On Simulation Pseudo-Bias and Truncation in the Modified Harmonic Mean Estimator*

Ryohei Oishi†

March 2026

Abstract

The modified harmonic mean estimator developed by Gelfand and Dey (1994) for the marginal likelihood exhibits a smaller computational bias—known as simulation pseudo-bias—than the original estimator proposed by Newton and Raftery (1994). Simulation pseudo-bias occurs in importance sampling when the simulated samples from the proposal distribution have narrower empirical support than the target distribution, even if their theoretical supports are identical. The modified estimator mitigates this pseudo-bias by changing the importance sampling target from the prior density to a weighting density parameterized by posterior samples. Furthermore, the pseudo-bias correction method provides a computational rationale for the truncation embedded in commonly used weighting densities. Empirically, this paper revisits the finding of Smets and Wouters (2007) that the marginal likelihood of their DSGE model is comparable to that of a VAR model. Using various harmonic-mean-type estimators, I confirm the robustness of their original finding, which was based on the local Laplace approximation. I also demonstrate that relying on pseudo-biased estimators leads to an erroneous preference for the DSGE model.

JEL Classification: C11, C14, C15, C52, E32

Keywords: Marginal likelihood, Simulation pseudo-bias, Importance sampling, Markov chain Monte Carlo, Sequential Monte Carlo, DSGE model

*I am sincerely grateful to my supervisors, Vincent Sterk and Morten O. Ravn, for their helpful comments and support. I thank O. Deniz Akyildiz, Ben Deaner, Wataru Hagio, Stephen Hansen, Samuel Livingstone, Joe Marlow (discussant), Yuxuan Ren, Raf Wouters, Andrei Zeleneev, seminar participants at UCL, and conference participants at IAAE 2025, Junior Milan Time Series Workshop, 12th MMF PhD Conference, 7th QMUL Economics and Finance Workshop, 15th RCEA Bayesian Econometrics Workshop, 4th Dynare Workshop for Advanced Users, and RES 2025 for their helpful comments and discussions.

†University College London, Gower Street, London, WC1E 6BT, UK (e-mail: ryohei.oishi.22@ucl.ac.uk)

1 Introduction

This paper studies the estimation of the marginal likelihood,[1] defined as

$$p(y|M) = \int_{\Theta_M} p(y|\theta_M, M)p(\theta_M|M)d\theta_M, \tag{1}$$

where y denotes the data, $p(y|\theta_M, M)$ and $p(\theta_M|M)$ are the likelihood and the prior distribution, respectively, for a model M parameterized by parameters $\theta_M \in \Theta_M$.[2] The marginal likelihood plays a crucial role in Bayesian statistics, particularly in model selection, model averaging, and hyperparameter estimation via marginal-likelihood maximization. For instance, in model selection, the model with the highest marginal likelihood is selected when prior model probabilities are equal. In model averaging, the marginal likelihood is used as a weight to average over models.

Despite its importance, accurately estimating the marginal likelihood is computationally challenging. Except for a few restrictive cases, a closed-form solution is not available because all parameters must be integrated out of a complex function. Furthermore, naive Monte Carlo estimators are rarely used in practice due to their computational inefficiency, especially in high-dimensional problems.

In response to these computational challenges, various estimators have been proposed in the literature. Among these, harmonic mean estimators (HMEs) are widely used due to their computational simplicity. The original HME proposed by Newton and Raftery (1994) estimates the reciprocal of the marginal likelihood as the harmonic mean of the likelihoods evaluated at posterior samples. While valued for its simplicity, it may exhibit infinite variance. The modified HME of Gelfand and Dey (1994) estimates the reciprocal using a weighted harmonic mean and remains one of the most widely used estimators. According to simulation studies, the modified HME is as accurate as more computationally intensive estimators (e.g., Chib and Jeliazkov, 2001).

In this paper, I show that the modified HME exhibits less computational bias—known as simulation pseudo-bias (Lenk, 2009)—than the original HME. Simulation pseudo-bias occurs in importance sampling when the simulated samples from the proposal distribution have narrower empirical support than the target distribution.[3] That is, although the absolute continuity condition holds in theory (and hence the estimator is unbiased), it is effectively violated *in simulation* due to the finite-sample nature of the simulation. Consequently, because the target distribution of the modified HME is a weighting density

[1] The marginal likelihood is also known as the marginal data density, integrated likelihood, or normalizing constant.

[2] Henceforth, conditioning on M is omitted unless necessary for clarity.

[3] The concept of pseudo-bias was developed in the literature on bootstrap methods (e.g., Ventura (2002)). A pseudo-biased estimator is theoretically unbiased, but becomes biased in practice due to the finite nature of simulation and computation.

parameterized using information from the posterior samples, it shares similar empirical support with the posterior distribution, which serves as the proposal distribution for both estimators. In contrast, the original HME suffers from larger pseudo-bias because the prior distribution, which is the target, is typically more dispersed than the posterior.

This finding offers a new theoretical explanation for the well-documented superior accuracy of the modified HME observed in simulation studies. It suggests that this accuracy stems not only from the estimator's finite variance and the computational efficiency gained through the weighting density, but also from its smaller simulation pseudo-bias.[4] This result is particularly critical because pseudo-bias is positive and tends to increase with model complexity, as reported by Lenk (2009) and confirmed by my own Monte Carlo simulations. Consequently, pseudo-bias leads us to incorrectly favor overly complex models. Furthermore, when comparing models evaluated via different types of estimators, a model evaluated using a pseudo-biased HME may gain an unfair advantage.[5] Finally, pseudo-bias distorts Bayesian model averaging by assigning disproportionately large weights to more complex models, except in the knife-edge case where all models exhibit identical pseudo-bias.

Furthermore, I introduce a method to correct for simulation pseudo-bias, as the modified HME is not entirely immune to it, depending on the choice of weighting density. Building upon the correction method proposed in Lenk (2009) for the original HME, I propose truncating the integral outside the region defined by the empirical support of the posterior sampler and adjusting for this modification ex post. This explicit adjustment restricts the integral to regions where posterior samples are actually drawn. Consequently, it ensures that the absolute continuity condition holds in simulation, thereby eliminating the cause of the pseudo-bias.

More importantly, this pseudo-bias correction method provides a computational rationale for the truncation embedded in commonly used weighting densities. For instance, Geweke's estimator uses a truncated normal density (Geweke, 1999), while the estimator developed by Sims, Waggoner and Zha (2008, hereafter SWZ) incorporates truncation based on the quantiles of the empirical unnormalized posterior density. Yet, neither paper discusses the computational benefits of such truncation or its connection to simulation pseudo-bias. This paper argues that truncation is required not only from a theoretical perspective to ensure finite moments, but also from a computational standpoint to improve estimation accuracy by mitigating pseudo-bias.

[4]Pseudo-bias and computational efficiency represent distinct concepts. While computational efficiency refers to the number of simulated samples required to achieve convergence, pseudo-bias defines the error that persists even after the estimate appears to have converged.

[5]For instance, even if two models have identical true marginal likelihoods, if one is evaluated using an HME affected by pseudo-bias and the other is computed analytically, the former may be selected solely because its estimate is inflated by the bias.

Monte Carlo simulations based on Bayesian linear regression and small-scale New Keynesian dynamic stochastic general equilibrium (DSGE) models confirm that various modified HMEs exhibit less simulation pseudo-bias than the original HME. For example, the Geweke estimator, arguably the most popular in macroeconomics, and the SWZ estimator prove highly accurate, with the pseudo-bias correction yielding negligible improvements. Specifically, across all simulations conducted in this paper, their root-mean-square errors are "barely worth mentioning" on the scale of evidence proposed by Kass and Raftery (1995).[6] Nevertheless, the pseudo-bias correction method offers moderate accuracy improvements for certain modified HMEs, such as those employing a uniform weighting density, particularly as model complexity increases.

Given these findings, I revisit Smets and Wouters (2007), who demonstrated that their DSGE model fits US macroeconomic data comparably to reduced-form vector autoregressive (VAR) models. Their conclusion rests primarily on the fact that the marginal likelihood of their medium-scale DSGE model estimated using the Laplace approximation is comparable to those of the VAR models, which are obtained analytically. While the Laplace approximation is fundamentally a local method that relies solely on the unnormalized posterior density and the Hessian at the mode, HMEs provide a global approximation based on the entire posterior distribution. I demonstrate that their original conclusion remains robust even when various (pseudo-bias corrected) HMEs are applied. Furthermore, I underscore the critical importance of the pseudo-bias correction; without it, one would erroneously conclude that the DSGE model fits the data *much better* than the VAR models.

Related literature. This paper contributes to the literature on marginal likelihood estimation using HMEs.[7] The most related work is Lenk (2009), which shows that the original HME is pseudo-biased and proposes a correction method. This paper contributes to the literature by showing that the modified HME exhibits smaller simulation pseudo-bias than the original HME, and by verifying this claim through a series of simulations that include macroeconomic models. This paper also introduces a pseudo-bias correction method that is applicable to the broader class of modified HMEs, whereas the correction method of Lenk (2009) is limited to the original HME.

[6] Kass and Raftery (1995) classify the strength of evidence as "barely worth mentioning", "positive", "strong", and "very strong", depending on the difference in marginal likelihoods.

[7] Various other estimation methods have been proposed in the literature. For example, Chib (1995) and Chib and Jeliazkov (2001) present estimators based on a different equality for the marginal likelihood, known as the basic marginal likelihood identity. Recent works employ state-of-the-art statistical methods, such as conditional Monte Carlo, adaptive importance sampling, variational Bayes, and power posteriors, or estimate the marginal likelihood as a by-product of a new sampling method. For instance, see Chan (2023), Chan et al. (2024), Li et al. (2023), Pajor (2017), and Waggoner et al. (2016). For recent surveys of marginal likelihood estimation, see Ardia et al. (2012), Friel and Wyse (2012), and Llorente et al. (2023).

The findings of this paper are relevant to the more recent works in this literature that focus on developing new weighting densities using state-of-the-art statistical techniques. For instance, Hajargasht and Woźniak (2020) use variational Bayes posterior output; McEwen et al. (2021) use a kernel density or a Gaussian mixture density estimated by using posterior samples; Polanska et al. (2023) apply normalizing flows; Metodiev et al. (2024) use a uniform density over a spherical set chosen to minimize the variance of the estimator.[8] This paper focuses on the general properties of the modified HMEs, and the pseudo-bias correction method developed in this paper is widely applicable to these studies.

This paper also contributes to the extensive literature on model comparison in macroeconomics. Recent studies using both the Geweke and SWZ estimators or solely the SWZ estimator are Bianchi (2013), Chen et al. (2022), Davig and Doh (2014), Gemma et al. (2023), Hubrich and Tetlow (2015), Inoue and Shintani (2018), Liu et al. (2016) among others.[9] For instance, Chen et al. (2022) and Hubrich and Tetlow (2015) estimate Markov-Switching (MS) DSGE models and MS-VAR models, respectively, and apply the SWZ estimator to select the number of regimes.[10] Inoue and Shintani (2018) estimate New Keynesian Phillips curves and DSGE models and use both the Geweke and SWZ estimators for model selection. This paper contributes to the literature by providing further support for the use of these estimators by demonstrating their accuracy through the lens of simulation pseudo-bias and by running simulations using a DSGE model. Another technical contribution is to introduce the concept of simulation pseudo-bias into econometrics, a topic that has received limited attention so far.

Structure of the paper. Section 2 introduces the original and modified HMEs, presents the main proposition that the modified estimators exhibit smaller pseudo-bias, explains the importance-sampling interpretation of this proposition, and offers an illustrative example. Section 3 develops a pseudo-bias correction method for the modified HMEs and discusses its implications for commonly used weighting densities. Section 4 provides some examples of weighting densities. Section 5 confirms the theoretical predictions through Monte Carlo simulations, while Section 6 revisits Smets and Wouters (2007). Finally, Section 7 concludes by summarizing the findings and outlining directions for future research.

[8] Also see Raftery et al. (2007) and Wang et al. (2018). Fuentes-Albero and Melosi (2013) take a slightly different approach by leveraging a special structure of a model in which the conditional posterior density of a certain parameter block can be analytically computed.

[9] Geweke's estimator is also widely used in the macroeconomic literature, and its applications span various models, including heterogeneous agent New Keynesian and structural VAR models. Recent examples include Adjemian et al. (2024), Bayer et al. (2024), Brunnermeier et al. (2021), Bianchi and Nicolò (2021), Breitenlechner et al. (2022), and Lenza and Primiceri (2022).

[10] Chen et al. (2022) report that the estimates of Geweke and SWZ are similar. In contrast, Davig and Doh (2014) estimate MS-DSGE models and report that the SWZ estimator was unstable in their case, leading to the use of Geweke's estimator.

Appendix A contains the proofs of the propositions, and the Supplementary Appendix provides derivations, algorithms, and additional information.

2 Simulation pseudo-bias in the (modified) harmonic mean estimators of marginal likelihood

How do we compute the marginal likelihood in practice when its closed-form solution is not available? Equation (1) implies that the marginal likelihood can be computed through Monte Carlo integration by drawing samples from the prior distribution and taking an average of the likelihood evaluated at these samples. Such a naive estimator is, however, rarely used in practice due to its computational inefficiency, particularly for high-dimensional problems. This inefficiency arises because many samples contribute negligibly to the integration, as the prior distribution is typically much more dispersed than the likelihood function. This computational challenge of the naive Monte Carlo estimator led to the development of alternative estimators, such as the original and modified HMEs. In this section, I first explain HMEs and then introduce the concept of simulation pseudo-bias.

2.1 The original and modified harmonic mean estimators

The original HME (Newton and Raftery, 1994) is valued for its computational simplicity. The original HME is based on the following identity:

$$p(y)^{-1} = E_{p(\theta|y)}\left[\frac{1}{p(y|\theta)}\right], \tag{2}$$

where $E_{p(\theta|y)}[\cdot]$ denotes expectation with respect to the posterior distribution $p(\theta|y)$. Therefore, the naive finite-sample analogue for this equality is

$$\hat{p}_{HME}(y)^{-1} \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{1}{p(y|\theta^{(i)})} \tag{3}$$

where $\theta^{(i)}$ is a sample from $p(\theta|y)$, N is the number of samples, and $p(y|\theta^{(i)})$ is the likelihood evaluated at the ith posterior sample. As these likelihoods are typically computed within a posterior sampling algorithm, no additional computations are needed.

The modified HME (Gelfand and Dey, 1994) is a generalization of the original HME. It is well known in the literature (e.g., Chib and Jeliazkov (2001)) that its accuracy is comparable to that of more computationally intensive estimators. It is based on the following identity:

$$p(y)^{-1} = E_{p(\theta|y)}\left[\frac{w(\theta)}{p(y|\theta)p(\theta)}\right], \tag{4}$$

where $w(\theta)$ is a weighting density that satisfies $\int_\Theta w(\theta)d\theta = 1$.[11] Some examples of $w(\theta)$ are summarized in Section 4. Then, similarly to the original HME, its sample analogue leads to the modified HME as follows:

$$\hat{p}_{MHME}(y)^{-1} \equiv \frac{1}{N}\sum_{i=1}^{N}\frac{w(\theta^{(i)})}{p(y|\theta^{(i)})p(\theta^{(i)})}. \tag{5}$$

Note that the original HME is a special case of the modified HME, in which the weighting density is set to be the prior distribution.

The flexibility of the weighting density enhances its accuracy for two reasons. First, it improves computational efficiency by approximating the posterior distribution. Second, it ensures a finite variance when the tail of the weighting density is thinner than that of the unnormalized posterior, as discussed in Gelfand and Dey (1994); Chib (1995); Geweke (1999). Nevertheless, I argue that these are not the only factors contributing to its well-established superior accuracy; computational bias, referred to as simulation pseudo-bias by Lenk (2009), also plays a significant role.

By restricting the integration limits of equation (4) to $A \subseteq \Theta$, we can derive an identity regarding the modified HME that provides the foundation for the pseudo-bias comparison between the original and modified HMEs and the pseudo-bias correction method.[12]

Theorem 1 (A modified HME identity). *For any subset $A \subseteq \Theta$ with $W(A) \equiv \int_A w(\theta)d\theta > 0$ and $p(\theta|y) > 0$ almost everywhere in A, we have*

$$p(y) = P(A)\left[\int_A \frac{w(\theta)}{p(y|\theta)p(\theta)}p(\theta|y)d\theta\right]^{-1}. \tag{6}$$

Proof. Equation (6) follows from Bayes' theorem $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$ and simple algebra. See Appendix A for details. □

2.2 Simulation pseudo-bias in the harmonic mean estimators

Simulation pseudo-bias. Let us now formally define the posterior simulation support and simulation pseudo-bias of the HMEs. Following Lenk (2009), the posterior simulation support of the parameter θ is defined as

$$\underline{A} = \{\theta \subseteq \Theta : p(y|\theta) > \underline{L}\}, \text{ where } \underline{L} = \min_{\theta\in\{\theta^{(i)}\}_{i=1}^{N}} p(y|\theta). \tag{7}$$

[11] See Supplementary Appendix D for the derivation.

[12] A similar expression is derived for the original HME by Lenk (2009) and Pajor and Osiewalski (2013).

The simulation support should equal the whole parameter space Θ when the sampler runs infinitely long, but in practice $\underline{A}$ is strictly smaller than Θ due to the finite-sample nature of simulation.

Since both the original and modified HMEs are computed using the posterior samples, what they actually compute is not the right-hand-side values in equations (2) and (4), but those integrated under the simulation support $\underline{A}$. Therefore, the original (3) and modified (5) HMEs can be written in integral form, respectively, as follows:

$$\hat{p}_{HME}(y)^{-1} \approx \int_{\underline{A}} \frac{1}{p(y|\theta)} p(\theta|y) d\theta \tag{8}$$

and

$$\hat{p}_{MHME}(y)^{-1} \approx \int_{\underline{A}} \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y) d\theta. \tag{9}$$

The difference between the true marginal likelihood and the HMEs (8) (9) is defined as the simulation pseudo-bias:

$$PB_{HME}(y) \equiv \frac{\left[\int_{\underline{A}} \frac{1}{p(y|\theta)} p(\theta|y) d\theta\right]^{-1}}{\int_{\Theta} p(y|\theta)p(\theta) d\theta} \tag{10}$$

and

$$PB_{MHME}(y) \equiv \frac{\left[\int_{\underline{A}} \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y) d\theta\right]^{-1}}{\int_{\Theta} p(y|\theta)p(\theta) d\theta}. \tag{11}$$

The main proposition of this paper is that the modified HME exhibits smaller simulation pseudo-bias than the original HME.

Proposition 1 (Simulation pseudo-bias in the original and modified HMEs)**.** *If the probability of the simulation support evaluated by the weighting density* $W(\underline{A}) \equiv \int_{\underline{A}} w(\theta) d\theta$ *is larger than the prior probability of the simulation support* $P(\underline{A}) \equiv \int_{\underline{A}} p(\theta) d\theta$*, the simulation pseudo-bias of the modified HME is smaller than that of the original HME.*

Proof. Theorem 1 and the definition of pseudo-bias for the original (10) and modified (11) HMEs give

$$PB_{MHME}(y) = W(\underline{A})^{-1} < P(\underline{A})^{-1} = PB_{HME}(y) \tag{12}$$

by the assumption that $W(\underline{A}) > P(\underline{A})$, and the proposition follows. □

The assumption of $W(A) > P(A)$ is usually satisfied in practice. This is because the weighting density is parameterized with information from posterior samples, whereas the prior distribution is much more dispersed than the posterior. It is common to use information from posterior samples to specify the weighting density, because the modified HME equals the exact marginal likelihood when the weighting density matches the posterior density and gains computational efficiency when the overlap is large. Therefore, in practice, the modified HME is more robust to simulation pseudo-bias than the original HME.[13]

Proposition 1 provides a new explanation for the well-known accuracy of modified HMEs based on simulation studies. Specifically, the proposition shows that the accuracy of the modified HME is attributable not only to its computational efficiency and finite variance, as is known in the literature, but also to a smaller pseudo-bias when the weighting density is carefully constructed.

Furthermore, Proposition 1 implies that the pseudo-bias is positive, because $P(\underline{A})^{-1} \geq 1$ and $W(\underline{A})^{-1} \geq 1$ by the definition of probability. The mechanism of the positive pseudo-bias will be clearly illustrated in the importance-sampling interpretation explained next.

Importance sampling interpretation. Rewriting the HMEs as importance sampling clearly illustrates the source of simulation pseudo-bias.[14] Lenk (2009) argued that the original HME suffers from the pseudo-bias because the empirical support of the prior distribution is wider than that of the posterior density, i.e., the absolute continuity condition of importance sampling does not hold *in simulation*. More formally, equation (2) can be rewritten as importance sampling by noting that

$$\begin{aligned} E_{p(\theta|y)}\left[\frac{1}{p(y|\theta)}\right] &= \int_{\Theta} \frac{1}{p(y|\theta)} p(\theta|y) d\theta \\ &= p(y)^{-1} \int_{\Theta} \frac{p(\theta)}{p(\theta|y)} p(\theta|y) d\theta. \end{aligned} \tag{13}$$

Equation (13) implies that the original HME is a form of importance sampling, where the target distribution is the prior and the proposal distribution is the posterior. Then, the absolute continuity condition is violated in simulation because the prior distribution is more dispersed than the posterior, even though they have the same support in theory.

On the other hand, for the modified HME, equation (4) can also be expressed as

[13]Note that this proposition does not claim that the modified HME has smaller pseudo-bias for all possible weighting density $w(\theta)$. One could construct an extreme example, such as a weighting density whose support is disjoint from the posterior distribution. However, in practice, since we aim to approximate the posterior distribution with $w(\theta)$, such cases are irrelevant.

[14]See Supplementary Appendix D, as well as Chan et al. (2020) and references therein, for more details on importance sampling.

importance sampling:

$$E_{p(\theta|y)}\left[\frac{w(\theta)}{p(y|\theta)p(\theta)}\right] = \int_\Theta \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y)d\theta, \\ = p(y)^{-1}\int \frac{w(\theta)}{p(\theta|y)} p(\theta|y)d\theta. \tag{14}$$

The second term in the last equation indicates importance sampling, where the target and proposal distributions are the weighting and posterior densities, respectively. Since the empirical support of the weighting density, parameterized by information from posterior samples, is closer to the posterior density, the absolute continuity condition is more likely to hold in simulation. Consequently, the simulation pseudo-bias of the modified HME is small.

As implied in Proposition 1 and equations (13) and (14), simulation pseudo-bias is positive. Take the original HME as an example. The overestimation occurs because the simulated samples from the posterior distribution are concentrated in regions where $\frac{p(\theta)}{p(\theta|y)} < 1$ in (13) when the prior is more dispersed than the posterior, leading to an underestimation of the reciprocal of the marginal likelihood.

2.3 An illustrative example

A simple example illustrates the mechanism underlying simulation pseudo-bias and the accuracy of the modified HMEs. Suppose that we observe $T = 100$ scalar data $\{y_t\}_{t=1}^T$ generated i.i.d. from a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$, where σ^2 is known and μ is the parameter of interest. We put a normal prior distribution with mean $\mu_0 = 0$ and variance $V_0 = 2$ on μ. Since this prior is conjugate, the posterior distribution is also normal with mean $\mu_T = \frac{T\bar{y}/\sigma^2+\mu_0/V_0}{T/\sigma^2+1/V_0}$ and variance $V_T = (T/\sigma^2 + 1/V_0)^{-1}$, where $\bar{y} = T^{-1}\sum_{t=1}^T y_t$. In this case, the marginal likelihood can be computed analytically using the following formula:[15]

$$p(y) = (2\pi\sigma^2)^{-T/2}\left(\frac{V_T}{V_0}\right)^{1/2} \exp\left\{-\frac{1}{2}\left(\frac{\sum_{t=1}^T y_t^2}{\sigma^2} + \frac{\mu_0^2}{V_0} - \frac{\mu_T^2}{V_T}\right)\right\}. \tag{15}$$

Figure 1 presents the posterior distribution, weighting densities including the prior distribution, and the marginal likelihood estimates. The black marker (plus) in the right panel represents the estimate by the original HME, which overestimates the marginal likelihood due to the pseudo-bias. Recall that the HME is a modified HME using the prior distribution as the weighting density. As shown in the left panel, the prior distribution

[15] See Supplementary Appendix B.1 for the derivation of the posterior distribution and the marginal likelihood.

is more dispersed than the posterior distribution, which violates the absolute continuity condition of importance sampling in simulation. To understand the source of the pseudo-bias more mechanically, note that the approximation of the second term in equation (13) is likely to be less than unity, i.e., $\int_\Theta \frac{p(\theta)}{p(\theta|y)} p(\theta|y) d\theta \approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{p(\theta^{(i)})}{p(\theta^{(i)}|y)}\right] < 1$, since most samples $\theta^{(i)}$ from the proposal distribution lie in the area where $\frac{p(\theta^{(i)})}{p(\theta^{(i)}|y)} < 1$. As a result, the original HME overestimates the marginal likelihood, since the reciprocal is underestimated.

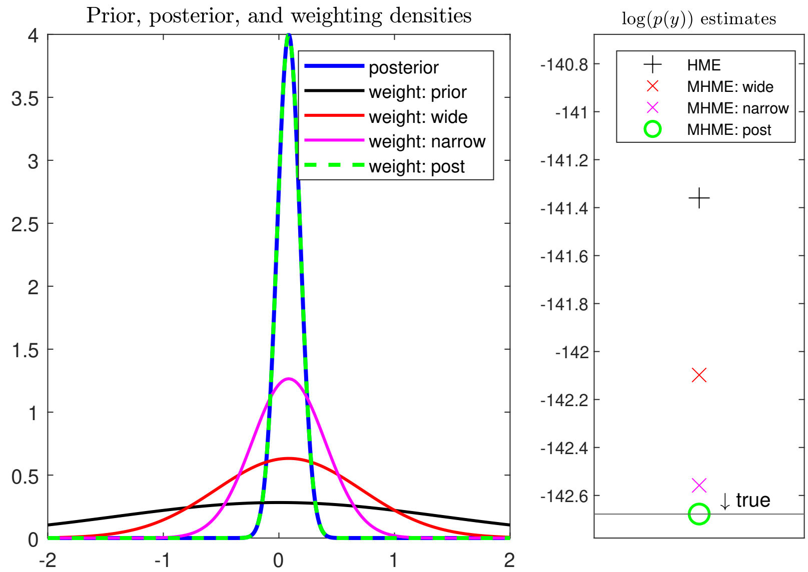


Figure 1: Densities and estimates of the marginal likelihood

Note: The left panel displays the posterior distribution and weighting densities. The black curve represents the prior distribution, which corresponds to the "weighting" density of the original HME. The red and magenta curves represent the weighting densities of the modified HMEs, with the former having a larger variance. The green dashed curve corresponds to the weighting density of the "exact" modified HME, where the weighting density matches the posterior distribution. The right panel shows the estimates, and the black line shows the true value computed by equation (15).

On the other hand, the estimates by the modified HME are closer to the true value, as shown by the red and magenta markers (cross) in the right panel. For these two HMEs, the weighting density is a normal distribution centered at μ_T with different variances: the red corresponds to a larger variance, and the magenta corresponds to a smaller variance. As the variance decreases, the weighting density approaches the posterior distribution, making the absolute continuity condition more likely to hold in simulation. Mechanically, the second term of equation (14) approaches the theoretical value of unity $\int \frac{w(\theta)}{p(\theta|y)} p(\theta|y) d\theta \approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{w(\theta^{(i)})}{p(\theta^{(i)}|y)}\right] \approx 1$ because samples from the posterior distribution are more likely to balance $\frac{w(\theta^{(i)})}{p(\theta^{(i)}|y)} > 1$ and $\frac{w(\theta^{(i)})}{p(\theta^{(i)}|y)} < 1$.

Finally, the estimate equals the true value, up to computation error, when the weighting density is set to the posterior density, as illustrated by the green marker (circle) in the right panel. This result follows from equation (14), as the modified HME recovers the exact marginal likelihood when the weighting density is set to the posterior distribution.

Note that using a weighting density that is more concentrated than the posterior is unproblematic from the perspective of simulation pseudo-bias, as it satisfies the absolute continuity condition. However, the concern now becomes computational efficiency: a larger number of samples is required for the estimator to converge because importance sampling approximates the integration using a more dispersed proposal distribution than the target. Once converged, however, there is no pseudo-bias for such estimators, whereas the HMEs that violate the absolute continuity condition in simulations still exhibit pseudo-bias even after convergence. Hence, there is a trade-off between pseudo-bias and computational efficiency. If sampling from the posterior distribution is more costly than evaluating the weighting density and implementing the pseudo-bias correction method proposed in the next section, it is better to use the correction method with a wider weighting density.

3 Pseudo-bias correction for the modified harmonic mean estimator

Although the modified HME has a smaller simulation pseudo-bias than the original HME, it is not entirely free of it and may still require a pseudo-bias correction method, depending on the specification of the weighting density. Proposition 2 provides a method to remove the pseudo-bias from the modified HME.

Proposition 2 (Simulation pseudo-bias correction). *The pseudo-bias-corrected modified HME can be computed by the following equality:*

$$p(y) = W(\underline{A}) \left[\int_{\underline{A}} \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y) d\theta \right]^{-1}. \tag{16}$$

The correction factor $W(\underline{A})$ can be estimated by, for instance, importance sampling:

$$\widehat{W}(\underline{A}) = \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}_{\underline{A}}(\theta^{(j)}) \frac{w(\theta^{(j)})}{q(\theta^{(j)})}, \tag{17}$$

where $\theta^{(j)}$ is a sample from the proposal density q and $\mathbb{1}_{\underline{A}}(a)$ is an indicator function taking the value of 1 if $a \in \underline{A}$ and 0 otherwise.

Proof. Putting $A = \underline{A}$ in Theorem 1 provides the equality (16). The estimation of $W(\underline{A})$ relies on importance sampling:

$$W(\underline{A}) = \int_{\underline{A}} w(\theta) d\theta = \int_{B} \mathbb{1}_{\underline{A}}(\theta) w(\theta) d\theta = \int_{B} \mathbb{1}_{A}(\theta) \frac{w(\theta)}{q(\theta)} q(\theta) d\theta,$$

where $q(\theta)$ denotes a proposal distribution with support B such that $A \subseteq B$. $\square$

Remark 1: The second term of (16) is not pseudo-biased because the limits of integration are restricted to the posterior simulation support. In other words, the absolute continuity condition in importance sampling always holds in the second term. Hence, the interpretation of this correction method is straightforward: first, restrict the limits of integration from Θ to its subset $\underline{A}$ such that the absolute continuity condition holds in simulation, and then adjust for this restriction by multiplying the scaling factor $W(\underline{A})$.

Remark 2: Proposition 2 is equivalent to the pseudo-bias correction method of Lenk (2009) in the case of the original HME. Replacing $w(\theta)$ with $p(\theta)$, we have

$$p(y) = P(\underline{A}) \left[\int_{\underline{A}} \frac{1}{p(y|\theta)} p(\theta|y) d\theta \right]^{-1},$$

which is equivalent to Equation 2.1 in Lenk (2009).

Remark 3: Theoretically, any subset of $\underline{A}$ removes the pseudo-bias. However, using $\underline{A}$ is computationally advantageous because the second term in (16) is equivalent to the uncorrected marginal likelihood estimated by the modified HME in (5), and hence, only the evaluation of $W(\underline{A})$ is required for the pseudo-bias correction.

Remark 4: When a direct sampling method is available for drawing samples from $w(\theta)$, $W(\underline{A})$ can also be calculated as the proportion of samples that lie within $\underline{A}$. This is indeed the approach of Sims et al. (2008), as explained in Supplementary Appendix A.

Discussion. This correction method has two implications for the weighting densities commonly used in practice. First, it provides a method to "robustify" a given weighting density. Specifically, once we choose a weighting density $\tilde{w}(\theta)$, the method provides another weighting density as $\tilde{\tilde{w}}(\theta) = 1_{\underline{A}}(\theta)\tilde{w}(\theta)/W(\underline{A})$ and it can be used as a weighting density in the modified HME in equation (5).

Second, and more importantly, it provides a theoretical rationale for some existing modified HMEs that incorporate truncation in their weighting densities. For instance, Geweke's estimator uses the truncated normal distribution as a weighting density. Furthermore, the SWZ estimator truncates the parameter space where the unnormalized posterior density falls below a specified quantile of the realized values. Hence, the SWZ estimator is already "robustified" against pseudo-bias. Both Geweke (1999); Sims et al. (2008) discuss the reason for truncation from the perspective of the absolute continuity condition in the population, which is a necessary condition for the modified HME to have finite moments (Gelfand and Dey, 1994), but not in simulation. Hence, the correction method (16) provides a formal computational rationale for the truncations commonly used in practice.

4 Examples of the modified harmonic mean estimators

In this section, I review three types of weighting densities that are commonly used in the literature and studied in the Monte Carlo simulations in Section 5 and the empirical application in Section 6.[16] The first two estimators, proposed by Geweke (1999) and Sims et al. (2008), are most frequently used in macroeconomic applications. Hence, studying their accuracy provides valuable insights into the macroeconomic literature. The other estimator is a modified HME using a uniform distribution as the weighting density. Although this weighting density offers a poor approximation of the posterior distribution, it serves as a useful benchmark in two respects. First, comparing it with the original HME allows us to assess how robust the modified HME is even under a poor approximation, thereby providing an upper bound on its sensitivity to pseudo-bias. Second, comparing it with the estimators proposed by Geweke (1999) and Sims and Zha (1998) highlights the importance of carefully selecting the weighting density.

Geweke's estimator. Geweke (1999) uses a truncated multivariate normal density as a weighting density:

$$w_{Geweke}(\theta) = \tau^{-1}(2\pi)^{-k/2}|\bar{\Omega}|^{-1/2} \exp\left(-\frac{1}{2}(\theta-\bar{\theta})'\bar{\Omega}^{-1}(\theta-\bar{\theta})\right) \mathbb{1}(\theta \in \bar{\Theta}),$$

where $\bar{\Theta} = \{\theta : (\theta-\bar{\theta})'\bar{\Omega}^{-1}(\theta-\bar{\theta}) \leq F^{-1}_{\chi^2_k}(\tau)\}$, $\bar{\theta}$ and $\bar{\Omega}$ are the posterior mean and covariance matrix of θ calculated by the posterior samples, k is the dimension of θ, $\tau \in (0,1)$ is a tuning parameter, and $F^{-1}_{\chi^2_k}(\tau)$ is an inverse of a χ^2 CDF with k degrees of freedom. I use $\tau = 0.9$ in the simulation. This estimator is particularly convenient because it only requires the output from the MCMC algorithm, such as the posterior samples and their corresponding unnormalized posterior densities.

Sims, Waggoner, and Zha's estimator. Sims et al. (2008) proposed a weighting density function $w_{SWZ}(\theta)$ to handle the non-Gaussian posterior distribution with potential multimodality by extending Geweke's estimator in three aspects. First, the SWZ estimator centers the distribution at the posterior mode, whereas Geweke's estimator centers at the posterior mean. Second, its weighting density is a more general form of an elliptical density function $g(\theta)$, while Geweke's weighting density is a truncated multivariate normal density. Third, it introduces an additional truncation criterion based on the unnormalized posterior density.

[16] For detailed algorithms to compute these estimators, see Supplementary Appendix A.

Formally, they propose the following weighting density:

$$w_{SWZ}(\theta) = q_L^{-1} g(\theta) \mathbb{1}(\theta \in \widehat{\Theta})$$

where

$$g(\theta) = \frac{\Gamma(k/2)}{2\pi^{k/2}|\widehat{S}|} \frac{f(r(\theta))}{r(\theta)^{k-1}}, \tag{18}$$

$r(\theta) = \sqrt{(\theta - \hat{\theta})'\widehat{\Omega}^{-1}(\theta - \hat{\theta})}$, $\hat{\theta}$ is a posterior mode, $\widehat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}(\theta^{(i)} - \hat{\theta})(\theta^{(i)} - \hat{\theta})'$, $\widehat{S}$ is the lower triangle of Cholesky decomposition of $\widehat{\Omega}$, $f(\cdot)$ is any one-dimensional probability density defined on the positive reals. Their truncation criterion is based on two conditions: $\widehat{\Theta} = \{\theta : p(y|\theta)p(\theta) > L_{1-q}, r(\theta) \in [a, b]\}$ where L_{1-q} is the$1 - q$-th percentile of the realized unnormalized posterior density $\{p(y|\theta^i)p(\theta^i)\}_{i=1}^{N}$ and Sims et al. (2008) recommend using $q = 0.9$. a, b, and ν are chosen to truncate approximately 10% of the tails of the posterior distribution, with their formal definitions provided below. q_L is the normalizing constant ensuring that $\int_\Theta w_{SWZ}(\theta)d\theta = 1$.

Since the elliptical distribution in equation (18) only requires $f(r)$ to be a one-dimensional probability density function defined on positive reals, Sims et al. (2008) use the following parametric form:

$$f(r) = \begin{cases} \frac{\nu r^{\nu-1}}{b^\nu - a^\nu} & \text{if } r \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

where a, b and ν are parameters defined as follows. Let c_1, c_{10} and c_{90} be the 1st, 10th, and 90th percentile values of the posterior draws $\{r^{(i)}\}_{i=1}^{N}$ computed as $r^{(i)} = \sqrt{(\theta^{(i)} - \hat{\theta})'\widehat{\Omega}^{-1}(\theta^{(i)} - \hat{\theta})}$. Then, set $\nu = \frac{\log(1/9)}{\log(c_{10}/c_{90})}$, $a = c_1$ and $b = \frac{c_{90}}{0.9^{\frac{1}{\nu}}}$ so that $r \in [a, b]$ contains approximately 90% of the posterior samples.

Computing the normalizing constant q_L requires additional simulations. In this paper, I use the direct sampling algorithm proposed by Sims et al. (2008) to draw from $g(\theta)$, supplemented by inversion sampling to draw from $f(r)$. Then, q_L can be evaluated as the proportion of samples from $g(\theta)$ that lie in $\widehat{\Theta}$.[17] Supplementary Appendix A.3 provides a detailed algorithm to implement the SWZ estimator, including the computation of q_L.[18]

Note that the SWZ estimator already incorporates pseudo-bias correction, as it truncates the parameter space based on the realized unnormalized posterior density, which

[17] I also applied importance sampling to evaluate q_L, but my computational experience suggests that direct sampling is computationally more stable and accurate than importance sampling.

[18] As far as I know, a complete algorithm of the SWZ estimator is not available in the literature. For instance, the description in Sims et al. (2008) omits the algorithm to sample from $f(r)$ and is somewhat complicated, while Herbst and Schorfheide (2016) assumes that we can sample from $g(\theta)$ without explaining how such a sampling method can be implemented. The algorithm provided in Supplementary Appendix A.3 should address these gaps and make the SWZ estimator more accessible to researchers.

forms the first condition in $\widehat{\Theta}$. The authors justify truncation as a way to prevent the likelihood from approaching zero in the interior of the parameter space, which can make estimates unreliable. In contrast, this paper offers a different perspective: such truncation criteria make the estimator less affected by simulation pseudo-bias.

A uniform weighting density. We can also use a much simpler probability distribution as a weighting density, such as a uniform distribution.[19] Specifically, I consider a uniform density whose support is defined on a hyperrectangle based on the posterior samples, $\times_k[\min_{i=1}^N(\theta_k^{(i)}), \max_{i=1}^N(\theta_k^{(i)})]$, with some truncation of the edges. In this paper, I truncate 10% from each edge of the rectangle. This 10% truncation is chosen to illustrate the sensitivity to simulation pseudo-bias. In practice, we could truncate more to mitigate pseudo-bias, but note that this is not a free lunch. Larger truncation reduces the estimator's computational efficiency, requiring more samples from the posterior distribution for convergence. Therefore, this trade-off depends on the computational cost of implementing the pseudo-bias correction and the cost of generating samples from the posterior distribution.

5 Monte Carlo simulation evidence

In this section, I conduct two types of Monte Carlo simulations to investigate the accuracy and sensitivity to pseudo-bias of the modified HMEs in a practical setup. The first simulation is based on Bayesian linear regression models, and I change the number of observations and regressors to examine how model complexity affects pseudo-bias and estimation accuracy. I set a conjugate prior distribution so that the marginal likelihood can be computed analytically and used as a benchmark. The second simulation is based on small-scale New Keynesian DSGE models to study macroeconomic applications. Since no analytical formula for the marginal likelihood is available in this case, I apply a sequential Monte Carlo (SMC) sampler, which provides an estimate of the marginal likelihood as a by-product of the sampling procedure.[20]

[19] Lenk (2009) uses this type of weighting density as a poor approximation of the posterior distribution in its Monte Carlo simulation, and Metodiev et al. (2024) proposes a new estimator extending the uniform weighting density.

[20] I checked that the SMC estimate of the marginal likelihood is accurate even in a small-scale multimodal state-space model whose marginal likelihood can be well-approximated by naive Monte Carlo integration. For a survey of SMC in economics, see Creal (2012).

5.1 Bayesian linear regression models

Model. Suppose that data $\{y_t\}_{t=1}^T$ is generated by the following data-generating process:

$$y_i = X_i'\beta + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

for $i = 1, 2, ..., T$, or by stacking them,

$$y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I_T),$$

where β is a k-dimensional vector of regression coefficients and X is a $T \times k$ matrix of regressors. We set the following conjugate prior distributions on β and σ^2:

$$\beta|\sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 V_0), \quad \sigma^2 \sim IG(v_0, v_1).$$

Then, the (full conditional) posterior distributions and the marginal likelihood can be derived as follows:[21]

$$\beta|\sigma^2, y, X \sim \mathcal{N}(\beta_T, \sigma^2 V_T), \quad \sigma^2|y, X \sim IG(\tilde{v}_0, \tilde{v}_1)$$

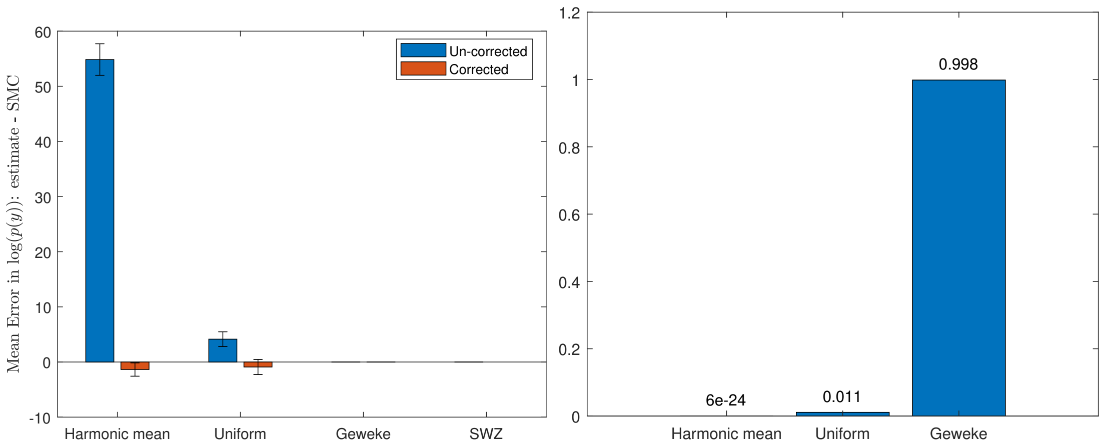
where $\tilde{v}_0 = T/2 + v_0$, $\tilde{v}_1 = \left[\frac{1}{v_1} + \frac{1}{2}(y'y + \beta_0' V_0^{-1}\beta_0 - \beta_T' V_T^{-1}\beta_T)\right]^{-1}$, $V_T = (X'X + V_0^{-1})^{-1}$, $\beta_T = (X'X + V_0^{-1})^{-1}(X'y + V_0^{-1}\beta_0) = V_T(X'y + V_0^{-1}\beta_0)$, and

$$p(y|X) = (2\pi)^{-T/2} \frac{|V_T|^{1/2}}{|V_0|^{1/2}} \frac{\Gamma(\tilde{v}_0)(\tilde{v}_1)^{\tilde{v}_0}}{\Gamma(v_0)(v_1)^{v_0}}.$$

Simulation setup. In each simulation, the true values of β and σ^2 are sampled from their prior distributions, and each element of X is randomly sampled from $\mathcal{N}(0, 1)$. I use different pairs of the number of observations (T) and the number of regressors (n_x) to evaluate the effect of model complexity: $(T, n_x) = (25, 3), (100, 3), (100, 10), (100, 20), (100, 40), (200, 100)$, following Pajor (2017). The parameters of the prior distribution are set as $\beta_0 = 0, V_0 = 7I_k, v_0 = 3, v_1 = 2.5$. I use the Gibbs sampler to draw $45,000$ samples from the posterior distribution, discarding the first $5,000$ samples as burn-in. I run 160 simulations and compare the estimates of various HMEs with the analytical marginal likelihood. The SWZ estimator requires a posterior mode, which can be obtained analytically in this case: the posterior modes of β and σ^2 are β_T and $\frac{1/\tilde{v}_1}{\tilde{v}_0+1}$, respectively.

[21] See Supplementary Appendix B.2, as well as Chan et al. (2019), for the derivation of the posterior distribution and the marginal likelihood.

Simulation results. The Monte Carlo simulation confirms the main proposition of the paper that the modified HME has a smaller simulation pseudo-bias.[22] Figure 2 presents the simulation results for a moderate-sized Bayesian linear regression model with 100 observations and 20 regressors, and shows that the accuracy improvement provided by the pseudo-bias correction is much smaller for the modified HMEs than for the original HME. Specifically, Figure 2a shows that pseudo-bias correction reduces the mean error of the original HME from 55 to almost zero, while the largest improvement among the modified HMEs is observed for the uniform weighting density, and the reduction is only about 4.[23] Geweke's estimator is already accurate without pseudo-bias correction, and the correction barely improves the accuracy. Note that there is no pseudo-bias-corrected SWZ estimator, as it implicitly incorporates pseudo-bias correction by construction.



(a) Mean errors

(b) The mean and CV of $W(A)$

Figure 2: The Bayesian linear regression model with $T = 100, n_x = 20$

Note: This figure summarizes the results of the Monte Carlo simulation exercise based on the Bayesian linear regression model with $T = 100, n_x = 20$, using results from 160 simulations. Harmonic Mean is the original harmonic mean estimator; Uniform is the modified harmonic mean estimator with the uniform weighting density; Geweke is the estimator proposed by Geweke (1999); and SWZ is the estimator proposed by Sims, Waggoner and Zha (2008). Figure 2a reports the mean error, computed as (estimate - true value), along with its one-standard-deviation band. *Corrected* indicates the pseudo-bias-corrected estimate obtained using method (16). Figure 2b presents the mean probability of the posterior simulation support evaluated by the weighting density w estimated through (17).

Nonetheless, pseudo-bias correction improves certain modified HMEs, implying that they are not entirely free from pseudo-bias. Specifically, the pseudo-bias of the modified HME with the uniform weighting density is classified as "strong" on the scale of Kass and Raftery (1995) as shown in Table 1. We can also observe the degree of pseudo-bias

[22]Supplementary Appendix C shows scatter plots of the errors and confirms the findings explained in the section, both for the regression and DSGE models.

[23]Given the high accuracy obtained by the pseudo-bias corrected original HME, some readers may conclude that it should always be preferred over modified HMEs because it does not require a specific choice of weighting density. However, this reasoning is not true, as the HME is not agnostic about the choice of the weighting density. Rather, it implicitly assumes that the weighting density is the prior distribution.

from a different perspective by examining the scaling factor for pseudo-bias correction, $W(A)$, which measures the probability of posterior simulation support evaluated by the weighting density. The closer the scaling factor is to unity, the less the estimator is affected by the simulation pseudo-bias. Figure 2b shows that the modified HMEs have larger mean estimates of the scaling factor than the original HME. For instance, Geweke's estimator has the largest scaling factor, which is almost unity, whereas the original HME has a value close to zero. The scaling factor for the modified HME with the uniform weighting density is 0.011, indicating its moderate sensitivity to pseudo-bias.

2 ln($\mathbf{B_{1,2}}$)	**Grades of Evidence**
0 to 2	Barely worth mentioning
2 to 6	Positive
6 to 10	Strong
> 10	Very strong

Table 1: Kass and Raftery (1995) scale

Note: This table summarizes the interpretation of evidence strength proposed by Kass and Raftery (1995), based on differences in marginal likelihoods when comparing two models.

Similar to Lenk (2009), we observe that more complex models tend to generate larger pseudo-bias, while the modified HMEs have smaller pseudo-bias than the original HME across all specifications. Table 2 summarizes the simulation results for various combinations of the number of observations and regressors. For all specifications, pseudo-bias correction improves the accuracy of the modified HMEs much less than that of the original HME, implying that the modified HME is less affected by the pseudo-bias. The mean of the estimated $W(A)$ also confirms this observation, because the $W(A)$ of the original HME is much smaller than that of the modified HMEs. Turning to the effect of model complexity on pseudo-bias, as the number of regressors grows, pseudo-bias increases for all estimators except Geweke's. For instance, while the pseudo-bias of the modified HME with the uniform weighting density is classified as "positive" with ten regressors, it becomes "strong" with twenty regressors and "very strong" with more than forty.

$(T,\, n_x)$		HM	C-HM	Unif	C-Unif	Geweke	C-Geweke	SWZ
(25, 3)	ME	5.48	-0.60	0.11	-0.66	0.00	-0.00	-0.00
	Std	1.37	0.70	0.33	0.42	0.01	0.01	0.00
	RMSE	5.65	0.92	0.35	0.79	0.01	0.01	0.00
	mean: $P(A)$	—	4e-03	—	0.49	—	1.00	—
(100, 3)	ME	8.35	-0.49	0.08	-0.24	-0.00	-0.00	-0.00
	Std	1.40	0.65	0.42	0.43	0.00	0.00	0.00
	RMSE	8.46	0.81	0.42	0.49	0.00	0.00	0.00
	mean: $P(A)$	—	3e-04	—	0.74	—	1.00	—
(100, 10)	ME	27.10	-0.95	1.30	-0.49	-0.00	-0.00	-0.00
	Std	2.33	0.98	0.83	0.85	0.00	0.00	0.00
	RMSE	27.20	1.36	1.54	0.98	0.00	0.01	0.00
	mean: $P(A)$	—	3e-12	—	0.19	—	1.00	—
(100, 20)	ME	54.85	-1.37	4.14	-0.91	-0.01	-0.01	-0.00
	Std	2.86	1.21	1.34	1.37	0.00	0.00	0.01
	RMSE	54.93	1.82	4.35	1.64	0.01	0.01	0.01
	mean: $P(A)$	—	6e-24	—	0.01	—	1.00	—
(100, 40)	ME	109.91	-2.05	14.12	-1.77	-0.02	-0.03	-0.02
	Std	3.45	1.74	1.91	2.19	0.02	0.02	0.01
	RMSE	109.96	2.69	14.25	2.82	0.03	0.03	0.02
	mean: $P(A)$	—	3e-47	—	3e-07	—	1.00	—
(200, 100)	ME	316.91	-1.58	50.19	0.07	-0.13	-0.34	-0.12
	Std	5.59	1.07	2.78	2.88	0.01	0.23	0.01
	RMSE	316.96	1.90	50.27	2.87	0.13	0.41	0.12
	mean: $P(A)$	—	1e-133	—	1e-19	—	0.83	—

Table 2: Summary of the Monte Carlo simulation using Bayesian linear regression models

Note: This table summarizes the mean error (ME, approximation - true), standard deviation of the error (Std), root mean squared error (RMSE), and the mean estimate of $P(A)$. T and n_x are the number of observations and regressors, respectively. HM is the original harmonic mean estimator; Unif is the modified harmonic mean estimator with uniform weighting density; Geweke is the estimator proposed by Geweke (1999); SWZ is the estimator proposed by Sims, Waggoner and Zha (2008). The prefix C indicates that the estimate is simulation pseudo-bias corrected. These values are based on 160 simulations.

Finally, Geweke's estimator remains accurate without pseudo-bias correction across all specifications. This is because the joint posterior density is nearly Gaussian: the posterior distribution of the regression coefficient is Gaussian, whereas only the posterior of the error variance follows an inverse-gamma distribution in this model. The SWZ estimator is also accurate for the same reason.

5.2 A small-scale New Keynesian dynamic stochastic general equilibrium model

The next simulation exercise employs a textbook-style three-equation New Keynesian DSGE model, consisting of the Euler equation, the hybrid New Keynesian Phillips curve, and the inertial Taylor rule with demand, supply, and monetary policy shocks.[24] This

[24]See, for instance, Walsh (2017) and Galí (2015) for a textbook treatment of this class of models.

model is relatively small compared to other workhorse DSGE models used in academic and policy studies, such as the one in Smets and Wouters (2007). Still, it captures the fundamental economic mechanisms common to most DSGE models, and its relatively small size is advantageous for Monte Carlo simulations, as we need to repeatedly estimate the model and compute the marginal likelihood.[25]

Model. We consider the following three-equation NK-DSGE model:

$$\begin{aligned} y_t &= E_t y_{t+1} - \gamma(i_t - E_t \pi_{t+1}) + e_t^d \\ \pi_t &= \gamma_b \pi_{t-1} + (1-\gamma_b) E_t \pi_{t+1} + \kappa y_t + e_t^s \\ i_t &= \rho_i i_{t-1} + \phi_\pi \pi_t + \phi_y y_t + e_t^m \end{aligned} \tag{19}$$

where $e_t^d = \rho_d e_{t-1}^d + \sigma_d \varepsilon_t^d, e_t^s = \sigma_s \varepsilon_t^s, e_t^m = \sigma_m \varepsilon_t^m$, $\beta \in (0,1), \kappa > 0, \phi_\pi > 1, \phi_y \geq 0$, $\rho_d, \gamma_b, \rho_i \in [0,1]$, and $\sigma_d, \sigma_s, \sigma_m > 0$. I assume $\varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^m \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. The solution of this model provides the state transition equation:

$$S_t = \Phi S_{t-1} + G \varepsilon_t \tag{20}$$

where $S_t = [y_t, \pi_t, i_t, e_t^d, e_t^s, e_t^m]'$ and $\varepsilon_t = [\varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^m]'$. The observation equation is

$$Y_t^o \equiv \begin{bmatrix} y_t^o \\ \pi_t^o \\ i_t^o \end{bmatrix} = H S_t + \Sigma_u^{1/2} u_t \tag{21}$$

where $u_t \overset{\text{iid}}{\sim} \mathcal{N}(0, I)$, Σ_u is the variance-covariance matrix of the observation error, and the specific structure of Φ, G and H is summarized in Supplementary Appendix B.3. Equations (20) and (21) provide a linear Gaussian state-space representation of the model, and I use the Kalman filter to evaluate its likelihood. The prior distributions for the parameters are standard and summarized in Table 3.

[25] For example, using a similar model, Cai et al. (2021) and Angelini et al. (2022) evaluate the performance of their SMC algorithm and bootstrap diagnostic tests, respectively, through Monte Carlo simulations.

Parameter	Description	Distribution	Mean	Std	Domain
γ	Elasticity of intertemporal substitutiton	Normal	1.5	0.375	$(0, \infty)$
ϕ_π	Taylor rule coefficient on inflation	Normal	1.5	0.25	$(1, \infty)$
ϕ_y	Taylor rule coefficient on output	Normal	0.12	0.05	$[0, \infty)$
κ	Slope of the NKPC	Gamma	0.1	0.1	$(0, \infty)$
σ_d	Std of a demand shock	Inv-Gamma	0.5	1	$(0, \infty)$
σ_s	Std of a supply shock	Inv-Gamma	0.5	1	$(0, \infty)$
σ_m	Std of a monetary policy shock	Inv-Gamma	0.5	1	$(0, \infty)$
ρ_d	Persistence of demand shock	Beta	0.5	0.2	[0,1]
γ_b	Inflation inertia	Beta	0.5	0.2	[0,1]
ρ_i	Taylor rule inertia	Beta	0.5	0.2	[0,1]

Table 3: Prior distribution for the three-equation New Keynesian models

Note: This table summarizes the prior distribution of the parameters in the three-equation New Keynesian DSGE model. Std denotes the standard deviation.

I consider two variants of the model in the simulation to study the effect of model complexity. One is the *simple* model that has no inertia, i.e., $\rho_d = \gamma_b = \rho_i = 0$. This model is simple, so the solution can be obtained analytically as shown in the Supplementary Appendix B.3. The other is the *extended* model, which is the full model in (19), and a requires numerical method, such as that of Uhlig (1995), to solve it.

When sampling from the posterior distribution, I use the sequential Monte Carlo (SMC) sampler. This SMC approach is preferred because it estimates the marginal likelihood as a by-product of the sampling, even though an analytical formula is not available for the model.[26] Specifically, following the algorithm described in Herbst and Schorfheide (2014) and Herbst and Schorfheide (2016), the estimate of the marginal likelihood by the SMC sampler is computed as follows:

$$\hat{p}_{SMC}(y) = \prod_{n=1}^{N_\phi} \left(\sum_{i=1}^{N} \tilde{w}_n^i W_{n-1}^i \right), \tag{22}$$

where $\tilde{w}_n^i$ is the unnormalized weight at the correction step, and W_n^i is the normalized weight at the selection step. I use this estimate as a benchmark when comparing the estimation accuracy of the HMEs.

Simulation setup. In the simulation, I generate a series of data $\{y_t, \pi_t, i_t\}_{t=1}^{T=160}$ from the state-space model (20) and (21), assuming that the initial state is at the steady state. The true parameters are drawn from their prior distributions, whose standard

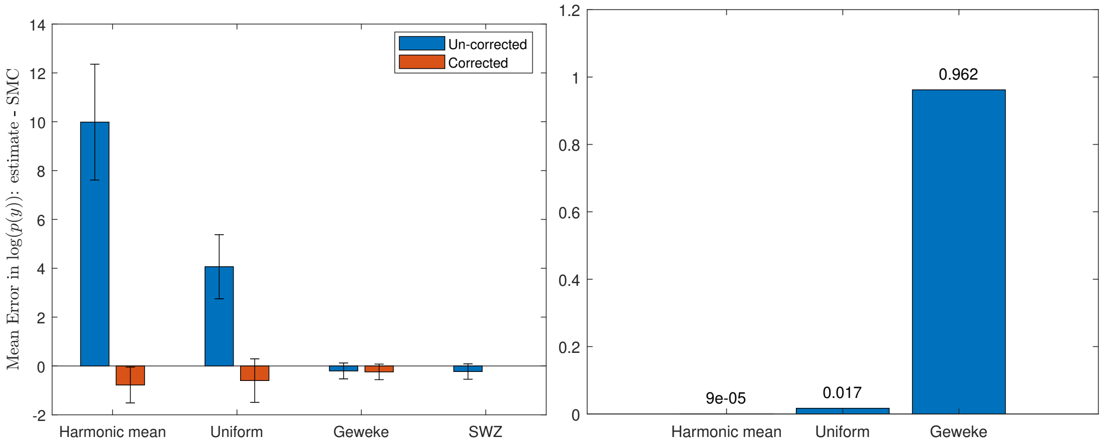
[26]The analytical formula is unavailable even for the simple model, despite the use of standard prior distributions such as the normal or inverse-gamma distributions. Note that the model is non-linear in parameters, even though it remains linear in states. Such nonlinearity eliminates the desirable property of standard prior distributions that allows parameters to be integrated out of the joint probability density. For a concrete example of the coefficient matrices, see equations (B.2) and (B.3) in Supplementary Appendix B.3.

deviations are deflated by a factor of 2, and I assume there is no observation error. The simulation is repeated 160 times. I use the SMC sampler to draw samples from the posterior distribution.[27] When computing the SWZ estimator, I obtain a posterior mode by maximizing the unnormalized log posterior density. Specifically, I iterate the optimization 100 times, setting the initial value to a random sample from the posterior particles. The result with the largest unnormalized log posterior density is selected as the posterior mode.

Simulation results. The simulation results from the DSGE models generally align with those from the Bayesian linear regression models and confirm that the modified HME is less affected by the pseudo-bias.

Figure 3 illustrates the simulation results using the extended DSGE model. Figure 3a shows that the modified HMEs are less affected by simulation pseudo-bias than the original HME, as the pseudo-bias correction improves their accuracy less than it does for the original HME. Specifically, the pseudo-bias correction reduces the mean error of the original HME from 10 to nearly zero, while the improvement for the modified HMEs is much smaller. Again, note that there is no pseudo-bias-corrected SWZ estimator, as it implicitly incorporates pseudo-bias correction by construction.

[27] In the SMC sampling, I generate 5120 particles, and the number of stages is 50. I use a non-linear tempering schedule, as described in Herbst and Schorfheide (2016), with a tempering parameter of $\lambda = 2$. I resample to avoid weight degeneracy issues when the effective sample size drops below half its initial size. The Metropolis-Hastings step is repeated twice during the mutation, and I use a Gaussian proposal distribution. The targeted acceptance ratio is set to 0.25, and the scaling parameter and the variance-covariance matrix of the proposal distribution are adaptively adjusted. For a more detailed explanation of the implementation, see Supplementary Appendix B.3.



(a) Mean errors

(b) Mean $W(A)$

Figure 3: The extended three-equation NK-DSGE model

Note: This figure summarizes the results of the Monte Carlo simulation exercise based on the extended three-equation NK-DSGE model of (19), using results from 160 simulations. Harmonic Mean is the original harmonic mean estimator; Uniform is the modified harmonic mean estimator with the uniform weighting density; Geweke is the estimator proposed by Geweke (1999); and SWZ is the estimator proposed by Sims, Waggoner and Zha (2008). Figure 3a reports the mean error, computed as (estimate - true value), along with its one-standard-deviation band. *Corrected* indicates the pseudo-bias-corrected estimate obtained using method (16). Figure 3b presents the mean probability of the posterior simulation support evaluated by the weighting density w estimated through (17).

Nonetheless, pseudo-bias correction improves the modified HME with the uniform weighting density, implying that the modified HME is not completely free from the pseudo-bias. Specifically, its mean error is reduced by about 4 with pseudo-bias correction, suggesting that the pseudo-bias is "positive" on the scale of Kass and Raftery (1995) shown in Table 1. Figure 3b also shows that it is moderately pseudo-biased because its mean $W(A)$ is far from unity, while that of Geweke's estimator is approximately unity.

The simulation also confirms that the pseudo-bias increases as the model becomes more complex. Table 4 shows that the pseudo-bias is larger for the extended model than the simple model. For instance, the pseudo-bias for the original HME is approximately 6 in the simple model and 10 in the extended model. Similarly, for the modified HME with the uniform weighting density, the error is "strong" in the extended model, whereas it is "positive" in the simple model, on the scale of Kass and Raftery (1995). This finding is also evident in the mean estimate of $W(A)$ because it is larger in the simple model than in the extended model for both the original and the modified HME with the uniform weighting density.

Model		HM	C-HM	Unif	C-Unif	Geweke	C-Geweke	SWZ
Extended	ME	9.98	-0.78	4.06	-0.60	-0.20	-0.24	-0.23
	Std	2.37	0.73	1.31	0.89	0.32	0.32	0.32
	RMSE	10.26	1.07	4.27	1.07	0.38	0.40	0.39
	mean: $P(A)$	—	9e-05	—	0.02	—	0.96	—
Simple	ME	6.20	-0.56	2.29	-0.50	-0.04	-0.11	-0.11
	Std	1.53	0.61	1.00	0.76	0.14	0.13	0.13
	RMSE	6.39	0.82	2.50	0.90	0.15	0.18	0.18
	mean: $P(A)$	—	2e-03	—	0.08	—	0.93	—

Table 4: Summary of the Monte Carlo simulation using the small-scale DSGE models

Note: This table summarizes the mean error (ME, approximation - SMC), standard deviation of the error (Std), root mean squared error (RMSE), and the mean estimate of $P(A)$ for each approximation method. HM is the original harmonic mean estimator; Unif is the modified harmonic mean estimator with uniform weighting density; Geweke is the estimator proposed by Geweke (1999); SWZ is the estimator proposed by Sims, Waggoner and Zha (2008). The prefix C indicates that the estimate is simulation pseudo-bias corrected. *Extended* is the full three-equation NK-DSGE model with inertia in Equation (19), and *Simple* is the version without inertia, i.e., $\rho_d = \gamma_b = \rho_i = 0$.

Finally, Table 4 shows that both Geweke's estimator and the SWZ estimator are accurate for both the simple and extended models. The superior performance of Geweke's estimator can be attributed to the standard shape of the posterior distribution of the model's parameters, as shown in Figure 4. The posterior distribution looks almost Gaussian and is unimodal. For these standard posterior distributions, the truncated Gaussian weighting density of Geweke's estimator approximates the posterior distribution well, resulting in an accurate estimate of the marginal likelihood. The SWZ estimator is also accurate for the same reason.[28] Furthermore, Herbst and Schorfheide (2014) show that Geweke's estimator provides a similar estimate to the SMC approximation for the DSGE model of Schmitt-Grohé and Uribe (2012) under the original prior distribution, although it tends to be slightly more volatile than the SMC estimate. The accuracy of Geweke's estimator observed in my simulations is consistent with their findings, and my results further indicate that the SWZ estimator is similarly accurate. This result supports the use of Geweke's estimator, which is arguably the most widely used in macroeconomics, as well as the SWZ estimator.

[28] If the posterior distribution is non-standard and multimodal, the SWZ estimator may outperform Geweke's estimator. Sims et al. (2008) argue that Geweke's weighting density provides a poor approximation for multimodal, non-Gaussian posteriors, which are often observed in Markov-Switching VAR models.

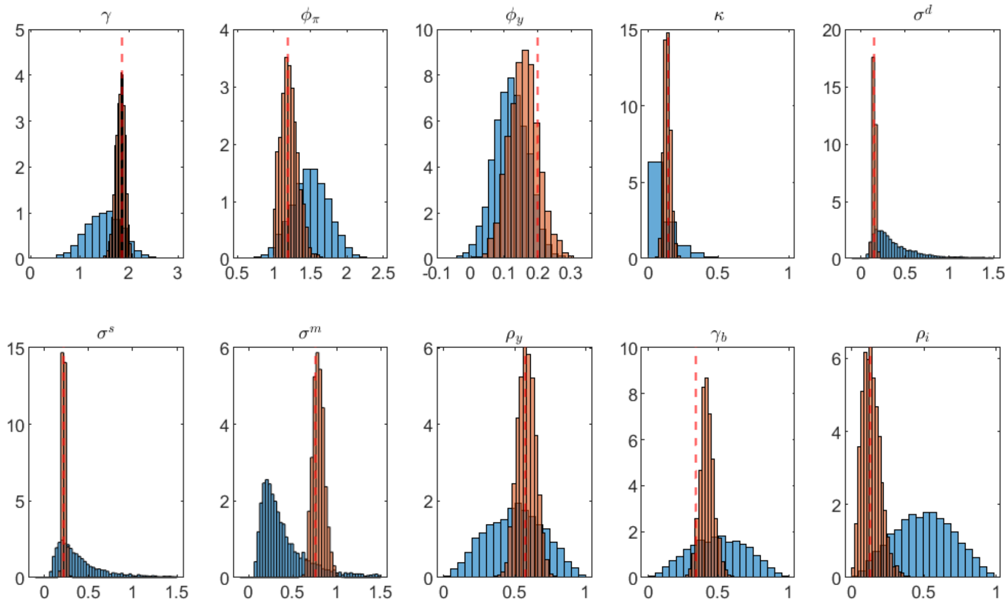


Figure 4: Posterior distribution in one simulation

Note: This figure displays posterior distributions from one of the Monte Carlo simulations. The orange and blue histograms represent the prior and posterior distributions, respectively, while the red dashed line indicates the true parameter value.

6 Revisiting Smets and Wouters (2007)

In this section, I revisit the well-known result of Smets and Wouters (2007), which shows that the marginal likelihood of their medium-scale DSGE model, estimated using the Laplace approximation, is comparable to the marginal likelihood of the VAR models, which is obtained analytically. Note that the Laplace approximation is a *local* method because it relies only on the unnormalized posterior density and the Hessian at the posterior mode. Hence, the estimate is sensitive to the optimization of the unnormalized posterior density, which is often numerically challenging, and to numerical errors in computing the Hessian. It also assumes that the posterior distribution is Gaussian, which is restrictive and generally not the case for DSGE models.

Therefore, it is worth re-examining their results using a more *global* estimation method, such as HMEs, instead of the local Laplace approximation. HMEs are global approximations because they use information from the entire posterior distribution, rather than just the information at a mode. Furthermore, unlike the Laplace approximation, they do not assume the posterior distribution is Gaussian, nor do they require optimization.[29]

[29] It is common to use information from the posterior mode and the Hessian to fine-tune MCMC algorithms when estimating DSGE models. While this pre-conditioning often improves sampling efficiency, it is not strictly necessary for posterior sampling. Additionally, although the SWZ estimator typically relies on the posterior mode, the posterior mean can be used instead to avoid the mode-finding step,

However, simply applying the HMEs is likely to give the DSGE model an unfair advantage compared to the VAR models due to the overestimation of the marginal likelihood caused by simulation pseudo-bias. To address this issue, I compute both the corrected and uncorrected HMEs for the DSGE model using the method in equation (16). For the VAR models, I estimate a conjugate VAR model with a Minnesota prior, after optimizing its hyperparameters by maximizing the marginal likelihood, following Giannone et al. (2015) and Ferroni and Canova (2021), and I report the marginal likelihood computed analytically. Additionally, I compare the marginal likelihoods of the two Bayesian VAR specifications from Smets and Wouters (2007), which can also be computed analytically.

I use the same dataset as Smets and Wouters (2007) and follow their method to compute the marginal likelihood for the full sample period from 1966:1 to 2004:4. Specifically, I calculate the marginal likelihood for this period by subtracting the estimate based on the training sample (1956:1–1965:4) from the estimate based on the extended sample (1956:1–2004:4). For reference, I also report the marginal likelihood estimates for the full sample period (1966:1–2004:4) without any subtraction for the DSGE model.[30]

Results. Table 5 reports the marginal likelihoods of three types of VAR models with lags varying from 1 to 5. The first column presents the marginal likelihoods of the optimized Minnesota conjugate VAR model I estimated, while the second and third columns present the values taken from Smets and Wouters (2007). We observe a similar pattern for the optimized Minnesota prior and the Sims and Zha (1998) prior: the marginal likelihoods peak at intermediate lag lengths, such as 3 or 4, and are generally higher than those for the "No other prior" specification. The marginal likelihoods for each prior specification with the optimal lag length fall between -907 and -928, and my own estimate is -914, which I use as my preferred estimate. We compare these values with the marginal likelihood of the DSGE model.

provided the posterior distribution is unimodal.

[30] I estimated the Smets and Wouters (2007) model running the replication code generously provided by Johannes Pfeifer, available at `https://github.com/johannespfeifer/dsge_mod`, with Dynare 6.2 of Adjemian et al. (2024).

VAR lags	Optimized Minnesota prior	No other prior	Sims and Zha (1998) prior
VAR(1)	-965.6	-928.0	-940.9
VAR(2)	-914.7	-966.6	-915.8
VAR(3)	-913.8	-1018.1	-908.7
VAR(4)	-917.9	-1131.2	-906.6
VAR(5)	-919.1	-	-907.7

Table 5: Marginal likelihoods of the VAR models

Note: This table presents the marginal likelihoods of VAR models with three different priors for five lags. The first column corresponds to the Minnesota prior, with hyperparameters optimized by maximizing the marginal likelihood following Ferroni and Canova (2021). The second and third columns are copied from Table 2 in Smets and Wouters (2007). The sample period is 1966:1-2004:4 and the period 1956:1-1965:4 is used as a training sample following Smets and Wouters (2007).

Table 6 reports the estimates of the marginal likelihood for the DSGE model across various sample periods using different HMEs and the Laplace estimator. The first row (1966:1—2004:4 (a)–(b)) presents the estimates comparable to those of the VARs in Table 5, and they are computed by subtracting the estimate of the training sample (1956:1–1965:4 (b)) from the extended sample period (1956:1–2004:4 (a)).[31] The second and third rows present the estimates used to obtain the results in the first row. The fourth row presents the estimate from the full sample period of 1966:1—2004:4 without any subtraction.

Before comparing the marginal likelihoods of the DSGE and VAR models, two points regarding Table 6 are worth mentioning. First, there is evidence of overestimation of the marginal likelihood due to simulation pseudo-bias. For example, the estimates of the original HME are larger than those of other estimators, such as Geweke's estimator and the SWZ estimator. This is evident in the second, third, and fourth rows of Table 6, where no subtraction is made between periods. After correcting for pseudo-bias, the original HME becomes comparable to the other estimators. Additionally, the modified HME with the uniform weighting density initially gives larger estimates than Geweke's estimator, but they become quantitatively similar after the pseudo-bias correction. Second, there are some differences between the Laplace approximation and other HMEs. For instance, the Laplace approximation for the sample period (b) 1956:1–1965:4 is smaller than the Geweke and SWZ estimators, and the difference is "positive" according to the scale in Kass and Raftery (1995), as shown in Table 1.

Now, let us compare the marginal likelihoods of the DSGE and VAR models. The

[31] The Laplace approximation reported in Table 6, −913.7, is different from the result reported by Smets and Wouters (2007) in Table 2, which is −905.8. Several factors may explain this difference. First, I re-estimate the posterior mode using the corresponding sample. Although Smets and Wouters (2007) provide optimization results, their optimization code is unavailable, so specific implementation details may differ. Second, I use Dynare version 6.2, which was not publicly available at the time Smets and Wouters (2007) was written. Smets and Wouters (2007) report Geweke's estimate for the full-sample period without pre-training as −923 in Table 4, which closely aligns with my estimate of −922.2.

Sample period	HM	C-HM	Unif	C-Unif	Geweke	C-Geweke	SWZ	(Laplace)
1966:1-2004:4 (a)-(b)	-875.7	-917.6	-909.0	-914.1	-914.9	-914.9	-915.1	-913.7
1956:1-2004:4 (a)	-1085.9	-1168.7	-1149.5	-1167.8	-1166.8	-1166.8	-1166.1	-1167.1
1956:1-1965:4 (b)	-210.2	-251.1	-240.5	-253.6	-251.9	-251.9	-251.1	-253.4
1966:1-2004:4 (no training)	-847.2	-920.9	-903.3	-921.2	-922.2	-922.2	-921.6	-923.0

Table 6: Marginal likelihoods of the Smets and Wouters (2007) model

Note: This table presents the marginal likelihood of the Smets and Wouters (2007) DSGE model, estimated using various harmonic-mean-type estimators, as well as the Laplace approximation employed by Smets and Wouters (2007). HM is the original harmonic mean estimator; Unif is the modified harmonic mean estimator with uniform weighting density; Geweke is the estimator proposed by Geweke (1999); SWZ is the estimator proposed by Sims, Waggoner and Zha (2008); and Laplace is based on the Laplace approximation. The prefix C indicates that the estimate is simulation pseudo-bias corrected. The first row shows the full sample estimate using the sample of 1956:1-1965:4 as a training sample. This row is comparable to Table 5 in this paper and Table 2 in Smets and Wouters (2007). The second and third rows show the results for each sample used to compute the second row. The fourth row shows the estimates without training periods.

estimation result confirms the robustness of the well-known result of Smets and Wouters (2007)—their DSGE model "is able to compete with Bayesian Vector Autoregression models"—when the marginal likelihood is evaluated using various HMEs, rather than the local Laplace approximation. The first row of Table 6 reports that, after correcting for pseudo-bias if needed, the marginal likelihood of the DSGE models is approximately −915. For instance, the marginal likelihoods of Geweke's estimator and the SWZ estimator are −914.9 and −915.1, respectively. The largest estimate is −914.1 by the pseudo-bias corrected modified HME with the uniform weighting density, while the lowest estimate is −917.6 by the pseudo-bias corrected original HME. These values are similar to the marginal likelihoods of the VAR models, which range from −907 to −928, and particularly close to my preferred estimate of −914.

This exercise also highlights the importance of correcting simulation pseudo-bias because, without correction, we would wrongly conclude that the DSGE model fits the data *much better* than the VAR models. For example, the uncorrected estimates from the original HME and the modified HME with the uniform weighting density are −875.7 and −909.0, respectively. They are considered "very strongly" or "strongly" higher than the marginal likelihood of the preferred VAR model, −914, on the scale of Kass and Raftery (1995).

7 Conclusion

In this paper, I demonstrated that the modified HME exhibits smaller simulation pseudo-bias than the original HME in practice. I also introduced a pseudo-bias correction method for the modified HMEs. Monte Carlo simulations confirmed the theoretical prediction and accuracy of the popular modified HMEs, such as those proposed by Geweke (1999) and Sims et al. (2008). Finally, I demonstrated that the well-known result of Smets and Wouters (2007) holds when various modified HMEs are applied instead of the Laplace approximation, and I highlighted the importance of pseudo-bias correction.

This study opens up several avenues for future research. First, while I focused on simulation pseudo-bias itself, its impact on model selection and model averaging requires further analysis. For instance, model selection might still be correct if the models under comparison induce quantitatively similar pseudo-bias. Furthermore, while the weights used for model averaging will be incorrect except for a knife-edge case where all the models exhibit the same pseudo-bias, their quantitative effect on the resulting model, such as on forecasting performance, remains unclear. Future research should examine how pseudo-bias and its correction affect model selection and averaging in practical macroeconomic applications, such as lag selection in VAR models, model comparison in larger, nested, and non-nested DSGE models, and forecasting comparisons. Second, comparing the performance of HMEs with alternative methods, such as the corrected arithmetic mean estimator (Pajor (2017)), bridge sampling (Frühwirth-Schnatter (2004)), and the variational Bayes weighting density (Hajargasht and Woźniak (2020)), in macroeconomic contexts remains a crucial area for future research. Third, examining the simulation pseudo-bias correction method from the bias-variance trade-off perspective will help clarify its properties. For instance, it may help guide the selection of the truncation region A to minimize the variance of the estimator.

A Proof of Theorem 1

The proof is a natural extension of Lenk (2009) and follows Pajor and Osiewalski (2013). Recall Bayes' theorem:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}.$$

Then we have,

$$\begin{aligned} W(A) &\equiv \int_A w(\theta)d\theta \\ &= \int_A \frac{w(\theta)}{p(\theta|y)} p(\theta|y)d\theta \\ &= \int_A \frac{w(\theta)}{p(y|\theta)p(\theta)/p(y)} p(\theta|y)d\theta \\ &= p(y) \int_A \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y)d\theta, \end{aligned}$$

where the assumption of $p(\theta|y) > 0$ almost everywhere is used in the second equality, and this leads to the equation in Proposition 2:

$$p(y) = P(A) \left[\int_A \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y)d\theta \right]^{-1}.$$

□

References

Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Rion, N. and Villemot, S. (2024), Dynare: Reference manual, version 6, Dynare Working Papers 80, CEPREMAP.

Angelini, G., Cavaliere, G. and Fanelli, L. (2022), 'Bootstrap inference and diagnostics in state space models: With applications to dynamic macro models', *Journal of Applied Econometrics* **37**(1), 3–22.

Ardia, D., Baştürk, N., Hoogerheide, L. and Van Dijk, H. K. (2012), 'A comparative study of monte carlo methods for efficient evaluation of marginal likelihood', *Computational Statistics & Data Analysis* **56**(11), 3398–3414.

Bayer, C., Born, B. and Luetticke, R. (2024), 'Shocks, frictions, and inequality in us business cycles', *American Economic Review* **114**(5), 1211–1247.

Bianchi, F. (2013), 'Regime switches, agents' beliefs, and post-world war ii us macroeconomic dynamics', *Review of Economic studies* **80**(2), 463–490.

Bianchi, F. and Nicolò, G. (2021), 'A generalized approach to indeterminacy in linear rational expectations models', *Quantitative Economics* **12**(3), 843–868.

Breitenlechner, M., Georgiadis, G. and Schumann, B. (2022), 'What goes around comes around: How large are spillbacks from us monetary policy?', *Journal of Monetary Economics* **131**, 45–60.

Brunnermeier, M., Palia, D., Sastry, K. A. and Sims, C. A. (2021), 'Feedbacks: financial markets and economic activity', *American Economic Review* **111**(6), 1845–1879.

Cai, M., Del Negro, M., Herbst, E., Matlin, E., Sarfati, R. and Schorfheide, F. (2021), 'Online estimation of dsge models', *The Econometrics Journal* **24**(1), C33–C58.

Chan, J. C. (2023), 'Comparing stochastic volatility specifications for large bayesian vars', *Journal of Econometrics* **235**(2), 1419–1446.

Chan, J. C., Hou, C. and Yang, T. T. (2020), Robust estimation and inference for importance sampling estimators with infinite variance, *in* 'Essays in Honor of Cheng Hsiao', Emerald Publishing Limited, pp. 255–285.

Chan, J. C., Yu, X. and Zhang, W. (2024), 'Bayesian model comparison for large bayesian vars after the covid-19 pandemic', *Available at SSRN 4913626* .

Chan, J., Koop, G., Poirier, D. J. and Tobias, J. L. (2019), *Bayesian econometric methods*, Vol. 7, Cambridge University Press.

Chen, X., Leeper, E. M. and Leith, C. (2022), 'Strategic interactions in us monetary and fiscal policies', *Quantitative Economics* **13**(2), 593–628.

Chib, S. (1995), 'Marginal likelihood from the gibbs output', *Journal of the american statistical association* **90**(432), 1313–1321.

Chib, S. and Jeliazkov, I. (2001), 'Marginal likelihood from the metropolis–hastings output', *Journal of the American statistical association* **96**(453), 270–281.

Creal, D. (2012), 'A survey of sequential monte carlo methods for economics and finance', *Econometric reviews* **31**(3), 245–296.

Davig, T. and Doh, T. (2014), 'Monetary policy regime shifts and inflation persistence', *Review of Economics and Statistics* **96**(5), 862–875.

Ferroni, F. and Canova, F. (2021), 'A hitchhiker's guide to empirical macro models'.

Friel, N. and Wyse, J. (2012), 'Estimating the evidence–a review', *Statistica Neerlandica* **66**(3), 288–308.

Frühwirth-Schnatter, S. (2004), 'Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques', *The Econometrics Journal* **7**(1), 143–167.

Fuentes-Albero, C. and Melosi, L. (2013), 'Methods for computing marginal data densities from the gibbs output', *Journal of Econometrics* **175**(2), 132–141.

Galí, J. (2015), *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.

Gelfand, A. E. and Dey, D. K. (1994), 'Bayesian model choice: asymptotics and exact calculations', *Journal of the Royal Statistical Society: Series B (Methodological)* **56**(3), 501–514.

Gemma, Y., Kurozumi, T. and Shintani, M. (2023), 'Trend inflation and evolving inflation dynamics: A bayesian gmm analysis', *Review of Economic Dynamics* **51**, 506–520.

Geweke, J. (1999), 'Using simulation methods for bayesian econometric models: inference, development, and communication', *Econometric reviews* **18**(1), 1–73.

Giannone, D., Lenza, M. and Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Hajargasht, G. and Woźniak, T. (2020), 'Accurate computation of marginal data densities using variational bayes', *arXiv preprint arXiv:1805.10036* .

Herbst, E. P. and Schorfheide, F. (2016), *Bayesian estimation of DSGE models*, Princeton University Press.

Herbst, E. and Schorfheide, F. (2014), 'Sequential monte carlo sampling for dsge models', *Journal of Applied Econometrics* **29**(7), 1073–1098.

Hubrich, K. and Tetlow, R. J. (2015), 'Financial stress and economic dynamics: The transmission of crises', *Journal of Monetary Economics* **70**, 100–115.

Inoue, A. and Shintani, M. (2018), 'Quasi-bayesian model selection', *Quantitative Economics* **9**(3), 1265–1297.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the american statistical association* **90**(430), 773–795.

Lenk, P. (2009), 'Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods', *Journal of Computational and Graphical Statistics* **18**(4), 941–960.

Lenza, M. and Primiceri, G. E. (2022), 'How to estimate a vector autoregression after march 2020', *Journal of Applied Econometrics* **37**(4), 688–699.

Li, Y., Wang, N. and Yu, J. (2023), 'Improved marginal likelihood estimation via power posteriors and importance sampling', *Journal of Econometrics* **234**(1), 28–52.

Liu, Z., Miao, J. and Zha, T. (2016), 'Land prices and unemployment', *Journal of Monetary Economics* **80**, 86–105.

Llorente, F., Martino, L., Delgado, D. and Lopez-Santiago, J. (2023), 'Marginal likelihood computation for model selection and hypothesis testing: an extensive review', *SIAM review* **65**(1), 3–58.

McEwen, J. D., Wallis, C. G., Price, M. A. and Mancini, A. S. (2021), 'Machine learning assisted bayesian model comparison: learnt harmonic mean estimator', *arXiv preprint arXiv:2111.12720* .

Metodiev, M., Perrot-Dockès, M., Ouadah, S., Irons, N. J., Latouche, P. and Raftery, A. E. (2024), 'Easily computed marginal likelihoods from posterior simulation using the thames estimator', *Bayesian Analysis* **1**(1), 1–28.

Newton, M. A. and Raftery, A. E. (1994), 'Approximate bayesian inference with the weighted likelihood bootstrap', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **56**(1), 3–26.

Pajor, A. (2017), 'Estimating the Marginal Likelihood Using the Arithmetic Mean Identity', *Bayesian Analysis* **12**(1), 261 – 287.

Pajor, A. and Osiewalski, J. (2013), 'A note on lenk's correction of the harmonic mean estimator', *Central European Journal of Economic Modelling and Econometrics* pp. 271–275.

Polanska, A., Price, M. A., Spurio Mancini, A. and McEwen, J. D. (2023), 'Learned harmonic mean estimation of the marginal likelihood with normalizing flows', *Physical Sciences Forum* **9**(1).

Raftery, A., Newton, M., Satagopan, J. and Krivitsky, P. (2007), 'Estimating the integrated likelihood via posterior simulation using the harmonic mean identity', *Bayesian statistics* **8**.

Schmitt-Grohé, S. and Uribe, M. (2012), 'What's news in business cycles', *Econometrica* **80**(6), 2733–2764.

Sims, C. A., Waggoner, D. F. and Zha, T. (2008), 'Methods for inference in large multiple-equation markov-switching models', *Journal of econometrics* **146**(2), 255–274.

Sims, C. A. and Zha, T. (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review* pp. 949–968.

Smets, F. and Wouters, R. (2007), 'Shocks and frictions in us business cycles: A bayesian dsge approach', *American economic review* **97**(3), 586–606.

Uhlig, H. (1995), 'A toolkit for analyzing nonlinear dynamic stochastic models easily'.

Ventura, V. (2002), 'Non-parametric bootstrap recycling', *Statistics and Computing* **12**, 261–273.

Waggoner, D. F., Wu, H. and Zha, T. (2016), 'Striated metropolis–hastings sampler for high-dimensional models', *Journal of Econometrics* **192**(2), 406–420.

Walsh, C. E. (2017), *Monetary theory and policy*, MIT press.

Wang, Y.-B., Chen, M.-H., Kuo, L. and Lewis, P. O. (2018), 'A new monte carlo method for estimating marginal likelihoods', *Bayesian analysis* **13**(2), 311.

Supplementary Appendix (Not for publication)

A Algorithms

This section first presents the algorithm to implement the modified HME in general, and then explains the algorithm for two specific modified HMEs proposed by Geweke (1999) and Sims et al. (2008). Note that it is recommended to take the natural log when computing these values in practice to mitigate computational errors, if possible. For instance, usually $\log(p(y))$ is computed and reported instead of $p(y)$, although I write the algorithm below without doing so for notational simplicity.

A.1 The modified HME in general

0. Decide what weighting density $w(\theta)$ to use. For each sample of the posterior distribution $\theta^{(i)}, i = 1, 2, ..., N$, do the following.

1. Compute $w^{(i)} = w(\theta^{(i)})$ and $p(y|\theta^{(i)})p(\theta^{(i)})$ for $i = 1, 2, ..., N$.

2. Approximate $p(y)$ by $p(y) = \left[\frac{1}{N}\sum_{i=1}^{N} \frac{w^{(i)}}{p(y|\theta^{(i)})p(\theta^{(i)})}\right]^{-1}$

A.2 Geweke's estimator

0. Decide τ. For instance, Geweke (1999) uses $\tau = 0.1, 0.5, 0.9$. For each sample of the posterior distribution $\theta^{(i)}, i = 1, 2, ..., N$, do the following.

1. Compute $w^{(i)}_{Geweke} = w_{Geweke}(\theta^{(i)})$ and $p(y|\theta^{(i)})p(\theta^{(i)})$ for $i = 1, 2, ..., N$ where

$$w_{Geweke}(\theta) = \tau^{-1}(2\pi)^{-k/2}|\bar{\Omega}|^{-1/2} \exp\left(-\frac{1}{2}(\theta - \bar{\theta})'\bar{\Omega}^{-1}(\theta - \bar{\theta})\right) \mathbb{1}(\theta \in \bar{\Theta}),$$

$\bar{\Theta} = \{\theta : (\theta - \bar{\theta})'\bar{\Omega}^{-1}(\theta - \bar{\theta}) \leq F^{-1}_{\chi^2_k}(\tau)\}$, $\bar{\theta}$ and $\bar{\Omega}$ are the posterior mean and covariance matrix of θ, k is the dimension of θ, $\tau \in (0, 1)$ is a tuning parameter, and $F^{-1}_{\chi^2_k}(\tau)$ is an inverse of a χ^2 CDF with the degree of freedom k.

2. Approximate $p(y)$ by $p(y) = \left[\frac{1}{N}\sum_{i=1}^{N} \frac{w^{(i)}_{Geweke}}{p(y|\theta^{(i)})p(\theta^{(i)})}\right]^{-1}$

A.3 Sims, Waggoner, and Zha's estimator

Recall the weighting density of the SWZ estimator:

$$w_{SWZ}(\theta) = q_L^{-1} g(\theta) \mathbb{1}(\theta \in \widehat{\Theta})$$

where

$$g(\theta) = \frac{\Gamma(k/2)}{2\pi^{k/2}|\widehat{S}|} \frac{f(r(\theta))}{r(\theta)^{k-1}},$$

q_L is the normalizing constant ensuring that $\int_{\Theta} w(\theta)_{SWZ} d\theta = 1$; $\widehat{\Theta}$ determines the tuncation as $\widehat{\Theta} = \{\theta : p(y|\theta)p(\theta) > L_{1-q}, r(\theta) \in [a, b]\}$, and L_{1-q} is $1-q$ percentile of the realized unnormalized posterior density $\{p(y|\theta^i)p(\theta^i)\}_{i=1}^N$, and Sims et al. (2008) recommend using $q = 0.9$; a, b, and ν are defined below; Γ is the gamma function; k is the number of parameters; $r(\theta)$ is a function defined as $r(\theta) = \sqrt{(\theta - \hat{\theta})'\widehat{\Omega}^{-1}(\theta - \hat{\theta})}$ where $\hat{\theta}$ is a posterior mode and $\widehat{\Omega}$ is a covariance matrix centered at the posterior mode: $\widehat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}(\theta^{(i)} - \hat{\theta})(\theta^{(i)} - \hat{\theta})'$; $\widehat{S}$ is the lower Cholesky decomposition of $\widehat{\Omega}$; $f()$ is a probability density function defined below. Let c_1, c_{10} and c_{90} be the 1st, 10th, and 90th percentile values of the posterior draws $\{r^{(i)}\}_{i=1}^N$ computed as $r^{(i)} = \sqrt{(\theta^{(i)} - \hat{\theta})'\widehat{\Omega}^{-1}(\theta^{(i)} - \hat{\theta})}$. Then, set $\nu = \frac{\log(1/9)}{\log(c_{10}/c_{90})}$, $a = c_1$ and $b = \frac{c_{90}}{0.9^{\frac{1}{\nu}}}$. The probability density function $f(r)$ is defined as follows:

$$f(r) = \begin{cases} \frac{\nu r^{\nu-1}}{b^\nu - a^\nu} & \text{if } r \in [a, b] \\ 0 & \text{otherwise.} \end{cases}$$

Mainly following Herbst and Schorfheide (2016), the SWZ estimator can be implemented as follows.

0. Compute a posterior mode $\hat{\theta}$ and a covariance matrix $\widehat{\Omega}$ centered at the posterior mode. Choose the threshold value q. Also compute other parameters, such as a, b

and ν.

1. Simulate N^{sim} series of i.i.d samples of $\{\theta^{(j)}\}_{j=1}^{N^{sim}}$ from $g(\theta)$ by the following algorithm.

   1-1. Sample $r^{(j)}$ from $f(r)$ by using the inversion sampling:

   $$r^{(j)} = \{(b^{\nu} - a^{\nu})u^{(j)} + a^{\nu}\}^{\frac{1}{\nu}},$$

   where $u^{(j)}$ is a random sample from the uniform distribution with support $[0, 1]$.

   1-2. Sample $\theta^{(j)}$ by

   $$\theta^{(j)} = \frac{r^{(j)}}{||x^{(j)}||}\widehat{S}x^{(j)} + \hat{\theta}$$

   where $||x||$ is the Euclidean norm, $\widehat{S}$ is the lower Cholesky decomposition of $\widehat{\Omega}$, and $x^{(j)}$ is a random sample from the k dimensional standard multivariate normal distribution.

2. Estimate the normalizing constant q_L by $q_L = \frac{1}{N^{sim}} \sum_{j=1}^{N^{sim}} \mathbb{1}(\theta^{(j)} \in \widehat{\Theta})$. Increase q if q_L is too small.[32]

3. For all $i = 1, 2, ..., N$, compute $w_{SWZ}^{(i)} = w_{SWZ}(\theta^{(i)})$ and $p(y|\theta^{(i)})p(\theta^{(i)})$.

4. Approximate $p(y)$ by $p(y) = \left[\frac{1}{N} \sum_{i=1}^{N} \frac{w_{SWZ}^{(i)}}{p(y|\theta^{(i)})p(\theta^{(i)})}\right]^{-1}$

Remark 1: The inversion sampling works as follows. First, note that if $X \sim F_X$, then the random variable following $F_X(U)^{-1}$, $U \sim$ Uniform$[0, 1]$ follows the same distribution as X because $P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x)$. Then, recall that we have $F_R(r) = \frac{r^{\nu} - a^{\nu}}{b^{\nu} - a^{\nu}}$ because of the definition of $f(r)$ in (4). Hence, by using $u \sim$ Uniform$[0, 1]$ and $u = \frac{r^{\nu} - a^{\nu}}{b^{\nu} - a^{\nu}}$, we have $r = \{(b^{\nu} - a^{\nu})u + a^{\nu}\}^{\frac{1}{\nu}}$, which enables us to sample directly from $f(f)$ by transforming a random uniform sample u.

[32] As a rule of thumb, Sims et al. (2008) recommends $q_L \geq 1.0e - 6$

Remark 2: When computing the natural log of the weighting density, using the following formula will help avoid computational instability:

$$\log\left(|\widehat{S}|\right) = \frac{1}{2}\log\left(|\widehat{\Omega}|\right) = \sum_{s=1}^{k}\log\left(\widehat{S}_{s,s}\right)$$

where $\widehat{S}_{s,s}$ denotes the sth diagonal element of $\widehat{S}$. The first equality follows because $|A| = |A'|$ for a square matrix A and $\Omega = SS'$. The second equality follows because S is a lower triangular matrix.

Remark 3: Note that the simulation evaluating q_L requires computing the likelihood $p(y|\theta^{(j)})$ for N^{sim} times, which could be computationally demanding if evaluating the likelihood is costly.

Remark 4: The above algorithm uses direct sampling from $g(\theta)$ to evaluate q_L, but importance sampling can also be used. However, my computational experience suggests that direct sampling is computationally more stable and accurate than importance sampling.

Remark 5: Some papers define the denominator of $g(\theta)$ as $(2\pi)^{k/2}$, not $2\pi^{k/2}$, but this definition distorts the evaluation of q_L if the direct sampling approach is employed.

Remark 6: Some papers skip the numerical evaluation of q_L and use some simplification. For instance, they set $q_L = 1 - q_{SWZ}$ or evaluate q_L by using the samples from the posterior distribution: $q_L = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{1}(p(y|\theta^{(i)})p(\theta^{(i)}) > L) \times \mathbb{1}(r^{(i)} \in [a,b])\right)$. However, such simplification leads to the violation of the assumption that $w_{SWZ}(\theta)$ is a probability density.

B Details of the models used in the illustrative example and the Monte Carlo simulation

This section explains the details of the three models used in the paper for the illustrative example and the Monte Carlo simulation. The first model is the conjugate normal model with known variance, and its posterior distribution and marginal likelihood can be derived analytically. The second model is a Bayesian linear regression model, and similarly to the first model, its (full conditional) posterior distribution and marginal likelihood are derived analytically, and the Gibbs sampling algorithm is explained. The third model is a small-scale New Keynesian dynamic stochastic general equilibrium (DSGE) model. I first derive the state-space representation of the model. Then, I provide the algorithm of the sequential Monte Carlo sampler to draw samples from the posterior distribution.

I use the following probability distributions below. First is the multivariate normal distribution: If $X \sim \mathcal{N}(\mu, \Sigma)$, then $p(x) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$ where k is the dimension of X. Second is the inverse-gamma distribution: If $y \sim IG(v_0, v_1)$, then $p(y) = [\Gamma(v_0)(v_1)^{v_0}]^{-1} y^{-(v_0+1)} \exp\{-\frac{1}{yv_1}\}$ where $\Gamma(\cdot)$ is a Gamma function.

B.1 A conjugate normal model

The posterior density can be written as

$$\begin{aligned}
p(\mu|y) &\propto p(y|\mu)p(\mu) \\
&= (2\pi)^{-T/2}(\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2\right\} \\
&\times (2\pi)^{-1/2}(V_0\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2V_0}(\mu-\mu_0)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2\right\} \exp\left\{-\frac{1}{2V_0}(\mu-\mu_0)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(T\mu^2 - 2T\bar{y}\mu) - \frac{1}{2V_0}(\mu^2 - 2\mu_0\mu)\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{T}{\sigma^2}+\frac{1}{V_0}\right)\mu^2 - 2\left(\frac{T\bar{y}}{\sigma^2}+\frac{\mu_0}{V_0}\right)\mu\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\frac{1}{(T/\sigma^2+1/V_0)^{-1}}\left[\mu - \frac{T\bar{y}/\sigma^2+\mu_0/V_0}{T/\sigma^2+1/V_0}\right]^2\right\}
\end{aligned}$$

Hence, we have $\mu|y \sim \mathcal{N}(\mu_T, V_T)$ where $V_T = (T/\sigma^2 + 1/V_0)^{-1}$ and $\mu_T = \frac{T\bar{y}/\sigma^2+\mu_0/V_0}{T/\sigma^2+1/V_0} = V_T(T\bar{y}/\sigma^2 + \mu_0/V_0)$.

The marginal likelihood can be obtained as follows:

$$
\begin{aligned}
p(y) &= \int p(y,\mu)d\mu \\
&= \int p(y|\mu)p(\mu)d\mu \\
&= \int (2\pi)^{-T/2}(\sigma^2)^{-T/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2\right\}\times(2\pi)^{-1/2}(V_0)^{-1/2}\exp\left\{-\frac{1}{2V_0}(\mu-\mu_0)^2\right\}d\mu \\
&= (2\pi)^{-(T+1)/2}(\sigma^2)^{-T/2}(V_0)^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{\sum_{t=1}^{T}y_t^2}{\sigma^2}+\frac{\mu_0^2}{V_0}\right)\right\} \\
&\times \int \exp\left\{-\frac{1}{2}\left[\left(\frac{T}{\sigma^2}+\frac{1}{V_0}\right)\mu^2-2\left(\frac{T\bar{y}}{\sigma^2}+\frac{\mu_0}{V_0}\right)\mu\right]\right\}d\mu \\
&= (2\pi)^{-(T+1)/2}(\sigma^2)^{-T/2}(V_0)^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{\sum_{t=1}^{T}y_t^2}{\sigma^2}+\frac{\mu_0^2}{V_0}\right)\right\} \\
&= \int \exp\left\{-\frac{1}{2V_T}(\mu-\mu_T)^2\right\}d\mu\times\exp\left\{\frac{\mu_T^2}{2V_T}\right\} \\
&= (2\pi)^{-(T+1)/2}(\sigma^2)^{-T/2}(V_0)^{-1/2}\exp\left\{-\frac{1}{2}\left(\frac{\sum_{t=1}^{T}y_t^2}{\sigma^2}+\frac{\mu_0^2}{V_0}\right)\right\} \\
&\times (2\pi)^{1/2}(V_T)^{1/2}\int(2\pi)^{-1/2}(V_T)^{-1/2}\exp\left\{-\frac{1}{2V_T}(\mu-\mu_T)^2\right\}d\mu \\
&= (2\pi\sigma^2)^{-T/2}\left(\frac{V_T}{V_0}\right)^{1/2}\exp\left\{-\frac{1}{2}\left(\frac{\sum_{t=1}^{T}y_t^2}{\sigma^2}+\frac{\mu_0^2}{V_0}-\frac{\mu_T^2}{V_T}\right)\right\}
\end{aligned}
$$

B.2 Bayesian linear regression models

The joint posterior density can be written as follows:

$$
\begin{aligned}
p(\beta,\sigma^2|y,X) &\propto p(y|\beta,\sigma^2,X)p(\beta|\sigma^2)p(\sigma^2) \\
&= (2\pi)^{-T/2}(\sigma^2)^{-T/2}\exp\left\{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right\} \\
&\times (2\pi)^{-k/2}|\sigma^2 V_0|^{-1/2}\exp\left\{-\frac{1}{2}(\beta-\beta_0)'(\sigma^2 V_0)^{-1}(\beta-\beta_0)\right\} \\
&\times (\Gamma(v_0)v_1^{v_0})^{-1}(\sigma^2)^{-(v_0+1)}\exp\left\{-\frac{1}{v_1\sigma^2}\right\} \\
&\propto (\sigma^2)^{-(T/2+k/2+v_0+1)}\exp\left\{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)\right\} \\
&\times \exp\left\{-\frac{1}{v_1\sigma^2}\right\}
\end{aligned}
$$

because $|\sigma^2 V_0|^{-1/2} = (\sigma^2)^{-k/2}|V_0|^{-1/2}$. Then, the full conditional posterior distribution of β can be obtained as follows:

$$
\begin{aligned}
p(\beta|\sigma^2,y,X) &= p(\sigma^2|y,X)^{-1}p(\beta,\sigma^2|y,X) \\
&\propto p(\beta,\sigma^2|y,X) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right\}\exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}(y'y-2y'X\beta+\beta'X'X\beta+\beta'V_0^{-1}\beta-2\beta_0'V_0^{-1}\beta+\beta_0'V_0^{-1}\beta_0)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\beta'X'X\beta-2\beta'X'y+\beta'V_0^{-1}\beta-2\beta_0'V_0^{-1}\beta)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\beta'X'X\beta-2\beta'X'y+\beta'V_0^{-1}\beta-2\beta_0'V_0^{-1}\beta)\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma^2}(X'y+V_0^{-1}\beta_0)'(X'X+V_0^{-1})^{-1}(X'y+V_0^{-1}\beta_0)\right\} \\
&= \exp\left\{-\frac{1}{2}(\beta-\beta_T)'(\sigma^2 V_T)^{-1}(\beta-\beta_T)\right\}
\end{aligned}
\tag{B.1}
$$

where $V_T = (X'X + V_0^{-1})^{-1}$ and $\beta_T = (X'X + V_0^{-1})^{-1}(X'y + V_0^{-1}\beta_0) = V_T(X'y + V_0^{-1}\beta_0)$. Hence, it follows the multivariate normal distribution: $\beta|\sigma^2, y, X \sim \mathcal{N}(\beta_T, \sigma^2 V_T)$.

The marginal posterior distribution of σ^2 can be computed as follows:

$$
\begin{aligned}
p(\sigma^2|y,X) &= \int p(\sigma^2, \beta|X, y)d\beta \\
&\propto (\sigma^2)^{-(T/2+k/2+v_0+1)} \exp\left\{-\frac{1}{v_1\sigma^2}\right\} \\
&\times \int \exp\left\{-\frac{1}{2\sigma^2}[(y - X\beta)'(y - X\beta) + (\beta - \beta_0)'V_0^{-1}(\beta - \beta_0)]\right\} d\beta \\
&= (\sigma^2)^{-(T/2+k/2+v_0+1)} \exp\left\{-\frac{1}{v_1\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}(y'y + \beta_0'V_0^{-1}\beta_0 - \beta_T'V_T^{-1}\beta_T)\right\} \\
&\times (2\pi)^{k/2}|\sigma^2 V_T|^{1/2} \int (2\pi)^{-k/2}|\sigma^2 V_T|^{-1/2} \exp\left\{-\frac{1}{2}(\beta - \beta_T)'(\sigma^2 V_T)^{-1}(\beta - \beta_T)\right\} d\beta \\
&\propto (\sigma^2)^{-(T/2+v_0+1)} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{v_1} + \frac{1}{2}(y'y + \beta_0'V_0^{-1}\beta_0 - \beta_T'V_T^{-1}\beta_T)\right]\right\}
\end{aligned}
$$

The last equation follows because the probability density function of multivariate normal distribution integrates to unity. Hence, we have $\sigma^2|y, X \sim IG(\tilde{v}_0, \tilde{v}_1)$ where $\tilde{v}_0 = T/2 + v_0$ and $\tilde{v}_1 = \left[\frac{1}{v_1} + \frac{1}{2}(y'y + \beta_0'V_0^{-1}\beta_0 - \beta_T'V_T^{-1}\beta_T)\right]^{-1}$.

To calculate the marginal likelihood, we follow a two-step approach. First, we calculate $p(y|X, \sigma^2) = \int p(y, \beta|X, \sigma^2)d\beta = \int p(y|\beta, X, \sigma^2)p(\beta|\sigma^2)d\beta$. Then, we calculate the marginal likelihood as $p(y|X) = \int p(y, \sigma^2|X) d\sigma^2 = \int p(y|X, \sigma^2)p(\sigma^2) d\sigma^2$. First, we have

$$
\begin{aligned}
p(y|X, \sigma^2) &= \int p(y, \beta|X, \sigma^2) d\beta \\
&= \int p(y|\beta, \sigma^2, X) p(\beta|\sigma^2) d\beta \\
&= \int (2\pi)^{-T/2} (\sigma^2)^{-T/2} \exp\left\{ -\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right\} \\
&\times (2\pi)^{-k/2} (\sigma^2)^{-k/2} |V_0|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (\beta - \beta_0)' V_0^{-1} (\beta - \beta_0) \right\} d\beta \\
&= (2\pi)^{-T/2-k/2} (\sigma^2)^{-T/2-k/2} |V_0|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} [(y - X\beta)'(y - X\beta) + (\beta - \beta_0)' V_0^{-1} (\beta - \beta_0)] \right\} \\
&= (2\pi)^{-T/2-k/2} (\sigma^2)^{-T/2-k/2} |V_0|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (y'y + \beta_0' V_0^{-1} \beta_0) \right\} \\
&\times \exp\left\{ \frac{1}{2\sigma^2} (X'y + V_0^{-1}\beta_0)'(X'X + V_0^{-1})^{-1}(X'y + V_0^{-1}\beta_0) \right\} \\
&\times \int \exp\left\{ -\frac{1}{2} (\beta - \beta_T)'(\sigma^2 V_T)^{-1}(\beta - \beta_T) \right\} d\beta \\
&= (2\pi)^{-T/2-k/2} (\sigma^2)^{-T/2-k/2} |V_0|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (y'y + \beta_0' V_0^{-1} \beta_0 - \beta_T' V_T^{-1} \beta_T) \right\} \\
&\times (2\pi)^{k/2} |\sigma^2 V_T|^{1/2} \int (2\pi)^{-k/2} |\sigma^2 V_T|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (\beta - \beta_T)'(\sigma^2 V_T)^{-1}(\beta - \beta_T) \right\} d\beta \\
&= (2\pi)^{-T/2} (\sigma^2)^{-T/2} |V_0|^{-1/2} |V_T|^{1/2} \exp\left\{ -\frac{1}{2\sigma^2} (y'y + \beta_0' V_0^{-1} \beta_0 - \beta_T' V_T^{-1} \beta_T) \right\}
\end{aligned}
$$

The fifth equation follows by recalling the derivation of equation (B.1).

Then, the marginal likelihood can be calculated as follows.

$$\begin{aligned}
p(y|X) &= \int p(y|X,\sigma^2)p(\sigma^2)d\sigma^2 \\
&= (2\pi)^{-T/2}|V_0|^{-1/2}|V_T|^{1/2}\int(\sigma^2)^{-T/2}\exp\left\{-\frac{1}{2\sigma^2}(y'y+\beta_0'V_0^{-1}\beta_0-\beta_T'V_T^{-1}\beta_T)\right\} \\
&\times(\Gamma(v_0)(v_1)^{v_0})^{-1}(\sigma^2)^{-(v_0+1)}\exp\left\{-\frac{1}{v_1\sigma^2}\right\}d\sigma^2 \\
&= (2\pi)^{-T/2}|V_0|^{-1/2}|V_T|^{1/2}(\Gamma(v_0)(v_1)^{v_0})^{-1} \\
&\times\int(\sigma^2)^{-(T/2+v_0+1)}\exp\left\{-\frac{1}{\sigma^2}\left(\frac{1}{v_1}+\frac{1}{2}(y'y+\beta_0'V_0^{-1}\beta_0-\beta_T'V_T^{-1}\beta_T)\right)\right\}d\sigma^2 \\
&= (2\pi)^{-T/2}|V_0|^{-1/2}|V_T|^{1/2}(\Gamma(v_0)(v_1)^{v_0})^{-1}(\Gamma(\tilde{v}_0)(\tilde{v}_1)^{\tilde{v}_0}) \\
&\times\int(\Gamma(\tilde{v}_0)(\tilde{v}_1))^{-1}(\sigma^2)^{-(\tilde{v}_0+1)}\exp\left\{-\frac{1}{\tilde{v}_1\sigma^2}\right\}d\sigma^2 \\
&= (2\pi)^{-T/2}\frac{|V_T|^{1/2}}{|V_0|^{-1/2}}\frac{\Gamma(\tilde{v}_0)(\tilde{v}_1)^{\tilde{v}_0}}{\Gamma(v_0)(v_1)^{v_0}}.
\end{aligned}$$

where $\tilde{v}_0 = T/2 + v_0$ and $\tilde{v}_1 = \left[\frac{1}{v_1}+\frac{1}{2}(y'y+\beta_0'V_0^{-1}\beta_0-\beta_T'V_T^{-1}\beta_T)\right]^{-1}$. The last equation follows because the probability density function of the inverse-gamma distribution integrates to unity.

Sampling method. We can use the Gibbs sampler to draw samples from the posterior distribution.

0. Iterate this procedure for $i = 1, 2, ..., N$.
1. Sample $\sigma^{2,(j)}$ from $IG(\tilde{v}_0, \tilde{v}_1)$
2. Given $\sigma^{2,(j)}$, sample $\beta^{(j)}$ from $\mathcal{N}(\beta_T, \sigma^{2,(j)}V_T)$

B.3 The three-equation New Keynesian DSGE model

The state-space representation of the extended three-equation NK-DSGE model. By applying the Uhlig (1995) toolkit to the model (19), we obtain the solution of the model as follows:

$$state_t = P(\theta)state_{t-1} + Q(\theta)exo_t$$

$$nstate_t = R(\theta)state_{t-1} + S(\theta)exo_t$$

$$exo_t = N(\theta)exo_{t-1} + \Sigma^{1/2}\varepsilon_t$$

where θ is a vector collecting all the parameters of the model, $state_t = [\pi_t, i_t]', nstate_t = y_t, exo_t = [e_t^d, e_t^s, e_t^m], \varepsilon_t = [\varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^m]'$, and $\Sigma^{1/2} = \text{diag}(\sigma_d, \sigma_s, \sigma_m)$. By setting $S_t = [y_t, \pi_t, i_t, e_t^d, e_t^s, e_t^m]'$, the state transition equation can be written as

$$S_t = \begin{bmatrix} y_t \\ \pi_t \\ i_t \\ e_t^d \\ e_t^s \\ e_t^m \end{bmatrix} = \begin{bmatrix} 0 & R & SN \\ \mathbf{0}_{2\times1} & P & QN \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times2} & N \end{bmatrix} S_{t-1} + \begin{bmatrix} S\Sigma^{1/2} \\ Q\Sigma^{1/2} \\ \Sigma^{1/2} \end{bmatrix} \varepsilon_t, \tag{B.2}$$

which provides equation (20). The observation equation can be written as

$$\begin{bmatrix} y_t^o \\ \pi_t^o \\ i_t^o \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} S_t + \Sigma_u^{1/2} u_t, \tag{B.3}$$

which provides equation (21).

The state-space representation of the simple three-equation NK-DSGE model. When there is no inertia, i.e., $\rho_d = \gamma_b = \rho_i = 0$, there exists a simple solution to the model (19). More specifically, under this condition, the model (19) can be written as follows:

$$y_t = E_t y_{t+1} - \gamma(i_t - E_t \pi_{t+1}) + \sigma_d \varepsilon_t^d$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa y_t + \sigma_s \varepsilon_t^s$$

$$i_t = \phi_\pi \pi_t + \phi_y y_t + \sigma_m \varepsilon_t^m$$

where $\beta \in (0,1), \kappa > 0, \phi_\pi > 1, \phi_y \geq 0$, and $\sigma_d, \sigma_s, \sigma_m > 0$. I also assume $\varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^m \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Then, we can use a method of undetermined coefficients, and the solution for this system of the expectational difference equations can be expressed as follows:

$$S_t \equiv \begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix} = \frac{1}{1+\gamma(\phi_\pi \kappa + \phi_y)} \begin{bmatrix} \sigma_d & -\phi_\pi \sigma_s & -\sigma_m \\ \kappa \sigma_d & (1+\phi_y)\sigma_s & -\kappa \sigma_m \\ (\phi_y + \phi_\pi \kappa)\sigma_d & \phi_\pi \sigma_s & \sigma_m \end{bmatrix} \begin{bmatrix} \varepsilon_t^d \\ \varepsilon_t^s \\ \varepsilon_t^m \end{bmatrix}$$

This solution directly provides the state-transition equation, and the observation equation is

$$Y_t^o \equiv \begin{bmatrix} y_t^o \\ \pi_t^o \\ i_t^o \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} S_t + \Sigma_u^{1/2} u_t$$

where $u_t \overset{\text{iid}}{\sim} \mathcal{N}(0, I)$.

The sequential Monte Carlo sampler. The algorithm of the sequential Monte Carlo sampler I use in this paper closely follows Herbst and Schorfheide (2014) and Herbst and Schorfheide (2016). This sampler approximates the posterior distribution by iteratively perturbing the likelihood function and gradually updating the prior distribution. Specifically, consider

$$\pi_n(\theta) = \frac{p(Y|\theta)^{\phi_n} p(\theta)}{\int p(Y|\theta)^{\phi_n} p(\theta) d\theta}$$

where $n = 1, ..., \phi_{N_\phi}$, ϕ_{N_ϕ} is the total number of iteration, and ϕ_n is an increasing sequence with $\phi_1 = 0$ and $\phi_{N_\phi} = 1$. Note that $\pi_n(\theta)$ corresponds to the posterior distribution when $n = N_\phi$.

In the algorithm below, I use slightly different notations from the main paper to make the explanation clearer. For more details, see Herbst and Schorfheide (2016).

- θ: A vector of parameters of the model.
- N: Number of particles (i). N_{MH}: number of the Metropolis-Hastings steps in the mutation step m.
- $\{\rho_n\}_{n=1}^{N_\phi}$: Resampling schedule, taking either 1 meaning "do resampling", and 0 meaning "no resampling", at each stage n.
- λ: A parameter controlling the tempering schedule: $\phi_n = \left(\frac{n}{N_\phi}\right)^\lambda$.
- α: The targeted acceptance ratio of the Metropolis-Hastings steps in the mutation.
- $\{\zeta_n\}_{n=1}^{N_\phi}$: A vector of tuning parameters used in the mutation step. They include a scaling parameter of the covariance matrix c_n and a covariance matrix of the multivariate Gaussian proposal distribution Σ_n^*. It also includes c^*, the initial scaling parameter.
- Generally, w and W denote the unnormalized and normalized weights of the importance sampling. $p(y|\theta)$ is the likelihood, and one can compute it by using the Kalman filter for the DSGE models in this paper because they are formulated as a linear Gaussian state space model. $p(\theta)$ is the prior density.

Algorithm.

0. Decide $\lambda, N, N_\phi, N_{MH}$ and c^* and compute $\{\phi_n\}_{n=1}^{N_\phi}$. $\{\rho_n\}_{n=1}^{N_\phi}$ and $\{\zeta_n\}_{n=1}^{N_\phi}$ are determined adaptively during the sampling procedure.
1. Initialization (ϕ_0=0): Generate first particles from prior distributions.

   Draw $\theta_0^i \overset{\text{iid}}{\sim} p(\theta)$ for $i = 1, 2, ..., N$. Set $W_0^i = 1/N$ for $i = 1, 2, ..., N$.
2. Recursion: Do the following for $n = 1, 2, ..., N_\phi$.

2-1. Correction: Computing particle weights $\tilde{W}_n^i$.

Reweight particles from stage $n-1$ by

$$\tilde{w}_n^i = [p(y|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}}$$

and

$$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}$$

for $i = 1, 2, ..., N$.

2-2. Recursion: Resampling particles and set equal weights to avoid weight degeneracy issues.

2-2-1. Set $\rho_n = 1$ if $\widehat{ESS}_n < T/2$ where the effective sample size is computed as $\widehat{ESS}_n = \frac{1}{\sum_{i=1}^N (\tilde{W}_i^n)^2}$. The resampling happens when the weights are so dispersed that the effective sample size becomes small, such as half of the number of particles.

2-2-2. If $\rho_n = 1$, resample by multinomial resampling. That is, for each $i = 1, 2, ..., N$, independently draw $\hat{\theta}_n^i$ from multinomial distribution with support points $\{\theta_{n-1}^i\}_{i=1}^N$ and probabilities $\{\tilde{W}_n^i\}_{i=1}^N$. Set $W_n^i = 1/N$ for $i = 1, 2, ..., N$.

If $\rho_n = 0$, no-resampling is needed, so set $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$.

2-3. Mutation: Propagating the particles $\{\hat{\theta}_n^i, W_n^i\}_{i=1}^N$ by the Metropolis-Hastings algorithm.

2-3-1. Compute ζ_n: c_n and Σ_n^*. The scaling parameter is computed by

$$c_n = c_{n-1} f(R_{n-1})$$

where $f(x) = 0.95 + 0.1\frac{e^{16(x-\alpha)}}{1+e^{16(x-\alpha)}}$ and R_{n-1} is the realized acceptance ratio at stage $n-1$, so that $\{c_n\}_{n=1}^{N_\phi}$ will be adjusted to achieve the targeted acceptance ration α.[33] Set $\Sigma_n^* = V_{\pi_n}[\theta] \approx \sum_{i=1}^N (\hat{\theta}_n^i - \bar{\hat{\theta}}_n^i) W_n^i (\hat{\theta}_n^i - \bar{\hat{\theta}}_n^i)'$ where

[33] We assume that the initial acceptance ratio is 0.5. This value is needed to compute c_1.

$\bar{\hat{\theta}}_n^i = \sum_{i=1}^N W_n^i \cdot \hat{\theta}_n^i$.

2-3-2. For $m = 1, ..., N_{MH}$ repeat the following Metropolis-Hastings step. First, draw the candidate $\theta_n^{*i,m}$ from the Gaussian proposal distribution by

$$\theta_n^{*i,m} \sim \mathcal{N}(\theta_n^{i,m-1}, c_n^2 \tilde{\Sigma}_n)$$

where $\theta_n^{i,0} = \hat{\theta}_n^i$.

Second, set

$$\theta_n^{i,m} = \begin{cases} \theta_n^{*i,m} & \text{if } \log(\alpha_n^{i,m}) > \log(u_n^{i,m}) \\ \theta_n^{i,m-1} & \text{otherwise} \end{cases}$$

where $u_n^{i,m} \sim \text{Unif}[0,1]$ and

$$\alpha_n^{i,m} = \frac{p(y|\theta_n^{*i,m})^{\phi_n} p(\theta_n^{*i,m})}{p(y|\theta_n^{i,m-1})^{\phi_n} p(\theta_n^{i,m-1})}$$

The average acceptance ratio can be computed as

$$R_n = \frac{1}{N} \sum_{i=1}^N 1\{\log(\alpha_n^{i,N_{MH}}) > u_n^{i,N_{MH}}\}$$

by using the acceptance results for each particle at the last step.

3. The approximation of $E_\pi[w(\theta)]$ is given by

$$\hat{h}_{N_\phi,N} = \frac{1}{N} \sum_{i=1}^N W_{N_\phi}^i w(\hat{\theta}_{N_\phi}^i)$$

Remark 1: In general, we can employ a random block Metropolis-Hastings algorithm in the mutation step. However, the single-block Metropolis-Hastings is sufficient in this paper, as the model has at most ten parameters.

Remark 2: In general, we can use a mixture of distributions as a proposal, as done in Herbst and Schorfheide (2014). However, the simple random walk Metropolis-Hastings is sufficient in this paper, as our model is much smaller than their medium-scale DSGE models.

C Additional Figures for the Simulation

C.1 Bayesian linear regression model

Figure C.1 presents the estimation errors for the original HME, the modified HME with a uniform weighting density, Geweke's estimator, and the SWZ estimator based on 160 simulations with the Bayesian linear regression model in Section 5.

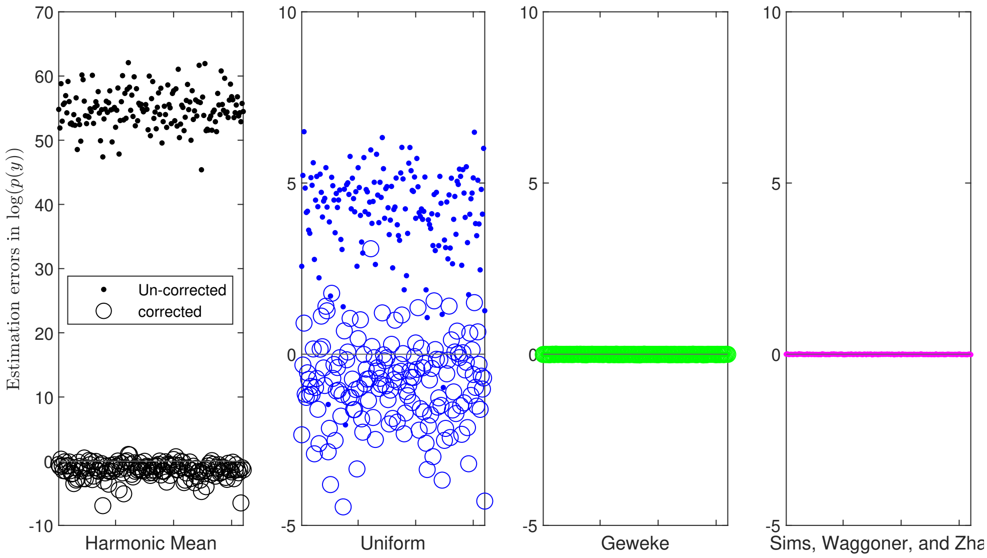


Figure C.1: Estimation errors in a Bayesian linear regression model

Note: This figure presents the estimation errors of the marginal likelihood from the Monte Carlo simulation in the Bayesian linear regression model described in Section 5, where the number of regressors is $n_x = 20$, and the the number of observations is $T = 100$. The errors are measured as deviations from the marginal likelihood computed analytically. Dots represent uncorrected estimates, while circles indicate pseudo-bias corrected estimates. From the left, the plots correspond to the original HME, the modified HME with the uniform weighting density, Geweke (1999)'s estimator, and Sims et al. (2008)'s estimator, respectively. These values are based on 160 simulations.

C.2 The three-equation NK-DSGE model

Figure C.2 presents the estimation errors for the original HME, the modified HME with a uniform weighting density, Geweke's estimator, and the SWZ estimator based on 160 simulations with the DSGE model in Section 5. In line with the findings of Lenk (2009), the original HME tends to overestimate the marginal likelihood, and the pseudo-bias correction significantly improves its accuracy. In contrast, even the modified HME using a simple uniform weighting density is less affected by the pseudo-bias than the original HME, as the improvement from the pseudo-bias correction is much smaller. Geweke's estimator is barely affected by the pseudo-bias, as applying the pseudo-bias correction does not improve its already superior accuracy. The SWZ estimator is also accurate and does not require a separate pseudo-bias correction, as it implicitly incorporates the correction by construction.

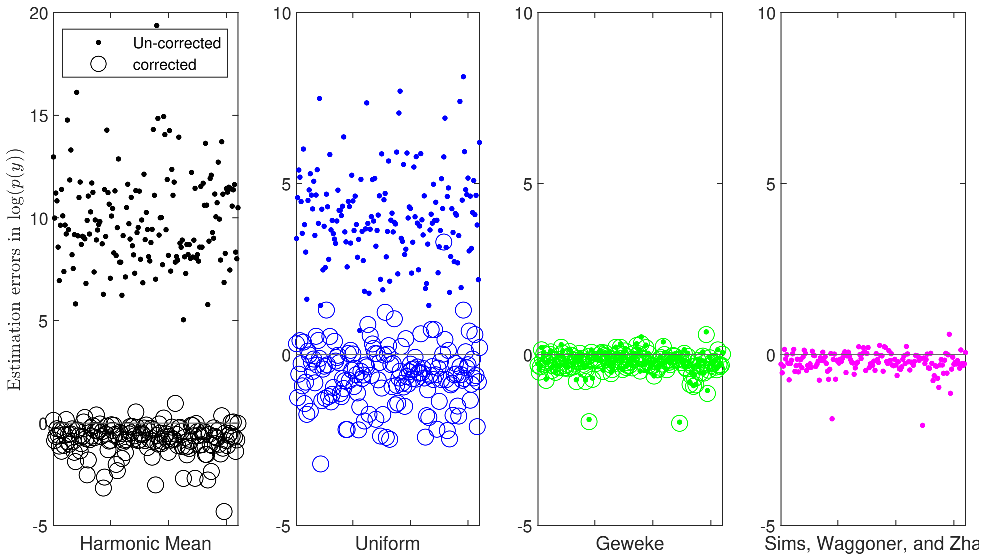


Figure C.2: Estimation errors in a small-scale DSGE model

Note: This figure presents the estimation errors of the marginal likelihood from the Monte Carlo simulation in the extended three-equation New Keynesian DSGE model described in Section 5. The errors are measured as deviations from the sequential Monte Carlo approximation. Dots represent uncorrected estimates, while circles indicate pseudo-bias corrected estimates. From the left, the plots correspond to the original harmonic mean estimator, the modified harmonic mean estimator with the uniform weighting density, Geweke (1999)'s estimator, and Sims et al. (2008)'s estimator, respectively. These values are based on 160 simulations.

D Additional information

Importance sampling. Suppose that you want to numerically compute

$$E_\pi[\phi(\theta)] = \int \phi(\theta)\pi(\theta)d\theta,$$

where ϕ is a deterministic function $\phi : \mathbb{R}^{dim(\theta)} \to \mathbb{R}$, but you cannot sample from the probability distribution π (target distribution). Then, suppose that you can sample from another probability distribution q (proposal distribution). Then, using samples from q, you can compute the expectation as

$$\begin{aligned} E_\pi[\phi(\theta)] &= \int \phi(\theta)\pi(\theta)d\theta \\ &= \int \phi(\theta)\frac{\pi(\theta)}{q(\theta)}q(\theta)d\theta \\ &= E_q[\phi(\theta)W(\theta)] \end{aligned}$$

where $W(\theta) = \frac{\pi(\theta)}{q(\theta)}$. Note that π needs to be absolutely continuous with respect to q, i.e., $\forall\theta, q(\theta) = 0 \Rightarrow \pi(\theta) = 0$.

Derivation of the modified HME. We can derive the modified HME as follows:

$$\begin{aligned} p(y)^{-1} &= \int \frac{w(\theta)}{p(y)}d\theta \\ &= \int \frac{w(\theta)}{p(y|\theta)p(\theta)}\frac{p(y|\theta)p(\theta)}{p(y)}d\theta \\ &= \int \frac{w(\theta)}{p(y|\theta)p(\theta)}\frac{p(y,\theta)}{p(y)}d\theta \\ &= \int \frac{w(\theta)}{p(y|\theta)p(\theta)}p(\theta|y)d\theta \\ &= E_{p(\theta|y)}\left[\frac{w(\theta)}{p(y|\theta)p(\theta)}\right] \end{aligned}$$

where the first equation follows because $w(\theta)$ integrates to unity, and the fourth equation follows directly from the definition of conditional probability. By setting $w(\theta) = p(\theta)$, this estimator becomes equivalent to the original HME.