A Monte Carlo procedure for checking identification in DSGE models*

Vo Phuong Mai Le (Cardiff University)†

Patrick Minford (Cardiff University and CEPR)‡

Michael Wickens (Cardiff University, University of York and CEPR)§

September 2014

Abstract

We propose a numerical method, based on indirect inference, for checking the identification of a DSGE model. Monte Carlo samples are generated from the model's true structural parameters and a VAR approximation to the reduced form estimated for each sample. We then search for a different set of structural parameters that could potentially also generate these VAR parameters. If we can find such a set, the model is not identified.

1 Introduction

In a recent paper Canova and Sala (2009) have argued that DSGE models may not be identified. They give examples of models in which the reduced form properties of DSGE models of different sorts are hard to distinguish and argue that a weak form of observational equivalence between DSGE models is widespread. They recommend careful exploration of these issues prior to estimation and testing of a particular DSGE model. Schorfheide (2011), however, notes that the extreme nonlinearity of DSGE model solutions in their structural parameters makes checking identification difficult except by numerical methods. In this paper we propose a numerical Monte Carlo method for checking the identification of a DSGE model which is simple to implement. We illustrate its application with two widely-used DSGE models.

Like Iskrev (2010) and Komunjer and Ng (2011) we offer a 'yes-or-no' approach. We do not consider the 'strength' of identification, a particular concern of Canova and Sala. Both the power of different estimation methods, which has been examined by Le et al (2012), and more precise estimation, which may overcome weak identification, have a bearing on the strength of identification. Iskrev and Komunjer and Ng base their yes-or-no methods of identification on the rank of matrices of 'reduced form' coefficients of various types using mostly numerical methods. The main advantage of our method, which is a substitute, is computational speed and convenience since the identification can be routinely carried out using the same methods required for the estimation and testing of the DSGE model.

*We are grateful for useful comments to Huw Dixon, Vito Polito, David Staines, Akos Valentinyi and other participants in the economics workshop, Cardiff University.

†Levp@cardiff.ac.uk; Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

‡Patrick.minford@btinternet.com; Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

§Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

An economic model is said to be exactly identified if and only if all of its coefficients can be derived uniquely from its reduced-form solution; it is over identified if there is more than one set of structural coefficients that can be derived from the reduced-form solution (if the reduced form is exactly correct, then these sets will coincide); and it is under identified (not identified) if it is not possible to derive all of the structural coefficients from the reduced-form solution. This includes situations where it may be possible to derive a sub-set of structural coefficients - either uniquely or not uniquely - from the reduced-form solution. Which of these situations prevails is determined prior to estimation.

These principles also apply to DSGE models. There is, however, an extra feature arising from the need to take account of the conditional expectations of future endogenous variables. This involves first solving the model to eliminate these expectational variables. If the DSGE model is over identified, the solution is, in effect, a restricted reduced form; if the DSGE model is exactly identified then it is identical to an unrestricted reduced form; and if the DSGE model is under identified then it is not possible to derive all of the structural parameters from the unrestricted reduced form.

Our proposed procedure for determining identification involves Monte Carlo simulation and the use of indirect inference. The method of indirect inference allows the testing of models to be focused on key data features of interest to the modeller/policymaker; the method also has very considerable power, rather greater than that of standard likelihood ratio tests when like is compared with like- see Le at al (2012). Le et al (2011) sets out the full methodology of indirect inference. Indirect Inference involves comparing some model describing the 'facts' with the structural DSGE model assumed to be generating them. Therefore it is well set up for answering the question of whether a given structural model is unique in generating the facts that it does. In what follows we will call descriptions of the facts 'reduced forms'.

Our numerical identification procedure is based on comparing the estimates of the coefficients of an auxiliary model derived using simulated data from the DSGE model whose identification is being assessed, and whose parameter values are given to us, with simulated data from the same DSGE model but with different parameter values obtained by disturbing the original parameter values. The comparison is made by testing the null hypothesis that the two sets of estimates of the coefficients of the auxiliary model are the same; this comparison can be made as numerically exact as desired at little extra computaional cost.

The test is conducted by first constructing a numerical distribution for the estimates of the coefficients of the auxiliary model based on simulations using the original DSGE parameters, and then determining whether the estimates of the coefficients of the auxiliary model based on each simulation of the disturbed parameter set could have been drawn from this distribution. If the model is not identified then the proportion of times that the test statistic computed for each sample is significant will be equal to the proportion of times that the estimates of the coefficients of the auxiliary model based on simulations using the original DSGE parameters are significant, i.e. the size of the test. If the proportion of rejections based on the disturbed parameters exceeds this critical value then we conclude that the DSGE is identified.

To illustrate this procedure we consider two models. The first is a standard three-equation New Keynesian model of the US, with a Phillips Curve, an IS curve (taken from the Euler equation and a market-clearing equation) and a Taylor rule. We consider this model in a variety of forms: with and without autocorrelated disturbances and with and without persistence in the Phillips Curve (through indexation) and the Taylor Rule (through interest rate smoothing). The versions with autocorrelated disturbances, with or without persistence, we can show analytically are over-identified; the version with neither autocorrelated errors nor persistence we can show is under-identified. Our procedure can therefore be checked against these two analytical results.

The second model we consider is the Smets-Wouters model of the US (Smets and Wouters, 2007), based on Christiano et al. (2005). This we cannot check analytically as the model is too large; however, we conjecture by analogy from other less complex rational expectations models (such as the 3-equation New Keynesian model above) that this model is over-identified- this is also consistent with Iskrev's findings when he applies his methods to this model. Here we use our numerical method to add information about this case.

The paper proceeds as follows. In section 2 we consider the analytics of identification for the three-equation model in its various forms. In section 3 we explain our numerical method and apply it in detail to this model, to establish that it confirms these analytical results. In section 4 we apply our method both to the 3-equation New Keynesian model of Clarida, Gali and Gertler and also to the Smets and Wouters (2007) model of the US. Section 5 concludes.

2 Indirect Inference as a method for evaluating identification

The general DSGE model can be written as $f(y_t, E_t y_{t+1}, y_{t-1}, \epsilon_t) = 0$ where y_t are p endogenous variables and e_t are p i.i.d. structural errors. We assume that the only exogenous variables are error processes, as is the case for all the models we deal with here. Our concern is with local identification of the loglinearised version of this. We assume that after log-linearisation a DSGE model can be written in the form:

$$A(L)y_t = BE_t y_{t+1} + D(L)e_t \tag{A1}$$

L denotes the lag operator $z_{t-s} = L^s z_t$ and $A(L)$, $D(L)$ are invertible polynomial functions.

The general solution of y_t is (Wickens, 2011, e.g. 506-8), after substituting out the expectations,

$$E(L)y_t = F(L)e_t \tag{A1}$$

which is a VARMA:

$$y_t = G(L)y_{t-1} + F(L)e_t. \tag{A3}$$

It can also usually be written as a VAR:

$$y_t = H(L)y_{t-1} + \eta_t \tag{A1}$$

In the next section we illustrate the solution forms with the 3-equation New Keynesian model.

2.1 Solution and analysis of the three equation New Keynesian model

The model (Model 1) consists of the following equations:

$$\pi_t = \omega E_t \pi_{t+1} + (1-\omega)\pi_{t-1} + \lambda y_t + \varepsilon_t^{\pi} \tag{1}$$

$$y_t = E_t y_{t+1} - \frac{1}{\sigma}(r_t - E_t \pi_{t+1}) + \varepsilon_t^{y} \tag{2}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\gamma \pi_t + \eta y_t + \psi(y_t - y_{t-1})) + \varepsilon_t^{r} \tag{3}$$

where the shocks follow the univariate AR(1) processes

$$\begin{aligned} \varepsilon_t^{\pi} &= \rho_{\pi}\varepsilon_{t-1}^{\pi} + u_t^{\pi} \\ \varepsilon_t^{y} &= \rho_{y}\varepsilon_{t-1}^{y} + u_t^{y} \\ \varepsilon_t^{r} &= \rho_{r}\varepsilon_{t-1}^{r} + u_t^{r} \end{aligned}$$

The first equation is the New-Keynesian Phillips curve. If $\omega = 0$ it is a backward-looking Phillips Curve and, if $\omega = 1$, it is a forward-looking Phillips Curve. The second equation is the aggregate demand equation and the last equation is an interest rate rule where the interest rate is 'smoothed' by the parameter ρ.

This model is similar to that used originally by Clarida, Gali and Gertler (1999) as their prototype New Keynesian model. The Phillips Curve at the heart of this model has been the subject of fierce econometric dispute between those who maintain that it is exclusively forward looking and those who argue that it is partly, or even substantially, backward looking. There is also the issue of the specification of the error processes and whether or not they are serially correlated. An issue of the Journal of Monetary Economics (Volume 52, 6, 2005) was devoted largely to this question, with papers on both sides of the debate (e.g. Gali et al, 2005; Rudd and Whelan, 2005). At the heart of this dispute there is an identification problem.

To illustrate this consider a simpler version of this model (Model 2):

$$\begin{aligned}
\pi_t &= \omega E_t \pi_{t+1} + \lambda y_t + e_{\pi t}, \quad \omega < 1 && (4)\\
y_t &= E_t y_{t+1} - \frac{1}{\sigma}(r_t - E_t \pi_{t+1}) + e_{yt} && (5)\\
r_t &= \gamma \pi_t + \eta y_t + e_{rt} && (6)\\
e_{it} &= \rho_i e_{i,t-1} + \varepsilon_{it} \quad (i = \pi, y, r)
\end{aligned}$$

The model therefore has 5 structural coefficients and 3 autoregressive coefficients. Re-writing the model using the lag operator $E_t x_{t+1} = L^{-1} x_t$ gives

$$\begin{bmatrix} 1-\omega L^{-1} & -\lambda & 0 \\ -\frac{1}{\sigma}L^{-1} & 1-L^{-1} & \frac{1}{\sigma} \\ -\gamma & -\eta & 1 \end{bmatrix} \begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \begin{bmatrix} e_{yt} \\ e_{\pi t} \\ e_{rt} \end{bmatrix}.$$

The solution is therefore

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \frac{1}{\Delta(L)} \begin{bmatrix} 1+\frac{\eta}{\sigma}-L^{-1} & \lambda & -\frac{\lambda}{\sigma} \\ -\frac{1}{\sigma}(\gamma - L^{-1}) & 1-\omega L^{-1} & -\frac{1}{\sigma}(1-\omega L^{-1}) \\ \gamma - (\gamma - \frac{\eta}{\sigma})L^{-1} & \lambda\gamma + \eta - \eta\omega L^{-1} & (1-\omega L^{-1})(1-L^{-1}) - \frac{\lambda}{\sigma}L^{-1} \end{bmatrix} \times \begin{bmatrix} e_{\pi t} \\ e_{yt} \\ e_{rt} \end{bmatrix}$$

where

$$\begin{aligned}
\Delta(L) &= \frac{\lambda}{\sigma}(\gamma - L^{-1}) + (1-\omega L^{-1})(1 + \frac{\eta}{\sigma} - L^{-1}) \\
&= (1 + \frac{\eta + \lambda\gamma}{\sigma}) - [\frac{\lambda}{\sigma} + \omega(1+\frac{\eta}{\sigma})]L^{-1} + \omega L^{-2} \\
&= [1 + \frac{\eta + \lambda\gamma}{\sigma}](1 - \lambda_1 L^{-1})(1 - \lambda_2 L^{-1})
\end{aligned}$$

As $\omega \leq 1$ and $\gamma \succ 1$, $\lambda_1\lambda_2 \prec 1$ and $\lambda_1 + \lambda_2 \prec 1$ we have $\lambda_1, \lambda_2 \prec 1$. Using successive forward substitution, the solution can be shown to be

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \begin{bmatrix} 1+\frac{\eta}{\sigma}-\rho_\pi & \lambda & -\frac{\lambda}{\sigma} \\ -\frac{1}{\sigma}(\gamma - \rho_\pi) & 1-\omega\rho_y & -\frac{1}{\sigma}(1-\omega\rho_r) \\ \gamma - (\gamma - \frac{\eta}{\sigma})\rho_\pi & \lambda\gamma + \eta - \eta\omega\rho_y & 1 - (1+\omega+\frac{\lambda}{\sigma})\rho_r + \omega\rho_r^2 \end{bmatrix} \times \begin{bmatrix} \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma} - [\frac{\lambda}{\sigma} + \omega(1+\frac{\eta}{\sigma})]\rho_\pi + \omega\rho_\pi^2} & 0 & 0 \\ 0 & \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma} - [\frac{\lambda}{\sigma} + \omega(1+\frac{\eta}{\sigma})]\rho_y + \omega\rho_y^2} & 0 \\ 0 & 0 & \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma} - [\frac{\lambda}{\sigma} + \omega(1+\frac{\eta}{\sigma})]\rho_r + \omega\rho_r^2} \end{bmatrix} \begin{bmatrix} e_{\pi t} \\ e_{yt} \\ e_{rt} \end{bmatrix}$$

or

$$z_t = Ae_t$$

where $z_t' = [\pi_t, y_t, r_t]$, $e_t' = [e_{\pi t}, e_{yt}, e_{rt}]$. Thus the matrix A is restricted, having 9 elements but consisting of only 5 structural coefficients (the ρ_i can be recovered directly from the error processes), implying that the model is over identified according to the (necessary) order condition.

If $\rho_i = 0$ for all i (Model 3) then the solution becomes

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma}} \begin{bmatrix} 1+\frac{\eta}{\sigma} & \lambda & -\frac{\lambda}{\sigma} \\ -\frac{1}{\sigma}\gamma & 1 & -\frac{1}{\sigma} \\ -\gamma & \lambda\gamma+\eta & 1 \end{bmatrix} \begin{bmatrix} e_{\pi t} \\ e_{yt} \\ e_{rt} \end{bmatrix}$$

which does not involve ω. Hence, ω is not identified and the other coefficients are over identified. This shows the important role of the error dynamics in identifying ω.In effect, without error dynamics, the expected future variables in the model are always zero, and so effectively they do not appear in the model; thus their coefficients disappear from both the structural and the reduced form- a rather special case of non-identification due to non-presence. While we look at this case below numerically, we regard it as quite atypical. [1]

The model we examine here- Model (1)- is somewhat more complex than Model (2), as noted above. The solution for this fuller three equation model has a similar form to the simplified one above but now in addition has two backward roots, due to the smoothing coefficient in the Taylor Rule and the indexation lag in the Phillips Curve. The determinant now becomes $\Delta(L) = (\frac{\sigma\rho+\psi}{\sigma})(1-\omega)L^2 - [(\frac{\sigma\rho+\psi}{\sigma}) + (1-\omega)(1+\rho+\frac{\eta+\psi}{\sigma})]L + (1+\rho+(\frac{\eta+\psi}{\sigma}) + \omega(\frac{\sigma\rho+\psi}{\sigma}) + \frac{\rho\lambda+(1-\rho)\lambda\gamma}{\sigma} + 1 - \omega) - [\frac{\lambda}{\sigma} + \omega(1+\frac{\eta}{\sigma})]L^{-1} + \omega L^{-2}$ Normally it will have two forward roots and two backward roots, all inside the unit circle. For example with $\omega = 0.8, \rho = 0.4, \sigma = 3.5, \lambda = 0.8, \gamma = 1.2, \eta = 0.9, \psi = \rho 0.1$, we obtain the two forward roots as 0.929 and 0.583, and the two backward roots are the complex pair $0.215 \pm 0.109i$. The model has 7 structural coefficients (the ρ_i we can as before establish from the errors directly) and is over identified. The unrestricted reduced form (after substituting out the forward roots) has 24 coefficients- 6 coefficients on the lagged endogenous variables, and 18 coefficients in the e_{it}. Again the order condition indicates over-identification.

As noted in section 1, the reduced form of a DSGE model can take a variety of forms. We now illustrate this with the simpler Model 2 above whose solution we could write as $z_t = Ae_t$. Since the errors each have a univariate AR coefficient we can easily transform this into a VARMA(3,2) where:

$$z_t = (\sum \rho_i) z_{t-1} - (\sum \rho_i \rho_j) z_{t-2} + (\prod \rho_i) z_{t-3} + A \begin{bmatrix} (1-\rho_y L)(1-\rho_r L)\epsilon_{\pi t} \\ (1-\rho_\pi L)(1-\rho_r L)\epsilon_{yt} \\ (1-\rho_y L)(1-\rho_\pi L)\epsilon_{rt} \end{bmatrix} \qquad (i = \pi, y, r)$$

One can also obtain a (different) VARMA(3,2) by substituting the solutions for $E_t\pi_{t+1}, E_t y_{t+1}$ into the Model 2 structural equations and rearrange this as a set of equations with the e_{it} errors as exogenous variables. The same transformation using the errors' AR coefficients will give the VARMA(3,2).

We can also write the equations as

[1] DSGE models have errors, indirectly observed from the model and the data, that embody exogenous variables (deliberately) omitted from the model; these factors will generally have a degree of persistence because of their nature- eg productivity, mark-ups, tax distortions- and hence the parameters of the time-series process used to model them are regarded as structural parameters like those relating the endogenous variables. We generally impose on them a univariate structure on the grounds that they are independently determined (by their own past, though we may allow for simultaneous correlation in their innovations). Each model, given the sample data, implies a different behaviour for these errors and so implies a different parameter set for the error processes. By choosing a DSGE model one is therefore also choosing a set of errors and an implied set of time-series parameters for their structure. These latter parameters are therefore part of the DSGE model structural parameters.

$$\begin{matrix} I & -A & z_t \\ 0 & I & e_t \end{matrix} = \begin{matrix} 0 & 0 & z_{t-1} \\ 0 & R & e_{t-1} \end{matrix}$$ where R is the matrix with ρ_i along the diagonal and zeros elsewhere.

Finally we can transform the model into a VAR by using the Fernandez-Villaverde et al (2007) ABCD method:

$z_t = Ae_{t-1} + Be_t$

$e_t = Ce_{t-1} + D\varepsilon_t$

where ε is the vector of innovations in e and $C = P = \begin{bmatrix} \rho_\pi & 0 & 0 \\ 0 & \rho_y & 0 \\ 0 & 0 & \rho_r \end{bmatrix}; A = QP; B = Q; D = I$

It follows since $e_{t-1} = Q^{-1}z_{t-1}$ that the VAR solution is:

$z_t = QPQ^{-1}z_{t-1} + Qe_t = Ez_{t-1} + \eta_t$. The order of the VAR solution is raised when lags are added to the structural model.

We see here the various ways that the solution may be represented. We can also note that the impulse response functions of some or all shocks can be calculated from these forms, as also for example the moments and cross-moments.

2.2 Model estimation and evaluation by indirect inference

Indirect inference has been widely used in the estimation of structural models (e.g., Smith, 1993, Gregory and Smith, 1991, 1993, Gourieroux, Monfort and Renault, 1993, Gourieroux and Monfort, 1996 and Canova, 2005). Here we make a further use of indirect inference to evaluate an already estimated or calibrated (DSGE) macroeconomic model using classical statistical inference. This is related to, but is different from, estimating a macroeconomic model by indirect inference. The common feature is the use of an auxiliary model in addition to the structural macroeconomic model. For a full description of the method of indirect inference see also Le et al. (2011). In addition to testing a particular prior numerical specification of the DSGE model, we examine how we might compare and test alternative numerical specifications of the model. Next we set out the main features of indirect inference.

2.2.1 Estimation

Estimation by indirect inference chooses the parameters of the macroeconomic model so that when this model is simulated it generates estimates of the auxiliary model similar to those obtained from the observed data. The optimal choice of parameters for the macroeconomic model are those that minimize the distance between a given function of the two sets of estimated coefficients of the auxiliary model. Common choices of this function are (i) the actual coefficients, (ii) the scores, and (iii) the impulse response functions. In effect, estimation by indirect inference provides an optimal calibration.

Suppose that y_t is an $m \times 1$ vector of observed data, $t = 1, ..., T$, $x_t(\theta)$ is an $m \times 1$ vector of simulated time series generated from the structural macroeconomic model, θ is a $k \times 1$ vector of the parameters of the macroeconomic model and $x_t(\theta)$ and y_t are assumed to be stationary and ergodic. The auxiliary model is $f[y_t, \alpha]$. We assume that there exists a particular value of θ given by θ_0 such that $\{x_t(\theta_0)\}_{s=1}^S$ and $\{y_t\}_{t=1}^T$ share the same distribution, i.e.

$$f[x_t(\theta_0), a] = f[y_t, \alpha]$$

where α is the vector of parameters of the auxiliary model, and the existence of a binding function relating θ to α.

The likelihood function for the auxiliary model defined for the observed data $\{y_t\}_{t=1}^T$ is

$$\mathcal{L}_T(y_t; \alpha) = \Sigma_{t=1}^T \log f[y_t, \alpha]$$

The maximum likelihood estimator of α is then

$$a_T = \arg\max_{\alpha} \mathcal{L}_T(y_t; \alpha)$$

The corresponding likelihood function based on the simulated data $\{x_t(\theta)\}_{s=1}^{S}$ is

$$\mathcal{L}_S[x_t(\theta); \alpha] = \Sigma_{t=1}^{S} \log f[x_t(\theta), \alpha]$$

with

$$a_S(\theta) = \arg\max_{a} \mathcal{L}_S[x_t(\theta); \alpha]$$

The simulated quasi maximum likelihood estimator (SQMLE) of θ is

$$\theta_{T,S} = \arg\max_{\theta} \mathcal{L}_T[y_t; \alpha_S(\theta)]$$

This value of θ corresponds to the value of α that maximises the likelihood function using the observed data. Further, as $x_t(\theta)$ and y_t are assumed to be stationary and ergodic, from Canova (2005),

$$plim\ a_T = plim\ a_S(\theta) = \alpha.$$

it can then be shown that

$$\begin{aligned} T^{1/2}(a_S(\theta) - \alpha) &\rightarrow N[0, \Omega(\theta)] \\ \Omega(\theta) &= E[-\frac{\partial^2 \mathcal{L}[\alpha(\theta)]}{\partial \alpha^2}]^{-1} E[\frac{\partial \mathcal{L}[\alpha(\theta)]}{\partial \alpha} \frac{\partial \mathcal{L}[\alpha(\theta)]}{\partial \alpha}']E[-\frac{\partial^2 \mathcal{L}[\alpha(\theta)]}{\partial \alpha^2}]^{-1} \end{aligned}$$

The covariance matrix can be obtained either analytically or by bootstrapping the simulations.

The method of simulated moments estimator (EMSME) may be extended to estimating a function $g(\theta)$ of θ. Let $g(a_T)$ and $g(\alpha_S(\theta))$ denote a continuous $p \times 1$ vector of functions which could, for example, be moments or scores, and let the mean functions be $G_T(a_T) = \frac{1}{T}\Sigma_{t=1}^{T} g(a_T)$ and $G_S(\alpha_S(\theta)) = \frac{1}{S}\Sigma_{s=1}^{S} g(\alpha_S(\theta))$. We require that $a_T \rightarrow \alpha_S$ in probability and that $G_T(a_T) \rightarrow G_S(\alpha_S(\theta))$ in probability for each θ. The EMSME is

$$\theta_{T,S} = \arg\min_{\theta} [G_T(a_T) - G_S(\alpha_S(\theta))]' W(\theta) [G(a_T) - G_S(\alpha_S(\theta))]$$

where $W(\theta_0)$ is the inverse of the variance-covariance matrix of the distribution of $G_S(a_S) - \overline{G[a_S(\theta_0)]}$. The estimator is consistent and asymptotically normal- Smith (1993), Gourieroux et al (1993) and Canova (2005).

It should be noted here that the properties of the estimator are not dependent on the precise auxiliary model chosen. It is assumed merely that this model is some set of 'description of facts' that can be found in the data. Thus it need not be the full reduced from solution of the model (e.g. the full VARMA form above); it may be some VAR approximation or as noted some impulse responses or some moments and cross-moments; we will call all these descriptions of the facts 'reduced forms', in the sense that they are all deriveable from the model's solution. The essential point is that whatever the description there must be a correspondence between it and the simulated description coming from the structural DSGE model. On the assumption that this model is the true one the II estimator finds the parameters that give the maximum degree of correspondence.

2.3 Model evaluation

In model evaluation indirect inference is used in a different way. The aim here is to compare the performance of an auxiliary model based on observed data with its performance based on data simulated from a calibrated or previously estimated macroeconomic model. We choose the auxiliary model to be a VAR and base our test on a function of the VAR coefficients. The test statistic is formed from the minimand of the EMSME evaluated using estimates of α derived from observed data and data simulated from the given numerically specified DSGE model. The distribution of this Wald type of test statistic is obtained numerically through bootstrapping.

Non-rejection of the null hypothesis is taken to indicate that dynamic behaviour of the macroeconomic model is not significantly different from that of the observed data. Rejection is taken to imply that the macroeconomic model is incorrectly specified. Comparison of the impulse response functions of the observed and simulated data should reveal in what respects the macroeconomic model fails to capture the auxiliary model.

A formal statement of the inferential problem is as follows. Using the same notation as before, we define y_t an $m \times 1$ vector of observed data ($t = 1, ..., T$), $x_t(\theta)$ an $m \times 1$ vector of simulated time series of S observations generated from the structural macroeconomic model, θ a $k \times 1$ vector of the parameters of the macroeconomic model. $x_t(\theta)$ and y_t are assumed to be stationary and ergodic. We set $S = T$ since we require that the actual data sample be regarded as a potential replication from the population of bootstrapped samples. The auxiliary model is $f[y_t, \alpha]$; an example is the $VAR(p)$ $y_t = \Sigma_{i=1}^{p} A_i y_{t-i} + \eta_t$ where α is a vector comprising elements of the A_i and of the covariance matrix of y_t. Under the null hypothesis H_0: $\theta = \theta_0$, the stated values of θ whether obtained by calibration or estimation; the auxiliary model is then $f[x_t(\theta_0), \alpha(\theta_0)] = f[y_t, \alpha]$. We wish to test the null hypothesis through the $q \times 1$ vector of continuous functions $g(\alpha)$. Such a formulation includes impulse response functions. Under $H_0 : g(\alpha) = g[\alpha(\theta_0)]$.

If a_T denotes the estimator of α using actual data and $a_S(\theta_0)$ is the estimator of α based on simulated data for θ_0, we may obtain $g(a_T)$ and $g[a_S(\theta_0)]$. Using N independent sets of simulated data obtained using the bootstrap we can also define the bootstrap mean of the $g[a_S(\theta)]$, $\overline{g[a_S(\theta_0)]} = \frac{1}{N}\Sigma_{k=1}^{N} g_k[a_S(\theta_0)]$. The Wald test statistic is based on the distribution of $g(a_T) - \overline{g[a_S(\theta_0)]}$ where we assume that $g(a_T) - \overline{g[a_S(\theta_0)]} \xrightarrow{p} 0$**.** The resulting Wald statistic (WS) may be written as

$$WS = (g(a_T) - \overline{g[a_S(\theta_0)]})' W(\theta_0)(g(a_T) - \overline{g[a_S(\theta_0)]})$$

where $W(\theta_0)$ is the inverse of the variance-covariance matrix of the distribution of $g(a_T) - \overline{g[a_S(\theta_0)]}$. $W(\theta_0)^{-1}$ can be obtained from the asymptotic distribution of $g(a_T) - \overline{g[a_S(\theta_0)]}$ and the asymptotic distribution of the Wald statistic would then be chi-squared. The empirical distribution of the Wald statistic is derived using bootstrap methods as follows.

Step 1: Determine the errors of the economic model conditional on the observed data and $\widehat{\theta}$*.*

Solve the DSGE macroeconomic model for the structural the errors ε_t given $\widehat{\theta}$ and the observed data[2]. The number of independent structural errors is taken to be less than or equal to the number of endogenous variables. The errors are not assumed to be normal.

Step 2: Construct the empirical distribution of the structural errors

On the null hypothesis the $\{\varepsilon_t\}_{t=1}^{T}$ errors are omitted variables. Their empirical distribution is assumed to be given by these structural errors. The simulated disturbances are drawn from these errors. In some DSGE models the structural errors are assumed to be generated by autoregressive processes. This is the case with the New Keynesian model here; we discuss below the precise assumptions made.

[2]Some equations may involve calculation of expectations. The method we use here is the robust instrumental variables estimation suggested by McCallum (1976) and Wickens (1982): we set the lagged endogenous data as instruments and calculate the fitted values from a VAR(1)—this also being the auxiliary model chosen in what follows.

2.4 Use of Indirect Inference to check identification

It can be seen from this description of the Indirect Inference testing method that it is checking whether a certain assumed model could be the one generating the data description (the 'reduced form') we find. If we now allow for an infinite amount of data, we can exploit the asymptotic features of the procedure in large samples of which we have a large number in parallel; we may think of this as having access to a very large panel sample. These features imply that any reduced form will allow asymptotically exact estimates of the true structural parameters and that the true structural model if tested against this reduced form, will be rejected at exactly the correct size of the test- e.g. 5% for a 95% confidence test. It is this idea that we exploit to check for identification. By this is meant that there is an exact and unique correspondence between the parameters of the structural model and the parameters of the reduced form of the data it generates. For the structural model under question we seek to find some alternative set of parameters and accompanying error terms such that we can generate the same reduced form as the true model and its true error terms. Suppose we have some true model and errors and some data sample from these. To find the reduced form for some alternative set we take these alternative parameters and find the error terms that would enable it to replicate this data sample. This gives us the alternative structural representation of the model consistent with this data sample. We repeat this for many data samples from the true model, so that we have no shortage of data with which to estimate the reduced form of both the true model and the alternative one. We then test the hypothesis that the two parameter sets are the same on all these samples, giving us a very large number of tests- we do this by indirect inference. We know that if they are the same, then our test at say 95% confidence will reject 5% of the time. We check whether this is the case.

If we can find a parameter set for which there is no difference, we conclude the model is not identified. If we can find no such set, we conclude that it is.[3] We search in an area close to the true parameters; thus our check is for local identification in the vicinity of the model's true parameters. Of course one should therefore carry out this check for the final estimated parameters of the model.

In practice the reduced form of a DSGE model can take a variety of forms, as we have seen. How exactly we measure the data behaviour does not matter for the test's validity, provided we measure it in the same way for both True and False models. The only effect on the test would be on its power which would be reduced by a highly inaccurate measure; but since we are dealing with asymptotic results under very large data sets power can be made as large as we wish. While it is true that DSGE models can be given exact VARMA(i,j) representations, which will be highly accurate if well estimated for the appropriate orders i and j, it is usual to estimate a VAR representation on the grounds that it is both accurate and less prone to estimation difficulties (as posed by the MA components of VARMAs). Here we follow this practice and use VAR representations for our tests. It turns out that the test has high power against False models using VARs.

3 The numerical identification procedure

The idea behind our numerical procedure starts from the proposition that as data samples get larger and more numerous, the VARs estimated on them have mean VAR coefficients that tend to the true values- thus the true values are the probability limit of the actual values. This also implies that in the limit the distribution of the VARs generated by the DSGE model will converge on the true VAR distribution.

[3] It might be asked why we do not treat the errors as exogenously known, generate a lot of data samples for both sets of parameters and test whether their reduced forms are the same using the same indirect inference method. The answer is that the power of the test will be reduced since we are eliminating the altered error implied by the alternative parameters. When we include the altered error, the structural errors entering the reduced form will be identical only in the case where there is lack of identification; whereas if we always use the true errors, the structural errors will always be the same.

We desire the power of the test to be as high as possible to generate the clearest possible distinction between the reduced form parameters.

Thus with many samples of large size we can test by indirect inference whether a DSGE model is generating these VARs. At 5% significance the VAR distribution generated by the true DSGE model will be rejected 5% of the time; if there is another DSGE model that can produce the same VAR coefficients then it too will be rejected only 5% of the time (the choice of 5% is arbitrary; any significance level could be chosen). However if it cannot produce these VAR coefficients, it will be rejected much, even 100%, of the time.

What is found with Indirect Inference is that the power of the Wald test at any significance level can be raised as high as one wishes by a) increasing the order of the VAR description b) increasing the amount of data used by raising the sample size c) increasing the number of samples. Thus the test of identification can be made as precise as one wishes.

Le et al (2012) illustrate a) on the Smets-Wouters model. The figure below is taken from their paper and shows a Monte Carlo experiment where the coefficients of the true Smets-Wouters model are moved by x-percent up or down (alternately) from their true values; the false coefficients are then tested at the 5% level by the Indirect Inference Wald test. It can be seen that the power of the test as measured by the percentage of rejections rises sharply as the VAR order rises.

VAR — no of coeffs	TRUE	1%	3%	5%	7%	10%	15%	20%
3 variable VAR(1)-9	5.00	19.76	52.14	87.30	99.38	100.00	100.00	100.00
3 variable VAR(2)-18	5.00	38.24	68.56	84.10	99.64	100.00	100.00	100.00
3 variable VAR(3)-27	5.00	38.22	65.56	92.28	99.30	100.00	100.00	100.00
5 variable VAR(1)-25	5.00	28.40	77.54	97.18	99.78	100.00	100.00	100.00
7 variable VAR(3)-147	5.00	75.10	99.16	99.96	100.00	100.00	100.00	100.00

Table 1: Indirect Inference Rejection Rates at 95% level for varying VARs

To illustrate b) we carried out the same experiment on this same model, but this time holding the VAR order constant and raising the sample size steadily. We raise the number of variables in the VAR to the maximum of the 7 observables and the order (3) that best describes the data; in doing this we increase the power of the Wald test to its maximum. We provide 3000 sample draws from the True model; we then test whether x% deviations from the true parameters and the true error moments generate rejection. Our results are shown in Fig 1. .

What we see is that for a sample size of 100, the false model is rejected (at 95% confidence) 100% of the time when x reaches 15%. But when the sample size rises to 225, this point is reached when x is 10%. At a sample size of 400, it is reached when x is 5%; furthermore when x is only 1%, the rejection rate is 90%. What this shows rather clearly is that as the sample size rises only values very close to the true parameters can fail to be rejected; we can make the region of non-rejection as tight as we like by raising the sample size.

Finally, we illustrate c) from the numbers shown above. We carried out the power analysis shown in the last figure with only 3000 draws. However the analysis in Table 1 used 10000 sample draws of size 225. It can be seen there that the rejection rate at x=1% rises to 75% from 50% as the number of draws rises; while at x=3% it rises to 99% also from around 50%. Thus the power also rises with the number of sample draws.

Thus with unlimited data we can exactly recover the true model parameters from the parameters of the VARs on the data, because no other set of structural parameters can fail to be rejected. We can think of this as 'negative estimation' where we discard potential parameters that are rejected, isolating the only possible ones- here the true ones.

Thus we follow the following numerical procedure:

a) we generate by Monte Carlo sampling a large number of samples of large size from the true DSGE model being checked

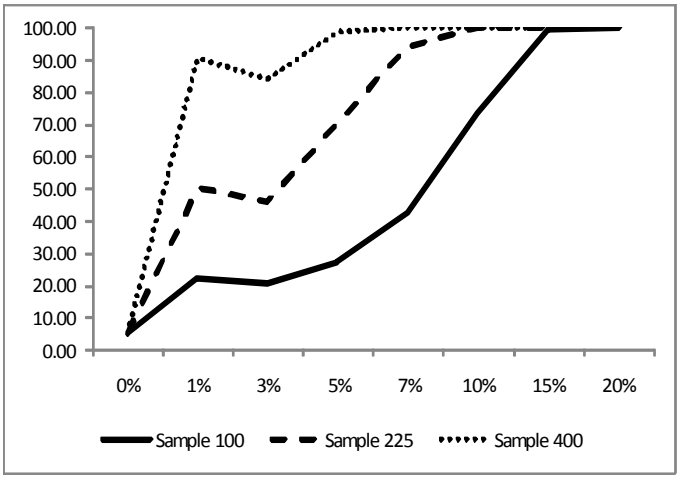

Figure 1: Power Functions for 95% (VAR(3) for 7 variables)

b) we compute the VAR distribution implied by these samples for a high order VAR on the maximum number of variables

c) we carry out a Wald test to check whether there are DSGE models in the region of the true model that are unrejected; if not we regard the DSGE model as identified.

Thus in this procedure we are combining testing with an estimation search for a DSGE model that can compete with the true model. This search should throw up other models if the model is not identified, since clearly changing the unidentified parameters away from their true values will not change the model's reduced form and so the VAR distribution it implies should be the same as that of the true model.

The procedure establishes local identification. Because the reduced form is non-linear in the DSGE model's parameters, it is possible for the rank condition to be satisfied locally yet fail to be satisfied for a different set of parameter values. Nevertheless it is important to establish local idetification as it implies that the model is identified in the region of its estimated parameter values.

We now carry out this procedure on:

a) the 3-equation model with autocorrelated disturbances which we have shown to be over-identified

b) the same model but where the errors are all $i.i.d.$ and ψ, ρ are both set to zero. In this case, the final one of the last section, we know the model is under-identified because ω cannot be retrieved.

These two cases are a check on how well the procedure can deal with the variety of identification possibilities including those where we know there is lack of identification.

3.1 a) The full New Keynesian 3-equation model

For our Monte Carlo experiment we use Model 1 and choose the parameter values shown in Table 2. These values are assumed to be those for the True model.

Using the parameter values set out in Table 2 (denoted by θ), we generate 1000 Monte Carlo samples of 500 observations each- the 'true data samples' from this DSGE model. We estimate a VAR(3) for the three variables, output, inflation and interest rates, on all 1000 samples. We use the resulting 1000 coefficient vectors, α_T, to construct the variance-covariance matrix Ω of the DSGE model's implied distribution for these coefficients and also for the Wald statistic

$$WS(\theta) = (a_T - \overline{a_T(\theta)})' W(\theta)(a_T - \overline{a_T(\theta)})$$

Parameter	Value	Parameter	Value
ω	0.7640	η	0.8830
σ	3.4550	ψ	0.0727
λ	0.0997	ρ_y	0.8654
ρ	0.4029	ρ_π	0.7999
γ	1.1624	ρ_r	0.7829

Table 2: Parameters of True model used in Monte Carlo simulations

where $\overline{a_T(\theta)}$ is the mean of the 1000 vectors and $W(\theta) = \Omega(\theta)^{-1}$ is the inverse of the variance-covariance matrix.

Next we examine the number of times a DSGE model with parameter vector θ_i is rejected by the true samples at the 5% level. Plainly if $\theta_i = \theta$ the rejection rate is 5%, using the Wald statistic just shown, by construction. For each other θ_i we take 1000 samples by Monte Carlo simulation and construct the Wald statistic distribution implied by that DSGE model as:

$$WS(\theta_i) = (a_S(\theta_i) - \overline{a_S(\theta_i)})' W(\theta_i)(a_S(\theta_i) - \overline{a_S(\theta_i)})$$

Now we calculate how often this DSGE model with vector θ_i is rejected on the 1000 true samples, whose VAR coefficient vectors are given by α_T. In effect this is done by counting the percent of values of

$$WS_T = (a_T - \overline{a_S(\theta_i)})' W(\theta_i)(a_T - \overline{a_S(\theta_i)})$$

that are greater than the 5% critical value of $WS(\theta_i)$.

We are looking for a vector θ_i that could generate the same reduced form coefficient vector, α_T, as the true vector θ; if so we could say that this θ_i could be 'confused with' the true vector θ. Numerically, we require that its rejection rate be at or arbitrarily close to 5%; in this case this model cannot be rejected as the true model any more than the true model itself. Since it is not the true model, this can only occur if it behaves just like the true model- ie the true model is not identified.

There is an important detail to be clarified about θ_i. This vector includes the error moments. Now if we choose the other parameters in θ_i, which are of course false, then the errors implied by this model are given by the true data samples interacting with these parameters. Thus we can extract the moments of these errors from the true data and the false other parameters; of course if we did not, then the θ_i would be rejected by the true sample data directly, as the error moments would be incompatible with both the data and the other structural parameters. Thus the error moments cannot be chosen freely given the true data and these other structural parameters.

To deal with this important detail we tackle the search for another θ_i with rejection rate at 5% in two stages.

First, we assume that the error moments are the true ones, and search on this assumption; lack of identification of a structural parameter will leave the structural errors unchanged. For example in the three-equation model ω is not identified because the term it multiplies in the model is always zero and so whatever it is the error term in that equation is unchanged. So this assumption would hold under lack of identification and searching for a set of parameters assuming it does could yield a set that could be confused with the true set.

Second, having found a candidate parameter set that satisfies our test under this assumption, we then extract the implied error moments. We do this by an exact iterative method, in which we generate the expectations in the DSGE model using the lagged errors and the other θ_i, extract the implied errors from the model and the data, and then reuse these as lagged errors to generate the expectations again, until convergence. We then redo the test using these extracted error moments- equivalently we draw repeatedly

Parameter	True	Found
ρ	0.4029	0.4057
γ	1.1624	1.1678
η	0.8830	0.9077
ψ	0.0727	0.0718
ω	0.784	0.7651
σ	3.4550	3.414
λ	0.0997	0.0981
ρ_y	0.8654	0.8727
ρ_π	0.7999	0.8124
ρ_r	0.7829	0.7782

Table 3: Table Caption

	Parameter		Rejection rate (at 5%) MC replications = 3000 Sample = 500	Parameter	Rejection rate (at 5%) MC replications = 6000 Sample = 400
	True	Varied		Varied	
λ	0.0997	0.0969	9.2000	0.0970	19.9000
σ	3.4553	3.3670	8.5667	3.4540	15.2000
γ	1.1624	1.1738	9.8333	1.1579	19.0000
η	0.8830	0.8728	10.033	0.8652	20.3333
ω	0.7640	0.6195	5.0000	0.7696	5.0000

Table 4: Checking on the non-identification of individual parameters in Model 3

with replacement from these extracted errors, of which we have in principle 500000 observations. At this stage if the false parameters have changed the error moments then rejection should result.

Our results for the 3-equation model were as follows.

We searched within a 3% region around the true θ, assuming that the error moments remained the true ones. We found the following set of parameters that gave a rejection rate of 5%:

We did find other sets but they only differed from this set by numerically tiny amounts. We went on to extract the error moments for this set and redid the test. The rejection rate rose to 97%. What this indicates is that had we extracted the implied error moments for each parameter set as we searched we would have found no set satisfying the test. However, this would have been a much more time-consuming search than our two-stage procedure.[4]

Now we had already established analytically above that this model was over-identified. Thus our numerical approach correctly confirms what we already know.

3.2 b) The New Keynesian model without persistence

We now turn to the version of the New Keynesian model where all persistence is removed from both the model and its error processes- Model 3. This leads to lack of identification of the parameter ω, as we have seen above. We should find that our test discovers this under-identification. We vary each parameter by a a small amount and see whether this variation can be rejected.

In the following Table we show for a small variation in each parameter from the true value how the rejection rate alters.

[4] We looked at this combined procedure for the SW model, as reported below (footnote to SW section), and found no set that failed to be rejected at high frequencies.

What this reveals is that indeed each parameter is identified except ω, as established analytically. Column 3 shows that the rejection rate roughly doubles the nominal rate with the false parameter with a sample size of 500, and 3000 Monte Carlo replications. With 6000 replications and a sample of 400 the rate rises further, roughly doubling again. This indicates clearly that as we raise the combination of the number of replications and the sample size we can push up the rejection rate. We may reasonably assume that we can push it up as far as we like by constantly raising both numbers.

3.3 c) Conclusions from the New Keynesian 3-equation models

What we have shown in these two exercises is that our numerical method accurately captures the identification of model parameters where we already know their status. Thus in the full model version where we know there is over-identification of all the parameters the method indeed finds that no other local parameter values can be accepted.

In the model version with no persistence again the method rightly finds that the ω parameter is not identified, as values for it other than the true value are not rejected at more than the 5% level chosen but that alternative values of all the other parameters are clearly rejected at higher rates and so are identified.

These results suggest that the method is reliable. We now go on to apply it to a case where we are unable to establish identification analytically.

4 Identification of the Smets-Wouters model

We now discuss an application of these procedures to the Smets-Wouters model, which we believe to be over-identified by the cross-equation restrictions imposed on the model by the RE assumption. However the model is far too large and complex to check identification analytically; and so we cannot be sure. Therefore in this section we use our suggested method to check it instead. This can be considered the first application where we try to bring fresh information about the identification of a major model in current use.

The Smets-Wouters model (2007) marks a major development in macroeconometric modelling based on DSGE models. Its main aim is to construct and estimate a DSGE model for the United States in which prices and wages, and hence real wages, are sticky due to nominal and real frictions arising from Calvo pricing in both the goods and labour markets, and to examine the consequent effects of monetary policy which is set through a Taylor rule. It may be said, therefore, to be a New Keynesian model. They combine both calibration and Bayesian estimation methods and use data for the period 1966Q1–2004Q4.

Unusually, the SW model contains a full range of structural shocks. In the EU version — Smets and Wouters (2003) — on which the US version is based, there are ten structural shocks. These are reduced to seven in the US version: for total factor productivity, the risk premium, investment-specific technology, the wage mark-up, the price mark-up, exogenous spending and monetary policy. These shocks are generally assumed to have an autoregressive structure. The model finds that aggregate demand has hump-shaped responses to nominal and real shocks. The model and its empirical performance is discussed in detail in Le et al (2011). Le et al (2011) find that the model does not fit well in the whole post-war sample- a key reason being that it generates too little inflation variability and too much output variability. They create a version in which there are competitive segments of the labour and product markets that are weighted with the imperfectly-competitive segments to create a hybrid model; such a model can fit the data post-1984 quite well and for the whole sample manages to pass the test at least at the 1% level. For our tests of identification here we assume that the SW model with high flexibility in the labour and product market is the true one; it does not seem to be critical to identification what version is used.

Our results for this model again revealed at the first stage, where we assumed the true error moments, that there was a candidate alternative set within 3% of the true parameters also giving a 5% rejection

Parameter	True	Varied	Parameter	True	Varied
φ	5.7400	5.7060	$\bar{\pi}$	0.7800	0.7643
σ_c	1.3800	1.3714	β	0.1600	0.1620
λ	0.7100	0.7086	$\overline{L}$	0.5300	0.5296
ξ_w	0.7000	0.7052	$\bar{\gamma}$	0.4300	0.4298
σ_L	1.8300	1.7861	α	0.1900	0.1906
ξ_p	0.6600	0.6634	ρ_g	0.9435	0.9434
ι_w	0.5800	0.5726	ρ_{gy}	0.6947	0.6986
ι_p	0.2400	0.2377	ρ_c	−0.1224	−0.1226
ψ	0.5400	0.5471	ρ_{inv}	0.3872	0.3853
Φ	1.5000	1.4843	ρ_{mon}	0.2387	0.2414
r_p	2.0400	2.0259	ρ_{prod}	0.9449	0.9411
ρ	0.8100	0.8072	ρ_π	0.1824	0.1792
r_y	0.0800	0.0802	ρ_w	0.0664	0.0656
$r_{\Delta y}$	0.2200	0.2262			

Table 5: Table Caption

rate.

We then went on to the next stage and extracted the implied error moments. Redoing the test raised the rejection rate to 97%.

This Monte Carlo experiment on the SW model again reveals that the model is identified, in the sense that no other set of parameters other than the true can fail to be rejected by the data generated by the model as the amount of this data is expanded without limit.

What we have just done is to search for a complete set of alternative values of the parameters that could fail to be rejected at more than the nominal (5%) rate. We have found none. Our search includes all sets where any subset is varied and the rest held constant; thus in principle it checks all combinations of variations in subsets of the parameters. [5]

5 Reconciliation with Canova and Sala

In their paper Canova and Sala drew attention to three problems: a) the possibility of 'non-appearance' of structural parameters in the reduced form b) the possibility of finding, with an infinite sample, that very similar impulse response coefficients emerge from maximum likelihood estimates generated from different sets of structural parameters c) that this problem is substantially worse with small samples. In this paper we have nothing to say about c); we have purely focused on identification with potentially infinite samples.

We have replicated a) in the case of the 3-equation New Keynesian model without persistence, Model 3. Here the coefficients on the expectations of future events disappear from the structural model because the expected variables become zero; they thus of course also disappear from the reduced form. We agree entirely with this potential problem and our numerical method can also in principle pick this problem up, as we show here in our numerical examination of it. In practice it might not be picked up by a search that was not carefully directed at particular parameters suspected of such disappearance. Thus one can entirely accept the need they urge to be conscious of such problems. Nevertheless, as we show, it is quite

[5] We also looked at combining stages a) and b), so that for each set of parameters searched through we back out the implied errors. This yielded no set of parameters at all that was not rejected at well over 5%, thus essentially confirming the results reported here.

unlikely that the 3-equation model will have this particular problem since it is rarely, if ever, that the model will have no persistence at all.

As for b), we have plainly not looked at partial reduced forms such as impulse response coefficients. However, as we pointed out in Le et al (2011, section 4.5, Table 10), the evaluation of impulse responses by our Wald test has much less power than evaluation of the full VAR coefficient set, which between them imply IRFs for all variables with all shocks. Thus the Table there shows that the SW hybrid model, which fails on the Wald test against the full set of VAR coefficients, nevertheless easily passes on the Wald test for certain small selections of IRFs. In effect the information contained in these IRFs is far less than that contained in the VAR coefficients. Thus by the same argument one would expect several 'false' models to generate similar IRFs to that coming from some 'True' model. Canova and Sala have thus shown that indeed this is so for certain IRFs popularly chosen to evaluate DSGE model performance. Thus it would seem rather likely that IRFs contain too little information to identify all structural parameters that should, according to our findings, be nevertheless identified by a full set of VAR coefficients.

A further element of reconciliation may be in the numerical procedure. We understand the Canova-Sala procedure for the infinite sample (population) case to be: generate a very large sample of data from some True model - estimate the likelihood of the true set of structural parameters using the Kalman filter- reestimate the likelihood for variations in the parameter values- identify areas/ridges where the likelihood is flat, these being where identification fails. Our procedure is: generate a large number of samples, each of large size, from a True model- estimate the VAR (reduced form approximation) on each- test using the Wald statistic whether varied parameter values generate VAR estimates that are the same as those in these samples- parameter sets where the rate of rejection is no worse than for the True model are where identification fails. The procedures differ in two main ways: a) the use of likelihood (ie closeness of the model 'predictions' to the data) versus the use of the Wald statistic (ie closeness of the VAR coefficients of the model to those in the data) b) how 'infinite' sampling is done- in their case with a very large single sample, in ours with many large samples. While b) does not seem to make any serious difference, a) could be important. Thus different structural parameters might get equally close to a set of data (ie have similar reduced form residuals), but might not share the same reduced form (ie have similar reduced form coefficients). This seems to be consistent with a finding of Le et al (2012) using Monte Carlo sampling for a true model of the SW type that the power of likelihood ratio tests was substantially less than the power of Wald tests of the type we use here.

To summarise, it would seem that our findings can be reconciled with the relevant ones of Canova and Sala. We agree with the possibility of 'disappearing' structural parameters that they identify; in principle these can be picked up by our method, but it would help the search to give the model a careful inspection for such a problem and use this to guide the search. We suggest the problem will be fairly rare in practice. We also agree with Canova and Sala that IRFs could fail to contain sufficient information to identify a DSGE model's structural parameters. However, what our findings suggest is that even so the DSGE models we have looked at are identified if a full VAR is used. Finally, it seems likely that the Wald statistic for VAR coefficients gives more discrimination between structural parameter sets than the likelihood criterion.

Of course with small samples and VARs of low order and few variables it still may be hard to distinguish different models using our available estimation methods. But this, we would suggest, is not a new problem, and to be distinguished from the pure identification issue per se.

6 Conclusion

Whether a DSGE model is identified is a matter of ongoing concern, and recently it has been suggested by Canova and Sala (2009) that identification may be a problem. As noted by Schorfheide (2011),

theoretical checks for identification are difficult to apply in practice; nevertheless both Iskrev (2010) and Komunjer and Ng (2011) have suggested practical methods using the rank condition. Here we suggest a numerical method based on estimation and testing by indirect inference for checking identification. The idea is to set up a Monte Carlo experiment in which a True DSGE model generates data samples; each of these has a VAR estimated on it and indirect estimation and testing is used to search for a set of False DSGE parameters that could fail to be rejected as generating these estimates. If the DSGE model is identified then no such False parameters can be found. If so, then we will have established that the reduced form (VAR) parameters implied by the DSGE model uniquely imply the True model parameters generating them: the condition for identification that the two parameter sets map exclusively into each other.

We have used the three equation model focused on by Canova and Sala as one example of this method in application. This is a good example since there are reasons to be concerned about identification in this model: the forward and lagged inflation terms and the autoregressive error term in the Phillips Curve can all substitute for each other in generating similar inflation behaviour. The cross-equation restrictions created by rational expectations nevertheless ensure over-identification, as we show analytically for this model.

Our numerical example confirms identification. In the Monte Carlo samples, no alternative parameter values fail to be rejected at levels sharply higher than the nominal 5% of the test statistic.

We also looked at the case where all dynamics in the model are suppressed, in which case the forward-looking parameter, ω, in the Phillips curve fails to be identified, whereas all the others are. Again our numerical procedure confirmed this, with the rejection rates rising sharply for small deviations from the true parameter values for all parameters other than ω.

We also looked at another example where identification seems assured by over-identification but we cannot establish this analytically because of the model's size, the Smets-Wouters model of the US economy. Here our numerical check again showed that as parameters leave their true values rejection rates rise rapidly.

We found that we could reconcile our findings with those of Canova and Sala, by appealing to three considerations: first, 'disappearing' parameters are possible but may be rare in DSGE models because of lag coefficients both in the model and in the error processes; second, impulse response functions may not contain as much information for identification as the full set of VAR coefficients we use here; third, the likelihood they use seems to be less well-determined than the Wald statistic we use here.

Plainly there are limitations to our results. For example, our method only establishes local identification of the DSGE model - i.e. identification in the region of its estimated parameters. Furthermore, our numerical results can always be vulnerable to holes in our search algorithms; though we have found a high degree of robustness so far, there is plainly much further work to be done to ensure that the parameter space has been thoroughly covered. Finally, we cannot deal in general with the possibility that a given DSGE model could be confused with an entirely different model, though of course any particular suggested model can be checked ad hoc by the method proposed here.

Nevertheless, we suggest that this numerical procedure could usefully be applied in empirical work to DSGE models when identification is in doubt. When we applied it here to two DSGE models widely used in applied macroeconomics, we found that they were identified. This suggests that, much as was generally believed prior to Canova and Sala, DSGE models are over-identified by virtue of the cross-equation restrictions imposed by rational expectations.

References

[1] Canova, F. and Sala, L. (2009). "Back to Square One: Identification Issues in DSGE Models", Journal of Monetary Economics, 56(4), pp. 431-449.

[2] Clarida, R., Gali, J. and Gertler, M. (1999). "The Science of Monetary Policy: A New Keynesian Perspective", Journal of Economic Literature, 37(4), pp. 1661-1707.

[3] Davidson, J., Meenagh, D.,Minf ord, P. and Wickens, M. (2010). "Why Crises Happen- Nonstationary Macroeconomics", Cardiff University, mimeo.

[4] Gali, J., Gertler, M. and Lopez-Solido, D (2005). "Robustness of the estimates of the hybrid New Keynesian Phillip Curve", Journal of Monetary Economics, 52(6), pp. 1107-1118.

[5] Komunjer, I. and Ng, S. (2011). "Dynamic Identification of Dynamic Stochastic General Equilibrium Models", Econometrica, 79 (6), 1995-2032.

[6] Iskrev, N. (2010). "Local Identification in DSGE Models", Journal of Monetary Economics, 57, 189-202.

[7] Le, V. P. M., Meenagh, D., Minford, P. and Wickens, M. (2011). " How Much Nominal Rigidity is There in the US Economy? Testing a New Keynesian DSGE Model using Indirect Inference", Journal of Economic Dynamics and Control, 35(12), pp.2078-2104.

[8] ———————————(2012). "Testing DSGE models by Indirect Inference and other methods: some Monte Carlo experiments", Cardiff Economics Working paper 2012/15; also CEPR discussion paper 9056.

[9] Rudd, J. and Whelan, K. (2005). "New tests of the New Keynesian Phillips Curve", Journal of Monetary Economics, 52(6), pp. 1167-1181

[10] Schorfheide, F. (2011). "Estimation and evaluation of DSGE models: progress and challenges", Working paper 11-7, Federal Reserve Bank of Philadelphia.

[11] Smets, F. and R. Wouters (2003). 'An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area,' Journal of the European Economic Association, 1(5),1123–1175.

[12] ———————————-(2007), 'Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach', *American Economic Review*, 97(3), 586-606.

[13] Wickens, M. (2011) Macroeconomic Theory- A Dynamic General equilibrium approach, Princeton University Press, 2nd edition.

[14] Working, E. (1927), 'What do statistical 'demand' curves show?' QJE, 51, 212-35.