Inference in Weakly Identified DSGE Models*

Pablo Guerron-Quintana	Atsushi Inoue	Lutz Kilian
North Carolina State University	North Carolina State University	University of Michigan CEPR

Preliminary and Incomplete

April 20, 2009

Abstract

We show that in weakly identified models (1) the posterior mode will not be a consistent estimator of the true parameter vector, (2) the posterior distribution will not be Gaussian even asymptotically, and (3) Bayesian credible sets and frequentist confidence sets will not coincide asymptotically. This means that Bayesian DSGE estimation should not be interpreted merely as a convenient device for obtaining asymptotically valid point estimates and confidence sets from the posterior distribution. As an alternative, we develop a new class of frequentist confidence sets for structural DSGE model parameters that remains asymptotically valid regardless of the strength of the identification. The proposed set correctly reflects the uncertainty about the structural parameters even when the likelihood is flat, it protects the researcher from spurious inference, and it is invariant to the prior in the case of weak identification by construction.

JEL Classification Codes: C32, C52, E30, E50.

KEYWORDS: DSGE models; Bayesian estimation; Identification; Inference; Confidence sets; Bayes factor.

---

*We thank Marco del Negro and Frank Schorfheide for providing access to their data. We thank Yanqin Fan for helpful conversations and participants at Vanderbilt University and at the Triangle Econometrics conference for helpful comments.

1 Introduction

In recent years, there has been growing interest in the estimation of dynamic stochastic general equilibrium (DSGE) models by Bayesian methods. One of the chief advantages of the Bayesian approach compared to the frequentist approach is that the use of prior information allows the researcher to estimate models that otherwise would be computationally intractable or may fail to converge. This feature has made these methods popular even among researchers who think of these methods merely as a convenient device for obtaining model estimates, but would not consider themselves Bayesians otherwise.

At the same time, there is growing evidence that many DSGE models used in empirical macroeconomics are only weakly identified (see, e.g., Canova and Sala 2008). Weak identification manifests itself in a likelihood that is nearly flat across the parameter space. For example, Del Negro and Schorfheide (2008) document that DSGE models that have very different policy implications may fit the data equally well. In particular a New Keynesian model with moderate price rigidities and trivial wage rigidities is observationally equivalent to a model with high wage and price rigidities. As a result, the posterior becomes highly dependent on the priors used by the researcher. For example, Smets and Wouters (2007a, p. 594) note that for their main behavioral parameters "the mean of the posterior distribution is typically relatively close to the mean of the prior assumptions". While this fact does not necessarily pose a problem for genuine Bayesians, it is especially troublesome for frequentist users of these methods because it suggests that we learn nothing from the data.

In this paper, we make two contributions. First, we show that in weakly identified models the usual asymptotic equivalence between Bayesian and frequentist estimation and inference breaks down.[1] The problem is that under weak identification the likelihood no longer asymptotically dominates the posterior, which helps explain the sensitivity of Bayesian DSGE estimates to the prior in practice. As a result, one cannot

[1]See Le Cam and Yang (2000, chapter 8) and the references therein for the large sample correspondence Bayesian and frequentist approaches. For more recent results in the econometrics literature, see Andrews (1994), Chernozhukov and Hong (2003) and Hahn (1997), for example, and Kim (1998) and Phillips and Ploberger (1996) for the nonstationary case.

interpret posterior modes (or means or medians) as frequentist point estimates or treat Bayesian credible sets effectively as frequentist confidence sets. In particular, it is not possible to construct confidence intervals from the quantiles of the posterior distribution or by adding multiples of posterior standard deviations to the posterior mean. Specifically. we show (1) that the posterior mode will not be a consistent estimator of the true parameter vector, (2) that the posterior distribution will not be Gaussian even asymptotically, and (3) that Bayesian credible sets and frequentist confidence sets will not coincide asymptotically. This means that Bayesian DSGE estimation should not be interpreted merely as a convenient device for obtaining asymptotically valid point estimates and confidence sets from the posterior distribution.

Second, as an alternative, we develop a new class of frequentist confidence sets for the structural parameters of DSGE models that remain valid asymptotically regardless of the strength of the identification. In general, the strength of identification is a matter of degree and there is no well-defined threshold that separates strongly identified from unidentified models (see, e.g., Canova and Sala 2008, Iskrev 2008). There is little hope of constructing pre-tests for strong identification nor is it clear that pre-testing would be an appropriate strategy in this context. Our approach is instead based on the premise that the structural model parameters are weakly identified in the sense that the component of the likelihood function that depends on the structural parameter vector is local to zero. As in the weak instruments literature, we think of this assumption as a device that reflects our inability to determine the strength of the identification from the data. The proposed confidence set is obtained by inverting the Bayes factor and does not depend on the prior asymptotically. It is conservative in that a $(1-\alpha)\%$ confidence set has at least $(1-\alpha)\%$ coverage probability asymptotically. The proposed set (correctly reflects the uncertainty about the structural parameters even when the likelihood is flat, it protects the researcher from spurious inference, and it is invariant to the prior in the case of weak identification by construction. Since the Bayes factor is the ratio of the posterior odds to the prior odds, if the likelihood is flat and hence the prior dominates the posterior, the numerator and the denominator of the ratio will tend to cancel, making the proposed confidence set robust to alternative priors.

The work most closely related to our analysis is Moon and Schorfheide's (2009) comparison of frequentist and Bayesian inference in partially identified models. In such models the structural parameter vector of interest can be bounded, but the set of admissible parameter values cannot be narrowed down to a point. Thus, the best a researcher can hope for is to identify the set of parameter values that is consistent with the data. Moon and Schorfheide establish that in partially identified models, Bayesian credible sets tend to be smaller than frequentist confidence sets. This finding is in contrast with the conventional point identified case, in which Bayesian and frequentist sets coincide asymptotically, enabling users to reinterpret Bayesian credible sets as frequentist confidence sets.

Like Moon and Schorfheide we find that Bayesian credible sets and frequentist confidence sets need not coincide asymptotically. In particular, the usual Bayesian credible set does not have the correct asymptotic coverage probability in weakly identified models, preventing its interpretation as a frequentist confidence set. Our analysis differs from Moon and Schorfheide's work, first, in that we focus on weakly point identified parameters rather than set identified parameters. The second difference is that we do not stop at documenting these differences, but propose an alternative confidence set that remains valid regardless of the strength of the identification.

The remainder of the paper is organized as follows. In section 2 we investigate the asymptotic behavior of the posterior distribution in weakly identified models. We establish the failure of the conventional frequentist interpretation of Bayesian posterior estimates. We propose an alternative confidence set based on the inversion of the Bayes factor and prove its asymptotic validity from a frequentist point of view. In section 3 we investigate the finite-sample performance of traditional pseudo-Bayesian methods by simulation. We focus on a commonly used New Keynesian model consisting of a Phillips curve, an investment-savings equation, and a Taylor rule. We demonstrate that the practice of constructing confidende intervals from the posterior of the structural parameters by adding $+/-1.96$ posterior standard deviations to the posterior mode (or mean) results in intervals with serious coverage deficiencies. In some cases, coverage rates of nominal 95% intervals for commonly used sample sizes drop to 44%. Moreover, in our example, coverage accuracy tends to worsen, as the sample size increases. In contrast, the conservative interval proposed in this paper in the simulation has more accurate coverage for all parameters and sample sizes. In section 4, we investigate an empirical example based on a larger scale DSGE model widely used in the DSGE literature (see, e.g., Del Negro and Schorfheide 2008). We focus on the question of what the relative role of wage and price rigidities. We also illustrate the robustness of the proposed confidence sets to alternative choices of priors. The concluding remarks are in section 5.

2 Asymptotic Theory

2.1 Asymptotic Behavior of the Posterior Distribution When Parameters are Weakly Identified

When parameters are strongly identified, the posterior distribution is degenerate about the true parameter value and asymptotically normal after suitable scaling. The latter result is called the Bernstein-von Mises theorem in the Bayesian literature. We will restate a version of the Bernstein-von Mises theorem for the multiparameter non-iid case for expository purposes.[2]

Proposition 1 (Bernstein-von Mises Theorem): Denote the log-likelihood function by $\ell_T(\theta) = \ln L_T(X_1, ..., \theta) = \ln f(X_1, X_2, ..., X_T|\theta)$. Suppose that the following conditions hold:

(a) $\theta_0 \in \text{int}(\Theta) \subset \Re^k$.

(b) The prior density $\pi(\theta)$ is continuous on Θ and $\pi(\theta_0) > 0$.

(c) The likelihood function $L_T(\theta)$ is twice continuously differentiable in a neighborhood of θ_0.

(d) For all $\delta > 0$ there exists $\varepsilon(\delta) > 0$ such that

$$\lim_{T\to\infty} P_{\theta_0}\left[\sup_{\theta\in\Theta\cap B_\delta(\theta_0)} [\ell_T(\theta) - \ell_T(\theta_0)] \le -\varepsilon(\delta)T\right] = 1, \quad (1)$$

where $B_\delta(\theta_0) \equiv \{\theta \in \Theta : \|\theta - \theta_0\| \le \delta\}$.

[2]There are stronger versions of this result. See Bickel and Doksum (2007) and Le Cam and Yang (2000) for more detailed treatments and different versions of this theorem.

(e) There is a matrix valued function $H(\theta)$ such that $\lim_{T\to\infty}\sup_{\theta\in\Theta}|-\frac{1}{T}\nabla_{\theta\theta}\ell_T(\theta)-H(\theta)| \overset{p}{\to} 0$ and $H(\theta_0)$ is positive definite where $\nabla_{\theta\theta}\ell_T(\theta)$ is the Hessian of the log-likelihood function $\ell_T(\theta)$.

(f) The maximum likelihood estimator (MLE) $\hat{\theta}_T$ of θ_0 is strongly consistent, i.e., $\hat{\theta}_T \to \theta_0$ almost surely.

Then for any compact set A

$$\int_{B_T} P(\theta|X_1, X_2, ..., X_T)d\theta \overset{P_{\theta_0}}{\to} P(z \in A). \tag{2}$$

where $B_T = \{\theta \in \Theta : [\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)]^{1/2}(\theta - \hat{\theta}_T) \subset A\}$ and $z \sim N(0_{k\times 1}, I_k)$.

The Bernstein-von Mises Theorem allows a classical interpretation of Bayesian confidence sets. In other words, Bayesian credible sets for $\hat{\theta}_T$ can be viewed as valid classical confidence sets for θ_0 asymptotically. This fact is important because it allows econometricians who are not Bayesians to use the Bayesian apparatus to estimate DSGE models, while interepreting the results in a classical fashion. However, recent research has shown that DSGE models are often only weakly identified (see, e.g., Del Negro and Schorfheide 2008, Canova and Sala 2008). Thus the Bernstein-von Mises theorem does not apply because assumptions (d), (e) and (f) will fail when parameters are not strongly identified. The next result shows formally that the classical interpretation of Bayesian credible sets breaks down when the model is not strongly identified.

Proposition 2 (Posterior Distributions of Exponential Families Under Weak Identification). Consider an exponential family:

$$L_T(x|\theta) \;=\; [\Pi_{t=1}^T h(x_t)] \exp\left[\sum_{j=1}^k \eta_j(\theta) \sum_{t=1}^T T_j(x_t) - TB(\theta)\right] \tag{3}$$

Suppose that

(a)

$$\eta_j(\theta) \;=\; \frac{1}{T}q_j(\theta) + o\left(\frac{1}{T}\right) \tag{4}$$

$$B(\theta) \;=\; \frac{1}{T}r(\theta) + o\left(\frac{1}{T}\right). \tag{5}$$

(b) The likelihood function (3) is correctly specified.

(c) $(1/T)\sum_{t=1}^{T} T_j(x_t) \to E(T_j(x_t))$ almost surely for $j = 1, 2, ..., k$.

Then when a conjugate prior is used, the posterior density almost surely converges to

$$\frac{[\exp(\sum_{j=1}^{k} q_j(\theta)E(T_j(x)) - r(\theta))]}{\int_\Theta E[\exp(\sum_{j=1}^{k} q_j(\theta)T_j(x) - r(\theta))]d\theta} \tag{6}$$

When a more general, not necessarily conjugate, prior $\pi(\theta)$ is used, the posterior density almost surely converges to

$$\frac{\pi(\theta)[\exp(\sum_{j=1}^{k} q_j(\theta)E(T_j(x)) - r(\theta))]}{\int_\Theta \pi(\theta)[\exp(\sum_{j=1}^{k} q_j(\theta)E(T_j(x)) - r(\theta))]d\theta} \tag{7}$$

If there is no unique $\theta_0 \in \Theta$ that maximizes $\sum_{j=1}^{k} \eta_j(\theta)q_j(\theta)E[T_j(x_i)] - r(\theta)$, it can be shown that the maximum likelihood estimator is inconsistent under the stated assumptions and has a nonstandard limiting distribution. Proposition 2 shows that (i) the posterior distribution is not degenerate around the true parameter value when the parameter is weakly identified; (ii) that it is not Gaussian; and (iii) that the limit of the posterior distribution depends on the prior. In other words, the effect of the prior on the posterior will not die out asymptotically, invalidating the usual classical interpretation of Bayesian credible sets. This result is intuitive because information does not accumulate even when the sample size grows when parameters are weakly identified. This means first that the posterior mode no longer coincides with the mean or median. Second, this means that, when the econometrician follows the standard procedure for strongly identified DSGE models and computes the mean (or median or mode) of the posterior distribution as the best guess for the parameter value, the resulting estimator will be inconsistent for the true parameter value.

Condition (a) is an extension of Stock and Wright's (2000) concept of weak identification in GMM to exponential families. It is useful to contrast our notion of weak identification in condition (a) of Proposition (2) to the limiting cases of strong identification and no identification. Strong identification would require that the terms $\eta_j(\theta)$ and $B(\theta)$ in the likelihood function take on values that allow us to solve uniquely for the maximum likelihood estimator. In contrast, lack of identification would correspond to $\eta_j(\theta) = 0$ and $B(\theta) = 0$ such that the likelihood function does not depend on θ. The intermediate case embodied in assumption (a) is that the component of the likelihood function that depends on θ is local to zero. This assumption is designed to represent our inability to determine which of the two limiting cases is a better approximation of reality.[3]

2.2 Bayes Factors and Asymptotically Valid Confidence Sets

As a practical alternative, we propose a frequentist confidence set for parameters in DSGE models that is valid regardless of the strength of identification. Consider testing

$$H_0 : \theta \in B_{\delta_T}(\theta_0)$$

against

$$H_1 : \theta \notin B_{\delta_T}(\theta_0)$$

where $B_\epsilon(\theta_0) = \{\theta \in \Theta : |\theta - \theta_0| \leq \delta_{T,j} \text{ for } j = 1, 2, ..., p\}$, $\Theta \subset \Re^p$ and $\delta_T = [\delta_{T,1}, ..., \delta_{T,p}]' \to 0_{p\times 1}$ as $T \to \infty$.

We define the *Bayes factor* (BF) in favor of H_1 by

$$\text{Bayes Factor}(\theta_0) = \frac{\pi(H_0)p(H_1|X)}{\pi(H_1)p(H_0|X)} \tag{8}$$

where $\pi(H_i)$ and $p(H_i|X)$ are the prior and posterior probabilities of H_i, respectively.

The reduced-form parameters Π are functions of the structural parameters of interest θ:

$$\Pi = g(\theta) \tag{9}$$

where $g : \Theta \to \Re^{dim(\Pi)}$. In DSGE models, Π is the vector of parameters of the state-space model,

$$x_{t+1} = Ax_t + Bw_t, \tag{10}$$

$$y_t = Cx_t + Dw_t, \tag{11}$$

[3]Our assumptions include the possibility of no identification, as discussed in Kadane (1975) and Poirier (1988), as a special case.

where x_t is a vector of possibly unobserved state variables, y_t is a vector of observed variables, $w_t \overset{iid}{\sim} N(0, I)$. The reduced-form parameters A, B, C and D are functions of structural parameters θ (see Fernández-Villaverde, Rubio-Ramírez, Sargent and Watson 2009).

Theorem 1 (Asymptotically Valid Confidence Sets Under Weak Identification)

(a) Θ is non-empty and compact in $\Re^k$.

(b) $\pi : \Theta \to \Re_+$ is continuous on Θ.

(c) The log-likelihood function $\ell_T(\Pi)$ is correctly specified and twice continuously differentiable in θ.

(d) There is a maximum likelihood estimator $\hat{\Pi}_T$ such that $\sqrt{T}(\hat{\Pi}_T - \Pi_0) \overset{d}{\to} N(0, V_\Pi)$ where $V_\Pi = -\text{plim}_{T\to\infty}[(1/T)\nabla_{\Pi\Pi}\ell_T(\Pi_0)]^{-1}$.

(e) $g : \Theta \to \Re^{dim(\Pi)}$ is continuously differentiable and $g(\theta_0) = \Pi_0$.

(f) $\delta_T = [\delta_{T,1}, ..., \delta_{T,p}]'$ satisfies the following condition: If $|\theta_j - \theta_{0,j}| \le \delta_{T,j}$ for $i = 1, ..., p$ then $Dg(\theta)(\theta - \theta_0) = o(T^{-1/2})$.

When $\theta = \theta_0$, then

$$P\left(\text{Bayes Factor}(\theta_0) \le e^{\frac{z'z}{2}}\right) = 1 \tag{12}$$

where $z \sim N(0_{k\times 1}, I_k)$.

Remarks.

1. Assumption (d) only requires the existence of a maximum likelihood estimator of the reduced-form parameters. We do not need to compute the maximum likelihood estimator of Π to obtain the Bayes factor.

2. Weak identification will arise when the rank of the Jacobian of the function g is less than k in the limit. Assumption (e) allows for this possibility as well as the possiblility of strong identification. Therefore, (12) holds true regardless of the strength of the identification.

3. As an example in which assumptions (d) and (e) are satisfied, consider a Taylor rule for monetary policy (Woodford, 2003):

$$R_t = \rho_r R_{t-1} + (1-\rho_r)\phi_\pi \pi_t + (1-\rho_r)\phi_x x_t + \xi_t$$

   where R_t is interest rate, π_t is inflation and x_t is output gap. As the interest rate becomes persistent ($\rho_r \to 1$), the parameters ϕ_π and ϕ_x becomes weakly identified in the sense of the zero-information-limit condition of Nelson and Startz (2007). The reduced form parameters, ρ_r, $(1-\rho_r)\phi_\pi$ and $(1-\rho_r)\phi_\pi$, are strongly identified. Therefore, assumptions (d) and (e) are satisfied.

4. Because Θ is compact and $g(\cdot)$ is continuously differentiable, Assumption (f) is always satisfied if $\delta_T = o(T^{-1/2})$. The condition $\delta_T = o(T^{-1/2})$ can be relaxed if a subset of θ, say α, is weakly identified in the sense that

$$g(\theta) = g_1(\beta) + \frac{1}{T^{\frac{1}{2}}} g_2(\theta).$$

   where $\theta = [\alpha' \ \beta']'$, α is a $p_1 \times 1$ weakly identified parameter vector and β is a $p_2 \times 1$ strongly identified parameter vector.[4] Then if $\delta_{T,j} = o(1)$ for $j = 1, ..., p_1$ and $\delta_{T,j} = o(T^{-1/2})$ for $j = p_1 + 1, ..., p$.

5. Theorem 1 implies that one can obtain level $(1-\alpha)$ confidence sets by inverting the Bayes factor:

$$\{\theta \in \Theta : \text{Bayes Factor}(\theta_0) \leq e^{\frac{\chi_k^2(1-\alpha)}{2}}\} \tag{13}$$

   which satisfies

$$\lim_{T\to\infty} P(\Theta_0) \geq 1-\alpha \tag{14}$$

Note that our approach does not allow the construction of point estimates of θ. In practice, the projection method can be used to construct confidence intervals for individual elements of θ (see Dufour and Taamouti, 2005, and Chaudhuri and Zivot, 2008, for the projection method in linear IV and GMM models, respectively. The level $(1-\alpha)$ confidence interval for the ith parameter θ_j is $(\underline{\theta}_j, \bar{\theta}_j)$ where the lower and upper

[4] This is an extention of Stock and Wright's (2000) concept of weak identification in GMM to that in classical minimum distance estimation.

confidence bounds are

$$\underline{\theta}_j = \min\left\{\theta_j \in \Theta_j : \min_{\theta_{-j}\in\Theta_{-j}} \text{Bayes Factor}((\theta_j, \theta_{-j})) \le e^{\frac{\chi_k^2(1-\alpha)}{2}}\right\}, \quad (15)$$

$$\bar{\theta}_j = \max\left\{\theta_j \in \Theta_j : \min_{\theta_{-j}\in\Theta_{-j}} \text{Bayes Factor}((\theta_j, \theta_{-j})) \le e^{\frac{\chi_k^2(1-\alpha)}{2}}\right\}, \quad (16)$$

and θ_{-j} is the parameter vector excluding θ_j and Θ_{-j} is the parameter space excluding the parameter space for θ_j. These confidence intervals have confidence level $1-\alpha$ by construction. Because the Bayes factor is not differentiable in θ when it is computed via simulation and because the number of parameters of a typical DSGE model is large, evaluation of (15) and (16) is computationally challenging. We replace Θ in (15) and (16) by the set of Monte Carlo realizations which saves the computational burden. Such practice can be justified because the set of Monte Carlo realizations becomes dense in the parameter space as the number of Monte Carlo simulations grows.

In practice one has to choose the radius of the neighborhood $B_{\delta_T}(\theta_0)$. We suggest the following data-dependent method for choosing δ_T. Because $\delta_T \to 0_{p\times 1}$, we have $\pi(H_0) \to 0$, $\pi(H_1) \to 1$, $P(H_0|X) \to 0$ and $P(H_1|X) \to 1$. Thus,

$$BayesFactor(\theta_0) \approx \frac{\pi(H_0)}{P(H_0|X)} = \frac{\frac{1}{|\delta_T|}\pi(H_0)}{\frac{1}{|\delta_T|}P(H_0|X)}$$

where $|\delta_T| = \Pi_{i=1}^{p}\delta_{T,i}$. We typically compute $\pi(H_0)$ and $P(H_0|X)$ by Monte Carlo simulation:

$$\hat{\pi}(H_0) = \frac{1}{M}\sum_{j=1}^{M} I(\theta^{(j)} \in B_{\delta_T}(\theta_0)),$$

$$\hat{P}(H_0|X) = \frac{1}{M}\sum_{j=1}^{M} I(\tilde{\theta}^{(j)} \in B_{\delta_T}(\theta_0)),$$

where M is the number of Monte Carlo realizations, $\theta^{(j)}$ is the jth Monte Carlo realization from the prior distribution and $\tilde{\theta}^{(j)}$ is the jth realization from the posterior distribution. Thus

$$\frac{1}{|\delta_T|}\hat{\pi}(H_0) = \frac{1}{M}\sum_{j=1}^{M} I(\theta^{(j)} \in B_{\delta_T}(\theta_0)), \quad (17)$$

$$\frac{1}{|\delta_T|}\hat{P}(H_0|X) = \frac{1}{M}\sum_{j=1}^{M} I(\tilde{\theta}^{(j)} \in B_{\delta_T}(\theta_0)), \quad (18)$$

Note that the right hand sides of (17) and (18) can be interpreted as a multivariate density estimator based on uniform kernel so that δ_T can be interpreted as the bandwidth. Consider a multivariate version of Silverman's rule of thumb:

$$\delta_{T,j} \;=\; \hat{\sigma}_j \left(\frac{4}{(p+2)T} \right)^{\frac{1}{p+4}} \tag{19}$$

where $\hat{\sigma}_j$ is the standard deviation of the posterior distribution of θ_j. [5] Because the prior and posterior distributions are not necessarily normal and the kernel is not normal, (19) is not necessarily optimal but satisfies assumption (f). Note that if θ_j is strongly identified, $\hat{\sigma}_j = o_p(1)$ and thus $\delta_{T,j} = o_p(T^{-1/2})$; and if θ_j is weakly identified, $\hat{\sigma}_j = O_p(1)$ and $\delta_{T,j} = o_p(1)$. Thus it follows from the discussion in remark 3 for theorem 1 that the resulting choce of δ_T satisfies assumption (f).

3 An Illustrative Example

We investigate the accuracy of both traditional pseudo-Bayesian methods and the proposed alternative by Monte Carlo simulation. Given the computational complexity of applying these econometric methods repeatedly, we select as an illustrative example a small-scale New Keynesian model, which is often used as an example in the related literature (see, e.g. Canova and Sala 2008).

3.1 Simulation Design

Our model setup is from Woodford (2003, pp: 246). The economy consists of a Phillips curve, a Taylor Rule, an investment-savings relationship, and the exogenous driving processes z_t and ξ_t:

$$\pi_t = \kappa x_t + \beta \mathbb{E}_t \pi_{t+1}, \tag{PC}$$

$$R_t = \rho_r R_{t-1} + (1-\rho_r)\phi_\pi \pi_t + (1-\rho_r)\phi_x x_t + \xi_t, \tag{TR}$$

$$x_t = \mathbb{E}_t x_{t+1} - \sigma \left(R_t - \mathbb{E}_t \pi_{t+1} - z_t \right), \tag{IS}$$

$$z_t = \rho_z z_{t-1} + \sigma^z \varepsilon_t^z,$$

$$\xi_t = \sigma^r \varepsilon_t^r.$$

[5] See Wand and Jones, 1995, p.111, for example.

where x_t, π_t and R_t denote the output gap, inflation rate and interest rate, respectively. The shocks ε_t^z and ε_t^r are assumed to be distributed $\mathcal{NID}(0,1)$. The parameters in the model are as follows: β discount factor, σ intertemporal elasticity of substitution, α probability of not adjusting prices for a given firm, θ elasticity of substitution across good varieties, ω parameter controlling disutility of labor supply, ϕ_π and ϕ_x capture the central bank's reaction to changes in inflation and rthe output gap, and $\kappa = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}\frac{\omega+\sigma}{\sigma(\omega+\theta)}$.

Clearly, the parameters contained in κ are not separately identified. This is especially problematic for the parameter α, which measures the degree of price stickiness in the economy and is critical for the analysis of monetary policy. In practice, macroeconomists often fix some parameters such as β, ω and sometimes θ, to allow estimation of α based on κ (see, e.g., Eichenbaum and Fisher 2007), but that procedure is not recommended (see Canova and Sala 2008).

Our Monte Carlo experiment consists of the following steps:

1. We generate $1,000$ synthetic data sets from this New Keynesian model. In generating the data, we set $\alpha = 0.75$, $\beta = 0.99$, $\phi_\pi = 1.5$, $\phi_x = 0.125$, $\omega = 1$,$\rho_r = 0.75$, $\rho_z = 0.90$, $\theta = 6$. These parameter values are standard choices in the macroeconomics literature (see An and Schorfheide 2007, Woodford 2003). We consider two sample sizes: $T = 96$ and $T = 188$. The smaller sample corresponds to the length of quarterly time series starting with the Great Moderation period in 1984 (see Stock and Watson 2002).

2. For each synthetic dataset, we estimate a total of eight parameters: $\Phi = [\alpha\ \phi_\pi\ \phi_x\ \theta\ \rho_r\ \rho_z\ \sigma^r\ \sigma^z]$. The estimation is carried out using Bayesian estimation methods for DSGE models. We characterize the posterior distribution of the parameters of interest using the Random Walk Metropolis-Hasting algorithm documented in An and Schorfheide (2007). We select a flat prior for Φ and denote its probability with $p(\Phi)$. The algorithm involves three steps:

   a. Let $\mathcal{L}(\Phi|Y)$ denote the likelihood of the data conditional on the parameters. Obtain the posterior mode $\widetilde{\Phi} = \arg\max[\ln p(\Phi) + \ln \mathcal{L}(\Phi|Y)]$ using a suitable maximization routine. Let $\widetilde{\sum}$ be the inverse Hessian evaluated at the posterior mode.

b. Draw $\Phi^{(0)}$ from a normal distribution with mean $\widetilde{\Phi}$ and covariance matrix $\varkappa^2\widetilde{\sum}$, where $\varkappa^2$ is a scaling parameter.

c. For $k = 1, ..., N_s$ draw ϑ from the proposal density $\mathcal{N}\left(\Phi^{(k-1)}, \varkappa^2\widetilde{\sum}\right)$. The new draw $\Phi^{(k)} = \vartheta$ is accepted with probability $\min\{1, q\}$ and rejected otherwise. The probability r is given by

$$q = \frac{\mathcal{L}\left(\vartheta|Y\right)p\left(\vartheta\right)}{\mathcal{L}\left(\Phi^{(k-1)}|Y\right)p\left(\Phi^{(k-1)}\right)}.$$

The posterior distributions are characterized using $N_s = 100,000$ iterations after discarding an initial burn-in phase of $5,000$ draws. The scaling parameter was chosen to set the aceptance rate at the values suggested by Roberts et al. (1997).

3.2 Simulation Results

Table 1 compares the coverage accuracy of alternative confidence sets for the model discussed above. The upper panel is for $T = 96$ which corresponds to the sample size of post Great moderation quarterly data. The lower panel for $T = 188$ which corresponds to the standard sample period between 1960 and 2006. We visually inspected draws from the posterior distribution and discarded data sets in which convergence seems to fail. That left 700 synthetic data sets for $T = 96$ and 600 synthetic data sets for $T = 188$. The nominal coverage probability is 0.90. The tuning parameter is chosen by the data-dependent method discussed in section 2.2. To save the computational cost the resutls are based on 5000 draws randomly chosen from the 100,000 draws from the posterior distribution.

The first row of the upper panel focuses on the traditional asymptotic confidence interval that a frequentist user might construct from the posterior mode (or mean or median) by adding $+/-1.96$ posterior standard errors. Some effective coverage rates are well below the nominal rates, and the coverage rate can be as low as 52.5%. Alternatively, a frequentist user may focus on the $1-\alpha$ equal-tailed percentile interval based on the posterior distribution. Such intervals also are frequently reported in the literature (see, e.g., Balke, Brown, and Yücel 2008). For the percentile interval, the coverage rates may drop as low as 27.8%. If we construct the interval by inverting the

Bayes factor, in contrast, all intervals have coverage rates of at least 89.4%.

As the sample size is increased in the second panel, the accuracy of the traditional asymptotic interval improves but the coverage rate can be as low as 71.5% depending on the parameter. The corresponding percentile intervals have coverage rate can be as low as 43.3%. The intervals based on inverting the Bayes factor in all cases have more accurate coverage than the traditional intervals.

We conclude from the illustrative example in Table 1 that traditional interval estimates for Bayesian DSGE model estimates are not reliable and that the proposed alternative interval has the potential to achieve substantial improvements in accuracy.

4 Empirical Application

The model for the empirical section builds on the recent literature on dynamic stochastic general equilibrium models (see, e.g., Altig et al. 2005; Smets and Wouters 2007a,b). Our specification follows very closely that of Del Negro et al. (2007) and Del Negro and Schorfheide (2008), who in turn build on Smets and Wouters (2003) and Christiano, Eichenbaum and Evans (2005). Since this type of environment has been extensively discussed in the literature, we provide only a brief discussion that omits lengthy derivations. The main features of the model can be summarized as follows: The economy grows along a stochastic path; prices and wages are assumed to be sticky à la Calvo; preferences display internal habit formation; investment is costly; and finally, there are five sources of uncertainty: neutral and capital embodied technology shocks, preference shocks, government expenditure shocks, and monetary shocks. Additional details on the formulation and estimation of DSGE models can be found in Fernandez-Villaverde et al. (2009).

4.1 Firms

There is a continuum of monopolistically competitive firms indexed by $j \in [0,1]$ each producing an intermediate good out of capital services, k_j, and labor services, $L_{j,t}$. The technology function is given by

$$Y_{j,t} = k_{j,t}^{\alpha} \left(Z_t L_{j,t}\right)^{1-\alpha} - Z_t \psi,$$

where ψ makes profits equal to zero in the steady state. The neutral technology shock, Z_t, grows at rate $z_t = \log(Z_t/Z_{t-1})$ which is assumed to follow the process[6]

$$z_t = (1 - \rho_z)\gamma + \rho_z z_{t-1} + \sigma_z \epsilon_{z,t},$$

where $\epsilon_{z,t}$ is distributed $\mathcal{NID}(0,1)$. Firms rent capital and labor in perfectly competitive factor markets.

Firms choose prices to maximize the present value of profits; prices are set in a Calvo fashion; that is, each period, firms optimally revise their prices with an exogenous probability $1 - \zeta_p$. If, instead, a firm does not re-optimize its price, then the price is updated according to the rule: $P_{j,t} = (\pi_{t-1})^{\iota_p} (\pi_*)^{1-\iota_p} P_{j,t-1}$, where π_{t-1} is the economy-wide inflation in the previous period, π_* is steady state inflation and $\iota_p \in [0,1]$.

There is a competitive firm that produces the final good using intermediate goods according to the technology

$$Y_t = \left[\int_0^1 Y_{j,t}^{1/(1+\lambda_{f,t})} dj\right]^{1+\lambda_{f,t}}.$$

Here $\lambda_{f,t}$ is the degree of monopoly power and evolves according to the process $\log \lambda_{f,t} = (1 - \rho_{\lambda_f}) \log \lambda_f + \rho_{\lambda_f} \log \lambda_{f,t-1} + \sigma_{\lambda_f} \epsilon_{\lambda,t}$. The shock $\epsilon_{\lambda,t}$ is assumed to be $\mathcal{NID}(0,1)$.

4.2 Households

The economy is populated by a continuum of households indexed by i. Every period households must decide how much to consume, work, and invest. In addition, they must choose the amount of money to be sent to a financial intermediary. Agents in the economy have access to complete markets; such an assumption is needed to eliminate wealth differentials arising from wage heterogeneity. Households maximize the expected present discounted value of utility

$$\mathbb{E}_0^i \sum_{t=0}^{\infty} \beta^t \left[\log(C_{i,t} - hC_{i,t-1}) - \phi_t \frac{L_{i,t}^{1+v_l}}{1+v_l}\right] \tag{20}$$

subject to

$$P_t C_{i,t} + P_t \left(I_{i,t} + a(u_{i,t})\overline{K}_{i,t}\right) + B_{i,t+1} = R_t^K u_{i,t}\overline{K}_{i,t} + W_{i,t}L_{i,t} + R_{t-1}B_{i,t} + A_{i,t} + \Pi_t + T_{i,t},$$

[6] The growth term is needed to have a well-defined steady state around which we can solve the model.

and

$$\overline{K}_{i,t+1} = (1-\delta)\overline{K}_{i,t} + I_{i,t}\left(1 - \Gamma(\frac{I_{i,t}}{I_{i,t-1}})\right).$$

Here, $\mathbb{E}_t^i$ is the time t expectation operator conditional on the information set of household i; ϕ_t is a preference shifter that follows the process $\log\phi_t = (1-\rho_\phi)\log\phi + \rho_\phi\log\phi_{t-1} + \sigma_\phi\epsilon_{\phi,t}$ with $\varepsilon_{\phi,t}$ distributed $\mathcal{N}(0,1)$; preferences display internal habit formation measured by $h \in (0,1)$; and Γ is a function reflecting the costs associated with adjusting the investment portfolio. This function is assumed to be increasing and convex satisfying $\Gamma(e^\gamma) = \Gamma'(e^\gamma) = 0$ and $\Gamma''(e^\gamma) > 0$ in the steady state. $T_{j,t}$ corresponds to lump-sum transfers from the government to household i. $B_{i,t}$ is the individual demand for one-period government bonds, which pay the gross nominal interest rate R_t. As in the related literature, it is assumed that physical capital can be used at different intensities (see, e.g., Altig et al. 2005). Furthermore, using the capital with intensity $u_{i,t}$ yields the return $R_t^K u_{i,t}\overline{K}_{i,t}$ but entails the cost $a(u_{i,t})$, which satisfies $a(1) = 0; a''(1) > 0; a'(1) > 0$. Finally, the term $A_{i,t}$ captures net payments from complete markets while Π_t corresponds to profits from producers.

4.3 Wage Setting

Following Erceg et al. (2000), we assume that each household is a monopolistic supplier of a differentiated labor service, $L_{i,t}$. Households sell these labor services to a competitive firm that aggregates labor and sells it to final firms. The technology used by the aggregator is

$$L_t = \left[\int_0^1 L_{i,t}^{1/(1+\lambda_w)} dj\right]^{1+\lambda_w}, \quad 0 < \lambda_w < \infty.$$

It is straightforward to show that the relationship between the labor aggregate and the aggregate wage, W_t, is given by

$$L_{i,t} = \left[\frac{W_{i,t}}{W_t}\right]^{-(1+\lambda_w)/\lambda_w} L_t.$$

To induce wage sluggishness, we assume that households set their wages in Calvo fashion. In particular, with exogenous probability ζ_w a household does not re-optimize wages each period. Hence, wages are set according to the rule of thumb $W_{i,t} = (\pi_* e^\gamma)^{1-\iota}(\pi_{t-1}e^{z_{t-1}})^{\iota_w} W_{i,t-1}$.

4.4 Government

As in most of the recent New Keynesian literature, we assume a cashless economy (Woodford, 2003). The monetary authority sets the short-term interest rate according to a Taylor rule. In particular, the central bank smoothes interest rates and responds to deviations of actual inflation from steady-state inflation, π_*, and deviations of output from its target level, Y_t^*.

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_r} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}\right]^{1-\rho_r} \exp(\sigma_r \epsilon_{r,t}). \tag{21}$$

The term $\epsilon_{r,t}$ is a random shock to the systematic component of monetary policy and is assumed to be standard normal; σ_r is the size of the monetary shock. This is the same Taylor rule used in Del Negro et al. (2006, 2008). R^* corresponds to the steady-state gross nominal interest rate.

Finally, we assume that government spending is given by $G_t = (1 - 1/g_t)\, Y_t$ where g_t follows the exogenous process $\log g_t = (1-\rho_g)\log g + \rho_g \log g_{t-1} + \sigma_g \epsilon_{g,t}$. The government uses taxes and one-period bonds to finance its purchases.

4.5 Data and Estimation

We follow Del Negro and Schorfheide (2008) in estimating the model using five observables: real output growth, per capita hours worked, labor share, inflation (annualized), and nominal interest rates (annualized). We use their dataset, which are quarterly and correspond to the period $1982.Q1 - 2005.Q4$. We also set our priors to the "Non-dogmatic Agnostic," "Low-rigidities," and "High-rigidities" priors employed in Del Negro and Schorfheide (see Tables 1 through 3 in their paper).

The parameter space is divided into two sets: $\Theta_1 = [\alpha\ \delta\ g\ L^*\ \psi]$, which is not estimated, and $\Theta_2 = [r_*\ \gamma\ \lambda_f\ \pi_*\ \zeta_p\ \iota_p\ \zeta_w\ \iota_w\ \lambda_w\ \Gamma''\ h\ a''\ v_l\ \psi_1\ \psi_2\ \rho_r\ \rho_z\ \sigma_z\ \rho_\phi\ \sigma_\phi\ \rho_\lambda\ \sigma_\lambda\ \rho_g\ \sigma_g\ \sigma_r\ L_{adj}]$, which is. The following values are used for the first set of parameters: $\alpha = 0.33$, $\delta = 0.025$, $g = 0.22$, $L^* = 1$, $\psi = 0$. Although these values are standard choices in the DSGE literature, some clarifications are in order. As in Del Negro and Schorfheide (2008), our parametrization imposes that firms make zero profits in steady state. We also assume that households work one unit of time in steady state. This assumption in turn has two implications: First, the parameter ϕ is endogenously determined by the optimality conditions in the model. Second, because hours worked have a mean different from that in the data, the measurement equation in the state space representation is

$$\log L_t(data) = \log L_t(model) + \log L_{adj},$$

Here, the term L_{adj} is required to match the mean observed in the data. Rather than imposing priors on the great ratios as in the Del Negro and Schorfheide, we follow the standard practice (Christiano et al. 2005) of fixing the capital share, α, the depreciation rate, δ, and the share of government expenditure on production, g.

The posterior distribution of the parameters in the set Θ_2 are characterized using the Random walk-Metropolis-Hasting algorithm outlined in Section 3. A total of three independent chains each of length $100,000$ were run. We conducted standard tests to check the convergence of each chain (see Gelman et al. 2004).

4.6 What Do the Data Tell Us about the Relative Importance of Wage and Price Rigidities?

Table 2 summarizes the posterior means, medians and modes as well as the posterior standard deviations, as shown in Table 6 of Del Negro and Schorfheide (2008).[7] For each structural parameter, we also show the 90% Bayesian credible interval and the proposed 90% confidence interval based on inverting the Bayes factor (*BF interval*).

For our purposes, the parameters of greatest interest are ζ_p and ζ_w, which quantify the degree of price and wage rigidities, respectively. Del Negro and Schorfheide found that the posterior of these parameters was heavily influenced by their prior, so a researcher entering a prior favoring one of these rigidities, would inevitably arrive at a posterior favoring that same rigidity. This finding suggests that a properly constructed confidence band should be wider. A researcher naïvely interpreting the credible sets as frequentist confidence sets would have concluded that these same parameters are fairly

[7]The attentive reader may notice that our posteriors somehow differ from those in Del Negro and Schorfheide (2008). This is so because, as previously explained, we opt not to use priors on the great ratios. For the discussion below, these differences are immaterial.

tightly estimated.

To put our findings in context, note that microdata-based estimates suggest that price contracts last on average about 2.3 quarters (Klenow and Kryvtsov, 2008). Based on the frequentist confidence intervals, a researcher would conclude that the length of those price spells is incompatible with the macro evidence in Table 2. In contrast, a researcher equipped with the *BF interval* would view Klenow and Kryvtsov's findings as perfectly consistent with the results from the Bayesian estimation exercise (the lower bound of the interval implies that prices are reset every 1.9 quarters).[8] When we turn to wage stickiness, the frequentialist method favors a model with a fairly flexible wage setting (the longest wage contract only lasts for 1.5 quarters). Our approach, however, suggests that the data are compatible with a model displaying wage contracts of length up to 2.3 quarters.

Tables 3 and 4 provides evidence that the *BF interval* is not very sensitive to the choice of prior. We compare the low rigidity and high rigidity priors explored by Del Negro and Schorfheide (2008). The *BF interval* is designed to help protect researchers from overly optimistic inferences. It allows applied users who do not self-identify as Bayesians to compute asymptotically valid confidence sets from DSGE models estimated by Bayesian methods, even when conventional methods relying on the asymptotic equivalence of Bayesian and frequentist estimation and inference would be invalid.

5 Concluding Remarks

An attractive feature of Bayesian DSGE estimation methods is that they facilitate the estimation of models that are too large to be estimated reliably by conventional maximum likelihood methods. This feature has made these methods popular even among researchers who think of these methods merely as a convenient device for obtaining model estimates, but would not consider themselves Bayesians otherwise. If the DSGE model is only weakly identified, however, Bayesian posterior estimates tend to be dominated by the prior and the usual asymptotic equivalence between frequentist and

[8]The lenght of price contracts is defined as $\frac{1}{1-\zeta_p}$, where ζ_p is the probability of not re-optimizing prices today.

Bayesian methods of estimation and inference breaks down. We showed that attempts to construct classical confidence sets from the posterior, whether based on percentile intervals or by adding multiples of posterior standard deviations to posterior modes, are invalid if the model is weakly identified. Moreover, the posterior mode, mean and median is an inconsistent estimator of the true structural parameter values. Given mounting evidence that many DSGE models used in the literature suffer from weak identification problems, this finding suggests caution in interpreting posterior modes as traditional point estimates and highlights the limitations of traditional confidence sets constructed from the posterior.

We proposed an alternative frequentist confidence set that remains asymptotically valid regardless of the strength of the identification. We showed that the proposed confidence set tends to have higher coverage accuracy than the alternative methods we showed to be theoretically invalid. The proposed set is designed to help applied users separate the information conveyed by the data from the information conveyed by the prior. This is an especially useful feature for non-Bayesian users of Bayesian DSGE estimation methods, given recent evidence that DSGE models with very different policy implications may be observationally equivalent. This means that the posterior tends to move nearly one for one with the prior. In such cases, one would like a frequentist confidence set to reflect the fact that there is essentially no information about the structural parameter in the data. This is indeed what we found in several examples based on the recent literature. While the intervals tend to be appropriately wide, we also showed that it is not necessarily the case that the proposed intervals include all possible values. Moreover, the strength of the identification and hence the width of the confidence set may differ from one structural parameter to the next.

At the same time, the proposed confidence set takes full advantage of Bayesian estimation methods in that it is based on the inversion of the Bayes factor. Our method has two attractive features. One is that it is circumvents the problems of estimating DSGE models by classical maximum likelihood methods by utilizing the Bayesian estimation framework in constructing the Bayes factor. The other is that by construction the proposed confidence set is invariant to the choice of prior in the case of weak identification. Since the Bayes factor is the ratio of the posterior odds to the prior odds, if the likelihood is flat and hence the prior dominates the posterior, the numerator and the denominator of the ratio will tend to cancel, making the proposed confidence set robust to alternative priors. We illustrated this point in the context of the question of the relative importance of wage and price rigodities in a New Keynesian model.

Appendix

Proof of Proposition 1: Let $B_T = \{\theta \in \Theta : [\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)]^{1/2}(\theta - \hat{\theta}_T) \in A\}$. Note that $B_T \overset{P_{\theta_0}}{\to} \{\theta_0\}$ because A is compact, $\hat{\theta}_T$ is strongly consistent and $\nabla_{\theta\theta}\ell_T(\theta)$ is diverging. Define $I_{1,T}$ and $I_{2,T}$ by

$$\begin{aligned}
& \frac{|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}}{\pi(\theta_0)L_T(X_1,...,X_T|\hat{\theta}_T)}\int_{\Theta}\pi(\theta)L_T(X_1,...,X_T|\theta)d\theta \\
= & \frac{|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}}{\pi(\theta_0)L_T(X_1,...,X_T|\hat{\theta}_T)}\int_{\Theta\cap B_\delta(\theta_0)^c}\pi(\theta)L_T(X_1,...,X_T|\theta)d\theta \\
& +\frac{|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}}{\pi(\theta_0)L_T(X_1,...,X_T|\hat{\theta}_T)}\int_{\Theta\cap B_\delta(\theta_0)}\pi(\theta)L_T(X_1,...,X_T|\theta)d\theta \\
= & I_{1,T} + I_{2,T}. \qquad (22)
\end{aligned}$$

$$\begin{aligned}
I_{1,T} &= \frac{1}{\pi(\theta_0)}\exp(\ell_T(\theta_0) - \ell_T(\hat{\theta}_T))|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}} \\
&\quad \times \int_{\Theta\cap B_\delta(\theta_0)^c}\pi(\theta)\exp(\ell_T(\theta) - \ell_T(\theta_0))d\theta \\
&\leq \frac{1}{\pi(\theta_0)}|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}\exp(-\varepsilon(\delta)T) \\
&\to 0, \qquad (23)
\end{aligned}$$

where the first inequality follows from $\exp(\ell_T(\theta_0) - \ell_T(\hat{\theta}_T)) \leq 1$ and Assumption (f) and the last convergence follows from Assumptions (e) and (f).

$$\begin{aligned}
I_{2,T} &= |\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}\int_{B_\delta(\theta_0)}\frac{\pi(\theta)}{\pi(\theta_0)}\exp(\ell_T(\theta) - \ell_T(\hat{\theta}_T))d\theta \\
&= |\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}\int_{B_\delta(\theta_0)}\exp(\ell_T(\theta) - \ell_T(\hat{\theta}_T))d\theta + O(\delta) \\
&= |\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}\int_{B_\delta(\theta_0)}\exp\left[-\frac{1}{2}(\theta-\hat{\theta}_T)'\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)(\theta-\hat{\theta}_T)\right]\exp(R_T(\theta))d\theta \\
&\quad +O(\delta)\ \ P_{\theta_0}\text{-a.s.} \\
&= |\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}\int_{B_\delta(\theta_0)}\exp\left[-\frac{1}{2}(\theta-\hat{\theta}_T)'\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)(\theta-\hat{\theta}_T)\right]d\theta \\
&\quad +O(\delta) + o(1)\ \ P_{\theta_0}\text{-a.s.} \\
&\to (2\pi)^{\frac{1}{2}}, \qquad (24)
\end{aligned}$$

as $\delta \to 0$, where $R_T(\theta) = (\theta - \hat{\theta}_T)'\nabla_{\theta\theta}\ell_T(\bar{\theta}_T)(\theta - \hat{\theta}_T) - (\theta - \hat{\theta}_T)'\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)(\theta - \hat{\theta}_T)$, $\bar{\theta}_T$ is a point between $\hat{\theta}_T$ and θ_0, the second equality follows from Assumption (b) and

$\exp(\ell_T(\theta) - \ell_T(\hat{\theta}_T)) \leq 1$, the third follows from the Taylor's theorem, the fourth from Assumption (c) and the last from Assumption (f). It follows from (27), (28) and (29) that

$$\frac{|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}}{\pi(\theta_0)L_T(X_1,...,X_T|\hat{\theta}_T)}\int_{\Theta}\pi(\theta)L_T(X_1,...,X_T|\theta)d\theta \;\rightarrow\; (2\pi)^{\frac{1}{2}} \tag{25}$$

For sufficiently large T, $B_T(\hat{\theta}_T) \subset B_\delta(\theta_0)$. By repeating arguments we obtain

$$\frac{|\nabla_{\theta\theta}\ell_T(\hat{\theta}_T)|^{\frac{1}{2}}}{\pi(\theta_0)L_T(X_1,...,X_T|\hat{\theta}_T)}\int_{\Theta\cap B_T(\hat{\theta}_T)}\pi(\theta)L_T(X_1,...,X_T|\theta)d\theta \;\rightarrow\; (2\pi)^{\frac{1}{2}}P(z \in A) \tag{26}$$

where $z \sim N(0, I(\theta_0))$. The desired result follows from (30) and (31).

Proof of Theorem 1: It follows from assumption (c), the Taylor theorem and the first order condition for MLE that

$$\begin{aligned} I_3 &\equiv \int_{B_{\delta_T}(\theta_0)}\pi(\theta)\exp(\ell_T(g(\theta)) - \ell_T(\hat{\Pi}_T))d\theta \\ &= \int_{B_{\delta_T}(\theta_0)}\pi(\theta)\exp\left(\frac{1}{2}(g(\theta)-\hat{\Pi}_T)^\top\nabla_{\theta\theta}\ell_T(\bar{\Pi}_T)(g(\theta)-\hat{\Pi}_T)\right)d\theta, \end{aligned} \tag{27}$$

where $\bar{\Pi}_T$ is between Π_0 and $\hat{\Pi}_T$. It follows from assumptions (e) and (f) that

$$\begin{aligned} g(\theta) &= g(\theta_0) + Dg(\bar{\theta}(\theta))(\theta - \theta_0) \\ &= g(\theta_0) + o(T^{-1/2}), \end{aligned} \tag{28}$$

where $\bar{\theta}(\theta)$ is a point between θ and θ_0. It follows from assumptions (a) and (b), (27) and (28) that

$$\begin{aligned} I_3 &= \int_{B_{\delta_T}(\theta_0)}\pi(\theta)\exp\left(\frac{1}{2}(g(\theta_0)-\hat{\Pi}_T)^\top\nabla_{\Pi\Pi}\ell_T(\bar{\Pi}_T)(g(\theta_0)-\hat{\Pi}_T)\right)d\theta + o_p(1) \\ &= \int_{B_{\delta_T}(\theta_0)}\pi(\theta)d\theta\exp\left(\frac{1}{2}(g(\theta_0)-\hat{\Pi}_T)^\top\nabla_{\Pi\Pi}\ell_T(\bar{\Pi}_T)(g(\theta_0)-\hat{\Pi}_T)\right) + o_p(1) \end{aligned} \tag{29}$$

It follows from assumption (d) and (29) that

$$I_3 = \int_{B_{\delta_T}(\theta_0)}\pi(\theta)d\theta\exp\left(-\frac{1}{2}z'z\right) + o_p(1) \tag{30}$$

where z denote the $k \times 1$ standard normal random vector, i.e., $z \sim N(0_{k\times 1}, I_k)$.

Let

$$I_4 = \int_{\Theta\backslash B_{\delta_T}(\theta_0)}\pi(\theta)\exp(\ell_T(g(\theta)) - \ell_T(\hat{\Pi}_T))d\theta, \tag{31}$$

where $A\backslash B = \{x : x \in A \text{ and } x \notin B\}$. Since $\ell_T(g(\theta)) \leq \ell_T(\hat{\Pi}_T)$ by the definition of MLE, it follows from (31) that

$$I_4 \;\leq\; \int_{\Theta\backslash B_{\delta_T}} \pi(\theta)d\theta. \tag{32}$$

Because $\int_\Theta \pi(\theta)\exp(\ell_T(g(\theta)))d\theta$ cancels out, the Bayes factor in favor of H_1 can be written as

$$\begin{aligned} \text{Bayes Factor } (H_1|H_0) &= \frac{\int_{B_{\delta_T}(\theta_0)} \pi(\theta)d\theta}{\int_{\Theta\backslash B_{\delta_T}(\theta_0)} \pi(\theta)d\theta} \frac{I_4}{I_3} \\ &\leq \exp\left(\frac{1}{2}z'z\right) + o_p(1) \end{aligned} \tag{33}$$

where the inequality follows from (32) and the last equality from (30). Therefore it follows from (33) that

$$2\ln(\text{Bayes Factor } (H_1|H_0)) \;\leq\; z'z + o_p(1) \tag{34}$$

from which we obtain the desired result.

References

1. Altig, David, Lawrence Christiano, Martin Eichenbaum, and Jesper Linde (2005), "Firm-Specific Capital, Nominal Rigidities and the Business Cycles," NBER working paper 11034.

2. An, Sungbae, and Frank Schorfheide (2007), "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113-172.

3. Andrews, Donald W.K. (1994), "The Large Sample Correspondence Between Classical Hypothesis Tests and Bayesian Posterior Odds Tests," *Econometrica* 62, 1207–1232.

4. Balke, Nathan S., Stephen P.A. Brown, and Mine K. Yücel (2008), "Globalization, Oil Price Shocks and U.S. Economic Activity," unpublished manuscript, Federal Reserve Bank of Dallas.

5. Bickel, Peter J. and Kjell Doksum (2006), *Mathematical Statistics: Basic Concepts and Selected Ideas,* Vol. I, Second Edition, Prentice Hall: Upper Saddle River, NJ.

6. Canova, Fabio, and Luca Sala (2008), "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, forthcoming.

7. Chaudhuri, Saraswata, and Eric Zivot (2008), "A New Method of Projection-Based Inference in GMM with Weakly Identified Nuisance Parameters," unpublished manuscript, University of North Carolina, Chapel Hill, and University of Washington, Seattle.

8. Chernozhukov, Victor and Han Hong (1998), "An MCMC Approach to Classical Estimation," *Journal of Econometrics,* 115, 293–346.

9. Christiano, Lawrence, Martin Eichenbaum, and Charles Evans (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1-45.

10. Del Negro, Marco, and Frank Schorfheide (2008), "Forming Priors for DSGE models (and How It Affects the Assessment of Nominal Rigidities)," *Journal of Monetary Economics*, 55, 1191-1208.

11. Del Negro, Marco, Frank Schorfheide, Frank Smets and Rafael Wouters (2007), "On the Fit and Forecasting Performance of New Keynesian Models," *Journal of Business and Economic Statistics*, 25, 123–162.

12. Doob, J.L. (1948), "Application of the Theory of Martingales," *Colloques internationaux du Centre National de la Recherche Scientifique,* 13, 23– 27.

13. Jean-Marie Dufour and Mohamed Taamouti, (2005), "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica,* 73, 1351–1365.

14. Eichenbaum, Martin, and Jones Fisher (2007), "Estimating the Frequency of Price Re-optimization in Calvo-style Models," *Journal of Monetary Economics*, 54, 2023-2047.

15. Erceg, C., D. Henderson, and A. Levin (2000), "Optimal Monetary Policy with Staggered Wage and Price Contracts," *Journal of Monetary Economics,* 46, 281-313.

16. Fernández-Villaverde, Jesús, Pablo A. Guerron-Quintana, Juan F. Rubio-Ramírez (2009), "The New Macroeconometrics: A Bayesian Approach," *Handbook of Applied Bayesian Analysis*, forthcoming.

17. Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent and Mark W. Watson (2009), "A, B, C's (and D's) for Understanding VARs," *American Economic Review,* forthcoming.

18. Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin (2004), *Bayesian Data Analysis,* second edition, Chapman & Hall/CRC: Boca Raton, FL.

19. Hahn, Jinyong (1997), "Bayesian Bootstrap of the Quantile Regression Estimator: A Large Sample Study," *International Economic Review,* 38, 795–808.

20. Heyde, C.C. and I.M. Johnston (1979), "On Asymptotic Posterior Normality for Stochastic Processes," *Journal of Royal Statistical Society*, Series B, 41, 184–189.

21. Iskrev, Nikolay (2008), "How Much Do We Learn from the Estimation of DSGE models? A Case Study of Identification Issues in a New Keynesian Business Cycle Model," unpublished manuscript, University of Michigan.

22. Kadane, Joseph B. (1975), "The Role of Identification in Bayesian Theory," in: Stephen E. Fienberg and Arnold Zellner (eds.), *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage*, North-Holland: Amsterdam, 175-191.

23. Klenow, Peter J. and Oleksiy Kryvtsov, "State-Dependent or Time-Dependent Pricing: Does it Matter for Recent U.S. Inflation?" *Quarterly Journal of Economics*, 123, 863-904.

24. Kim, Jae-Young (1998), "Large Sample Properties of Posterior Densities, Bayesian Information Criterion and the Likelihood Principle in Nonstationary Time Series Models," *Econometrica,* 66, 359–380.

25. Le Cam, Lucien and Grace Lo Yang (2000), *Asymptotics in Statistics: Some Basic Concepts*, Second Edition, Springer-Verlag: New York, NY.

26. Moon, Hyungsik Roger, and Frank Schorfheide (2009), "Bayesian and Frequentist Inference in Partially Identified Models," unpublished manuscript, University of Southern California and University of Pennsylvania.

27. Nelson, Charles R. and Richard Startz (2007), "The Zero-Information-Limit Condition and Spurious Inference in Weakly Identified Models," *Journal of Econometrics,* 138, 47–62.

28. Ploberger, Werner and Peter C.B. Phillips (1996), "An Asymptotic Theory of Bayesian Inference For Time Series," *Econometrica,* 64, 381–412.

29. Poirier, Dale J. (1998), "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14, 483-509.

30. Smets, F., and R. Wouters (2003), "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123-1175.

31. Smets, Frank, and Rafael Wouters (2007a), "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586-606.

32. Smets, Frank, and Rafael Wouters (2007b), "On the Fit and Forecasting Performance of New Keynesian models," *Journal of Business and Economic Statistics*, 25, 123-162.

33. Stock, James H. and Mark Watson (2002), "Has the Business Cycle Changed and Why?" *NBER Macroeconomics Annual 2002*, 159-218.

34. Stock, James H. and Jonathan H. Wright (2000), "GMM with Weak Identification," *Econometrica,* 68, 1055–1096.

35. Wand, M.P. and M.C. Jones (1995), *Kernel Smoothing*, Chapman & Hall: London, UK.

36. Woodford, Michael M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy,* Princeton University Press: Princeton, NJ.

Table 1: Effective Coverage Rates of Traditional Intervals Based on the Posterior and the Bayes Factor Intervals: Small-Scale New Keynesian Model

$T = 96$									
	ϕ_π	ϕ_x	α	θ	ρ_z	ρ_τ	σ_z	σ_r	Joint
Mean$\pm 1.645SD$	0.996	0.971	0.923	1.000	0.793	0.858	0.593	0.525	
Median$\pm 1.645SD$	0.990	0.971	0.793	1.000	0.790	0.858	0.623	0.642	
Mode$\pm 1.645SD$	0.926	0.983	0.846	0.772	0.864	0.861	0.891	0.946	
Percentile	0.954	0.867	0.803	1.000	0.657	0.741	0.364	0.278	
BF Interval	1.000	0.994	1.000	1.000	0.970	0.972	0.991	0.972	0.894

$T = 188$									
	ϕ_π	ϕ_x	α	θ	ρ_z	ρ_τ	σ_z	σ_r	Joint
Mean$\pm 1.645SD$	0.995	0.985	0.977	0.998	0.832	0.900	0.715	0.733	
Median$\pm 1.645SD$	0.995	0.987	0.943	0.998	0.832	0.898	0.748	0.778	
Mode$\pm 1.645SD$	0.973	0.977	0.880	0.782	0.880	0.893	0.925	0.962	
Percentile	0.977	0.957	0.948	0.998	0.695	0.807	0.505	0.433	
BF Interval	0.998	0.993	0.998	0.998	0.992	0.993	0.985	0.977	0.932

Notes: The numbers are coverage probabilities of the level 90% confidence intervals and set.

Table 2: Traditional Interval Estimates Based on the Posterior and the Bayes Factor Interval Estimates: Medium-Scale New Keynesian Model

Agnostic Priors

	Rigidity Parameters					
	Posterior				Credible	BF confidence
	means	medians	modes	SD	intervals	intervals
ζ_p	0.694	0.695	0.669	0.046	[0.615,0.769]	[0.479, 0.840]
ζ_w	0.219	0.214	0.211	0.071	[0.112,0.341]	[0.036, 0.563]
	Other Endogenous Propagation Parameters					
	Posterior				Credible	BF confidence
	means	medians	modes	SD	intervals	intervals
ν_l	1.767	1.699	1.757	0.530	[1.026,2.749]	[0.458, 4.661]
ψ_1	2.382	2.376	2.412	0.253	[1.983,2.806]	[1.278, 3.390]
ψ_2	0.074	0.073	0.042	0.024	[0.038,0.116]	[0.007, 0.180]
ρ_r	0.722	0.724	0.708	0.038	[0.657,0.781]	[0.547, 0.840]
$\imath_p$	0.096	0.074	0.004	0.083	[0.007,0.260]	[0.000, 0.562]
$\imath_w$	0.279	0.269	0.451	0.121	[0.102,0.489]	[0.000, 0.947]
s''	9.146	8.986	6.966	1.973	[6.177,12.639]	[3.036, 17.459]
h	0.753	0.756	0.769	0.056	[0.657,0.840]	[0.512, 0.905]
a''	0.230	0.214	0.258	0.101	[0.096,0.418]	[0.035, 0.888]
	Exogenous Propagation Parameters					
	Posterior				Credible	BF confidence
	means	medians	modes	SD	intervals	intervals
ρ_z	0.232	0.222	0.432	0.117	[0.052,0.445]	[0.001, 0.676]
ρ_ϕ	0.958	0.960	0.981	0.018	[0.926,0.984]	[0.867, 0.999]
ρ_{λ_f}	0.942	0.949	0.939	0.033	[0.875,0.982]	[0.771, 0.998]
ρ_g	0.916	0.916	0.912	0.027	[0.870,0.959]	[0.781, 0.992]

Table 3: Traditional Interval Estimates Based on the Posterior and the Bayes Factor Interval Estimates: Medium-Scale New Keynesian Model

Low-Rigidities Priors

	Posterior means	Posterior medians	Posterior modes	SD	Credible intervals	BF confidence intervals
Rigidity Parameters						
ζ_p	0.658	0.660	0.731	0.045	[0.580,0.730]	[0.466, 0.793]
ζ_w	0.266	0.263	0.417	0.057	[0.177,0.364]	[0.089, 0.551]
Other Endogenous Propagation Parameters						
ν_l	1.856	1.776	1.329	0.597	[1.018,2.961]	[0.426, 5.016]
ψ_1	2.417	2.412	2.262	0.249	[2.014,2.838]	[1.611, 3.692]
ψ_2	0.074	0.072	0.072	0.023	[0.038,0.115]	[0.007, 0.178]
ρ_r	0.724	0.726	0.772	0.037	[0.659,0.780]	[0.556, 0.836]
$\imath_p$	0.154	0.128	0.167	0.116	[0.014,0.381]	[0.000, 0.684]
$\imath_w$	0.270	0.261	0.139	0.112	[0.101,0.467]	[0.001, 0.770]
s''	8.955	8.788	10.805	2.007	[5.966,12.497]	[3.799, 20.434]
h	0.740	0.742	0.772	0.055	[0.648,0.826]	[0.535, 0.914]
a''	0.242	0.226	0.143	0.103	[0.103,0.430]	[0.035, 0.858]
Exogenous Propagation Parameters						
ρ_z	0.217	0.207	0.031	0.115	[0.042,0.421]	[0.000, 0.704]
ρ_ϕ	0.956	0.957	0.947	0.018	[0.924,0.982]	[0.853, 0.998]
ρ_{λ_f}	0.950	0.955	0.979	0.027	[0.897,0.984]	[0.782, 1.000]
ρ_g	0.911	0.912	0.909	0.027	[0.866,0.955]	[0.791, 0.996]

Table 4: Traditional Interval Estimates Based on the Posterior and the Bayes Factor Interval Estimates: Medium-Scale New Keynesian Model

High-rigidity Priors

	Rigidity Parameters					
		Posterior			Credible	BF confidence
	means	medians	modes	SD	intervals	intervals
ζ_p	0.773	0.774	0.828	0.056	[0.678,0.855]	[0.580, 0.905]
ζ_w	0.446	0.427	0.749	0.115	[0.288,0.670]	[0.155, 0.848]
	Other Endogenous Propagation Parameters					
		Posterior			Credible	BF confidence
	means	medians	modes	SD	intervals	intervals
ν_l	1.520	1.466	0.247	0.667	[0.493,2.709]	[0.133, 4.794]
ψ_1	2.322	2.316	2.382	0.241	[1.944,2.733]	[1.521, 3.374]
ψ_2	0.057	0.055	0.064	0.021	[0.025,0.095]	[0.007, 0.171]
ρ_r	0.744	0.746	0.794	0.036	[0.682,0.800]	[0.577, 0.857]
$\imath_p$	0.082	0.063	0.173	0.072	[0.006,0.224]	[0.000, 0.492]
$\imath_w$	0.188	0.181	0.147	0.091	[0.047,0.347]	[0.001, 0.611]
s''	9.996	9.860	10.167	2.073	[6.857,13.592]	[3.983, 20.630]
h	0.795	0.801	0.850	0.052	[0.701,0.870]	[0.562, 0.935]
a''	0.217	0.202	0.201	0.103	[0.082,0.407]	[0.018, 0.747]
	Exogenous Propagation Parameters					
		Posterior			Credible	BF confidence
	means	medians	modes	SD	intervals	intervals
ρ_z	0.235	0.223	0.414	0.125	[0.052,0.463]	[0.000, 0.730]
ρ_ϕ	0.900	0.918	0.626	0.066	[0.749,0.966]	[0.607, 0.996]
ρ_{λ_f}	0.852	0.888	0.626	0.106	[0.647,0.970]	[0.492, 0.999]
ρ_g	0.927	0.926	0.997	0.029	[0.880,0.978]	[0.822, 0.999]