Quantitative Economics 4 (2013), 197–229 1759-7331/20130197

Frequentist inference in weakly identified dynamic stochastic general equilibrium models

Pablo Guerron-Quintana
Federal Reserve Bank of Philadelphia

Atsushi Inoue
North Carolina State University

Lutz Kilian
University of Michigan and CEPR

A common problem in estimating dynamic stochastic general equilibrium models is that the structural parameters of economic interest are only weakly identified. As a result, classical confidence sets and Bayesian credible sets will not coincide even asymptotically, and the mean, mode, or median of the posterior distribution of the structural parameters can no longer be viewed as a consistent estimator. We propose two methods of constructing confidence intervals for structural model parameters that are asymptotically valid from a frequentist point of view regardless of the strength of identification. One involves inverting a likelihood ratio test statistic, whereas the other involves inverting a Bayes factor statistic. A simulation study shows that both methods have more accurate coverage than alternative methods of inference. An empirical study of the degree of wage and price rigidities in the U.S. economy illustrates that the data may contain useful information about structural model parameters even when these parameters are only weakly identified.

Keywords. DSGE models, identification, inference, confidence sets, Bayes factor, likelihood ratio.

JEL classification. C32, C52, E30, E50.

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

Pablo Guerron-Quintana: pguerron@gmail.com
Atsushi Inoue: atsushi@ncsu.edu
Lutz Kilian: lkilian@umich.edu

We thank Marco del Negro and Frank Schorfheide for providing access to their data. We thank four anonymous referees, Fabio Canova, Yanqin Fan, Ulrich Müller, Frank Schorfheide, Jim Stock, and numerous seminar and conference participants for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Philadelphia or of the Federal Reserve System.

DOI: 10.3982/QE306

of the Bayesian approach compared to the frequentist approach is that the use of prior information allows the researcher to estimate structural models that otherwise would be computationally intractable or would produce economically implausible estimates. 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.

There is growing evidence, however, that many DSGE models used in empirical macroeconomics are only weakly identified (see, e.g., Canova and Sala (2009)). Weak identification manifests itself in a likelihood that is nearly flat across the parameter space. As a result, the posterior of the structural model parameters becomes highly dependent on the priors used by the researcher.[1] The weak identification of the structural parameters of DSGE models causes the usual asymptotic equivalence between Bayesian and frequentist estimation and inference to break down.[2] Given that the effect of the prior on the posterior does not die out asymptotically, it can be shown that the posterior mode will not be a consistent estimator of the true parameter vector, and that Bayesian credible sets and frequentist confidence sets based on Gaussian approximations will not coincide even asymptotically, removing the rationale for constructing confidence intervals from the quantiles of the posterior distribution or by adding multiples of posterior standard deviations to the posterior mean.[3]

In this paper, we develop alternative methods of constructing confidence sets for the structural parameters of DSGE models that remain valid asymptotically from a frequentist point of view, regardless of the strength of the identification. Lack of identification manifests itself in a likelihood function that lacks curvature. Our approach allows the structural parameters of the model to be weakly identified in the sense that the (scaled) slope of the log-likelihood function is local to zero. As in the weak-instruments literature, we think of the local-to-zero model as a device that reflects our inability to determine the strength of the identification from the data (see, e.g., Canova and Sala (2009), Iskrev (2010), Koop, Pesaran, and Smith (2011)). The proposed confidence set is obtained by inverting either a suitably defined likelihood ratio (LR) statistic or a Bayes factor (BF) statistic. The implied LR confidence set can be constructed using classical estimation methods and has $1-\alpha$ coverage probability asymptotically, regardless of the strength of identification, whereas the implied BF confidence set is conservative in that a $1-\alpha$

[1] For example, Smets and Wouters (2007, p. 594) noted that for their main behavioral model parameters, "the mean of the posterior distribution is typically relatively close to the mean of the prior." A similar finding is reported in Del Negro and Schorfheide (2008), who documented that DSGE models that have very different policy implications may fit the data equally well. In particular, a New Keynesian model with low price rigidities and low wage rigidities is observationally equivalent to a model with high wage rigidities and high price rigidities.

[2] See Le Cam and Yang (2000, Chapter 8) and the references therein for the large-sample correspondence between Bayesian and frequentist approaches. For more recent results in the econometrics literature, see Andrews (1994) and Chernozhukov and Hong (2003), for example, as well as Kim (1998) and Phillips and Ploberger (1996).

[3] Our results are reminiscent of Moon and Schorfheide's (2012) recent finding that Bayesian credible sets and conventional frequentist confidence sets differ asymptotically in set-identified (as opposed to weakly identified) models. The key difference is that our analysis does not deal with set identification.

confidence set has at least $1 - \alpha$ coverage probability asymptotically. Both methods allow the construction of confidence intervals for individual structural parameters by the projection method.

The BF confidence set is asymptotically invariant to the prior in the case of weak identification. One advantage of the BF interval is that it may be computed from the output produced by existing Bayesian code for DSGE model estimation. No additional numerical estimation is required. This feature facilitates its adoption by applied users of Bayesian methods of estimating DSGE models. In contrast, construction of the LR confidence set dispenses with Bayesian estimation methods altogether. It requires instead the explicit derivation of the state-space representation of the unrestricted reduced form of the DSGE model on a case-by-case basis and the numerical estimation of its parameters. Our theoretical analysis shows that the asymptotic distribution of the BF test statistic is first-order stochastically dominated by that of the LR test statistic. Moreover, the BF test statistic does not have power against local alternatives when the model contains some strongly identified structural parameters, whereas the LR test has local power even in that case.

Related work includes Komunjer and Ng (2009), who established conditions for identifying structural parameters in DSGE models from autocovariance structures, and Fukač, Waggoner, and Zha (2007), who contrasted local and global identification. While these procedures are helpful in assessing the identifiability of structural model parameters, they are not informative about the strength of identification, suggesting that there remains a need for approaches such as ours that are robust to weak identification. Our approach builds on a large literature that exploits the observation that one can invert an asymptotically pivotal statistic to construct confidence intervals that remain asymptotically valid under weak identification.[4] The key difference is that we consider a different class of models and estimators, and that we invert a different type of statistic than previous studies. Our approach relies on inverting an LR test statistic or a BF statistic. In subsequent work, Qu (2011) developed an alternative frequency-domain approach to constructing confidence sets for structural parameters of DSGE models, and Andrews and Mikusheva (2012) developed an identification-robust LM test. The simulation evidence in the latter paper for a small-scale DSGE model suggests that our methods have power comparable to the alternative methods proposed in Qu (2011) and in Andrews and Mikusheva (2012). Finally, in closely related work, Dufour, Khalaf, and Kichian (2009) proposed a multivariate extension of the Anderson–Rubin (AR) test statistic for multiequation macroeconomic models with weakly identified structural parameters. The key difference from our paper is that their approach is limited to DSGE models with a finite-order vector autoregressive (VAR) representation in the observables (which is the exception rather than the rule), whereas our approach is more widely applicable. Moreover, Andrews and Mikusheva's simulation study shows that the Dufour, Khalaf, and Kichian (2009) AR-type test statistic has lower power for joint inference than alternative test statistics including the LR test.

[4]Examples include Stock and Wright (2000), Kleibergen and Mavroeidis (2009), and Andrews and Cheng (2012).

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 illustrate the failure of the conventional frequentist interpretation of Bayesian posterior estimates. We propose the LR and BF confidence sets, establish their asymptotic validity from a frequentist point of view, and compare the asymptotic power of the LR and BF tests. Section 3 focuses on implementation issues. We discuss the derivation of the degrees-of-freedom parameter in the asymptotic distribution of the proposed test statistics and we show how the projection method can be used to construct confidence intervals for individual structural parameters from the joint confidence set.

In Section 4, we investigate the finite-sample accuracy of the proposed procedures in the context of two small-scale New Keynesian macroeconomic models. We show by simulation that the joint LR confidence set has between 86% and 91% coverage for realistic sample sizes, whereas the joint BF confidence set for the same sample sizes has coverage rates ranging from 79% to 99%, depending on the model and the choice of prior, highlighting that the accuracy of the BF method can be sensitive to the choice of the prior in small samples. We also demonstrate that the practice of constructing confidence intervals from the posterior distribution of individual structural parameters by adding $+/-1.645$ posterior standard deviations to the posterior mode (or mean) or by computing posterior percentiles may result in intervals with serious coverage deficiencies. In some cases, the coverage rate of nominal 90% intervals drops as low as 8%. In contrast, the LR interval has coverage rates of at least 94% for commonly used sample sizes, while the BF interval has corresponding coverage rates of at least 97%.

An empirical illustration in Section 5 focuses on the question of the importance of wage and price rigidities in the U.S. economy. This empirical example involves a medium-scale DSGE model widely used in the literature (see, e.g., Del Negro and Schorfheide (2008)). We provide evidence that wage rigidities are much smaller than price rigidities. We also show that estimates of the aggregate degree of price rigidity are consistent with microeconomic evidence on price rigidities when using the LR interval, but not when using pseudo-Bayesian interval estimates of the type that a frequentist user may construct from the posterior distribution produced by Bayesian estimation procedures for DSGE models. This example demonstrates that the LR test approach may have enough power to be practically useful. The BF interval estimates, in contrast, are sensitive to the choice of the prior, consistent with the simulation evidence. The concluding remarks are in the Appendix.

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 in the limit and asymptotically normal after suitable scaling. The latter result is called the Bernstein–von Mises theorem in the Bayesian literature (see, e.g., Bickel and Doksum (2006), Le Cam and Yang (2000)). The Bernstein–von Mises theorem allows a classical interpretation of Bayesian confidence sets. In other words,

Bayesian credible sets can be viewed as asymptotically valid classical confidence sets for the parameter of interest. This fact is important because it allows econometricians who are not Bayesians to use the Bayesian apparatus to estimate DSGE models, taking advantage of its superior convergence properties, while interpreting the results in a classical fashion. Under weak identification, however, the Bernstein–von Mises theorem does not apply. The following proposition illustrates that the classical interpretation of Bayesian credible sets breaks down when the model is not strongly identified.

PROPOSITION 1 (Posterior Distributions in Normal Linear Regression Models Under Weak Identification). *Suppose that $y_t = \gamma x_t + u_t$, where $\{x_t\}$ is an independent and identically distributed (i.i.d.) scalar random variable with $E(x_t^2) < \infty$ and is independent of the scalar random variable $\{u_t\}$, $u_t \overset{i.i.d.}{\sim} N(0,1)$, and $T^{-1/2}\sum_{t=1}^{T} x_t u_t \xrightarrow{d} N(0, E(x_t^2))$. In addition, suppose that θ is the parameter of interest such that $\gamma = \theta T^{-1/2}$. Consider the prior distribution $\pi(\theta) \propto \exp(-(\theta - \bar{\theta})^2/2\bar{\sigma}^2)$, where $\bar{\theta}$ is the prior mean and $\bar{\sigma}^2$ is the prior variance. Then the posterior distribution of θ is given by*

$$p\big(\theta|\{x_t\}_{t=1}^T\big) = \sqrt{T^{-1}\sum_{t=1}^{T} x_t^2 + 1/\bar{\sigma}^2}\,\phi\left(\frac{\theta - \dfrac{T^{-1/2}\displaystyle\sum_{t=1}^{T} x_t y_t + \bar{\theta}/\bar{\sigma}^2}{T^{-1}\displaystyle\sum_{t=1}^{T} x_t^2 + 1/\bar{\sigma}^2}}{\dfrac{1}{\sqrt{T^{-1}\displaystyle\sum_{t=1}^{T} x_t^2 + 1/\bar{\sigma}^2}}}\right),$$

where $\phi(\cdot)$ denotes the probability density function (p.d.f.) of a standard normal random variable. Taking the limit as $T \to \infty$, the right-hand side converges in probability to

$$\sqrt{E\big(x_t^2\big) + 1/\bar{\sigma}^2}\,\phi\left(\frac{\theta - \dfrac{\sqrt{E(x_t^2)}z + \bar{\theta}/\bar{\sigma}^2}{E(x_t^2) + 1/\bar{\sigma}^2}}{\sqrt{\dfrac{1}{E(x_t^2) + 1/\bar{\sigma}^2}}}\right),$$

where z satisfies $T^{-1/2}\sum_{t=1}^{T} x_t y_t / \sqrt{E(x_t^2)} \xrightarrow{d} z$.

Proposition 1 illustrates that (i) the posterior distribution is not degenerate in the limit about the true parameter value when the parameter is weakly identified; (ii) that, unconditionally, the asymptotic limit of the posterior distribution is a mixture of normals; (iii) that the posterior mean is unconditionally random even in the asymptotic limit; and (iv) 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 as the sample size grows, if parameters are weakly identified. Hence, the posterior mode no longer coincides with the mean or median. This also means that the mean (or median or mode) of the posterior distribution is not a consistent estimator of the true parameter value.

Although this section focuses on Bayesian estimation methods, it is worth stressing that analogous problems arise if one uses frequentist maximum likelihood methods of estimating the parameter of interest. It can be shown that under these conditions, the unrestricted maximum likelihood estimator (MLE) of θ is inconsistent and has a nonstandard limiting distribution. Intuitively, this problem arises because there is insufficient curvature in the likelihood function.

It is useful to contrast our notion of weak identification in Proposition 1 to the limiting cases of strong identification and no identification. Suppose that $\gamma = g(\theta)$. Strong identification requires not only that the true parameter value θ_0 uniquely maximizes the population objective function

$$-\frac{1}{2}\ln(2\pi) - \frac{1}{2}E\big(x_t^2\big)\big(g(\theta) - g(\theta_0)\big)^2 - \frac{1}{2}, \tag{1}$$

but that the curvature of the likelihood is strong enough for the rank condition for identification,

$$\operatorname{rank}\big(E\big(x_t^2\big)\big[g'(\theta_0)\big]^2\big) = \dim(\theta_0) = 1, \tag{2}$$

to be satisfied. In contrast, lack of identification, as discussed in Kadane (1975) and Poirier (1998), would correspond to $g(\theta) = \gamma_0$ for all $\theta \in \Theta$ such that the likelihood function does not depend on θ even in finite samples. In that case, the likelihood is perfectly flat in θ.

Of particular interest in empirical work is the intermediate case in which the likelihood is nearly flat. A natural way to allow for this situation in the example above is to model the slope of the log-likelihood function with respect to θ as local to zero,

$$-\frac{1}{2}\ln(2\pi) - \frac{1}{2T}E\big(x_t^2\big)(\theta - \theta_0)^2 - \frac{1}{2}, \tag{3}$$

where we imposed the assumption in Proposition 1 that $\gamma = \theta T^{-1/2}$ on the population objective function.

This approach is designed to represent our inability in finite samples to determine with a reasonable degree of accuracy which of the two limiting cases is a better approximation of reality. By analogy to the problems of weak instruments and weak identification in the generalized method of moments (GMM) literature, modelling the slope of the log likelihood as local to zero may be viewed as a statistical device for obtaining a more accurate asymptotic approximation to the distribution of θ.

2.2 *Likelihood ratio tests*

In this paper, we propose two frequentist confidence sets for parameters in DSGE models that are asymptotically valid regardless of the strength of identification. One is based

on the likelihood ratio test statistic and the other is based on the Bayes factor. Our starting point is the reduced-form representation of the DSGE model. Let γ denote a vector of reduced-form parameters of the state-space model

$$x_{t+1} = Ax_t + Bw_t, \tag{4}$$

$$y_t = Cx_t + Dw_t, \tag{5}$$

where x_t is a vector of possibly unobserved state variables, y_t is a vector of observed variables, and $w_t \overset{\text{i.i.d.}}{\sim} N(0, I)$. Whereas C is a matrix of known constants, the reduced-form parameters A, B, and D are typically functions of structural parameters θ (see Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007)).[5] Denote such relationships between the reduced-form and the structural parameters by $f_T(\gamma, \theta) = 0_{r\times 1}$. The dimension of f_T, r, can be smaller than the dimension of γ if some reduced-form parameters are not tied to the structural parameter vector θ.

Consider the likelihood ratio statistic for testing

$$H_0 : f_T(\gamma, \theta_0) = 0_{r\times 1} \quad \text{for some } \theta_0 \in \Theta$$

and denote it by $\mathrm{LR}_T(\theta_0) = 2(\ell_T(\hat{\gamma}_T) - \ell_T(\tilde{\gamma}(\theta_0)))$, where $\ell_T(\gamma)$ is the log-likelihood function for γ, $\hat{\gamma}_T$ is the unconstrained MLE of γ, and $\tilde{\gamma}_T$ is the constrained MLE of γ subject to the restriction $f_T(\gamma, \theta_0) = 0$. We assume that the reduced-form parameter γ is strongly identified, while allowing some or all of the elements of the structural parameter vector θ not to be strongly identified. Weak identification of all structural parameters in this model would imply that the constraint function is local to zero for all γ and θ, that is, $f_T(\gamma, \theta_0) = T^{-1} f_2(\gamma, \theta)$. In other words, even when the reduced-form parameters are strongly identified, the structural parameters θ cannot be inferred from the restrictions. Our analysis is similar to Stock and Wright (2000) and Antoine and Renault (2009) in that we do not require prior knowledge of which structural parameters in θ are strongly identified or weakly identified. In fact, our results hold even when all structural parameters are weakly identified or all structural parameters are strongly identified.

On the other hand, the assumption that the reduced-form parameter γ in the minimal state-space representation is strongly identified is essential to our approach. In many DSGE models some elements of the reduced-form parameter vector γ are unidentified. If the likelihood function is a function of γ_1/γ_2 only, for example, γ_1 and γ_2 are not jointly identified, only the ratio γ_1/γ_2 is identified. Analogously to Komunjer and Ng's (2009) analysis of structural DSGE model parameters, we assume that in such situations the state-space model has been reparameterized such that γ is strongly identified. Moreover, our assumption that the reduced-form parameters are strongly identified rules out situations in which some reduced-form parameters are unidentified because of a near root cancellation in the vector autoregressive moving average (VARMA) representation of the DSGE model (see, e.g., Schorfheide (2011) and Andrews and Cheng (2012)). Since our paper was written, alternative methods of inference for DSGE models that do not

[5] By redefining the variables and coefficient matrices, (4) and (5) can be written in the form used in Fernández-Villaverde et al. (2007): $x^*_{t+1} = A^* x^*_t + B w^*_{t+1}$ and $y_{t+1} = C^* x^*_t + D^* w^*_{t+1}$.

require the assumption of strongly identified reduced-form parameters have been developed by Qu (2011) and Andrews and Mikusheva (2012). How practically important this assumption is remains an open question.

PROPOSITION 2 (Asymptotic Distribution of the LR Test Statistic). *Suppose the following stipulations.*

(a) *The log-likelihood function* $\ell_T(\gamma)$ *is correctly specified and twice continuously differentiable in* γ.

(b) *The unconstrained and constrained MLE are consistent, that is,* $\hat{\gamma}_T - \gamma_{0,T} = o_p(1)$ *and* $\tilde{\gamma}_T(\theta_0) - \gamma_{0,T} = o_p(1)$, *and* $T^{-1/2}\nabla_\gamma \ell_T(\gamma_{0,T}) \xrightarrow{d} N(0_{\dim(\gamma)\times 1}, V_\gamma)$, *where* $\gamma_{0,T}$ *satisfies* $f_T(\gamma_{0,T}, \theta_0) = 0_{r\times 1}$ *and* $V_\gamma = -\operatorname{plim}_{T\to\infty}[(1/T)\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})]^{-1}$ *is positive definite.*

(c) *The function* $f_T(\gamma, \theta_0)$ *is continuously differentiable in* γ *and* $\operatorname{rank}(D_\gamma f_T(\gamma_{0,T}, \theta_0)) = \dim(f_T) = r$.

Then

$$\mathrm{LR}_T(\theta_0) \xrightarrow{d} \chi^2_r. \tag{6}$$

REMARKS.

1. It should be noted that there is a close link between our LR approach and the Anderson–Rubin approach in the weak-instrument literature. Suppose that the reduced form is given by

$$y_i = \gamma_1' z_i + v_{1i},$$
$$x_i = \gamma_2' z_i + v_{2i},$$

whereas the structural equation of interest is

$$y_i = \theta x_i + u_i.$$

Then the structural parameter θ and the reduced-form parameters γ are subject to the restriction $f(\gamma, \theta) = \gamma_1 - \theta\gamma_2$. The Wald test based on $\sqrt{T} f(\hat{\gamma}, \theta_0)$ is the Anderson–Rubin statistic with the degrees of freedom given by the number of instruments.

2. Proposition 2 shows that the LR test statistic can be used to construct confidence intervals with $1 - \vartheta$ coverage probability asymptotically, regardless of the strength of identification. In practice, we proceed in four steps.

Step 1. Estimate the reduced-form parameters γ in the state-space model (6) and (7) by Gaussian MLE using the Kalman filter.

Step 2. Define a set of points in the space of structural parameters, θ. This may be accomplished, for example, by defining a grid of points in the parameter space or by drawing at random from a suitable distribution such as a truncated uniform distribution, the prior distribution, or the posterior distribution.

Step 3. For each of these points, compute $\mathrm{LR}_T(\theta) = 2(\ell_T(\hat{\gamma}_T) - \ell_T(\tilde{\gamma}_T(\theta)))$ and check the inequality

$$\left\{\theta \in \Theta : \mathrm{LR}_T(\theta) \leq \chi^2_{r,1-\vartheta}\right\}.$$

Step 4. The set of the points that satisfy this inequality is the level $1-\vartheta$ confidence set.

3. It should be noted that we do not establish that

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} \Pr[\theta \in \mathrm{CS}_T] \to 1 - \vartheta$$

as in Mikusheva (2007) or Andrews and Cheng (2012), where CS_T stands for confidence set. Rather, the proposed confidence set for θ will be valid only over the subset of Θ for which there are no weak identification issues in γ. Establishing the uniform validity of these confidence intervals rather than their pointwise validity is difficult, loosely speaking, because $\sqrt{T}(\hat{\gamma} - g(\theta))$, where $\gamma \equiv g(\theta)$, may not be asymptotically normally distributed for all possible values of θ. It can be shown that our inference problem is not a special case of the framework studied in Andrews and Cheng (2012). Their proof of uniform asymptotic validity relies on assumptions that do not apply in our context.

4. Proposition 2 does not take a stand on the strength of the identification of θ. If one knew which structural parameters are strongly identified, one could partition the parameter vector θ into the vector β of strongly identified parameters and the vector α of weakly identified parameters. Weak identification here refers to the likelihood being nearly flat with respect to α, which may be modelled as $f_T(\gamma, \theta) = f_1(\gamma, \beta) + T^{-1/2} f_2(\gamma, \theta)$. This means that the rank condition for identification is nearly violated. This additional information could be used to obtain a tighter confidence set for the weakly identified parameters based on the likelihood obtained by concentrating out β, given a hypothesized value of α, as shown in Proposition 3a below. The power of the test is increased in this case because the degrees of freedom only depend on $\dim(\alpha)$ rather than r.

This approach differs from the common practice among DSGE users of imposing consistent estimates of structural parameters that can be recovered from long-run averages of the data (such as the aggregate depreciation rate or the share of labor income) when estimating the remaining model parameters. The difference is that this conventional procedure ignores the estimation uncertainty of the first-stage estimate in deriving the asymptotic distribution, whereas the proposition below incorporates that uncertainty.

Proposition 3a (Asymptotic Distribution of the LR Test Statistic When Some Structural Parameters Are Known to Be Strongly Identified). *Let $\tilde{\gamma}_T(\alpha_0)$ and $\tilde{\beta}_T(\alpha_0)$ denote the constrained MLE of γ and β, respectively, given α_0. Suppose that assumptions* (a), (b), *and* (c) *of Proposition 2 hold with $\tilde{\gamma}_T(\theta_0)$ replaced by $\tilde{\gamma}_T(\alpha_0)$. Further suppose that $\tilde{\beta}_T(\alpha_0)$ is consistent for β_0 and that $D_\beta f(\gamma_{0,T}, \theta_0)$ has rank k_2. Then*

$$\mathrm{LR}_T(\theta_0) \xrightarrow{d} \chi^2_{r-k_2}. \tag{7}$$

Although Propositions 2 and 3a are not surprising from a technical point of view, they provide a powerful tool for dealing with problems of weak identification of structural parameters in DSGE models. Our analysis shows that inference on these parameters may be conducted without ever estimating the structural model. Only estimates of the reduced form are required. As a result, we can dispense with Bayesian methods of estimating the structural parameters altogether. The construction of the LR confidence set requires instead the explicit derivation of the state-space representation of the unrestricted reduced form of the DSGE model on a case-by-case basis and numerical maximum likelihood estimation of its parameters. Next, we consider an alternative approach based on the inversion of the Bayes factor. Although the Bayes factor statistic can be constructed from the output of commonly used Bayesian estimation routines, we evaluate this statistic from a frequentist point of view.

2.3 *Bayes factors*

Consider testing $H_0 : \theta \in B_{\delta_T}(\theta_0)$ against $H_1 : \theta \notin B_{\delta_T}(\theta_0)$, where $B_{\delta_T}(\theta_0) = \{\theta \in \Theta : |\theta - \theta_0| \leq \delta_{T,j} \text{ for } j = 1, 2, \ldots, p\}$, $\Theta \subset \Re^p$, and $\delta_T = [\delta_{T,1}, \ldots, \delta_{T,k}]' \to 0_{k\times 1}$ as $T \to \infty$. We define the Bayes factor 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 evaluation of the posterior probabilities requires specification of the likelihood function. For consistency with the analysis of the LR statistic, we impose the same restrictions on γ under the null hypothesis as before. If $f_T(\gamma, \theta) = \gamma - g(\theta) = 0$ as is typical for DSGE models, the log-likelihood function for θ can be written as $\ell_T(g(\theta))$. Otherwise, the analysis must be based on the concentrated log-likelihood function $\ell_T(\tilde{\gamma}_T(\theta))$.

Theorem 1 states the asymptotic distribution of the BF statistic, again under the premise that all elements of γ are fully identified, but that it is not known which elements of θ are strongly identified or weakly identified.

THEOREM 1 (Asymptotic Distribution of the Bayes Factor). *In addition to assumptions* (a)–(c) *in Proposition 2, suppose the following assumptions*:

(a) *The prior density* $\pi : \Theta \to \Re_+$ *is continuous on* $\Theta \subset \Re^k$.

(b) $\tilde{\gamma}_T(\theta)$ *is continuous in a neighborhood of* θ_0.

(c) $\delta_T = o(1)$.

Then

$$\lim_{T\to\infty} \Pr\big(\text{Bayes factor}(\theta_0) \leq e^{z_T' P_T z_T/2}\big) = 1, \tag{9}$$

where

$$P_T = \big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1/2} R_T' \big\{R_T\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1} R_T'\big\}^{-1} R_T\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1/2},$$

$$z_T = \big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1/2} \nabla_\gamma \ell_T(\gamma_{0,T}),$$

and $R_T = D_\gamma f_T(\gamma_{0,T}, \theta_0)$. *In other words, the distribution of* $2\ln[\text{Bayes factor}(\theta_0)]$ *is asymptotically first-order stochastically dominated by* χ^2_r.[6]

REMARKS.

1. To build intuition for Theorem 1, consider the normal linear regression model example of Section 2.1. In that case, $P_T = 1$ and $z_T = -(\sum_{t=1}^T x_t^2)^{-1/2} \sum_{t=1}^T x_t u_t$. The latter converges in distribution to a standard normal random variable. Thus $2\ln[\text{Bayes factor}(\theta_0)]$ is asymptotically bounded by χ^2_1.

2. Theorem 1 requires only the existence of an asymptotically normally distributed MLE of a transformation of the reduced-form parameters. We do not need to compute the MLE of γ to obtain the Bayes factor.

3. If $\dim(f) = \dim(\gamma)$ and f_T is continuously differentiable in γ and θ with $\operatorname{rank}(D_\gamma f_T(\gamma_{0,T}, \theta_0)) = \dim(\gamma)$, then assumption (b) can be shown to be satisfied using the implicit function theorem.

4. Theorem 1 implies that by inverting the Bayes factor, one can obtain a level $(1 - \vartheta)$ confidence set

$$\Theta_0 = \left\{\theta \in \Theta : \text{Bayes factor}(\theta) \leq e^{\chi^2_{r,1-\vartheta}/2}\right\}. \tag{10}$$

This set satisfies

$$\lim_{T\to\infty} \Pr(\theta_0 \in \Theta_0) \geq 1 - \vartheta, \tag{11}$$

where $1 - \vartheta$ is the coverage probability.

5. The fact that we focus on the Bayes factor in favor of the alternative hypothesis (as opposed to the Bayes factor in favor of the null hypothesis) is not innocuous. If we reverse the numerator and denominator in (10), under strong identification, an additional $\log(T)$ term will emerge in (10) that makes it impossible to derive the asymptotic bounds on the distribution of the Bayes factor.

As in the case of the LR test statistic, tighter confidence intervals can be obtained, if we know which structural parameters are strongly identified, as shown in Proposition 3b.

PROPOSITION 3b (Asymptotic Distribution of the BF Test Statistic When Some Structural Parameters Are Known to Be Strongly Identified). *In addition to assumptions* (a), (b), *and* (c) *of Proposition 2, and assumptions* (a), (b), *and* (c) *of Theorem 1 with* $\tilde{\gamma}_T(\theta_0)$ *replaced by* $\tilde{\gamma}_T(\alpha_0)$, *suppose that* $\tilde{\beta}_T(\alpha_0)$ *is consistent for* β_0 *and that* $D_\beta f(\gamma_{0,T}, \theta_0)$ *has rank* k_2. *Then, conditional on the strongly identified parameters,*

$$\lim_{T\to\infty} \Pr\left(\text{Bayes factor}(\alpha_0) \leq e^{z_T' Q_T z_T/2}\right) = 1, \tag{12}$$

[6]The idea of deriving bounds for the asymptotic distribution of the Bayes factor is not without precedent. Similar ideas can be found in Edwards, Lindman, and Savage (1963) and Berger and Sellke (1987), for example.

where

$$\begin{aligned} Q_T = {} & H_T^{-1/2} R' \big(R H_T^{-1} R'\big)^{-1} R H_T^{-1/2} \\ & - H_T^{-1/2} R' \big(R H_T^{-1} R'\big)^{-1} R_\beta \big[R_\beta' \big(R H_T^{-1} R'\big)^{-1} R_\beta\big]^{-1} \\ & \times R_\beta' \big(R H_T^{-1} R'\big)^{-1} R H_T^{-1/2} \end{aligned} \tag{13}$$

is idempotent and has rank $r - k_2$, $H_T = \nabla_{\gamma\gamma} \ell_T(\gamma_{0,T})$, *and* $R_\beta = D_\beta f_T(\gamma_{0,T}, \theta_0)$.

2.4 *A power comparison of the LR and BF tests*

It is useful to compare the asymptotic properties of the LR and BF intervals. The LR confidence set has $1 - \vartheta$ coverage probability asymptotically. The BF interval is more conservative in that a $1 - \vartheta$ confidence set has at least a $1 - \vartheta$ coverage probability asymptotically. This difference arises because, unlike the LR statistic, $2\ln(\text{Bayes factor}(\theta_0))$ in (10) is merely bounded by a random variable with a χ^2_r distribution. Therefore, asymptotically, we would expect the LR interval to be tighter.

To formally analyze the asymptotic power of the LR and BF tests, we partition Θ into $\mathcal{A}$ and $\mathcal{B}$, where $\alpha \in \mathcal{A}$ denotes the weakly identified elements in the parameter vector $\theta \in \Theta$ and $\beta \in \mathcal{B}$ denotes the strongly identified elements. Suppose that the true parameter value is $\theta_{1,T} = [\alpha_1' \;\; \beta_{1,T}']'$ and that $\beta_{1,T} = \beta_0 + T^{-1/2} c$, where $c \in \Re^{k_2}$ and $\theta_0 = [\alpha_0' \;\; \beta_0']'$ is the hypothesized parameter value.

Theorem 2 (Asymptotic Power of the LR and BF Tests). *Suppose that assumptions* (a) *and* (b) *in Proposition 2 hold, with* $\gamma_{0,T}$ *in assumption* (b) *replaced by* $\gamma_T(\theta)$, *and that assumption* (a) *in Theorem 1 holds. In addition, make the following assumptions*:

(a) $f_T(\gamma, \theta) = f_1(\gamma, \beta) + T^{-1/2} f_2(\gamma, \theta)$, $f_1 : \mathcal{B} \to \Re^{\dim(\gamma)}$ *and* $f_2 : \Theta \to \Re^{\dim(\gamma)}$ *are continuously differentiable,* $\Theta = \mathcal{A} \times \mathcal{B} \subset \Re^{k_1} \times \Re^{k_2}$, $\mathcal{A}$ *and* $\mathcal{B}$ *are compact in* $\Re^{k_1}$ *and* $\Re^{k_2}$, *respectively, and* $D_\gamma f_1(\gamma, \beta_0)$ *is nonsingular for all* γ.

(b) *Let*

$$\delta_{T,j} = \begin{cases} o(1), & j = 1, \ldots, k_1, \\ o\big(T^{-1/2}\big), & j = k_1 + 1, \ldots, k. \end{cases}$$

A practical method for choosing $\delta_{T,j}$ *subject to these constraints will be discussed later in this section.*

(c) *If* $k_2 > 0$, *then*

$$T^{k_2/2} \sup_{\beta \in \mathcal{B}: \neg\exists \bar{c} \in \bar{C}_T \text{ such that } \beta = \beta_0 + \bar{c} T^{-1/2}} \exp\big(\ell_T\big(\tilde{\gamma}_T(\theta)\big) - \ell_T(\hat{\gamma}_T)\big) \xrightarrow{p} 0,$$

where $\bar{C}_T = \{\bar{c} \in \Re^{k_2} : -c_{\min} T^\eta \le c \le c_{\max} T^\eta\}$, $c_{\max} > 0$, $c_{\min} > 0$, *and* $\eta \in (0, 1/2)$.

Then we have

$$\mathrm{LR}_T(\theta_0) \xrightarrow{d} \big(d(\alpha_0, \alpha_1, \beta_0, c) + z\big)' P \big(d(\alpha_0, \alpha_1, \beta_0, c) + z\big) \tag{14}$$

and

$$
\begin{aligned}
&T^{k_2/2}\,\text{Bayes factor}(\theta_0)\\
&\xrightarrow{d}\int_{\mathcal{A}\times\Re^{k_2}}\pi(\alpha,\beta_0)\\
&\quad\times\exp\biggl(-\frac{1}{2}\bigl(d(\alpha,\alpha_1,\beta_0,c)+z\bigr)'P\bigl(d(\alpha,\alpha_1,\beta_0,c)+z\bigr)\biggr)d\begin{bmatrix}\alpha' & c'\end{bmatrix}'\\
&\quad\Big/\exp\biggl(-\frac{1}{2}\bigl(d(\alpha_0,\alpha_1,\beta_0,0)+z\bigr)'P\bigl(d(\alpha_0,\alpha_1,\beta_0,0)+z\bigr)\biggr),
\end{aligned}
\tag{15}
$$

where $z\sim N(0_{r\times1},I_r)$, $d(\alpha,\alpha_1,\beta_0,c)=V_\gamma^{-1/2}G(\alpha,\alpha_1,\beta)[c'\ (\alpha-\alpha_1)']'$, $G(\alpha,\alpha_1,\beta_0)=\lim_{T\to\infty}\{[D_\gamma f_1(\gamma_{1,T},\beta_0)]^{-1}D_\beta f_1(\gamma_{1,T},\beta_0)\ [D_\gamma f_1(\gamma_{1,T},\beta_0)]^{-1}D_\alpha f_2(\gamma_{1,T},\bar\alpha)\}$, $[\bar\gamma_T'\ \bar\theta_T']'=[\bar\gamma_T'\ \bar\alpha_T'\ \bar\beta_T']'$ *is a point between* $[\gamma_{1,T}'\ \alpha_1'\ \beta_0'+T^{-1/2}c']'$ *and* $[\gamma_{0,T}'\ \alpha_0'\ \beta_0']'$, *and* $P=V_\gamma^{1/2}R'(RV_\gamma R')^{-1}RV_\gamma^{1/2}$.

Remarks.

1. It is useful to relate Theorem 2 to the normal linear regression model example of Section 2.1. In that example, there is no strongly identified structural parameter ($\theta=\alpha$ and $k_2=0$). Because $d(\alpha,\alpha_1)=-\sqrt{E(x_t^2)}(\alpha-\alpha_1)$ and $P=1$, by Theorem 2,

$$
\mathrm{LR}_T(\alpha_0)\xrightarrow{d}\Bigl(-\sqrt{E\bigl(x_t^2\bigr)}(\alpha_0-\alpha_1)+z\Bigr)^2,
\tag{16}
$$

$$
\begin{aligned}
&2\ln\bigl(\text{Bayes factor}(\alpha_0)\bigr)\\
&\xrightarrow{d}\Bigl(-\sqrt{E\bigl(x_t^2\bigr)}(\alpha_0-\alpha_1)+z\Bigr)^2\\
&\quad+2\ln\biggl\{\int_{\Re}\frac{1}{\sqrt{2\pi\bar\sigma^2}}\exp\biggl(-\frac{(\alpha-\bar\alpha)^2}{2\bar\sigma^2}\biggr)\\
&\quad\times\exp\biggl[-\frac{1}{2}\Bigl(-\sqrt{E\bigl(x_t^2\bigr)}(\alpha-\alpha_1)+z\Bigr)^2\biggr]\biggr\}d\alpha\\
&\overset{d}{\equiv}\Bigl(-\sqrt{E\bigl(x_t^2\bigr)}(\alpha_0-\alpha_1)+z\Bigr)^2-\ln\bigl(1+\bar\sigma^2E\bigl(x_t^2\bigr)\bigr)\\
&\quad-\frac{1}{1+\bar\sigma^2E(x_t^2)}\Bigl(-\sqrt{E\bigl(x_t^2\bigr)}(\bar\alpha-\alpha_1)+z\Bigr)^2.
\end{aligned}
\tag{17}
$$

The last two terms explain the power loss of the BF relative to the LR test. The relative power loss is increasing in the distance between the prior mean and the true parameter value. The effect of the prior variance on the power of the test is mixed. When the prior variance is larger, the effect of prior beliefs on the last term in (19) is smaller, increasing the power, but increased uncertainty also reduces the power by lowering the second-to-last term. Note that even under the null hypothesis ($\alpha_1=\alpha_0$), the last two terms in (19) are still present, which explains the conservative nature of the BF test.

2. In assumption (a) of Theorem 2, we model f_T such that the part of f_T that depends on weakly identified parameters vanishes asymptotically and the number of restrictions is the same as the number of reduced-form parameters. By the implicit function theorem,

$$\begin{aligned}&\begin{bmatrix} \dfrac{\partial \gamma_{0,T}}{\partial \alpha'} & \dfrac{\partial \gamma_{0,T}}{\partial \beta'} \end{bmatrix} \\ &\quad = \big[D_\gamma f_T(\gamma_{0,T}, \theta_0)\big]^{-1} \begin{bmatrix} T^{-1/2} \dfrac{\partial f_2(\gamma_{0,T}, \theta_0)}{\partial \alpha'} & \dfrac{\partial f_T(\gamma_{0,T}, \theta_0)}{\partial \beta'} \end{bmatrix},\end{aligned} \tag{18}$$

which allows us to model the scaled slope of the log likelihood with respect to α as local to zero.

3. Assumption (a) allows for the case in which the parameters are all weakly identified such that $\theta = \alpha$ and $f_T(\theta) = T^{-1/2} g_2(\alpha)$, and the case of $k = k_1$ such that $0 < k_2 < k$ and $f_2(\theta) \equiv 0$ for all θ, and the case in which they are all strongly identified such that $\theta = \beta$, $f_T(\theta) = f_2(\beta)$, and $k = k_2$.

4. The assumption that the Jacobian is nonsingular on the space of reduced-form parameters is trivially satisfied if $f_T(\gamma, \theta)$ equals $\gamma - g_T(\theta)$, as in the conventional representation of the A, B, C, D matrices.

5. Assumption (c) says that if the value of strongly identified parameters is not in a neighborhood of the true parameter value, the difference between the value of the likelihood at that parameter value and the maximized value of the likelihood diverges at rate T^c for some $c > 0$. A sufficient condition is that the MLE of the strongly identified parameters converges at rate $T^{1/2}$.

6. Expression (15) implies that

$$\begin{aligned}&2\ln\big(\text{Bayes factor}(\theta_0)\big) - k_2 \ln(T) \\ &\quad \xrightarrow{d} \big(d(\alpha_0, \alpha_1, \beta_0, 0) + Pz\big)' \big(d(\alpha_0, \alpha_1, \beta_0, 0) + Pz\big) \\ &\qquad + 2\ln\Bigg(\int_{\mathcal{A} \times \Re^{k_2}} \pi(\alpha, \beta_0) \\ &\qquad \times \exp\Bigg(-\frac{1}{2} \big(d(\alpha, \alpha_1, \beta_0, c) + Pz\big)' \big(d(\alpha, \alpha_1, \beta_0, c) + Pz\big) \Bigg) d\begin{bmatrix} \alpha' & c' \end{bmatrix}' \Bigg).\end{aligned} \tag{19}$$

Note that the first term of (19) is the asymptotic noncentral chi-squared distribution of the LR test statistic and that the second term is negative with probability 1 because it is the log of a number that is less than 1 with probability 1. Thus, even if k_2 is known, the LR test is more powerful than the BF test when the chi-squared critical values are used, generalizing the intuition provided by the example of Section 2.1.

To summarize, the LR test has power against local alternatives for strongly identified parameters and against global alternatives for weakly identified parameters. The BF test has power against global alternatives for weakly identified parameters only in the

absence of strongly identified parameters. The BF test also lacks power against local alternatives for the strongly identified structural parameters, but both the LR test and the BF test are consistent against global alternatives for the strongly identified parameters, as shown in Proposition 4.

PROPOSITION 4 (Consistency of the LR and BF Tests). *Suppose that assumptions* (a) *and* (b) *in Proposition 2, assumption* (a) *in Theorem 1, and assumptions* (a) *and* (b) *in Theorem 2 hold with the* $\gamma_{0,T}$ *in assumption* (b) *of Proposition 2 replaced by* $\gamma_T(\theta)$. *Then for* $\beta_1 \neq \beta_0$,

$$\mathrm{LR}_T(\beta_1) \xrightarrow{p} \infty, \tag{20}$$

$$\text{Bayes factor}(\beta_1) \xrightarrow{p} \infty. \tag{21}$$

2.5 *Implementation issues*

2.5.1 *Determining the degrees of freedom* The construction of valid confidence sets requires knowledge of the degrees-of-freedom parameter r in the asymptotic distribution of the LR and BF statistics. Our approach to determining the number of identified reduced-form parameters exploits the similarity transform, similar to the approach taken in Komunjer and Ng (2009) in the context of evaluating the identifiability of structural DSGE model parameters. Recall that the linear approximation to the DSGE model can be written in state-space format (6) and (7). If there is a nonsingular matrix T such that $C^* = CT^{-1}$, $x_t^* = Tx_t$, $A^* = TAT^{-1}$, and $B^* = TB$ satisfy the same restrictions as C, A, and B, then there exists another observationally equivalent state-space representation

$$x_{t+1}^* = A^* x_t^* + B^* w_t, \tag{22}$$

$$y_t = C^* x_t^* + D w_t. \tag{23}$$

Our objective is to find the minimal state-space representation among the set of equivalent representations. A state-space representation (A, B, C, D) is minimal if there is no equivalent state-space representation involving fewer state variables. In practice, finding the minimal state-space representation involves trial and error on a model-by-model basis. This involves checking the rank conditions that (i) $[C' \;\; A'C' \;\; \cdots \;\; A^{(n-1)\prime}C']$ has rank n (observability) and (ii) that $[B \;\; AB \;\; A^2B \;\; \cdots \;\; A^{n-1}B]$ has rank n (reachability), where n is the number of state variables. For most models, these rank conditions must be evaluated numerically at a large number of randomly chosen structural parameter values. The state-space representation is minimal if and only if it is observable and reachable. If any one of the two rank conditions fails, we need to search for a minimal representation with fewer state variables.

Conditional on having derived the minimal state-space representation, if

$$\begin{bmatrix} A' \otimes I_n - I_n \otimes \bar{A} \\ B' \otimes I_n \\ -I_n \otimes \bar{C} \end{bmatrix} \tag{24}$$

has rank n^2 for any distinct feasible pairs (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ (see Proposition 1 of Glover and Willems (1974, pp. 643–644) and Proposition 1 of Komunjer and Ng (2009)), then the dimension of γ corresponds to the number of free elements in A, B, C, and D. Further details on the derivation of the degrees-of-freedom parameter for each of our simulation designs and for the empirical example are provided in the Online Appendix, available in a supplementary file on the journal website, http://www.qeconomics.org/supp/306/supplement.pdf.

2.5.2 *The projection method* Although our approach does not allow the construction of point estimates of θ, the projection method can be used to construct confidence intervals for individual elements of θ from the LR and BF joint confidence sets (see, e.g., Dufour and Taamouti (2005), Chaudhuri and Zivot (2011) for the use of the projection method in linear instrumental variable (IV) and GMM models, respectively). Here we focus on the BF approach without loss of generality. The level $1-\vartheta$ confidence interval for the ith parameter θ_j is $(\underline{\theta}_j, \bar{\theta}_j)$, where the lower and upper confidence bounds are

$$\underline{\theta}_j = \min\Big\{\theta_j \in \Theta_j : \max_{\theta_{-j}\in\Theta_{-j}} \text{Bayes factor}\big((\theta_j, \theta_{-j})\big) \le e^{\chi^2_{\gamma,1-\vartheta}/2}\Big\}, \tag{25}$$

$$\bar{\theta}_j = \max\Big\{\theta_j \in \Theta_j : \max_{\theta_{-j}\in\Theta_{-j}} \text{Bayes factor}\big((\theta_j, \theta_{-j})\big) \le e^{\chi^2_{\gamma,1-\vartheta}/2}\Big\}, \tag{26}$$

where θ_{-j} is the parameter vector that excludes θ_j and Θ_{-j} is the parameter space that excludes the parameter space for θ_j. 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 (25) and (26) is computationally challenging. In practice, we replace Θ in (25) and (26) by the set of Monte Carlo realizations of Θ, which reduces the computational burden. This approach is justified because the set of Monte Carlo realizations becomes dense in the parameter space, as the number of Monte Carlo draws increases.

To implement the BF method, 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,

$$\text{Bayes factor}(\theta_0) \approx \frac{\pi(H_0)}{P(H_0|X)} = \frac{\dfrac{1}{|\delta_T|}\pi(H_0)}{\dfrac{1}{|\delta_T|}P(H_0|X)},$$

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

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

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

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,

$$\frac{1}{|\delta_T|}\hat{\pi}(H_0) = \frac{1}{|\delta_T| M}\sum_{j=1}^{M} I\big(\theta^{(j)} \in B_{\delta_T}(\theta_0)\big), \tag{27}$$

$$\frac{1}{|\delta_T|}\hat{P}(H_0|X) = \frac{1}{|\delta_T| M}\sum_{j=1}^{M} I\big(\tilde{\theta}^{(j)} \in B_{\delta_T}(\theta_0)\big). \tag{28}$$

Note that the right-hand sides of (27) and (28) can be interpreted as a multivariate density estimator based on a uniform kernel with δ_T as the bandwidth. Let

$$\delta_{T,j} = \hat{\sigma}_j\left(\frac{1}{T}\right)^{1/(p+4)}, \tag{29}$$

where $\hat{\sigma}_j$ is the standard deviation of the posterior distribution of θ_j (see, e.g., Scott (1992, p. 152)). Because the prior and posterior distributions are not necessarily normal and the kernel is not normal, (29) need not be optimal, but it nevertheless satisfies assumption (b) of Theorem 2. Note that if θ_j is strongly identified, $\hat{\sigma}_j = o_p(1)$ and, thus, $\delta_{T,j} = o_p(T^{-1/2})$; if θ_j is weakly identified, $\hat{\sigma}_j = O_p(1)$ and $\delta_{T,j} = o_p(1)$. Hence, this choice for δ_T satisfies assumption (b) and does not affect the limiting distribution of the BF test statistic.

3. Small-sample accuracy

We are interested in comparing the small-sample accuracy of the LR and BF methods and of pseudo-Bayesian methods that reinterpret posterior estimates from a frequentist point of view. We consider two data generating processes for this Monte Carlo study. As an illustrative example, we first focus on a small-scale New Keynesian model similar to Woodford (2003, p. 246) that has previously been used as an example in the related literature (see, e.g., Canova and Sala (2009)).

3.1 *Simulation design 1*

The economy consists of a Phillips curve, a Taylor rule, an investment–savings relationship, and the exogenous driving processes z_t and ξ_t:

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

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

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

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

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

where x_t, π_t, and R_t denote the output gap, the inflation rate, and the interest rate, respectively. The shocks ε_t^z and ε_t^r are assumed to be distributed NID$(0,1)$. The model parameters are the discount factor β, the intertemporal elasticity of substitution σ, the probability α of not adjusting prices for a given firm, the elasticity of substitution across varieties of good, θ, and the parameter ω that controls the disutility of labor supply; ϕ_π and ϕ_x capture the central bank's reaction to changes in inflation and the output gap, respectively, 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. In particular, α and θ are at most partially identified. The parameters of this data generating process (DGP) are $\sigma = 1$, $\alpha = 0.75$, $\beta = 0.99$, $\phi_\pi = 1.5$, $\phi_x = 0.125$, $\omega = 1$, $\rho_r = 0.75$, $\rho_z = 0.90$, $\theta = 6$, $\sigma^z = 0.30$, and $\sigma^r = 0.20$. These parameter values are standard choices in the macroeconomics literature (see An and Schorfheide (2007), Woodford (2003)). As shown in the Online Appendix, the degrees-of-freedom parameter of the LR statistic and of the BF statistic for this small-scale New Keynesian model is 8.

Our Monte Carlo study consists of the following steps:

Step 1. We generate 1,000 synthetic data sets of length T for output and inflation using the New Keynesian model as the DGP. 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)). The larger sample corresponds to the period between 1960 and 2006. For each synthetic data set, we treat output and inflation as our observables and estimate a total of eight parameters: $\Phi = [\alpha \ \ \phi_\pi \ \ \phi_x \ \ \theta \ \ \rho_r \ \ \rho_z \ \ \sigma^r \ \ \sigma^z]$. The remaining parameters are treated as known in the estimation.

Step 2. For each synthetic data set, we estimate the structural parameters of interest and construct joint and pointwise confidence sets by the BF and LR methods.

Step 2a. For the BF method, 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). As the baseline, we consider two types of priors, which are summarized in Table 1. For the uniform priors, we impose boundary restrictions to make the priors proper and to avoid implausible values (e.g., negative variances, persistence parameters outside the unit circle, and indeterminacy of the model). As an alternative, we consider the informative priors proposed in An and Schorfheide that are centered around the true values in our DGP with loose standard deviations (see Table 1). The algorithm involves five steps:

(i) Let $L(\Phi|Y)$ and $p(\Phi)$ denote the likelihood of the data conditional on the parameters and the prior probability, respectively. Obtain the posterior mode $\widetilde{\Phi} = \arg\max[\ln p(\Phi) + \ln L(\Phi|Y)]$ using a suitable maximization routine. To ensure that we find the maximum, we provide our maximization procedure with 10 randomly selected starting points, which gives us a set of potential maxima $\{\widetilde{\Phi}_i\}_{i=1}^{10}$. The mode corresponds to the candidate that achieves the highest value among the 10 potential candidates.

TABLE 1. Prior specification for parameters of the small-scale New Keynesian model.

Uniform Priors			
Parameters	Distributions	Lower Bound	Upper Bound
ϕ_p	Uniform	1	5
ϕ_x	Uniform	0	5
α	Uniform	0	1
θ	Uniform	1	15
ρ_z	Uniform	0	1
ρ_r	Uniform	0	1
σ_z	Uniform	0	1
σ_r	Uniform	0	1
Informative Priors			
Parameters	Distributions	Mean	Standard Deviations
ϕ_p	Gamma	1.5	0.25
ϕ_x	Gamma	0.125	0.1
α	Beta	0.75	0.2
θ	Normal	6	2
ρ_z	Beta	0.9	0.2
ρ_r	Beta	0.75	0.2
σ_z	Inverse gamma	0.3	0.2
σ_r	Inverse gamma	0.2	0.2

(ii) Let $\widetilde{\Sigma}$ be the inverse Hessian evaluated at the posterior mode. Draw $\Phi^{(0)}$ from a normal distribution with mean $\widetilde{\Phi}$ and covariance matrix $\varkappa^2\widetilde{\Sigma}$, where $\varkappa^2$ is a scaling parameter.

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

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

The posterior distributions are characterized using $M = 100{,}000$ iterations after discarding an initial burn-in phase of 1,000 draws. In light of the computational cost, the confidence intervals discussed below are based on 5,000 draws randomly chosen from these 100,000 draws. Selecting $\varkappa^2$ is a delicate issue. Ideally, one should fine-tune that parameter for each synthetic data set, so that the acceptance rate falls within the values suggested by Roberts, Gelman, and Gilks (1997). Given the scale of our experiment, hand picking $\varkappa^2$ for each synthetic data set is prohibitively expensive. Instead, we set one common scaling parameter for our exercise. We obtain this value by fine-tuning $\varkappa^2$ based on 10 separate Monte Carlo replications and then taking the average.

(iv) Finally, we use Gelfand and Smith's (1990) approach to check the convergence of the Metropolis–Hasting sampler for each synthetic data set.

(v) On the basis of the posterior distribution, we construct the joint and pointwise BF confidence sets for Φ.

STEP 2b. For the LR method, we use the Kalman filter to estimate the reduced-form representation of the New Keynesian model, we construct and invert the LR test statistic, and we generate joint and pointwise LR confidence sets for Φ.

STEP 3. For each method, we evaluate the coverage accuracy of the joint and pointwise 90% confidence sets in repeated sampling. We also report median interval lengths for the pointwise intervals.

3.1.1 *Coverage accuracy and median interval lengths* Table 2 shows the finite-sample coverage accuracy of the joint confidence sets obtained from the LR and BF approach. The LR results do not depend on the prior of course. For the BF method, we focus on the results for a uniform and for an informative prior for now. For a nominal 90% confidence set, both sets should have at least 90% coverage asymptotically, with the BF interval being more conservative. Table 2 shows that for $T = 96$, the coverage accuracy of the BF confidence sets is between 97% and 99%, whereas that of the LR confidence set is 87%. For $T = 188$, the BF confidence sets have coverage rates of about 99%, whereas the coverage accuracy of the LR confidence set is 91%. Thus, both methods appear to be accurate under the null hypothesis, but the LR method comes closer to attaining the nominal coverage probability.

Table 3 shows the corresponding coverage rates and median interval lengths for individual structural parameters. The full results are shown in the Online Appendix. Table 3(a) shows that, for $T = 96$, the LR coverage rates are between 94% and 97%. For $T = 188$, the coverage accuracy of the LR test is between 95% and 98%. It is useful to contrast these results with the coverage accuracy of pseudo-Bayesian intervals. The first three entries in each panel of Table 3(b) and (c) focus on the traditional asymptotic confidence interval that a frequentist user might construct from the posterior mode, mean, or median by adding $+/-1.645$ posterior standard errors. Some of the effective coverage rates are well below the nominal rates. For $T = 96$, the coverage probability may be as low as 56%. Alternatively, a frequentist user may focus on the nominal 90% equal-tailed percentile interval based on the posterior distribution, as in the fourth row (see, e.g., Balke, Brown, and Yücel (2008)). The coverage rate of this percentile interval may drop as low as 42%. In contrast, if we construct the interval by inverting the Bayes factor (BF

TABLE 2. Effective coverage rates of nominal 90% confidence sets.

		$T = 96$	$T = 188$
Small-Scale New Keynesian Model With Two Observables			
LR Set		0.873	0.908
BF Set	Uniform prior	0.965	0.985
BF Set	Informative prior	0.988	0.986
BF Set	Asymmetric prior	0.931	0.952

TABLE 3. Coverage rates and median lengths of nominal 90% confidence intervals.

		α		θ		σ_z		σ_r	
T		Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
	(a) Small-Scale New Keynesian Model With Two Observables: LR								
96	LR	0.962	0.25	0.970	5.24	0.942	0.24	0.939	0.31
188	LR	0.973	0.25	0.976	5.80	0.953	0.22	0.960	0.30
	(b) Small-Scale New Keynesian Model With Two Observables: Uniform Priors								
96	Median ± 1.645SD	0.834	0.22	1.000	12.8	0.649	0.49	0.629	0.39
	Mean ± 1.645SD	0.906	0.22	1.000	12.8	0.622	0.49	0.563	0.39
	Mode ± 1.645SD	0.823	0.22	0.772	12.8	0.893	0.49	0.917	0.39
	Percentile	0.975	0.21	1.000	12.3	0.548	0.49	0.416	0.39
	BF	1.000	0.32	1.000	14.0	0.994	0.76	0.990	0.79
188	Median ± 1.645SD	0.927	0.21	1.000	12.8	0.728	0.34	0.735	0.26
	Mean ± 1.645SD	0.965	0.21	1.000	12.8	0.694	0.34	0.681	0.26
	Mode ± 1.645SD	0.878	0.21	0.795	12.8	0.922	0.34	0.947	0.26
	Percentile	0.987	0.20	1.000	12.3	0.628	0.34	0.581	0.26
	BF	1.000	0.30	1.000	14.0	0.997	0.74	0.998	0.69
	(c) Small-Scale New Keynesian Model With Two Observables: Informative Priors								
96	Median ± 1.645SD	0.997	0.12	1.000	6.41	0.895	0.21	0.909	0.14
	Mean ± 1.645SD	0.996	0.12	1.000	6.41	0.916	0.21	0.921	0.14
	Mode ± 1.645SD	0.946	0.12	1.000	6.41	0.636	0.21	0.708	0.14
	Percentile	0.996	0.12	1.000	6.44	0.939	0.21	0.944	0.14
	BF	1.000	0.25	1.000	12.5	0.999	0.38	1.000	0.18
188	Median ± 1.645SD	1.000	0.12	1.000	6.39	0.906	0.18	0.907	0.12
	Mean ± 1.645SD	1.000	0.12	1.000	6.39	0.914	0.18	0.935	0.12
	Mode ± 1.645SD	0.983	0.12	1.000	6.39	0.669	0.18	0.711	0.12
	Percentile	1.000	0.12	1.000	6.41	0.932	0.18	0.952	0.12
	BF	1.000	0.22	1.000	12.3	1.000	0.32	1.000	0.19
	(d) Small-Scale New Keynesian Model With Two Observables: Asymmetric Priors								
96	Median ± 1.645SD	0.103	0.14	0.400	8.96	0.636	0.49	0.619	0.41
	Mean ± 1.645SD	0.109	0.14	0.713	8.96	0.592	0.49	0.542	0.41
	Mode ± 1.645SD	0.521	0.14	0.504	8.96	0.874	0.49	0.919	0.41
	Percentile	0.118	0.14	0.185	8.51	0.519	0.50	0.384	0.41
	BF	0.968	0.24	1.000	9.49	0.994	0.78	0.981	0.79
188	Median ± 1.645SD	0.163	0.13	0.342	8.95	0.704	0.35	0.756	0.26
	Mean ± 1.645SD	0.170	0.13	0.646	8.95	0.677	0.35	0.696	0.26
	Mode ± 1.645SD	0.550	0.13	0.500	8.95	0.899	0.35	0.949	0.26
	Percentile	0.176	0.13	0.145	8.51	0.599	0.35	0.566	0.26
	BF	0.978	0.23	1.000	9.50	0.995	0.73	0.994	0.70

interval), as shown in the last row, all intervals for individual parameters have coverage rates of at least 99%, well in excess of the required coverage accuracy. As the sample size is increased, the accuracy of the pseudo-Bayesian delta method intervals improves, but may remain as low as 67%, depending on the parameter. The corresponding percentile intervals have coverage rates as low as 58%. The conservative intervals based on inverting the Bayes factor in all cases have essentially 100% coverage probability.

The results in Table 3(b) and (c) also indicate that the accuracy of some pseudo-Bayesian intervals for the structural parameters α and θ can be quite good, even when those parameters are weakly identified as in our experiment. The reason is as follows: In the weakly identified case, the posterior distribution essentially replicates the prior distribution. Thus, a natural conjecture is that the symmetry of the priors for α and θ about their true values is responsible for the relatively high accuracy of the traditional methods because it makes it more likely that the credible interval includes the true parameter value. To verify our conjecture, we repeated the Monte Carlo experiment with uniform priors with bounds of $[0, 0.8]$ and of $[5.5, 15]$ for α and θ, respectively. Under this alternative asymmetric prior, the true values are close to the boundary of the support of the priors. As Table 3(d) shows, in that case, the coverage rates for the traditional confidence intervals for α decline to values as low as 10% for $T = 96$, and as low as 16% for $T = 188$. For θ, the coverage rate may drop as lows as 19% for $T = 96$ and 15% for $T = 188$. Even under the most optimistic scenario based on the mode, the accuracy for those parameters is only about 50% and 50%, respectively. On the other hand, the BF approach remains quite robust to the new priors, delivering coverage rates of at least 97% for the same parameters. We conclude that pseudo-Bayesian interval estimates of the type a frequentist may construct from Bayesian posterior estimates for the parameters of Bayesian DSGE models are not reliable, and that LR and BF intervals have the potential to achieve substantial improvements in coverage accuracy.

An obvious concern is that the higher coverage of the LR and BF intervals for individual parameters reflects a substantial increase in length. For example, in the limiting case of no identification, one would expect appropriately sized intervals to cover the support of the structural parameter. A comparison of the median interval length for each method and parameter in Table 3 shows, first, that robust LR and BF intervals are often wide, but not necessarily excessively so. Second, there is no clear ranking between the median length of BF and LR intervals. This is not surprising because prior information in small samples may help reduce the interval length. Only asymptotically would one expect the power advantages of the LR intervals to be decisive. Third, although allowing for weak identification tends to increase interval length, in some cases pointwise LR intervals may be shorter than the corresponding pseudo-Bayesian intervals that a frequentist user might have constructed.

3.2 *Simulation design 2*

Whereas the reduced form of the first simulation design can be expressed as a finite-order VAR model, we now focus on a DGP that does not have a finite-order VAR representation (see the Online Appendix). The model is obtained by including interest rates as an additional observable. This modification requires the introduction of an additional structural shock, mk_t, in the Phillips curve to avoid a stochastic singularity. We allow all shocks to follow AR(1) processes. The new DGP is

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

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

$$x_t = \mathbb{E}_t x_{t+1} - \sigma(R_t - \mathbb{E}_t \pi_{t+1} - z_t),$$
$$mk_t = \rho_{mk} mk_{t-1} + \sigma^{mk} \varepsilon_t^{mk},$$
$$z_t = \rho_z z_{t-1} + \sigma^z \varepsilon_t^z,$$
$$\xi_t = \rho_\xi \xi_{t-1} + \sigma^\xi \varepsilon_t^\xi.$$

The DGP parameters for this design are the same as for the first design with the following DGP parameters added: $\rho_\xi = 0.5$, $\sigma^\xi = 0.2$, $\rho_{mk} = 0.9$, and $\sigma^{mk} = 0.14$. The priors are augmented as follows: For the case of uninformative priors, we postulate uniform $[0, 1]$ distributions for ρ_{mk}, ρ_ξ, and σ^{mk}. For the informative priors, we postulate beta prior distributions $\mathcal{B}(0.9, 0.2)$ for ρ_{mk} and $\mathcal{B}(0.5, 0.2)$ for ρ_ξ (the numbers in parentheses are the mean and standard deviation), whereas for the scale parameter σ^{mk} and σ^ξ, we postulate an inverse gamma prior distribution with mean 0.14 and standard deviation 0.2, and with mean 0.2 and standard deviation 0.2, respectively. The model is estimated directly in state-space form. As shown in the Online Appendix, the degrees-of-freedom parameter for the LR and BF tests is 18 for this modified model.

3.2.1 *Coverage accuracy and median interval lengths* Tables C.4 and C.5 in Appendix C show that, as in design 1, the coverage accuracy of the joint LR confidence set is somewhat lower than in the first example, with 86% for $T = 96$ and 87% with $T = 188$. Further analysis showed that these small size distortions vanish for higher T. The corresponding joint BF confidence sets are somewhat sensitive to the prior. For the uniform prior joint, the coverage rates are 94% for $T = 96$ and 91% for $T = 188$. Likewise, the results for the asymmetric prior yield acceptable coverage rates of 90% for $T = 96$ and 88% for $T = 188$. On the other hand, for the informative prior, the corresponding rates are 83% and 79%. This result illustrates that the joint coverage rates of the BF method in finite samples can be somewhat sensitive to the choice of prior. This is not the case for the individual coverage rates, however. The LR intervals have coverage rates between 94% and 99%; the BF intervals have coverage rates of essentially 100% regardless of the choice of prior. In contrast, the pseudo-Bayesian intervals have coverage rates anywhere between 8% and 100%. The results on median interval length are substantively identical to those for the first simulation design. We conclude that both the LR and BF intervals can substantially improve the accuracy of confidence sets for structural model parameters when one or more parameters are weakly identified.

4. Empirical application: Quantifying wage and price rigidities

To illustrate the practical usefulness of our methodology, we now construct LR and BF confidence intervals in a medium-scale state-of-the-art DSGE framework. Our model specification follows very closely that of Del Negro and Schorfheide (2008), who in turn build on Smets and Wouters (2007) and Christiano, Eichenbaum, and Evans (2005). Since this type of environment has been extensively discussed in the literature, we omit a discussion of the model. 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 Fernández-Villaverde, Guerron-Quintana, and Rubio-Ramírez (2010).

4.1 *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 the nominal interest rate (annualized). We use their quarterly data set for the period 1982.Q1–2005.Q4. We set our priors alternatively to the agnostic prior, the low-rigidities prior, and the high-rigidities priors employed in Del Negro and Schorfheide (see Tables 1–3 in their paper).

The parameter space is partitioned into two sets:

$$\Theta_1 = \begin{bmatrix} \alpha & \delta & g & L^* & \psi \end{bmatrix},$$

which is not estimated, and

$$\Theta_2 = \begin{bmatrix} r_* & \gamma & \lambda_f & \pi_* & \zeta_p & \iota_p & \zeta_w & \iota_w & \lambda_w & S'' & h & a'' & \upsilon_l & \psi_1 & \psi_2 \\ \rho_r & \rho_z & \sigma_z & \rho_\phi & \sigma_\phi & \rho_\lambda & \sigma_\lambda & \rho_g & \sigma_g & \sigma_r & L_{\text{adj}} \end{bmatrix},$$

which is. For definitions of these parameters, the reader is referred to Del Negro and Schorfheide (2008). The following values are used for the first set of parameters: $\alpha = 0.33$, $\delta = 0.025$, $g = 0.22$, $L^* = 1$, and $\psi = 0$. Although these values are standard choices in the DSGE literature, some clarifications are in order. As in Del Negro and Schorfheide (2008), our parameterization imposes the constraint that firms make zero profits in the 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

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

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

The posterior distributions of the parameters in the set Θ_2 are characterized using the random-walk Metropolis–Hasting algorithm outlined in Section 3.1. 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, Carlin, Stern, and Rubin (2004)). The

degrees-of-freedom parameter in the limiting distribution of the LR and BF test statistics is 99, as shown in the Online Appendix.

4.2 *The relative importance of wage and price rigidities in the U.S. economy*

Table 4(a) summarizes the posterior means, medians, and modes as well as the posterior standard deviations of selected structural parameters, along with the 90% credible interval (obtained from the percentiles of the posterior distribution) and the 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. These parameters represent the probabilities of not reoptimizing prices and wages, respectively. The length of price contracts is defined as $\frac{1}{1-\zeta_p}$ quarters, where ζ_p is the probability of not reoptimizing prices today. By analogy, the length of the wage contract is $\frac{1}{1-\zeta_w}$ quarters. The remaining results can be found in the Online Appendix.

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. In situations such as this, intervals that allow for weak identification are essential to protect the econometrician from mischaracterizing the degree of stickiness in the economy, as illustrated in Table 4. There is an active literature on measuring the degree of price rigidity at the microlevel (see, e.g., Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008)). For example, Klenow and Kryvtsov (2008) found that price contracts last, on average, about 2.3 quarters. Based on the percentile intervals in Table 4(a), a researcher would have concluded that the length of those price spells is incompatible with the macroevidence. The lower bound of the percentile interval corresponds to a price spell of 2.6 quarters, which is inconsistent with the microevidence at the 10% significance level. In contrast,

TABLE 4. The 90% intervals for rigidity parameters in the medium-scale New Keynesian model.

	Posterior				Percentile	BF	LR
	Means	Medians	Modes	SD	Interval	Interval	Interval
				(a) Agnostic Priors			
ζ_p	0.692	0.693	0.695	0.047	[0.611, 0.767]	[0.512, 0.833]	
ζ_w	0.222	0.217	0.164	0.073	[0.113, 0.350]	[0.035, 0.504]	
				(b) Low-Rigidity Priors			
ζ_p	0.659	0.661	0.695	0.045	[0.581, 0.729]	[0.470, 0.796]	
ζ_w	0.266	0.264	0.269	0.057	[0.177, 0.364]	[0.080, 0.515]	
				(c) High-Rigidity Priors			
ζ_p	0.772	0.770	0.786	0.058	[0.676, 0.855]	[0.626, 0.887]	
ζ_w	0.446	0.428	0.391	0.114	[0.286, 0.647]	[0.239, 0.714]	
				(d) LR Method			
ζ_p							[0.543, 0.854]
ζ_w							[0.017, 0.470]

a researcher relying on the BF interval in Table 4(a) would have viewed 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 2.0 quarters.

Table 4(b) and (c) provides additional evidence on the sensitivity of this conclusion to the choice of prior. We replace the agnostic prior by the low-rigidity and high-rigidity priors explored in Del Negro and Schorfheide (2008). The BF interval suggests that Klenow and Kryvstov's findings are plausible under priors that favor low price rigidity (Table 4(b)), but no longer plausible under a prior that imposes substantial price rigidity (Table 4(c)). In the latter case, the lower bound on the length of the price spell is 2.7, which is higher than the microdata estimate of 2.3. This example illustrates once again that in small samples the BF interval is not invariant to the choice of prior. This fact suggests caution in interpreting the BF results.

It is useful to contrast these results with those for the LR interval, which does not depend on any priors. The LR interval in Table 4(d) implies a lower bound on the length of a price spell of 2.2 quarters, which is consistent with the microevidence. We conclude that the macroevidence on price rigidities is compatible with the microevidence, contrary to what the percentile interval would have suggested under any of the three priors we examined. Regarding the wage rigidity parameter, Table 4(d) shows that there is not much divergence in the BF and LR estimates on the lower bound of the length of a wage spell (with all estimates between 1 and 1.3 quarters), but there is greater dispersion in the estimates of the upper bound (which range from 1.5 to 3.5 quarters). The LR method implies that wage spells shorter than 1 quarter and longer than 1.9 quarters can be rejected at the 10% level. Finally, all BF and LR methods agree that there is much higher price rigidity than wage rigidity in the U.S. economy, but these differences are most pronounced when using the LR method.

5. Concluding remarks

This paper made five distinct contributions. First, we illustrated that the usual asymptotic equivalence between frequentist and Bayesian methods of estimation and inference breaks down if the structural parameters of the DSGE model are only weakly identified. This fact invalidates the interpretation of posterior modes, medians, or means as point estimates and invalidates frequentist confidence sets constructed from the posterior. Second, we proposed two alternative frequentist confidence sets that remain asymptotically valid regardless of the strength of identification. One is constructed by inverting an LR statistic; the other involves inverting the Bayes factor statistic. Third, we contrasted the relative merits of these statistics from a theoretical point of view. In particular, we showed that the LR test statistic has power against local alternatives, whereas the BF test statistic does not. Fourth, we provided simulation evidence that both LR and BF intervals tend to be more accurate than pseudo-Bayesian confidence intervals, which often have very poor coverage accuracy. Whereas the LR method produces reasonably accurate confidence sets for realistic sample sizes, we found that the accuracy of the corresponding BF confidence sets can be sensitive to the choice of the prior. Finally, we

provided an empirical example that shows that weak identification does not necessarily mean that there is no information in the data about the structural parameters of interest.

Appendix: Proofs

The proofs of Propositions 1, 2, 3b, and 4 are omitted to conserve space. The remaining proofs are sketched below. The full proofs of all propositions are available in the Online Appendix.

Proof of Theorem 1. It follows from assumption (a) in Proposition 2 and assumptions (a), (b), and (c) of Theorem 1, the Taylor theorem, the first-order condition for the unconstrained MLE, and the result in Proposition 2 that $[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})]^{1/2}(\tilde{\gamma}_T(\theta_0) - \hat{\gamma}_T) = P_T z_T + o_p(1)$, that

$$\begin{aligned} I_{3,T}i &\equiv \int_{B_{\delta_T}(\theta_0)} \pi(\theta)\exp\bigl(\ell_T\bigl(\tilde{\gamma}_T(\theta)\bigr) - \ell_T(\hat{\gamma}_T)\bigr)\,d\theta \\ &= \int_{B_{\delta_T}(\theta_0)} \pi(\theta)\,d\theta \exp\left(-\frac{1}{2}z_T' z_T\right) + o_p(1), \end{aligned} \tag{30}$$

where $\bar{\gamma}_T(\theta)$ is a point between $\tilde{\gamma}_T(\theta)$ and $\hat{\gamma}_T$. Let

$$I_{4,T} = \int_{\Theta\setminus B_{\delta_T}(\theta_0)} \pi(\theta)\exp\bigl(\ell_T\bigl(\tilde{\gamma}_T(\theta)\bigr) - \ell_T(\hat{\gamma}_T)\bigr)\,d\theta. \tag{31}$$

Since $\ell_T(\tilde{\gamma}_T(\theta)) \le \ell_T(\hat{\gamma}_T)$ by the definition of MLE, it follows from (31) that

$$I_{4,T} \le \int_{\Theta\setminus B_{\delta_T}(\theta_0)} \pi(\theta)\,d\theta. \tag{32}$$

Combining these results, the Bayes factor in favor of H_1 can be written as

$$\text{Bayes factor}(\theta_0) = \frac{\displaystyle\int_{B_{\delta_T}(\theta_0)} \pi(\theta)\,d\theta}{\displaystyle\int_{\Theta\setminus B_{\delta_T}(\theta_0)} \pi(\theta)\,d\theta}\frac{I_{4,T}}{I_{3,T}} \le \exp\left(\frac{1}{2}z_T' z_T\right) + o_p(1), \tag{33}$$

where the inequality follows from (32). □

Proof of Proposition 3a. By assumption (a) in Proposition 2, the constrained MLE $\tilde{\gamma}_T(\alpha_0)$ satisfies the first-order conditions

$$\nabla\ell_T\bigl(\tilde{\gamma}_T(\theta_0)\bigr) + D_\gamma f\bigl(\tilde{\gamma}_T(\alpha_0), \alpha_0, \tilde{\beta}(\alpha_0)\bigr)'\tilde{\lambda}_T(\alpha_0) = 0_{\dim(\gamma)\times 1}, \tag{34}$$

$$D_\beta f\bigl(\tilde{\gamma}_T(\alpha_0), \alpha_0, \tilde{\beta}_T(\alpha_0)\bigr) = 0_{k_2\times 1}, \tag{35}$$

$$f\bigl(\tilde{\gamma}_T(\alpha_0), \alpha_0, \tilde{\beta}_T(\alpha_0)\bigr) = 0_{r\times 1}, \tag{36}$$

where $\tilde{\lambda}_T(\alpha_0)$ is the $r \times 1$ vector of Lagrange multipliers. A Taylor series expansion of these first-order conditions about $[\gamma_{0,T}' \ \beta_0' \ 0_{1\times r}]'$ yields

$$\begin{bmatrix} \nabla_{\gamma\gamma}\ell_T(\gamma_{0,T}) & 0_{\dim(\gamma)\times k_2} & D_\gamma f(\gamma_{0,T},\theta_0)' \\ 0_{k_2\times\dim(\gamma)} & 0_{k_2\times k_2} & D_\beta f(\gamma_{0,T},\theta_0)' \\ D_\gamma f(\gamma_{0,T},\theta_0) & D_\beta f(\gamma_{0,T},\theta_0) & 0_{r\times r} \end{bmatrix} \begin{bmatrix} \tilde{\gamma}_T(\alpha_0)-\gamma_{0,T} \\ \tilde{\beta}_T(\alpha_0)-\beta_0 \\ \tilde{\lambda}_T(\alpha_0) \end{bmatrix} = \begin{bmatrix} -\nabla\ell_T(\gamma_{0,T}) \\ 0_{k_2\times 1} \\ 0_{r\times 1} \end{bmatrix} + o_p\big(T^{-1/2}\big).$$

After solving these equations and some further manipulations, we obtain

$$\begin{aligned} &\tilde{\gamma}_T(\alpha_0)-\hat{\gamma}_T \\ &\quad = H_T^{-1}R'\big(RH_T^{-1}R'\big)^{-1}RH_T^{-1}\nabla\ell_T(\gamma_{0,T}) \\ &\qquad - H_T^{-1}R'\big(RH_T^{-1}R'\big)^{-1}R_\beta\big[R_\beta'\big(RH_T^{-1}R'\big)^{-1}R_\beta\big]^{-1} \\ &\qquad \times R_\beta'\big(RH_T^{-1}R'\big)^{-1}RH_T^{-1}\nabla_\gamma\ell_T(\gamma_{0,T}) \\ &\qquad + o_p(1), \end{aligned} \tag{37}$$

where $H_T = \nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})$ and $R_\beta = D_\beta f_T(\gamma_{0,T},\theta_0)$. Hence, we can write

$$\begin{aligned} \mathrm{LR}_T(\alpha_0) &= \big(\tilde{\gamma}_T(\alpha_0)-\hat{\gamma}_T\big)'\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]\big(\tilde{\gamma}_T(\alpha_0)-\hat{\gamma}_T\big)+o_p(1) \\ &= z_T'Q_Tz_T+o_p(1). \end{aligned} \tag{38}$$

Because Q_T is idempotent and has rank $r-k_2$, we obtain the desired result. □

Proof of Theorem 2. First we prove (14). An application of the implicit function theorem to $f_T(\gamma,\alpha)=0$ yields

$$\frac{\partial\gamma}{\partial\alpha'} = -T^{-1/2}\big[D_\gamma f_1(\gamma,\beta)\big]^{-1}D_\alpha f_2(\gamma,\theta)+o\big(T^{-1/2}\big), \tag{39}$$

$$\frac{\partial\gamma}{\partial\beta'} = -\big[D_\gamma f_T(\gamma,\theta)\big]^{-1}D_\beta f_T(\gamma,\theta). \tag{40}$$

Thus, the mean value theorem implies

$$\begin{aligned} \gamma_{1,T}-\gamma_{0,T} = &-T^{-1/2}\big[D_\gamma f_1(\bar{\gamma}_T,\bar{\beta}_T)\big]^{-1}D_\beta f_1(\bar{\gamma}_T,\bar{\beta}_T)c \\ &- T^{-1/2}\big[D_\gamma f_1(\bar{\gamma}_T,\beta_1)\big]^{-1}D_\alpha f_2(\bar{\gamma}_T,\bar{\alpha})(\alpha_1-\alpha_0) \\ &+ o\big(T^{-1/2}\big), \end{aligned} \tag{41}$$

where $[\bar{\gamma}_T' \ \bar{\theta}_T']' = [\bar{\gamma}_T' \ \bar{\alpha}_T' \ \bar{\beta}_T']'$ is a point between $[\gamma_{1,T}' \ \alpha_1' \ \beta_0' + T^{-1/2}c']'$ and $[\gamma_{0,T}' \ \alpha_0' \ \beta_0']'$. Because $\gamma_{1,T}-\gamma_{0,T} = O(T^{-1/2})$, and f_1 and f_2 are continuously differ-

entiable, we can write (41) as

$$\gamma_{1,T} - \gamma_{0,T} = T^{-1/2} G(\alpha_0, \alpha_1, \beta_0)\big[\, c' \quad (\alpha_1 - \alpha_0)' \,\big]' + o\big(T^{-1/2}\big), \tag{42}$$

where $G(\alpha_0, \alpha_1, \beta_1) = \lim_{T\to\infty}\{[D_\gamma f_1(\gamma_{1,T}, \beta_0)]^{-1}[D_\beta f_1(\gamma_{1,T}, \beta_0) \;\; D_\alpha f_2(\gamma_{1,T}, \bar{\alpha}_T)]\}$.

Using (42) and arguments analogous to those in the proof of Theorem 1, we can show $\hat{\gamma}_T - \gamma_{1,T} = -[\nabla_{\gamma\gamma}\ell_T(\gamma_{1,T})]^{-1}\nabla_\gamma \ell_T(\gamma_{1,T}) + o_p(T^{-1/2})$ and

$$\begin{aligned}
&\tilde{\gamma}_T(\theta_0) - \gamma_{0,T} \\
&\quad = \big\{-\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1} \\
&\qquad + \big[\nabla_{\gamma\gamma}\ell_T(\gamma_0)\big]^{-1} R_T'\big\{R_T\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1} R_T'\big\}^{-1} R_T\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T})\big]^{-1}\big\} \\
&\qquad \times \big\{\nabla_\gamma \ell_T(\gamma_{1,T}) + T^{-1/2}\nabla_{\gamma\gamma}\ell_T(\gamma_{0,T}) G(\alpha_0, \alpha_1, \beta_0)\big[\, c' \quad (\alpha_1 - \alpha_0)' \,\big]'\big\} \\
&\qquad + o_p\big(T^{-1/2}\big).
\end{aligned} \tag{43}$$

Let $R = \lim_{T\to\infty} R_T$, $d(\alpha, \alpha_1, \beta_0, c) = V_\gamma^{-1/2} G(\alpha, \alpha_1, \beta_0)[c' \;\; (\alpha_1 - \alpha)']'$, and $P = V_\gamma^{1/2} \times R'(RV_\gamma R')^{-1} RV_\gamma^{1/2}$. It follows from (42), (43), the twice continuous differentiability of $\ell_T(\cdot)$, and assumption (a) in Theorem 2 that

$$\big[\nabla_{\gamma\gamma}\ell_T(\gamma_{1,T})\big]^{1/2}\big(\tilde{\gamma}_T(\theta_0) - \hat{\gamma}_T\big) = P\big[d(\alpha_0, \alpha_1, \beta_0, c) + z_T\big] + o_p(1). \tag{44}$$

Because P is idempotent and has rank r, (A.1) and a second-order Taylor series expansion of the LR test statistic around $\hat{\gamma}_T$ yield (14).

Next we prove (15). Define

$$I_{j,T} = T^{k_2/2} \int_{\Theta_{j,T}} \pi(\theta) \exp\big(\ell_T\big(\tilde{\gamma}_T(\theta)\big) - \ell_T(\hat{\gamma}_T)\big)\, d\theta \tag{45}$$

for $j = 5, 6, 7$, where $G_T(\gamma_{1,T}, \theta_{1,T}) = -[D_\gamma f_T(\gamma_{1,T}, \theta_{1,T})]^{-1} D_\theta(\gamma_{1,T}, \theta_{1,T})$,

$$\begin{aligned}
\Theta_{5,T} &= \big\{\theta \in \Theta : |\theta_j - \theta_{0,j}| < \delta_{T,j} \text{ for } j = 1, \ldots, k\big\}, \\
\Theta_{6,T} &= \big\{\theta \in \Theta : |\theta_j - \theta_{0,j}| \geq \delta_{T,j}, \beta = \beta_0 + \bar{c}T^{-1/2} \text{ for some } \bar{c} \in \bar{C}_T\big\}, \\
\Theta_{7,T} &= \big\{\theta \in \Theta : |\theta_j - \theta_{0,j}| \geq \delta_{T,j}, \neg\exists \bar{c} \in \bar{C}_T \text{ such that } \beta = \beta_0 + \bar{c}_T T^{-1/2}\big\},
\end{aligned}$$

$\bar{C}_T = \{\bar{c} \in \Re^{k_2} : -c_{\min}T^\eta \leq c \leq c_{\max}T^\eta\}$, $c_{\max} > 0$, $c_{\min} > 0$, and $\eta \in (0, 1/2)$. Define

$$\begin{aligned}
\bar{\Theta}_{5,T} &= \big\{\big[\, \alpha' \quad c' \,\big] \in \mathcal{A} \times \Re^{k_2} : |\alpha_j - \alpha_{0,j}| < \delta_{T,j} \text{ for } j = 1, \ldots, k_1, \\
&\qquad |c_j| < \delta_{T,j+k_1} T^{1/2} \text{ for } j = 1, \ldots, k_2\big\}, \\
\bar{\Theta}_{6,T} &= \big\{\theta \in \Theta : |\theta_j - \theta_{0,j}| \geq \delta_{T,j}, \beta = \beta_0 + \bar{c}T^{-1/2} \text{ for some } \bar{c} \in \bar{C}_T\big\}, \\
\bar{\Theta}_{7,T} &= \big\{\theta \in \Theta : |\theta_j - \theta_{0,j}| \geq \delta_{T,j}, \neg\exists \bar{c} \in \bar{C}_T \text{ such that } \beta = \beta_0 + \bar{c}_T T^{-1/2}\big\}.
\end{aligned}$$

It follows from assumptions (a) and (b) in Proposition 2, Taylor's theorem, a change of variables, and (44), that $I_{5,T}$ equals

$$\begin{aligned} & T^{k_2/2}\int_{\Theta_{5,T}} \pi(\theta)\exp\left(\frac{1}{2}\big(\tilde{\gamma}_T(\theta)-\hat{\gamma}_T\big)'\nabla_{\gamma\gamma}\ell_T(\bar{\gamma}_T)\big(\tilde{\gamma}_T(\theta)-\hat{\gamma}_T\big)\right)d\theta \\ &= \int_{\bar{\Theta}_{5,T}} \pi\big(\big[\alpha' \quad c'\big]'\big)d\big[\alpha' \quad c'\big]' \\ &\quad \times \exp\left\{-\frac{1}{2}\big[d(\alpha_0,\alpha_1,\beta_0,0)+z_T\big]'P\big[d(\alpha_0,\alpha_1,\beta_0,0)+z_T\big]\right\}+o_p(1), \end{aligned} \tag{46}$$

where $\bar{\gamma}_T$ is a point between $\hat{\gamma}_T$ and $\tilde{\gamma}_T(\theta)$.

Note that the arguments used to derive (44) are valid not only for a particular value of α_0 and c, but also for all α and c. The compactness of α, the continuous differentiability of f_T, and the continuity of $\nabla_{\gamma\gamma}\ell_T$ imply that (43) holds uniformly in $\alpha\in\mathcal{A}$, which in turn implies that (44) holds uniformly in $\alpha\in\mathcal{A}$. Thus, $I_{6,T}$ equals

$$T^{k_2/2}\int_{\Theta_{6,T}} \pi(\theta)\exp\left(\frac{1}{2}\big(\tilde{\gamma}_T(\theta)-\hat{\gamma}_T\big)'\nabla_{\gamma\gamma}\ell_T(\bar{\gamma}_T)\big(\tilde{\gamma}_T(\theta)-\hat{\gamma}_T\big)\right)d\theta \tag{47}$$

$$\begin{aligned} &= \int_{\mathcal{A}\times\bar{C}_T} \pi\big(\big[\alpha',\beta_0'\big]'\big) \\ &\quad \times \exp\left\{\frac{1}{2}z_T'\big[P-PV_\gamma^{-1/2}G(\alpha,\alpha_1,\beta_0)\big(G(\alpha,\alpha_1,\beta_0)' \right. \\ &\quad \left. \times V_\gamma^{-1/2}PV_\gamma^{-1/2}G(\alpha,\alpha_1,\beta_0)\big)^{-1}G(\alpha,\alpha_1,\beta_0)'V_\gamma^{-1/2}P\big]z_T\right\}d\alpha \\ &\quad +o_p(1). \end{aligned} \tag{48}$$

Moreover,

$$I_{7,T}=T^{k_2/2}\int_{\Theta_{7,T}}\pi(\theta)\exp\big(\ell_T\big(\tilde{\gamma}_T(\theta)\big)-\ell_T(\hat{\gamma}_T)\big)\,d\theta=o_p(1), \tag{49}$$

where the last equality follows from assumption (c) in Theorem 2.

Because $\int_{\bar{\Theta}_{5,T}}\pi([\alpha',\beta_0'+T^{-1/2}c']')\,d[\alpha'\ \ c']'=\int_{\Theta_{5,T}}\pi([\alpha',\beta']')\,d[\alpha',\beta']'$, it follows from (46), (48), and (49) that the Bayes factor in favor of H_1 can be written as

$$\begin{aligned} & T^{k_2/2}\,\text{Bayes factor}(\theta_0) \\ &= T^{k_2/2}\frac{\displaystyle\int_{B_{\delta_T}(\theta_0)}\pi(\theta)\,d\theta}{\displaystyle\int_{\Theta\setminus B_{\delta_T}(\theta_0)}\pi(\theta)\,d\theta}\frac{I_{6,T}+I_{7,T}}{I_{5,T}} \\ &= \int_{\mathcal{A}\times\bar{C}_T}\pi(\alpha,\beta_0) \end{aligned}$$

$$\times \exp\biggl(-\frac{1}{2}\bigl(d(\alpha,\alpha_1,\beta_0,c)+Pz_T\bigr)'\bigl(d(\alpha,\alpha_1,\beta_0,c)+Pz_T\bigr)\biggr)\,d\begin{bmatrix}\alpha' & c'\end{bmatrix}'$$

$$\Big/ \exp\biggl(-\frac{1}{2}\bigl(d(\alpha_0,\alpha_1,\beta_0,0)+Pz_T\bigr)'\bigl(d(\alpha_0,\alpha_1,\beta_0,0)+Pz_T\bigr)\biggr)$$

$$+ o_p(1),$$

which completes the proof of (15). □

References

An, S. and F. Schorfheide (2007), "Bayesian analysis of DSGE models." *Econometric Reviews*, 26, 113–172. [214]

Andrews, D. W. K. (1994), "The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests." *Econometrica*, 62, 1207–1232. [198]

Andrews, D. W. K. and X. Cheng (2012), "Estimation and inference with weak, semi-strong, and strong identification." *Econometrica*, 80, 2153–2211. [199, 203, 205]

Andrews, I. and A. Mikusheva (2012), "Maximum likelihood inference in weakly identified models." Report, MIT. [199, 204]

Antoine, B. and E. Renault (2009), "Efficient GMM with nearly-weak identification." *The Econometrics Journal*, 12, 135–171. [203]

Balke, N. S., S. P. A. Brown, and M. K. Yücel (2008), "Globalization, oil price shocks and U.S. economic activity." Unpublished manuscript, Federal Reserve Bank of Dallas. [216]

Berger, J. O. and T. Sellke (1987), "Testing a point null hypothesis: Irreconcilability of p values and evidence." *Journal of the American Statistical Association*, 82, 112–139. [207]

Bickel, P. J. and K. Doksum (2006), *Mathematical Statistics: Basic Concepts and Selected Ideas*, Vol. I, second edition. Prentice Hall, Upper Saddle River, NJ. [200]

Canova, F. and L. Sala (2009), "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56, 431–449. [198, 213]

Chaudhuri, S. and E. Zivot (2011), "A new method of projection-based inference in GMM with weakly identified nuisance parameters." *Journal of Econometrics*, 164, 239–251. [212]

Chernozhukov, V. and H. Hong (2003), "An MCMC approach to classical estimation." *Journal of Econometrics*, 115, 293–346. [198]

Christiano, L., M. Eichenbaum, and C. Evans (2005), "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of Political Economy*, 113, 1–45. [219]

Del Negro, M. and F. Schorfheide (2008), "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)." *Journal of Monetary Economics*, 55, 1191–1208. [198, 200, 219, 220, 222]

Dufour, J.-M., L. Khalaf, and M. Kichian (2009), "Structural multi-equation macroeconomic models: Identification-robust estimation and fit." Working Paper 2009-19, Bank of Canada. [199]

Dufour, J.-M. and M. Taamouti (2005), "Projection-based statistical inference in linear structural models with possibly weak instruments." *Econometrica*, 73, 1351–1365. [212]

Edwards, W., H. Lindman, and L. J. Savage (1963), "Bayesian statistical inference for psychological research." *Psychological Review*, 70, 193–242. [207]

Fernández-Villaverde, J., P. A. Guerron-Quintana, and J. F. Rubio-Ramírez (2010), "The new macroeconometrics: A Bayesian approach." In *Handbook of Applied Bayesian Analysis*, 366–399, Oxford University Press, Oxford. [220]

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007), "*A*, *B*, *C*'s (and *D*'s) for understanding VARs." *American Economic Review*, 97, 1021–1026. [203]

Fukač, M., D. F. Waggoner, and T. Zha (2007), "Local and global identification of DSGE models: A simultaneous-equation approach." Unpublished manuscript, Reserve Bank of New Zealand and Federal Reserve Bank of Atlanta. [199]

Gelfand, A. and A. Smith (1990), "Sampling based approaches to calculating marginal densities." *Journal of American Statistical Association*, 85, 398–409. [215]

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, second edition. Chapman & Hall/CRC, Boca Raton, FL. [220]

Glover, K. and J. C. Willems (1974), "Parameterizations of linear dynamical systems: Canonical forms and identifiability." *IEEE Transactions on Automatic Control*, AC-19, 640–645. [212]

Iskrev, N. (2010), "Local identification in DSGE models." *Journal of Monetary Economics*, 57, 189–202. [198]

Kadane, J. B. (1975), "The role of identification in Bayesian theory." In *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage* (S. E. Fienberg and A. Zellner, eds.), 175–191, North-Holland, Amsterdam. [202]

Kim, J.-Y. (1998), "Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models." *Econometrica*, 66, 359–380. [198]

Kleibergen, F. and S. Mavroeidis (2009), "Weak instrument robust tests in GMM and the new Keynesian Phillips curve." *Journal of Business and Economic Statistics*, 27, 293–311. [199]

Klenow, P. J. and O. Kryvtsov (2008), "State-dependent or time-dependent pricing: Does it matter for recent U.S. inflation?" *Quarterly Journal of Economics*, 123, 863–904. [221]

Komunjer, I. and S. Ng (2009), "Dynamic identification of DSGE models." Unpublished manuscript, University of California, San Diego and Columbia University. [199, 203, 211, 212]

Koop, G., M. H. Pesaran, and R. P. Smith (2011), "On identification of Bayesian DSGE models." Unpublished manuscript, Cambridge University. [198]

Le Cam, L. and G. L. Yang (2000), *Asymptotics in Statistics: Some Basic Concepts*, second edition. Springer-Verlag, New York. [198, 200]

Mikusheva, A. (2007), "Uniform inference in autoregressive models." *Econometrica*, 75, 1411–1452. [205]

Moon, H. R. and F. Schorfheide (2012), "Bayesian and frequentist inference in partially identified models." *Econometrica*, 80 (2012), 755–782. [198]

Nakamura, E. and J. Steinsson (2008), "Five facts about prices: A reevaluation of menu cost models." *Quarterly Journal of Economics*, 123, 1415–1464. [221]

Phillips, P. C. B. and W. Ploberger (1996), "An asymptotic theory of Bayesian inference for time series." *Econometrica*, 64, 381–412. [198]

Poirier, D. J. (1998), "Revising beliefs in nonidentified models." *Econometric Theory*, 14, 483–509. [202]

Qu, Z. (2011), "Inference and specification testing in DSGE models with possible weak identification." Report, Boston University. [199, 204]

Roberts, G., A. Gelman, and W. Gilks (1997), "Weak convergence and optimal scaling of random walk Metropolis algorithms." *The Annals of Applied Probability*, 7, 110–120. [215]

Schorfheide, F. (2011), "Estimation and evaluation of DSGE models: Progress and challenges." Working Paper 16781, NBER. [203]

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York. [213]

Smets, F. and R. Wouters (2007), "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American Economic Review*, 97, 586–606. [198, 219]

Stock, J. H. and M. Watson (2002), "Has the business cycle changed and why?" *NBER Macroeconomics Annual*, 2002, 159–218. [214]

Stock, J. H. and J. H. Wright (2000), "GMM with weak identification." *Econometrica*, 68, 1055–1096. [199, 203]

Woodford, M. M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ. [213, 214]

Submitted September, 2012. Final version accepted October, 2012.