Econometric Issues Arising from DSGE Models *

Martin Fukac[†] Adrian Pagan[‡] and Vlad Pavlov[§]

May 30, 2006

Contents

1 Introduction **2**

2 Some preliminaries **5**

2.1 Solving the Model 5

2.2 An Example of Re-formulating the Euler Equations 7

3 Identification **10**

4 Estimation Issues **13**

4.1 Analysing Systems Estimators of a DSGE Model 15

4.2 Euler Equation Estimation 22

5 Evaluation **23**

5.1 Examining the Euler Equations 24

5.2 How Well Does the Model Track the Data? 24

6 References **27**

*Research of the second author was supported by ESRC Grant No. 000 23-0244.

† Centre for Economic Research and Graduate Education in Prague and Czech National Bank

‡ Australian National University and Queensland University of Technology.

§ Queensland University of Technology

1 Introduction

DSGE models are becoming widely used in both academic and central bank research. In the case of the latter there is naturally great interest in the ability of the models to adequately represent the data. Moreover, a natural question that arises is whether it is possible to recover reliable estimates of the parameters of the model from such data sets. At one extreme this question is about identification i.e. whether it is possible to recover unique estimates of the parameters even with an infinite sample. A literature has emerged on this in the DSGE model context that we attempt to elucidate in section 3 of the paper. By producing a classification of identification issues we see that there are situations in which it may be impossible to identify the DSGE model parameters but that this is not important for policy use.

Increasing attention has also been paid to methods for estimating their parameters from data sets, as well as the ability of these models to represent selected characteristics of the data. Estimation of the model parameters was initially done by applying instrumental variable estimators (GMM) to the Euler equations underlying them. This approach aimed to account for the presence of endogenous variables and future expectations that appear in these relations. But this strategy fell out of favour as simulations showed that the methods for summarizing the uncertainty in parameter estimators were unreliable due to weak instruments - see Mavroeidis (2004) for example. Although the parameters were rarely unidentified, very large samples might be needed to produce useful inferences. This led to the development of a literature that tried to produce better performance for the indices of uncertainty when samples of the size encountered in macroeconomics were only available e.g. Staiger and Stock (1997), Poskitt and Skeels(2005) and Andrews and Stock (2005). That literature is still evolving, particularly if more than two endogenous variables are involved in an Euler equation. An alternative to improving the finite sample inference has been to select other estimators such as MLE. But the fact that MLE is an instrumental variables estimator in the simultaneous equations set up- see Hausman (1975) and Hendry (1976) - means that the issue cannot be that one is using a different estimator in that case but rather that there has just been a different method of constructing instruments. Indeed this has been shown in examples such as Fuhrer and Olivei (2004) where one could improve on standard IV by utilizing the forecasts from a derived reduced form rather than using the

OLS estimates, as is done with standard IV estimators such as 2SLS.[1] This method has a long history. In the context of simultaneous equations Brundy and Jorgenson (1971) called these the FIVE and LIVE estimators. Such estimators used as instruments the predictions from the derived reduced form after imposing either all the restrictions on it (FIVE) or the sub-set of them stemming from the equation being estimated (LIVE) as instruments (since any set of weights attached to the instruments produce consistent estimators we are free to choose these weights however we wish) and Fuhrer and Olivei (2004) essentially applied these estimators when estimating NKPM systems.

Whilst MLE has often been seen to produce better results than GMM in estimating DSGE models it has also been criticized by authors such as Ahn and Schorfheide, 2005, who refer to the "dilemma of absurd parameter estimates" found when applying MLE to DSGE models. They argue that Bayesian methods often produce more acceptable parameter estimates, and there is little doubt that these methods have become increasingly popular. Indeed, some such as Sims (2005) believes that the failure to use such methods is a reason why macro-econometric modelling in central banks can be greatly improved. We therefore look at whether Bayesian methods really are superior in section 4 and whether one should apply systems estimators when one is unsure of the specification of the complete system. Since the vast majority of DSGE models are driven by a single $I(1)$ common factor we outline a way of transforming these models to such a form that estimation could proceed in a single-equation fashion by using the Euler equations that underlie these models. Such analysis sometimes gives insights into specification issues with DSGE models, since it does not require a correct specification of the complete system. We examine estimation issues in the context of an open-economy model proposed by Lubik and Schforheide (2005) and fitted to U.K. data. Throughout this paper we use this model to illustrate our arguments.

As well as getting values for the parameters that are inputs into a DSGE model, central bank policy analysts are clearly going to be interested in the outputs. Much early work on these models looked at only a few outputs, such as the variances and covariances of a select group of variables. But, as the models have been increasingly used by central bank researchers, there has been a growing interest in evaluating the models in other ways. Of course

[1] It should be observed that FIML would use restrictions that shocks are uncorrelated whereas most proposed IV estimators do not, but the difference in performance do not seem to stem from this difference.

there have been many suggestions of how to evaluate DSGE models other than using a few moments e.g. Canova et al (1994) suggested that one examine the implied VARs, while others have formalized such tests in different ways e.g. del Negro et al. (2004). In Fukac and Pagan (2006) we proposed a structured way of doing this testing. Here we augment the methods proposed there in two ways. First, we argue that, once the model parameters have been estimated, one should enquire into the ability of the model to satisfy the assumptions used in the derivation of the Euler equations. Normally this is complicated by the presence of I(1) factors driving the models, but using the transformation mentioned above enables us to circumvent that problem. Second we look at the tracking performance of the model. Deriving the predicted path of variables implied by a DSGE model is complex. In most instances researchers have not been willing to assume that the DSGE model will be an accurate rendition of the economy and have thought that there is a wedge between the model outputs and the data. This has sometimes been handled by adding an observation shock on to the model output e.g. Altug (1989) and Ireland (2004). If one assumes that it is zero then one is effectively replacing the model variables with data. To avoid that situation it is necessary to make some assumption about the nature of the observation shocks and how they relate to the model shocks. If one assumes that the observation shocks are uncorrelated with the model shocks then one can utilize the Kalman filter to extract the predictions made by the model about the data, but this a strong assumption. There seems no reason to think that they should be uncorrelated. We therefore follow Watson (1993) and propose that one allow for a correlation between the shocks, with the extent of the correlation being controlled by the need to maintain the model covariance characteristics while getting as close to the data as possible on some specified dimension e.g. this might involve a covariance matrix or a spectrum over certain frequencies (Watson used these). Rather than solve the problem the way Watson did we formulate the problem as one that involves an application of the Kalman filter, but with the model and observation shocks being correlated. We illustrate the method with the Lubik and Schorfheide model.

2 Some preliminaries

2.1 Solving the Model

Consider the following stylized version of an economic system of the form

$$B_0 y_t = B_1 y_{t-1} + Dx_t + CE_t y_{t+1} + u_t \tag{1}$$

where y_t is a vector of $n \times 1$ variables, x_t is a set of observable and u_t a set of unobservable shocks. There are p observable and less than or equal to n unobservable shocks. If there were more than n of the latter we would be looking at factor models and we side step that issue in this paper. By observable we will mean that the shocks can be recovered from a statistical model. By unobservable we will mean that the shocks are defined by the economic model. Later however we will allow for another class of unobservable shocks that are not defined by an economic model but which are added simply to produce a better tracking of the data; the latter we will call observation shocks.

To find a representation that eliminates the expectations we follow Binder and Pesaran (1995) and write $\xi_t = y_t - Py_{t-1}$, which is then substituted to obtain

$$\begin{aligned} B_0(\xi_t + Py_{t-1}) &= B_1 y_{t-1} + Dx_t + CE_t(\xi_{t+1} + Py_t) + u_t \\ &= B_1 y_{t-1} + Dx_t + CE_t(\xi_{t+1} + P(\xi_t + Py_{t-1})) + u_t \\ &= B_1 y_{t-1} + Dx_t + CE_t(\xi_{t+1}) + CP\xi_t + CP^2 y_{t-1} + u_t, \end{aligned}$$

so that we need $B_0P - B_1 - CP^2 = 0$ to eliminate the y_{t-1} term and to produce

$$B_0\xi_t = CE_t(\xi_{t+1}) + Dx_t + CP\xi_t + u_t.$$

This then implies

$$\begin{aligned} \xi_t &= (B_0 - CP)^{-1}CE_t(\xi_{t+1}) + (B_0 - CP)^{-1}Dx_t + (B_0 - CP)^{-1}u_t \\ &= \Pi_1 E_t \xi_{t+1} + \Pi_2 x_t + \Pi_3 u_t, \end{aligned}$$

and the solution to the latter would be

$$\xi_t = \sum_{j=0}^{\infty} \Pi_1^j E_t(\Pi_2 x_{t+j} + \Pi_3 u_{t+j}).$$

Thus

$$y_t = Py_{t-1} + \sum_{j=0}^{\infty} \Pi_1^j (\Pi_2 E_t x_{t+j} + \Pi_3 E_t u_{t+j})$$

and we need to specify the nature of x_t and u_t. In the case where the x_t and u_t are AR(1) processes we would get a Vector Autoregression with Exogenous Variables (VARX) system for y_t :

$$y_t = Py_{t-1} + D_0 x_t + G_0 u_t. \tag{2}$$

Now the economic theory here is a statement about P. D_0 involves both P and a statistical process for x_t, with the latter capable of being inferred from the data independently of the model. This is not so for u_t since, although one might estimate a process for it, this can only be done by estimating the complete model.

The discussion above has proceeded as if y_t was a stationary random variable. Where the situation become more complex is if the observed data is an $I(1)$ process, as this means that the DSGE model must be driven by an $I(1)$ factor. In that instance the model variables are generally measured as a deviation of the observed variables from this factor i.e. $\tilde{y}_t = y_t - a_t$, where A_t is the $I(1)$ factor and the solution is then for $\tilde{y}_t$.[2] For this reason (2) is then estimated with a Kalman filter approach, with the observations y_t being related to the model variables as $y_t = \tilde{y}_t + a_t$. If one then wanted to estimate the DSGE model parameters using the Euler equations in (1) there would be a complication since

$$B_0 \tilde{y}_t = B_1 \tilde{y}_{t-1} + Dx_t + CE_t \tilde{y}_{t+1} + u_t \tag{3}$$

so that, substituting y_t for the $\tilde{y}_t$ using $y_t = \tilde{y}_t + v_t$, would result in

$$\begin{aligned} B_0 y_t &= B_1 y_{t-1} + Dx_t + CE_t(y_{t+1}) + B_0 v_t - B_1 v_{t-1} \\ &\quad + CE_t v_{t+1} + Gu_t. \end{aligned}$$

Now $E_t v_{t+1}$ is often either v_t if (say) A_t is a random walk or zero (for those y_t that are I(0)) and so it is clear that the error term in these Euler equations is a composite one and therefore it follows a Vector MA process. This poses estimation difficulties.

[2] In most instances one of the u_t is an $I(1)$ shock. If there are more shocks then variables the discussion of this paper needs to be re-considered. But few DSGE models have that property at the moment.

Clearly it would be best if we could estimate $\tilde{y}_t$ and use these observations in the original Euler equations (1). One way to do this is to measure $\overline{y}_t = y_t - y_t^p$, where y_t^p is the permanent component of y_t. A reason for doing this is that the permanent component of a series can be estimated using standard co-integration methods. Once estimated $\overline{y}_t$ can be formed and used to estimate the Euler equations. We give an example of how to perform the transformation in the next sub-section.

2.2 An Example of Re-formulating the Euler Equations

The model we utilize is in Lubik and Schorfheide (2005). It is a small four equation model of an open economy. The IS curve describes output y_t and is specified in their paper in terms of the transformed variable $\tilde{y}_t = y_t - A_t$. The same transform is applied to (unobservable) foreign output to produce $\tilde{y}_t^*$.

$$\begin{aligned} \tilde{y}_t &= E_t\tilde{y}_{t+1} - [\tau + \theta](R_t - E_t\pi_{t+1}) - \alpha(\tau+\theta)\rho_q\Delta q_t \\ &\quad -\rho_A\Delta A_t - \frac{\theta}{\tau}(1-\rho_{y^*})\tilde{y}_t^*, \qquad (4) \\ 0 &< \alpha < 1, \tau^{-1} > 0 \end{aligned}$$

In the IS equation q_t is the observable terms of trade, A_t is the log level of (unobservable) technology, α is the import share, τ is the intertemporal elasticity of substitution and $\theta = \alpha(2-\alpha)(1-\tau)$.

Their open economy Phillips curve is

$$\begin{aligned} \pi_t &= \beta E_t\pi_{t+1} - \alpha(1-\beta\rho_q)\Delta q_t + \frac{\kappa}{\tau+\theta}\tilde{y}_t \\ &\quad + \frac{\kappa\theta}{\tau[\tau+\theta]}\tilde{y}_t^*, \qquad (5) \end{aligned}$$

where π_t is the domestic inflation rate, β is the discount factor and κ is a "price stickiness" parameter.

The exchange rate equation is

$$\Delta e_t - \pi_t = -(1-\alpha)\Delta q_t - \pi_t^*, \qquad (6)$$

where e_t is the log of the exchange rate and π_t^* is the (unobservable) foreign inflation rate.

The policy rule for the nominal interest rate (R_t) is

$$R_t = \rho_R R_{t-1} + (1-\rho_R)\left[\psi_1 \pi_t + \psi_2 \tilde{y}_t + \psi_3 \Delta e_t\right] + \varepsilon_t^R. \quad (7)$$

Exogenous variables evolve as

$$\begin{aligned}
\Delta q_t &= \rho_q \Delta q_{t-1} + \varepsilon_t^q. && (8)\\
\Delta A_t &= \rho_a \Delta A_{t-1} + \varepsilon_t^a && (9)\\
\tilde{y}_t^* &= \rho_{y^*} \tilde{y}_{t-1}^* + \varepsilon_t^{y^*}, && (10)\\
\pi_t^* &= \rho_{\pi^*} \pi_{t-1}^* + \varepsilon_t^{\pi^*}, && (11)
\end{aligned}$$

Now, as we have mentioned above, we want to transform these equations to incorporate variables that are deviations from permanent components rather than from the unobservable factor A_t. Therefore, define $\overline{y}_t = y_t - y_t^p$, where the "p" indicates the permanent component. Then

$$\begin{aligned}
\tilde{y}_t &= y_t - A_t = y_t - y_t^p + y_t^p - A_t \\
&= y_t - y_t^p + A_t^p - A_t \\
&= \overline{y}_t + \frac{\rho_a}{1-\rho_a} \Delta A_t
\end{aligned}$$

given the nature of A_t (see Morley(2002)). Thus the IS equation becomes

$$\begin{aligned}
\overline{y}_t + \frac{\rho_a}{1-\rho_a}\Delta A_t &= E_t(\overline{y}_{t+1} + \frac{\rho_a}{1-\rho_a}\Delta A_{t+1}) - [\tau+\theta](R_t - E_t\pi_{t+1}) - \alpha\theta\rho_q\Delta q_t \\
&\quad -\rho_a \Delta A_t - \frac{\theta}{\tau}(1-\rho_{y^*})\tilde{y}_t^*.
\end{aligned}$$

Collecting terms gives

$$\begin{aligned}
\overline{y}_t &= E_t(\overline{y}_{t+1}) - [\tau+\theta](R_t - E_t\pi_{t+1}) - \alpha\theta\rho_q\Delta q_t \\
&\quad -2\rho_a\Delta A_t - \frac{\theta}{\tau}(1-\rho_{y^*})\tilde{y}_t^* \\
&= E_t(\overline{y}_{t+1}) - [\tau+\theta](R_t - E_t\pi_{t+1}) - \alpha\theta\rho_q\Delta q_t \\
&\quad -\frac{2\rho_a(1-\rho_a)}{(1-\rho_a L)}\Delta y_t^p - \frac{\theta}{\tau}(1-\rho_{y^*})\tilde{y}_t^*, \quad (12)
\end{aligned}$$

since

$$\Delta y_t^p = \Delta A_t^p = \frac{\varepsilon_t^a}{1-\rho_a} = \frac{(1-\rho_a L)\Delta A_t}{(1-\rho_a)}.$$

In a similar way we can re-write the Phillips curve and interest rate rules as[3]

$$\begin{aligned}\pi_t &= \beta E_t \pi_{t+1} - \alpha(1-\beta\rho_q)\Delta q_t + \frac{\kappa}{\tau+\theta}\overline{y}_t \\ &\quad + \frac{\kappa\rho_a}{(\tau+\theta)(1-\rho_a)}\Delta A_t + \frac{\kappa\theta}{\tau[\tau+\theta]}\tilde{y}_t^* \end{aligned} \tag{13}$$

$$\begin{aligned}R_t &= \rho_R R_{t-1} + (1-\rho_R)\left[(\psi_1+\psi_3)\pi_t + \psi_2\overline{y}_t + \psi_3(\Delta e_t - \pi_t)\right] \\ &\quad + \frac{(1-\rho_R)\rho_a\Delta A_t}{(1-\rho_a)} + \varepsilon_t^R\end{aligned}$$

The solution to this model has the form

$$z_t = \Gamma_1 R_{t-1} + \Gamma_2 \Delta q_t + \Gamma_3 v_t$$

where $z_t = \begin{bmatrix} \overline{y}_t \\ \pi_t \\ \Delta e_t - \pi_t \\ R_t \end{bmatrix}$ and $v_t = \begin{bmatrix} \Delta A_t \\ \tilde{y}_t^* \\ \pi_t^* \\ \varepsilon_t^R \end{bmatrix}$. Since $v_t = \Phi v_{t-1} + \varepsilon_t$ we will have

$$\begin{aligned} z_t &= \Gamma_1 R_{t-1} + \Gamma_2 \Delta q_t + \Gamma_3 \Phi v_{t-1} + \Gamma_3 \varepsilon_t \\ &= \Gamma_1 R_{t-1} + \Gamma_2 \Delta q_t + \Gamma_3 \Phi \Gamma_3^{-1}(z_{t-1} - \Gamma_1 R_{t-2} - \Gamma_2 \Delta q_{t-1}) + \Gamma_3 \varepsilon_t \end{aligned}$$

and so the equation for $\overline{y}_t = S z_t$, where $S = \begin{matrix} 1 & 0 & 0 & 0 \end{matrix}$ become

$$\overline{y}_t = w_t'\delta + S\Gamma_2 \Delta q_t + S\Gamma_3 \varepsilon_t$$

where $w_t' = \begin{bmatrix} R_{t-1} & \overline{y}_{t-1} & \pi_{t-1} & \Delta e_{t-1} - \pi_{t-1} & R_{t-2} & \Delta q_{t-1} \end{bmatrix}$. Hence

$$E_t(\overline{y}_{t+1}) = w_{t+1}'\delta + S\Gamma_2 \rho_q \Delta q_t,$$

showing that $E_t(\overline{y}_{t+1})$ can be recovered as the predictions from the regression of $\overline{y}_{t+1}$ against w_{t+1} and Δq_t. The same is true of $E_t(\pi_{t+1})$.

[3] We can see some interesting features of this model. If $\rho_a = 0$ then the structural errors in the IS and Phillips curves are proportional to $\tilde{y}_t^*$ and so they are perfectly correlated. This means that there may be singularity problems in the system and this would create difficulties for MLE.

3 Identification

The first item that needs to be addressed is whether it is possible to learn about the values of the parameters of DSGE models. This literature generally goes under the heading of "identification". In extreme cases the model may be unidentified and so one can learn nothing about the parameter values from the data. In most cases however there is "weak identification" in which the data is fairly uninformative about the parameter value. In such instances it turns out to be difficult to produce a precise measure of how uninformative it is. It will be useful to look at identification issues in a staged way since this clarifies some of the discussion in the literature e.g. that in Canova and Sala (2005).

To begin let the parameters in the DSGE model be θ and the unknown parameters in $B_0, B_1, D,$ and C be η. It is often the case that $\dim(\theta) < \dim(\eta)$.Now the first distinction that one needs to make is between *model and structural identification* - see Preston (1978). The VARX system represents the DGP and so describes the observations. Suppose then that the shocks in a DSGE model are white noise and that there are no restrictions upon B_1 and D. Then P and D_0 are of full rank with no restrictions upon them. Varying B_0 i.e. changing the model, then only affects the covariance matrix of the errors in the VARX system i.e. G_0.Hence, if the shocks are uncorrelated, fixing any $[n \times (n-1)]/2$ elements in B_0 will produce identical VARX systems i.e. all such models are observationally equivalent. From this analysis it is unlikely that we will ever have a unique DSGE model. The point could be considered a little trite since we have been familiar for many years with the fact that different orderings of variables i.e. different triangular representations of B_0, are observationally equivalent. To distinguish between them requires extra information, such as prior ideas about the signs and magnitudes of impulse responses.[4] Thus a paper such as Kim(2003) has no implications for the type of identification issues we are normally concerned with, as it just demonstrates that there are a number of models that are observationally equivalent.

Structural identification is concerned with the ability to learn about parameters of a given model from the VARX. To look at this it is useful to ask

[4] The impulse responses to the shocks u_t in (**??**) are weighted averages of the impulse responses to $v_t = G_0 u_t,$ where the weights depend on G_0 and, hence, B_0. The impulse responses to v_t are invariant to the models when there is observational equivalence, but B_0 will vary with the model.

three distinct questions

1. Can we identify η?

2. Can we identify θ from η when the latter is identified?

3. If η is not identified can we identify θ without using the mapping between η and θ?

The reason for the distinction is that knowledge of η determines the impulse responses and therefore, for most policy purposes, what we wish to learn about is η, and not θ. Only if our policy actions involve changing θ would we be concerned to identify the latter.

The New Keynesian Policy Model (NKPM) is a good example of the issues that arise in attempting to estimate η. In its simplest form it has a Phillips curve, an IS curve and an interest rate rule

$$\pi_t = \eta_2 E_t(\pi_{t+1}) + \eta_3 \xi_t + u_{St} \quad (14)$$
$$\xi_t = \eta_5 E_t(\xi_{t+1}) + \eta_6 (r_t - E_t(\pi_{t+1})) + u_{Dt} \quad (15)$$
$$r_t = \eta_7 r_{t-1} + \eta_8 \xi_t + \eta_9 \pi_t + u_{It}, \quad (16)$$

where π_t is inflation, ξ_t is demand and r_t is an interest rate. We might have $E_t(\xi_{t+1})$ and $E_t(\pi_{t+1})$ in place of the current values in the policy rule without changing any of the discussion. There are no observable shocks in this system.

Consider the estimation of the parameters of this system in the case where there is no serial correlation in the shocks. First, it is clear that the rank of P in (2) is only one, since r_{t-1} is the only pre-determined variable in the VARX. Second, there are two endogenous variables on the RHS of (14), so two instruments are needed to estimate the parameters η_2 and η_3.[5] This therefore leads to identification problems . In contrast, when the equations become "hybrid" ones in which π_{t-1} and ξ_{t-1} appear in each equation, two extra instruments will become available and the order condition for identification of the Phillips curve parameters will be satisfied.

Now what happens if there is serial correlation in the shocks of the first two equations of the NKPM in (**??**)-(16) (a very common assumption)? The

[5]Since $E_t(\xi_{t+1}) = w_t'\delta$, where $w_t' = \begin{bmatrix} \xi_t & \pi_t & r_t & r_{t-1} \end{bmatrix}$, we see that it involves endogenous variables. Of course this reflects the dating of information.

expectations are constructed differently but instruments are still needed for them. However now we can transform the equation to eliminate the serial correlation e.g. the inflation equation with an AR(1) for its shock of the form

$$u_{St} = \rho_S u_{St-1} + e_{St}$$

would become

$$\begin{aligned}\pi_t &= \rho_S \pi_{t-1} + \eta_2 E_t \pi_{t+1} - \eta_2 \rho_S E_{t-1} \pi_t + \\ &\quad \eta_3 \xi_t - \eta_3 \rho_S \xi_{t-1} + e_{St},\end{aligned}$$

Since the same transformation applies to the first two equations this clearly means that there will now be three instruments available to estimate this equation viz. ξ_{t-1}, r_{t-1}, and π_{t-1}. Hence the assumption that the shocks have an AR structure generates enough instruments for the estimation of ρ, η_2 and η_3 in the inflation equation, and this is also true of the remaining equations. Of course this does not come from the structure of the model but is simply a consequence of an extraneous assumption about shocks.

Now in many instances the model parameters appear in the η_j but are not equal to them. Thus in Gali and Gertler's (1999) Phillips curve $\eta_2 = \frac{(1-\alpha)(1-\beta\alpha)}{\alpha}$ and $\eta_3 = \beta$. Suppose one wishes to estimate α (the fraction of firms that do not adjust their price at time t). Is it identified? Now it may be that η_2 cannot be identified in the system and so α would not be, but that depends on more than the nature of Phillips curve, so we will assume that it can be recovered, and hence the question becomes one of whether there is a unique mapping between α and η_2. As Ma(2005) points out there isn't one. This is clearly because the equation connecting α and η_2 is effectively a quadratic and so there will generally be two values for α for any values of η_2. Examination of the quadratic shows that the solution is real and that there is no reason to expect that the two values of α are the same. Of course from an operational perspective it does not matter that α cannot be uniquely estimated, as the impulse responses depend upon η_j. This is a good example of why one wants to distinguish between the η and θ parameters.

To take another example that is in the literature, Canova and Sala (2005) look at the following version of the Phillips curve in the NKPM

$$\begin{aligned}\pi_t &= \frac{\omega}{1+\omega\beta}\pi_{t-1} + \frac{\beta}{1+\omega\beta}E_t(\pi_{t+1}) + \\ &\quad \frac{(\varphi+\nu)(1-\zeta\beta)(1-\zeta)}{(1+\beta\omega)\zeta}\xi_t + u_{2t} \\ &= \eta_1\pi_{t-1} + \eta_2 E_t(\pi_{t+1}) + \eta_3\xi_t + u_{2t}\end{aligned}$$

Now suppose that the shocks in the system, u_{jt} were white noise. Then we can potentially estimate the η_j provided the IS curve and Phillips curves in the NKPM are of the hybrid form, since ξ_{t-1}, π_{t-1} would be available as instruments. However, even if the η_j are identified and β, ω are known, it is immediately obvious that this is not true of ν and ξ, since $\eta_3 = \frac{(\varphi+\upsilon)(1-\zeta\beta)(1-\zeta)}{(1+\beta\omega)\zeta}$ is the only η_j that involves the two parameters ν and ζ. Now in Canova and Sala v_{1t} and v_{2t} are AR(1) processes and, as we might expect from the discussion above, this aids identification a great deal. Elimination of the serial correlation in the second equation introduces terms $\frac{\rho(\varphi+\upsilon)(1-\zeta\beta)(1-\zeta)}{(1+\beta\omega)\zeta}\xi_{t-1}$ and $\frac{\rho\omega}{1+\omega\beta}\pi_{t-1}$ and creates two extra η_j to use to estimate the three parameters ρ, ν and ζ. Thus identification is once again being achieved by the assumption of the shocks being AR processes, which is not part of the economic model.

When $\dim(\theta) < \dim(\eta)$ it is possible that there is identification of θ even if η cannot be identified.

The examples show that it is not entirely clear that one needs to be able to identify the DSGE model parameters unless one wanted to perform experiments in which they were changed. Sometimes this is done in policy uses of DSGE models but mostly it is not. Standard uses generally require η, so that it is unclear how much emphasis should be placed upon demonstrations of the difficulty of reliably estimating DSGE model parameters e.g. as in Canova and Sala (2005). It would seem useful to begin any analysis by asking whether the η are identified, and that means one is naturally led to consider the Euler equations.

4 Estimation Issues

There are various formal methods of estimation, differentiated largely by the extent of how much credence is to be placed upon the complete DSGE model.

Single equation method of moments estimators like GMM, which work off the moments coming from Euler equations, utilize the complete system only to the extent of suggesting what would be reasonable instruments. Maximum likelihood methods, which maximize a log likelihood, $L(\theta)$, with respect to the model parameters θ, try to improve on the precision of GMM by using the precise structure of the DSGE model.

As has been known for a long time, such efficiency can come at the expense of bias and inconsistency of estimators, unless the complete system is an adequate representation of the data. As Johansen (2005) has pointed out, this is a price of MLE, and it should not be assumed that the DSGE model has that property. Again this calls for a proper examination of the extent to which the DSGE model is capable of capturing the main characteristics of the data.

Bayesian methods have also become increasingly popular. To get point estimates of θ comparable to MLE, one can maximize $L(\theta) + \ln p(\theta)$, where $p(\theta)$ is the prior on θ. The resulting estimate of θ is often referred to as the mode of the posterior. An advantage of the Bayesian method is that there is often information about the range of possible values for θ, either from constraints such as the need to have a steady state or from past knowledge that has accumulated among researchers. Imposing this information upon the MLE is rarely easy. It can be done by penalty functions, but often these make estimation quite difficult. Adding on $\ln p(\theta)$ to the log likelihood generally means that the function being maximized is quite smooth in θ, and so estimation becomes much easier. We think that this advantage has been borne out in practice; the number of parameters being estimated in DSGE models like Smets and Wouters (2003) is quite large, and one suspects that MLE estimation would be quite difficult.

There is however a cost to Bayesian methods. Unlike penalty functions the use of a prior changes the shape of the function being optimized. If $L(\theta)$ is relatively flat in θ then the choice of prior will become very important in determining the estimated parameter values.[6] In DSGE models this seems

[6] The mode of the posterior is generally used to begin a process of simulating realizations from the posterior density for θ. Often the method used is that set out in Schforheide (2004). One wonders how useful the posteriors being reported are, since being able to characterize a high dimensional density accurately requires huge numbers of realizations from it - the empty-space phenomenon. To illustrate this consider estimating the height at the origin of a multi-dimensional density. Table 4.2 of Silverman (1986) vividly illustrates the fact that, when the density is $N(0, I_d)$, a 90% accuracy for the estimate requires a

Table 1: Bayesian and FIML Parameter Estimation Results

	Prior distribution			Bayes Mean		FIML	
	Density	Mean	std	Mean	90% Interval	Mean	90% Interval
ψ_1	Gamma	1.50	.20	1.36	[1.06, 1.63]	1.62	[0.33, 2.91]
ψ_2	Gamma	0.25	.125	0.22	[0.10, 0.33]	0.32	[-0.21, 0.85]
ψ_3	Gamma	0.25	.125	0.13	[0.07, 0.20]	-0.07	[-0.19, 0.05]
ρ_R	Beta	0.50	.20	0.77	[0.71, 0.82]	0.90	[0.85, 0.95]
α	Beta	0.20	.05	0.10	[0.06, 0.13]	.0003	[.0002, .0004]
κ	Gamma	0.50	.15	0.58	[0.34, 0.78]	0.004	[.0001, .008]
τ	Gamma	0.50	.20	0.18	[0.10, 0.27]	.002	[-.0006, .03]
ρ_q	Norm	-.20	.2	-0.18	[-.30, -0.07]	-.15	[-.32, .005]
ρ_A	Beta	0.20	.10	0.55	[0.47, 0.62]	.39	[.33, .44]
ρ_{y^*}	Beta	0.90	.05	0.96	[0.94, 0.99]	.99	[.97, 1.01]
ρ_{π^*}	Beta	0.70	.10	0.41	[0.28, 0.53]	.24	[.03, .46]
σ_R	InvGamma	0.50	4	0.31	[0.24, 0.37]	.20	[.17, .23]
σ_g	InvGamma	1.50	4	1.33	[1.16, 1.50]	1.26	[1.11, 1.41]
σ_A	InvGamma	1.50	4	0.45	[0.33, 0.57]	.61	[.43 .79]
σ_{y^*}	InvGamma	1.50	4	0.71	[0.40, 1.01]	1.45	[-.71, 3.61]
σ_{π^*}	InvGamma	0.55	4	3.28	[2.87, 3.75]	3.56	[3.00, 4.12]

likely to become an issue.

4.1 Analysing Systems Estimators of a DSGE Model

Let us look at some of these issues in the context of LS's model. Bayesian and MLE estimates of LS's model parameters are given in Table 1.[7]

To get these we utilize UK data provided on Schorfheide's web page. This contains series on the quarterly real output growth Δy_t, the annualized quarterly inflation $\pi 4_t$, the annualized nominal interest rate $R4_t$, quarterly exchange rate change, Δe_t, and terms of trade growth, Δq_t. The series were

sample size of 4 when $d = 1$ but one of $842,000$ when $d = 10$. There are often far more parameters in DSGE models than ten.

[7] To estimate the model we employ DYNARE version 3.042, by S. Adjemian, M. Juillard and O.Kamenik.

de-meaned and the data are related to the model variables as follows

$$[\Delta y_t, \pi 4_t, R4_t, \Delta e_t, \Delta q_t] = [\Delta \tilde{y}_t + \Delta A_t, 4 \times \pi_t, 4 \times R_t, \Delta e_t, \Delta q_t].$$

The priors used were those in LS except for the parameter ρ_q. As noted in Fukac and Pagan (2006) Δq_t is an observable exogenous variable and therefore ρ_q can be estimated by regressing Δq_t on Δq_{t-1}. Since the resulting estimate was significantly negative the LS prior that forced it tto be positive seemed inappropriate. In the case of MLE no constraints were placed on the sign of coefficients. It is clear from this table that there are some major differences in the estimates - in particular for the parameters $\psi_3, \alpha, \rho_A, \kappa, \tau$ and ρ_{π^*}. Apart from the fact that the Bayesian estimates all have the "right" sign by design, their magnitude is well away from zero. The standard approach of comparing the prior and posterior means would suggest that the data plays some role in the Bayesian estimates as the estimate of α is pulled towards the MLE value of zero, although there is a peculiar exception for ρ_A. Against this the Bayesian 90% intervals are far shorter than those obtained with the asymptotic MLE results and they suggest quite a high degree of precision in estimating the parameters.

As with all Bayesian estimation it is hard to know whether one sees these differences as a good or a bad thing. In the event that one does have extremely good prior information one does expect that there will be much more precision about the possible parameter values, but few priors in macroeconometric models can be thought of in this way. Often it is only the signs that would be known. For example what is the basis for thinking that the price stickiness parameter (κ) is .5 or that the first order serial correlation in unobservable foreign inflation is .7? Of course these can be experimented with, but experimentation through addition and subtraction of variables was always a major objection by Bayesians to frequentist approaches. Their argument has been that such experimentation makes it hard to know whether the supplied measures of uncertainty about the parameter values are appropriate. Once one starts varying priors it is unclear what one learns from the reported posteriors. Essentially the Bayesian estimates are being designed to match the researcher's attitudes to what would be attractive parameter estimates. Our impression is that a lot of Bayesian estimation in macroeconomics involves extensive searching over both the type and form of priors.

It is worth looking at this issue in a bit more detail in conjunction with two of the parameters, namely α and ρ_{π^*}. These appear in the exchange rate equation. Under the assumptions made in this model Δq_t and π_t^* are strongly exogenous processes, and so (6) is actually a regression equation, with $\Delta e_t - \pi_t$ as dependent variable, Δq_t as the regressor and with first order serially correlated errors. We know therefore that one can estimate α from this equation by simply using the MLE.

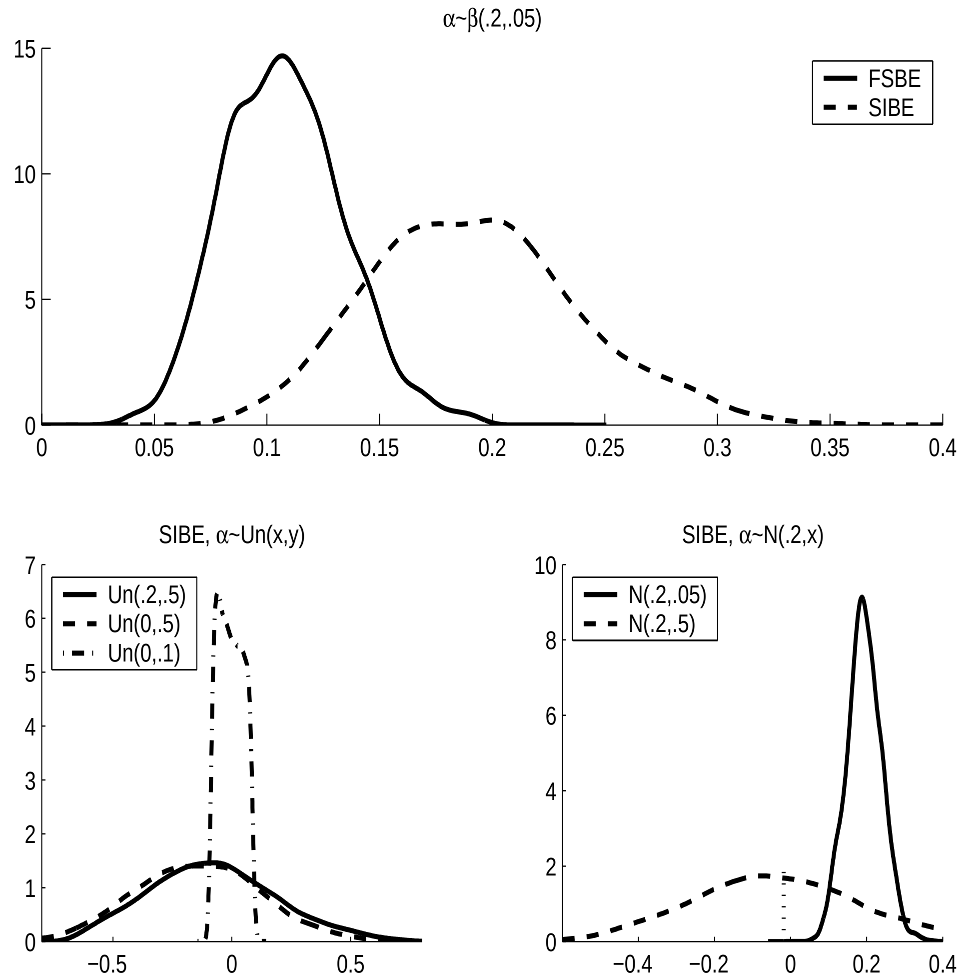
The exact MLE estimates of the parameters of (6) are found using MicroFit Version 5. We will call this the LIML estimator since it is a single equation estimator. The parameter estimates are found in Table 2 .

Table 2 LIML Estimates of the Parameters of (6)				
	est	std dev	t	
α	-.113	.26	.43	
ρ_{π^*}	.073	.11	.65	
σ_{π^*}	3.195	.36		

Now it is clear that these are very different to the Bayesian estimates. To look at this in more detail we stand back from the system and simply estimate the exchange rate equation using Bayesian methods. We first look at α. A negative value for α is certainly unattractive since it is meant to be an import share, but the implication of the MLE estimates is more that one can't estimate it with any precision. In times past a "wrong sign" might well have suggested to an investigator that there are specification problems with the equation. But one does not get any such feeling from the Bayesian estimates. So how does the Bayesian mode of .1 come about? Figure 1 shows that there is a large difference between the posteriors of the system (FSBE) and single equation (SIBE) Bayesian estimators when the beta prior is that used by LS. The imprecision that is indicated by the MLE is present in the single equation results. Moreover the fact that the mean of the posterior for SIBE is virtually the same as the prior tells one that there is very little information in the sample about α. This point is made very clear when one looks at Figure 2 which shows how the likelihood changes with α and what the criterion being used to get the mode of the posterior is. It's evident that the augmented criterion is dominated by the prior component. Given this any answers reflect the location of the prior mean. If it is centered on zero one will get results very similar to the MLE. To show this we impose a normal prior with a zero mean and the same standard deviation as for the beta density. The results are in the bottom part of Figure 1. Although there is nothing surprising in these, the point is that the Bayesian *system* estimates suggest the opposite i.e. that there is a good deal of information in the sample, as shown by the difference in the means of priors and posteriors in Table 1. The extra information is in fact not from the data but from the imposition of cross-equation restrictions due to the presence in the system of forward expectations and the assumption that the shocks in the LS model are uncorrelated. Hence, the estimate of α must be shifting because of these. But for these restrictions to be benign the complete system must be correctly specified. This seems a big assumption and we will see later that there is strong evidence against it. One wonders at the wisdom of using the complete system to estimate parameters that can be estimated without reference to it.

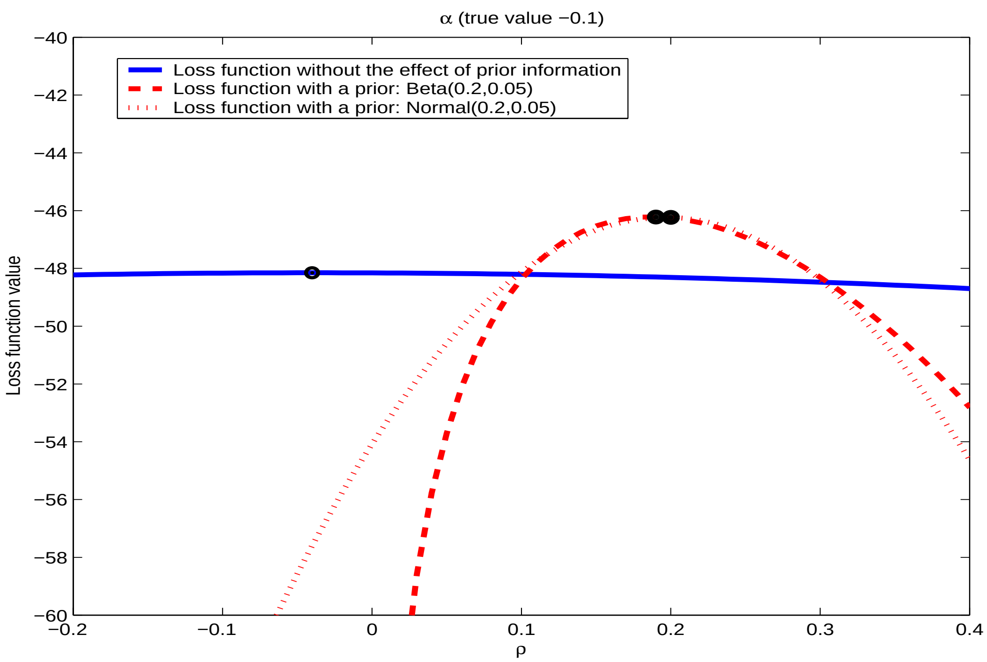
Turning to ρ_{π^*}, Figure 3 that shows the SIBE and FSBE are much the same. Again it would appear that there is a good deal of information about ρ_{π^*} in the sample, since the means of priors and posteriors are quite different. But this is very dependent on the type of prior and its location. If we choose a uniform density that has the same mean and variance as the beta density we basically get the MLE. Moreover, if we choose a normal prior with the same standard deviation as the beta prior, and also one that is five times higher, than we would again get the MLE evidence. Thus the parameterization of a given prior is now the principal determinant of the modal estimates and the range of uncertainty about the parameter. Again this is not something one learns from Table 1. In this simple model one can discover these difficulties but in more complex applications it seems unlikely.

The suggestion by Canova and Sala (2005) that one should increase the variance of the prior and see whether the prior and posterior cohere strongly throughout such an operation is a good one, since one would expect that they would do so if there was little information in the sample. But why not just compare the Bayesian mode to the MLE as that is equivalent to a very high variance in the prior and also does not depend upon the nature of the prior? One disturbing thing about many studies with Bayesian methods is that the standard deviations of the estimates are incredibly small and suggest t ratios of around 30-60. Getting such outcomes with macroeconomic data should be regarded as incredible and should alert the reader and researcher that something has gone wrong. It may also be that the MLE produces "absurd parameter estimates", but often that this is best interpreted as a warning rather than something that is to be suppressed



α~β(.2,.05)
FSBE
SIBE
15
10
5
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
SIBE, α~Un(x,y)
Un(.2,.5)
Un(0,.5)
Un(0,.1)
7
6
5
4
3
2
1
0
−0.5
0
0.5
SIBE, α~N(.2,x)
N(.2,.05)
N(.2,.5)
10
8
6
4
2
0
−0.4
−0.2
0
0.2
0.4

Figure 1:



α (true value −0.1)
Loss function without the effect of prior information
Loss function with a prior: Beta(0.2,0.05)
Loss function with a prior: Normal(0.2,0.05)
Loss function value
ρ
−40
−42
−44
−46
−48
−50
−52
−54
−56
−58
−60
−0.2
−0.1
0
0.1
0.2
0.3
0.4

Figure 2:

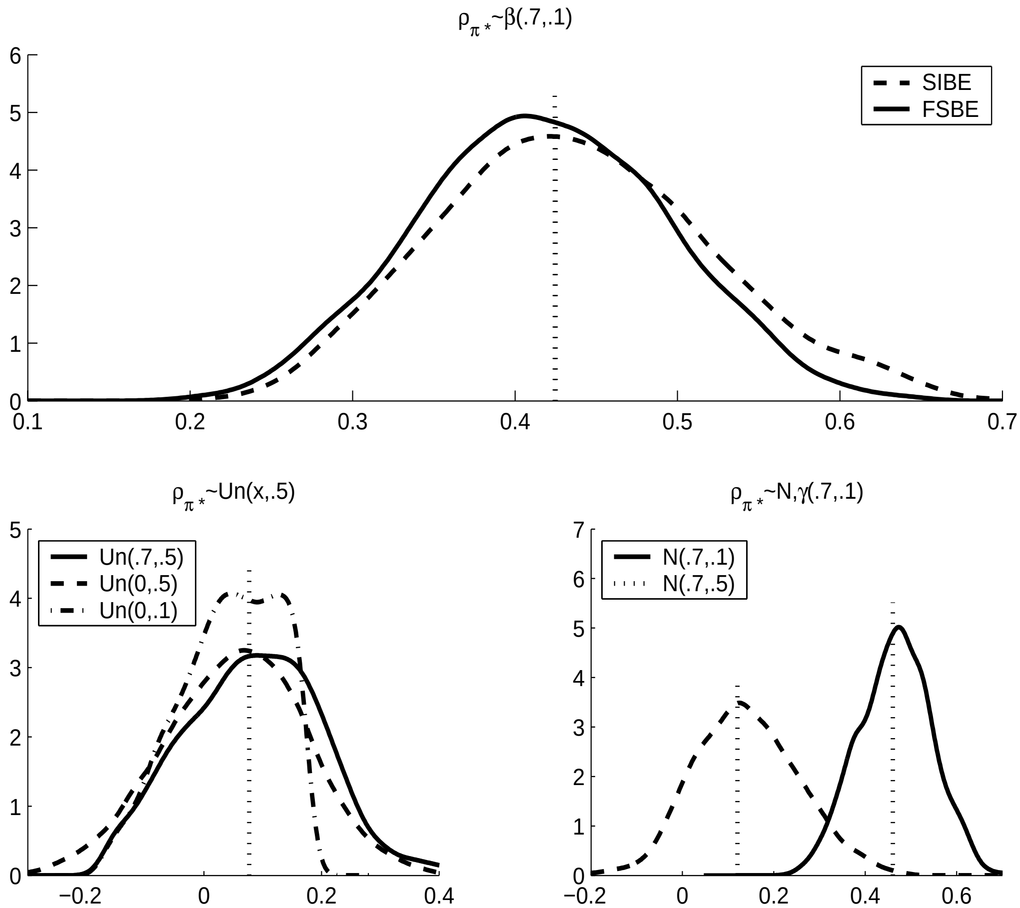


Figure 3:

4.2 Euler Equation Estimation

As we argued in the section on identification it is instructive to estimate the Euler equation parameters η rather than the DSGE model parameters θ. We therefore look at estimating the IS equation in LS's model from this perspective. To estimate (12) we need to measure y_t^p. This can be done by applying the Beveridge-Nelson decomposition to data on y_t. To do this we assumed that Δy_t was an AR(2), although the answers are not sensitive to making it an AR(1) or higher order. Once this has been found $\overline{y}_t$ can be constructed. The next step is to estimate $E_t\overline{y}_{t+1}$, and this involves getting the predictions of $\overline{y}_{t+1}$ from the regression of that variable against $\overline{y}_t, R_t, R_{t-1}, \Delta e_t - \pi_t, \pi_t$ and Δq_t. $E_t\pi_{t+1}$ can be found in the same way.

Now (12) can be written as

$$\begin{aligned} \zeta_t &= \overline{y}_t - E_t\overline{y}_{t+1} = \eta_1(R_t - E_t\pi_{t+1}) + \eta_2\frac{\Delta y_t^p}{(1-\rho_a L)} + \eta_3\Delta q_t + u_t \quad (17) \\ u_t &= \frac{e_t}{(1-\rho_{y^*}L)} \end{aligned}$$

and e_t is white noise. Writing $\frac{\Delta y_t^p}{(1-\rho_a L)} = \Delta y_t^p + \rho_a \Delta y_{t-1}^p + ...$, we can estimate the equation above by using an IV estimator and correcting for first order serial correlation in the error.[8] Under the assumptions of the LS model Δy_t^p is proportional to ε_{at} and so is uncorrelated with u_t. If we run an OLS regression ignoring the serial correlation, then the residuals from that regression, $\hat{u}_t$, can be lagged and used as instruments in the AR corrected regression. As instruments we therefore use $\hat{u}_{t-1}, \Delta y_t^p, \Delta y_{t-1}^p, \Delta y_{t-2}^p, \Delta q_t, \zeta_{t-1}, \overline{y}_{t-1}$ along with the lagged real interest rate and inflation expectations. Results are in Table 3.

[8] We truncated the expansion at two lags in the results presented below but our conclusions are not sensitive to higher order truncations.

Table 3 Estimates of IS Curve					
Coeff		Est	t	Est	t
$R_t - E_t(\pi_{t+1})$		-.22	-7.2	-.19	-8.6
Δq_t		-.01	-.7	-.04	-4.2
Δy_t^p		-.24	-20.8	-.34	-28.0
Δy_{t-1}^p		-.17	-14.4	-.03	-2.0
Δy_{t-2}^p		-.07	-5.2	.02	1.2
$\overline{y}_{t-1} - E_t(\overline{y}_{t+1})$				.42	9.0
AR(1)		.53	3.7	-.36	-1.4

Now a question that arises is whether the specification of this IS equation is satisfactory. In particular the assumption that there is only forward looking behaviour in it. We therefore replace $E_t\overline{y}_{t+1}$ with $(1-\phi)E_t\overline{y}_{t+1} + \phi\overline{y}_{t-1}$ and add the variable $\overline{y}_{t-1} - E_t\overline{y}_{t+1}$into (17). A test of whether one needs the more general specification can be performed by testing if $\phi = 0$. Table 3 shows that there is strong evidence that ϕ is not zero. It is noticeable that the implied estimates of ρ_a (ratio of coefficients on Δy_{t-1}^p and Δy_t^p) and ρ_{y^*} are close to zero and it suggests that the incorrect specification of the IS relation may have biassed the estimates of these coefficients in the MLE and Bayesian estimates.

5 Evaluation

Estimation approaches in DSGE models involve either formal or informal uses of the data. The latter are often termed "calibration" and often constitute a wide range of procedures - matching of moments, use of opinions and intuition, evidence from previous micro and macroeconometric work etc. Informal methods are rarely uninformed by data. There is a case that they can be highly effective - they can often provide a filter against errors in data and can combine together quite a lot of information in a useful way. The issue shouldn't really be whether informal methods are "bad" estimation methods, but rather whether one performs an adequate evaluation of any model whose parameters have been quantified by such an approach.

5.1 Examining the Euler Equations

Using the parameter values found with (say) Bayesian estimation we can determine how well the Euler equations are satisfied. Our strategy will be to use the estimated DSGE model parameter values to determine y_t^p via a Beveridge-Nelson decomposition as well as $E_t(y_{t+1})$ and $E_t(\pi_{t+1})$. It is then possible to compute what the residuals would be from the IS, Phillips curve and interest rate rule equations. These should be proportional to y_t^*, y_t^* and ε_t^R respectively, so that we can check the serial correlation assumptions made about these shocks in estimating the DSGE models by Bayesian and MLE methods. It is also the case that the residuals from the first two equations should be perfectly correlated while that from the last equation should be uncorrelated with the others. Table 4 shows what happens if we fit an AR(2) to the residuals from each of the three equations.

Table 4 AR(2) Models Fitted
Euler Equation Residuals

Eq		AR(1)	t		AR(2)	t
y_t		-.05	-.45		-.35	-3.2
π_t		.59	5.18		-.06	-.54
R_t		-.47	-4.2		-.18	-1.61

It is clear that the assumptions about these shocks are not compatible with the Bayesian parameter estimates, with the closest being for ρ_{y^*}, although the degree of persistence in the shocks is well outside the prior range in Table 1. Indeed the monetary policy shock seems to be negatively correlated. The prediction of LS's model that there should be a perfect correlation between the IS and Phillips curve shocks is resoundingly rejected with a correlation between these of -.10. In contrast, the correlation between the Phillips curve and interest rate rule shocks, which should be zero, is significantly different from that value at .39.

5.2 How Well Does the Model Track the Data?

In the history of macroeconometric modelling a primary way of assessing the quality of models was via historical simulation of them using a set of observed values of exogenous variables. The maxim among the proprietors of such models was "simulate early and simulate often", as that enabled the system properties to be viewed and was a complement to single equation tests of adequacy such as serial correlation. It seems important that we see such model tracking exercises for DSGE models, as the plots of the paths are often very revealing about model performance, far more than might be found from just an examination of a few serial correlation coefficients and bivariate correlations, which have been the standard way of looking at DSGE output to date.[9] It is not that one should avoid computing moments for comparison, but it seems to have been overdone, in comparison to tests that focus more on the uses of these models such as forecasting (which is effectively what the tracking exercise is about).

Now as mentioned in the introduction there is a problem with producing such exercises for DSGE models. Using y_{t-1}^D in place of y_{t-1}^* when making a prediction is unappealing given that the model may well be mis-specified, and so there is need for a wedge between y_t^* and y_t^D. Altug (1989) pioneered one way of doing this by writing $y_t = y_t^* + \eta_t$ and then assuming that the η_t were i.i.d. and uncorrelated with model shocks. One can then estimate the $var(\eta_t)$ and extract estimates of y_t^* using Kalman filtering methods. Ireland (2004) has a generalization of this where η_t can be serially correlated. But there seems no reason to think that these residuals should be uncorrelated with model shocks and it is easy to construct cases where they would not be.

An alternative approach was developed by Watson (1993), in which he asked what was the smallest η_t that one needed to reconcile the DSGE characteristics with the same characteristics in the data. Thus, when y_t is a single variable and both y_t^* and y_t are *i.i.d.*, one can show that the smallest variance of η_t will be $(var(y_t^*) - var(y_t))^2$, and the values of y_t^* which are consistent with this minimal variance will be equal to $\left[\frac{var(y_t^*)}{var(y_t)}\right]^{1/2}$.

If the data and model are not *i.i.d.* then one needs to somehow solve the same problem allowing for the serial correlation. Watson's suggestion was to find the shock that would minimize the gap between the spectra of y_t^* and y_t. He then showed that the value of y_t^* could be reconstructed as $y_t^* = \Xi(L)y_t$, where $\Xi(L)$ has both backward and forward elements.

It is worth looking at this issue and our resolution of it in the context of a

[9] One problem with such moment comparisons is when parameters are estimated from the data and these involve moments of shocks. In a regression model this would mean that the variance of the regression error (shock) can be chosen to perfectly match the variance of the variable being explained. Thus the comparison of moments is often best when parameters have not been estimated.

simple application in which there is just a single variable y_t that is observed. The "DSGE model" is then an equation for a variable y_t^* which is the model equivalent to y_t. There is a model shock v_t and an observation shock u_t, and the system has the State Space Form (SSF)

$$\begin{aligned} y_t &= x_t + u_t \\ y_t^* &= v_t. \end{aligned}$$

The variances of y_t and y_t^* are known – the latter since the DSGE model has been calibrated. Our aim is to estimate y_t^* as closely as possible given the data. We assume that the shocks u_t, v_t are bivariate normal with zero means.

From the bivariate normal we know that the best estimate of y_t^* given the observations is $E(y_t^*|y_t) = \frac{cov(y_t^*,y_t)}{var(y_t)} y_t$. But we don't know what $cov(y_t^*, u_t)$ is and so we don't know the numerator of the weight to be applied to y_t. Some value needs to be selected for it and to do that we need a criterion. A simple one would be to minimize $var(y_t - E(y_t^*|y_t))$ i.e. to minimize

$$\begin{aligned} & E(y_t - \frac{cov(y_t^*, y_t)}{var(y_t^*)} y_t)^2 \\ = & \ (1 - \frac{cov(y_t^*, y_t)}{var(y_t)})^2 var(y_t) \end{aligned}$$

Since $var(y_t)$ is fixed this is minimized by maximizing $cov(y_t^*, y_t)$. Now, because $cov(y_t^*, y_t) \leq std(y_t^*) \times std(y_t)$, it is clear that the optimal estimate of y_t^* will be

$$\frac{std(y_t) std(y_t^*)}{var(y_t)} y_t = \frac{std(y_t^*)}{std(y_t)} y_t$$

which is Watson's (1993) result.

Instead of working with $cov(y_t^*, y_t)$ we could work with $cov(v_t, u_t)$ since the difference between these involves $var(y_t^*)$ which is known. But then we can recognize that $E(y_t^*|y_t)$ would be the estimate we would make when applying the Kalman filter to the SSF and allowing for the covariance between model and observation shocks. For a given value of the latter we can compute $E(y_t^*|y_t)$ and then this may be chosen by minimizing the sum of squares $\sum_{t=1}^{T} (y_t - E(y_t^*|y_t))^2$.

Now DSGE models are more complex and have the state space form

$$
\begin{aligned}
y_t &= F y_t^* + u_t \\
y_t^* &= M y_{t-1}^* + v_t
\end{aligned}
$$

but we can estimate $E(y_t^*|y_0...y_t)$ with the Kalman filter assuming a given $cov(u_t, v_t)$. Again this will be unknown so we have to have a criterion to choose it. One would be to just do what we did above i.e. minimize the variance of $y_t - FE(y_t^*|y_0...y_t)$, but one could also minimize a weighted average of autocovariances of this quantity, which is what a spectral approach does. In doing this minimization the only free parameters are those in $cov(u_t, v_t)$ since

$$var(u_t) = var(y_t) - 2Fcov(v_t, u_t).$$

Since $var(u_t) \geq 0$, this restriction must be enforced and that imposes constraints upon the allowable values for $cov(v_t, u_t)$.

6 References

Adolfson, M, S. Laseen, J. Linde and M. Villani (2005), "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model", *Sveriges Riksbank Working Paper 190*

Altug, S., 1989, "Time to Build and Aggregate Fluctuations: Some New Evidence" , *International Economic Review* 30, 889-920.

Andrews,D.W.K. and J.H.Stock (2005), "Inference with Weak Instruments", Discussion Paper No1530, Cowles Foundation for Research in Economics, Yale University

Ahn, S. and F. Schorfheide (2005), "Bayesian Analysis of DSGE Models", mimeo, University of Pennsylvania

Binder,M and M.H.. Pesaran (1995), "Multivariate Rational Expectations Models and Macroeconomic Modelling: A Review and Some New Results" in M.H. Pesaran and M Wickens (eds) *Handbook of Applied Econometrics: Macroeconomics,* Basil Blackwell, Oxford.

Breunig, R. S. Najarian and A.R. Pagan "Specification Testing of Markov Switching Models" , *Oxford Bulletin of Economics and Statistics*, 65, 703-725.

Brundy, J.M. and D.W. Jorgenson (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables", *Review of Economics and Statistics,* 53,207234.

Canova, F. (1994), "Statistical Inference in Calibrated Models", *Journal of Applied Econometrics*, 9 (supplement), S123-S144.

Canova, F. and L. Sala (2005), "Learning about the Parameters and the Dynamics of DSGE Models: Identification and Estimation", paper presented to the Model Evaluation Conference, Oslo, May 2005

Canova,F, M. Finn and A.Pagan (1994) "Evaluating a Real Business Cycle Model" in C. Hargreaves (ed) *Non-Stationary Time Series Analysis and Co-Intergration* (Oxford University Press), 225-255.

Del Negro, M., F. Schorfheide, F. Smets and R. Wouters (2004) "On the Fit and Forecasting Performance of New Keynesian Models" *Federal Reserve Bank of Atlanta Working Paper 2004-37*

Fuhrer, J.C. and G.P. Olivei (2004), "Estimating Forward-Looking Euler Equations with GMM and Maximum Likelihood Estimators: An Optimal Instruments Approach", paper presented to the Conference on Models and Monetary Policy, Federal Reserve Bank Board of Governors, March

Fukac, M.and A.R.Pagan (2006) "Issues in Adopting DSGE Models for Use in the Policy Process", *CAMA Working Paper 10/2006*, Australian National University

Gali, J.and M.Gertler (1999), "Inflation Dynamics: A Structural Econometric Analysis " *Journal of Monetary Economics*, 44, 195-222.

Gali, J. and T. Monachelli (2002), "Monetary Policy and Exchange Rate Volatility in a Small Open Economy", *NBER Working Paper 8850*

Hausman, J.A. (1975), "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models", *Econometrica*, 43, 727-738.

Hendry, D.F. (1976), " The Structure of Simultaneous Equations Estimators", *Journal of Econometrics*, 4, 51-88.

Ireland, P. (2004), "A Method for Taking Models to the Data" *Journal of Economic Dynamics and Control,* 28, 1205-1226.

Johansen, S. (2005), "What is the Price of Maximum Likelihood", paper presented to the Model Evaluation Conference, Oslo, May 2005

Kim, J. (2003), "Functional Equivalence Between Intertemporal and Multisectoral Investment Adjustment Costs", *Journal of Economic Dynamics and Control*, 27, 533-549.

Lubik, T.A. and F. Schorfheide (2005) " Do Central Banks respond to exchange Rate Movements? A Structural Investigation", *Journal of Monetary Economics,* (forthcoming)

Mavroeidis, S. (2004), "Weak Identification of Forward-looking Models in Monetary Economics", *Oxford Bulletin of Economics and Statistics*, 66, 609-635.

Morley, J.C. (2002), ”A state-space approach to calculating the Beveridge-Nelson decomposition”, *Economics Letters* 75, 123-127

Nason, J.M. and G.W. Smith (2005), "Identifying the New Keynesian Phillips Curve", *Working Paper 2005-1*, Federal Reserve Bank of Atlanta

Poskitt, D.S. and C.L. Skeels (2005), "Approximating the Distribution of the Instrumental Variables Estimator when the Concentration Parameter is Small", *Journal of Econometrics* (forthcoming)

Preston, A. J. (1978) " Concepts of Structure and Model Identifiability for Econometric Systsems: in A. R. Bergstrom et al., *Stability and Inflation* (Wiley, 1978), 275–297.

Schorfheide, F. (2000), "Loss Function-Based Evaluation of DSGE Models", *Journal of Applied Econometrics*, 15, 645-670.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.

Sims, C. A. (2005). "Improving Monetary Policy Models", paper presented to the IRFMP/IMP Conference on DSGE Modeling at Policymaking Institutions: Progress and Prospects, Washington, December

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

Staiger, D. and J.H. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, 557-586.

Watson, M.W. (1993), "Measures of Fit for Calibrated Models", *Journal of Political Economy*, 101, 1011-1041.