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

Ryohei Oishi

University College London

March 2026

What is the marginal likelihood? How can we compute it?

- The **marginal likelihood** is a key quantity in Bayesian statistics. Used for model selection, model averaging and so on
- Smets and Wouters (2007): *"The NNS (DSGE) model has a fit comparable to that of Bayesian VAR models"*
- The marginal likelihood:

$$p(y|M) = \int_{\Theta_M} p(y|\theta, M)p(\theta|M)d\theta$$

  How can we compute it? In practice, we often cannot solve this integration analytically

- A naive Monte Carlo integration is computationally inefficient and rarely used

Harmonic mean estimators (HMEs)

- One popular method is the original harmonic mean estimator [Newton and Raftery (1994, JRSSB)]

$$p(y)^{-1} = \int_{\Theta} \frac{1}{p(y|\theta)} p(\theta|y) d\theta$$

However, it has computational bias (simulation pseudo-bias) [Lenk (2009, JCGS)]

- Gelfand and Dey's modified HME is known to be accurate [Gelfand and Dey (1994, JRSSB), Chib and Jeliazkov (2001, JASA)]

$$p(y)^{-1} = \int_{\Theta} \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y) d\theta$$

where $w(\theta)$ can be any density function Derivation

- Popular choices of $w(\theta)$ in macroeconometrics: Geweke (1999) and Sims, Waggoner, and Zha (2008, SWZ)

Three contributions of this paper

1. Shows that the modified HMEs are more robust to simulation pseudo-bias than the original HME

   $\Rightarrow$ Offer a novel explanation why the modified HMEs are accurate

2. Proposes a pseudo-bias correction method for the modified HMEs

   $\Rightarrow$ Provides a computational rationale for the tail truncation of $w(\theta)$ commonly used in practice

3. Confirms these theoretical arguments by Monte Carlo simulations and an empirical application

   $\Rightarrow$ Both the Geweke and SWZ estimators are accurate

   $\Rightarrow$ Revisiting Smets and Wouters (2007) highlights the importance of the pseudo-bias correction

Related Literature

- Estimation of the marginal likelihood by HMEs:

  Gelfand and Dey (1994), Geweke (1999), Sims, Waggoner and Zha (2008), Lenk (2009), Fuentes-Albero and Melosi (2013), Chan and Grant (2015), Hajargasht and Woźniak (2020), Polanska, Price, Spurio Mancini and McEwen (2023), McEwen, Wallis, Price and Mancini (2024), Metodiev, Perrot-Dockes, Ouadah, Irons, Latouche and Raftery (2024)

  Robustness and bias correction apply broadly to these papers

- Model comparison in macroeconomics:

  Smets and Wouters (2007), Hubrich and Tetlow (2015), Inoue and Shintai (2018), Chen, Leeper, and Leith (2022), Bayer, Born, and Luetticke (2024), Dynare, to name a few.

  Support the use of modified HMEs through simulation

  Introduce simulation pseudo-bias in macroeconomic setups

Harmonic mean estimator and simulation pseudo-bias

- Lenk (2009): the harmonic mean estimator overestimates the marginal likelihood due to the simulation pseudo-bias
- The pseudo-bias is best understood by the *importance sampling* interpretation

Detour: Importance sampling

- Suppose you want to numerically compute

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

  but you cannot sample from π, the *target* distribution

- By sampling from q, the *proposal* distribution, you can compute

$$E_\pi[\phi(\theta)] = \int_\Theta \phi(\theta)\frac{\pi(\theta)}{q(\theta)}q(\theta)d\theta = E_q[\phi(\theta)W(\theta)]$$

  where $W(\theta) = \frac{\pi(\theta)}{q(\theta)}$

- The target must be absolutely continuous wrt the proposal
  i.e., $q(\theta) = 0 \Rightarrow \pi(\theta) = 0$
  i.e., The support of π must be a subset of the support of q

Simulation pseudo-bias in the harmonic mean estimator

- The original HME is importance sampling Derivation

$$p(y)^{-1} = \int_\Theta \frac{1}{p(y|\theta)} p(\theta|y)d\theta = p(y)^{-1} \int_\Theta \frac{p(\theta)}{p(\theta|y)} p(\theta|y)d\theta$$

- The target is the prior $p(\theta)$ and the proposal is the posterior $p(\theta|y)$
- $p(\theta|y)$ is much more concentrated than $p(\theta)$
  $\Rightarrow$ Absolute continuity fails **in simulation**

The modified HMEs are more robust to the pseudo-bias

- The modified HME is also importance sampling Derivation

$$p(y)^{-1} = \int_\Theta \frac{w(\theta)}{p(y|\theta)p(\theta)} p(\theta|y) d\theta = p(y)^{-1} \int_\Theta \frac{w(\theta)}{p(\theta|y)} p(\theta|y) d\theta$$

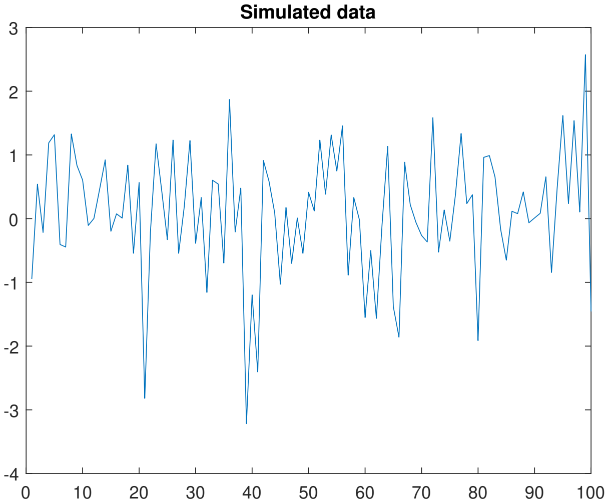
- Now, the target is $w(\theta)$ and the proposal is $p(\theta|y)$
- In practice, $w(\theta)$ is set to approximate $p(\theta|y)$ using the information of the posterior draws; thus they have similar support both in theory and *in simulation*

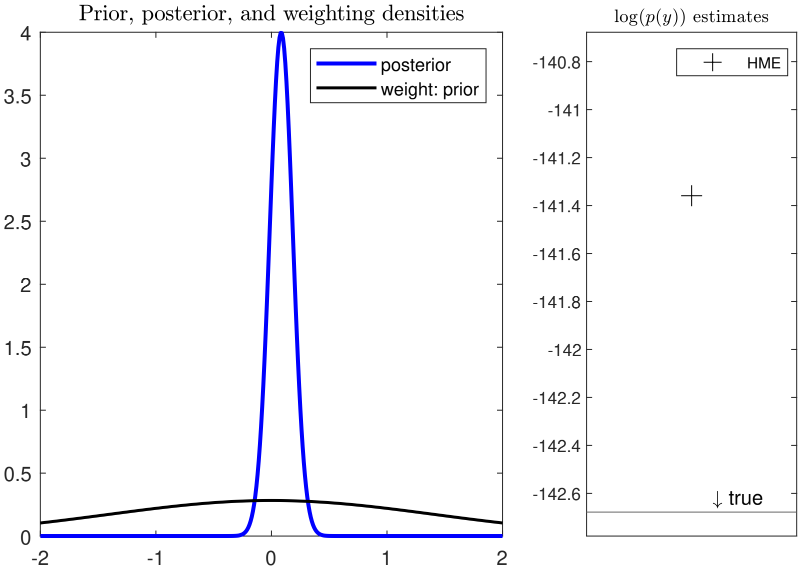
  $\Rightarrow$ The modified HMEs are more robust to the pseudo-bias

An illustrative example

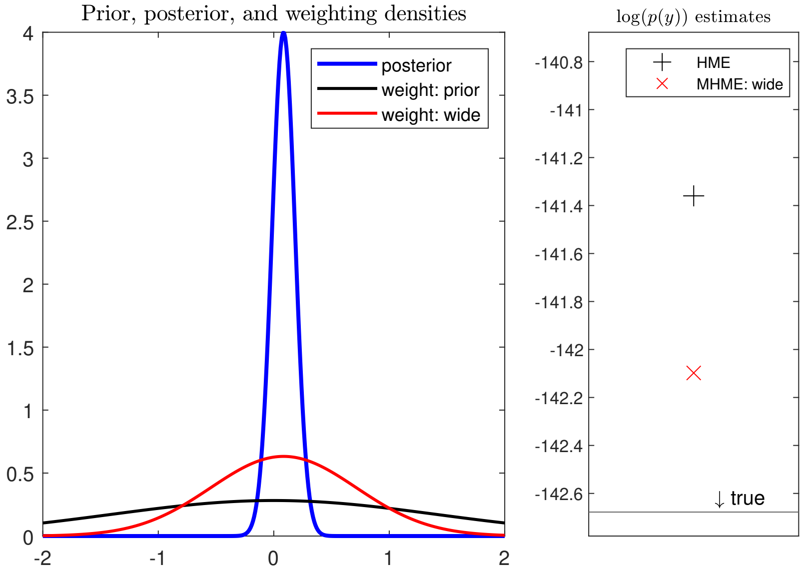
Estimate the mean of $\{y_t\}_{t=1}^{T}$: $y_t|\mu \overset{\text{iid}}{\sim} N(\mu, 1)$

Prior is $\mu \sim N(0, 2)$. $\mu = 0$ for the true DGP.

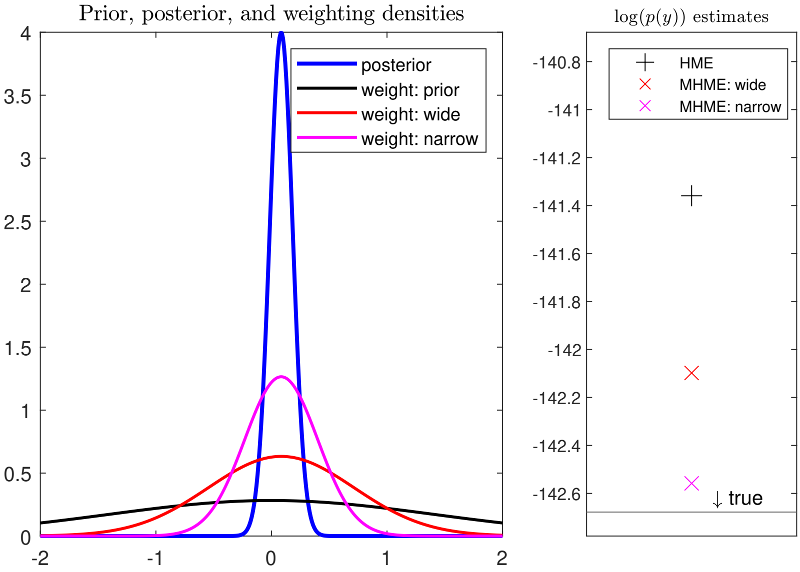




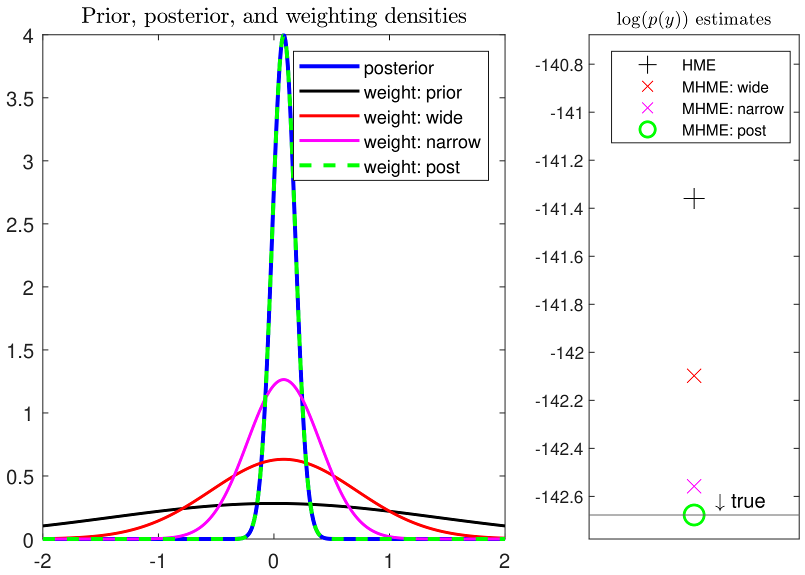
The original HME (Importance sampling) : $p^{-1}(y) = p(y)^{-1} \int_{\Theta} \frac{p(\theta)}{p(\theta|y)} p(\theta|y) d\theta$



A modified HME (importance sampling): $p(y)^{-1} = p(y)^{-1} \int_{\Theta} \frac{w(\theta)}{p(\theta|y)} p(\theta|y) d\theta$



A modified HME (importance sampling): $p(y)^{-1} = p(y)^{-1} \int_\Theta \frac{w(\theta)}{p(\theta|y)} p(\theta|y) d\theta$



A modified HME (importance sampling): $p(y)^{-1} = p(y)^{-1} \int_{\Theta} \frac{w(\theta)}{p(\theta|y)} p(\theta|y) d\theta$

Simulation pseudo-bias correction for the *modified* HMEs

For any subset $A \subseteq \Theta$ with $P(A) \equiv \int_A w(\theta)d\theta > 0$ and $p(\theta|y) > 0$ almost everywhere in A, we have

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

In practice, A can be chosen as

$$A = \{\theta : p(y|\theta)p(\theta) \geq \underline{L}\}, \quad \underline{L} = \min_{\theta \in \{\theta^{(i)}\}_{i=1}^N} p(y|\theta)p(\theta).$$ proof

Discussion:

- The choice of A is flexible. The above choice is convenient because the second term is an uncorrected estimate
- Robustifying $\tilde{w}(\theta)$ by $\tilde{\tilde{w}}(\theta) = 1_A(\theta)P(A)\tilde{w}(\theta)$
- In the literature, $w(\theta)$ often incorporates truncation. This correction method gives a new rationale for such truncation

Skip simulation

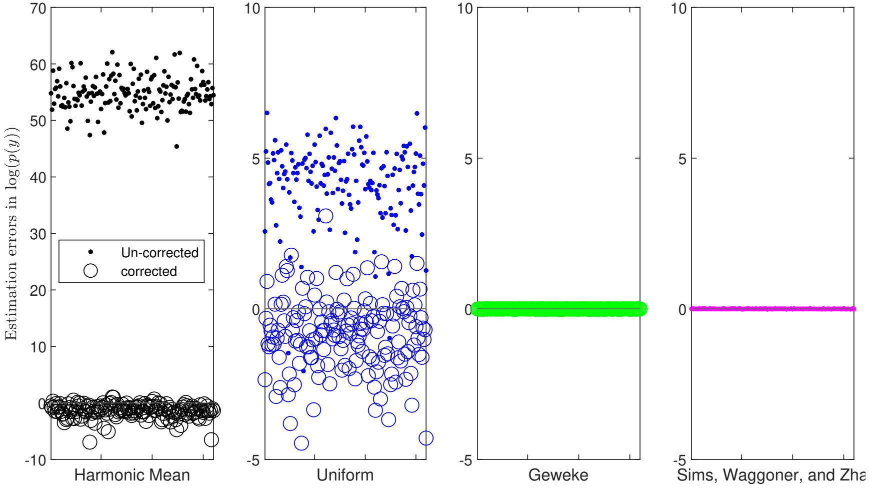
Simulation 1: Bayesian linear regression models

- A Bayesian linear regression model

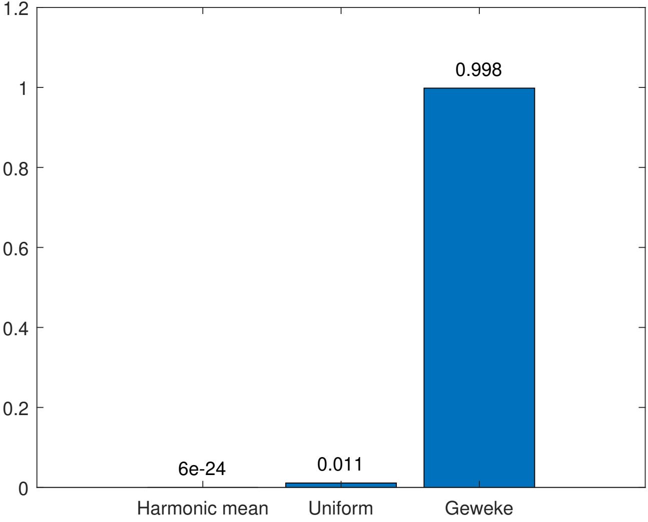
$$y_i = X_i'\beta + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$
$$\beta|\sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 V_0)$$
$$\sigma^2 \sim IG(v_0, v_1)$$

- Simulate $\{y\}_{t=1}^{T}$ for 160 times. The true β and σ^2 are drawn from the prior distribution. $X_i \overset{\text{iid}}{\sim} N(0, I)$
- Change T and n_x
- The marginal likelihood can be computed analytically

Estimation errors (HME - True) when $T = 100, n_x = 20$



Mean $P(A)$



Models with more regressors have larger pseudo-bias

# of X	HM	C-HM	Uniform	C-Uniform	Geweke	SWZ
n_x = 10	27.17	-0.95	1.30	-0.49	-0.00	-0.00
n_x = 20	54.85	-1.37	4.14	-0.91	-0.01	-0.01
n_x = 40	109.79	-2.05	14.12	-1.77	-0.02	-0.02

Table: Mean error when $T = 100$

Scale

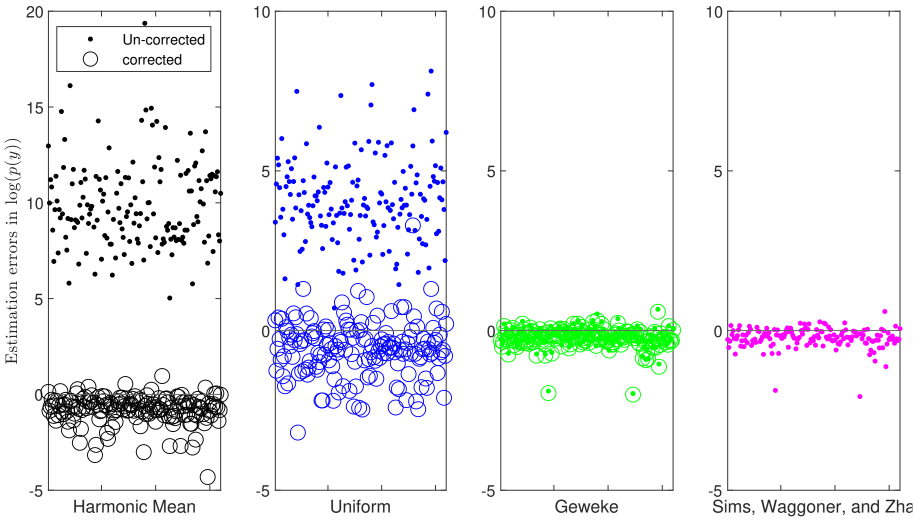
Simulation 2: 3 Eq New Keynesian DSGE Model

$$
\begin{aligned}
y_t &= E_t y_{t+1} - \gamma(i_t - E_t \pi_{t+1}) + e_t^d \\
\pi_t &= \gamma_b \pi_{t-1} + (1-\gamma_b) E_t \pi_{t+1} + \kappa y_t + e_t^s \\
i_t &= \rho_i i_{t-1} + \phi_\pi \pi_t + \phi_y y_t + e_t^m \\
e_t^d &= \rho_d e_{t-1}^d + \sigma_d \varepsilon_t^d, \quad e_t^s = \sigma_s \varepsilon_t^s, \quad e_t^m = \sigma_m \varepsilon_t^m
\end{aligned}
$$

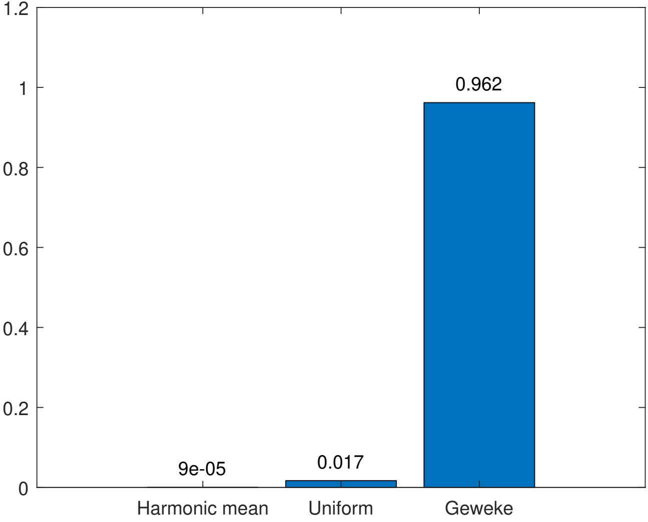
where $\varepsilon_t^d, \varepsilon_t^s, \varepsilon_t^m \sim \mathcal{N}(0,1)$

- Simulate 160 samples of y_t, π_t, i_t from the model. True parameters are drawn from their prior distribution
- I use a sequential Monte Carlo sampler for estimation. It provides a ML estimate as a by-product of the sampling [Herbst and Schorfheide (2014)]

Estimation errors (HME - SMC)



Mean $P(A)$



Smaller pseudo-bias in the simple model

The simple model has no inertia: $\rho_d = \gamma_b = \rho_i = 0$

Model	HM	C-HM	Uniform	C-Uniform	Geweke	SWZ
Extended	9.98	-0.78	4.06	-0.20	-0.24	-0.23
Simple	6.20	-0.56	2.29	-0.50	-0.04	-0.11

Posterior distribution Scale

Revisiting Smets and Wouters (2007)

- Their conclusion: *"The NNS (DSGE) model has a fit comparable to that of Bayesian VAR models"*
- This conclusion is based on a comparison of marginal likelihood
- They use the Laplace approximation, a **local** approximation, which is vulnerable to changes in the mode and computational error in the Hessian at the mode
- Is their conclusion robust when a **global** approximation method, such as the harmonic mean estimators, is used?
- Using HMEs may be unfairly advantageous to the DSGE model because of the pseudo-bias, so bias correction is needed

The ML of the DSGE model is similar to that of the VAR model. Wrong conclusion without pseudo-bias correction

VAR lags	Optimized Minnesota prior
VAR(1)	-965.6
VAR(2)	-914.7
VAR(3)	-913.8
VAR(4)	-917.9
VAR(5)	-919.1

Table: Marginal likelihoods of the VAR models

Sample period	HM	C-HM	Unif	C-Unif	Geweke	SWZ
1966:1-2004:4	-875.7	-917.6	-909.0	-914.1	-914.9	-915.1
w/o subtraction	-847.2	-920.9	-903.3	-921.2	-922.2	-921.5

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

*The marginal likelihoods are computed by
ML(1966:1-2004:4) = ML(1956:1-2004:4) - ML(1956:1-1965:4)

Conclusion

- The modified harmonic mean estimators are more robust to the simulation pseudo-bias
- A proposed pseudo-bias correction method gives rationale for truncation
- Both the Geweke and SWZ estimators are accurate in conventional macroeconometric setups
- The result of Smets and Wouters (2007) is robust and highlights the importance of pseudo-bias correction

Thank You!

Definition of the simulation pseudo-bias in Lenk (2009)

A simulation estimator of the posterior mean of $T(\theta)$ is pseudo-biased when $\exists A \subseteq \Theta$ and small ε such that

$$
\begin{aligned}
&P(\cap_{i=1}^{N}\{\theta^{(i)} \in A|Y\}) > 1-\varepsilon \\
&E[T(\theta)|\theta \in A, Y] \neq E[T(\theta)|Y] \\
&Var[T(\theta)|\theta \in A, Y] < Var[T(\theta)|Y]
\end{aligned}
$$

Kass and Rafetry scale

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

Table: Kass and Raftery (1995) scale

Note: This table summarizes the interpretation of evidence strength proposed by [**?**], based on differences in marginal likelihoods when comparing two models.

Back1 Back2

Derivation of the modified HME

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

By setting $w(\theta) = p(\theta)$, you obtain the harmonic mean estimator

Back

IS interpretation of the HME

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

Back

IS interpretation of the modified HME

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

Back

Derivation of the bias-correction method Back

Recall Bayes' theorem:

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

Then we have,

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

which leads to the bias correction equation

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

Geweke'sestimator

- Geweke (1999) uses the *truncated multivariate normal density* for $w(\theta)$:

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

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, 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 a degree of freedom k

Sims, Waggoner, and Zha's estimator

- SWZ extends Geweke's $h_G(\theta)$ in three ways

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

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

$$\widehat{\Theta} = \{\theta : p(y|\theta)p(\theta) > L_{1-q}, r(\theta) \in [a, b]\}$$

and L_{1-q} is $(1-q)$ percentile of the (unnormalized) posterior density $\{p(y|\theta^i)p(\theta^i)\}_{i=1}^N$, $r(\theta) = \sqrt{(\theta - \widehat{\theta})\widehat{\Omega}^{-1}(\theta - \widehat{\theta})}$, $\hat{\theta}$ is the posterior mode, $f()$ is any one-dimensional probability density defined on the positive reals, $\widehat{S} = \sqrt{\widehat{\Omega}}$ and a and b are set to truncate approximately 10% of the tail

Prior distributions for the 3eq NK model

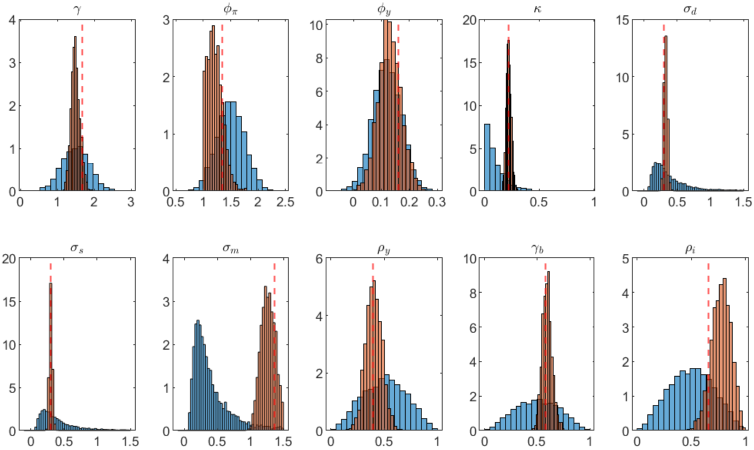
Parameter	Description	Distribution	Mean	Std	Domain
γ	Elasticity of intertemporal substitutiton	Normal	1.5	0.375	$(0, \infty)$
ϕ_π	Taylor rule coefficient on inflation	Normal	1.5	0.25	$(1, \infty)$
ϕ_y	Taylor rule coefficient on output	Normal	0.12	0.05	$[0, \infty)$
κ	Slope of the NKPC	Gamma	0.1	0.1	$(0, \infty)$
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: 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.

Back

Posterior distribution of the 3eq NK model



Back