On Identification of Bayesian DSGE Models

Gary Koop, University of Strathclyde
M. Hashem Pesaran, University of Southern California and Trinity College, Cambridge
Ron P. Smith, Birkbeck College

August 6, 2012

Abstract

This paper is concerned with identification of dynamic stochastic general equilibrium (DSGE) models from a Bayesian perspective, and proposes two Bayesian indicators. The first indicator follows a suggestion by Poirier of comparing the posterior density of the parameter of interest with the posterior expectation of its prior conditional on the remaining parameters, as opposed to comparing the posterior distribution to its prior as is usually done. The second indicator examines the rate at which the posterior precision of the parameter gets updated with the sample size, using simulated data. For identified parameters the posterior precision increases at rate T. We show that for parameters that are either unidentified or are weakly identified the posterior precision may be updated but its rate of update will be slower than T. We use empirical examples to demonstrate that these methods are useful in practice.

JEL Classifications: C11, C15, E17

Key Words: Bayesian identification, weak identification, DSGE models, posterior updating.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are typically estimated by Bayesian methods (DeJong, Ingram and Whiteman, 2000, Smets and Wouters, 2003, 2007, and An and Schorfheide, 2007). The application of Bayesian techniques to DSGE models, made possible through the use of Markov Chain Monte Carlo (MCMC) algorithms presents a natural progression from the earlier calibrated DSGE models popularized by Kydland and Prescott (1996). The Bayesian approach also seems to conveniently circumvent the problems often encountered when estimating DSGE models by maximum likelihood, where the likelihood function in terms of the structural parameters of DSGE models was often found to be badly behaved. Although, solutions of DSGE models are usually in the forms of linear VAR or VARMA models, the structural parameters are non-linear functions of the solution parameters, often involving complicated cross-equation restrictions. Unlike the simpler simultaneous equations model (SEM) the non-linear nature of the cross equation restrictions in DSGE models and the role of unobserved shocks makes it more difficult to analytically check identification or to judge identification from the information matrix or curvature of the likelihood function.

The early literature on identification of rational expectations (RE) models started with the work of Sargent (1976) and McCallum (1979) on observational equivalence and was extended to more general set ups by Wallis (1980), Pesaran (1981, 1987), and Pudney (1982). Interest in identification of RE models then waned as models tended to be calibrated rather than estimated. But recent interest in estimation of DSGE models has prompted a return to the problem of identification of RE models in general and that of DSGE models in particular. The issues in identification of the new Keynesian Phillips curve has been discussed by Mavroeidis, (2005), Nason and Smith, (2008), Kleibergen and Mavroeidis, (2009), Dees et al., (2009); and of the Taylor rule by Cochrane (2011). More generally Canova and Sala (2009) conclude "it appears that a large class of popular DSGE structures are only very weakly identified" and Iskrev (2010b) concludes "the results indicate that the parameters of the Smets and Wouters (2007) model are

quite poorly identified in most of the parameter space". Other recent papers which consider determining the identification of DSGE models are Andrle (2010), Iskrev (2010a), Iskrev and Ratto (2010), and Komunjer and Ng (2011).

Whereas papers like Iskrev (2010a,b) and Komunjer and Ng (2011) provide classical or frequentist procedures for determining identification based on the rank of particular matrices, our objective is to provide Bayesian indicators of the identification of one or more parameters. This seems useful both because these models are usually estimated by Bayesian methods and because the issues raised by identification are rather different in a Bayesian context.

Given an informative prior, such that a well behaved marginal prior exists for the parameter of interest, then there is a well defined posterior distribution, whether or not the parameter is identified. Thus an individual can like Lindley, in one of the classic Bayesian texts, conclude "that unidentifiability causes no real difficulty in the Bayesian approach" (Lindley, 1971, page 46 footnote 34). However, for social learning, where different people may have different priors, sensitivity to the choice of priors is an important issue. If the parameter is not identified, one cannot learn about the parameter directly from the data and even with an infinite sample of data the posterior would be determined by the priors.

Within a Bayesian context, learning is interpreted as a changing posterior distribution, and a common practice in DSGE estimation is to judge identification by a comparison of the prior and posterior distributions for a parameter. As we discuss, not only can the posterior distribution differ from the prior even when the parameter is unidentified, but in addition a changing posterior, apparent learning, need not be informative about identification. This can happen because, for instance, the requirement for a determinate solution of a DSGE model puts restrictions on the joint parameter space, which may create dependence between identified and unidentified parameters, even if their priors are independent. What proves to be informative in a Bayesian context is the rate at which learning takes place (posterior precision increases) as more data becomes available.

This paper suggests two Bayesian indicators of identification. The first, like the classical

procedures, indicates non-identification while the second, which is likely to be more useful in practice, indicates either non-identification or weak identification. Like most of the literature our analysis is local in the sense that we investigate identification at a given point in the feasible parameter space. Although our indicators can be applied to any point in the parameter space, in the Bayesian context prior means seem a natural start point. If the parameters are identified at their prior means then other points could be investigated. Also, whilst we focus on DSGE models, our analysis is more generally applicable.

The first indicator, to be referred to as the 'Bayesian comparison indicator', is based on Proposition 2 of Poirier (1998) and considers identification of the $p_1 \times 1$ vector of parameters, θ_1, assuming that the remaining $p_2 \times 1$ vector of parameters, θ_2, is identified. It compares the posterior distribution of θ_1 with the posterior expectation of its prior distribution conditional on θ_2, and concludes that θ_1 is unidentified if the two distributions coincide. This contrasts to the direct comparison of the prior of θ_1 with its posterior. These could differ from one another even if θ_1 is unidentified. Like the classical indicators based on the rank of a matrix, the outcome of this Bayesian indicator is a yes/no answer, but in practice will depend on the numerical accuracy of the MCMC procedures used to compute the posterior distributions. The MCMC approximation error, however, can be made as small as required, unlike the classical procedures which require determining whether a particular numerical eigenvalue is zero. Komunjer and Ng (2011) note that the magnitude of the eigenvalues depends on the units of measurement and discuss the choice of numerical cut-off used to determine whether the eigenvalues are sufficiently small.

Despite its generality, the application of the Bayesian comparison indicator to DSGE models can be problematic, since it is often difficult to suitably partition the parameters of the model such that there exists a sub-set which is known to be identified. Furthermore, in many applications the main empirical issue of interest is not a yes/no response to an identification question, but if one or more parameters of the model are weakly identified, in the sense discussed, for example, by Stock, Wright and Yogo (2002), and Andrews and Cheng (2011), in the classical literature. Accordingly,

we also propose a second indicator, which we refer to as the 'Bayesian learning rate indicator', that examines the rate at which the posterior precision of a given parameter gets updated with the sample size, T, using simulated data. For identified parameters the posterior precision increases at the rate T. But for parameters that are either not identified or weakly identified the posterior precision may be updated but its rate of update will be slower than T. Notice that this procedure simulates sample of increasing size and does not require the size of the available realized data to be large. To our knowledge this paper is the first to present a Bayesian treatment of weak identification.

Empirical illustrations show the usefulness of both of these indicators for checking the presence and strength of identification in Bayesian contexts. They are easy to implement in standard software packages such as Dynare (http://www.dynare.org/). In a recent paper Caglar et al (2012) applies the learning rate indicator to examine the identification of the parameters of the Bayesian DSGE model of Smets and Wouters (2007), and find that many parameters of this widely used model do not appear to be well identified.

The paper is organized as follows. Section 2 discusses the identification issues that arise in linear rational expectations DSGE models, and provides two simple examples that will be used for illustration. General issues in Bayesian identification are considered in Section 3, whilst Section 4 considers issues that are more specific to Bayesian identification of DSGE models. The two Bayesian indicators of identification advanced in the paper are outlined in sub-sections 4.1 and 4.2. The main theoretical results are summarized in two propositions with proofs provided in an Appendix. Section 5 presents several empirical illustrations of the methods developed and discussed in Sections 2 and 4. Section 6 concludes.

2 Rational Expectations DSGE Models

2.1 A General Framework

Most macroeconomic DSGE models are constructed by linearizing an underlying non-linear rational expectations, RE, model around its steady state. A typical log-linearized RE model can be

written as

$$A_0(\theta)y_t = A_1(\theta)E_t(y_{t+1}) + A_2(\theta)y_{t-1} + A_3(\theta)x_t + u_t, \tag{1}$$

$$x_t = \Phi_x x_{t-1} + v_t,\ u_t = \Phi_u u_{t-1} + \varepsilon_t,$$

where y_t is an $n \times 1$ vector of deviations from the steady states, x_t is an $m \times 1$ vector of observed exogenous variables, u_t is an $n \times 1$ vector of unobserved variables, and ε_t is the $n \times 1$ vector of structural shocks, assumed to be serially uncorrelated with zero mean, the covariance matrix, $E(\varepsilon_t \varepsilon_t') = \Omega(\theta)$. For Bayesian or maximum likelihood estimation, ε_t is also typically assumed to be normally distributed. The expectations $E_t(y_{t+1}) = E(y_{t+1} \mid \Im_t)$ are assumed to be rationally formed with respect to the information set, $\Im_t = (y_t, x_t, y_{t-1}, x_{t-1}, ...)$. To simplify the exposition it is assumed that both the exogenous and unobserved variables follow VAR(1) processes. The parameters of interest are the $p \times 1$ vector of structural parameters, θ, and the remaining (reduced form) parameters Φ_x and Φ_u are assumed as given. It is also assumed that there are no feedbacks from y_t to x_t or u_t. To identify the structural shocks it is common in the literature to assume that $\Omega(\theta) = I_n$.

If $A_0(\theta)$ is nonsingular, then (1) can be written

$$y_t = A_0(\theta)^{-1}A_1(\theta)E_t(y_{t+1}) + A_0(\theta)^{-1}A_2(\theta)y_{t-1} \tag{2}$$

$$+ A_0(\theta)^{-1}A_3(\theta)x_t + A_0(\theta)^{-1}u_t.$$

The solution of such a system is discussed in Binder and Pesaran (1995, 1997) and Sims (2002) and if the unique stationary solution exists, it takes the form:

$$y_t = C(\theta)y_{t-1} + G_1(\theta, \phi_x)x_t + G_2(\theta, \phi_u)u_t, \tag{3}$$

where $\phi_i = vec(\Phi_i)$, $i = x, u$. The matrices $G_i(\theta, \phi_i)$ $i = x, u$, can be obtained using the method of undetermined coefficients (see Blinder and Pesaran, 1997, for details). Notice that the coefficient matrix for the lagged dependent variable vector is just a function of θ, and not ϕ_x or ϕ_u.

If $\Phi_u = 0$, this is just a VAR with exogenous variables and the likelihood function for the reduced form parameters is easily obtained. In the general case where the unobserved components of the model are serially correlated, the rational expectations solution will involve moving average components and it is more convenient to write the model as a state space model where Kalman filtering techniques can be used to evaluate the likelihood function. In such cases the reduced form parameters may not be identified. Below we use some simple special cases of DSGE models where the RE solution is available analytically to demonstrate the identification issues.

2.2 DSGE models without lags

Abstracting from lagged values and exogenous regressors and for notational simplicity, not making the dependence on θ explicit, (1) simplifies to

$$A_0 y_t = A_1 E_t(y_{t+1}) + \varepsilon_t, \tag{4}$$

$$E(\varepsilon_t) = 0,\ E(\varepsilon_t \varepsilon_t') = \Omega.$$

If A_0 is non-singular using (4) we have

$$y_t = A_0^{-1} A_1 E_t(y_{t+1}) + A_0^{-1}\varepsilon_t = Q E_t(y_{t+1}) + A_0^{-1}\varepsilon_t. \tag{5}$$

The regular case, where there is a unique stationary solution, arises if the non-zero eigenvalues of $Q = A_0^{-1}A_1$ lie within the unit circle. In this case, the unique solution of the model is given by

$$y_t = \sum_{j=0}^{\infty} Q^j A_0^{-1} E_t(\varepsilon_{t+j}). \tag{6}$$

Since $E_t(\varepsilon_{t+j}) = 0$ for $j \geq 0$, then $E_t(y_{t+1}) = 0$ and the solution simplifies to

$$A_0 y_t = \varepsilon_t, \tag{7}$$

or

$$y_t = A_0^{-1}\varepsilon_t = u_t,\ E(u_t u_t') = \Sigma = A_0^{-1}\Omega A_0^{-1\prime}. \tag{8}$$

Notice that (8) provides us with a likelihood function which does not depend on A_1 and, therefore, the parameters that are unique to A_1 (i.e. the coefficients that are specific to the forward variables)

are not identified. Furthermore, the RE model is observationally equivalent to a model without forward variables which takes the form of (7). Since what can be estimated from the data, namely Σ, is not a function of A_1, all possible choices of A_1 are observationally equivalent in the sense that they lead to the same observed data covariance matrix. Although the coefficients in the forward solution (6) are functions of A_1, this does not identify them because $E_t(\varepsilon_{t+j}) = 0$. Elements of A_1 could be identified by certain sorts of *a priori* restrictions, but these are likely to be rather special, rather limited in number and cannot be tested.

If the parameters of the DSGE model were thought to be known *a priori* from calibration, there would be no identification problem and the structural errors ε_{it} could be recovered and used, for instance, in calculating impulse response functions, IRFs. However, suppose someone else believed that the true model was just a set of random errors $y_t = u_t$, with different IRFs. There is no information in the data that a proponent of the DSGE could use to persuade the other person that the DSGE model was correct relative to the random error model. This is exactly the same point that Sargent (1976) made with respect to "natural and unnatural rate theories".

The above result generalizes to higher order RE models. Consider for example the model

$$A_0 y_t = \sum_{i=1}^{q} A_i E_t(y_{t+i}) + \varepsilon_t.$$

Once again the unique stable solution of this model is also given by $A_0 y_t = \varepsilon_t$, and none of the elements of $A_1, A_2, ..., A_q$ that are variation free with respect to the elements of A_0 are identified.

2.2.1 Example 1. A simple New Keynesian (NK) system

As an illustration consider a standard three equation NK-DSGE model used in Benati (2010) that involves only current and future variables:

$$R_t = \psi \pi_t + \varepsilon_{1t}, \tag{9}$$

$$x_t = E_t(x_{t+1}) - \sigma(R_t - E_t(\pi_{t+1})) + \varepsilon_{2t}, \tag{10}$$

$$\pi_t = \beta E_t(\pi_{t+1}) + \gamma x_t + \varepsilon_{3t.} \tag{11}$$

where $E_t(x_{t+1}) = E(x_{t+1} \mid \Im_t)$. The model contains a monetary policy rule determining the interest rate, R_t, an IS curve determining the output gap, x_t, and a Phillips Curve determining inflation, π_t, all measured as deviations from their steady states. The errors, which are assumed to be white noise, are a monetary policy shock, ε_{1t}, a demand shock, ε_{2t}, and a supply or cost shock, ε_{3t}. These are assumed to be orthogonal. The discount factor is β and σ is the inter-temporal elasticity of substitution. This is a highly restricted system with many parameters set to zero *a priori.* For instance, output does not appear in the monetary policy rule and the coefficient of future output is exactly equal to unity in the IS equation. In terms of (4), $y_t = (R_t, x_t, \pi_t)'$ and

$$A_0 = \begin{pmatrix} 1 & 0 & -\psi \\ \sigma & 1 & 0 \\ 0 & -\gamma & 1 \end{pmatrix}, \; A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & \sigma \\ 0 & 0 & \beta \end{pmatrix}.$$

Hence

$$A_0^{-1} = \frac{1}{\gamma\sigma\psi + 1} \begin{pmatrix} 1 & \gamma\psi & \psi \\ -\sigma & 1 & -\sigma\psi \\ -\gamma\sigma & \gamma & 1 \end{pmatrix}$$

$$\begin{aligned} Q &= A_0^{-1} A_1 \\ &= \frac{1}{\gamma\sigma\psi + 1} \begin{pmatrix} 0 & \gamma\psi & \psi(\beta + \gamma\sigma) \\ 0 & 1 & \sigma(1 - \beta\psi) \\ 0 & \gamma & \beta + \gamma\sigma \end{pmatrix} \end{aligned}$$

and the two non-zero eigenvalues of Q are

$$\begin{aligned} \kappa_1 &= \frac{1}{2\left(\gamma\sigma\psi + 1\right)} \left(1 + \beta + \gamma\sigma + \zeta\right), \\ \kappa_2 &= \frac{1}{2\left(\gamma\sigma\psi + 1\right)} \left(1 + \beta + \gamma\sigma - \zeta\right), \\ \zeta &= \sqrt{\beta^2 - 2\beta + \gamma^2\sigma^2 + 2\gamma\sigma + 2\gamma\sigma\beta - 4\gamma\sigma\beta\psi + 1}. \end{aligned} \tag{12}$$

Assuming that $|\kappa_i| < 1$ for $i = 1, 2$ then the solution is given by (7), which in the present example can be written as:

$$R_t = \psi\pi_t + \varepsilon_{1t}, \; x_t = -\sigma R_t + \varepsilon_{2t}, \; \pi_t = \gamma x_t + \varepsilon_{3t}. \tag{13}$$

This illustrates some of the features of DSGE models. First, the RE model parameter matrices, A_0 and A_1, are written in terms of deeper parameters $\theta = (\gamma, \sigma, \psi, \beta)'$. Second, the parameters

which appear only in A_1 do not enter the RE solution and, thus, do not enter the likelihood function. In this example, β does not appear in the likelihood function, though, σ which appears in A_1 does appear in the likelihood function because it also appears in A_0. Third, the restrictions necessary to ensure regularity (i.e. $|\kappa_i| < 1$ for $i = 1, 2$), imply bounds involving the structural parameters, including the unidentified β. Thus, the parameter space is not variation free. Fourth, if β is fixed at some pre-selected value for the discount rate (as would be done by a calibrator), then the model is identified. Canova and Sala (2009) make similar points with a similar model.

2.3 DSGE models with lags

In order to reproduce the dynamics that are typically observed with macroeconomic data, most empirical DSGE models include lagged values of endogenous or exogenous (observed or unobserved) variables. For instance Clarida, Gali and Gertler (1999) assume that the errors in the IS and Phillips curve equations follow AR(1) processes and derive an optimal feedback policy for the interest rate based on the forecasts from these autoregressions. In this case, there is a predictable component in expected inflation because of the serial correlation in the equation errors.

Consider the special case of (1), where $A_3 = \Phi_u = 0$ so that the model only contains lagged endogenous variables

$$A_0 y_t = A_1 E_t(y_{t+1}) + A_2 y_{t-1} + \varepsilon_t. \tag{14}$$

In this case the unique solution is given by

$$y_t = C y_{t-1} + A_0^{-1} \varepsilon_t, \tag{15}$$

where C solves the quadratic matrix equation $A_1 C^2 - A_0 C + A_2 = 0$. The solution is unique and stationary if all the eigenvalues of C and $(I - A_1 C)^{-1} A_1$ lie strictly inside the unit circle. Therefore, the RE solution is observationally equivalent to the non-RE simultaneous equations model, SEM:

$$A_0 y_t = A_2 y_{t-1} + \varepsilon_t,$$

where, in the case of the SEM, $C = A_0^{-1} A_2$.

Again whereas the order condition for identification of the SEM requires n^2 restrictions, the RE model requires $2n^2$ restrictions. Not only is the RE model observationally equivalent to a purely backward looking SEM, it is observationally equivalent (in the sense of having the same reduced form), to any other model of expectations where in (14) $E_t(y_{t+1})$ is replaced by Dy_{t-1}. More specifically, knowing the form of the solution, (15), does not, on its own, provide information on the cross-equation parametric restrictions. In either case, the identifying cross-equation restrictions are lost.

Thus, in models with lags, the same problem of observational equivalence between RE and other models recurs. One may be able to distinguish the reduced forms of particular RE models from other observationally equivalent models, because the RE models impose particular types of cross-equation restriction on the reduced form, which arise from the nature of the rational expectations. But such restrictions are subject to the objection made by Sims (1980), who criticized identification by 'incredible' dynamic restrictions on the coefficients and lag lengths. RE models, which depend on restrictions on the form of the dynamics, such as AR(1) errors, are equally vulnerable to such objections.

2.3.1 Example 2: A New Keynesian Phillips Curve

An important example which has been widely discussed in the literature is the New Keynesian Phillips Curve (NKPC). The NKPC determines inflation, π_t, by expected inflation and an exogenous driving process, such as the output gap, x_t:

$$\pi_t = \beta E_{t-1}\pi_{t+1} + \gamma x_t + \varepsilon_t, \tag{16}$$

where β is the discount factor and ε_t is a martingale difference process and $E_{t-1}\pi_{t+1} = E(\pi_{t+1} \mid \mathfrak{I}_{t-1})$, where $\mathfrak{I}_{t-1}$ is information available at time $t-1$. Note that in this model expectations are conditioned on $\mathfrak{I}_{t-1}$, rather than on $\mathfrak{I}_t$. It is assumed that there is no feedback from π_t to x_t, and x_t follows a stationary AR(2) process

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + v_t; \;\; \sim IID(0, \sigma_v^2). \tag{17}$$

The RE solution of the NKPC is given in this case by

$$\pi_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_{\pi t}; \;\; \sim IID(0, \sigma^2_{\varepsilon\pi}) \tag{18}$$

where

$$\alpha_1 = \frac{\gamma\left(\rho_1 + \beta\rho_2\right)}{1 - \beta\rho_1 - \beta^2\rho_2}, \; \alpha_2 = \frac{\gamma\rho_2}{1 - \beta\rho_1 - \beta^2\rho_2}. \tag{19}$$

The reduced form parameters $\phi = (\rho_1, \rho_2, \alpha_1, \alpha_2)' = (\rho', \alpha')'$ can be obtained by estimating the system of equations (17) and (18) in $y_t = (\pi_t, x_t)'$. Assuming that x_t is weakly exogenous and hence uncorrelated with $v_{\pi t}$, ϕ is identified and can be estimated by OLS on each equation. Identification of the structural parameters $\theta = (\beta, \gamma, \rho_1, \rho_2)'$ will then involve inverting the mapping from ϕ to θ, given by (19).

As noted originally in Pesaran (1981, 1987, Ch. 7) and emphasized recently by Mavroeidis (2005) and Nason and Smith (2008) among others, identification of the structural parameters critically depends on the process generating x_t. Assuming that $\rho_2 \neq 0, \gamma \neq 0$ and the denominator in (19) $1 - \beta\rho_1 - \beta^2\rho_2 \neq 0$ then

$$\begin{aligned}
\beta &= \frac{\alpha_1\rho_2 - \alpha_2\rho_1}{\rho_2\alpha_2}, \text{ for } \rho_2\alpha_2 \neq 0, \\
\gamma &= \frac{\alpha_1\left(1 - \beta\rho_1 - \beta^2\rho_2\right)}{\rho_1 + \beta\rho_2}, \text{ for } 1 - \beta\rho_1 - \beta^2\rho_2 \neq 0.
\end{aligned}$$

Within the classical framework, the matrix of derivatives of the reduced form parameters with respect to the structural parameters, $R(\theta)$, plays an important role in identification. In this example this matrix is given by

$$R(\theta) = \frac{\partial\alpha}{\partial\theta'} = \frac{1}{1 - \beta\rho_1 - \beta^2\rho_2}\begin{bmatrix} \gamma\left(\rho_2 + \frac{\rho_1 + 2\beta\rho_2}{1 - \beta\rho_1 - \beta^2\rho_2}\right) & \rho_1 + \beta\rho_2 \\ \gamma\frac{\rho_2(\rho_1 + 2\beta\rho_2)}{1 - \beta\rho_1 - \beta^2\rho_2} & \rho_2 \end{bmatrix}.$$

In a classical set up a "yes/no" answer to to the question of whether a particular value of θ is identified is given by investigating if the rank of $R(\theta)$, evaluated at that particular value, is full. Therefore, necessary conditions for identification are $1 - \beta\rho_1 - \beta^2\rho_2 \neq 0, \gamma \neq 0$ and $\rho_2 \neq 0$. This

matrix will also play a role in the Bayesian analysis below. We shall also examine the weakly identified case, where $1-\beta\rho_1-\beta^2\rho_2\neq 0, \gamma\neq 0$, but ρ_2 is replaced by $\rho_{2T}=\delta/\sqrt{T}$.

3 Identification: General Issues

We begin with a brief overview of identification in a general context, before focussing on DSGE models. We use notation where $\theta=(\theta_1,\theta_2)'$ lies in a region Φ, $p(\theta)$ is the prior, $p(\theta|y)$ is the posterior and $L(\theta;y)$ is the likelihood function. We define identification as follows: θ is identified if $L\left(\theta^{(1)};y\right)=L\left(\theta^{(2)};y\right)$ implies that $\theta^{(1)}=\theta^{(2)}$ for all y. See, for example, Hsiao (1983, pp. 226-227). It can be seen that non-identification can occur in several ways as was illustrated in Examples 1 and 2. As we shall stress below, some of the traditional Bayesian results relate only to particular types of non-identification.

It is worth emphasizing that θ is our notation for the structural parameters and, in many cases one may have other parameters, π, such as reduced form parameters, which are functions of the structural parameters, where π could be identified, but elements of θ are not. For instance if $\pi=\theta_1\theta_2$.

Poirier (1998), building on earlier contributions (e.g. Kadane, 1974) sets out a framework for discussing identification in Bayesian models and we describe here a few of his key results which we will use later in our discussion of identification in DSGE models. Consider the case where the parameters in θ_2 are identified but the scalar, θ_1, is not. In DSGE models, the range of one parameter often depends on another because of the determinacy conditions, so care must be taken with the bounds of the parameter space. Formally, let $\theta=(\theta_1',\theta_2')'\in\Phi\subset\mathbb{R}^p$, and suppose that $\theta_1\in\Phi_1\subset\mathbb{R}^{p_1}$ and $\theta_2\in\Phi_2\subset\mathbb{R}^{p_2}$, then the parameter space over θ_1 and θ_2 is said to be variation free if $\Phi=\Phi_1\times\Phi_2$, namely if the parameter space is a product space.

Result 1: Suppose θ_1 is not identified in the sense that $L(\theta;y)$ is flat over $\theta_1\in\Phi_1(\theta_2)$. In this case, the likelihood function can be written as depending only on θ_2 (although θ_1 can still enter the bounds of the parameter space as $\Phi_2(\theta_1)$).

It is straightforward to use Result 1 and Bayes' theorem to show:

Result 2: If there is prior independence between θ_1 and θ_2 such that $p(\theta_1, \theta_2) = p(\theta_1) p(\theta_2)$ and the parameter space is a product space (i.e. $\Phi = \Phi_1 \times \Phi_2$) then $p(\theta_1|y) = p(\theta_1)$.

This is the commonly cited result that "posterior equals prior for unidentified parameters". Note, however, that it only holds for the particular type of non-identification defined in Result 1 and assumes prior independence and a variation free parameter space. If any of these conditions is not satisfied then $p(\theta_1|y) \neq p(\theta_1)$. Informally speaking, data based learning about θ_2 can "spill over" onto the unidentified θ_1 (see Koop and Poirier, 1997, for an example).

As we shall discuss below, a better metric for investigating identification can be constructed based on Proposition 2 of Poirier (1998) which we state here.

Result 3: Assume the conditions of Result 1 apply and let $p(\theta_1, \theta_2) = p(\theta_1|\theta_2) p(\theta_2)$ be the joint prior (which may exhibit correlation between θ_1 and θ_2), then the following holds for all y:

$$\begin{array}{l} p(\theta_1|y) = \int_{\Phi(\theta_2)} p(\theta_1|\theta_2, y) p(\theta_2|y) \, d\theta_2 \\ = \int_{\Phi(\theta_2)} p(\theta_1|\theta_2) p(\theta_2|y) \, d\theta_2 \\ = E_{\theta_2|y} [p(\theta_1|\theta_2)] . \end{array} \quad . \tag{20}$$

In words, the marginal posterior for the unidentified θ_1 will always be the posterior expectation of the conditional prior, $p(\theta_1|\theta_2)$.

The concepts discussed so far can be used with any econometric model, but we will use them below with DSGE models.

4 Bayesian Identification in DSGE Models

From the material in Section 2, it can be seen that some types of DSGE models are either simultaneous equations models, or closely related to them. For such models, of course, identification issues are well-understood. In the Bayesian literature on identification in the simultaneous equations model influential papers include Drèze (1976), Drèze and Richard (1983) and Kleibergen and van Dijk (1998). And Bayesian instrumental variable methods are well established (see, among many others, Kleibergen and Zivot, 2003, Hoogerheide, Kleibergen and van Dijk, 2007). Insofar as the DSGE model can be written as a conventional SEM, conventional methods can be used

for Bayesian estimation and checking for identification. The NK-DSGE and NKPC models above fall in this category. There is also a literature relating to specific models such as the NKPC (e.g. Mavroeidis, 2005 and Kleibergen and Mavroeidis, 2009, 2010). For DSGEs which can be written in structural VAR form Rubio-Ramirez, Waggoner and Zha (2010) provide an exhaustive treatment.

Bayesians typically use posterior simulation algorithms to estimate DSGE models. Our first proposed indicator can be calculated as part of such a posterior simulation algorithm without the need for additional steps such as the coding of analytical derivatives. Our second indicator involves using artificial data but it, as well, will involve standard posterior simulation algorithms. The Bayesian who uses proper priors will (under weak conditions) obtain a proper posterior, allowing for valid statistical inference. However, if a parameter is not identified, then there is the possibility that there is no data-based learning about it and its posterior can solely reflect prior information. In complicated models such as DSGEs, where it can be hard to analytically disentangle identification issues, this can lead to the case where the researcher believes she is presenting posterior estimates but is really simply reproducing her prior.

Even if parameters are identified, weak identification can lead to relatively flat regions of the likelihood function where the prior is extremely influential. Such concerns have lead to a recent interest in identification issues in Bayesian DSGE modelling. Consider, for instance, Canova (2007, page 190) which states "while it is hard to 'cheat' in a classical framework, it is not very difficult to give the impression that identification problems are absent in a Bayesian framework by choosing tight enough priors, presenting well-behaved posterior distributions and entirely side-stepping the comparison between priors and posteriors". In response to this, an increasingly common practice is to compare priors and posteriors for structural parameters, a practice which Canova (2007, page 191) refers to as "necessary [but] by no means sufficient" to reveal identification problems in DSGE models. We will draw on our earlier discussion of Bayesian identification (see Section 3.1) to discuss why this is so and introduce an alternative method for investigating identification

in Bayesian DSGE models.

Result 2 of Section 3 underlies some informal discussion of identification in the Bayesian DSGE literature. For instance, An and Schorfheide (2007, page 127) say that: "A direct comparison of priors and posteriors can often provide valuable insights about the extent to which data provide information about parameters of interest." This is true, but can be an imperfect way of formally investigating identification issues, since the posterior for an unidentified parameter can differ substantially from its prior if the unidentified parameter is, a priori, correlated with identified ones or if the parameter space is not a product space.

Nevertheless, it is common in the Bayesian DSGE literature to use such informal comparisons of priors and posteriors, as the quote from An and Schorheide above indicates. Among many others, Smets and Wouters (2007, page 594) compare prior and posteriors and note that the mean of the posterior distribution is typically quite close to the mean of the prior assumptions and later note that "It appears that the data are quite informative on the behavioral parameters, as indicated by the lower variance of the posterior distribution relative to the prior distribution."

4.1 A Bayesian Comparison Indicator for 'Yes/No' Identification

Result 3 of Section 3 offers a promising way of formally investigating identification issues. In cases where a subset of structural parameters, say θ_2, is known to be identified, but there is doubt regarding the identification of another sub-set, θ_1, then $p(\theta_1|y)$ should be equal to $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ where we use the notation of Section 3. For DSGE modelers interested in indicators that may shed light on identification issues, we would recommend comparing the properties of $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ in addition to (or instead of) comparing $p(\theta_1|y)$ to $p(\theta_1)$.

In terms of computation, note that our proposed indicator is typically easy to calculate. That is, the Bayesian DSGE modeler will typically be using an MCMC algorithm and, thus, posterior draws of θ_2 will be available. Calculation of $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ simply involves evaluating $p(\theta_1|\theta_2)$ at each draw of θ_2 at a grid of values for θ_1 and averaging across the posterior draws of θ_2. In many cases an analytical form for $p(\theta_1|\theta_2)$ will be available. For instance, if $p(\theta_1, \theta_2)$ is normal

then $p(\theta_1|\theta_2)$ is also normal with textbook formula for its mean and variance. For priors which do not admit of analytical results, adding a prior simulation step at each posterior draw would only slightly add to the computational burden.

Formally, if a parameter is unidentified then $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ should be identical, apart from MCMC approximation error. Hence, the two densities cannot be used as a test for identification. That is, any difference between $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ beyond MCMC approximation error means identification is present. We will illustrate the usefulness of this indicator below. However, for DSGE models, it has one substantive drawback. For the theory underlying Result 3 to hold, the parameters in θ_1 must not enter the likelihood function (this point is stressed on page 489 of Poirier, 1998). When working with a DSGE model, we would like to simply set θ to be the structural parameters. But typically we will not be able to do so (in the sense that the indicator defined in this way will not necessarily be zero for unidentified parameters). This is because, in general, identification problems can arise involving both the parameters in θ_1 and θ_2.

To see how this can happen in practice, we return to our example involving the NKPC in subsection 2.3.1. If $\rho_2 = 0$, there is only one reduced form parameter in the inflation equation, a_1, and it depends on two structural parameters β and γ and, thus, there is an identification problem involving these structural parameters. In contrast to Example 1, the identification problem does not manifest itself simply in terms of a single structural parameter which does not enter the likelihood function. Using the notation of Result 3, we might be tempted to set $\theta_1 = \beta$ and $\theta_2 = (\gamma, \rho)$ in order to investigate the identification of β. However, it can be easily seen that the derivations in (20) used to prove Result 3 are no longer valid since β still appears in the likelihood function. In such cases, we will not have $p(\theta_1|y) = E_{\theta_2|y}[p(\theta_1|\theta_2)]$, even though θ_1 is only partially identified.

The advice given by Poirier (1998) in such cases is to re-parameterize the model so that θ_2 contains only identified parameters. In some DSGE cases, a simple way of choosing θ_2 suggests

itself: let θ_2 be some or all of the reduced form parameters and θ_1 be one of the structural parameters. We know the reduced form parameters are identified and, thus, the conditions under which Result 3 can be used as an identification indicator are satisfied. In such cases, we can recommend a comparison of $p(\theta_1|y)$ and $E_{\theta_2|y}[p(\theta_1|\theta_2)]$ as shedding light on the identification of θ_1.

This identification indicator aims at providing a "yes/no" answer to the question of whether a given parameter is identified. In this sense it is comparable to the classical indicators of identification proposed in the literature by Iskrev (2010a) and Komunjer and Ng (2011). However, it is worth noting that these classical indicators of identification are used with linearized DSGE models, whereas our indicator is not restricted to linearized models. We also note that the MCMC approximation error in our indicator can be made as small as desired, unlike the classical procedures of Iskrev (2010a) and Komunjer and Ng (2011) that require determining whether a particular numerical matrix has rank zero (the model is not identified). Komunjer and Ng (2011) recognize this and note that the magnitude of the eigenvalues of the relevant matrix depends on the units of measurement and that rank routines in numerical packages use a cut-off to determine if the eigenvalues are sufficiently small. They have an *ad hoc* argument for the choice of tolerance. Furthermore, these alternative indictors face the additional difficulty that the rank condition tells you whether the system is identified, but it does not identify which parameter is causing the trouble. In contrast the indicator we propose is not subject to this problem.

4.2 A Bayesian Learning Rate Indicator for Weak Identification

In this sub-section we consider our second Bayesian indicator that focusses on the rate at which posterior precision rises with T. The advantage of this indicator is two-fold. First, it can be applied to any parameter of interest and is relatively easy to compute. Second, and more importantly, it can be used to detect evidence of weak identification which is likely to be more prevalent in the empirical Bayesian analysis. Our theoretical results also show that in cases where identification is weak the Bayesian outcomes are likely to be more sensitive to the choice of the priors even if the

sample size under consideration is relatively large. Empirically, this indicator involves simulating artificial data sets of increasing size and then estimating the DSGE model using these data sets. Since the generation of simulated data is fairly standard in the DSGE literature, this strategy fits in well with existing empirical methodologies.

To explain the theory underlying this second indicator, note that standard Bayesian results for stationary models (see, e.g., Berger, 1985, page 224), imply that, under certain regularity conditions, the Bayesian asymptotic theory relating to the posterior is identical to the asymptotic distribution theory for the maximum likelihood estimator. Thus, for instance, the posterior for θ will asymptotically converge to to its true value and the role of the prior will vanish. One of the regularity conditions is that θ is identified. In this sub-section, we relax this assumption and show that this asymptotic convergence will not occur. To this end we make the following assumptions.

Assumption 1 *Let $\theta = (\theta_1, \theta_2, ..., \theta_p)' \in \Theta \subset \mathbb{R}^p$ be a $p \times 1$ vector of structural parameters of interest in a DSGE model of the type set out in Section 2, and suppose that the likelihood function of the DSGE model for a sample of T observations can be written as $L_T(\alpha(\theta); y)$, where $\alpha(\theta) = (\alpha_1(\theta), \alpha_2(\theta), ..., \alpha_k(\theta))'$ is a $k \times 1$ vector-valued function of θ. Further suppose that $\alpha_s(\theta)$, for $s = 1, 2, ..., k$ are differentiable at each point of the open convex set Θ^0 containing Θ, and denote the first derivatives of $\alpha_s(\theta)$ by $\partial\alpha_s(\theta)/\partial\theta_i$, $i = 1, 2, ..., p$, and the $k \times p$ matrix of derivatives by*

$$R(\theta) = \frac{\partial\alpha(\theta)}{\partial\theta'} = (\partial\alpha_s(\theta)/\partial\theta_i). \tag{21}$$

Assumption 2 *Let $\ell_T(\alpha) = \ln[L_T(\alpha; y)]$, and assume that*

$$Q_T(\alpha) = \frac{-1}{T}\frac{\partial^2\ell_T(\alpha)}{\partial\alpha\partial\alpha'} \tag{22}$$

is a positive definite matrix for all values of $\alpha \in A \subset \mathbb{R}^k$. Denote the maximum likelihood (ML) estimator of α by $\hat{\alpha}_T$ and assume that

$$\sqrt{T}(\hat{\alpha}_T - a_0) \rightarrow_d N(0, Q^{-1}), \tag{23}$$

where α_0 is the true value of $\alpha \in A$, and $Q = plim_{T\rightarrow\infty}Q_T(\alpha_0)$ is a symmetric positive definite

matrix. Similarly, corresponding to α_0 *select* $\theta_0 \in \Theta$ *such that* $\alpha(\theta_0) = \alpha_0$*. The choice of* θ_0 *need not be unique so long as it satisfies the equation system,* $\alpha_0 = \alpha(\theta_0)$*.*

Assumption 3 *The prior distribution of* θ *is given by the multivariate normal density*

$$p(\theta) = (2\pi)^{-p/2} \left|\underline{H}\right|^{1/2} \exp -\frac{1}{2}(\theta - \underline{\theta})' \underline{H}(\theta - \underline{\theta}), \tag{24}$$

where $\underline{\theta}$ *and* $\underline{H}$ *are prior mean vector and prior precision matrix of* θ*.*

Remark 1 *Assumption 1 is standard and postulates a well defined likelihood function for the model under consideration, and sets out the relationships that exist between the structural,* θ*, and the reduced form,* α*, parameters. Assumption 2 is again standard and assumes that* α *is identified and its ML estimator is asymptotically normally distributed. Since our focus is on identification of the structural parameters, it seems reasonable to begin with the standard case where the reduced form parameters are identified, although as we shall see below our analyses allow for the possibility of weak identification of the structural parameters. Assumption 3 is made for analytical convenience and can be relaxed at the expense of further mathematical complications. In the theoretical derivations we focus on the multivariate normal prior, although in simulations we also consider truncated normal densities for some of the priors.*

Remark 2 *In contrast to the identified case where the likelihood eventually dominates the prior (assuming that the prior is nonprejudicial), as we shall see, the choice of the prior is important in the unidentified and weakly identified cases, even if* $T \to \infty$*. Despite this, our Bayesian learning rate indicator continues to be applicable even if Assumption 3 is relaxed and priors other than multivariate normal are assumed. Although, for non-normal priors the exact expression for the limit of the posterior precision in the unidentified or weakly identified cases would be different from those provided below, nevertheless it still follows that the rate of increase of the posterior precision would be slower than* T *in the unidentified or weakly identified cases. Therefore, the proposed Bayesian learning rate indicator is applicable more generally and does not depend on the particular choice of the priors.*

Given that by assumption α is identified, then in the classical sense θ is globally identified if $\text{Rank}(R(\theta_0)) = k \geq p$ for all $\theta_0 \in \Theta$, and θ is identified locally in the neighborhood of $\underline{\theta}$ if $\text{Rank}(R(\underline{\theta})) = k \geq p$. A necessary condition for identification is given by $p \leq k$. Our analysis focusses on the case where the rank condition is not met, and hence one or more elements of θ might not identified. Nevertheless, there exists $\hat{\theta}_T$ such that $\hat{\alpha}_T = \alpha(\hat{\theta}_T)$. The choice of $\hat{\theta}_T$ need *not* be unique, but as can be seen from the following proposition this does not affect the posterior precision of θ which is uniquely defined irrespective of the rank of R.

Proposition 1 *Under Assumptions 1, 2, and 3, the posterior distribution of θ is approximately normal with mean $\bar{\theta}_T$ and the precision matrix $\bar{H}_T$, where*

$$\bar{\theta}_T = (T\ S_T + \underline{H})^{-1}\,(T\ S_T\ \hat{\theta}_T + \underline{H}\ \underline{\theta}), \tag{25}$$

$$\bar{H}_T = T\ S_T + \underline{H}, \tag{26}$$

$$S_T = R' Q_T R, \tag{27}$$

$R = R(\theta)$, for all $\theta \in \Theta^0$, with $R(\theta)$ defined by (21), for all choices of $\hat{\theta}_T$ such that $\hat{\alpha}_T = \alpha(\hat{\theta}_T)$.

A proof is provided in the Appendix.

It is clear that the posterior precision, $\bar{H}_T$, is unique and well defined even if R, or equivalently S_T, fails the rank condition. In what follows we use this result and consider the limiting behaviour of the posterior precision of one or more elements of θ. To simplify the exposition consider the case where $p = k$, and partition θ into a $p_1 \times 1$ vector, θ_1, whose identification is under consideration on the assumption that the remaining $p_2 \times 1$ parameter vector, θ_2, is identified. Since the marginals of the multivariate normal distribution are also normally distributed, the posterior of θ_1, up to terms of order $O_p(T^{-1/2})$, is normally distributed with the posterior precision given by

$$\bar{H}_{11,T} = \bar{H}_{11,T} - \bar{H}_{12,T}\bar{H}_{22,T}^{-1}\bar{H}_{21,T},$$

where, using (26),

$$\bar{H}_T = \begin{pmatrix} \bar{H}_{11,T} & \bar{H}_{12,T} \\ \bar{H}_{21,T} & \bar{H}_{22,T} \end{pmatrix} = T \begin{pmatrix} S_{11,T} & S_{12,T} \\ S_{21,T} & S_{22,T} \end{pmatrix} + \begin{pmatrix} \underline{H}_{11} & \underline{H}_{12} \\ \underline{H}_{21} & \underline{H}_{22} \end{pmatrix}.$$

Hence

$$\bar{H}_{11,T} = (TS_{11,T} + \underline{\mathrm{H}}_{11}) - (TS_{12,T} + \underline{\mathrm{H}}_{12})(TS_{22,T} + \underline{\mathrm{H}}_{22})^{-1}(TS_{21,T} + \underline{\mathrm{H}}_{21}). \tag{28}$$

For a finite T, $\bar{H}_{11,T}$ is well defined irrespective of whether R is a full rank matrix. Thus in small samples identification is not an issue in Bayesian analysis, although in public discourse the sensitivity of the posterior means and precisions to the choice of the priors will be an issue. It is for answering this question that it is important to investigate the limiting properties of $\bar{H}_{11,T}$ under different assumptions about the rank of S_T.

4.2.1 The identified case

Initially, we consider the identified case where S_T is bounded in T, and has a full rank for all T including as $T \to \infty$. In this case $S_{11,T} - S_{12,T}S_{22,T}^{-1}S_{21,T}$ is a positive definite matrix for all T, and we have

$$\begin{aligned} T^{-1}\bar{H}_{11,T} &= (S_{11,T} + T^{-1}\underline{\mathrm{H}}_{11}) - \left(S_{12,T} + T^{-1}\underline{\mathrm{H}}_{12}\right)\left(S_{22,T} + T^{-1}\underline{\mathrm{H}}_{22}\right)^{-1}\left(S_{21,T} + T^{-1}\underline{\mathrm{H}}_{21}\right) \\ &= S_{11,T} + T^{-1}\underline{\mathrm{H}}_{11} - \left(S_{12,T} + T^{-1}\underline{\mathrm{H}}_{12}\right)S_{22,T}^{-1}\left(I + T^{-1}\underline{\mathrm{H}}_{22}S_{22}^{-1}\right)^{-1}\left(S_{21,T} + T^{-1}\underline{\mathrm{H}}_{21}\right). \end{aligned} \tag{29}$$

Furthermore, all eigenvalues of $\underline{\mathrm{H}}_{22}S_{22,T}^{-1}$ will also be bounded in T, and we have

$$\left(I_{p_2} + T^{-1}\underline{\mathrm{H}}_{22}S_{22,T}^{-1}\right)^{-1} = I_{p_2} - T^{-1}\underline{\mathrm{H}}_{22}S_{22,T}^{-1} + O\left(T^{-2}\right).$$

Using this result in (29) now yields

$$\begin{aligned} T^{-1}\bar{H}_{11,T} &= \left(S_{11,T} - S_{12,T}S_{22,T}^{-1}S_{21,T}\right) \\ &\quad + T^{-1}\left(\underline{\mathrm{H}}_{11} - \underline{\mathrm{H}}_{12}S_{22,T}^{-1}S_{21,T} - S_{12,T}S_{22,T}^{-1}\underline{\mathrm{H}}_{21} + S_{12,T}S_{22,T}^{-1}\underline{\mathrm{H}}_{22}S_{22}^{-1}S_{21,T}\right) + O\left(T^{-2}\right). \end{aligned} \tag{30}$$

Therefore, in the identified case the dependence of the average posterior precision on the priors diminish at the rate of T^{-1} and eventually vanish for T sufficiently large. In the limit we have

$$\lim_{T\to\infty}\left(T^{-1}\bar{H}_{11,T}\right) = \lim_{T\to\infty}\left(S_{11,T} - S_{12,T}S_{22,T}^{-1}S_{21,T}\right) > 0.$$

4.2.2 Unidentified/weakly identified cases

We now turn to the weakly identified case. Following the classical literature on weak identification (see, e.g., Stock, Wright and Yogo, 2002), we allow the elements of $\partial\alpha_s(\theta)/\partial\theta_i$ to vary with T, so that R_T could be full rank for a fixed T but not as $T \to \infty$. Specifically, we set

$$R_T = \begin{pmatrix} R_{11,T} & R_{12,T} \\ R_{21,T} & R_{22,T} \end{pmatrix}, \tag{31}$$

and define weak identification by

$$R_{11,T} = T^{-1/2}\Delta_{11,T}\ ,\ R_{12,T} = T^{-1/2}\Delta_{12,T}, \tag{32}$$

where $\Delta_{11,T}$ and $\Delta_{12,T}$ are respectively $p_1 \times p_1$ and $p_1 \times p_2$ matrices bounded in T. Additional terms of order T^{-1} can also be included in specification of $R_{11,T}$ and $R_{12,T}$, by setting $R_{11,T} = T^{-1/2}\Delta_{11,T} + O\left(T^{-1}\right)$, and $R_{12,T} = T^{-1/2}\Delta_{12,T} + O\left(T^{-1}\right)$, with no effects on the limiting behaviour of the posterior precision. We also allow $R_{21,T}$ and $R_{22,T}$ to vary with T, but require $R_{22,T}$ to have full rank uniformly in T (for all T and as $T \to \infty$). The unidentified case then arises as a special case when $\Delta_{11,T} = 0$ and $\Delta_{12,T} = 0$.

Since under Assumption 2, Q_T is a symmetric positive definite matrix bounded in T, then we can write $Q_T = U_T'U_T$, where U_T, the Cholesky factor of Q_T, is a lower triangular matrix. Define P_T as

$$\begin{aligned} P_T = U_T R_T &= \begin{pmatrix} U_{11,T} & 0 \\ U_{21,T} & U_{22,T} \end{pmatrix} \begin{pmatrix} R_{11,T} & R_{12,T} \\ R_{21,T} & R_{22,T} \end{pmatrix} \\ &= \begin{pmatrix} U_{11,T}R_{11,T} & U_{11,T}R_{12,T} \\ U_{21,T}R_{11,T} + U_{22,T}R_{21,T} & U_{21,T}R_{12,T} + U_{22,T}R_{22,T} \end{pmatrix} = \begin{pmatrix} P_{11,T} & P_{12,T} \\ P_{21,T} & P_{22,T} \end{pmatrix}, \end{aligned} \tag{33}$$

and note that under weak identification we have

$$P_{11,T} = T^{-1/2}U_{11,T}\Delta_{11,T} = T^{-1/2}D_{11,T},\ P_{12,T} = T^{-1/2}U_{11,T}\Delta_{12,T} = T^{-1/2}D_{12,T}.$$

Note that $D_{11,T}$ and $D_{12,T}$ continue to be bounded in T. It can also be seen that $P_{22,T} = U_{21,T}R_{12,T} + U_{22,T}R_{22,T}$ is a full rank matrix for all T. Recall that $U_{22,T}$ and $R_{22,T}$ are both $p_2 \times p_2$ full rank matrices and are bounded in T. The second main theoretical result of the paper can now be summarized in the following proposition

Proposition 2 *Suppose that Assumptions 1-3 hold, and θ_1, the $p_1 \times 1$ sub-vector of the structural parameters, $\theta = (\theta_1', \theta_2')'$, is weakly identified as defined by (32). Then the posterior precision of θ_1 is given by*

$$\begin{aligned}\bar{H}_{11,T} = &\left(D_{11,T} - D_{12,T}P_{22,T}^{-1}P_{21,T}\right)' \left(D_{11,T} - D_{12,T}P_{22,T}^{-1}P_{21,T}\right) \\ &+ \underline{H}_{11} - P_{21,T}'P_{22,T}'^{-1} - \underline{H}_{12}P_{22,T}^{-1}P_{21,T} + P_{21,T}'P_{22,T}'^{-1}\underline{H}_{22}P_{22,T}^{-1}P_{21,T} + O(T^{-1}).\end{aligned} \tag{34}$$

where

$$D_{11,T} = U_{11,T}\Delta_{11,T},\ D_{12,T} = U_{11,T}\Delta_{12,T},$$

$$P_{22,T} = U_{21,T}R_{12,T} + U_{22,T}R_{22,T},\ \ P_{21,T} = U_{21,T}R_{11,T} + U_{22,T}R_{21,T},$$

$\underline{H}_{11}$, $\underline{H}_{21}$, and $\underline{H}_{22}$, are the prior precisions associated with the partitioned vector, $\theta = (\theta_1', \theta_2')'$, and $U_{11,T}$, $U_{21,T}$, and $U_{22,T}$ are defined by U_T, the lower triangular Cholesky factor of Q_T.

See the Appendix for a proof.

This proposition establishes that the posterior precision of the weakly identified θ_1 does not rise with T and is bounded in the prior precisions ($\underline{\mathrm{H}}_{11}$,$\underline{\mathrm{H}}_{21}$, $\underline{\mathrm{H}}_{22}$), and the data, through the matrices $D_{11,T}, D_{12,T}, P_{22,T}$, and $P_{21,T}$. Using (34), it also readily follows that, in the unidentified case where $\Delta_{11,T} = 0$ and $\Delta_{12,T} = 0$, the posterior precision of θ_1 reduces to

$$\bar{H}_{11,T} = \underline{\mathrm{H}}_{11} - P_{21,T}'P_{22,T}'^{-1}\underline{\mathrm{H}}_{21} - \underline{\mathrm{H}}_{12}P_{22,T}^{-1}P_{21,T} + P_{21,T}'P_{22,T}'^{-1}\underline{\mathrm{H}}_{22}P_{22,T}^{-1}P_{21,T} + O(T^{-1}), \tag{35}$$

and shows that, in general, the prior and posterior precisions of θ_1 differ even if θ_1 is unidentified. For the lim of $\bar{H}_{11,T}$ to coincide with $\underline{\mathrm{H}}_{11}$, we also need $\lim_{T\to\infty} P_{21,T} = lim_{T\to\infty}\left(U_{21,T}R_{11,T} + U_{22,T}R_{21,T}\right) = 0$. Sufficient conditions for this latter condition to hold are

$$\lim_{T\to\infty} \frac{1}{T}\frac{\partial^2 \ell_T(\alpha)}{\partial\alpha_i\partial\alpha_j'} = 0, \text{ and } \frac{\partial\alpha_i(\theta)}{\partial\theta_j} = 0, \text{for } i = 1, 2...., p_1; j = p_1 + 1, p_2 + 2, ..., p.$$

Namely, the posterior precisions of unidentified parameters tend to their prior precisions if they are isolated from the rest of the structural parameters, and can be estimated using reduced form

parameters that are asymptotically uncorrelated with the remaining reduced form parameters used for estimation of the structural parameters that are identified.

Finally, it is interesting to note that the average precision, $T^{-1}\bar{H}_{11,T}$, tends to zero irrespective of whether θ_1 is unidentified or weakly identified. But, the limiting value of the posterior precision of θ_1 under the weakly identified case can never be smaller than the limiting value of the posterior precision under the unidentified case. More specifically, denoting the limit values of $D_{11,T}$, $D_{12,T}$, $P_{22,T}$, etc. by $D_{11}, D_{12}, P_{22}, ...$, we have

$$\begin{aligned}&\lim_{T\to\infty}\left(\bar{H}_{11,T}\,|\theta_1 \text{ is weakly identified}\right) - \lim_{T\to\infty}\left(\bar{H}_{11,T}\,|\theta_1 \text{ is unidentified}\right)\\ &= \left(D_{11} - D_{12}P_{22}^{-1}P_{21}\right)'\left(D_{11} - D_{12}P_{22}^{-1}P_{21}\right) \geq 0. \end{aligned} \tag{36}$$

In terms of the original parameterization recall that $D_{11} = U_{11}\Delta_{11}$, $D_{12} = U_{11}\Delta_{12}$, $P_{22} = U_{21}R_{12} + U_{22}R_{22}$, and $P_{21} = U_{21}R_{11} + U_{22}R_{21}$, where

$$U = \begin{pmatrix} U_{11} & 0 \\ U_{21} & U_{22} \end{pmatrix},$$

is the Cholesky factor of $Q = p\lim_{T\to\infty}(Q_T)$, and $\Delta_{11}, \Delta_{12}, R_{11}, ...$ are the limits of $\Delta_{11,T}, \Delta_{12,T}, R_{11,T}, ...$ as $T \to \infty$.

Implications for practice: The above results suggest using the limiting behavior of posterior precision as an indicator of weak identification in Bayesian analysis. The problem of whether a parameter is identified or not (a yes/no answer) can be provided either with the help of classical procedures of the type proposed by Iskrev (2010a) and Komunjer and Ng (2011), or by using our first indicator when possible. Once, we are satisfied that a parameter of interest is (locally) identified, the posterior precision of the parameter in question can then be computed using simulated data of increasing size. More specifically, suppose that it is of interest to investigate if one or more elements of θ are weakly identified, for example in the locality of the prior mean vector, $\underline{\theta}$. First, the researcher should generate an artificial data set of size T from the DSGE model at $\theta = \underline{\theta}$. T should be chosen to be a large value where asymptotic results are expected to be very good approximations (e.g. the empirical illustrations below set $T = 10,000$). Second, the

researcher should estimate the posterior precision of the parameters of the DSGE model using sample sizes $\tau = cT$ for a grid of values for c (e.g. $c = 0.001, 0.01, 0.1, 1.0$). By comparing the behavior of the posterior precision over different sample sizes, the researcher can see which parameters are weakly identified and which are not. For instance, the posterior precision should be rising with the sample size for a well identified parameter, but will not be doing so for a weakly identified parameter, while the posterior precision, divided by the sample size will go to a constant for identified parameters, and go to zero for unidentified parameters. Notice that for identified parameters, the posterior precision need not rise monotonically with T, the posterior precision may fall before rising depending on the priors.

This approach provides an indicator of weak identification at a locality of a point in the sample space. Hence, the researcher may wish to carry out the procedure for various artificial data sets generated with different parameter values. This strategy is comparable to the one used by Iskrev (2010a), who draws parameter values from the prior and checks identification at each of the draws.

4.3 Example 2 (continued): The posterior precisions for the NKPC under weak identification

To illustrate the case of weak identification we derive the posterior precision of β in the NKPC example in sub-section 2.3.1 from the general derivations in sub-section 4.2.2. It proves convenient to define α as $\alpha = (\alpha_2, \alpha_1)'$, namely switching the position of α_1 and α_2 in α. But as before we set $\theta = (\beta, \gamma)'$ and recall that

$$\alpha_1 = \frac{\gamma\left(\rho_1 + \beta\rho_2\right)}{1 - \beta\rho_1 - \beta^2\rho_2},\ \alpha_2 = \frac{\gamma\rho_2}{1 - \beta\rho_1 - \beta^2\rho_2}.$$

The other parameters in the model are the coefficients for the AR(2) process for the output gap, ρ_1 and ρ_2. We assume weak identification through the specification $\rho_2 = \rho_{2T} = \delta/\sqrt{T}$. Under this specification we have

$$R_T(\theta) = \frac{\partial\alpha}{\partial\theta'} = \frac{1}{1 - \beta\rho_1 - \beta^2\rho_{2T}} \begin{bmatrix} \gamma\frac{\rho_{2T}(\rho_1 + 2\beta\rho_{2T})}{1-\beta\rho_1-\beta^2\rho_{2T}} & \rho_{2T} \\ \gamma\left(\rho_{2T} + \frac{\rho_1 + 2\beta\rho_{2T}}{1-\beta\rho_1-\beta^2\rho_{2T}}\right) & \rho_1 + \beta\rho_{2T} \end{bmatrix}.$$

and

$$Q_T = \frac{1}{\hat{\sigma}^2_{\varepsilon\pi}} \begin{pmatrix} T^{-1}\sum_{t=3}^{T} x^2_{t-1} & T^{-1}\sum_{t=3}^{T} x_{t-1}x_{t-1} \\ T^{-1}\sum_{t=3}^{T} x_{t-1}x_{t-2} & T^{-1}\sum_{t=3}^{T} x^2_{t-2} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}^2_1 & \hat{\sigma}_1\hat{\sigma}^2_2\hat{\lambda} \\ \hat{\sigma}_1\hat{\sigma}^2_2\hat{\lambda} & \hat{\sigma}^2_2 \end{pmatrix},$$

where $\hat{\sigma}^2_{\varepsilon\pi} = T^{-1}\sum_{t=3}^{T}(\pi_t - \hat{\alpha}_1 x_{t-1} - \hat{\alpha}_2 x_{t-2})^2$, and $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are OLS estimators of α_1 and α_2 in the regression (18), which we repeat:

$$\pi_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_{\pi t},\ \varepsilon_{\pi t} \sim IID(0, \sigma^2_{\varepsilon\pi}).$$

Assuming that x_t is a stationary process we have

$$Q = p\lim_{T\to\infty}(Q_T) = \frac{1}{\sigma^2_{\varepsilon\pi}} \begin{pmatrix} \gamma_x(0) & \gamma_x(1) \\ \gamma_x(1) & \gamma_x(0) \end{pmatrix},$$

where $\gamma_x(s)$ is the autocovariance of x_t of order s, and $\sigma^2_{\varepsilon\pi} = Var(\varepsilon_{\pi t})$. Hence,

$$\sigma^2_1 = \sigma^2_2 = \gamma_x(0)/\sigma^2_{\varepsilon\pi},\ \text{and}\ \lambda = \gamma_x(1)/\gamma_x(0) = \rho_x(1), \tag{37}$$

where $\rho_x(1)$ is the first order autocorrelation coefficient of x_t. We note for reference, that using familiar results for the autocovariances of an $AR(2)$ process, e.g. Hamilton(1994, p. 58), we have

$$\gamma_x(0) = \frac{\sigma^2_v(1-\rho_{2T})}{(1+\rho_{2T})\left[(1-\rho_{2T})^2 - \rho^2_1\right]}, \gamma_x(1) = \left(\frac{\rho_1}{1-\rho_{2T}}\right)\gamma_x(0). \tag{38}$$

Since x_t is stationary, the limit value of Q_T, does not change as a result of switching the positions of α_1 and α_2 in α, and we have

$$Q = \frac{\gamma_x(0)}{\sigma^2_{\varepsilon\pi}} \begin{pmatrix} 1 & \rho_x(1) \\ \rho_x(1) & 1 \end{pmatrix} = U'U,$$

with the Cholesky factor of Q given by

$$U = \frac{\sqrt{\gamma_x(0)}}{\sigma_{\varepsilon\pi}} \begin{pmatrix} \sqrt{1-\rho^2_x(1)} & 0 \\ \rho_x(1) & 1 \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}.$$

Under weak identification where $\rho_{2T} = \delta/\sqrt{T}$, we obtain

$$R_{11,T} = T^{-1/2}\frac{\gamma\delta\rho_1}{(1-\beta\rho_1)^2} + O(T^{-1}),\ R_{12,T} = T^{-1/2}\frac{\delta}{1-\beta\rho_1} + O(T^{-1}),$$

and in terms of our general notation in (32) we have

$$\Delta_{11} = \frac{\gamma\delta\rho_1}{(1-\beta\rho_1)^2},\ \text{and}\ \Delta_{12} = \frac{\delta}{1-\beta\rho_1}.$$

$$R = \frac{1}{1-\beta\rho_1}\begin{pmatrix} 0 & 0 \\ \frac{\gamma\rho_1}{1-\beta\rho_1} & \rho_1 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}.$$

Also, using U,

$$\begin{aligned} D_{11} &= U_{11}\Delta_{11} = \frac{\sqrt{\gamma_x(0)}}{\sigma_{\varepsilon\pi}}\frac{\sqrt{1-\rho_x^2(1)}\gamma\delta\rho_1}{(1-\beta\rho_1)^2} \\ D_{12} &= U_{11}\Delta_{12} = \frac{\sqrt{\gamma_x(0)}}{\sigma_{\varepsilon\pi}}\frac{\sqrt{1-\rho_x^2(1)}\delta}{1-\beta\rho_1} \end{aligned} \tag{39}$$

Similarly,

$$\begin{aligned} P_{22} &= U_{21}R_{12} + U_{22}R_{22} = \rho_x(1)\times 0 + \frac{\rho_1}{1-\beta\rho_1} = \frac{\rho_1}{1-\beta\rho_1} \\ P_{21} &= U_{21}R_{11} + U_{22}R_{21} = \rho_x(1)\times 0 + \frac{\gamma\rho_1}{(1-\beta\rho_1)^2} = \frac{\gamma\rho_1}{(1-\beta\rho_1)^2} \end{aligned} \tag{40}$$

From our general results in (34)

$$\begin{aligned} &\lim_{T\to\infty}\left(\bar{H}_{11,T}\left|\rho_2 = \delta/\sqrt{T}\right.\right) = \left(D_{11} - D_{12}P_{22}^{-1}P_{21}\right)'\left(D_{11} - D_{12}P_{22}^{-1}P_{21}\right) \\ &\underline{\mathrm{H}}_{11} - P_{21}'P_{22}'^{-1}\underline{\mathrm{H}}_{21} - \underline{\mathrm{H}}_{12}P_{22}^{-1}P_{21} + P_{21}'P_{22}'^{-1}\underline{\mathrm{H}}_{22}P_{22}^{-1}P_{21} \end{aligned}$$

But since P_{12} and P_{22} are scalars we have $\underline{\mathrm{H}}_{21} = \underline{\mathrm{H}}_{12}$

$$\begin{aligned} &\lim_{T\to\infty}\left(\bar{H}_{11,T}\left|\rho_2 = \delta/\sqrt{T}\right.\right) = \left(D_{11} - D_{12}\frac{P_{21}}{P_{22}}\right)'\left(D_{11} - \frac{P_{21}}{P_{22}}D_{12}\right) \\ &+ \underline{\mathrm{H}}_{11} - 2\underline{\mathrm{H}}_{12}\frac{P_{21}}{P_{22}} + \underline{\mathrm{H}}_{22}\left(\frac{P_{21}}{P_{22}}\right)^2. \end{aligned}$$

Now using (39) and (40) we have $P_{21}/P_{22} = \gamma/\left(1-\beta\rho_1\right)$,and

$$\begin{aligned} D_{11} - D_{12}\frac{P_{21}}{P_{22}} &= \frac{\sqrt{\gamma_x(0)}}{\sigma_{\varepsilon\pi}}\frac{\sqrt{1-\rho_x^2(1)}\gamma\delta\rho_1}{(1-\beta\rho_1)^2} - \frac{\sqrt{\gamma_x(0)}}{\sigma_{\varepsilon\pi}}\frac{\sqrt{1-\rho_x^2(1)}\delta}{1-\beta\rho_1}\frac{\gamma}{1-\beta\rho_1} \\ &= \frac{\sqrt{\gamma_x(0)\left[1-\rho_x^2(1)\right]}}{\sigma_{\varepsilon\pi}}\frac{\gamma\delta\left(\rho_1-1\right)}{\left(1-\beta\rho_1\right)^2}. \end{aligned}$$

Hence

$$\begin{aligned} &\lim_{T\to\infty}\left(\bar{H}_{11,T}\left|\rho_2 = \delta/\sqrt{T}\right.\right) = \frac{\gamma^2\delta^2\left(1-\rho_1\right)^2}{\left(1-\beta\rho_1\right)^4}\left(\frac{\gamma_x(0)}{\sigma_{\varepsilon\pi}^2}\right)\left[1-\rho_x^2(1)\right] \\ &+ \underline{\mathrm{H}}_{11} - 2\underline{\mathrm{H}}_{12}\left(\frac{\gamma}{1-\beta\rho_1}\right) + \underline{\mathrm{H}}_{22}\left(\frac{\gamma}{1-\beta\rho_1}\right)^2. \end{aligned} \tag{41}$$

Noting that in the present example all the precisions are scalars, we set $\bar{H}_{11,T} = \bar{h}_{T,\beta\beta}$, $\underline{H}_{11} = \underline{h}_{\beta\beta}$, $\underline{H}_{22} = \underline{h}_{\gamma\gamma}$, $\underline{H}_{12} = \underline{h}_{\beta\gamma}$, and note that

$$\bar{h}_{\beta\beta}(\delta) = \lim_{T\to\infty} \bar{h}_{T,\beta\beta}\left(\rho_{2T} = \delta/\sqrt{T}\right) = \frac{\gamma^2\delta^2\left(1-\rho_1\right)^2}{\left(1-\beta\rho_1\right)^4}\frac{\gamma_x(0)}{\sigma_{\varepsilon\pi}^2}\left[1-\rho_x^2(1)\right] + \left(\underline{h}_{\beta\beta} + \kappa^2\underline{h}_{\gamma\gamma} - 2\kappa\underline{h}_{\beta\gamma}\right).$$

where $\kappa = P_{21}/P_{22} = \gamma/\left(1-\beta\rho_1\right)$.

Using (37) and (38) this is

$$\bar{h}_{\beta\beta}(\delta) = \lim_{T\to\infty} \bar{h}_{T,\beta\beta}\left(\rho_{2T} = \delta/\sqrt{T}\right) = \kappa^2\delta^2\left(\frac{\sigma_v^2}{\sigma_{\varepsilon\pi}^2}\right)\frac{\left(1-\rho_1\right)^2}{\left(1-\beta\rho_1\right)^2} + \left[\underline{h}_{\beta\beta} + \kappa^2\underline{h}_{\gamma\gamma} - 2\kappa\underline{h}_{\beta\gamma}\right]. \quad (42)$$

A similar procedure can be used to show that

$$\bar{h}_{\gamma\gamma}(\delta) = \lim_{T\to\infty} \bar{h}_{T,\gamma\gamma}\left(\rho_{2T} = \delta/\sqrt{T}\right) = \delta^2\left(\frac{\sigma_v^2}{\sigma_{\varepsilon\pi}^2}\right)\frac{\left(1-\rho_1\right)^2}{\left(1-\beta\rho_1\right)^2} + \left[\kappa^{-2}\underline{h}_{\beta\beta} + \underline{h}_{\gamma\gamma} - 2\kappa^{-1}\underline{h}_{\beta\gamma}\right]. \quad (43)$$

Therefore, as to be expected from the general theoretical results, in the weakly identified case the posterior precisions do not rise with T, and tend to a finite limit the size of which depend on the prior precisions, the parameters of the underlying model, and the strength of identification as measured by δ.

In line with the analysis in sub-section 4.2, we will now give simulations below in sub-section 5.2 that illustrate the effect of weak identification on posterior precision for the NKPC parameters.

5 Applications

In this section, we illustrate both of our Bayesian identification indicators in the context of the two examples of DSGE models introduced in Section 2. These are the NK-DSGE (see Section 2.2.1) and the NKPC (see Section 2.3.1).

5.1 Example 1 (cont.): Bayesian identification of the simple NK-DSGE model

Previously, we introduced a simple NK-DSGE in (9), (10) and (11). We will illustrate some issues relating to Bayesian inference and identification in this simple and easily understood model where the identification of the model can be immediately seen. This example involves four structural parameters, σ, γ, ψ and β. The rational expectations solution given in (13) does not involve β so

this parameter is unidentified. However, the bounds given in (12) which ensure regularity such that there is a unique stationary solution do involve β.

We generated one artificial data set of $T = 10,000$ observations from (13) with $\sigma = 0.4$, $\gamma = 0.75$ and $\psi = 2.0$. These values were chosen so as to be not too far from the boundaries given in (12), but also not too near. The posterior simulation algorithm rejected 3.1% of the draws for violating the bounds. The errors, ε_{jt} for $j = 1, 2, 3$ are all standard normal and independent of one another.

We estimate the model using different sample sizes and two different priors. Both priors are normal with prior means: $E(\sigma) = 0.4$, $E(\gamma) = 0.75$, $E(\psi) = 2.0$ and $E(\beta) = 0.9$. The two priors differ in their prior variances. Let $\theta = (\sigma, \gamma, \psi, \beta)'$. The first prior (which we call the Independent Prior) has $var(\theta) = I_4$. The second prior (the Dependent Prior) has the same prior covariance matrix except for a single element: this is the covariance between β and σ which is set to 0.9. These priors are combined with the likelihood function based on the three equation system in (13). We use a random walk Metropolis-Hastings algorithm to do posterior simulation in this model. In this small model, with only four parameters, this algorithm works well. In larger models, depending on the form of the prior, more efficient posterior simulation algorithms could be used.

We begin by illustrating the properties of our first Bayesian identification indicator with $T = 100$ (i.e. we use only the first 100 of the artificially generated observations). Figures 1 and 2 graph various priors and posteriors for β and σ, respectively, for the Independent Prior. Figures 3 and 4 do the same for the Dependent Prior. For the sake of brevity, we only present graphs for one identified and one unidentified parameter and, thus, do not present graphs for γ and ψ.

Consider first the priors and posteriors for β. Since β is unidentified, a naive researcher may expect its posterior to equal its prior. For the reasons discussed in Section 3, this may not be the case. The top panels of Figures 1 and 3 illustrate this empirically. Even with the Independent Prior (where β is, a priori, uncorrelated with the other parameters in the model), the fact that β

enters the bounds for the regularity region given in (12) has an appreciable impact on the posterior in Figure 1. In Figure 3 (which uses a prior where the unidentified β is strongly correlated with the identified σ), this effect is even more noticeable. The posterior for β has a much smaller variance than its prior, indicating how information about σ is spilling over onto β.

The priors and posterior for σ show (as expected) that learning is occurring about this identified parameter. The posteriors in the top panels of Figures 2 and 4 are concentrated near the true value used to generate the data.

This example illustrates an important point we have made previously: An informal comparison of priors and posteriors of structural parameters in DSGE models can be a useful way of investigating if learning is occurring about a parameter. However, such a comparison will not tell the researcher why the learning is occurring. Our figures show posteriors can differ from priors, even for parameters which do not enter the likelihood function, either when the parameter space is not variation free or through prior correlations with identified parameters. Since DSGE models will often exhibit such features, this illustration shows how caution should be taken when interpreting comparisons of priors with posteriors.

In Section 3.2, we recommended using an alternative indicator based on (20). If interest centers on identification issues relating to θ_1 then this indicator involved comparing $p(\theta_1|y)$ to $E_{\theta_2|y}[p(\theta_1|\theta_2)]$. The bottom panels of Figures 1 through 4 present such a comparison for β and σ for our two priors. Clearly our indicator is working well. For the unidentified parameter, $p(\beta|y)$ and $E_{\sigma,\gamma,\psi|y}[p(\beta|\sigma,\gamma,\psi)]$ are the same density. For the identified parameter, $p(\sigma|y)$ and $E_{\beta,\gamma,\psi|y}[p(\sigma|\beta,\gamma,\psi)]$ are massively different, indicating the parameter is identified.

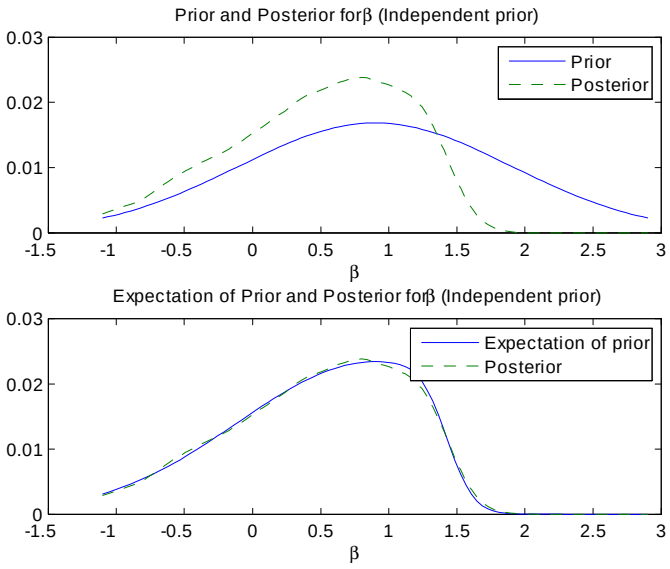

Figure 1

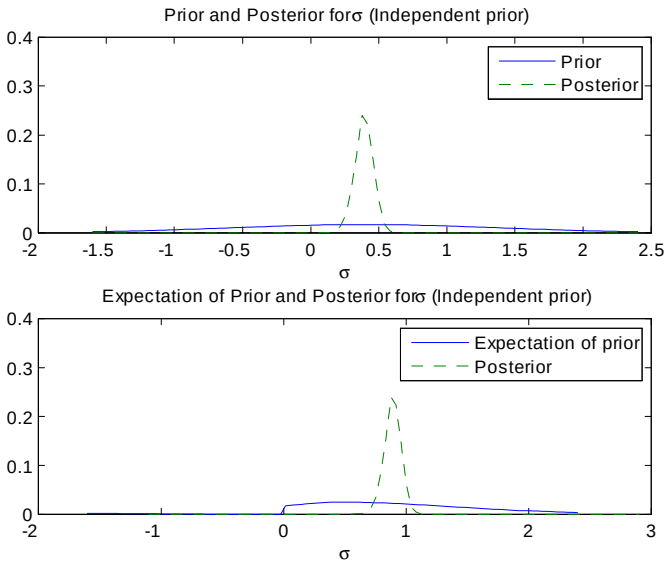

Figure 2

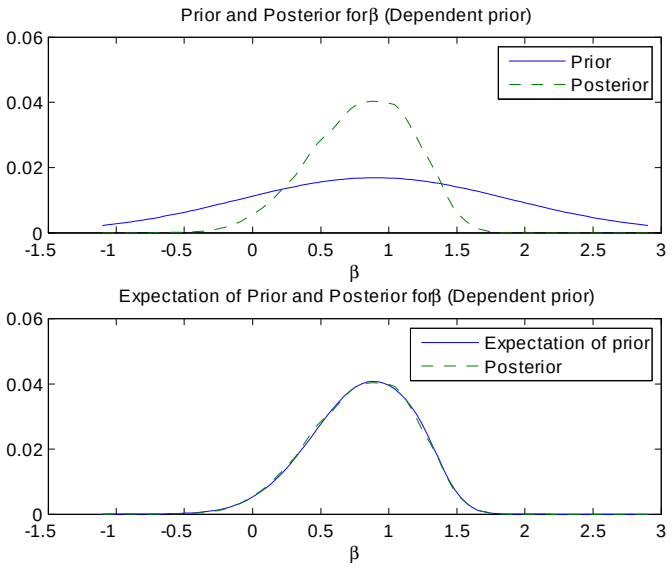

Figure 3

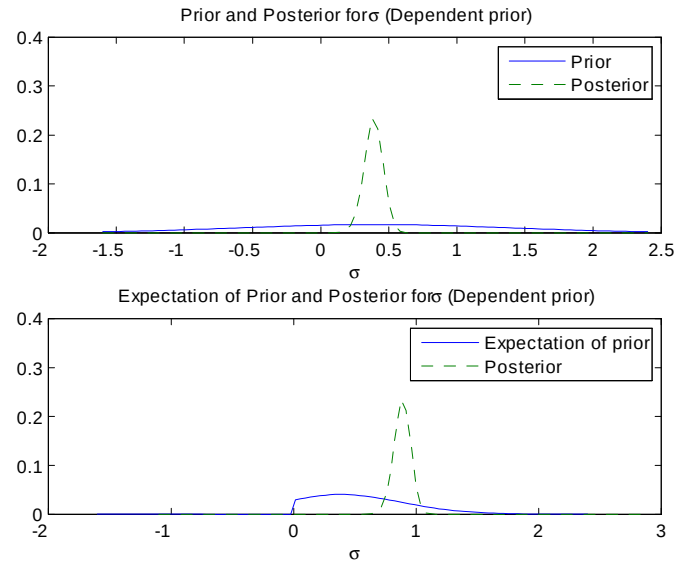

Figure 4

To illustrate our second Bayesian identification indicator, based on large sample derivations, Table 1 presents the posterior precisions of the parameters divided by T in the model using larger and larger data sets. Remember that the theoretical derivations underlying our second Bayesian identification indicator imply that the posterior precision of the identified parameters, σ, γ and ψ, should be increasing at a rate of T. But the posterior precision of the unidentified parameter,

β, will be increasing (if at all) at a slower rate. These properties can be clearly seen in Table 1. In contrast to the identified parameters, the posterior precision of β divided by sample size is heading towards zero. This result holds irrespective of whether the prior distribution of β depends on the other parameters or not. For example, in the case of independent priors, when $T = 10,000$ the average posterior precision of β is 2×10^{-4} as compared to 2.941, 0.909 and 0.689 for σ, γ and ψ, respectively.

Table 1: Posterior Precisions Divided by Sample Size				
Number of observations	σ	γ	ψ	β
	Independent Prior			
T=20	1.191	0.233	0.625	0.107
T=50	2.105	0.741	0.571	0.044
T=100	2.222	0.769	0.588	0.023
T=1,000	2.439	0.833	0.556	0.002
T=10,000	2.941	0.909	0.689	2×10^{-4}
	Dependent Prior			
T=20	1.163	0.225	0.625	0.263
T=50	1.539	1.333	0.556	0.131
T=100	1.961	0.741	0.588	0.072
T=1,000	2.941	1.052	0.526	0.008
T=10,000	2.778	1.102	0.476	8×10^{-4}

5.2 Example 2 (cont.): The NKPC model

For the reasons discussed in Section 4.1, our first Bayesian indicator of identification will not work reliably when we work with the NKPC and parameterize the model in terms of its structural parameters. However, our second Bayesian identification indicator of Section 4.2, based on the rate at which the posterior precision is updated, should still work. Accordingly, we use the NKPC to investigate the performance of this second identification indicator. We simulate artificial data under various assumptions about ρ_2. We first consider where ρ_2 is a constant, not a function of T; then $\rho_2 = 0$ gives the unidentified case; $\rho_2 \neq 0$ gives the identified case. We then consider the weakly identified case where ρ_2 is a function of T so that $\rho_2 = \delta/\sqrt{T}$ for various values of δ.

Unidentified and Identified Cases for Fixed Values of ρ_2 We generated one artificial data set of $10,000$ observations from the NKPC with $\beta = 0.6$, $\gamma = 0.9, \rho_1 = 0.3$ and the three values of $\rho_2 = 0, 0.3, 0.6$. In addition, u_t is i.i.d. $N(0, 0.25)$ and v_t is i.i.d. $N(0, 1)$. The priors for all

the parameters are normal: $N(0.5, 0.1I_4)$. This prior is chosen so that the prior means are a bit different (but not too different) from the true values and the prior variance is fairly informative. Note that, for β, a unique RE solution exists when $0 < \beta < 1$ and we impose this on the model. The prior is combined with the likelihood function based on the two equations for π_t and x_t. We use a random walk Metropolis-Hastings algorithm to do posterior simulation using the first T of the artificially generated observations for $T = 20, 50, 100, 1,000$ and $10,000$. In order to make these data sets as comparable as possible, we use the same seed for the random number generator for all DGPs.

Table 2 reports posterior precisions divided by T for the case where $\rho_2 = 0$. The average precision for γ and β go to zero, those for ρ_1 and ρ_2 do not. The average precision seems to go to zero faster for β than for γ. To illustrate what happens in the identified case, Table 3 gives results for $\rho_2 = 0.30$ and Table 4 for $\rho_2 = 0.60$.

Table 2: Posterior Precisions Divided by T (DGP: $\rho_2 = 0$)

Number of observations	ρ_1	ρ_2	γ	β
T=20	2.778	3.333	0.820	0.633
T=50	2.564	3.356	0.364	0.256
T=100	2.326	3.360	0.213	0.128
T=1,000	1.017	4.167	0.050	0.013
T=10,000	1.091	4.348	0.009	0.001

Table 3: Posterior Precisions Divided by T (DGP: $\rho_2 = 0.30$)

Number of observations	ρ_1	ρ_2	γ	β
T=20	2.671	3.910	0.839	1.331
T=50	2.417	2.621	0.389	0.702
T=100	2.219	2.377	0.231	0.428
T=1,000	1.401	1.670	0.039	0.062
T=10,000	1.148	1.121	0.018	0.030

Table 4: Posterior Precisions Divided by T (DGP: $\rho_2 = 0.60$)

Number of observations	ρ_1	ρ_2	γ	β
T=20	3.321	4.167	0.877	2.273
T=50	3.309	3.287	0.466	1.429
T=100	2.459	2.478	0.303	0.987
T=1,000	1.144	1.098	0.083	0.376
T=10,000	1.760	1.603	0.098	0.251

The Weakly Identified Case To investigate our results relating to weak identification, we use the same data generating process and prior as in the preceding section except that we set $\rho_2 = \delta/\sqrt{T}$ and use various values of δ to control the strength of identification. Note that this

implies there are different data generating processes for different choices of T. Thus, unlike the preceding artificial data illustrations, we cannot simulate a single artificial data set of size $T = 10,000$ and then take sub-sets of it to produce results for different sample sizes. The need to simulate completely different artificial data sets (using the same seed in the random number generator) for different sample sizes adds another cause for results to differ across sample size in the table below. Furthermore, for some sample sizes, our chosen values of δ lead to explosive AR(2) processes for the output gap. These are omitted in the table below.

Our aim is to investigate the quality of the asymptotic approximations obtained under the assumption of weak identification. The asymptotic approximations to the posterior precisions of β and γ are given in (42) and (43), respectively. Table 5 calculates these formulae, evaluated at the true values for the parameters. By definition, these will be the same for all T. We also estimate these posterior precisions using the artificially generated data and our random walk Metropolis-Hastings algorithm. Note here we give posterior precision, rather than posterior precision divided by T.

There are several reasons why the estimated results can differ from the asymptotic results including small sample effects, the fact that the estimated results are based on a single artificial data set and parameter uncertainty (i.e. the asymptotic results use the true values of the parameters, whereas the estimated results do not). But still, Table 7 indicates that our weak identification asymptotic results are at least roughly reliable. In general, estimated precisions are somewhat less than asymptotic precisions, which is due to parameter uncertainty. That is, the need to estimate parameters (as opposed to knowing their true values) slightly lowers precision.

It is interesting to note that the estimated precisions for γ tend to slightly increase with T but the opposite is true for β. This is consistent with findings in Table 2 for the unidentified NKPC, though that table gives average precisions rather than precision. Note that there we are finding the posterior precision of γ to rise with sample size (although this rise was not at a rate of T) to a greater extent than occurred for β. This finding is repeated in the estimated results in Table 5,

but does not occur with the asymptotic results.

Table 5 : Posterior Precisions: Asymptotic and Estimated				
Number of obs	Asymptotic β	Estimated β	Asymptotic γ	Estimated γ
	DGP: $\rho_2 = 0$			
T=20	22.046	18.490	18.301	15.919
T=50	22.046	17.226	18.301	18.269
T=100	22.046	17.399	18.301	21.875
T=1,000	22.046	18.797	18.301	55.211
T=10,000	22.046	18.169	18.301	122.415
	DGP: $\rho_2 = \frac{2}{\sqrt{T}}$			
T=20	36.092	37.202	29.961	17.781
T=50	36.092	34.224	29.961	19.578
T=100	36.092	28.939	29.961	21.730
T=1,000	36.092	22.175	29.961	42.282
T=10,000	36.092	18.907	29.961	102.344
	DGP: $\rho_2 = \frac{4}{\sqrt{T}}$			
T=50	78.229	69.292	64.940	19.875
T=100	78.229	60.524	64.940	27.105
T=1,000	78.229	29.316	64.940	36.346
T=10,000	78.229	20.448	64.940	77.017
	DGP: $\rho_2 = \frac{6}{\sqrt{T}}$			
T=100	148.458	96.640	123.239	27.921
T=1,000	148.458	37.861	123.239	37.861
T=10,000	148.458	21.591	123.239	61.395

5.3 Discussion

In this section, we have illustrated the properties of our proposed Bayesian identification indicators in simple models using artificially generated data. The advantages of such an approach are that insights can be obtained in a simple and well-understood context where the nature of the identification problem is clear. However, the researcher will often be working with much more complicated DSGE models. The question of how our large sample Bayesian identification indicator performs in such a case is addressed in Caglar et al (2012). This paper uses the popular medium size DSGE model of Smets and Wouters (2007) which involves 14 endogenous variables and 36 structural parameters. The identification indicator proposed reveals that this model is poorly identified. In particular, it finds that parameters relating to price and wage stickiness, the monetary policy rule and the steady states in some equations are poorly identified. These findings are similar to those presented in Iskrev (2010a). Caglar et al (2012) shows how our identification

indicator can be used to obtain sensible results in a substantive DSGE model.

6 Concluding Remarks

This paper has examined the identification of the parameters of DSGE models, in the light of the widespread concern in the literature that the parameters may be either not identified or only weakly identified. In purely forward looking models, with no lags, the coefficients of the expectational variables are not identified since they do not enter the likelihood function. In forward looking models with lags, identification is dependent on the assumed structure of the dynamics, making it vulnerable to the Sims (1980) critique of 'incredible' identifying restrictions. In more complicated models with unobserved variables and no analytical solution, it is difficult to determine whether the models are identified. When the DSGE models are estimated by Bayesian methods, this lack of identification may not be evident since the posterior may differ from the prior even if the parameter is not identified and the posterior for unidentified parameters may also be updated as the sample size increases. These properties have been demonstrated both analytically and numerically, using familiar examples of unidentified or weakly identified rational expectations DSGE models.

We propose two Bayesian identification indicators. The first involves comparing the marginal posterior of a parameter with the posterior expectation of the prior for that parameter conditional on the other parameters. This can be computed as part of the MCMC estimation of a DSGE model using whatever real data set the researcher is working with. However, this indicator can be applied only in situations where parameters can be partitioned into a set that are known to be identified and another set for which identification is uncertain. This may not be possible when the researcher is working with the structural parameters of a DSGE model. Our second Bayesian indicator is more generally applicable and considers the rate at which the posterior precision gets updated as the sample size (T) is increased. For identified parameters the posterior precision rises with T, whilst for an unidentified or weakly identified parameter its posterior

precision may be updated but its rate of update will be slower than T. This result assumes that the identified parameters are $\sqrt{T}$-consistent, but similar differential rates of updates for identified and unidentified parameters can be established in the case of weak (or super) consistent estimators. This suggests a strategy where the researcher simulates larger and larger data sets and observes the behavior of the posterior as sample size increases.

We present an empirical illustration which shows the effectiveness of the first Bayesian identification indicator, in cases where it is applicable. Further empirical illustrations show the usefulness of our second Bayesian identification indicator, both for checking for the presence and the strength of identification.

Acknowledgements

We are grateful for helpful comments from the editor, an associate editor and two referees as well as from Cheng Hsiao, Dale Poirier, Ivan Jeliazkov and participants at seminars at the University of California - Irvine, American, Montreal and York Universities and the NBER summer institute on DSGE models.

Appendix

Proof of Proposition 1

The posterior density of θ, evaluated at θ_0, can be written as

$$p(\theta_0 \,|y\,) = \exp\{\ln[p(\theta_0)] + \ell_T(\theta_0) - \ln(p(y))\}. \tag{44}$$

Under Assumption 3, the prior density is given by (24). To obtain $\ell_T(\alpha(\theta_0)) = \ell_T(\theta_0)$, we first note that under Assumption 2 using standard results from $\sqrt{T}$-consistent ML estimators, we have (noting that $\partial \ell_T(\hat{\alpha}_T)/\partial \alpha_s = 0$)

$$\ell_T(\alpha_0) = \ell_T(\hat{\alpha}_T) + \frac{1}{2}\sum_{s=1}^{k}\sum_{r=1}^{k}\frac{\partial^2 \ell_T(\hat{\alpha}_T)}{\partial \alpha_r \partial \alpha_s}(\alpha_{0r} - \hat{\alpha}_{rT})(\alpha_{0s} - \hat{\alpha}_{sT}) + O_p(T^{-1/2}). \tag{45}$$

Further, under Assumption 1 and by the mean value theorem for all pairs θ^a and $\theta^b \in \Theta^0$ we have (see, for example, Jennrich (1969, lemma 3), and Davidson (1994, p. 340))

$$\alpha_s(\theta^a) - \alpha_s(\theta^b) = \sum_{i=1}^{p}\frac{\partial \alpha_s(\mathring{\theta})}{\partial \theta_i}(\theta_i^a - \theta_i^b), \tag{46}$$

where $\mathring{\theta} \in \Theta^0$ is a point on the line segment joining θ^a and θ^b. Note that Jennrich formulation of the mean value theorem allows θ^a or θ^b to depend on y, as will be required in our application. Although θ is not identified, there exists $\hat{\theta}_T$ such that $\hat{\alpha}_T = \alpha(\hat{\theta}_T)$. Using the mean value theorem in (46) for $\hat{\theta}_T$ and $\theta_0 \in \Theta^0$, we have

$$\alpha_s(\theta_0) - \alpha_s(\hat{\theta}_T) = \sum_{i=1}^{p} \frac{\partial \alpha_s(\mathring{\theta})}{\partial \theta_i}(\theta_{0i} - \hat{\theta}_{iT}),$$

where $\mathring{\theta}$ is on the line segment joining $\hat{\theta}_T$ and θ_0. Using this result in (45) we have (recalling that $\alpha_0 = \alpha(\theta_0)$)

$$\ell_T(\theta_0) = \ell_T(\hat{\theta}_T) + \frac{1}{2}\sum_{s=1}^{k}\sum_{r=1}^{k}\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial^2 \ell_T(\hat{\alpha}_T)}{\partial \alpha_r \partial \alpha_s} \frac{\partial \alpha_s(\mathring{\theta})}{\partial \theta_i} \frac{\partial \alpha_r(\mathring{\theta})}{\partial \theta_j}(\theta_{0i} - \hat{\theta}_{iT})(\theta_{0j} - \hat{\theta}_{jT}) + O_p(T^{-1/2}).$$

Written more compactly we have

$$\ell_T(\theta_0) = \ell_T(\hat{\theta}_T) - \frac{T}{2}(\theta_0 - \hat{\theta}_T)' \mathring{S}_T (\theta_0 - \hat{\theta}_T) + O_p(T^{-1/2}),$$

where $\mathring{S}_T = \mathring{R}' Q_T \mathring{R}$ and $\mathring{R} = R(\mathring{\theta})$, for all θ_0 and $\mathring{\theta} \in \Theta^0$. Using this result and (24) in (44) we have

$$p(\theta_0 \,|y) \propto \exp\{-\frac{1}{2}(\theta_0 - \underline{\theta})' \underline{\mathrm{H}}(\theta_0 - \underline{\theta}) - \frac{T}{2}(\theta_0 - \hat{\theta}_T)' \mathring{S}_T (\theta_0 - \hat{\theta}_T) + O_p(T^{-1/2})\}, \tag{47}$$

which establishes the desired result.

Proof of Proposition 2

Note that $S_T = R_T' Q_T R_T = R_T' U_T' U_T R_T = P_T' P_T$, where $P_T = U_T R_T$, and consider the following partitioned form of S_T:

$$S_T = \begin{pmatrix} P_{11,T}' P_{11,T} + P_{21,T}' P_{21,T} & P_{11,T}' P_{12,T} + P_{21,T}' P_{22,T} \\ P_{12,T}' P_{11,T} + P_{22,T}' P_{21,T} & P_{12,T}' P_{12,T} + P_{22,T}' P_{22,T} \end{pmatrix}.$$

Then using (28) we have

$$\bar{H}_{11,T} = TP_{11,T}' P_{11,T} + TP_{21,T}' P_{21,T} + \underline{\mathrm{H}}_{11} - \left(TP_{11,T}' P_{12,T} + TP_{21,T}' P_{22,T} + \underline{\mathrm{H}}_{12}\right) \times$$

$$\left(TP_{12,T}' P_{12,T} + TP_{22,T}' P_{22,T} + \underline{\mathrm{H}}_{22}\right)^{-1} \left(TP_{12,T}' P_{11,T} + TP_{22,T}' P_{21,T} + \underline{\mathrm{H}}_{21}\right),$$

and under weak identification we obtain

$$\bar{H}_{11,T} = TP'_{21,T}P_{21,T}+G_{11,T}-\left(TP'_{21,T}P_{22,T}+G_{12,T}\right)\times$$
$$\left(P'_{22,T}P_{22,T}+T^{-1}G_{22,T}\right)^{-1}\left(P'_{22,T}P_{21,T}+T^{-1}G_{21,T}\right) \tag{48}$$

where

$$G_{11,T}= D'_{11,T}D_{11,T}+\underline{\mathrm{H}}_{11},\ G_{12,T}= D'_{11,T}D_{12,T}+\underline{\mathrm{H}}_{12}, \tag{49}$$

$$G_{21,T}= D'_{12,T}D_{11,T}+\underline{\mathrm{H}}_{21},\ G_{22,T}= D'_{12,T}D_{12,T}+\underline{\mathrm{H}}_{22}. \tag{50}$$

Note that matrices $G_{ij,T}$, for $i,j=1,2$, and $P_{22,T}$ are all bounded in T. Then

$$\begin{aligned}\left(P'_{22,T}P_{22,T}+T^{-1}G_{22,T}\right)^{-1} &= \left[P'_{22,T}\left(I_{p_2}+T^{-1}P'^{-1}_{22,T}G_{22,T}P^{-1}_{22,T}\right)P_{22,T}\right]^{-1}\\ &= P^{-1}_{22,T}\left(I_{p_2}+T^{-1}P'^{-1}_{22,T}G_{22,T}P^{-1}_{22,T}\right)^{-1}P'^{-1}_{22,T}\\ &= P^{-1}_{22,T}\left[I_{p_2}-T^{-1}P'^{-1}_{22,T}G_{22,T}P^{-1}_{22,T}+O\left(T^{-2}\right)\right]P'^{-1}_{22,T}\\ &= A_{22,T}-T^{-1}A_{22.T}G_{22,T}A_{22,T}+O\left(T^{-2}\right)\end{aligned}$$

where $A_{22,T}=\left(P'_{22,T}P_{22,T}\right)^{-1}$. Using this result in (48) we have

$$\bar{H}_{11,T}= TP'_{21,T}P_{21,T}+G_{11,T}-\left(TP'_{21,T}P_{22,T}+G_{12,T}\right)\times$$
$$\left[A_{22,T}-T^{-1}A_{22,T}G_{22,T}A_{22,T}+O\left(T^{-2}\right)\right]\left(P'_{22,T}P_{21,T}+T^{-1}G_{21,T}\right).$$

However, since $P_{22,T}$ is a square full rank matrix, then $P_{22,T}\left(P'_{22,T}P_{22,T}\right)^{-1}P'_{22,T}=I_{p_2}$, and the terms of order T cancel out and we are left with

$$\bar{H}_{11,T}= G_{11,T}-P'_{21,T}P'^{-1}_{22,T}G_{21,T}-G_{12,T}P^{-1}_{22,T}P_{21,T}+P'_{21,T}P'^{-1}_{22,T}G_{22,T}P^{-1}_{22,T}P_{21,T}+O(T^{-1}).$$

Substituting for $G_{ij,T}$ from (49) and (50), and after some algebra we obtain (34).

References

An, S. and F. Schorfheide (2007) Bayesian analysis of DSGE models, Econometric Reviews, 26(2-4), 113-172.

Andrews, D. and X. Cheng (2011) Estimation and inference with weak, semi-strong and strong identification, Cowles Foundation Discussion Paper 1773R.

Andrle, M. (2010) A note on identification patterns in DSGE models, ECB Working Paper 1235.

Benati, L. (2010) Are policy counterfactuals based on structural VARs reliable?, ECB Working Paper 1188.

Berger, J. (1985) Statistical Decision Theory and Bayesian Analysis, second edition, Springer-Verlag: New York.

Binder, M. and M.H. Pesaran (1995) Multivariate rational expectations models and macroeconometric modelling: A review and some new results. In M.H. Pesaran & M. Wickens (eds.), Handbook of Applied Econometrics: Macroeconomics, pp. 139-187. Oxford: Basil Blackwell.

Binder, M. and M.H. Pesaran (1997) Multivariate linear rational expectations models: characterization of the nature of the solutions and their fully recursive computation, Econometric Theory, 13, 877-888.

Caglar, E., J.S. Chadha and K. Shibayama (2012) Bayesian estimation of DSGE models: is the workhorse model identified? Koc-University-Tusiad Economic Research Forum, working paper number 1205.

Canova, F. (2007) Comment on Bayesian analysis of DSGE models by S. An and F. Schorfheide, Econometric Reviews, 26, 113-172.

Canova, F. and Sala, L. (2009) Back to square one: identification issues in DSGE models, Journal of Monetary Economics, 56, 431-449.

Clarida, R., J. Gali and M. Gertler (1999) The science of monetary policy: a new Keynesian perspective, Journal of Economic Literature, 37, 1661-1707.

Cochrane, J. (2011) Determinacy and identification with Taylor rules, Journal of Political Economy, 119(3), 565-615.

Davidson, J. (1994), Stochastic Limit Theory, Oxford University Press: Oxford.

Dees, S., M.H. Pesaran, L.V. Smith and R.P. Smith (2009) Identification of New Keynesian Phillips Curves from a Global Perspective, Journal of Money Credit and Banking, 41(7).

DeJong, D., B. Ingram and C. Whiteman (2000) A Bayesian approach to dynamic macroeconomics, Journal of Econometrics, 98, 203-223.

Drèze, J. (1976) Bayesian limited information analysis of the simultaneous equations model, Econometrica, 44, 1045–1075.

Drèze, J. and J.-F. Richard (1983). Bayesian analysis of simultaneous equation systems, in Z. Grilliches and M.D. Intrilligator, (eds.), Handbook of Econometrics, Vol. 1.

Hamilton, J. (1994) Time Series Analysis, Princeton University Press: Princeton.

Hoogerheide, L., F. Kleibergen and H. van Dijk (2007) Natural conjugate priors for the instrumental variables regression model applied to the Angrist-Krueger data, Journal of Econometrics, 138, 63-103.

Hsiao, C. (1983) Identification, Chapter 4 in Handbook of Econometrics, Volume 1, Z. Griliches and M.D. Intriligator (Eds.), North-Holland.

Iskrev, N. (2010a) Local identification in DSGE models, Journal of Monetary Economics, 57, 189-202.

Iskrev, N. (2010b) Evaluating the strength of identification in DSGE models. An a priori approach, manuscript.

Iskrev N. and M. Ratto (2010) Analysing identification issues in DSGE models, manuscript,

Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators, The Annals of Mathematical Statistics, 40, 663-643.

Kadane, J.B. (1974) The role of identification in Bayesian theory. In S.E. Fienberg & A. Zellner (eds.), Studies in Bayesian Econometrics and Statistics, pp. 175-191. Amsterdam: North-

Holland.

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.

Kleibergen F. and S. Mavroeidis (2010) Identification-pathology robust Bayesian analysis of DSGE models, manuscript.

Kleibergen, F. and H. van Dijk (1998) Bayesian simultaneous equations analysis using reduced rank structures, Econometric Theory, 14, 699-744.

Kleibergen, F. and E. Zivot (2003). Bayesian and classical approaches to instrumental variable regression, Journal of Econometrics, 114, 29-72.

Komunjer, I. and S. Ng (2011). Dynamic Identification of DSGE Models, Econometrica, 79, 1995-2032.

Koop, G. and D. Poirier (1997) Learning about the cross-regime correlation in switching regression models, Journal of Econometrics, 78, 217-227.

Kydland, F. and E. Prescott (1996) The computational experiment: an econometric tool, Journal of Economic Perspectives, 10, 69-85.

Lindley, D.V. (1971) Bayesian Statistics: A Review, Philadelphia, SIAM.

Mavroeidis, S. (2005) Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve, Journal of Money, Credit, and Banking, 37, 421-448.

McCallum, B. (1979) On the observational inequivalence of classical and Keynesian models, Journal of Political Economy, 87, 395-402.

Nason, J. and G.W. Smith (2008) Identifying the new Keynesian Phillips curve, Journal of Applied Econometrics, 23, 525-51.

Pesaran, M.H. (1981) Identification of Rational Expectations Models, Journal of Econometrics, 16, 375-98.

Pesaran, M.H. (1987) The Limits to Rational Expectations, Blackwell: Oxford.

Poirier, D.J. (1998) Revising beliefs in nonidentified models, Econometric Theory 14, 483-509.

Pudney. S. (1982) The identification of rational expectations models under structural neutrality, Journal of Economic Dynamics and Control, 4, 117-121.

Rubio-Ramirez, J., D. Waggoner and T. Zha (2010) Structural vector autoregressions: Theory of identification and algorithms for inference, Review of Economic Studies, 77, 665-696.

Sargent T.J. (1976) The observational equivalence of natural and unnatural rate theories of macroeconomics, Journal of Political Economy, 84, 631-40.

Sims, C. (1980) Macroeconomics and reality, Econometrica, 48, 1-48.

Sims, C. (2002) Solving linear rational expectations models, Computational Economics, 20, 1–20.

Smets, F. and R. Wouters (2003) An estimated dynamic stochastic general equilibrium model of the Euro area, Journal of the European Economic Association, 1, 1123-1175.

Smets, F. and R. Wouters (2007) Shocks and frictions in US business cycles: A Bayesian DSGE approach, American Economic Review, 97, 588-606.

Stock, J., J. Wright and M. Yogo. (2002) A survey of weak instruments and weak identification in generalized method of moments, Journal of Business and Economic Statistics, 20 , 518-529.

Wallis, K.F. (1980) Econometric implications of the rational expectations hypothesis, Econometrica, 48, 49-73.