A Monte Carlo procedure for checking identification in DSGE models

Vo Phuong Mai Le (University of Sheffield)*

Patrick Minford (Cardiff University and CEPR)†

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

March 2013

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.

Keywords: Identification; DSGE model; Monte Carlo; Indirect Inference

JEL Codes: C13; C51; C52; E32

*V.Le@sheffield.ac.uk; University of Sheffield, Department of Economics, 9 Mappin Street, Sheffield, UK. Tel: +44 (0)1142223405

†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

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.

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 et al. (2012). Le et al. (2011) sets out the full methodology of indirect inference.

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.

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. Here we use the numerical method to add information about this case.

The paper proceeds as follows. In section 2 we review the early work on identification by Working (1927) and show how it is altered by application to a DSGE model. In section 3 we consider the analytics of identification for the three-equation model in its various forms. In section 4 we explain our numerical method and apply it in detail to this model, to establish that it confirms these analytical results. In section 5 we apply our method to the Smets-Wouters model of the US. In Section 6 we reconcile our results with those of Canova and Sala (2009) and section 6 concludes.

2 Identification — the basics and the application to DSGE models:

We begin by considering the original approach to identification (originally due to Working (1927)). This turns out to be extremely helpful in understanding identification for DSGE models and also in explaining the essence of our proposed testing method.

Suppose that the demand and supply for a commodity is given by:

$$q_t = a_{11}p_t + a_{12}y_t + u_{dt} \tag{1}$$

$$q_t = a_{21}p_t + a_{22}z_t + u_{st} \tag{2}$$

where the as are constants, p is price, q quantity, and y and z are exogenous variables which affect demand and supply respectively. Demand and supply are assumed to be equal. Constants are omitted for simplicity. The equations are treated here as structural or behavioural equations and the as as structural coefficients. The issue of identification is: given that these two equations are both operative, can each be confused with the other? To put it another way, is it possible to distinguish between each equation and a linear combination of the two equations together?

The first point to note is that it is normal to suppress the error terms in this case and to treat the model as if it is deterministic; all variables take their expected values. The reason for this is that data is assumed to be available in as large quantities as needed to achieve exact estimation by whatever method is chosen. Thus the estimates, which have the status of means, will have a variance (due to the errors) that tends to zero, so that we may ignore the errors.

Form a linear combination of the two equations ignoring errors:

$$\mu q_t + (1-\mu)q_t = q_t = [\mu a_{11} + (1-\mu)a_{21}]p_t + \mu a_{12}y_t + (1-\mu)a_{22}z_t \tag{3}$$

Plainly the demand equation without its error is different from it because it has no term in z_t; similarly the supply equation without its error because it has no term in y_t. So if we fit these two equations to the data, we will get estimates of the four as.

Now suppose instead that the supply equation was:

$$q_t = a_{21}p_t + a_{22}y_t \tag{4}$$

Our linear combination of the two equations would then be:

$$q_t = [\mu a_{11} + (1-\mu)a_{21}]p_t + [\mu a_{12t} + (1-\mu)a_{22}]y_t \tag{5}$$

It is clear from inspection that now neither the demand nor the supply equation can be distinguished from this. When you estimate the demand equation you could be picking up the coefficients of the supply equation or the demand equation or any combination of the two; it is impossible to know what you have estimated across this set of possibilities. The same applies when you estimate the supply equation. Notice that it matters not what technique of estimation you use (TSLS, FIML, indirect via the reduced form); the point is you just do not know what you are picking up, however sophisticated your estimation method and however much data you have.

Although the error terms are different, we have no way of knowing this. We say that neither equation is identified.

Now suppose the supply equation was:

$$q_t = a_{21}p_t + a_{22}z_t + a_{23}y_t \tag{6}$$

Our linear combination would be:

$$q_t = [\mu a_{11} + (1-\mu)a_{21}]p_t + [\mu a_{12t} + (1-\mu)a_{22}]y_t + (1-\mu)a_{23}z_t \tag{7}$$

It is now clear that the demand equation is distinct from this (and so identified) while the supply equation is not (and so unidentified). What has enabled an equation, in this case the demand one, to be identified is a restriction it imposes; viz in this case the exclusion restriction on one variable.

Now consider the case of the model with this last supply equation. We obtain its reduced form as:

$$\begin{aligned} p_t &= \frac{1}{a_{11}-a_{21}}[a_{22}z_t + (a_{23}-a_{12})y_t] \\ q_t &= \frac{1}{a_{11}-a_{21}}[a_{11}a_{22}z_t + (a_{11}a_{23}-a_{12}a_{21})y_t] \end{aligned}$$

We can see that the two structural coefficients of the supply equation cannot be derived from the reduced form.

We can also see that if we substitute our linear combination of the two equations for the true supply equation we obtain the identical reduced form. Thus we may substitute the following parameters $a'_{21}, a'_{22}, a'_{23}$ which are given by the linear combinations above and we will get the identical reduced form.

Notice that if we substitute just one altered coefficient in the supply equation, even if we alter the error term so that it still generates the data for q, we will alter the reduced form.

Thus what we have established for the underidentified parameters is that there exists some set of alternative values for these parameters, different from the true parameters, that yield the exact same reduced form when the error term is altered consistently so that the altered equation is still equal to the supply data. We have also established that there is no such set for models where all parameters are identified or over-identified; any variation whatsoever in the structural parameters of such models, accompanied by the necessary changes in the error terms to fit the same data, must change the reduced form.

2.1 Identification with a DSGE model

Now we must see how this identification problem changes when we have a DSGE model.

Formally, we find that the typical DSGE model has few if any exogenous variables. Also it has error terms which are no longer the result of the model's 'inaccuracy' (i.e. due to the model's mistakes) but are rather exogenous variables deliberately omitted in order to focus only on the key causative influences in behaviour. Thus now the errors become the key exogenous variables of the model — usually the only ones. In principle we now just have exogenous variables for the model, and no 'mistakes'/'errors'; if unusually the model has normal exogenous variables directly observed from the data these are just

another exogenous variable, treated just the same as the error-exogenous variables. For simplicity we will assume from now on that all the exogenous variables are error terms.

Notice that this completely transforms the treatment of 'errors' given a mass of data. Since these errors are exogenous variables there are no longer any 'errors' that can be disregarded with a mass of data. All that a mass of data does is to give us a mass of potential paths for exogenous variables, just as we would for the Working model with the potential paths for exogenous variables there. There we examined identification conditional on knowing the exogenous variables. Essentially we can think of the reduced form solution being a function of the current exogenous variables (and of their expected future values, which are also a function of the current and lagged values). Thus we can ask, just as with the Working model above, whether, conditional on the exogenous variables, the model is identified.

But now notice that the exogenous errors are only observed indirectly, by extraction from the model and the data, given the model parameters. Thus if we postulate an alternative set of values, this will also generate a different set of assumed exogenous errors. Thus with DSGE models the exogenous errors are themselves functions of the parameters of the model, given the data.

To fix ideas, rewrite the Working model as follows solely in terms of errors:

$$q_t = a_{11}p_t + u_{dt} \tag{8}$$

$$q_t = a_{21}p_t + u_{st} \tag{9}$$

Here if we regard the errors as directly observed exogenous variables, then plainly the model is identified, since no linear combination of both equations can be confused with either equation — simply because their (exogenous) 'error terms' will be different.

Now suppose the demand equation is the same but that the supply equation becomes:

$$q_t = a_{21}p_t + a_{22}u_{dt} + a_{23}u_{st} \tag{10}$$

It is now clear that a linear combination of the two equations is indistinguishable from the supply equation: thus if we substitute for the true supply equation above the linear combination

$$q_t = [\lambda a_{11} + (1-\lambda)a_{21}]p_t + [\lambda + (1-\lambda)a_{22}]u_{dt} + (1-\lambda)a_{23}u_{st} \tag{11}$$

we obtain exactly the same reduced form which is:

$$\begin{bmatrix} p_t \\ q_t \end{bmatrix} = \frac{1}{a_{11} - a_{21}} \begin{bmatrix} -(1-a_{22}) & a_{23} \\ -(a_{21} - a_{11}a_{22}) & a_{11}a_{23} \end{bmatrix} \begin{bmatrix} u_{dt} \\ u_{st} \end{bmatrix}$$

This exactly replicates Working's result, that when the supply equation does not 'exclude' the demand error it fails to be identified.

If we knew the true errors, u_{dt} and u_{st}, by direct observation, this would exactly parallel the analysis of the Working model with the errors now acting as exogenous, directly observable, variables. However, with a DSGE model we can only observe the errors indirectly, by extracting them from the structural equation, its parameters and the data. If we change the supply equation, we must therefore change our indirectly observed supply error.

Thus, if the true model is as above, but we now create a linear combination of the two equations and substitute it for the true supply equation, we obtain the following implied supply equation:

$$q_t = a'_{21}p_t + a'_{22}u_{dt} + a'_{23}\widehat{u}_{st} \tag{12}$$

where $a'_{21} = [\lambda a_{11} + (1-\lambda)a_{21}], a'_{22} = [\lambda + (1-\lambda)a_{22}], a'_{23} = (1-\lambda)a_{23}$
and:

$$\widehat{u}_{st} = \frac{1}{a'_{23}}\{a_{23}u_{st} - [\lambda(a_{11} - a_{21})]p_t - \lambda(1-a_{22})u_{dt}\} \tag{13}$$

The reduced form of this model is now expressed as:

$$\begin{bmatrix} p_t \\ q_t \end{bmatrix} = \frac{1}{a_{11} - a'_{21}} \begin{bmatrix} -(1 - a'_{22}) & a'_{23} \\ -(a'_{21} - a_{11}a'_{22}) & a_{11}a'_{23} \end{bmatrix} \begin{bmatrix} u_{dt} \\ \widehat{u}_{st} \end{bmatrix}$$

But in the case where the parameters $a'_{21} = [\lambda a_{11} + (1-\lambda)a_{21}]$, $a'_{22} = [\lambda + (1-\lambda)a_{22}]$, $a'_{23} = (1-\lambda)a_{23}$, i.e. exactly in the linear combination that can be confused with the true supply equation, then we can verify that this collapses to:

$$\begin{bmatrix} p_t \\ q_t \end{bmatrix} = \frac{1}{a_{11} - a_{21}} \begin{bmatrix} -(1 - a_{22}) & a_{23} \\ -(a_{21} - a_{11}a_{22}) & a_{11}a_{23} \end{bmatrix} \begin{bmatrix} u_{dt} \\ u_{st} \end{bmatrix}$$

Thus we see here that the lack of identification shows up as an identical reduced form, even though it was generated by this different set of structural parameters and the implied different extracted error.

It is this idea that we exploit for our test. For any model 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.[1] We search in an area close to the True parameters, exploiting the feature of linear models that local implies global identification; thus the presence of a linear combination of equations that can be confused with one of the model's equations will be detected locally and if not present locally will not be present globally either.

In practice the reduced form of a DSGE model can take a variety of forms; we illustrate this below with the 3-equation New Keynesian model. The role of the reduced form, however it is expressed, is to capture the behaviour found in the data generated by the structural model. Identification then fails if another structural model can generate data that has the same behaviour. Our test asks whether another False model could generate the data behaviour that was generated by the True model; it does this by finding via simulation what the distribution is of the data behaviour coefficients for the False model compared with what it is for the True model; if these distributions are not different according to our test, we declare a lack of identification. Effectively we are testing whether the False parameters can be regarded as True according to the Indirect Inference test. 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. 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

[1] 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.

by the MA components of $VARMA$s). 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 Analytical identification of the three equation New Keynesian model

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

$$\pi_t = \omega E_t \pi_{t+1} + (1-\omega)\pi_{t-1} + \lambda y_t + \varepsilon_t^\pi \tag{14}$$

$$y_t = E_t y_{t+1} - \frac{1}{\sigma}(r_t - E_t \pi_{t+1}) + \varepsilon_t^y \tag{15}$$

$$r_t = \rho r_{t-1} + (1-\rho)(\gamma \pi_t + \eta y_t + \psi (y_t - y_{t-1})) + \varepsilon_t^r \tag{16}$$

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. A recent 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):

$$\pi_t = \omega E_t \pi_{t+1} + \lambda y_t + e_{\pi t}, \quad \omega < 1 \tag{17}$$

$$y_t = E_t y_{t+1} - \frac{1}{\sigma}(r_t - E_t \pi_{t+1}) + e_{yt} \tag{18}$$

$$r_t = \gamma \pi_t + \eta y_t + e_{rt} \tag{19}$$

$$e_{it} = \rho_i e_{i,t-1} + \varepsilon_{it} \quad (i = \pi, y, r)$$

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 $\gamma > 1$, $\lambda_1\lambda_2 < 1$ and $\lambda_1+\lambda_2 < 1$ we have $\lambda_1, \lambda_2 < 1$. 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.

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 fundamentally atypical. [2]

[2] 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 — e.g.

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

$$\begin{aligned}\Delta(L) &= \left(\frac{\sigma\rho+\psi}{\sigma}\right)(1-\omega)L^2 - \left[\left(\frac{\sigma\rho+\psi}{\sigma}\right)+(1-\omega)\left(1+\rho+\frac{\eta+\psi}{\sigma}\right)\right]L \\ &\quad + \left(1+\rho+\left(\frac{\eta+\psi}{\sigma}\right)+\omega\left(\frac{\sigma\rho+\psi}{\sigma}\right)+\frac{\rho\lambda+(1-\rho)\lambda\gamma}{\sigma}+1-\omega\right) \\ &\quad - \left[\frac{\lambda}{\sigma}+\omega\left(1+\frac{\eta}{\sigma}\right)\right]L^{-1}+\omega L^{-2}\end{aligned}$$

Normally it will have two forward roots and two backward roots, all inside the unit circle. For example with $\omega = 0.8$, $\rho = 0.4$, $\sigma = 3.5$, $\lambda = 0.8$, $\gamma = 1.2$, $\eta = 0.9$, $\psi = 0.1$, 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}.

These examples illustrate that, while in principle we can establish the identification of a DSGE model analytically, and in the case of this 3-equation model we can do so in practice, in general as the number of equations rise this will rapidly become impracticable. Thus with the Smets-Wouters model in loglinearised form, our other example in this paper, analytical solution is not a practical matter, as the model is too large. This is our motivation for seeking a numerical way of resolving identification. We base our procedure on indirect inference about the structural parameters. Canova and Sala (2009) also take a numerical route in their discussion of identification which is based on the properties of the impulse responses implied by the data that are estimated by maximum likelihood methods. We argue below that the choice of model features to estimate is important for a numerical approach to weak identification. We choose a VAR to characterise the data, and the VAR coefficients as the relevant data properties; and we use indirect inference as the basis of our estimation procedure. We argue that this enables us to check the identification of DSGE models rather precisely.

As noted in section 1, the reduced from 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

$$\begin{bmatrix} I & -A \\ 0 & I \end{bmatrix}\begin{bmatrix} z_t \\ e_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix}\begin{bmatrix} z_{t-1} \\ e_{t-1} \end{bmatrix}$$

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.

where R is the matrix with ρ_i along the diagonal and zeros elsewhere.

Finally we can transform the model into a VAR by approximating each of the MA errors by an MA(2) in a weighted average of the three errors; then divide this MA(2) process into the left-hand side AR(3) process. The resulting ratio of polynomials in L, which gives rise to an infinite order AR process can be approximated by a finite order AR process to yield the VAR. Typically this is found to give an accurate approximation of the VARMA.

4 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.

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 mean VAR coefficients then it too will be rejected only 5% of the time. However if it cannot produce these mean VAR coefficients, it will be rejected much, even 100%, of the time.

The choice of 5% is arbitrary; any significance level could be chosen. 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.

Le et al. (2012) illustrate a) on the Smets-Wouters model. The table 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.

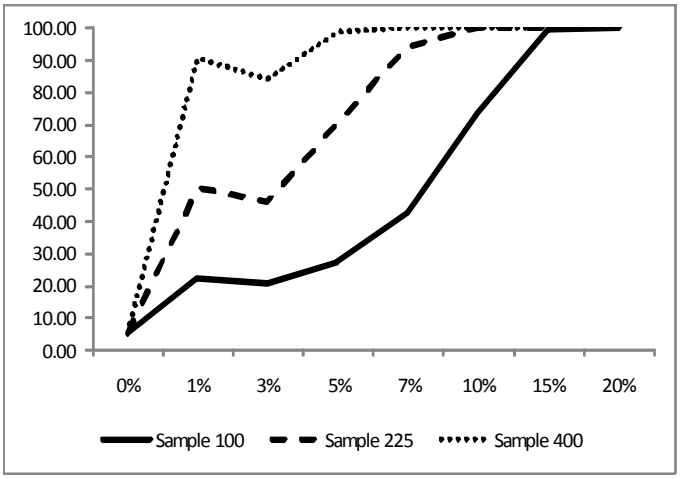

Figure 1: Power Functions for 95% (VAR(3) for 7 variables)

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 10,000 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

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 identification locally; this is necessary and sufficient for a linear model to be identified globally when the reduced form representation is a structural VAR (i.e. one whose errors are linear combinations of the structural errors). This is the case we are examining here.

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.

4.1 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.

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

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

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

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 — i.e. 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 with replacement from these extracted errors, of which we have in principle 500,000 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%:

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: 3-Equation Model Parameter Values

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.[3]

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

4.2 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.

[3] 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.

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

	Parameter		Rejection rate (at 5%)	Parameter	Rejection rate (at 5%)
	True	Varied	MC replications = 3000 Sample = 500	Varied	MC replications = 6000 Sample = 400
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

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.

4.3 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.

5 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.

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: Smets-Wouters Model Parameter Values

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 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.[4]

[4] 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.

6 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 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 (i.e. closeness of the model 'predictions' to the data) versus the use of the Wald statistic (i.e. 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 (i.e. have similar reduced form residuals), but might not share the same reduced form (i.e. 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.

7 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. Thus in practice identification has to be checked by whether one can retrieve the structural parameters from estimation; this in turn places the method of estimation in a key role. Here we suggest a numerical method based on estimation and testing by indirect inference for checking whether a DSGE model is identified. 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, we only consider linear (or linearised) models for which local and global identification coincide. 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.

8 Acknowledgements

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

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] Christiano, L.J., Eichenbaum, M., Evans, C.L. (2005). "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", *Journal of Political Economy*, 113(1), 1-45.

[3] 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.

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

[5] 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.

[6] 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.

[7] Le, V. P. M., Meenagh, D., Minford, P. and Wickens, M. (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.

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

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

[10] 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.

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

[12] Working, E. (1927), 'What do statistical 'demand' curves show?' *Quarterly Journal of Economics*, 51, 212-35.