INTERNATIONAL ECONOMIC REVIEW
Vol. 57, No. 2, May 2016

CONSISTENT VARIANCE OF THE LAPLACE-TYPE ESTIMATORS: APPLICATION TO DSGE MODELS*

By Anna Kormilitsina and Denis Nekipelov[1]

Southern Methodist University, U.S.A.; University of Virginia, U.S.A.

The Laplace-type estimator has become popular in applied macroeconomics, in particular for estimation of dynamic stochastic general equilibrium (DSGE) models. It is often obtained as the mean and variance of a parameter's quasi-posterior distribution, which is defined using a classical estimation objective. We demonstrate that the objective must be properly scaled; otherwise, arbitrarily small confidence intervals can be obtained if calculated directly from the quasi-posterior distribution. We estimate a standard DSGE model and find that scaling up the objective may be useful in estimation with problematic parameter identification. It this case, however, it is important to adjust the quasi-posterior variance to obtain valid confidence intervals.

1. INTRODUCTION

In spite of the popularity of medium-scale dynamic stochastic general equilibrium (DSGE) models in empirical macroeconomic research, their estimation is often associated with practical difficulties. For an applied researcher, the problems with estimation range from the possibility of multiple local solutions to poor identification of model parameters due to the flatness of the objective function in the vicinity of the extremum. In estimations with classical objectives, it has become popular to rely on Bayesian methods by using the Laplace-type estimator (LTE). (See Christiano et al., 2010; Coibion and Gorodnichenko, 2011; Kormilitsina, 2011; Schmitt-Grohé and Uribe, 2011, among others.) The LTE is a Bayesian alternative to the classical extremum estimators. It consists in formulating the so-called "quasi-likelihood" function based on a prespecified statistical criterion, which could be derived from the general method of moments (GMMs) objective, the maximum likelihood, or another classical estimator. The quasi-likelihood function implies the quasi-posterior distribution of model parameters, which can be evaluated using Markov Chain Monte Carlo (MCMC) algorithms, and the estimate is then obtained as the mean or a quantile of the quasi-posterior distribution.

The popularity of the LTE is largely due to the result in Chernozhukov and Hong (2003), who demonstrate that the estimator is both theoretically and computationally attractive. From the computational perspective, the LTE allows one to overcome the curse of dimensionality problem related to the search of the extremum in classical estimation, because it relies on MCMC methods instead of costly search procedures. From the theoretical point of view, Chernozhukov and Hong (2003) establish that under mild assumptions, the LTE is asymptotically equivalent to the corresponding frequentist extremum estimator. Moreover, if the generalized information equality (GIE) holds, then the variance of the quasi-posterior distribution provides a consistent estimate for the variance of the corresponding frequentist estimator. However, if the GIE does not hold, then the variance of the parameter estimate cannot be approximated by the

*Manuscript received February 2014; revised November 2014.

[1] This article has benefited from discussions with Han Hong, Atsushi Inoue, Frank Schorfheide, and James Stock. We would also like to thank the editor, Jesus Fernández-Villaverde, and two anonymous referees for their insightful comments. Please address correspondence to: Anna Kormiltsina, Southern Methodist University, 3300 Dyer Street, Suite 301, Umphrey Lee, Dallas, TX 75275. E-mail: *annak@smu.edu*.

quasi-posterior distribution. Instead, one should transform the quasi-posterior variance using the "sandwich formula" in Chernozhukov and Hong (2003).[2]

In this article, the focus is on situations where the GIE is not satisfied. More specifically, we study the LTE derived using a GMM objective. These estimators are popular in empirical macroeconomic research; however, it is often difficult to ensure the GIE in these problems, because an efficient weighting matrix cannot be reliably computed given the sample size in these applications.[3] Because relying on efficient weighting may significantly hinder the small sample performance of the estimator, researchers often resort to diagonal or other inefficient weighting matrices in formulating the GMM objective.

Within the class of GMM problems, our contribution is the following: First, we demonstrate that even when the weighting matrix is efficient, the GIE may fail if the objective function is not scaled correctly. We show that although in classical GMM estimation, the scaling of the objective function is not essential for the calculation of variance, proper scaling is crucial in LTE, as it modifies the quasi-posterior distribution. In particular, larger scaling implies smaller variance of the quasi-posterior distribution. We therefore conclude that one can calculate confidence intervals directly from quasi-posterior distributions only in efficient estimation problems with proper scaling of the objective function.

Our second contribution is of practical nature. We find that in empirical applications, it may be optimal to force deviation from the GIE by scaling up the objective function. In an empirical exercise, we estimate a simple DSGE model using real and simulated data. We first document that the variance of the quasi-posterior distribution is generally inversely proportional to the scaling parameter. Moreover, the variance of the LTE calculated by properly transforming the variance of the quasi-posterior distribution is robust to the choice of the scaling parameter. However, we find that these conclusions fail when the scaling is absent ($\mu = 1$). In this case, both the variance of the MCMC chains and the variance of the estimator are usually greater than those at $\mu > 1$, contrary to the predictions of theory. This result is indicative of the poor performance of the unscaled LTE, which we relate to the presence of poorly identified parameters and small samples. We confirm this idea in a Monte Carlo experiment where we repeatedly estimate the model using artificially generated data sets. We find that increasing the scaling parameter of the objective function allows one to reduce both the bias and variance of parameter estimates. We therefore conclude that in empirical applications, the scaling of the objective can be used as an instrument to improve the outcome of estimation. It has to be emphasized, however, that confidence intervals of the estimator in this case must be obtained by appropriately transforming the variance of the quasi-posterior distribution.

Implementation of the LTE parallels that in the Bayesian estimation, which has also become a popular approach in empirical macroeconomics (see, for example, An and Schorfheide, 2007; Fernández-Villaverde, 2010; Aruoba and Schorfheide, 2011; Fernández-Villaverde et al., 2012; and references therein). The uniqueness of the LTE, however, is that it relies on Bayesian methods to address alternative, classical estimation problems.[4] The LTE based on the maximum likelihood estimator is most similar to the Bayesian estimation methods commonly used to estimate DSGE models. Both the LTE and the Bayesian approach therefore face similar difficulties in empirical applications, stemming from problematic parameter identification and short data samples. However, although scaling of the quasi-likelihood function may help resolve these problems for LTE, it cannot be helpful for Bayesian estimation. The reason is that the Bayesian approach assumes that the structural parameters are of a stochastic, instead of a deterministic, nature. This means that a Bayesian economist is interested in evaluating the

[2] See theorems 2 and 4 in Chernozhukov and Hong (2003).

[3] This is usually the case in minimum-distance estimation problems that aim to match a large number of impulse responses or moments of the model and data. See Christiano et al. (2010), Kormilitsina (2011), and DiCecio (2009).

[4] Although in this article, we focus on LTE based on GMM objective, our results can be easily extended to include other classical estimation methods where the LTE is commonly applied, for example, extremum estimators that contain nonparametric plug-in components. See Altonji and Segal (1996), Windmeijer (2005), and Newey and Windmeijer (2009), among many others.

whole posterior distribution instead of the finite number of moments. The scaling of the objective function modifies the moments of the quasi-posterior distribution of the LTE in a known manner, which allows one to derive an explicit relationship for the reverse transformation of moments. In Bayesian estimation, the reverse transformation of the overall distribution is required, which cannot be obtained so easily, because scaling the log of likelihood implies the power transformation of the likelihood function.

The article proceeds as follows: In Section 2, we derive the theoretical relationship between variances of the LTE and the GMM estimator in the presence of scaling parameter and test it by estimating a simple stochastic process. In Section 3, we estimate a typical DSGE model using real and artificial data to investigate the effect of the scaling parameter on the variance of the estimated parameter. Finally, Section 4 summarizes the results for a conclusion.

2. LTE FOR MOMENT-BASED MODELS

We consider a standard GMM setting where a model is defined by the moment function $\rho(x,\theta) : \mathcal{X} \times \Theta \mapsto \mathbb{R}^r$, where $\mathcal{X}$ is a subset of $\mathbb{R}^l$ and Θ is a compact convex subset of $\mathbb{R}^p$. r is the number of moment conditions, l is the dimension of the data, and p is the number of estimated parameters.

Parameter of interest θ is identified from the unconditional moment vector $m(\theta) = E[\rho(X,\theta)] = 0$, where function $\rho(\cdot,\cdot)$ might be discontinuous. We assume that the solution to this system exists and is uniquely identified by θ_0 according to the following assumption.[5]

ASSUMPTION 1. *$m(\theta) = 0$ iff $\theta = \theta_0$, where $\theta_0 \in int(\Theta)$.*

Suppose that the data form an i.i.d. sample $\{x_i\}_{i=1}^n$ from the distribution of random variable X. The sample analog of the unconditional moment function is $m_n(\theta) = \frac{1}{n}\sum_{i=1}^n \rho(x_i,\theta)$. The GMM estimator is then defined as $\hat{\theta}^{GMM} = \underset{\theta\in\theta}{\operatorname{argsup}}(Q_n(\theta))t$ where

$$Q_n(\theta) = -\frac{n}{2} m_n(\theta)' \, W \, m_n(\theta) \tag{1}$$

and W is a positive-definite weighting matrix. We assume that the moment function has a first mean square derivative.[6]

ASSUMPTION 2. *There exists a continuous function $\dot{m}(\theta)$ such that*

$$R(\theta,\delta) = m_n(\theta+\delta) - m_n(\theta) - \dot{m}(\theta)\delta$$

satisfies $E[R(\theta,\delta)^2]/\delta^2 \to 0$ for θ in some fixed neighborhood of θ_0 and δ in some fixed neighborhood of the origin.

The sample moment function is assumed to be stochastically equicontinuous.

[5] If the function $m(\theta)$ is differentiable, then a necessary condition for Assumption 1 to hold is that the Jacobi matrix $\partial m(\theta)/\partial\theta$ has rank p.

[6] This assumption gives a high-level condition that assures that the sample moment function is differentiable in mean square. Chen et al. (2003) demonstrate that a local linear representation of a sample function holds for nonsmooth functions as well. In chapter 3.2. of "Weak convergence and empirical processes" by Van Der Vaart and Wellner (1996), the authors give primitive conditions for such expansions to hold; they generally require a "reasonable" bound on the entropy of the class of functions $\{\rho(\cdot,\theta),\ \theta\in\Theta\}$ and the smoothness of the expectation $E[\rho(X,\theta)]$ in θ.

ASSUMPTION 3. *For some fixed neighborhood*[7] of θ_0, $U(\theta_0)$,

$$\sup_{\theta \in U(\theta_0)} \frac{\sqrt{n}\,\|m_n(\theta) - m_n(\theta_0) - m(\theta)\|}{1 + \sqrt{n}\|\theta - \theta_0\|} = o_p(1).$$

The behavior of the GMM estimator is derived from the local linear representation of the objective function using Assumptions 2 and 3. If $m_n(\theta_0)$ satisfies the Lindeberg condition, then the GMM estimator is asymptotically normal with asymptotic variance

$$V_\theta = \left(\dot{m}(\theta_0)'\, W\, \dot{m}(\theta_0)\right)^{-1} \dot{m}(\theta_0)'\, W\, V\, W\, \dot{m}(\theta_0)\ \left(\dot{m}(\theta_0)'\, W\, \dot{m}(\theta_0)\right)^{-1}, \tag{2}$$

where $V = \mathrm{Var}(\rho(X, \theta_0))$.[8]

It is important to note that asymptotic properties of the GMM estimator are robust to the scaling of the objective function $Q_n(\theta)$. This means that for an alternative objective function $\tilde{Q}_n(\theta) \equiv \mu Q_n(\theta)$, the asymptotic properties of the GMM estimator are exactly the same as for the original estimator. It is easy to see from Formula (2) if one thinks of the scaling parameter as embedded into the weighting matrix W. For two weighting matrices W_1 and W_2 that are proportional to each other so that $W_1 = \mu W_2$, the values of V_θ will be identical, because the scaling factor μ will cancel as a result of multiplication of W and its inverse in Formula (2).

Now we consider the LTE. For any $h \in \mathbb{R}^p$, consider the local parameter sequence $\{\theta_{(n)}\}_{n=1}^{\infty}$ in the neighborhood of θ_0 such that each element of the sequence is defined as

$$\theta_{(n)} = \theta_0 + \frac{h}{\sqrt{n}} - \frac{1}{2}\left(\dot{m}(\theta_0)' W \dot{m}(\theta_0)\right)^{-1} \dot{m}(\theta_0)' W m_n(\theta_0).$$

Denote the third term on the right-hand side $T_n/\sqrt{n}$, so the formula above takes the form

$$\theta_{(n)} = \theta_0 + \frac{h}{\sqrt{n}} + \frac{T_n}{\sqrt{n}},$$

where

$$T_n = -\frac{\sqrt{n}}{2}\left(\dot{m}(\theta_0)' W \dot{m}(\theta_0)\right)^{-1} \dot{m}(\theta_0)' W m_n(\theta_0).$$

If $m_n(\theta_0) = O_p(1/\sqrt{n})$, then sequence $\{\theta_{(n)}\}_{n=1}^{\infty}$ concentrates at θ_0 as $n \to \infty$. It is recentered to account for the location of the minimum of the sample objective function. The second-order expansion of the objective function evaluated at an element $\theta_{(n)}$ can then be written as follows:

$$\begin{aligned} Q_n(\theta_{(n)}) = Q_n(\theta_0) &- \frac{n}{4} m_n(\theta_0)' W \dot{m}(\theta_0) \left(\dot{m}(\theta_0)' W \dot{m}(\theta_0)\right)^{-1} \dot{m}(\theta_0)' W m_n(\theta_0) \\ &+ \frac{1}{2} h' \dot{m}(\theta_0)' W \dot{m}(\theta_0)\, h + o_p(1). \end{aligned} \tag{3}$$

Define the quasi-likelihood function of the LTE using the GMM objective (1) as follows:

$$L_n^{\mu}(\theta) \propto e^{\mu Q_n(\theta)},$$

[7] We assume the standard Euclidean norm in $\mathbb{R}^p$ and take into account the fact that $m(\theta_0) = 0$ according to Assumption 1.

[8] We assume that the moments are not degenerate or collinear, meaning that V is a positive definite matrix.

where μ is the scaling parameter. Although in Chernozhukov and Hong (2003) $\mu = 1$, we allow for a general value of μ to emphasize the importance of choosing this parameter correctly. Moreover, in empirical applications, it might be useful to scale the quasi-likelihood differently.[9]

Given some prior distribution $\pi(\theta)$, the quasi-posterior distribution of parameter θ is defined as

$$p_n(\theta) = \frac{\exp\left(\mu Q_n(\theta)\right)\, \pi(\theta)}{\int_\Theta \exp\left(\mu Q_n(\theta)\right)\, \pi(\theta)\, d\theta}. \tag{4}$$

Evaluating the quasi-posterior distribution at an element $\theta_{(n)}$, one can find that because in Equation (3), the first two terms on the right-hand side do not depend on h, they mutually cancel in the numerator and in the denominator in Formula (4). This means that for nondegenerate prior densities, asymptotically, the posterior distribution in Equation (4) will be dominated by the quadratic term in the expansion $\propto \exp(-\mu\frac{1}{2}h'\dot{m}(\theta_0)'\, W\, \dot{m}(\theta_0)h)$. The latter is a multivariate Gaussian density with variance V_μ^{LTE}:

$$V_\mu^{LTE} = \mu^{-1}\left(\dot{m}(\theta_0)'\, W\, \dot{m}(\theta_0)\right)^{-1}. \tag{5}$$

As $n \to \infty$, the quasi-posterior distribution converges to the likelihood in regular models, and its variance converges to V_μ^{LTE}. Therefore, multiplication of the objective function by a constant μ proportionally reduces the variance of the quasi-posterior distribution. This result follows from Theorem 1 presented below.

THEOREM 1. *Consider the total variation of moments norm defined as*

$$\|f\|_{TVM(\alpha)} = \int (1 + |h|^\alpha) f(h)\, dh.$$

Under Assumptions 2 and 3, the posterior distribution along the selected parameter subsequence $\theta_0 + h/\sqrt{n} + T_n/\sqrt{n}$ converges in total variation norm to the Gaussian distribution with covariance matrix V_μ^{LTE}:

$$\left\| \frac{1}{\sqrt{n}} p_n\left(\theta_0 + \frac{h}{\sqrt{n}} + \frac{T_n}{\sqrt{n}}\right) - \frac{\exp\left(-\frac{1}{2}h'\, \left(V_\mu^{LTE}\right)^{-1}\, h\right)}{\left(2\pi\, det\left(V_\mu^{LTE}\right)^{-1}\right)^{\frac{p}{2}}} \right\|_{TVM(\alpha)} \xrightarrow{p} 0. \tag{6}$$

The result of this theorem follows from the proof of theorem 1 in Chernozhukov and Hong (2003). Theorem 1 implies that the asymptotic variance of the quasi-posterior distribution used to obtain the variance of the LTE does not necessarily coincide with the asymptotic variance of the GMM estimator. As a result, the standard errors from the generated posterior distribution are not asymptotically valid. Nevertheless, by comparing Equations (2) and (5), one can see the following relationship between V_θ and V_μ^{LTE}:[10]

$$V_\theta = \mu^2\, V_\mu^{LTE}\, \dot{m}(\theta_0)\, W\, V\, W\, \dot{m}(\theta_0)'\, V_\mu^{LTE}. \tag{7}$$

This formula provides an estimate of V_θ that can be obtained from V_μ^{LTE}. This estimate is asymptotically equivalent to the corresponding GMM estimate of the parameter variance. Because the GMM estimate is independent of the scaling parameter μ, the expression on the

[9] We demonstrate potential benefits of increasing the scaling parameter in Section 3.

[10] This relationship is also implied by theorem 2 in Chernozhukov and Hong (2003).

right-hand side of this equation is expected to be robust to the choice of parameter μ. Sometimes, as happens in the empirical application in Section 3, the right-hand side of Equation (7) may be sensitive to the choice of the scaling parameter. This can be explained by one of the following:

(i) The scale of the Hessian of $\mu Q_n(\theta)$ is compatible with $1/\sqrt{n}$, meaning that the conditions in Pakes and Pollard (1989) are violated and, as a result, Monte Carlo approximation for the variance is not valid. This may generate a problem calculating V_θ with low values of μ.

(ii) Large negative values of $\mu Q_n(\theta)$ lead to quantities exceeding machine infinity for some draws of θ after exponentiation. This means that a significant portion of the MCMC sample consists of highly noisy observations. This may affect V_θ obtained in estimation with large values of the scaling parameter μ.

(iii) The scale of the proposal density is incompatible with the standard deviation of the quasi-likelihood function, meaning that the convergence of the Monte-Carlo approximation to the true posterior distribution may be slow (requiring the order of the number of simulation draws B to be exponential in the sample size n). In other words, V_θ will be incorrectly calculated and may vary with the scaling parameter μ, if the estimation procedure is not performed correctly, so that Markov chains fail to converge for a given length of the chain.

2.1. *Example.* We illustrate the results of Section 2 in the following simple example, where we estimate the mean and variance of the i.i.d. stochastic process generated by independent draws from the normal distribution with mean a and variance σ^2. The estimated parameters are grouped in a vector of interest $\theta = [a; \sigma^2]$. The moment conditions are derived from the definition of the first and the second moments:

$$\rho(X, \theta) = \begin{bmatrix} X - \theta(1) \\ (X - \theta(1))^2 - \theta(2) \end{bmatrix}.$$

For the sample of size n, the model quasi-likelihood function is

$$L_n^\mu(\theta) \sim e^{-\mu \frac{n}{2} m_n(\theta)' W m_n(\theta)},$$

where μ is the scaling parameter, W is a moment weighting function, and the empirical vector of moments $m_n(\theta)$ is defined as

$$m_n(\theta) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta(1)) \\ \frac{1}{n}\sum_{i=1}^{n}\left((x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2 - \theta(2)\right), \end{bmatrix}.$$

Note that the Jacobian matrix of the moment conditions is

$$\dot{m}(\theta_0) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

and the variance of the moment conditions V is

$$V = \begin{bmatrix} \theta(2) & 0 \\ 0 & 2\theta(2)^2 \end{bmatrix}.$$

We generate a sample of $n = 200$ observations from this model parameterized with $\theta = [0.5; 0.25]$. We estimate the model assuming different specifications for the quasi-likelihood function and objective. First, we use either identity or efficient weighting matrix to form

TABLE 1
SIMPLE EXAMPLE: ESTIMATES OF PARAMETER'S VARIANCE–COVARIANCE MATRIX

	$V_a = 0.25$		$Cov(a, \sigma^2) = 0$		$V_{\sigma^2} = 0.125$	
μ	V_μ^{LTE}	V_θ	V_μ^{LTE}	V_θ	V_μ^{LTE}	V_θ
Identity Weighting Matrix						
1	0.28	0.28	0.0081	0.0076	0.13	0.13
5	0.055	0.27	0.0017	0.0082	0.025	0.12
10	0.027	0.27	0.0009	0.0094	0.013	0.13
50	0.0055	0.28	0.00018	0.0096	0.0025	0.13
100	0.0027	0.27	8.4e-005	0.0082	0.0013	0.13
500	0.00055	0.27	1.9e-005	0.01	0.00026	0.13
1000	0.00028	0.28	1e-005	0.012	0.00013	0.13
Efficient Weighting Matrix						
1	1	0.28	−0.00039	0.0083	0.97	0.12
5	0.2	0.28	0.00013	0.0089	0.2	0.13
10	0.1	0.28	−0.00012	0.0082	0.1	0.13
50	0.02	0.28	−1.7e-006	0.0087	0.02	0.13
100	0.01	0.27	−4.5e-005	0.0068	0.01	0.13
500	0.002	0.27	2.3e-006	0.009	0.002	0.13
1000	0.001	0.28	1.9e-007	0.0088	0.001	0.13

NOTES: The first column of the table displays the choice for the parameter μ. Columns 2, 4, and 6 report the elements of the variance–covariance matrix V_μ^{LTE}, and columns 3, 5, and 7 show the elements of the implied matrix V_θ. The first row indicates the true values of the elements of variance–covariance matrix of the GMM estimator.

the distance function. Second, we vary the scaling parameter μ, from 1 to 1000. The quasi-posterior distribution for the model parameter is approximated with the standard random walk Metropolis–Hastings algorithm that produces a chain $\{\theta^{(i)}\}_{i=1}^{B}$ where $B = 10^6$. We calculate the asymptotic variance of the quasi-posterior distribution using the formula

$$\hat{V}_\mu^{LTE} = \frac{n}{B}\sum_{i=1}^{B}(\theta^{(i)} - \hat{\theta})(\theta^{(i)} - \hat{\theta})', \tag{8}$$

where $\hat{\theta} = \frac{1}{B}\sum_{i=1}^{B}\theta^{(i)}$. Then, we evaluate V_θ according to Formula (7). In estimation with identity weighting matrix, $W = I$, this formula produces the estimate

$$\hat{V}_\theta = \mu^2 \hat{V}_\mu^{LTE}\, V\, \hat{V}_\mu^{LTE},$$

and when we use the efficient weighting matrix, $W = V^{-1}$,

$$\hat{V}_\theta = \mu^2 \hat{V}_\mu^{LTE} V^{-1}\, \hat{V}_\mu^{LTE}.$$

Table 1 presents the estimates $\hat{V}_\mu^{LTE}$ and implied $\hat{V}_\theta$ for different values of the scaling parameter μ. The first column of the table displays the choice for the parameter μ. Columns 2, 4, and 6 report the elements of the variance–covariance matrix $\hat{V}_\mu^{LTE}$, and columns 3, 5, and 7 show the elements of the implied matrix $\hat{V}_\theta$. The first row records the true values of the elements of V_θ.

The results reported in Table 1 verify the theoretical relationship in Equation (7). First, columns 2, 4, and 6 of Table 1 reveal that the elements of $\hat{V}_\mu^{LTE}$ are inversely proportional with parameter μ. On the opposite, the estimates $\hat{V}_\theta$ are robust to the choice of the scaling parameter. All estimates of $\hat{V}_\theta$ match the true variance V_θ very closely. One can also notice that estimation involving an efficient weighting matrix provides almost identical estimates $\hat{V}_\mu^{LTE}$ and $\hat{V}_\theta$ when the objective function is not scaled ($\mu = 1$). However, this is not the case for the estimation based on the identity weighting matrix. One can see that the estimate of the parameter variance

TABLE 2
SIMPLE EXAMPLE: COVERAGE PROBABILITIES

	Identity Weighting Matrix				Efficient Weighting Matrix			
	I		II		I		II	
μ	a	σ^2	a	σ^2	a	σ^2	a	σ^2
1	100	99.9	94.7	93.9	93.6	95.4	94.2	94.3
5	91.7	98.9	93.9	94.4	58.5	65.3	95.2	95.1
10	78	91.3	94.2	92.9	46.8	43.4	95.3	93.1
50	43.4	53.2	95.7	92.4	19.7	20.6	94.2	93.9
100	30.4	42.6	94.9	93.6	15.1	18.2	95.4	93.7
500	13.7	18.8	93.3	92.4	7.6	7.1	94.1	93.6
1000	11.2	14.5	93	91.6	5.3	5.4	94.7	93.2

NOTES: The table reports 95% coverage probabilities as percentages. Each number is based on 1000 estimations using randomly generated data sets of 200 observations. In columns I, we use variance V_μ^{LTE} to calculate coverage probability, whereas columns II rely on V_θ.

obtained with the LTE procedure is always different from the variance of the GMM. At the same time, when $\mu = 1$, the estimate of V_μ^{LTE} is close to identity matrix, which is consistent with Formula (5).

An asymptotic covariance matrix is often used by practitioners to construct asymptotic confidence intervals and test hypotheses. Although Chernozhukov and Hong (2003) demonstrate that LTE provides a good coverage of asymptotic confidence intervals, it is important to verify that the coverage is robust to the choice of the scaling parameter μ. Table 2 reports actual coverage probabilities of the 95% asymptotic confidence intervals obtained with the LTE. The table shows a 95% coverage probability for parameters a and σ^2 for estimations using the two choices of the weighting matrix and different scaling parameters μ. In columns I, the coverage probability is calculated using the variance estimate $\hat{V}_\mu^{LTE}$, whereas columns II report coverage probability when using the implied parameter variance $\hat{V}_\theta$. Not surprisingly, the larger the scaling parameter μ in columns I, the smaller is coverage. In this case, the Markov chains become more concentrated around the estimate, resulting in a smaller probability that the confidence interval contains the true parameter value. At the same time, the coverage probabilities calculated using $\hat{V}_\theta$ are just below or at 95% and robust to the choice of the scaling parameter μ. This indicates that intentional scaling of the objective function does not affect the distribution of the estimates, and therefore is not going to influence the outcome of hypothesis testing.

To summarize, the simple estimation exercise presented demonstrates the following: First, the variance of the quasi-posterior distribution V_μ^{LTE} is inversely related with μ. Therefore, unlike in the GMM, researchers must pay attention to the choice of the scaling parameter for the objective function.

Second, the variance of the quasi-posterior distribution V_μ^{LTE} is only equivalent to V_θ when the objective function is scaled appropriately and uses efficient weighting. In estimation with inefficient weighting, the variance of the quasi-posterior distribution does not provide valid standard errors for the estimated parameters. For example, in problems where the GMM objective is based on identity or diagonal weighting matrices, such standard errors are meaningless and cannot be relied on. However, even when there are reasons that prevent a researcher from using the efficient estimation objective, for example, when the moment variance–covariance matrix is poorly defined, the proper variance of the estimate can still be obtained from the variance–covariance matrix V_μ^{LTE} according to Formula (7). Finally, we demonstrate that standard errors obtained using this formula are robust to the choice of the scaling parameter μ. Although greater scaling does not modify the peaks of the quasi-likelihood, it increases its curvature and makes the peaks more pronounced. Therefore, the scaling parameter might become a useful tool in applications with irregular likelihood. The next section provides some confirming evidence of the usefulness of the scaling parameter in empirical applications.

3. LTES FOR DSGE MODELS

3.1. *Model and Estimation Strategy.* Because the estimator is especially popular within the empirical macroeconomic literature, we test the theoretical results in a simple DSGE model derived from a prototype New Keynesian macroeconomic model of a closed economy. This allows one to study the theoretical relationships in a more realistic environment, where the model is more complicated and the data set is relatively small.

The log-linear dynamics of the simple model is summarized by three expectational equations:

$$\begin{aligned}\hat{y}_t &= E_t\hat{y}_{t+1} - \hat{r}_t + E_t\hat{\pi}_{t+1} + \hat{\epsilon}_t,\\ \hat{\pi}_t &= \beta E_t\hat{\pi}_{t+1} + \kappa\hat{y}_t + \hat{\gamma}_t,\\ \hat{r}_t &= \alpha_R\hat{r}_{t-1} + \alpha_\pi\hat{\pi}_t + \alpha_Y\hat{y}_t + \hat{\zeta}_t,\end{aligned}$$

where a hat denotes log-deviation from a steady state, y_t, π_t, and r_t are output, inflation, and the interest rate, parameters β and κ are intertemporal discount factor and parameter of the Phillips curve, and α_R, α_π, and α_Y are the parameters of the monetary policy rule. Finally, ϵ_t, γ_t, and ζ_t are shock processes, and a hat denotes the log deviation from a steady state. The shocks evolve as $AR(1)$ processes:

$$\hat{z}_{t+1} = \rho_z\hat{z}_t + \upsilon^z_{t+1},$$

where $z = \{\epsilon, \gamma, \zeta, \}$, and υ^z_t denote i.i.d. zero mean processes with standard deviation σ_z.

We estimate parameter vector θ:

$$\theta = \{\alpha_R,\ \alpha_\pi,\ \alpha_Y,\ \kappa,\ \rho_\zeta,\ \rho_\gamma,\ \rho_\epsilon,\ 100\sigma_\zeta,\ 100\sigma_\gamma,\ 100\sigma_\epsilon\},$$

and calibrate the remaining parameters and steady-state quantities as follows: $\beta = 0.9$, $Y = 1$, $R = \pi/\beta$, $\pi = 1$, $\alpha_R = 0.7$, $\alpha_\pi = 0.5$, and $\alpha_Y = 0.15$. The vector of observable variables is $x_t = [r_t, y_t, \pi_t]$. We match the elements of the variance–covariance matrix $\mathrm{cov}(x_t, x_{t-l})$, for $l = 0, 1, \ldots, 4$. Because $\mathrm{cov}(x_t, x_t)$ is symmetric, we have only 42 covariance elements to match. The distance function is the quadratic form as defined in Equation (1). All sample moment conditions are summarized in a vector $m_n(\theta)$:

$$m_n(\theta) = [m_{1,n}(\theta); m_{2,n}(\theta); \ldots; m_{42,n}(\theta)],$$

where

$$m_{i,n}(\theta) = q_n^{hkl} - q^{hkl}(\theta),$$

$h, k = r, y, \pi$, and $l = 0, 1, \ldots, 4$ denotes the lag. Empirical estimates q_n^{hkl} are calculated as

$$q_n^{hkl} = \sum_{t=1}^{n}\frac{\hat{q}_t}{n},$$

where $\hat{q}_t = [\hat{q}_t^1, \hat{q}_t^2, \ldots, \hat{q}_t^{42}]$, and each moment $\hat{q}_t^i$ is identified by specific values of h, k, and l and is calculated as

$$\hat{q}_t^i = h_t k_{t-l} - \sum_{j=1}^{n}\frac{h_j}{n}\cdot\sum_{j=1}^{n}\frac{k_j}{n},$$

for $t = l+1, .., n$ and $i = 1, ..., 42$. Theoretical covariances, $q^{hkl}(\theta)$, are obtained as unconditional covariances of first-order approximations to the dynamic processes of model variables h and k_{-l}. In estimation of DSGE models that require higher order approximate solutions due to the importance of nonlinear characteristics,[11] the results in Andreasen et al. (2013) can be used to calculate the theoretical moments. These authors study the statistical properties of the pruned state-space system for second- and third-order approximations to the solutions of DSGE models and derive the closed-form solution to first and second moments and impulse response functions.

We use the standard Random Walk Metropolis Hastings algorithm to draw from the quasi-posterior distribution. In each estimation, we obtain the estimate in two steps. In each step, we create the MCMC chain of one million draws.[12] In the first step, we run the Metropolis Hastings algorithm with the purpose to obtain a better starting point. We use the mean value of the resulting Markov chain as the starting point for the second set of estimations as well as to obtain the weighting matrix W. In the Metropolis-Hastings algorithm, we specify the proposal distribution as multivariate zero-mean normal with variance $c\Sigma$, where Σ is the inverse of the numerical Hessian of the objective function evaluated at the starting element of the MCMC chain. We vary parameter c of the proposal distribution to achieve the acceptance rate in the range of 30–40%. For each estimation, we verify if the algorithm converges, and the resulting Markov Chain is stationary. With this purpose, we visually investigate the Markov chains, including the trace plots, autocorrelation functions, and cumulative sum plots.

We estimate the efficient weighting matrix using the Newey–West estimator. The variance of the quasi-posterior distribution is calculated as in Formula (8), and the variance of the vector of estimates V_θ is obtained from Formula (7), where $\dot{m}(\hat{\theta})$ is the gradient of the moment conditions $m(\theta)$, evaluated numerically at a quasi-posterior estimate $\hat{\theta}$.

3.2. *Results.* We first estimate the model using real data for the period from the third quarter of 1954 till the third quarter of 2010, with the total of 225 observations.[13] The quarterly data include real GDP divided by labor force, GDP deflator, obtained as the ratio of the nominal to real GDP, and the effective annualized federal funds rate. The data are detrended with the standard HP filter with a default smoothing parameter of 1600.

Table 3 presents the variance estimates for parameter θ. The upper part of the table shows the variances V_μ^{LTE} and V_θ for efficient estimation, and the lower part provides the results from estimation using the diagonal weighting matrix.[14] Each row in the table documents the results of an estimation that uses a specific scaling parameter μ to define the quasi-likelihood function. Column 1 specifies the choice for the scaling parameter for each set of estimations. We vary the scaling parameter from 1 to 1000.

The results of estimation using both efficient and diagonal weighting matrices demonstrate that generally, the estimated variance V_μ^{LTE} is indeed smaller for the larger values of the scaling parameter μ. Moreover, for values of μ larger than 10, the elements of V_μ^{LTE} are inversely proportional to μ as is expected from Equation (5). This is especially true for multipliers 10 and 100. When $\mu = 1000$, the inverse proportionality between V_μ^{LTE} and μ is not so clear for some parameters (α_R, κ). We explain this by the possibility of machine errors, resulting from the need to compare the exponents of large numbers. Similarly, the estimates of V_θ are very similar across μ in estimations where the scaling parameter $\mu > 10$. Although the estimation results generally agree with the theory, the theoretical relationship of inverse proportionality of V_μ^{LTE} with μ and robustness of V_θ cannot be validated when $\mu = 1$. In comparison to estimations with

[11] Higher order approximations are necessary in studies of the consequences of uncertainty shocks or macroeconomic determinants behind risk premia (see Fernández-Villaverde et al., 2011, or van Binsbergen et al., 2012).

[12] However, we save only every 100th draw to ensure that there is no autocorrelation between chain elements.

[13] The data are obtained from www.bea.gov, the FRED database, and www.bls.gov.

[14] The diagonal elements are inverse elements of the diagonal of the moment variance matrix. We only report the diagonal elements of the variance–covariance matrices.

TABLE 3
ESTIMATED VARIANCES: REAL DATA

μ	α_R	α_π	α_Y	κ	ρ_ζ	ρ_ϵ	ρ_γ	σ_ζ	σ_ϵ	σ_γ
Efficient Weighting Matrix: V_μ^{LTE}										
1	3.49	43.5	79.3	0.15	4.33	0.709	1.45	26.3	3.06	0.418
10	0.196	1.77	3.26	0.0112	0.242	0.0484	0.0621	1.35	0.232	0.0249
100	0.0124	0.174	0.303	0.00107	0.023	0.00486	0.00606	0.13	0.0231	0.00243
1000	0.000559	0.0192	0.0289	0.000306	0.00238	0.00062	0.000609	0.0133	0.00276	0.000254
V_θ										
1	3.8	96.4	156	0.143	14	1.13	3.5	52.2	5.26	0.737
10	1.04	18.3	40.7	0.0732	4.72	0.602	0.668	19.5	3.32	0.273
100	0.307	16.9	35.4	0.069	4.32	0.602	0.657	17.7	3.29	0.268
1000	0.0698	20.7	40.9	0.563	4.93	1.16	0.66	22	5.4	0.282
Diagonal Weighting Matrix: V_μ^{LTE}										
1	1.06	6.98	8.77	1.34×10^{-4}	0.143	0.603	0.55	1.31	10.2	1.16
10	0.0447	0.226	0.412	1.29×10^{-5}	3.71×10^{-4}	0.0757	0.0626	0.00737	1.18	0.112
100	0.00395	0.0181	0.0345	2.75×10^{-7}	4.31×10^{-5}	0.00699	0.00645	0.00145	0.109	0.0114
1000	0.000396	0.00181	0.00327	3.26×10^{-8}	1.07×10^{-5}	0.00068	0.000634	0.000196	0.0105	0.00112
V_θ										
1	0.978	10.5	24.7	6.02×10^{-6}	0.00885	0.637	1.85	0.762	5.02	1.42
10	0.291	2.02	4.6	9.98×10^{-5}	0.0115	0.375	0.735	0.308	6.62	0.706
100	0.241	1.62	3.73	1.99×10^{-5}	0.0228	0.37	0.671	1.66	6.35	0.702
1000	0.254	1.6	3.02	4.24×10^{-6}	0.000498	0.315	0.581	0.0248	6.09	0.662

NOTES: This table shows the asymptotic variance of the quasi-posterior distribution (V_μ^{LTE}) and the variance of the estimator (V_θ) in estimation with real data of $n = 225$ observations, when the estimated parameter includes the coefficients of the monetary policy rule ($n_\theta = 10$). The upper part uses efficient weighting of the objective, and the lower part uses diagonal weighting. Each row presents results from estimation with a specific objective function. The first column provides the scaling parameter for each objective function. The first and third groups of four rows provide the variance of the quasi-posterior distribution V_μ^{LTE}, whereas the second and fourth groups show the variance V_θ calculated using Formula (7).

TABLE 4
ESTIMATED VARIANCES AND PARAMETERS: $n_\theta = 7$, REAL DATA, W EFFICIENT

μ	κ	ρ_ζ	ρ_ϵ	ρ_γ	σ_ζ	σ_ϵ	σ_γ
Efficient Weighting Matrix: V_μ^{LTE}							
1	0.359	15.3	0.0494	1.72	0.172	0.164	0.324
10	0.0227	0.224	0.0049	0.168	0.0183	0.019	0.0346
100	0.00222	0.0227	0.000489	0.0169	0.00193	0.0019	0.00353
1000	0.00022	0.00225	5.01e-005	0.00167	0.000192	0.000193	0.000351
V_θ							
1	0.301	4.76	0.0559	2.51	0.0446	0.143	0.395
10	0.199	0.843	0.0388	2.07	0.0981	0.15	0.429
100	0.197	1.04	0.0375	2.13	0.116	0.147	0.45
1000	0.192	1.04	0.039	2.08	0.116	0.151	0.444
Parameter Estimates							
1	0.121	0.604	0.949	0.72	0.05	0.157	0.142
10	0.0998	0.884	0.954	0.734	0.0416	0.147	0.128
100	0.0979	0.894	0.954	0.739	0.0396	0.147	0.125
1000	0.0977	0.895	0.954	0.739	0.0394	0.147	0.125

NOTES: See notes to Table 3, with the exception that the estimated parameter excludes the coefficients of the monetary policy rule ($n_\theta = 7$).

$\mu > 1$, both the variance of the quasi-posterior distributions and V_θ in this case are significantly larger for some parameters than what we expect according to the theory. For example, the variance of the monetary policy rule parameters α_π and α_y are 96 and 156, respectively, when $\mu = 1$, whereas they concentrate around 18 and 40 for $\mu = 10$,100, and 1000.

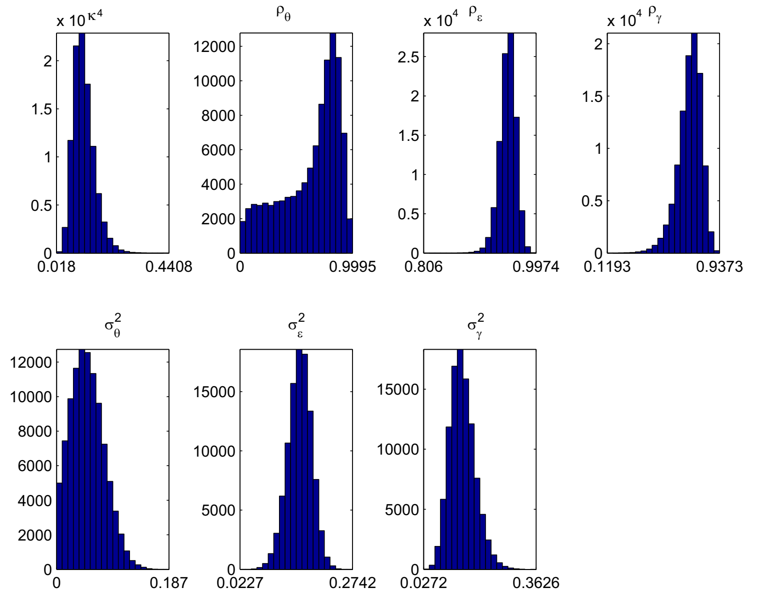
In an attempt to improve the estimation results for an unscaled model ($\mu = 1$), we eliminate from estimation the parameters of the Taylor-type monetary policy rule and reestimate the model with the remaining seven parameters. We set these parameters at values common in the literature: $\alpha_R = 0.7$, $\alpha_\pi = 0.5$, and $\alpha_Y = 0.15$. The resulting variance estimates are presented in Table 4. Comparing the estimated variances when $\mu = 1$ with other choices of μ, we find that the evidence of an inverse theoretical relationship between μ and V_μ^{LTE} improves for an unscaled model. Although in case of $\mu = 1$, the variance estimates are much more similar to those with larger μ, they are still greater than expected. For example, with no scaling, the variance of parameter ρ_ζ is 4.76, which is more than four times larger than those in estimations with $\mu > 10$ ($\rho_\zeta = 0.84$).

To shed more light on the noticeable difference in estimation without quasi-likelihood scaling, we report parameter estimates in this model in the lower part of Table 4.[15] It becomes evident that even in the model where seven parameters are estimated, the estimates of some parameters at $\mu = 1$ are noticeably different from the estimates obtained with larger scaling. It is important to emphasize that the estimates are very similar across estimations with $\mu \geq 10$.

Because increasing the scaling parameter does not change the (local) extrema of the quasi-likelihood,[16] the observed difference in the mean estimate might indicate the asymmetry of the quasi-posterior distribution. Figures 1 and 2 show the distributions of MCMC chains in estimation without scaling and with $\mu = 1000$ correspondingly. Figure 1 reveals that the quasi-posterior distributions of parameter estimates indeed do not look symmetric in estimation with $\mu = 1$. Figure 2 demonstrates that with substantial scaling ($\mu = 1000$), the quasi-posterior distributions are symmetric and have a Gaussian form, as in the asymptotic theory of Chernozhukov and Hong (2003).

There are two possible explanations for the effect we observe. First, the theory that establishes the Gaussian form of quasi-posterior distribution is asymptotic, and the asymmetry of the

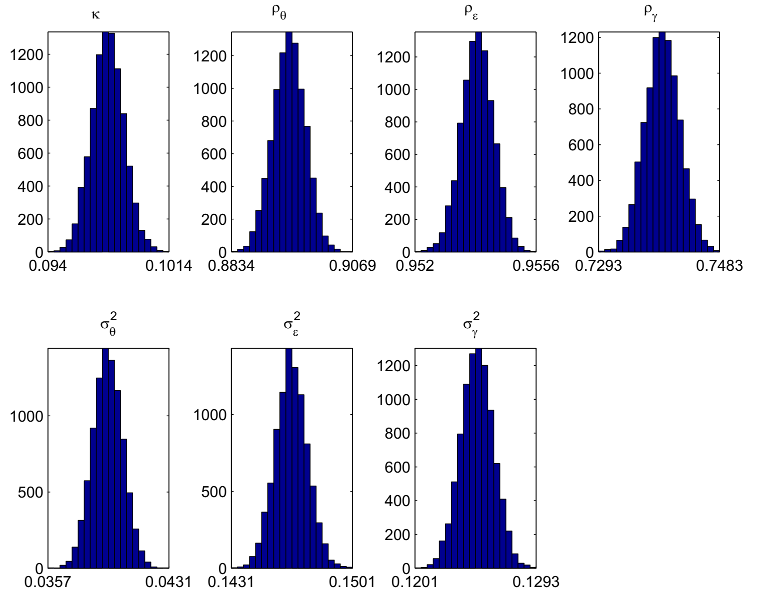
[15] Similar results are obtained when we estimate all 10 parameters including the monetary policy rule.
[16] Unless the weighting matrix changes significantly.



NOTES: The graphs show the quasi-posterior distributions for the estimation using the actual data. This graph is obtained using an MCMC chain of 10 million elements.

FIGURE 1

QUASI-POSTERIOR DISTRIBUTION IN ESTIMATION WITH $\mu = 1$



NOTES: The graphs show the quasi-posterior distributions for the estimation using the actual data. This graph is obtained using an MCMC chain of 1 million elements.

FIGURE 2

QUASI-POSTERIOR DISTRIBUTION IN ESTIMATION WITH $\mu = 1000$

quasi-posterior distributions may be due to the finite data sample. The finite sample problem may become less severe as we reduce the number of estimated parameters. Alternatively or in addition to this, the parameters of the monetary policy rule might be poorly identified. According to Canova and Sala (2009), problematic parameter identification leads to biased parameters and large and uninterpretable confidence intervals. Removing these parameters from estimation therefore may result in less biased estimates and more meaningful confidence intervals.

We now use artificial data sets to estimate the model and see whether the deviation from theory at $\mu = 1$ can be explained by small sample deficiencies or if it can be ascribed to identification problems in the population objective. We first reproduce the same results as the ones reported in Tables 3 and 4. With this purpose, we generate a short and a long data set, with lengths of $T = 200$ and 5000, respectively. The samples are generated by feeding the shock processes into the model dynamic equations. The model is calibrated by θ_0 as follows:

$$\theta_0 = \{0.7, 0.5, 0.15, 0.7,\ 0.8,\ 0.8,\ 0.8,\ 1,\ 1,\ 1\}.$$

We estimate 10 parameters using these data samples. To ensure that the results are not influenced by specific samples, we repeat estimation for 100 samples and consider the average values for the variance estimates. The resulting average variances V_μ^{LTE} and implied V_θ are presented in Tables 5 and 6. We use the star superscripts to report when the variance at no scaling ($\mu = 1$) falls within the 95% confidence intervals of the variance estimates obtained with $\mu >\geq 10$. The number of stars indicates the number of times the estimates fall outside the confidence bands. For example, a parameter with three stars indicates that its variance estimate at $\mu = 1$ is significantly different from all variance estimates at $\mu = 10, 100$, and 1000, whereas one star indicates that the variance at $\mu = 1$ is significantly different from one variance estimate at $\mu = 10, 100$, or 1000.

Table 5 provides estimates obtained with the small sample. The table reveals that the variance of the quasi-posterior distributions decreases at a rate of μ for $\mu > 1$. At the same time, the estimate's variance V_θ is robust to the choice of μ when scaling is present ($\mu > 1$); however, for some parameters, the variance is larger in the absence of scaling ($\mu = 1$) and statistically different from the estimates with $\mu > 1$. In efficient estimation, the largest discrepancy observed for parameter σ_ζ: 70.4 at $\mu = 1$ versus 20.2 at $\mu = 100$. This is the only parameter with statistically larger variance at no scaling than in the presence of scaling. When the diagonal weighting matrix is used to formulate the objective, the positive effect of scaling is more pronounced. Namely, 8 out of 10 estimated parameters demonstrate variance estimates that are statistically larger in the absence of scaling than when some scaling is present. In addition, the asymptotic variance of some parameters is substantially larger without scaling: for example, the variance of σ_ζ at $\mu = 1$ is more than 100 times larger than the one obtained with $\mu = 100$ (13,782 versus 134).

Table 6 provides estimates obtained with the long sample. Generally, the results are very similar to the results of Table 5. As can be seen from this table, even when the data set is large (5000 observations) and the objective uses efficient weighting, some problem remains with the theoretical justification of relationship (7) in the absence of quasi-likelihood scaling. Again, the effect is more pronounced with the diagonal weighting matrix. Overall, although to a smaller extent, the conclusions we draw from the results in Tables 5 and 6 resemble those obtained with the actual data.

We now restrict estimation to seven parameters and estimate the model using the long and short data sets, and diagonal or efficient weighting matrices in formulating the objective function. The resulting elements of the average variance matrix V_θ are presented in Table 7. First of all, we observe that for all four scenarios, the estimates are much more similar across μ, including $\mu = 1$, than in Tables 5 and 6. The strongest similarity of the estimates is observed in estimation with efficient weighting. In fact, we find that for both short and long samples and all parameters, the estimate of V_θ at $\mu = 1$ is not significantly different from any of its estimates at $\mu > 1$. Therefore, we observe that V_θ is robust to the choice of μ for all values of μ, including

TABLE 5

ESTIMATED VARIANCES: SHORT ARTIFICIAL SAMPLE

μ	α_R	α_π	α_Y	κ	ρ_ζ	ρ_ϵ	ρ_γ	σ_ζ	σ_ϵ	σ_γ
Efficient Weighting Matrix: V_μ^{LTE}										
1	1.24	6.11	3.4	1.34	0.208	0.443	0.27	72.8	5.08	2.16
10	0.104	0.225	0.166	0.138	0.03	0.0255	0.0198	2.5	0.443	0.203
100	0.00906	0.0189	0.0176	0.0128	0.00258	0.0028	0.00208	0.215	0.0541	0.0209
1000	0.0011	0.00167	0.00226	0.00144	0.000277	0.000259	0.000193	0.0214	0.00479	0.00215
V_θ										
1	0.795	5.98	2.74	1.62	0.272	0.534	0.286	70.4*	6.38	2.29
10	0.789	1.61	1.33	1.44	0.263	0.26	0.209	16.8	4.68	2.05
100	0.729	1.75	1.3	1.43	0.233	0.288	0.212	19.7	5.8	2.08
1000	1.03	1.76	2.63	1.67	0.257	0.28	0.209	20.2	5.18	2.27
Diagonal Weighting Matrix: V_μ^{LTE}										
1	4.29	207	132	6.89	0.27	7.81	3.45	1,850	199	48.7
10	0.723	2.3	2.74	0.943	0.0714	1.73	0.242	26.9	18.6	4.74
100	0.0966	0.198	0.154	0.113	0.00817	0.132	0.0219	2.75	1.56	0.51
1000	0.0124	0.0165	0.0226	0.00961	0.000802	0.00978	0.00203	0.293	0.151	0.0443
V_θ										
1	6.81	1881***	803***	87.4***	7.2**	10.9	4.95**	13,782***	229*	91*
10	10.8	48.3	82.6	23.7	2.94	32.7	3.14	413	222	50.6
100	6.55	15	17	15	0.965	3.8	1.22	134	83.6	24.7
1000	5.97	10.9	13.4	10.1	0.927	3.94	1.01	130	63.1	23.9

NOTES: Table shows the asymptotic variance of the quasi-posterior distribution (V_μ^{LTE}) and the variance of the estimator (V_θ) in estimation with a *short* data set of artificial data ($n = 200$), when the estimated parameter *includes* the coefficients of the monetary policy rule ($n_\theta = 10$). The upper part uses efficient weighting of the objective, and the lower part uses diagonal weighting. Each row presents results from estimation with a specific objective function. The first column provides the scaling parameter for each objective function. The first and third groups of four rows provide the variance of the quasi-posterior distribution V_μ^{LTE}, whereas the second and fourth groups show the variance V_θ calculated using Formula (7). The numbers represent variance estimates averaged over 100 model estimations, where each estimation uses a unique data set generated from the true model. Stars distinguish parameters with estimates of V_θ obtained with $\mu = 1$ that are significantly different from estimates of V_θ obtained with $\mu = 10$, 100, and 1000. The number of stars indicates the number of times the estimates fall outside the confidence bands. For example, a parameter with three stars indicates that its variance estimate at $\mu = 1$ is significantly different from all variance estimates at $\mu = 10$, 100, and 1000, whereas one star indicates the variance at $\mu = 1$ is significantly different from one variance estimate at $\mu = 10$, 100, or 1000.

TABLE 6
ESTIMATED VARIANCES: LONG ARTIFICIAL SAMPLE

μ	α_R	α_π	α_Y	κ	ρ_ζ	ρ_ϵ	ρ_Y	σ_ζ	σ_ϵ	σ_Y
Efficient Weighting Matrix: V_μ^{LTE}										
1	5.45	10.7	4.65	2.91	0.487	0.403	0.356	136	6.77	4.07
10	0.353	0.65	0.37	0.285	0.0504	0.0386	0.0352	7.63	0.694	0.402
100	0.0408	0.0763	0.0407	0.0286	0.00479	0.0038	0.00348	0.93	0.0666	0.0406
1000	0.0037	0.00668	0.00372	0.0028	0.000511	0.000369	0.000342	0.0805	0.00663	0.00399
V_θ										
1	6.38	12.7*	4.79	2.93	0.54	0.415	0.365	166*	6.99	4.05
10	3.66	6.74	3.75	2.91	0.533	0.392	0.362	78.6	7.05	4.09
100	4.22	7.91	4.12	2.91	0.498	0.384	0.355	95.6	6.72	4.13
1000	3.94	7.04	3.76	2.84	0.545	0.373	0.348	84.7	6.72	4.06
Diagonal Weighting Matrix: V_μ^{LTE}										
1	29.1	53.8	72.3	39.9	0.758	13.5	3.7	846	381	131
10	3.31	3.45	5.56	4.08	0.104	0.962	0.368	73.5	39.5	12.8
100	0.356	0.284	0.701	0.433	0.0106	0.0841	0.0359	7.51	4.07	1.19
1000	0.0392	0.0353	0.0686	0.0444	0.00106	0.00986	0.00374	0.853	0.438	0.126
V_θ										
1	89.9**	247***	812***	161***	6.25**	20.9***	3.09***	3,398***	866***	175**
10	26.4	39.1	74	37.9	2.61	5.63	1.28	585	246	78.4
100	10.7	21.4	28.1	24	1.08	2.4	1.02	201	76.5	43.5
1000	10.7	24.8	25.1	25.7	1	3.1	1.14	208	87.6	50.3

NOTES: See notes to Table 5, with the exception that the long data set is used for estimation ($n = 5000$).

1. However, some noticeable discrepancy at $\mu = 1$ is still present in estimation using diagonal weighting. Namely, the variance estimates at $\mu = 1$ are larger and statistically different from those obtained with $\mu > 1$ for at least two out of seven parameters, for both short and long data samples. Comparing these results with the results in Tables 5 and 6, we conclude that the sensitivity of parameter variance at $\mu = 1$ is more probably associated with poor identification of parameters, although the choice of the weighting matrix may also important, especially in relatively small samples.

The sensitivity of the variance estimates often manifests itself as increased variance at low scaling; therefore, scaling up of the objective function may help produce smaller confidence intervals. If scaling does not increase the bias of the estimate, then adjusting the scaling parameter can become a useful tool helping to improve the quality of confidence intervals. Table 8 reports the bias and the variance of the parameter estimate θ in estimation of 10 model parameters assuming diagonal weighting and a short data sample. The statistics are calculated using the data from the same experiment as the one that produces the lower part of Table 5. The bias is the absolute value of the average deviation of the parameter estimate from the true value, expressed in percentages relative to the true parameter value. The variance is the variance of the parameter estimate calculated over 100 estimations. The table demonstrates that scaling the objective does not have a negative effect on the quality of the estimates. On the contrary, we find that the average bias decreases as we scale up the objective function for all parameters except κ. The largest bias reduction is observed for parameter σ_ζ, where bias reduces more than 100 times. On average, the bias of each parameter decreases by the factor of 16. Besides the positive effect on bias of the estimate, we observe that parameters are estimated more precisely when the scaling is present. Namely, the variance of the parameter estimates decreases with μ for 6 parameters out of 10. For some parameters, such as α_π, α_Y, and σ_ζ, the variance decreases drastically when scaling parameter increases above $\mu = 1$. This is the case for parameters α_π and σ_ζ, where scaling of the objective allows one to reduce the variance by a factor of approximately 100. Therefore, precise estimation of those parameters in the absence of scaling is problematic, and increasing the scaling parameter definitely improves the outcome of estimation.

TABLE 7
V_θ IN A MODEL WITH SEVEN ESTIMATED PARAMETERS

μ	κ	ρ_ζ	ρ_ϵ	ρ_γ	σ_ζ	σ_ϵ	σ_γ
Efficient Weighting Matrix: Long Sample							
1	0.562	0.063	0.346	0.341	0.577	5.23	1.81
10	0.552	0.0621	0.338	0.327	0.576	5.16	1.8
100	0.551	0.061	0.335	0.331	0.581	5.13	1.79
1000	0.554	0.0623	0.337	0.341	0.578	5.19	1.82
Short Sample							
1	0.581	0.0621	0.669	0.326	0.377	4.21	1.7
10	0.482	0.0423	0.248	0.307	0.353	3.8	1.59
100	0.447	0.0423	0.244	0.25	0.344	3.47	1.5
1000	0.528	0.051	0.26	0.22	0.362	3.84	1.5
Diagonal Weighting Matrix: Long Sample							
1	9.09	0.278	14.2***	0.909	3.82	190***	21.8
10	9.49	0.355	2.9	0.844	3.59	99.5	23.6
100	9.7	0.368	1.98	0.831	3.43	84.3	24
1000	9.67	0.384	2.84	0.828	3.8	95	24.1
Short Sample							
1	11.9	1.63***	45.9*	3.2	12.2***	86.7	50
10	8.34	0.468	20.1	1.53	5.4	97.8	37
100	6.93	0.436	8.52	1.32	5.06	75.2	29.7
1000	8.26	0.434	2.84	1.16	5	70.3	30.2

NOTES: Table shows the asymptotic variance estimate (V_θ) in estimations with a *short* ($n = 200$) and *long* ($n = 5000$) data sets of artificial data, when the estimated parameter *excludes* the coefficients of the monetary policy rule ($n_\theta = 7$), using the efficient or diagonal weighting matrix in the objective, and various scaling levels ($\mu = 1$, 10, 100, and 1000). Each row presents results from estimation with a specific objective function. The numbers represent variance estimates averaged over 100 model estimations, where each estimation uses a unique data set generated from the true model. Stars distinguish parameters with estimates of V_θ obtained with $\mu = 1$ that are significantly different from estimates of V_θ obtained with $\mu = 10$, 100, and 1000. The number of stars indicates the number of times the estimates fall outside the confidence bands. For example, a parameter with three stars indicates that its variance estimate at $\mu = 1$ is significantly different from all variance estimates at $\mu = 10$, 100, and 1000, whereas one star indicates that the variance at $\mu = 1$ is significantly different from one variance estimate at $\mu = 10$, 100, or 1000.

TABLE 8
STATISTICAL PERFORMANCE OF PARAMETER ESTIMATE IN THE DSGE MODEL

μ	α_R	α_π	α_Y	κ	ρ_ζ	ρ_ϵ	ρ_γ	σ_ζ	σ_ϵ	σ_γ
Average Bias of the Estimate										
1	14	374	686	0.287	7.24	41.5	19.2	777	122	35.1
10	9.26	7.56	156	12.1	5.97	19.1	7.11	68.1	49.2	8.13
100	6.4	8.46	57.8	6.63	5.5	10	3.6	16.2	31.3	7.01
1000	0.557	14.5	68.4	11.9	5.72	8.87	4.17	6.4	31.8	2.06
Variance of the Estimate										
1	1.1	1.83×10^3	320	9.48	1.09	1.66	2.93	2.24×10^4	75.3	17.8
10	4.39	24.4	15.5	11.1	1.59	5.99	1.2	177	43.2	15.7
100	5.02	16.1	8.83	9.67	1.13	3.97	1.55	144	53.3	22.3
1000	4.55	11.7	9.42	11.8	1.62	2.46	1.5	58.3	47.1	20.5

NOTES: The table shows the average bias of the estimate and its variance across 100 simulations in a model with 10 estimated parameters and a short artificial data set ($n = 200$). The objective uses diagonal weighting, and the scaling of the objective is shown in the first column. Each row in the upper part of the table represents the mean value of the bias, in percentages relative to true value. The lower part rows provide the variance of the estimate calculated across 100 estimations. The results demonstrate that the bias and the variance generally decrease with scaling.

We evaluate the average bias and the variance for the remaining estimations in Tables 5–7 and find that the bias and the variance improve with μ to a larger extent for the results in Table 5 (short data set) than in Table 6 (long data set). In these tables, we estimate 10 model parameters and suspect poor parameter identification. The improvement in the bias is much less noticeable if it exists at all for the results in Table 7, where we believe that the parameters are well identified. Therefore, we conclude that the reliance on the objective scaling can be especially helpful in problems with the possibility of problematic parameter identification. Although we do not provide explicit rules on how to choose the scaling parameter, we recommend running several estimations assuming different scaling of the objective.[17] Then, any parameter μ can be chosen from the range of values for which V_θ is approximately constant across μ. We leave this for future research to develop rules for the optimal choice of objective scaling.

4. CONCLUSION

This article suggests that in empirical estimations using the LTE derived from the GMM, scaling of the objective function could improve the quality of the confidence intervals, especially when parameters are poorly identified. One reason for this is that the objective function may be relatively flat in the vicinity of the proposed estimate, and its scaling would increase the curvature without changing the peaks of the quasi-likelihood, therefore allowing one to estimate parameters more precisely. We confirm this idea by estimating a typical DSGE model from the empirical macroeconomic research. We find that without scaling, the variance of the estimate can be larger than expected in theory, especially when estimation involves parameters that are poorly identified.

It is important to remember, however, that if the GMM objective is inappropriately scaled, then it is no longer possible to obtain confidence intervals directly from the variance of the quasi-posterior distribution. In this article, we demonstrate that the variance of the quasi-posterior distribution and the scaling parameter are inversely related. Therefore, if the variance of the LTE is calculated directly as the variance of the quasi-posterior distribution, then arbitrarily small confidence intervals can be obtained by scaling up the objective. This finding is closely related with the result in Chernozhukov and Hong (2003), who show that if the GIE does not hold, then the variance of the quasi-posterior distribution is not a valid estimate of confidence intervals. Often, it is difficult to ensure the GIE in problems of empirical macroeconomics, where, often, moment conditions are highly correlated, which makes it difficult to obtain a quality estimate for the efficient weighting matrix. In this literature, the LTE is becoming more popular as an alternative to a classical GMM estimator because of its ease of use and asymptotic equivalence with GMM.

REFERENCES

ALTONJI, J. G., AND L. M. SEGAL, "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics* 14(3) (1996), 353–66.

AN, S., AND F. SCHORFHEIDE, "Bayesian Analysis of DSGE Models," *Econometric Reviews* 26(2) (2007), 113–72.

ANDREASEN, M. M., J. FERNÁNDEZ-VILLAVERDE, AND J. RUBIO-RAMÍREZ, "The Pruned State-Space System for Non-linear DSGE Models: Theory and Empirical Applications," NBER Working Papers 18983, National Bureau of Economic Research, Inc., 2013.

ARUOBA, S. B., AND F. SCHORFHEIDE, "Sticky Prices versus Monetary Frictions: An Estimation of Policy Trade-Offs," *American Economic Journal: Macroeconomics* 3(1) (2011), 60–90.

CANOVA, F., AND L. SALA, "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics* 56(4) (2009), 431–49.

[17] Producing several long enough MCMC chains using Matlab is very time-consuming; however, the estimation procedure is much faster if using a Fortran compiler. The Fortran codes to estimate the model in this article are available on request from the authors.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM, "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica* 71(5) (2003), 1591–608.
CHERNOZHUKOV, V., AND H. HONG, "An MCMC Approach to Classical Estimation," *Journal of Econometrics* 115 (2003), 293–346.
CHRISTIANO, L. J., M. TRABANDT, AND K. WALENTIN, "DSGE Models for Monetary Policy Analysis," in B. M. Friedman and M. Woodford, eds., *Handbook of Monetary Economics*, Volume 3, Chap. 7 (Amsterdam: Elsevier, 2010), 285–367.
COIBION, O., AND Y. GORODNICHENKO, "Strategic Interaction among Heterogeneous Price-Setters in an Estimated DSGE Model," *The Review of Economics and Statistics* 93(3) (2011), 920–40.
DICECIO, R., "Sticky Wages and Sectoral Labor Comovement," *Journal of Economic Dynamics and Control* 33(3) (2009), 538–53.
FERNÁNDEZ-VILLAVERDE, J., "The Econometrics of DSGE Models," *SERIEs* 1(1) (2010), 3–49.
———, P. GUERRON-QUINTANA, J. F. RUBIO-RAMÍREZ, AND M. URIBE, "Risk Matters: The Real Effects of Volatility Shocks," *American Economic Review* 101(6) (2011), 2530–61.
———, P. A. GUERRÓN-QUINTANA, AND J. RUBIO-RAMÍREZ, "Estimating Dynamic Equilibrium Models with Stochastic Volatility," NBER Working Papers 18399, National Bureau of Economic Research, Inc., 2012.
KORMILITSINA, A., "Oil Price Shocks and the Optimality of Monetary Policy," *Review of Economic Dynamics* 14(1) (2011), 199–223.
NEWEY, W., AND F. WINDMEIJER, "Generalized Method of Moments with Many Weak Moment Conditions," *Econometrica* 77(3) (2009), 687–719.
PAKES, A., AND D. POLLARD, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57(5) (1989), 1027–57.
SCHMITT-GROHÉ, S., AND M. URIBE, "Business Cycles with a Common Trend in Neutral and Investment-Specific Productivity," *Review of Economic Dynamics*, 14(1) (2011), 122–35.
VAN BINSBERGEN, J. H., J. FERNÁNDEZ-VILLAVERDE, R. S. KOIJEN, AND J. RUBIO-RAMÍREZ, "The Term Structure of Interest Rates in a DSGE Model with Recursive Preferences," *Journal of Monetary Economics* 59(7) (2012), 634–48.
VAN DER VAART, A. W., AND J. A. WELLNER, *Weak Convergence and Empirical Processes* (New York: Springer, 1996).
WINDMEIJER, F., "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics* 126(1) (2005), 25–51.