inferences on Incomplete Structural Models: The Minimal Econometric Interpretation

Takashi Kano

Hitotsubashi University

March 2023

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8668 http://hias.hit-u.ac.jp/

HIAS discussion papers can be downloaded without charge from:
https://hdl.handle.net/10086/27202
https://ideas.repec.org/s/hit/hiasdp.html

Posterior Inferences on Incomplete Structural Models: The Minimal Econometric Interpretation

Takashi Kano
Graduate School of Economics
Hitotsubashi University
Naka 2-1, Kunitachi, Tokyo
186-8601, JAPAN
Email: tkano@econ.hit-u.ac.jp

Current Draft: March 14, 2023

Abstract

The minimal econometric interpretation (MEI) of DSGE models provides a formal model evaluation and comparison of misspecified nonlinear dynamic stochastic general equilibrium (DSGE) models based on atheoretical reference models. The MEI approach recognizes DSGE models as incomplete econometric tools that provide only prior distributions of targeted population moments but have no implications for actual data and sample moments. This study, based on the MEI approach, develops a Bayesian posterior inference method. Prior distributions of targeted population moments simulated by the DSGE model restrict the hyperparameters of Dirichlet distributions. These are natural conjugate priors for multinomial distributions followed by corresponding posterior distributions estimated by the reference model. The Pólya marginal likelihood of the resulting restricted Dirichlet-multinomial model has a tractive approximated log-linear representation of the Jensen-Shannon divergence, which the proposed distribution-matching posterior inference uses as the limited information likelihood function. Monte Carlo experiments indicate that the MEI posterior sampler correctly infers calibrated structural parameters of an equilibrium asset pricing model and detects the true model with posterior odds ratios.

Key Words : *Bayesian posterior inference, Minimum econometric interpretation, Nonlinear DSGE model, Model misspecification, Equilibrium asset pricing model*

JEL Classification Number : C11, C52, E37

† I wish to thank Gianni Amisano, Toru Kitagawa, Jim Nason, and Toshi Watanabe and conference participants from the 4th Hitotsubashi Summer Institute (HSI 2019) and 16th International Conference on Computational and Financial Econometrics (CFE 2022) for their insightful discussions and helpful comments. I wish to express my gratitude for the grant-in-aid for scientific research from the Japan Society for the Promotion of Science (Nos. 24330060, 17H02542, 17H00985, and 20H00073) and financial support from the Hitotsubashi Institute for Advanced Study. I am solely responsible for any errors and misinterpretations in this study.

‡*Contact information:* Graduate School of Economics, Hitotsubashi University, Naka 2-1, Kunitachi-city, Tokyo, 186-8601, Japan. *Emails:* tkano@econ.hit-u.ac.jp.

"And now here is my secret, a very simple secret: It is only with the heart that one can see rightly; what is essential is invisible to the eye." (Saint-Exupéry, The Little Prince).

"Assuming the population moment is equal to the sample moment can be treacherous." (Geweke 2010).

1. Introduction

In his seminal study, Geweke (2010) develops the minimal econometric interpretation (MEI) of dynamic stochastic general equilibrium (DSGE) models, generalizing the Bayesian calibration approach of DeJong et al. (1996) as a crucial predecessor. What makes the MEI approach sharply distinct from the conventional likelihood-based inferences and standard moment-matching calibration exercises is its central premise. In the MEI approach, DSGE models only need to provide prior distributions of unobservable population moments. These are dimensions of the macroeconomic reality that DSGE modelers intend to mimic. The MEI approach treats DSGE models as incomplete econometric tools, assuming that they have no direct implications on either actual data or sample moments. This perspective of MEI on incomplete structural models enables formal implementation of Bayesian evaluation and comparison of nonlinear DSGE models even under potential model misspecification (e.g., stochastic singularity).[1]

However, population moments are unobservable. Hence, the MEI approach requires a densely parameterized atheoretical econometric model as a statistical reference to establish the posterior distributions of targeted population moments from observed data. Essentially, the reference model acts as a hypothetical bridge between DSGE models and actual data through posterior distributions of targeted population moments.

The MEI approach evaluates a DSGE model by measuring the degree of overlap between prior distributions of the targeted population moments implied by the DSGE model and the corresponding posterior distributions estimated by the reference model. The higher the degree of overlap, the better the fit of the DSGE model to the macroeconomic reality of interest. Hence, when conducting formal model evaluation and comparison of the DSGE model, the MEI approach utilizes

[1]Canova (2007), DeJong and Dave (2011), Del Negro and Schorfheide (2011), and Fernández-Villaverde et al. (2016) review the MEI approach as a Bayesian model comparison method of DSGE models subject to model misspecifciation.

the sampling variation of the targeted population moments implied by the statistical reference model and places much weaker weight on the corresponding sample moments and the underlying observed data.[2] This characteristic of the MEI approach formalizes the critical point made by Eichenbaum (1991), and reemphasized by Geweke (2010), with the epigram for empirical DSGE research that assuming that the population moment is equal to the sample moment can be treacherous; observed data and macroeconomic reality of interest are distinct objects.[3]

This study aims to develop a posterior inference method based on the MEI approach. The key idea behind the proposed method is to update the prior distributions of the structural parameters by matching the synthetic distributions of the targeted population moments implied by the DSGE model (hereafter, theoretical moment distribution) with the corresponding posterior distributions estimated using the statistical reference model (hereafter, empirical moment distribution) as closely as possible. Essentially, this study formalizes distribution-matching, not moment-matching posterior inference of the DSGE model.

While the MEI approach provides a formal, explicit, simple, and easy model evaluation, comparison, and development method, a practical concern for the intrinsic nature of MEI is that the resulting decision-making could be affected by implausibly assumed prior distributions of the structural parameters of the DSGE model. For instance, when facing a significant deviation in the synthetic distribution of a targeted population moment simulated by the DSGE model from the corresponding posterior distribution estimated with the reference model, an MEI user cannot distinguish *a priori* whether this failure stems from the misspecification of the proposed DSGE model or implausible prior distributions of its structural parameters. As a formal Bayesian decision-making scheme, the MEI approach requires a parameter-updating process to avoid this ambiguity.

The proposed distribution-matching posterior inference method starts by discretizing the em-

[2]Close cousins of the MEI approach are the prior-predictive analysis (Box 1980; Canova 1994; Lancaster 2004; Geweke 2005) and posterior-predictive analysis (Gelman et al. 2003; Faust and Gupta 2012). These specification tests measure how well prior and posterior distributions of the targeted sample moments (i.e., the checking functions) implied by the DSGE models contain their observed values with data, calculating prior and posterior predictive p-values. Geweke (2010), however, argues that the prior predictive check is Bayesian but the posterior predictive check is not (see also Lancaster 2004). Faust and Gupta (2012) discusses the practical advantage of the posterior predictive check for DSGE modelers and policy makers.

[3]The MEI approach is applied by Nason and Rogers (2006) and Kano and Nason (2014) to model evaluation and comparisons among small open-economy models and new Keynesian monetary models, respectively. Loria et al. (2022) also exploit the MEI approach to construct from calibrated DSGE models mixture priors for population parameters of statistical unobserved component models.

pirical and theoretical moment distributions over finite grid points. The discretized empirical distribution of a targeted population moment is assumed to follow a multinomial distribution conditional on a mass probability vector attached to finite grid points. This study assumes that the sole role of the DSGE model is to provide the prior distribution of the mass probability vector. In Bayesian literature, a Dirichlet distribution with hyper-parameters considered the concentration parameters acts as a natural conjugate prior for the multinomial distribution's mass probability vector. The DSGE model then imposes restrictions reflected in the discretized theoretical moment distribution on Dirichlet concentration parameters. Hence, based on the MEI approach, this study emphasizes that the DSGE model has no direct implications for the actual data and sample moments.

The resulting restricted Dirichlet distribution and empirical multinomial distribution jointly constructs the Dirichlet-multinomial (DM) model. Importantly, after integrating out the mass probability vector, the DM model yields the marginal likelihood of the empirical moment distribution, conditional on the theoretical counterpart with a Pólya distribution. This DM marginal likelihood plays a crucial role as a likelihood function in the proposed MEI posterior inference method. Moreover, as a mathematical proposition, this study shows that the logarithmic kernel of the resulting DM marginal likelihood has a tractable approximated representation of the Jensen-Shannon (JS) divergence between the empirical and theoretical moment distributions.

The shape of the JS divergence depends on the relative size of the simulated draws for the theoretical moment distribution compared to that of the empirical one. When the number of draws for the theoretical moment distribution is sufficiently large relative to that of the empirical one, JS divergence converges to the Kullback-Leibler (KL) divergence of the empirical moment distribution from the theoretical one. This establishes a quasi-likelihood function calculated from the multinomial distribution with mass probabilities restricted by theoretical moment distribution. However, when the number of draws for the theoretical distribution is minimum and equal to one, JS divergence converges to the predictive density function of the DM model, which closely traces out the empirical moment distribution. Hence, JS divergence offers a density kernel for a minimum distance (MD) estimator in this extreme case. This novel finding extends the central propositions Del Negro and Schorfheide (2004) established in their DSGE-VAR framework toward more flexibly selected population moments that only nonlinear DSGE models can target.

This study proposes a two-step Markov chain Monte Carlo (MCMC) procedure as the MEI posterior inference method. This treats JS divergence as the logarithm of the likelihood function (hereafter, JS likelihood). The first step applies a conventional MCMC sampler for the reduced-form parameters of the statistical reference model and constructs posterior distributions of the targeted population moments as empirical moment distributions. Given the resulting empirical moment distributions, the second step implements a posterior sampler induced by the JS likelihood to jointly simulate the theoretical moment distributions and posterior distributions of the structural parameters of the underlying DSGE model. The proposed MEI posterior sampler draws targeted population moments from the DSGE model and calculates the JS likelihood. Premultiplied by suitable prior distributions of structural parameters, the constructed JS likelihood then yields the kernel of the posterior probability density for a random-walk Metropolis-Hastings (RW-MH) algorithm for the theoretical moment distributions and posterior distributions of structural parameters. Moreover, the MEI posterior sampler enables the estimation of the marginal likelihood of the underlying data, conditional on DSGE and reference models, leading to formal Bayesian posterior model evaluation, comparison, and criticism.

To investigate how the proposed MEI posterior sampler works, this study conducts Monte Carlo experiments with Labadie's (1989) equilibrium consumption-based asset-pricing model, which Geweke (2010) exercises with the MEI approach. Labadie's (1989) model is notable for two reasons. First, this single-shocked model is subjected to the stochastic singularity problem for conventional full-information likelihood inference with multiple time-series data. Second, the model is nonlinear in deriving the equilibrium price of the risky asset. Hence, the linear inference method (e.g., the DSGE-VAR method) cannot be applied.

Given the calibrated values of structural parameters, Labadie's (1989) model simulates synthetic time-series data of the risk-free rate, equity premium, and consumption growth rate with a realistic sample length. Following Geweke (2010), this study adopts the first-order vector autoregression (VAR) as the statistical reference model. In the first step of the MCMC procedure, six targeted population moments from the standard normal-inverted Wishart model are simulated and empirical moment distributions are constructed. Under correctly specified informative prior distributions of structural parameters, the MEI posterior sampler in the second step of the MCMC procedure generates theoretical moment distributions tightly overlapping their empirical counter-

parts. Hence, resulting posterior distributions of structural parameters were close to corresponding calibrated values. Moreover, the Monte Carlo investigation shows that this successful performance of the MEI posterior sampler is robust against two different settings of prior distributions of structural parameters: uniform and misspecified prior distributions. The estimated posterior odds ratios indicate that the model with correctly specified informative prior distributions dominates models with the other two prior distributions.

This study emphasizes two novel aspects of the proposed MEI posterior sampler for literature on the Bayesian inference of DSGE models. First, based on the MEI approach, the MEI posterior sampler considers the sampling variations of the targeted unobservable population moments but weakens dependence on actual data and sample moments. Essentially, the MEI posterior sampler can be interpreted as a data augmentation technique using empirical moment distributions to avoid overfitting actual data and enhance the generalization ability of the inferred DSGE model.

Second, the MEI posterior sampler is a limited-information Bayesian inference method that does not rely on the full-information likelihood function of the underlying DSGE model.[4] In the extreme case with the minimum number of draws for theoretical moment distributions, the MEI posterior sampler nests the approximated Bayesian computation with the MCMC sampler (MCMC-ABC) of Marjoram et al. (2003) and Forneron and Ng (2018). Hence, it offers a Bayesian distribution-matching indirect inference method, which can easily be implemented for invisible essential population characteristics of macroeconomic reality using potentially misspecified nonlinear DSGE models.

The remainder of this paper is organized as follows. Section 2 introduces the proposed MEI posterior inference method. Section 3 discusses the construction and implementation of the Monte Carlo experiments. Section 4 reports the results of the Monte Carlo experiments. Finally, Section 5 concludes the study.

[4]Kim (2002) developed a Bayesian posterior inference method with a limited information likelihood from the objective function of an extremum estimator. This resembles the Laplace type estimator of Chernozhukov and Hong (2003). Inoue and Shintani (2018) showed that the resulting simulated quasi marginal likelihood can be used for formal Bayesian model evaluation and comparison. Fernández-Villaverde et al. (2016) provided a compact survey of Bayesian limited-information inference methods.

2. The MEI posterior inference

Let A and E denote the DSGE and empirical models characterized by the parameter vectors θ_A and θ_E, respectively. These two models can simulate the vector of unobservable population moments $\mathbb{M}$ which a researcher explains as I distinct macroeconomic facts: $\mathbb{M} = \mathbf{E}(\mathbb{Z})$, where $\mathbb{Z}$ is the vector of the corresponding sample moments, and $\mathbf{E}$ is the mathematical expectations operator. Let m_i denote the i-th element of the population moment vector $\mathbb{M} = [m_1, m_2, \cdots, m_I]$. The DSGE model A is parsimonious and can be nonlinear when simulating the population moment vector. Let $m_{A,i}$ denote the i-th element of the population moment vector $\mathbb{M}_A$, which is simulated by the DSGE model A equipped with the structural parameter vector θ_A (hereafter the "theoretical" moment). The theoretical moment $m_{A,i}$ is a deterministic nonlinear function of the structural parameter vector θ_A, $m_{A,i}(\theta_A)$. The empirical model E is atheoretical with a higher degree of freedom to simulate the reduced-form posterior distribution of the population moment vector. Hence, let $m_{E,i}$ represent the i-th element of the population moment vector $\mathbb{M}_E$, which is estimated by the empirical model E equipped with the parameter vector θ_E (hereafter the "empirical" moment).

2.1. The Dirichlet-multinomial model for discretized population moment distributions

The most important device this study relies on is the Dirichlet-multinomial (DM) model, which characterizes the probability distributions of the discretized population moment distributions.[5] Assume that the population moment m_i has the finite support $\mathbf{S}_i = [\underline{s}_i, \bar{s}_i]$. Suppose that support $\mathbf{S}_i$ is decomposed into K mutually exclusive subintervals $\mathbf{s}_{k,i}$ for $k = 1, \cdots K$. Let $\mathbf{p}_{k,i} \geq 0$ denote the mass probability of the event wherein the population moment m_i drops to the k-th subinterval $\mathbf{s}_{k,i}$. Let $\mathbf{p}_i \equiv [\mathbf{p}_{1,i}, \mathbf{p}_{2,i}, \cdots, \mathbf{p}_{K,i}]$ denote a vector comprising $\mathbf{p}_{k,i}$ that satisfies the regularity condition $\sum_{k=1}^{K} \mathbf{p}_{k,i} = 1$.

Suppose that the two models A and E generate collections of theoretical and empirical moments, $\mathbf{m}_{A,i} \equiv \{m_{A,i}^j\}_{j=1}^M$ and $\mathbf{m}_{E,i} \equiv \{m_{E,i}^j\}_{j=1}^N$, where $M \geq 1$ and $N \geq 1$ are numbers of elements in the theoretical and empirical moment collections, respectively. The former and latter collections are equivalent to the theoretical and empirical moment distributions. Let Θ_A denote the collection of M structural parameter vectors, $\Theta_A \equiv \{\theta_A^j\}_{j=1}^M$. Moreover, I define collections $\mathbf{M}_A$ and $\mathbf{M}_E$ by horizontally stacking *I* theoretical and empirical moment distributions,

[5] The Dirichlet-multinomial model is introduced by Gelman et al. (2003) and Lancaster (2004).

$\mathbf{M}_A \equiv [\mathbf{m}_{A,1}, \cdots, \mathbf{m}_{A,I}]$, and $\mathbf{M}_E \equiv [\mathbf{m}_{E,1}, \cdots, \mathbf{m}_{E,I}]$, respectively.

Consider the following discretization of the empirical moment distribution $\mathbf{m}_{E,i}$ over mass probability vector $\mathbf{p}_i$ for $i = 1, \cdots, I$. Let $n_{k,i} \geq 0$ for $k = 1, \cdots, K$ denote the number of draws of $m_{E,i}$ that drop into the k-th subinterval $\mathbf{s}_{k,i}$. The condition $\sum_{k=1}^{K} n_{k,i} = N$ is satisfied. The probability of $\mathbf{m}_{E,i}$ conditional on $\mathbf{p}_i$ is then characterized by a multinomial distribution with parameter $n_i = [n_{1,i}, n_{2,i}, \cdots, n_{K,i}]$:

$$p(\mathbf{m}_{E,i}|\mathbf{p}_i) = \frac{\Gamma(N+1)}{\prod_{k=1}^{K}\Gamma(n_{k,i}+1)} \prod_{k=1}^{K} (\mathbf{p}_{k,i})^{n_{k,i}}, \tag{1}$$

where $\Gamma(.)$ is the Gamma function.

Let $\alpha_{k,i} \geq 1$ represent one plus the number of draws of the theoretical moment $m_{A,i}$, which drops into the k-th subinterval $\mathbf{s}_{k,i}$. Condition $\sum_{k=1}^{K} \alpha_{k,i} = M + K$ is then satisfied. Consider a joint event wherein $\alpha_{k,i} - 1$ draws of $m_{A,i}$ drop into the k-th subinterval $\mathbf{s}_{k,i}$ with probability $\mathbf{p}_{k,i}$ for $k = 1, \cdots, K$ simultaneously. The joint probability of $\mathbf{p}_i$ conditional on $\mathbf{m}_{A,i}$ is then characterized by the Dirichlet distribution with concentration parameter vector $\alpha_i = [\alpha_{1,i}, \alpha_{2,i}, \cdots, \alpha_{K,i}]$:

$$p(\mathbf{p}_i|\mathbf{m}_{A,i}) = \frac{\Gamma(M+K)}{\prod_{k=1}^{K}\Gamma(\alpha_{k,i})} \prod_{k=1}^{K} (\mathbf{p}_{k,i})^{\alpha_{k,i}-1}. \tag{2}$$

The DSGE model A imposes theoretical restrictions on the concentration parameter vector α_i.

The restricted Dirichlet distribution (2) acts as a natural conjugate prior distribution for multinomial distribution (1). The Dirichlet-multinomial (DM) model consisting of eqs.(1) and (2) imply that the conditional distribution of $\mathbf{p}_i$ on $\mathbf{m}_{A,i}$ and $\mathbf{m}_{E,i}$ is characterized by the Dirichlet distribution with parameter $n_i + \alpha_i$:

$$p(\mathbf{p}_i|\mathbf{m}_{A,i}, \mathbf{m}_{E,i}) \propto p(\mathbf{p}_i|\mathbf{m}_{A,i}) p(\mathbf{m}_{E,i}|\mathbf{p}_i) = \frac{\Gamma(N+M+K)}{\prod_{k=1}^{K}\Gamma(n_{k,i}+\alpha_{k,i})} \prod_{k=1}^{K} (\mathbf{p}_{k,i})^{n_{k,i}+\alpha_{k,i}-1}. \tag{3}$$

where $n_{k,i} + \alpha_{k,i} - 1 \geq 1$ for $k = 1, \cdots, K$ and $\sum_{k=1}^{K}(n_{k,i} + \alpha_{k,i}) = N + M + K$.

The DM model's important property is its analytical representation of the marginal likelihood function. Integrating out $\mathbf{p}_i$ from the DM model provides the analytical form of the marginal likelihood function with Pólya distribution

$$p(\mathbf{m}_{E,i}|\mathbf{m}_{A,i}) = \int p(\mathbf{m}_{E,i}|\mathbf{p}_i)p(\mathbf{p}_i|\mathbf{m}_{A,i})\mathbf{dp}_i = \frac{\Gamma(N+1)\Gamma(M+K)}{\Gamma(N+M+K)} \prod_{k=1}^{K} \frac{\Gamma(n_{k,i}+\alpha_{k,i})}{\Gamma(n_{k,i}+1)\Gamma(\alpha_{k,i})}. \tag{4}$$

The posterior inference of the DSGE model A stems from the DM marginal likelihood function (4).

The difficulty is that the DM marginal likelihood (4) becomes infinite for large values of N and M as Gamma functions explode. This study shows that the above DM marginal likelihood (4) has a tractable approximation with the density kernel of the Jensen-Shannon (JS) divergence between empirical and theoretical distributions.[6] To establish the main proposition, the ratio of the total number of draws in the theoretical moment distribution to that of the empirical one is defined as $\lambda(\equiv (M+K)/N)$. For $k = 1, \cdots, K$, let $\zeta_{k,i}$ and $q_{k,i}$ denote frequencies wherein the empirical and theoretical moments drop into the k-th subinterval, $\zeta_{k,i} \equiv n_{k,i}/N$ and $q_{k,i} \equiv \alpha_{k,i}/(M+K)$, respectively. Particularly, $\zeta_{k,i}$ is equivalent to the maximum likelihood (ML) estimate of $\mathbf{p}_{k,i}$ in multinomial distribution (1). Let $\boldsymbol{\zeta}_i$ and $\mathbf{q}_i$ denote empirical and theoretical probability mass vectors $\boldsymbol{\zeta}_i \equiv [\zeta_{1,i}, \zeta_{2,i}, \cdots, \zeta_{k,i}]$ and $\mathbf{q}_i \equiv [q_{1,i}, q_{2,i}, \cdots, q_{K,i}]$, respectively. Appendix A establishes the following proposition:

Proposition 1. *The logarithm of the marginal likelihood function of the DM model (4) is approximated as follows:*

$$\ln p_\lambda(\mathbf{m}_{E,i}|\mathbf{m}_{A,i}) \approx \ln N - (1+\lambda)N\, D_{JS}(\boldsymbol{\zeta}_i \,||\, \mathbf{q}_i), \tag{5}$$

where $D_{JS}(\boldsymbol{\zeta}_i \,||\, \mathbf{q}_i)$ denotes the JS divergence between the empirical and theoretical distributions.

$$D_{JS}(\boldsymbol{\zeta}_i \,||\, \mathbf{q}_i) = \frac{1}{1+\lambda} \sum_{k=1}^{K} \zeta_{k,i} \left\{ \ln \zeta_{k,i} - \ln \left(\frac{1}{1+\lambda}\zeta_{k,i} + \frac{\lambda}{1+\lambda} q_{k,i} \right) \right\} + \frac{\lambda}{1+\lambda} \sum_{k=1}^{K} q_{k,i} \left\{ \ln q_{k,i} - \ln \left(\frac{1}{1+\lambda}\zeta_{k,i} + \frac{\lambda}{1+\lambda} q_{k,i} \right) \right\}$$

with the standard regularity condition $0 \times \ln 0 = 0$.

The approximated DM marginal likelihood (i.e., the JS likelihood function (5)) has three important properties. First, as the JS divergence is nonnegative, the JS likelihood (5) is maximized when

[6] The JS divergence is a generalized form of the Kullback-Leibler (KL) divergence with symmetry. Lin (1991) discusses the properties of JS divergence.

the empirical and theoretical distributions match exactly: $\lim_{\boldsymbol{\zeta}_i \to \mathbf{q}_i} \ln p_\lambda(\mathbf{m}_{E,i}|\mathbf{m}_{A,i}) \to \ln N$.[7] Second, when the number of draws in the theoretical moment distribution, M, becomes sufficiently large relative to that of the empirical one, N (i.e., $\lambda \to \infty$), it is the case that

$$\lim_{\lambda \to \infty} \ln p_\lambda(\mathbf{m}_{E,i}|\mathbf{m}_{A,i}) \to \ln N - N \sum_{k=1}^{K} \zeta_{k,i} \left(\ln \zeta_{k,i} - \ln q_{k,i} \right) \propto -N D_{KL}(\boldsymbol{\zeta}_i \,||\, \mathbf{q}_i). \tag{6}$$

Hence, the JS likelihood becomes proportional to the negative of the Kullback-Leibler (KL) divergence of $\boldsymbol{\zeta}_i$ from $\mathbf{q}_i$, denoted by $D_{KL}(\boldsymbol{\zeta}_i \,||\, \mathbf{q}_i)$. As the logarithm of multinomial distribution (1) is represented by the KL divergence,

$$\ln p(\mathbf{m}_{E,i}|\mathbf{p}_i) \approx -N \sum_{k=1}^{K} \zeta_{k,i} (\ln \zeta_{k,i} - \ln \mathbf{p}_{k,i}) = -N D_{KL}(\boldsymbol{\zeta}_i \,||\, \mathbf{p}_i),$$

Function (6) is obtained by replacing the unrestricted probability mass vector $\mathbf{p}_i$ with its theoretical counterpart $\mathbf{q}_i$. Hence, when $\lambda \to \infty$, the JS likelihood (5) approaches the quasi-likelihood function constructed by the multinomial distribution imposing the theoretical restriction of the DSGE model A. Notably, this second property corresponds to the first proposition established by Del Negro and Schorfheide (2004, Proposition 1).

Third, as Appendix B shows, when the number of draws in the theoretical moment distribution M takes the minimum value of 1, or equivalently, $\lambda \to (K+1)/N$, the JS likelihood (5) becomes

$$\lim_{\lambda \to \frac{K+1}{N}} \ln p_\lambda(\mathbf{m}_{E,i}|m_{A,i}) \to \sum_{k=1}^{K} \mathbf{I}[m_{A,i} \in \mathbf{s}_{k,i}] \ln \left(\frac{n_{k,i}+2}{N+K+1} \right), \tag{7}$$

where $\mathbf{I}[m_{A,i} \in \mathbf{s}_{k,i}]$ is the indicator function that takes the value of 1 if $m_{A,i}$ drops into the subinterval $\mathbf{s}_{k,i}$ and 0 otherwise. The term $\left(\frac{n_{k,i}+2}{N+K+1} \right)$ is the predictive density of the DM model (3), at which the theoretical moment that draws $m_{A,i}$ drops into the k-th subinterval $\mathbf{s}_{k,i}$.[8] For a sufficiently large N, predictive density is approximated well by the frequency of the empirical moment distribution $\zeta_{k,i}$. Hence, the theoretical moment that draws $m_{A,i}$ from the JS likelihood (7) closely

[7] If the empirical and theoretical distributions are uniform, $\zeta_k = q_k = 1/K$, the JS divergence becomes minimized and the JS likelihood (5) is maximized to $\ln N$.

[8] The DM model (3) implies that the predictive density of a new draw $m_{A,i}$ conditional on $\mathbf{m}_{E,i}$ and $\mathbf{m}_{A,i}$ is

$$p_\lambda(m_{A,i} \in \mathbf{s}_{k,i}|\mathbf{m}_{E,i}, \mathbf{m}_{A,i}) = \frac{1}{1+\lambda}\zeta_{k,i} + \frac{\lambda}{1+\lambda} q_{k,i}.$$

The predictive probability is the weighted average between the ML estimate $\zeta_{k,i}$ of the multinomial distribution (1) and the expectation of the Dirichlet distribution (2) with relative weights $1/(1+\lambda)$ and $\lambda/(1+\lambda)$. As $q_{k,i} > 0$, the JS likelihood function (7) is well-defined even when the ML estimate $\zeta_{k,i}$ is zero.

traces the shape of the empirical moment distribution $\mathbf{m}_{E,i}$. This third property echoes the second proposition of Del Negro and Schorfheide (2004, Proposition 2). In this case, the posterior estimate of the structural parameter θ_A can be interpreted as a minimum distance (MD) estimate with KL divergence. This is obtained by fitting the theoretical moment distribution to the empirical counterpart as closely as possible.[9]

2.2. The posterior joint distributions of Θ_A, $\mathbf{m}_{A,i}$, $\mathbf{m}_{E,i}$, *and* $\mathbf{p}_i$

This subsection introduces the joint posterior distribution of Θ_A, $\mathbf{m}_{A,i}$, $\mathbf{m}_{E.i}$, and $\mathbf{p}_i$ conditional on data $\mathbf{y}$ and the empirical and DSGE models E and A. In doing so, this subsection borrows Condition 4.1 and Proposition 4.2 of Geweke (2010), which are cornerstones of the MEI approach:

Condition 4.1 (Geweke 2010). *Conditional on the empirical and DSGE models,* E *and* A*,*

$$p(m_i, \theta_E, \mathbf{y}|A, E) = p(m_i|A)p(\theta_E|m_i, E)p(\mathbf{y}|\theta_E, m_i, E).$$

Condition 4.1 of Geweke (2010) implies that the sole function of the DSGE model A is to provide a prior distribution of population moment m_i. Empirical model E is incomplete as it provides no proper prior distribution of m_i: $p(m_i|E) \propto const.$ Then, the following proposition holds:

Proposition 4.2 (Geweke 2010). *Under Condition 4.1,*

$$\begin{aligned} p(\mathbf{y}|m_i, A, E) &= \frac{\int p(m_i, \theta_E, \mathbf{y}|A, E)d\theta_E}{p(m_i|A, E)}, \\ &= \frac{\int p(m_i|A)p(\theta_E|m_i, E)p(\mathbf{y}|\theta_E, m_i, E)d\theta_E}{p(m_i|A)}, \\ &= \int p(\theta_E|m_i, E)p(\mathbf{y}|\theta_E, m_i, E)d\theta_E, \\ &= p(\mathbf{y}|m_i, E). \end{aligned}$$

This proposition guarantees the core argument of the MEI approach; the DSGE model A has no direct implication on the data probability $\mathbf{y}$.

The joint distribution of Θ_A, $\mathbf{m}_{A,i}$, $\mathbf{m}_{E,i}$, $\mathbf{p}_i$, and $\mathbf{y}$ conditional on the two models E and A then are characterized as follows:

[9] Appendix B also shows that function (7) is proportional to the KL divergence $D_{KL}(\mathbf{q}_i \,||\, \frac{1}{1+\lambda}\boldsymbol{\zeta}_i + \frac{\lambda}{1+\lambda}\mathbf{q}_i)$.

$$
\begin{aligned}
p(\Theta_A, \mathbf{m}_{A,i}, \mathbf{m}_{E,i}, \mathbf{p}_i, \mathbf{y}|A, E) &= p(\Theta_A|A)\; p(\mathbf{m}_{A,i}|\Theta_A, A)\; p(\mathbf{m}_{E,i}, \mathbf{p}_i, \mathbf{y}|\mathbf{m}_{A,i}, A, E) \\
&= p(\Theta_A|A)\; p(\mathbf{p}_i|\mathbf{m}_{A,i}(\Theta_A))\; p(\mathbf{m}_{E,i}, \mathbf{y}|\mathbf{p}_i, A, E) \\
&= p(\Theta_A|A)\; p(\mathbf{p}_i|\mathbf{m}_{A,i}(\Theta_A))\; p(\mathbf{m}_{E,i}|\mathbf{p}_i)\; p(\mathbf{y}|\mathbf{m}_{E,i}, E). \qquad (8)
\end{aligned}
$$

Particularly, the first equality stems from Condition 4.1, and the sole role of DSGE model A is to generate the prior distribution of the population moment $p(\mathbf{m}_{A,i}|A) = \int p(\mathbf{m}_{A,i}|\Theta_A, A)p(\Theta_A|A)d\Theta_A$. As the theoretical moment $m_{A,i}$ is provided as a deterministic nonlinear function of the structural parameter θ_A, the distribution $p(\mathbf{m}_{A,i}|\Theta_A, A)$ degenerates to a mass point. The second term of the second equality results from the distribution of $\mathbf{p}_i$ conditional on $\mathbf{m}_{A,i}$ and is given by the Dirichlet distribution (2). The third term of the third equality stems from multinomial distribution (1). Finally, the fourth term of the third equality reflects Proposition 4.2, which guarantees equality $p(\mathbf{y}|\mathbf{m}_{E,i}, A, E) = p(\mathbf{y}|\mathbf{m}_{E,i}, E)$.[10]

From Bayes' law, $p(\mathbf{y}|\mathbf{m}_{E,i}, E) = \frac{p(\mathbf{m}_{E,i}|\mathbf{y},E)p(\mathbf{y}|E)}{p(\mathbf{m}_{E,i}|E)}$, by substituting this result into eq.(8) and dividing the result by the marginal data density $p(\mathbf{y}|A, E)$ yields the posterior joint distribution of Θ_A, $\mathbf{m}_{A,i}$, $\mathbf{m}_{E,i}$, and $\mathbf{p}_i$ on the data $\mathbf{y}$ and the two models A and E:

$$
\begin{aligned}
p(\Theta_A, \mathbf{m}_{A,i}, \mathbf{m}_{E,i}, \mathbf{p}_i|\mathbf{y}, A, E) &= \frac{p(\Theta_A, \mathbf{m}_{A,i}, \mathbf{m}_{E,i}, \mathbf{p}_i, \mathbf{y}|A, E)}{p(\mathbf{y}|A, E)} \\
&= p(\Theta_A|A)\; p(\mathbf{p}_i|\mathbf{m}_{A,i}(\Theta_A))\; p(\mathbf{m}_{E,i}|\mathbf{p}_i)\; \frac{p(\mathbf{m}_{E,i}|\mathbf{y}, E)\; p(\mathbf{y}|E)}{p(\mathbf{m}_{E,i}|E)p(\mathbf{y}|A, E)} \\
&\propto p(\Theta_A|A)\; p(\mathbf{p}_i|\mathbf{m}_{A,i}(\Theta_A))\; p(\mathbf{m}_{E,i}|\mathbf{p}_i)\; p(\mathbf{m}_{E,i}|\mathbf{y}, E),
\end{aligned}
$$

[10] To show that the joint distribution (8) indeed leads to the MEI approach, confirm that marginalizing out Θ_A, $\mathbf{m}_{A,i}$, $\mathbf{m}_{E,i}$, and $\mathbf{p}_i$ from the joint distribution (8) yields the marginal data density of $\mathbf{y}$ conditional on the two models A and E:

$$
\begin{aligned}
& p(\mathbf{y}|A, E) \\
&= \int_{\mathbf{m}_{E,i}} \int_{\mathbf{m}_{A,i}} \int_{\mathbf{p}_i} \int_{\Theta_A} p(\Theta_A|A)p(\mathbf{m}_{A,i}|\Theta_A, A)p(\mathbf{p}_i|\mathbf{m}_{A,i})p(\mathbf{m}_{E,i}|\mathbf{p}_i)p(\mathbf{y}|\mathbf{m}_{E,i}, E)d\Theta_A d\mathbf{p}_i d\mathbf{m}_{A,i} d\mathbf{m}_{E,i}, \\
&= \int_{\mathbf{m}_{E,i}} \int_{\mathbf{m}_{A,i}} \int_{\mathbf{p}_i} p(\mathbf{m}_{A,i}|A)p(\mathbf{p}_i|\mathbf{m}_{A,i})p(\mathbf{m}_{E,i}|\mathbf{p}_i)p(\mathbf{y}|\mathbf{m}_{E,i}, E)d\mathbf{p}_i d\mathbf{m}_{A,i} d\mathbf{m}_{E,i}, \\
&= \int_{\mathbf{m}_{E,i}} \int_{\mathbf{m}_{A,i}} p(\mathbf{m}_{A,i}|A)p(\mathbf{m}_{E,i}|\mathbf{m}_{A,i})p(\mathbf{y}|\mathbf{m}_{E,i}, E)d\mathbf{m}_{A,i} d\mathbf{m}_{E,i}, \\
&= \int_{\mathbf{m}_{E,i}} p(\mathbf{m}_{E,i}|A)p(\mathbf{y}|\mathbf{m}_{E,i}, E)d\mathbf{m}_{E,i}.
\end{aligned}
$$

As established in Proposition 4.3 by Geweke (2010), marginal data density $p(\mathbf{y}|A, E)$ is given by the convolution of two densities. This is the same marginal data density that the MEI approach uses for formal model comparison.

where the proportionality in the third line results from $p(\mathbf{m}_{E,i}|E) \propto const.$ If $\mathbf{p}_i$ is marginalized out, then the joint distribution becomes the following:

$$p(\Theta_A, \mathbf{m}_{A,i}, \mathbf{m}_{E,i}|\mathbf{y}, A, E) \propto p(\Theta_A|A)\, p_\lambda(\mathbf{m}_{E,i}|\mathbf{m}_{A,i}(\Theta_A))\, p(\mathbf{m}_{E,i}|\mathbf{y}, E), \tag{9}$$

where the second term of the RHS results from the DM marginal likelihood (4) and JS likelihood (5) in Proposition 1.

2.3. An MCMC procedure for the posterior joint distribution

This study proposes the posterior inference of the MEI approach using the following two-step MCMC procedure:

Two-step MCMC procedure

Step 1. Given $\mathbf{y}$ and the empirical model E, draw $\mathbf{m}_{E,i}$ from $p(\mathbf{m}_{E,i}|\mathbf{y}, E)$ for $i = 1, \cdots, I$.

Step 2. Given $\mathbf{M}_E$, draw Θ_A and $\mathbf{M}_{A,}$ from $p(\Theta_A|A)\Pi_{i=1}^{I} p_\lambda(\mathbf{m}_{E,i}|\mathbf{m}_{A,i}(\Theta_A))$.

Resulting draws of Θ_A and $\mathbf{M}_A$ construct the posterior draws from $p(\Theta_A, \mathbf{M}_A|\mathbf{y}, A, E)$.

The posterior draws of $\mathbf{m}_{E,i}$ in **Step 1** are implemented using the Gibbs sampling procedure with an atheoretical empirical model. Let $p(\mathbf{y}|\theta_E, E)$ denote the likelihood function of empirical model E with parameter θ_E given data $\mathbf{y}$. Given prior distribution $p(\theta_E|E)$, parameter θ_E is drawn from the posterior kernel

$$p(\theta_E|\mathbf{y}, E) \propto p(\mathbf{y}|\theta_E, E)p(\theta_E|E).$$

Given a draw of θ_E, the empirical moment $m_{E,i}$ is simulated as a deterministic nonlinear function, $m_{E,i}(\theta_E)$. This process is repeated N times to construct a draw of $\mathbf{m}_{E,i}$ from the conditional distribution $p(\mathbf{m}_{E,i}|\mathbf{y}, E)$.

To implement **Step 2**, this study focuses on the case with $M = 1$ because it is computationally infeasible to simultaneously draw a large number of theoretical moments $\mathbf{m}_{A,i}$ by solving the underlying nonlinear DSGE model for different values of the structural parameters Θ_A. With $M = 1$, collection Θ_A degenerates to a single vector θ_A. Hence, the conditional density becomes:

$$p_\lambda(\theta_A|\mathbf{M}_E) = p(\theta_A|A)\Pi_{i=1}^{I} p_\lambda(\mathbf{m}_{E,i}|m_{A,i}(\theta_A)).$$

Because there is no analytical representation of $p_\lambda(\theta_A|\mathbf{M}_E)$, **Step 2** adopts the following RW-MH

algorithm.

The RW-MH algorithm in Step 2

Given the initial draw θ_A^{old} and corresponding conditional probability $p(\theta_A^{old}|\mathbf{M}_E)$,

2(a). Draw a new candidate of the structural parameter θ_A^{new} from

$$\theta_A^{new} = \theta_A^{old} + \mathbf{v}, \quad \mathbf{v} \sim i.i.d.(\mathbf{0}, \tau\Omega)$$

where, $\mathbf{v}$ is a random variate drawn from an i.i.d. with the mean of zero and diagonal variance-covariance matrix Ω accompanied by the configuration parameter τ.

2(b). Calculate the conditional probability $p_\lambda(\theta_A^{new}|\mathbf{M}_E)$. Compute

$$r(\theta_A^{new}|\theta_A^{old}) = \min\left\{1, \frac{p_\lambda(\theta_A^{new}|\mathbf{M}_E)}{p_\lambda(\theta_A^{old}|\mathbf{M}_E)}\right\}.$$

2(c). Draw a uniform random variate $u \sim U[0,1]$. Accept $\theta_A = \theta_A^{new}$ if $r(\theta_A^{new}|\theta_A^{old}) \geq u$. Keep $\theta_A = \theta_A^{old}$ otherwise.

2(d). Set $\theta_A^{old} = \theta_A$. Repeat 2(a)-(d) many times.

The proposed RW-MH procedure can be interpreted as a generalization of the approximated Bayesian computation with the MCMC sampler (MCMC-ABC) developed by Marjoram et al. (2003) and extended by Forneron and Ng (2018). To understand this interpretation, suppose that the empirical moment distribution $\mathbf{m}_{E,i}$ degenerates to a single scaler value of the sample moment $m_{E,i}$. The JS likelihood (7) then becomes:

$$\begin{aligned}\lim_{\lambda\to K+1} \ln p_\lambda(m_{E,i}|m_{A,i}) &\to \sum_{k=1}^{K} \mathbf{I}[m_{A,i} \cap m_{E,i} \in \mathbf{s}_{k,i}] \ln\left(\frac{3}{K+2}\right) \\ &= \mathbf{I}\left[|m_{A,i} - m_{E,i}| \leq \frac{\mathbf{S}_i}{K}\right] \ln\left(\frac{3}{K+2}\right).\end{aligned}$$

This function takes the value of $\ln\left(\frac{3}{K+2}\right)$ if the simulated moment $m_{A,i}$ drops to the same subinterval $\mathbf{s}_{k,i}$, where the sample moment $m_{E,i}$ resides, and 0 otherwise. Essentially, only when the absolute distance between $m_{A,i}$ and $m_{E,i}$ is less than the width of the subinterval $\mathbf{S}_i/K$ does the JS likelihood take a positive value of $\frac{3}{K+2}$. Given this particular distance function, the proposed RW-MH procedure yields the same Markov kernel as the MCMC-ABC.

2.4. The marginal likelihood estimation and model comparison

In the MEI approach, the marginal likelihood of the DSGE model A is evaluated relative to that of the empirical model E. This relative marginal likelihood $\psi_\lambda(\mathbf{y}|A,E) \equiv p(\mathbf{y}|A,E)/p(\mathbf{y}|E)$ is approximated by the modified harmonic mean estimator by Geweke (1999): for J posterior draws,

$$\hat{\psi}_\lambda(\mathbf{y}|A,E) = \left[\frac{1}{J}\sum_{j=1}^{J}\frac{f(\theta_A^j)}{p_\lambda(\theta_A^j|\mathbf{M}_E)}\right]^{-1},$$

where θ_A^j is drawn from the posterior joint distribution (9) using the MCMC procedure. Following the recommendation of Geweke (1999), this study uses, for the function $f(\theta_A)$, a truncated normal approximation of the posterior distribution for θ_A.

The formal model comparison between the two nonlinear DSGE models A_1 and A_2 is implemented with the estimated relative marginal likelihoods $\hat{\psi}_\lambda(\mathbf{y}|A_1,E)$ and $\hat{\psi}_\lambda(\mathbf{y}|A_2,E)$. Given the prior model probabilities $p(A_1)$ and $p(A_2)$, the posterior odds ratio of A_1 versus A_2 conditional on the reference model E is

$$\frac{p(A_1)\hat{\psi}_\lambda(\mathbf{y}|A_1,E)}{p(A_2)\hat{\psi}_\lambda(\mathbf{y}|A_2,E)} = \frac{p(A_1)p(\mathbf{y}|A_1,E)}{p(A_2)p(\mathbf{y}|A_2,E)}.$$

When greater than one, the posterior odds ratio is in favor of the model A_1. The nonlinear DSGE model A_1 can then replicate the empirical moment distributions estimated by the reference model E better than the other nonlinear DSGE model A_2.

3. Monte Carlo experiments on the MEI posterior sampler with an equilibrium asset-pricing model

To show how the proposed MEI posterior inference method works practically, this study conducts Monte Carlo experiments with an equilibrium consumption-based asset-pricing model by Labadie (1989) as a simple example. Generally, an equilibrium asset-pricing model prices a risky asset based on the conditional covariance between the stochastic discount factor and future risky returns. The implied multiplicity makes it infeasible for a linearized model to price a risky asset in equilibrium. As a result, drawing a Bayesian posterior inference of an equilibrium asset-pricing model conditional on time-series data of asset prices becomes quite difficult. This section shows that the proposed MEI posterior sampler easily draws a posterior inference of a nonlinear

equilibrium asset-pricing model, as in Labadie (1989).

The population moment vector $\mathbb{M}$ investigated in this section is a six-dimensional column vector. This includes the population means of the risk-free rate $\mathbf{E}(r_t^f)$, equity premium $\mathbf{E}(ep_t)$, and consumption growth rate $\mathbf{E}(\ln \kappa_t)$, the population variances of the equity premium $\mathbf{V}(ep_t)$ and consumption growth rate $\mathbf{V}(\ln \kappa_t)$, and the population first-order autocorrelation of the consumption growth rate $\mathbf{Cor}(\ln \kappa_t)$;

$$\mathbb{M} \equiv [m_i] = [\mathbf{E}(r_t^f), \mathbf{E}(ep_t), \mathbf{V}(ep_t), \mathbf{E}(\ln \kappa_t), \mathbf{V}(\ln \kappa_t), \mathbf{Cor}(\ln \kappa_t)]'.$$

Geweke (2010) focuses on the population means of the risk-free rate and equity premium. This is because the standard equilibrium asset pricing models, as in Mehra and Prescott (1985) and Labadie (1989), are not intended to explain higher-order population moments of two asset prices. Additionally, this study includes the population variance of the equity premium and the population first and second moments related to the consumption growth rate to more precisely identify the consumption growth rate process.

3.1. Labadie's (1989) model and the true distributions of population moments

Labadie's (1989) model is a continuous state-space generalization of Mehra and Prescott's (1985) equilibrium consumption-based asset-pricing model. While Labadie (1989) investigates the role of inflation tax in the equity premium puzzle, the Monte Carlo exercises focus only on the implications of the model for the real rates of return, following Geweke (2010).[11]

A representative investor maximizes the lifetime utility function

$$\mathbf{E}_t \sum_{i=0}^{\infty} \beta^i \frac{c_{t+i}^{1-\gamma} - 1}{1-\gamma}, \quad 0 < \beta < 1, \quad \gamma > 0, \tag{10}$$

where β, γ, and $\mathbf{E}_t$ are the subjective discount factor, degree of relative risk aversion, and mathematical conditional expectation operator, respectively. The objective function (10) is maximized, subject to the budget constraint as follows:

$$q_t z_{t+1} + p_t b_{t+1} + c_t \leq (q_t + e_t) z_t + b_t,$$

where c_t, e_t, q_t, p_t, z_t, and b_t are consumption, stochastic endowment, price of the risky equity,

[11] Geweke (2010) investigates four distinct consumption-based equilibrium asset-pricing models for the real rates of return of risk-free and risky assets.

price of the risk-free bond, amount of risky equity, and amount of risk-free bonds, respectively. The growth rate of endowment $\ln \kappa_t$ follows the AR(1) process:

$$\ln \kappa_t \equiv \ln e_t / e_{t-1} = \delta_0 + \delta_1 \ln \kappa_{t-1} + \upsilon_t, \quad \upsilon_t \sim i.i.d. N(0, \sigma_e^2), \tag{11}$$

where δ_0, δ_1, and υ_t are the unconditional mean and AR root of the endowment growth rate and endowment growth shock, respectively. Endowment growth rate shock υ_t is assumed to follow a normal distribution with a mean of zero and variance of σ_e^2. Thus, κ_t follows a lognormal process. Hence, the model contains five structural parameters: $\theta_A = [\beta, \gamma, \mu, \delta_0, \delta_1, \sigma_e^2]$.

In equilibrium, consumption should be equal to the endowment $c_t = e_t$. Imposing this market-clearing condition on the first-order necessary conditions (FONCs) for equity holdings and risk-free bond holdings gives the following:

$$q_t e_t^{-\gamma} = \beta \mathbf{E}_t e_{t+1}^{-\gamma} (q_{t+1} + e_{t+1}), \tag{12}$$

and

$$p_t e_t^{-\gamma} = \beta \mathbf{E}_t e_{t+1}^{-\gamma}. \tag{13}$$

If a new variable h_t is defined by $q_t e_t^{-\gamma}$, then the FONC for equity holdings (12) can be rewritten as follows:

$$h_t = \beta \mathbf{E}_t (h_{t+1} + e_{t+1}^{1-\gamma}). \tag{14}$$

Eq.(14) is a linear rational expectation model with respect to h_t. Solving eq.(14) for h_t by forward iterations yields the fundamental solution:

$$h_t = \sum_{i=1}^{\infty} \beta^i \mathbf{E}_t e_{t+i}^{\rho}, \tag{15}$$

where $\rho \equiv 1 - \gamma$.

Under the assumption of log-normality of the endowment growth rate, Labadie (1989) proves that if the long-run average growth rate of the marginal utility times the real dividend e_t^{ρ} is of exponential order less than β^{-1}, the variable h_t in eq.(15) can be solved with respect to the equity price q_t as follows:

$$q_t = e_t \sum_{j=1}^{\infty} A_j \kappa_t^{a_j}, \tag{16}$$

where $A_1 = \beta \exp[\rho(\delta_0 + 0.5\rho\sigma^2)]$ and $a_1 = \rho\delta_1$, and for $j > 1$,

$$A_{j+1} = A_j \beta \exp[(a_j + \rho)(\delta_0 + 0.5(a_j + \rho)\sigma^2)], \quad \text{and} \quad a_{j+1} = (a_j + \rho)\delta_1.$$

Labadie (1989) provides an algorithm to find fixed points A^*, a^*, and $H_t = A^*\kappa_t^{a^*}$ that satisfy $q_t = e_t H_t$ and solves the nonlinear equation (16). The real rate of equity return r_t^q is then given by the following:

$$1 + r_t^q = \frac{q_t + e_t}{q_{t-1}} = \kappa_t \frac{H_t + 1}{H_{t-1}} = \kappa_t \frac{A^*\kappa_t^{a^*} + 1}{A^*\kappa_{t-1}^{a^*}}. \tag{17}$$

From eq.(13), the risk-free rate r_t^f is as follows:

$$\frac{1}{1 + r_t^f} = p_t = \beta \mathbf{E}_t \kappa_{t+1}^{-\gamma} = \beta \exp(-\gamma\delta_0 + 0.5\gamma^2\sigma^2)\kappa_t^{-\gamma\delta_1}. \tag{18}$$

Equity premium ep_t is defined by the difference between the two rates of returns r_t^q and r_t^f: $ep_t \equiv r_t^q - r_t^f$. The rate of equity return (17) and risk-free rate (18) are nonlinear functions of the structural parameters θ_A and the state variable κ_t.

To construct the true model denoted by R, this study follows Labadie's (1989) calibration. The true value of β is set to 0.980, that of γ is 2.000, that of δ_0 is 0.017, that of δ_1 is 0.180, and that of σ_e^2 is 0.003. Four true moments $\mathbf{E}(r_t^f|R)$, $\mathbf{E}(\ln\kappa_t|R)$, $\mathbf{V}(\ln\kappa_t|R)$, and $\mathbf{Cor}(\ln\kappa_t|R)$ can be calculated using analytical values.

$$\mathbf{E}(r_t^f|R) = \beta^{-1}\exp\left(\frac{\gamma\delta_0}{1-\delta_1} + \frac{\gamma^2\sigma_e^2}{1-\delta_1^2}(\delta_1^2 - 0.5)\right) - 1,$$
$$\mathbf{E}(\ln\kappa_t|R) = \frac{\delta_0}{1-\delta_1}, \quad \mathbf{V}(\ln\kappa_t|R) = \frac{\sigma_e^2}{1-\delta_1^2}, \quad \text{and} \quad \mathbf{Cor}(\ln\kappa_t|R) = \delta_1.$$

The true population mean and variance of the equity premium, $\mathbf{E}(ep_t|R)$ and $\mathbf{V}(ep_t|R)$, have no analytical representations. Given the corresponding initial state variable $\ln\kappa_0 = \delta_0/(1-\delta_1)$, eqs.(11), (17), and (18) generate a synthetic time series of ep_t for $T_{true} = 1,000$ quarter-periods. Then, the true population moments $\mathbf{E}(ep_t|R)$ and $\mathbf{V}(ep_t|R)$ are approximated as the time-series mean and variance of the synthetic data of ep_t: $\mathbf{E}(ep_t|R) = T_{true}^{-1}\sum_{t=1}^{T_{true}} ep_t$ and $\mathbf{V}(ep_t|R) = T_{true}^{-1}\sum_{t=1}^{T_{true}}(ep_t - \mathbf{E}(ep_t|R))^2$. The true moment vector $\mathbb{M}_R$ is then constructed as follows:

$$\mathbb{M}_R \equiv 100 \times [\mathbf{E}(r_t^f|R), \mathbf{E}(ep_t|R), \mathbf{V}(ep_t|R), \mathbf{E}(\ln\kappa_t|R), \mathbf{V}(\ln\kappa_t|R), \mathbf{Cor}(\ln\kappa_t|R)].$$

The second column of Table 1 displays the true population moments $\mathbb{M}_R$. For example, the true model yields 5.725 % of a mean risk-free rate $\mathbf{E}(r_t^f|R)$, 0.493 % of the mean equity premium $\mathbf{E}(ep_t|R)$, and 27.5% of the variance in the equity premium $\mathbf{V}(ep_t|R)$. Particularly, the small mean equity premium of 0.493 % implies that Labadie's (1989) model suffers from an equity premium

puzzle.

3.2. Posterior inference of the empirical model E: ***Step 1***

Step 1 of the MEI posterior inference method employs empirical model E to infer the empirical moment distributions from the actual data. Following Geweke (2010), this study utilizes first-order vector autoregression (VAR) as the empirical model E to simulate the posterior distribution of the empirical moment vector $\mathbf{m}_E$. Information set $y_t = [r_t^f, ep_t, \ln \kappa_t]'$ follows a trivariate first-order VAR(1)

$$y_t - \mu = F(y_{t-1} - \mu) + u_t, \quad u_t \sim i.i.d.N(\mathbf{0}, \Sigma), \tag{19}$$

where μ is the 3×1 column vector of the unconditional mean of y_t, F is the 3×3 coefficient matrix, and u_t is the 3×1 reduced-form disturbance vector that follows the multivariate normal distribution with a mean of $\mathbf{0}$ and the variance-covariance matrix Σ. The empirical moment vector $\mathbb{M}_E$ is a linear mapping of VAR parameters μ, F, and Σ;

$$\begin{aligned}\mathbb{M}_E &= 100 \times [\mathbf{E}(r_t^f|E), \mathbf{E}(ep_t|E), \mathbf{V}(ep_t|E), \mathbf{E}(\ln \kappa_t|E), \mathbf{V}(\ln \kappa_t|E), \mathbf{Cor}(\ln \kappa_t|E)] \\ &= 100 \times [\mu_1, \mu_2, \sigma_{ep}^2, \mu_3, \sigma_{\ln \kappa}^2, \rho_{\ln \kappa}],\end{aligned}$$

where μ_i for $i = 1, 2, 3$ is the i-th element of the constant vector μ, and σ_{ep}^2 and $\sigma_{\ln \kappa}^2$ are the (2,2)th and (3,3)th elements of the variance-covariance matrix Σ_y, which is given as the solution of the discrete Lyapunov equation $\Sigma_y = F\Sigma_y F' + \Sigma$. Finally, $\rho_{\ln \kappa}$ is the (3,3)th element of the first-order autocorrelation matrix from the VAR (19).

This study implements the draws of the VAR parameters from the posterior distribution $p(\theta_E|\mathbf{y}, E)$ using the Gibbs sampling procedure for the standard Normal-inverted Wishart model. The prior joint distributions of μ, F, and Σ are set to be improper. The resulting posterior distribution of Σ follows an inverted Wishart distribution. Conditional on a draw of Σ, the posterior joint distribution of μ and F is a multivariate normal. The realistic sample length is set to $T_E = 200$ at a quarterly frequency. From synthetic data with 10,000 sample lengths simulated from the true model, the last 200 data points are extracted to construct data $\mathbf{y}$. The Gibbs sampling procedure for the Normal-inverted Wishart model simulates posterior distributions of θ_E and $\mathbf{m}_E$ conditional on the synthetic data $\mathbf{y}$. The number of Gibbs sampling draws N is set to 30,000.

Figure 1 shows the empirical moment distribution $\mathbf{M}_E$. Each window plots the kernel-smoothed densities of the posterior distribution of an empirical moment and the corresponding true

moment as solid gray and vertical dashed blue lines, respectively. The upper-left, upper-middle, upper-right, lower-left, lower-middle, and lower-right windows correspond to $\mathbf{E}(r_t^f|E)$, $\mathbf{E}(ep_t|E)$, $\mathbf{V}(ep_t|E)$, $\mathbf{V}(\ln\kappa_t|E)$, $\mathbf{V}(\ln\kappa_t|E)$, and $\mathbf{Cor}(\ln\kappa_t|E)$, respectively. The third and fourth columns of Table 2 report the posterior means and standard deviations of the empirical population moments, respectively.

Figure 1 shows that the empirical model E suffers from small-sample bias in inferring the true population moments from the synthetic data $\mathbf{y}$. The upper-left window of Figure 1 shows that the posterior mean of $\mathbf{E}(r_t^f|E)$ was slightly larger than the true value $\mathbf{E}(r_t^f|R)$. The third column of Table 2 reports that the posterior mean of $\mathbf{E}(r_t^f|E)$ is 5.813, which overestimates the true value of 5.725. However, the degree of the small-sample bias is not so severe: the posterior standard deviation of $\mathbf{E}(r_t^f|E)$ is 0.173. This implies that the 90 % Bayesian credible interval sufficiently includes the true value. Regarding the five other empirical moments, the upper-middle, upper-right, lower-left, lower-middle, and lower-right windows of Figure 1 show that the posterior distributions of $\mathbf{E}(ep_t|E)$, $\mathbf{E}(\ln\kappa_t|E)$, $\mathbf{V}(ep_t|E)$, $\mathbf{V}(\ln\kappa_t|E)$, and $\mathbf{Cor}(\ln\kappa_t|E)$ contain corresponding true values close to their modal values within the 90% Bayesian credible intervals. Thus, even with a small sample bias and posterior uncertainty, owing to the short length of data $\mathbf{y}$, empirical model (19) infers the true values of the targeted population moments with high precision. This property of the empirical model E is necessary for the successful inference of the true values of the structural parameters through the MEI posterior sampler.

3.3. Posterior inference of the structural model A: ***Step 2***

Step 2 implements the RW-MH algorithm to infer the DSGE model A. Given a draw of θ_A from the candidate distribution in **Step 2(a)**, the four theoretical moments $\mathbf{E}(r_t^f|A)$, $\mathbf{E}(\ln\kappa_t|A)$, $\mathbf{V}(\ln\kappa_t|A)$, and $\mathbf{Cor}(\ln\kappa_t|A)$ are calculated using their analytical values.

$$\mathbf{E}(r_t^f|A) = \beta^{-1}\exp\left(\frac{\gamma\delta_0}{1-\delta_1} + \frac{\gamma^2\sigma_e^2}{1-\delta_1^2}(\delta_1^2 - 0.5)\right) - 1,$$

$$\mathbf{E}(\ln\kappa_t|A) = \frac{\delta_0}{1-\delta_1}, \quad \mathbf{V}(\ln\kappa_t|A) = \frac{\sigma_e^2}{1-\delta_1^2}, \quad \text{and} \quad \mathbf{Cor}(\ln\kappa_t|A) = \delta_1.$$

The theoretical population mean and variance of the equity premium, $\mathbf{E}(ep_t|A)$ and $\mathbf{V}(ep_t|A)$, have no analytical representations. Given the corresponding initial state variable $\ln\kappa_0 = \delta_0/(1-\delta_1)$, eqs.(11), (17), and (18) generate a synthetic time series of ep_t for $T_A = 1,000$ sample periods.

Theoretical moments $\mathbf{E}(ep_t|A)$ and $\mathbf{V}(ep_t|A)$ are then approximated as the time-series mean and variance, respectively, of the synthetic data of ep_t; $\mathbf{E}(ep_t|A) = T_A^{-1}\sum_{t=1}^{T_A} ep_t$ and $\mathbf{V}(ep_t|A) = T_A^{-1}\sum_{t=1}^{T_A}(ep_t - \mathbf{E}(ep_t|A))^2$. The theoretical moment vector $\mathbb{M}_A(\theta_A)$ is given by the following:

$$\mathbb{M}_A(\theta_A) \equiv 100 \times [\mathbf{E}(r_t^f|A), \mathbf{E}(ep_t|A), \mathbf{V}(ep_t|A), \mathbf{E}(\ln\kappa_t|A), \mathbf{V}(\ln\kappa_t|A), \mathbf{Cor}(\ln\kappa_t|A)].$$

Eq.(7) provides the JS likelihood $\Sigma_i^I \ln p_\lambda(\mathbf{m}_{E,i}|m_{A,i}(\theta_A))$, given the empirical moment distribution $\mathbf{M}_E$.

To construct prior distributions of the structural parameters $p(\theta_A|A)$, this study adopts three strategies. The first strategy (the third-fifth columns of Table 2), denoted "Prior I," assumes the correctly specified informative prior distributions for the structural parameters θ_A with the prior means matching the true values; the prior distribution of β is the Beta distribution with a mean of 0.980 and standard deviation of 0.001; that of γ is the Gamma distribution with a mean of 2.000 and standard deviation of 1.500; that of δ_0 is the normal distribution with a mean of 0.017 and standard deviation of 0.005; that of δ_1 is the normal distribution with a mean of 0.180 and standard deviation of 0.100; and that of σ_e^2 is the inverted Gamma distribution with a mean of 0.003 and standard deviation of 0.001.

The second strategy assumes the uniform distributions with suitable supports for the five structural parameters (the sixth and seventh columns of Table 2, denoted "Prior II"). The second prior strategy checks how well the MEI posterior sampler recovers true values of the structural parameters θ_A without any prior information. Finally, the third strategy (the eighth-tenth columns of Table 2 denoted "Prior III"), sets the prior mean of the relative risk aversion parameter γ to an incorrectly high value of 5.000 and keeps prior distributions of the other structural parameters to the same as in Prior I. Hence, the third prior strategy scrutinizes how the MEI posterior sampler updates the misspecified prior position of γ toward the true value of 2.000 exploiting the posterior information of the empirical moment distributions $\mathbf{M}_E$.

4. Monte Carlo experiment results

This section reports the results of the Monte Carlo experiments with the MEI posterior sampler. Given the empirical posterior moments $\mathbf{M}_E$ (reported as solid gray lines in Figure 1), the RW-MH algorithm in **Step 2** simulates the posterior distributions of the structural parameters θ_A through the theoretical moments $\mathbb{M}_A$.

Initial values of $\theta_{A,0}$ are set to be far from the true values for initiating the RW-MH algorithms: $\beta_0 = 0.95$, $\gamma_0 = 1.5$, $\delta_{0,0} = 0.001$, $\delta_{1,0} = 0.001$, and $\sigma^2_{e,0} = 0.001$. Consequently, initial values of the theoretical population moments $m_{A,0}$ are distinct from their true values. Subsequent MCMC draws are monitored to check how the theoretical moment distributions $\mathbf{M}_A$ converge to stationary distributions overlapping the empirical moment distributions $\mathbf{M}_E$ and how closely simulated posterior distributions of the structural parameters θ_A recover the corresponding true values.

This study configures the MEI posterior sampler as follows. The finite support of $\mathbf{E}(r_t^f)$ is $[0.0,\ 20.0]$, that of $\mathbf{E}(ep_t)$ is $[-1.0,\ 3.0]$, that of $\mathbf{V}(ep_t)$ is $[0.0,\ 80.0]$, that of $\mathbf{E}(\ln \kappa_t)$ is $[0.0, 5.0]$, that of $\mathbf{V}(\ln \kappa_t)$ is $[0.0,\ 80.0]$, and that of $\mathbf{Cor}(\ln \kappa_t)$ is $[-0.5,\ 0.5]$. Grid number K for discretizing finite supports is set to 300. Number N of posterior draws for the empirical moment distribution is set to 30,000. Subsection 2.1 shows that this study sets the number M for the theoretical moment distribution to a minimum value of 1. In this case, weight λ becomes 0.01, and JS likelihood (5) degenerates to eq.(7). The length of the MCMC sampling is set to 200,000. The first 20,000 MCMC samples are burned in to guarantee convergence to stationary distributions. The RW-MH draws maintain a 40 % rejection rate by suitably setting the configuration parameter τ.[12]

4.1. MEI posterior inferences of θ_A and m_A under Prior I

Figure 2 plots the kernel-smoothed densities of the empirical moment distributions $\mathbf{M}_E$ as solid grey lines and simulated theoretical moment distributions $\mathbf{M}_A$ under Prior I as solid black lines. The upper-left window corresponds to the population mean of the risk-free rate, $\mathbf{E}(r_t^f)$; the upper-middle window corresponds to the population mean of the equity premium, $\mathbf{E}(ep_t)$; the upper-right window corresponds to the population variance of the equity premium, $\mathbf{V}(ep_t)$; the lower-left window corresponds to the population mean of the consumption growth rate, $\mathbf{E}(\ln \kappa_t)$; the lower-middle window corresponds to the population variance of the consumption growth rate, $\mathbf{V}(\ln \kappa_t)$; and the lower-right window corresponds to the population first-order autocorrelation of the consumption growth rate, $\mathbf{Cor}(\ln \kappa_t)$. The fifth and sixth columns of Table 1 report the posterior means and standard deviations of the theoretical moment distributions under Prior I.

The upper-left, upper-right, and lower-middle windows in Figure 2 clearly show that the the-

[12]This study executes the entire procedure of the proposed MEI posterior inference by Python 3.9 codes. A MacBook Air with the M2 chip and 24GB RAM provides sufficiently quick implementation taking 149 minutes. Program codes are available upon request.

oretical posterior distributions of $\mathbf{E}(r_t^f|A)$, $\mathbf{V}(ep_t|A)$, and $\mathbf{V}(\ln \kappa_t|A)$ overlap with their empirical counterparts of $\mathbf{E}(r_t^f|E)$, $\mathbf{V}(ep_t|E)$, and $\mathbf{V}(\ln \kappa_t|E)$. The third and fifth columns of Table 1 show that the posterior means of these theoretical population moments, 5.805, 24.86, and 28.04, are quite close to those of their empirical counterparts, 5.813, 24.86, and 29.12, respectively. Figure 2 also shows that the theoretical posterior distributions of the other population moments $\mathbf{E}(ep_t|A)$, $\mathbf{E}(\ln \kappa_t|A)$, and $\mathbf{Cor}(\ln \kappa_t|A)$ overlap with their empirical counterparts $\mathbf{E}(ep_t|E)$, $\mathbf{E}(\ln \kappa_t|E)$, and $\mathbf{Cor}(\ln \kappa_t|E)$. However, the cross-equation restrictions imposed by Labadie's (1989) model lead to fewer overlapping degrees. Note that in Table 1, for all theoretical moments, their posterior means are close to not only those of the empirical moments but also true values; indeed, one posterior standard deviation interval of the theoretical moment distributions contain the corresponding true values.

Figure 3 depicts posterior distributions of the structural parameters θ_A. The upper-left, upper-middle, upper-right, lower-left, and lower-middle windows correspond to β, γ, σ_e^2, δ_0, and δ_1, respectively. Each window plots the kernel-smoothed density of the posterior distribution of a structural parameter as a solid black line, the calibrated true value as the vertical dashed blue line, and the prior distribution as the dashed red line. The third and fourth columns of Table 3 report the posterior means and standard deviations of the structural parameters under Prior I, respectively, with the corresponding calibrated true values in the second column.

The windows in Figure 3 show that the MEI posterior sampler successfully detects the calibrated true values of the structural parameters. All posterior distributions of the structural parameters θ_A contain the corresponding true values close to their posterior means. Particularly, posterior modal values of β, γ, and δ_0 almost perfectly match the true values. The third column of Table 3 confirms that the posterior means of the structural parameters are almost identical to the true values. Moreover, posterior distributions of the structural parameters tend to be more symmetric and narrower than their corresponding prior distributions. Hence, the MEI posterior sampler correctly updates prior belief on the positions of the structural parameters toward the true values by exploiting the empirical moment distributions $\mathbf{M}_E$ effectively.

4.2. Convergence and efficiency of the MEI posterior sampler

To check the convergence and efficiency of the MEI posterior sampler, Figures 4 and 5 plot

the time series of the simulated draws of the theoretical moments $\mathbb{M}_A$ and structural parameters θ_A, respectively. The upper-left, upper-middle, upper-right, and lower-middle windows in Figure 4 show that the posterior draws of $\mathbf{E}(r_t^f|A)$, $\mathbf{E}(ep_t|A)$, $\mathbf{V}(ep_t|A)$, and $\mathbf{V}(\ln \kappa_t|A)$, represented by the gray-solid lines, quickly converge to stationary distributions around empirical means (black solid lines) and true values (dotted black lines) with negligible autocorrelations. The lower-left and lower-right windows of Figure 4 show that the posterior draws of $\mathbf{E}(\ln \kappa_t|A)$ and $\mathbf{Cor}(\ln \kappa_t|A)$ converge to stationary distributions, although they are less efficient with higher autocorrelations. Simulated posterior draws of the structural parameters β, γ, σ_e^2, and δ_0 reported in Figure 5 quickly converged to stationary distributions around the calibrated true values with small autocorrelations, except for the posterior draws of δ_1 subject to high autocorrelation.[13] Therefore, based on this visual evidence, the MEI posterior sampler achieves stable convergence to stationary posterior distributions. However, it yields different degrees of efficiency over structural parameters and population moments.

To check the convergence and efficiency of the MEI posterior sampler more formally, this study calculates the convergence criterion $|Z|$ proposed by Geweke (1992) and the inefficiency factor proposed by Chib (2001) for each of the structural parameters as well as the population moments.[14] The second and third columns in Table 4 report the former and latter statistics, respectively. For all structural parameters and population moments, convergence criteria are less than 1.959. Hence, the null of a convergence cannot be rejected at the conventional 5 % significance level. However, as shown in the fourth column, inefficiency factors are large for most structural parameters and population moments. The minimum and maximum values of the inefficiency factors are 91 and 14,754, respectively, for $\mathbf{E}(r_t^f|A)$ and $\mathbf{Cor}(\ln \kappa_t|A)$. Particularly, inefficiency factors are greater than 2,000 for structural parameters, except σ_e^2. Therefore, the proposed MEI posterior sampler suffers from a substantial inefficiency problem; correct posterior inferences require sufficiently long posterior simulations.

4.3. Robustness check with Priors II: Role of prior information

This subsection examines the extent to which successful posterior inferences of the MEI ap-

[13] By construction, simulated draws of δ_1 are equivalent to those of $\mathbf{Cor}(\ln \kappa_t|A)$.

[14] Particularly, in this study, Geweke's convergence criterion uses the first 20 % and last 40% posterior draws for the first and second distributions, respectively. The inefficient factor exploits the estimated autocorrelation coefficients up to 20,000 lags.

proach under Prior I depend on the correctly specified prior belief about the true values of structural parameters. The sixth and seventh columns of Table 2 show that Prior II adopts uniform prior distributions for all structural parameters. Hence, the results of the MEI posterior sampler under Prior II reflect only information of the empirical moment distributions $\mathbf{M}_E$ but are independent of any prior belief in the positions of the structural parameters. Figure 6 depicts the kernel-smoothed densities of the empirical moment distributions $\mathbf{M}_E$ as solid gray lines and simulated theoretical moment distributions $\mathbf{M}_A$ under Prior II as solid black lines. The seventh and eighth columns of Table 1 report the posterior means and standard deviations of the theoretical moment distributions under Prior II.

Notably, each window in Figure 6 is visually indistinguishable from the corresponding window in Figure 2. Particularly, the simulated posterior distributions of $\mathbf{E}(r_t^f|A)$, $\mathbf{V}(ep_t|A)$, and $\mathbf{V}(\ln \kappa_t|A)$ overlap with their empirical counterparts, $\mathbf{E}(r_t^f|E)$, $\mathbf{V}(ep_t|E)$, and $\mathbf{V}(\ln \kappa_t|E)$, as shown in Figure 2. The seventh and eighth columns of Table 1 also confirm this visual impression; the posterior means and standard deviations of the theoretical moment distributions under Prior II are close and quite similar to those under Prior I.

Results from Prior II imply that the MEI posterior sampler successfully traces the empirical moment distributions and, as a result, recovers the true positions of the structural parameters even without any prior information of the structural parameters. Figure 7 shows simulated posterior distributions of the structural parameters θ_A under Prior II. Construction of Figure 7 is identical to that of Figure 3 under Prior I. The MEI posterior sampler under Prior II implies the kernel-smoothed densities of the posterior distributions of the structural parameters (solid black lines) that contain the corresponding true values (vertical blue dashed lines) close to their posterior means, even with flat prior distributions (red dashed lines). However, several crucial differences in the posterior inferences of the structural parameters between Priors I and II are shown. The fifth and sixth columns of Table 3 show the posterior means and standard deviations of structural parameters under Prior II. First, the posterior mean of γ under Prior II is 2.415, which is much larger than 2.076 under Prior I and the calibrated true value of 2.000. Second, posterior standard deviations under Prior II were larger than those under Prior I.

These differences in the posterior inferences of the structural parameters are correctly re-

flected in the lower marginal likelihood under Prior II than under Prior I. The last two rows of Table 3 show the estimated log relative marginal likelihood $\ln \hat{\psi}_\lambda(\mathbf{y}|A, E)$ and the posterior odds ratio (relative to Prior I) for each specification. The estimated log relative marginal likelihood under Prior II is -719,685.2, which is much lower than that under Prior I (-719,673.8). The resulting posterior odds ratio of Prior II relative to Prior I is 0.000. The result of the formal model comparison validates the general conjecture that correct prior information leads to a higher posterior model probability.

4.4. Robustness check with Priors III: Updating misspecified prior information

A major concern of the MEI approach originally developed by Geweke (2010) is that the model evaluation and comparison depend on the prior information of the structural parameters because the original MEI approach is absent from the updating process of θ_A through data. Hence, if prior information is crucially misspecified, the resulting model evaluation, comparison, and criticism with the MEI approach are heavily distorted. The last subsection shows that the MEI posterior sampler proposed in this study updates misspecified prior information correctly using information of the empirical moment distributions. Simultaneously, the misspecified prior information of a particular structural parameter leads to biased posterior inferences of all other structural parameters through the cross-equation restrictions of the DSGE model. The posterior odds ratio identifies the distortion caused by the misspecification of prior information of the structural parameters.

Recall that Prior III sets the prior mean of γ to an incorrectly higher value of 5.000. Figure 8 plots the kernel-smoothed densities of the empirical and theoretical moment distributions, $\mathbf{M}_E$ and $\mathbf{M}_A$ under Prior III, as shown in Figures 2 and 6 for Priors I and II. As observed in the lower-right window, the MEI posterior sampler under Prior III results in the worst overlap of the theoretical moment distribution with the empirical one with respect to $\mathbf{E}(\ln \kappa_t)$. The last two columns of Table 1 also numerically confirm this property of Prior III; the posterior mean of $\mathbf{E}(\ln \kappa_t|A)$ under Prior III is 1.706, which is much lower than those of 2.119 and 2.007 under Priors I and II, respectively, and far from its empirical counterpart of 2.256 as well.

Figure 9 plots the kernel-smoothed densities of the posterior distributions of the structural parameters, as in Figures 3 and 7 for Priors I and II. The upper-middle window of Figure 9 confirms that the prior distribution of γ is misspecified with a larger prior mean of 5.000. Two important

results are worth mentioning. First, although the posterior distribution of γ remains biased to the right, the MEI posterior sampler updates the misspecified prior distribution of γ correctly, toward the true value of 2.000. The seventh column of Table 3 shows that the posterior mean of γ under Prior III is 3.672, which is close to the true value of 2.000. Second, the misspecification of the prior mean of γ results in biased posterior inferences on not only γ itself but also on other structural parameters. Particularly, the posterior means of the consumption growth rate parameters δ_0 and δ_1 are biased downward, as shown in the last two columns of Table 3. Hence, the misspecified prior belief about the position of a single structural parameter tends to distort the posterior inferences of the entire parameter space of the DSGE model through cross-equation restrictions. The estimated log relative marginal likelihood under Prior III is -719,675.3, which is significantly lower than that under Prior I. The resulting posterior odds ratio of Prior III relative to Prior I was 0.223. Thus, the formal Bayesian model comparison prefers Prior I to Prior III and correctly identifies the distorting effect of the misspecified prior belief of the structural parameters.

5. Concluding remarks

This study developed a distribution-matching Bayesian indirect inference method extending the MEI approach by Geweke (2010). Following the central premise of the MEI approach, the proposed Bayesian strategy assumes that DSGE models only provide prior distributions of unobservable population moments. Based on a statistical reference model to estimate empirical moment distributions, the developed MEI posterior sampler simulates the theoretical moment distributions to overlap empirical moments as tightly as possible and, as a result, infers the posterior distributions of the structural parameters of the DSGE model. Because deriving a full-information likelihood is unnecessary, this simulation-based method makes drawing posterior inferences of potentially misspecified nonlinear DSGE models with flexible exogenous impulses possible.

This study formalizes the MEI posterior sampler utilizing the restricted DM model and its approximated representation of JS divergence as a parametric probability model of mixing discretized empirical and theoretical moment distributions. As an inheritance of the original MEI approach, the proposed posterior inference method reflects sampling variations of the targeted population moments in the posterior inference of the DSGE model under investigation. Essentially, it acts as a data

augmentation technique that avoids overfitting actual data and enhances the generalization ability of the inferred DSGE model.

Monte Carlo experiments based on Labadie's (1989) nonlinear equilibrium asset-pricing model provide solid evidence that the proposed MEI posterior sampler successfully detects calibrated values of the structural parameters. Moreover, the estimated posterior odds ratios distinguish the model with correctly specified prior distributions of the structural parameters from those with uniform and misspecified prior distributions.

The MEI posterior inference scheme is not necessarily specific to DSGE models. Moreover, it is applicable to the estimation and inference problems of structural models in broad fields of economics. Particularly, its main characteristic of distribution matching enables researchers to draw posterior inferences of structural models with heterogeneous economic agents, which have stringent implications on distributions of household income and firm productivity.

Finally, the proposed MEI posterior inference scheme had several limitations. First, simultaneous draws of the theoretical population moments from a nonlinear DSGE model are computationally costly for a large number M. Hence, the MEI posterior sampler is practically infeasible over continuous variations in the weight λ in the JS likelihood. This limitation implies that an MEI user cannot infer the degree of misspecification of the DSGE model through the posterior odds ratios calculated for different values of λ, as exercised in the DSGE-VAR method by Del Negro and Schorfheide (2004). Second, more efficient candidate densities in the RW-MH algorithm are needed to improve the inefficiency of the MEI posterior sampler (Section 4.2). These are the main challenges that this study's Bayesian strategy needs to address in the future.

References

Box, G. E. P., 1980, Sampling and Bayes inference in scientific modeling and robustness , *Journal of the Royal Statistical Society*, A 143, 383-430.

Canova, F., 1994, Statistical inference in calibrated models, *Journal of Applied Econometrics* 9, S123-S144.

Canova, F., 2007, *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton, NJ.

Chernozhukov, V., Hong, Han, 2003, An MCMC approach to classical estimation, *Journal of Econometrics*, 115, 293-346.

Chib, S., 2001, Markov chain Monte Carlo methods: computation and inference , in Heckman, J. J., Leamer, E., eds., *Handbook of Econometrics*, vol.5, 3569-3649, North Holland, Amsterdam.

DeJong, D. N., Dave, C., 2011, *Structural Macroeconometrics*, 2nd ed. Princeton University Press, Princeton, NJ.

DeJong, D. N., Ingram, B. F., Whiteman, C. H., 1996, A Bayesian approach to calibration, *Journal of Business & Economic Statistics*, 14, 1-9.

Del Negro, M., Schorfheide, F., 2004, Priors from general equilibrium models for VARs, *International Economic Review*, 45, 643-673.

Del Negro, M., Schorfheide, F., 2011, Bayesian macroeconometrics, in Geweke, J., Koop, G., Van Dijk, H., eds, *The Oxford Handbook of Bayesian Econometrics*, 293-389, Oxford University Press.

Eichenbaum, M., 1991, Real business-cycle theory: wisdom or whimsy?, *Journal of Economic Dynamics and Control*, 15, 607-626.

Faust, J., and Gupta, A.., 2012, Posterior predictive analysis for evaluating DSGE models, NBER Working Paper, 17906.

Fernández-Villaverde, J., Rubio-Ramirez, J. F., Schorfheide, F., 2016, Solution and estimation methods for DSGE models, in Taylor, J. B., Uhlig, H. eds., *Handbook of Macroeconomics*, vol.2A, 527-724.

Forneron, J. J., Ng, S., 2018, The ABC of simulation estimation with auxiliary statistics, *Journal of Econometric*, 205, 112-139.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2003, *Bayesian Data Analysis*, second edition, Chapman & Hall/CRC, Boca Raton, FL.

Geweke, J., 1992, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, in Bernardo et al. eds *Bayesian Statistics*, vol.4, 169-194, Clarendon Press, Oxford, U.K.

Geweke, J., 1999, Using simulation methods for Bayesian econometric models: inference, development and communication, *Econometric Reviews*, 18, 1-126.

Geweke, J., 2010, *Complete and Incomplete Econometric Models*, Princeton University Press, Princeton, NJ.

Inoue, A., Shintani, M., 2018, Quasi Bayesian model selection, *Quantitative Economics*, 9, 1265-1297.

Kano, T., Nason, J. M., 2014, Business cycle implications of internal consumption habit in new Keynesian models, *Journal of Money, Credit, and Banking*, 49, 519-544.

Kim, J. Y., 2002, Limited information likelihood and Bayesian analysis, *Journal of Econometrics*, 107, 175-193.

Labadie, P., 1989, Stochastic inflation and the equity premium, *Journal of Monetary Economics*, 24, 277-298.

Lancaster, T., 2004, *An Introduction to Modern Bayesian Econometrics*, Blackwell, Malden, MA.

Lin, J., 1991, Divergence measures based on the Shannon entropy, *IEEE Transactions on Information Theory*, 37, 145-151.

Loria, F., Matthes, C., Wang, M. C., 2022, Economic theories and macroeconomic reality, *Journal of Monetary Economics*, 126, 105-117.

Marjoram, P., Molitor, J., Plagnol, V., Tavarè, S., 2003, Markov chain Monte Carlo without likelihoods, *Proceedings of the National Academy of Science*, 100, 15324-15328.

Mehra, R., Prescott, E. C., 1985, The equity premium: a puzzle, *Journal of Monetary Economics*, 15, 145-61.

Nason, J. M., Rogers, J. H., 2006, The present-value model of the current account has been rejected: round up the usual suspects, *Journal of International Economics*, 68, 159-187.

Appendix A: Proof of Proposition 1

In this appendix, I omit subscript i without loss of generality. To prove Proposition 1, first note that the Gamma function $\Gamma(x)$ is equal to the factorial $(x-1)!$ for variable x. Stirling's approximation implies that for a large number of x, $\ln x! \approx x \ln x - x + \frac{1}{2}\ln(2\pi x) = (x+1)\ln x - x + \frac{1}{2}\ln(2\pi/x)$. As the last term is dominated for large x, it is the case that

$$\ln \Gamma(x+1) = \ln x! \approx (x+1)\ln x - x.$$

Note that Pólya distribution (4) can be rewritten as

$$\begin{aligned} p(\mathbf{m}_E|\mathbf{m}_A) &= \frac{N!(M+K-1)!}{(N+M+K-1)!}\prod_{k=1}^{K}\frac{(n_k+\alpha_k-1)!}{n_k!(\alpha_k-1)!} \\ &= \frac{N!(M+K)!}{(N+M+K)!}\frac{N+M+K}{M+K}\prod_{k=1}^{K}\frac{(n_k+\alpha_k)!}{n_k!\alpha_k!}\frac{\alpha_k}{n_k+\alpha_k} \\ &= \frac{N!(M+K)!}{(N+M+K)!}\frac{1+\lambda}{\lambda}\prod_{k=1}^{K}\frac{(n_k+\alpha_k)!}{n_k!\alpha_k!}\frac{\alpha_k}{n_k+\alpha_k} \end{aligned}$$

Taking the logarithm of both sides of the above result and applying Stirling's approximation of the factorial yields:

$$\begin{aligned} \ln p(\mathbf{m}_E|\mathbf{m}_A) &= \ln\frac{1+\lambda}{\lambda} + \ln N! + \ln(M+K)! - \ln(N+M+K)! \\ &\quad + \sum_{k=1}^{K}\ln(n_k+\alpha_k)! - \sum_{k=1}^{K}\ln n_k! - \sum_{k=1}^{K}\ln\alpha_k! + \sum_{k=1}^{K}\ln\alpha_k - \sum_{k=1}^{K}\ln(n_k+\alpha_k) \\ &\approx \ln\frac{1+\lambda}{\lambda} + (N+1)\ln N + (M+K+1)\ln(M+K) - (N+M+K+1)\ln(N+M+K) \\ &\quad + \sum_{k=1}^{K}(n_k+\alpha_k)\ln(n_k+\alpha_k) - \sum_{k=1}^{K}n_k\ln n_k - \sum_{k=1}^{K}\alpha_k\ln\alpha_k \\ &= \ln N + \sum_{k=1}^{K}(n_k+\alpha_k)\ln\left(\frac{n_k+\alpha_k}{N+M+K}\right) \\ &\qquad - \sum_{k=1}^{K}n_k\ln\left(\frac{n_k}{N}\right) - \sum_{k=1}^{K}\alpha_k\ln\left(\frac{\alpha_k}{M+K}\right) \\ &= \ln N - \sum_{k=1}^{K}n_k\left[\ln\left(\frac{n_k}{N}\right) - \ln\left(\frac{n_k+\alpha_k}{N+M+K}\right)\right] - \sum_{k=1}^{K}\alpha_k\left[\ln\left(\frac{\alpha_k}{M+K}\right) - \ln\left(\frac{n_k+\alpha_k}{N+M+K}\right)\right] \end{aligned} \tag{A.1}$$

Note that the term $(n_k+\alpha_k)/(N+M+K)$ in the third and fourth terms on the RHS of the last equality in eq.(A.1) can be rewritten as the weighted average between ζ_k and q_k with weights $1/(1+\lambda)$ and $\lambda/(1+\lambda)$:

$$\left(\frac{n_k+\alpha_k}{N+M+K}\right) = \frac{1}{1+\lambda}\zeta_k + \frac{\lambda}{1+\lambda}q_k.$$

Substituting this relation into eq.(A.1) establishes Proposition 1:

$$\begin{aligned} \ln p(\mathbf{m}_E|\mathbf{m}_A) \approx \ln N &- (1+\lambda)N\sum_{k=1}^{K}\frac{1}{1+\lambda}\zeta_k\left[\ln(\zeta_k) - \ln\left(\frac{1}{1+\lambda}\zeta_k + \frac{\lambda}{1+\lambda}q_k\right)\right] \\ &- (1+\lambda)N\sum_{k=1}^{K}\frac{\lambda}{1+\lambda}q_k\left[\ln(q_k) - \ln\left(\frac{1}{1+\lambda}\zeta_k + \frac{\lambda}{1+\lambda}q_k\right)\right]. \end{aligned}$$

Appendix B: Approximated marginal density when $\lambda \to (K+1)/N$

Suppose that a single theoretical draw m_A drops into the k'-th subinterval $\mathbf{s}_{k'}$. For $k \neq k'$, $\alpha_k = 1$ and $q_k = 1/(K+1)$. For k', $\alpha_{k'} = 2$ and $q_{k'} = 2/(K+1)$. When $\lambda \to (K+1)/N$, eq.(A.1) becomes:

$$\begin{aligned}
\lim_{\lambda \to \frac{K+1}{N}} \ln p(\mathbf{m}_E | m_A \in \mathbf{s}_{k'}) \approx \ln N &+ \sum_{k \neq k'} n_k \left[\ln\left(\frac{n_k+1}{N+K+1}\right) - \ln\left(\frac{n_k}{N}\right) \right] + \sum_{k \neq k'} \left[\ln\left(\frac{n_k+1}{N+K+1}\right) - \ln\left(\frac{1}{K+1}\right) \right] \\
&+ n_{k'} \left[\ln\left(\frac{n_{k'}+2}{N+K+1}\right) - \ln\left(\frac{n_{k'}}{N}\right) \right] + 2 \left[\ln\left(\frac{n_{k'}+2}{N+K+1}\right) - \ln\left(\frac{1}{K+1}\right) \right] \\
= \ln N &+ \sum_{k=1}^{K} n_k \left[\ln\left(\frac{n_k+1}{N+K+1}\right) - \ln\left(\frac{n_k}{N}\right) \right] + \sum_{k \neq k'} \left[\ln\left(\frac{n_k+1}{N+K+1}\right) - \ln\left(\frac{1}{K+1}\right) \right] \\
&+ 2 \left[\ln\left(\frac{n_{k'}+2}{N+K+1}\right) - \ln\left(\frac{2}{K+1}\right) \right] + n_{k'} \left[\ln\left(\frac{n_{k'}+2}{N+K+1}\right) - \ln\left(\frac{n_{k'}+1}{N+K+1}\right) \right]
\end{aligned}$$

The first two terms on the RHS of the last equality are constant. The last term is approximately zero when N is sufficiently high. Then it is the case that

$$\begin{aligned}
\lim_{\lambda \to \frac{K+1}{N}} \ln p(\mathbf{m}_E | m_A \in \mathbf{s}_{k'}) &\propto \sum_{k \neq k'} \left[\ln\left(\frac{n_k+1}{N+K+1}\right) - \ln\left(\frac{1}{K+1}\right) \right] + 2 \left[\ln\left(\frac{n_{k'}+2}{N+K+1}\right) - \ln\left(\frac{2}{K+1}\right) \right] \\
&= -(K+1) \sum_{k=1}^{K} q_k \left[\ln(q_k) - \ln\left(\frac{1}{1+\lambda}\zeta_k + \frac{\lambda}{1+\lambda} q_k \right) \right] \\
&\propto -D_{KL}\left(\mathbf{q} \,\middle\|\, \frac{1}{1+\lambda}\boldsymbol{\zeta} + \frac{\lambda}{1+\lambda}\mathbf{q} \right) \\
&\propto \ln\left(\frac{n_{k'}+2}{N+K+1}\right)
\end{aligned}$$

where $D_{KL}\left(\mathbf{q} \,\middle\|\, \frac{1}{1+\lambda}\boldsymbol{\zeta} + \frac{\lambda}{1+\lambda}\mathbf{q} \right)$ denotes KL divergence from $\frac{1}{1+\lambda}\boldsymbol{\zeta} + \frac{\lambda}{1+\lambda}\mathbf{q}$ to $\mathbf{q}$. This implies

$$\begin{aligned}
\lim_{\lambda \to \frac{K+1}{N}} \ln p(\mathbf{m}_E | m_A) &\propto -\sum_{k=1}^{K} \mathbf{I}[m_A \in \mathbf{s}_k] D_{KL}\left(\mathbf{q} \,\middle\|\, \frac{1}{1+\lambda}\boldsymbol{\zeta} + \frac{\lambda}{1+\lambda}\mathbf{q} \right) \\
&\propto \sum_{k=1}^{K} \mathbf{I}[m_A \in \mathbf{s}_k] \ln\left(\frac{n_k+2}{N+K+1}\right)
\end{aligned}$$

where $\mathbf{I}[m_A \in \mathbf{s}_k]$ is the indicator function that takes the value of 1 if m_A drops into the k-th subinterval $\mathbf{s}_k$ and 0 otherwise. Note that when $\lambda \to (K+1)/N \approx 0$ the marginal likelihood is approximately proportional to the negative KL divergence from the empirical distribution $\boldsymbol{\zeta}$ to $\mathbf{q}$.

Appendix C: Stationarity of the MEI sampler

The proposed MEI sampler satisfies the reversibility (detailed balance) condition:

$$\begin{aligned}
p_\lambda(\theta_A^{new}|\mathbf{M}_E) r(\theta_A^{old}|\theta_A^{new}) &= p_\lambda(\theta_A^{new}|\mathbf{M}_E) \min\left\{1, \frac{p_\lambda(\theta_A^{old}|\mathbf{M}_E)}{p_\lambda(\theta_A^{new}|\mathbf{M}_E)}\right\} \\
&= \min\left\{p_\lambda(\theta_A^{new}|\mathbf{M}_E), p_\lambda(\theta_A^{old}|\mathbf{M}_E)\right\} \\
&= p_\lambda(\theta_A^{old}|\mathbf{M}_E) \min\left\{1, \frac{p_\lambda(\theta_A^{new}|\mathbf{M}_E)}{p_\lambda(\theta_A^{old}|\mathbf{M}_E)}\right\} \\
&= p_\lambda(\theta_A^{old}|\mathbf{M}_E) r(\theta_A^{new}|\theta_A^{old}). \qquad \text{(C.1)}
\end{aligned}$$

For the Markov kernel $K(\theta_A^{old}|\theta_A^{new})$ defined by

$$\begin{aligned} K(\theta_A^{old}|\theta_A^{new}) &= r(\theta_A^{old}|\theta_A^{new}) + \int [1 - r(\theta_A^{old}|\theta_A^{new})] d\theta_A^{old} \times \mathbf{I}[\theta_A^{old} = \theta_A^{new}], \\ &= r(\theta_A^{old}|\theta_A^{new}) + \alpha(\theta_A^{new}) \mathbf{I}[\theta_A^{old} = \theta_A^{new}] \end{aligned}$$

the posterior distribution satisfies

$$\begin{aligned} &\int K(\theta_A^{old}|\theta_A^{new}) p_\lambda(\theta_A^{new}|\mathbf{M}_E) d\theta_A^{new} \\ &= \int r(\theta_A^{old}|\theta_A^{new}) p_\lambda(\theta_A^{new}|\mathbf{M}_E) d\theta_A^{new} + \int \alpha(\theta_A^{new}) \mathbf{I}[\theta_A^{old} = \theta_A^{new}] p_\lambda(\theta_A^{new}|\mathbf{M}_E) d\theta_A^{new} \\ &= \int r(\theta_A^{old}|\theta_A^{new}) p_\lambda(\theta_A^{new}|\mathbf{M}_E) d\theta_A^{new} + \alpha(\theta_A^{old}) p_\lambda(\theta_A^{old}|\mathbf{M}_E) \\ &= \int r(\theta_A^{new}|\theta_A^{old}) p_\lambda(\theta_A^{old}|\mathbf{M}_E) d\theta_A^{new} + \alpha(\theta_A^{old}) p_\lambda(\theta_A^{old}|\mathbf{M}_E) \quad \text{from eq. (C.1)} \\ &= \int r(\theta_A^{new}|\theta_A^{old}) p_\lambda(\theta_A^{old}|\mathbf{M}_E) d\theta_A^{new} + \int [1 - r(\theta_A^{new}|\theta_A^{old})] d\theta_A^{new} p_\lambda(\theta_A^{old}|\mathbf{M}_E) \\ &= p_\lambda(\theta_A^{old}|\mathbf{M}_E) + \int r(\theta_A^{new}|\theta_A^{old}) d\theta_A^{new} p_\lambda(\theta_A^{old}|\mathbf{M}_E) - \int r(\theta_A^{new}|\theta_A^{old}) d\theta_A^{new} p_\lambda(\theta_A^{old}|\mathbf{M}_E) \\ &= p_\lambda(\theta_A^{old}|\mathbf{M}_E). \end{aligned}$$

Table 1: True, Empirical, and Theoretical Distributions of Population Moments $\mathbf{m}$

	True $\mathbf{m}_R$	Empirical $\mathbf{m}_E$		Theoretical $\mathbf{m}_A$ Prior I		Theoretical $\mathbf{m}_A$ Prior II		Theoretical $\mathbf{m}_A$ Prior III	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
$\mathbf{E}(r_t^f)$	5.725	5.813	0.173	5.805	0.196	5.821	0.330	5.836	0.203
$\mathbf{E}(ep_t)$	0.493	0.619	0.313	0.471	0.130	0.510	0.181	0.647	0.174
$\mathbf{V}(ep_t)$	27.50	24.86	2.603	24.86	2.027	27.17	2.197	24.93	2.198
$\mathbf{E}(\ln \kappa_t)$	2.075	2.256	0.456	2.119	0.312	2.007	0.436	1.706	0.213
$\mathbf{V}(\ln \kappa_t)$	31.03	29.12	3.143	28.04	2.269	28.27	2.326	28.03	2.252
$\mathbf{Cor}(\ln \kappa_t)$	0.180	0.129	0.077	0.178	0.047	0.174	0.052	0.135	0.039

Table 2: True Values and Prior Distributions of Structural Parameters

	True	Prior I			Prior II		Prior III		
		Dist.	Mean	S.D.	Dist.	Support	Dist.	Mean	S.D.
β	0.980	Beta	0.980	0.001	Uniform	[0.001, 0.999]	Beta	0.980	0.001
γ	2.000	Gamma	2.000	1.500	Uniform	[0.001, 10.00]	Gamma	5.000	1.500
δ_0	0.017	Normal	0.017	0.005	Uniform	[0.001, 0.500]	Normal	0.017	0.005
δ_1	0.180	Normal	0.180	0.100	Uniform	[0.001, 0.500]	Normal	0.180	0.100
σ_e^2	0.003	InvGam	0.003	0.001	Uniform	[0.001, 0.100]	InvGam	0.003	0.001

Table 3: Posterior Distributions of Structural Parameters

	True	Prior I		Prior II		Prior III	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
β	0.980	0.981	0.007	0.981	0.011	0.985	0.006
γ	2.000	2.076	0.533	2.415	1.044	3.672	0.856
δ_0	0.017	0.017	0.002	0.016	0.003	0.014	0.002
δ_1	0.180	0.178	0.047	0.174	0.052	0.137	0.040
σ_e^2	0.003	0.003	0.000	0.003	0.000	0.003	0.001
$\ln \hat{\psi}_\lambda(\mathbf{y}\|A, E)$		-719673.8		-719685.2		-719675.3	
Odds Ratio vs. Prior I		1		0.000		0.223	

TABLE 4: CONVERGENCE AND EFFICIENCY OF MEI POSTERIOR SIMULATOR

	Prior I		Prior II		Prior II	
	|Z|	Inefficiency	|Z|	Inefficiency	|Z|	Inefficiency
β	0.089	2,058	0.752	6,201	0.396	1,854
γ	0.017	2,139	0.581	3,371	0.158	1,769
δ_0	0.093	5,574	0.578	6,277	0.070	3,162
δ_1	0.091	14,478	0.889	12,906	0.718	5,288
σ_e^2	0.189	634	1.075	673	0.874	847
$\mathbf{E}(r_t^f \mid A)$	0.076	91	0.739	638	0.010	139
$\mathbf{E}(ep_t \mid A)$	0.197	3,851	0.880	5,850	1.050	3,889
$\mathbf{V}(ep_t \mid A)$	0.572	172	1.466	284	1.261	376
$\mathbf{E}(\ln \kappa_t \mid A)$	0.198	3,073	0.364	5,940	0.763	2,020
$\mathbf{V}(\ln \kappa_t \mid A)$	0.205	238	1.026	142	1.167	452
$\mathbf{Cor}(\ln \kappa_t \mid A)$	0.132	14,754	0.986	13,285	0.989	5,409

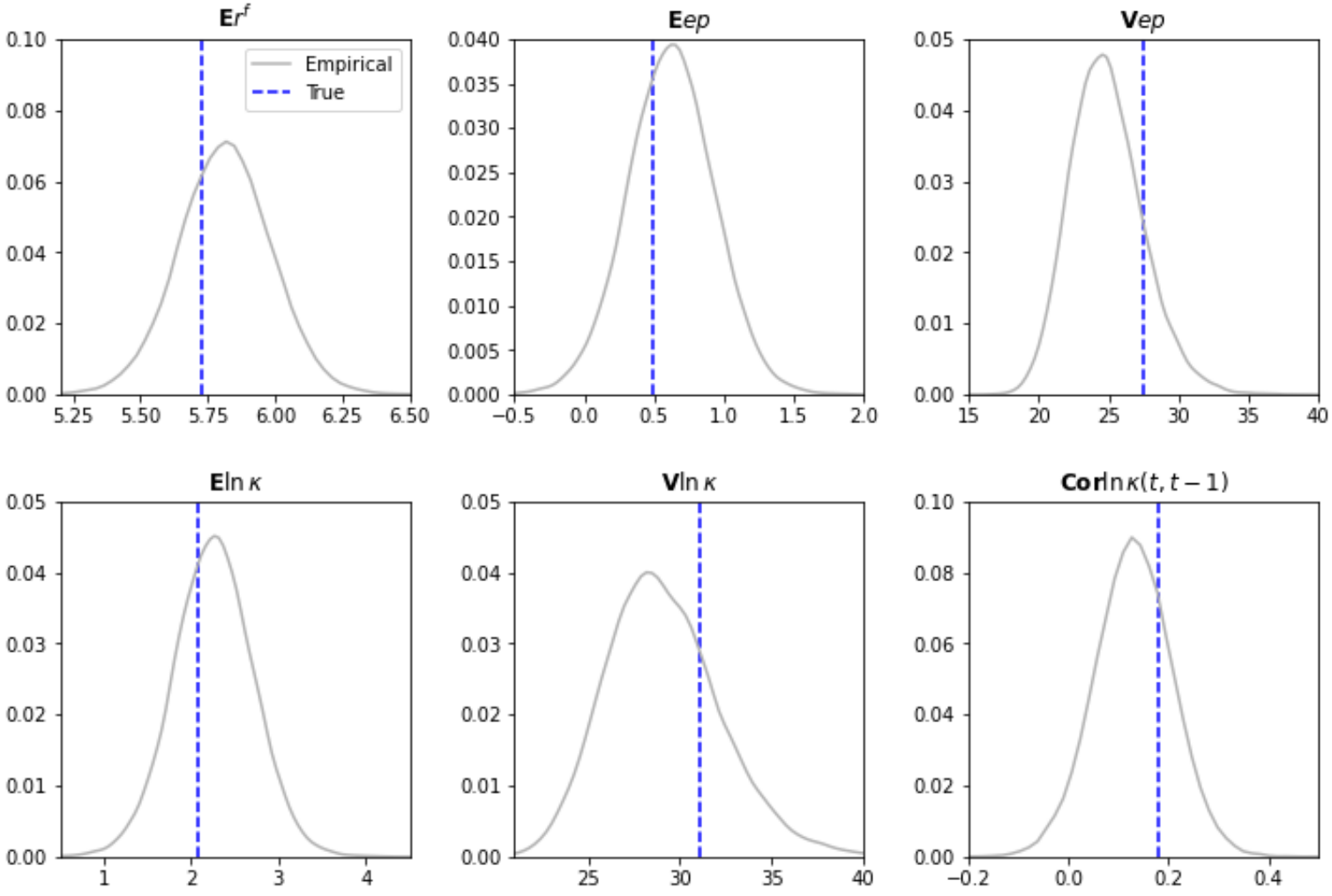

Figure 1: Empirical Distributions of Population Moments

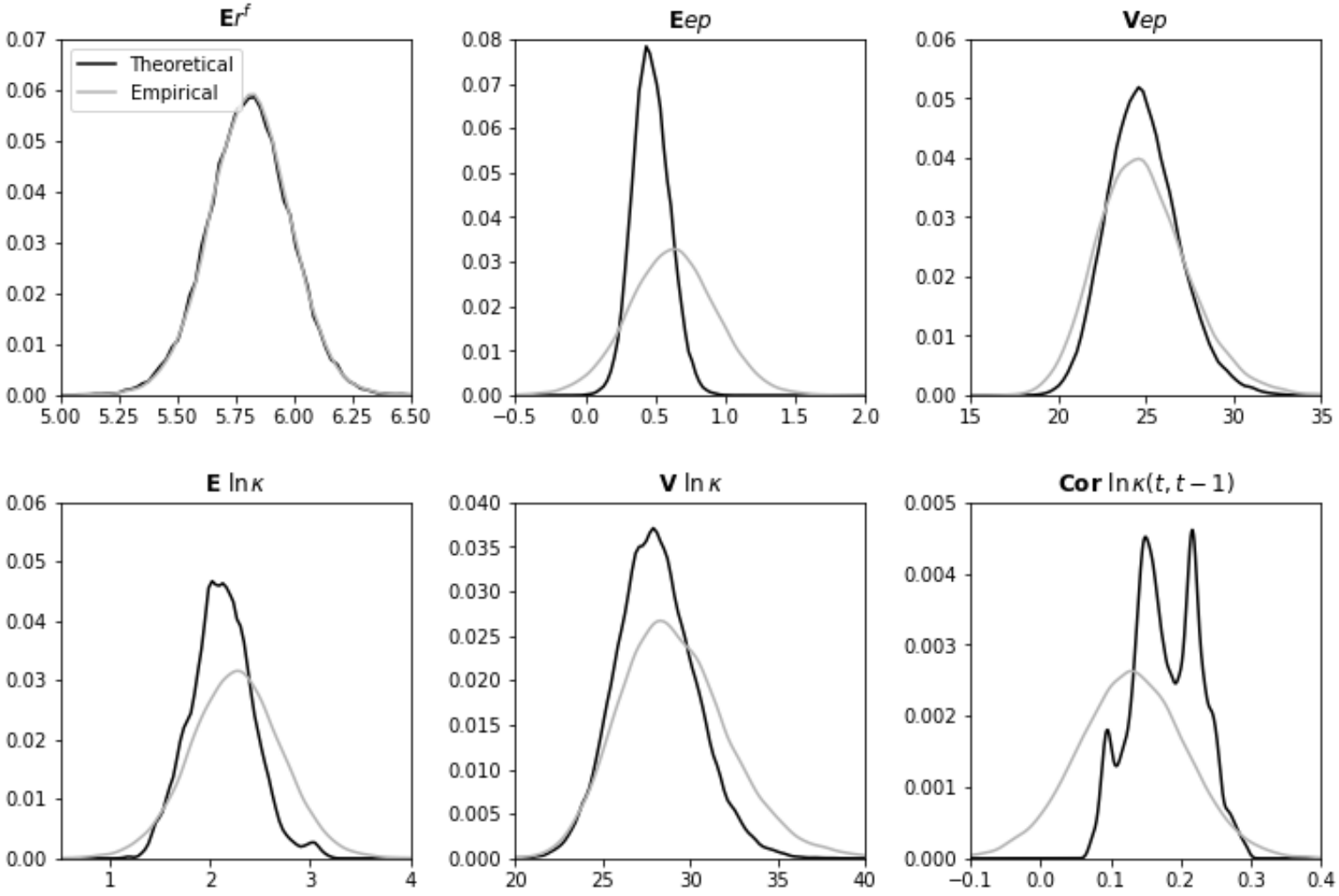

Figure 2: Empirical and Theoretical Distributions of Population Moments: Prior I

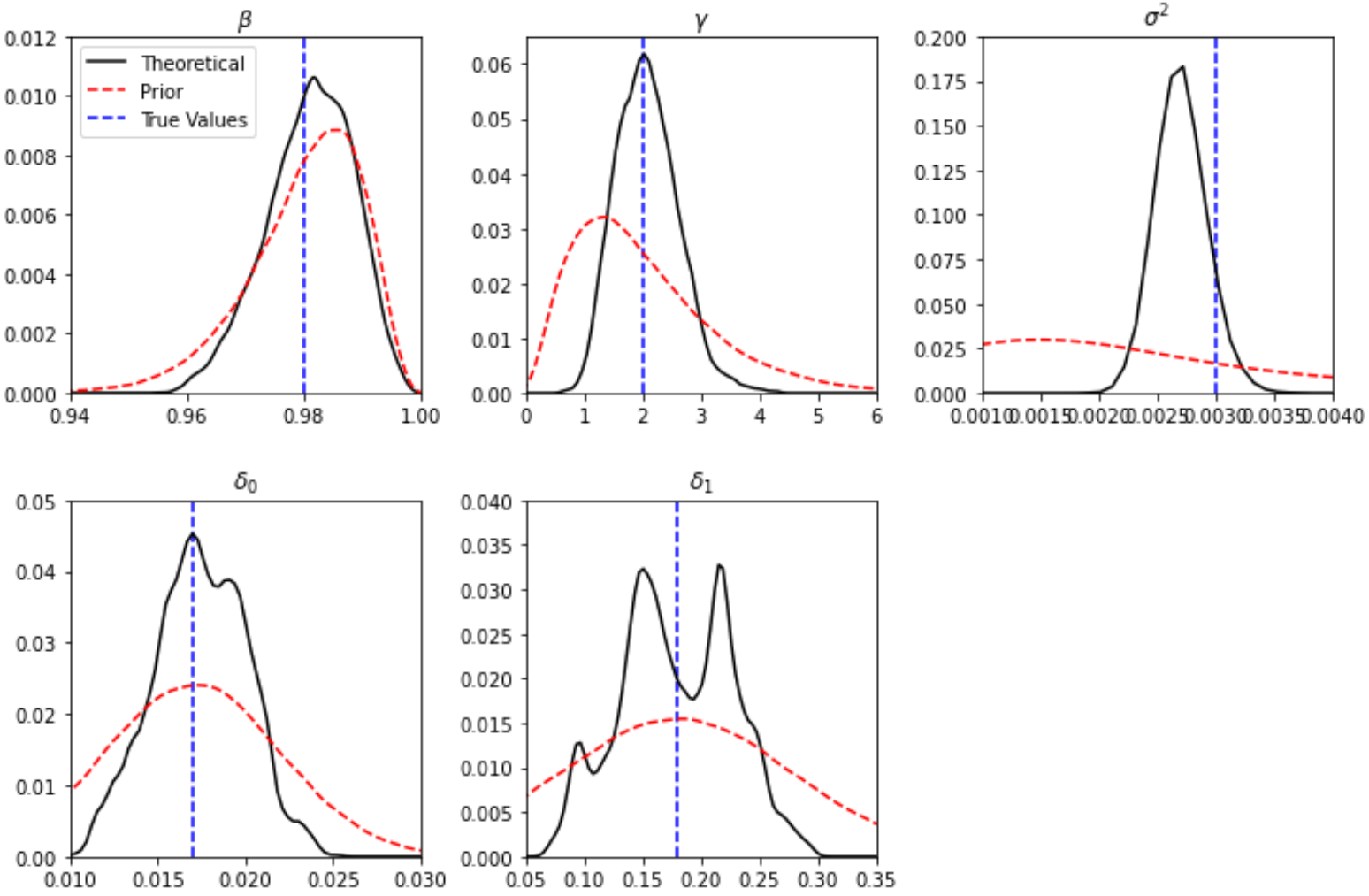

Figure 3: Posterior Distributions of Structural Parameters: Prior I

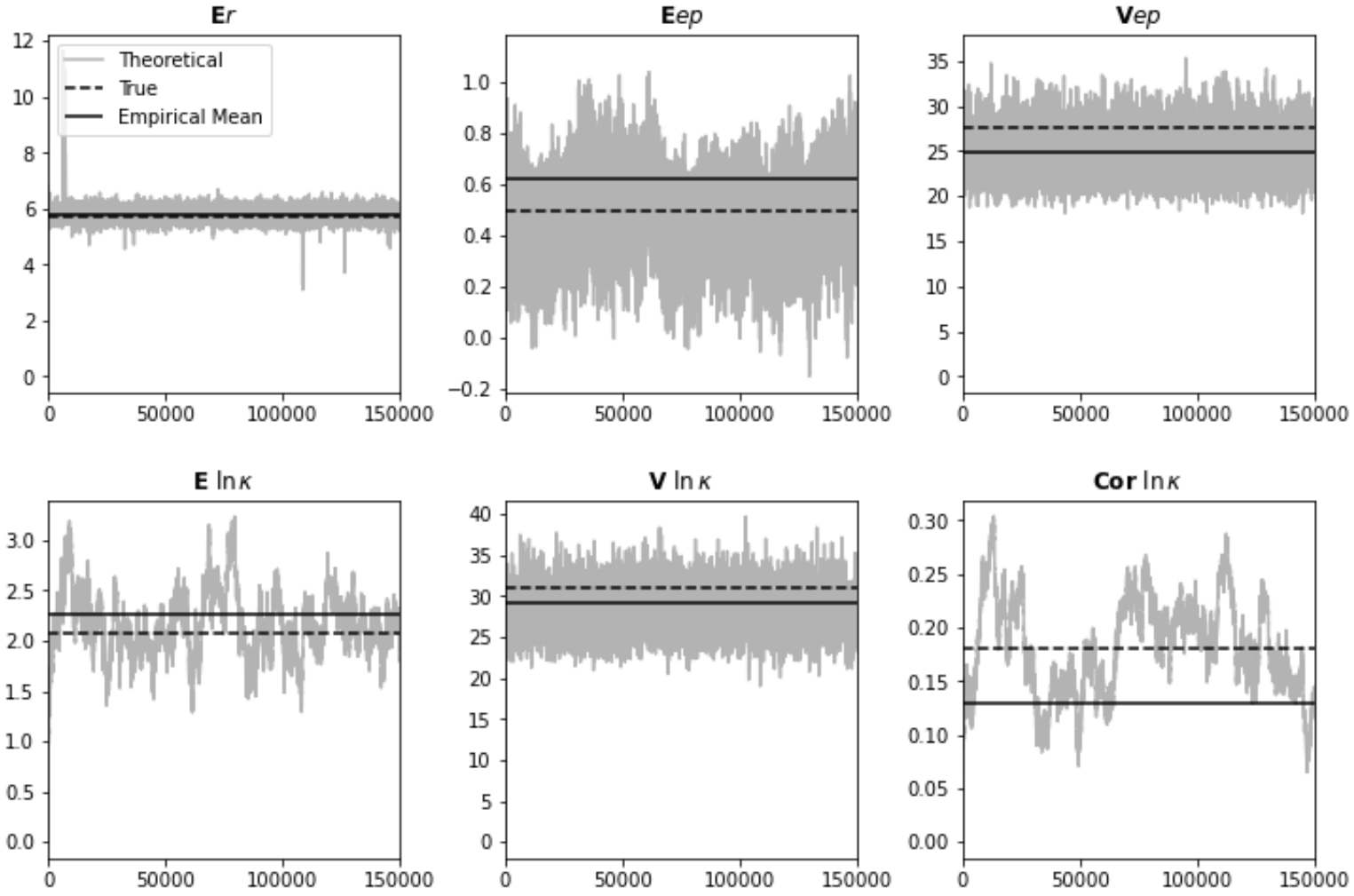

Figure 4: Convergence of Theoretical Moments m_A

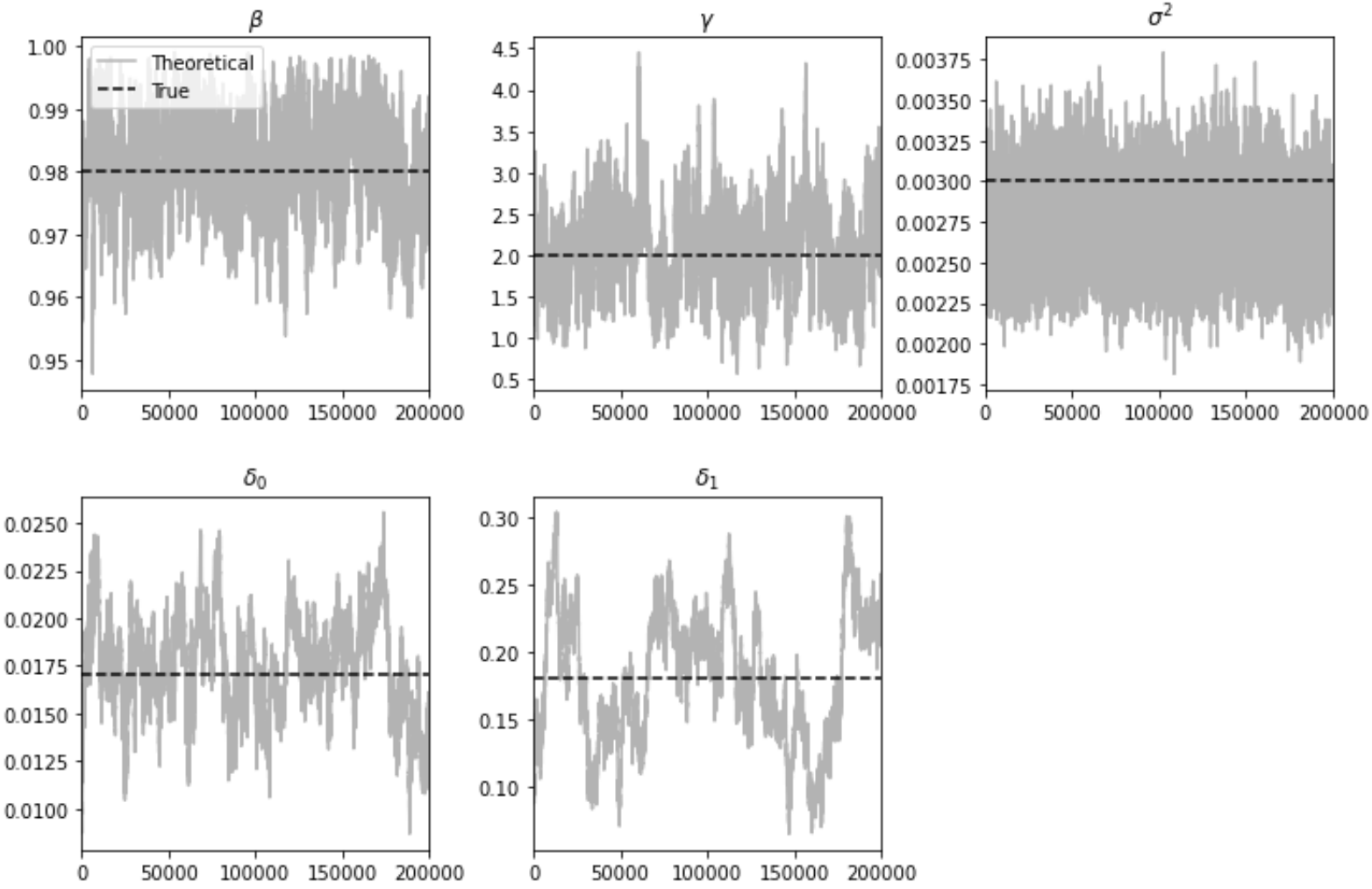

Figure 5: Convergence of Structural Parameters θ_A

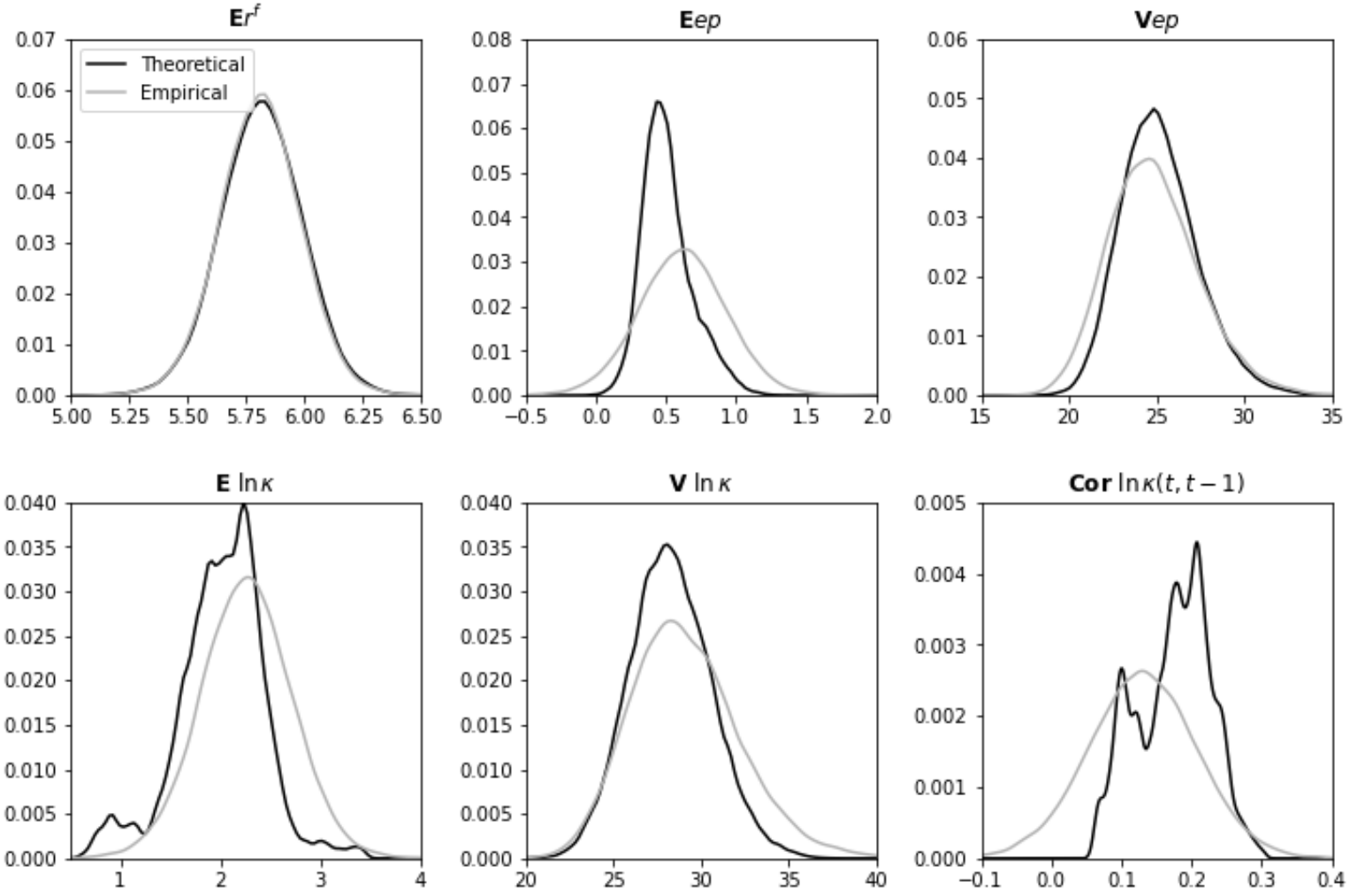

Figure 6: Empirical and Theoretical Distributions of Population Moments: Prior II

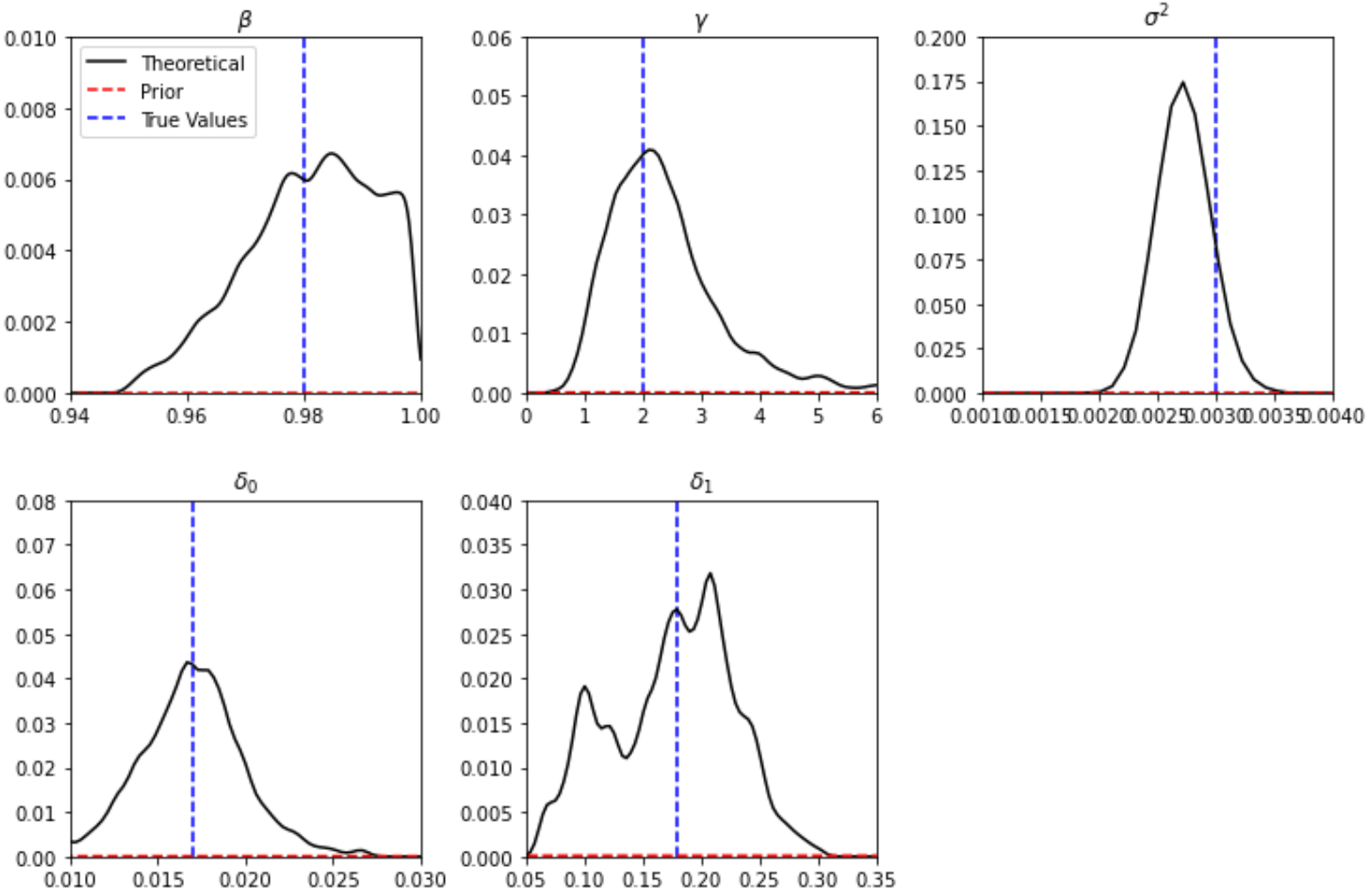

Figure 7: Posterior Distributions of Structural Parameters: Prior II

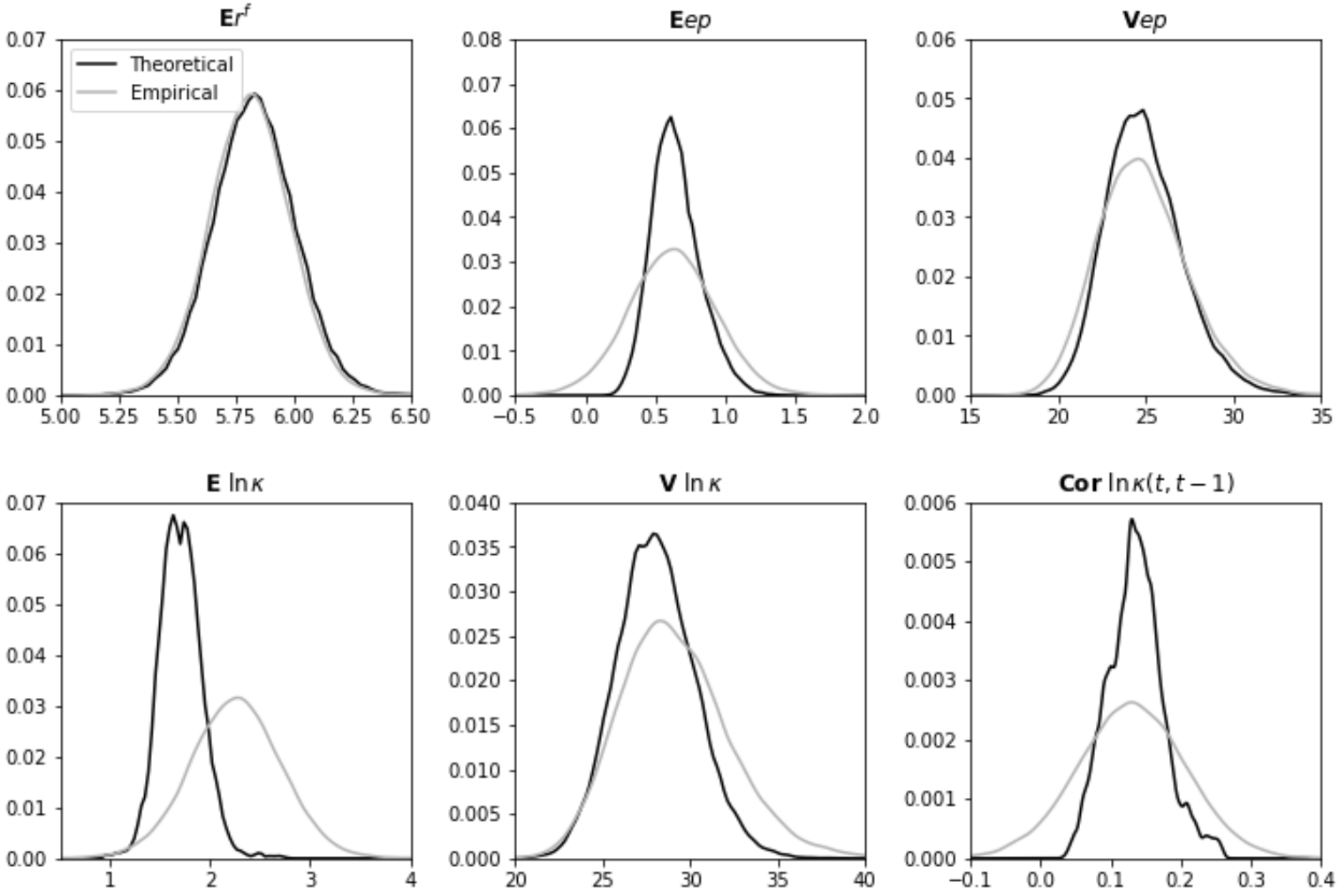

Figure 8: Empirical and Theoretical Distributions of Population Moments: Prior III

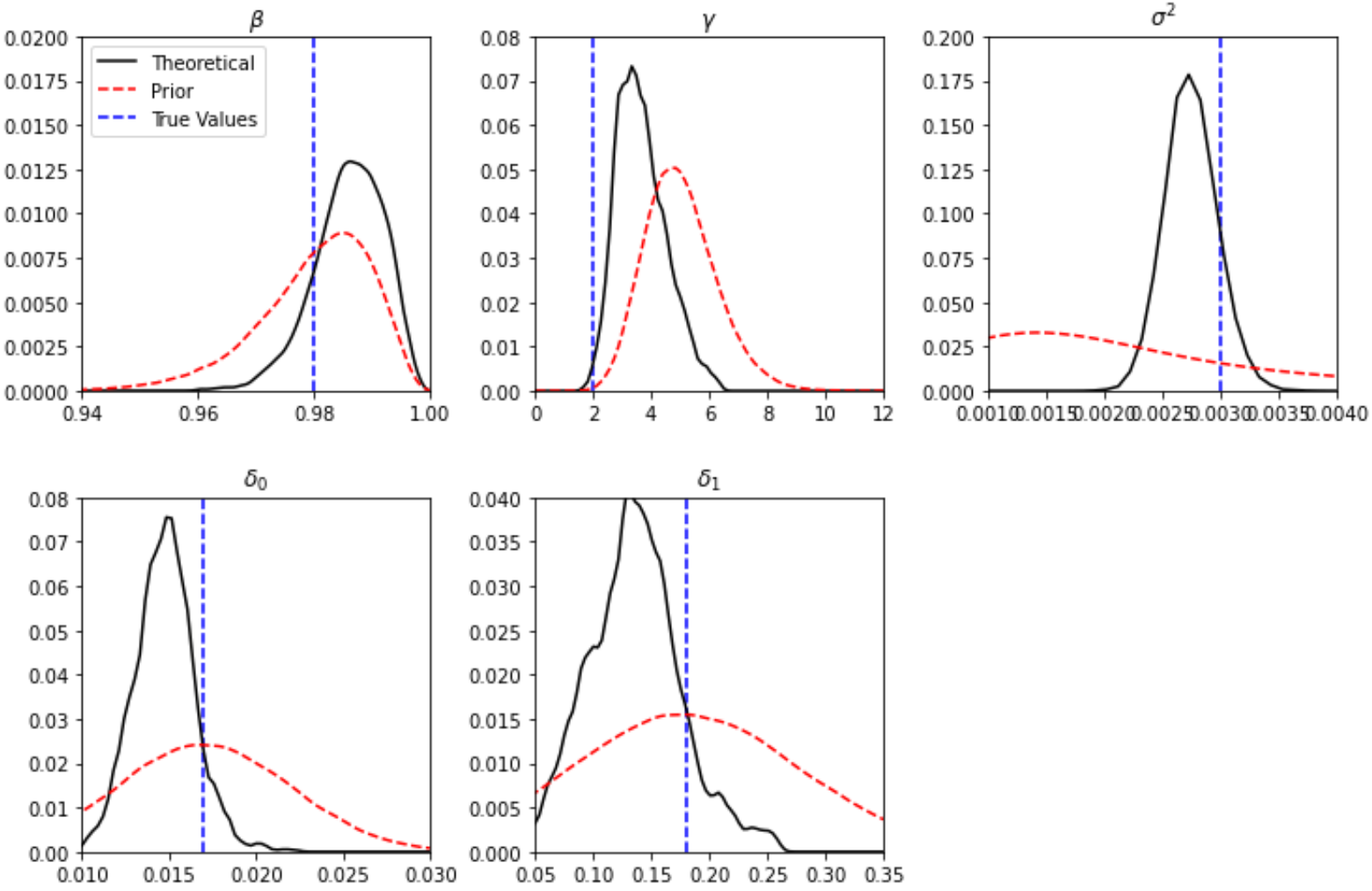

Figure 9: Posterior Distributions of Structural Parameters: Prior III