



# Testing DSGE Models by Indirect Inference: a Survey of Recent Findings

David Meenagh<sup>1</sup> · Patrick Minford<sup>1,2</sup> · Michael Wickens<sup>1,2,3</sup> · Yongdeng Xu<sup>1</sup>

Published online: 28 February 2019  
© The Author(s) 2019

## Abstract

We review recent findings in the application of indirect inference to DSGE models. We show that researchers should tailor the power of their test to the model under investigation in order to achieve a balance between high power and finding a robust model; this will involve choosing only a limited number of variables on whose behaviour they should focus. Also recent work reveals that it makes little difference which these variables are or how their behaviour is measured whether via a VAR, IRFs or moments. We also review identification issues, how to test part of a model and whether alternative evaluation methods such as forecasting or likelihood ratio tests are potentially helpful.

**Keywords** Pseudo-true inference · DSGE models · Indirect inference; Wald tests · Likelihood ratio tests; robustness

**JEL Classification** C12 · C32 · C52 · E1

---

✉ David Meenagh  
Meenaghd@cf.ac.uk

Patrick Minford  
Patrick.minford@btinternet.com

Yongdeng Xu  
Xuy16@cardiff.ac.uk

<sup>1</sup> Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff CF10 3EU, UK

<sup>2</sup> CEPR, Cardiff, UK

<sup>3</sup> University of York, York, UK

## 1 Introduction

Indirect Inference is a method for estimating and testing models of any size, complexity or nonlinearity by comparing the behaviour of data simulated from the model with that of selected observed data where behaviour is summarised by a set of descriptive features, known as the auxiliary model. In estimation the parameters of the structural model are chosen to best match its simulation with the selected data. In testing the simulation of the model with a given set of parameter estimates is compared with the selected data. This gives a joint test of the given parameter values and the structure of the model. Rather than using asymptotic theory, by simulating the model many times, in estimation, it is possible to derive numerical distributions of the estimated parameters and, in testing, the numerical distribution of the test statistic can be employed.

The use of indirect inference in estimating and testing macroeconomic and other structural models is not a well-known empirical tool among economists. It attracted attention in the 1990s as a way of estimating nonlinear models and a variety of papers were written then on its estimation properties, essentially seeking to discover whether it could mimic the asymptotic properties of maximum likelihood (it did). But its potential role in testing models, especially in small samples, and its properties as an estimator in small samples, was not explored until the work described here which began around 2000. It is particularly useful for testing DSGE models which, as they are commonly estimated by Bayesian methods, tend not to be subjected to any formal statistical test. Le et al. (2016a) have provided a survey of testing DSGE models using indirect inference. In this paper we extend their survey by covering recent research which addresses a number of questions that have arisen in the use of indirect inference.

The earlier survey discussed the power of indirect inference tests in small samples, comparing this with the power of tests based on frequentist likelihood methods. It also suggested ways in which modellers could use this power to determine the robustness of their model, especially when used for policy analysis. It was assumed in that survey that the auxiliary model would be a VAR of low order in a few variables while the structural model would be a DSGE model, whether of small size as in Clarida et al. (1999), or large as in Smets and Wouters (2007). Since then the method has been applied to microeconomic and trade models where the features of the structural model are quite different; we do not explicitly consider these models here, but the questions we examine and the answers we give can be easily extended to them.

We address the following issues. In section 2 we discuss some important methodological issues in testing DSGE macroeconomic models such as whether it is worth testing DSGE models when, as some claim, they are irredeemably mis-specified a priori and, if it is, how best to estimate and test DSGE models. In section 3 we compare the power of indirect inference tests with the main alternative frequentist method, the likelihood ratio test in small samples. Asymptotically, the two cannot be distinguished under the usual assumption that the model is true. In part, but for completeness, this section recaps material from our earlier survey. In section 4, we review the evidence related to a variety of issues that arise when carrying out the indirect inference test, such as how many and which variables should be used in the auxiliary model, and whether moments, IRFs or a VAR are preferable auxiliary models. In section 5 we review the relationship between indirect inference and identification. In section 6 we conclude with a review of our main findings. Some of these issues have been addressed in

published papers, including our own, in which case we have drawn on these in summary form to bring them into this conspectus of available work. We hope this will be helpful to applied economists using indirect inference.

## 2 Methodological Issues in Testing DSGE Macroeconomic Models

### 2.1 Why Test a Mis-Specified Model?

It is nearly 40 years since DSGE models (in the form of RBC models) were first used in macroeconomics and, arguably, are now the standard approach. Throughout this time, and continuing to this day, their use has been heavily criticised from a number of perspectives. We will consider some criticisms often made by classically-trained econometricians.

A philosophical question in testing DSGE models raised by some econometricians is whether these models should be dismissed because they are so ‘unrealistic’ that one must consider them as inherently ‘mis-specified’ and so cannot be treated as ‘true’ for testing purposes. There is an irony in this as the impetus behind the use of DSGE rather than traditional macroeconomic models was Lucas’s critique that the latter were essentially reduced form and not structural models and therefore could not be used for policy or control purposes.

Influenced by the Lucas critique, the focus in DSGE models is on structural modelling (estimating deep structural rather than “reduced-form” parameters). This has resulted in DSGE models being smaller and simpler than traditional macroeconomic models, especially in their dynamic specification. The smallness of DSGE models reflects the influence of Friedman’s (1953) advocacy of the use of simple rather than complicated models but it makes them more likely to be rejected by the data than traditional macroeconomic models. Traditional models usually have a flexible, and sometimes data-determined, dynamic specification which gives them a better chance of passing statistical tests, though it raises the question of whether this is due to data-mining. An example of a DSGE model that focuses on structural specification with hardly any structural dynamics is the Smets and Wouters (2007) DSGE model of the United States. Significantly, the disturbances in this model are very highly serially correlated, which could indicate severe dynamic mis-specification. The rejection of DSGE models using conventional testing procedures and the frequent presence of highly serially correlated structural disturbances, have all contributed to undermining the high ideals originally envisaged for the DSGE approach to macroeconomic modelling and may explain why some econometricians regard DSGE models as mis-specified and of little empirical relevance.

The issue of the ‘truth’ of models has a long history in economics and statistics. It was also addressed by Friedman (1953) who argued that the appropriate criterion for judging economic models is their predictive ability and not their ‘realism’ or ‘truth’. Rather than dismiss DSGE models as ‘incredible’, as some have done, or accept that there is no point in testing them because they would fail the test, it would be better to find a way of putting them on firmer statistical foundations. Thus a test of these models should not be about whether or not they are ‘true’ but whether they are ‘pseudo-true’ meaning able to match the data of interest. This idea was developed from testing non-

nested hypotheses by Cox (1961, 1962). Indirect inference tests proceed on this basis; the model is treated as though true and, if rejected, is assumed to be mis-specified. The usefulness of the test is judged by its power which, in the case of indirect inference tests, we have found to be very high and superior to likelihood based tests.

## 2.2 How Best to Test and Estimate DSGE Macroeconomic Models?

The issue of how best to test and estimate DSGE models also has a long history. In an interview with Evans and Honkapohja (2005) Tom Sargent, referring to rational expectations models, remarked that Lucas and Prescott thought that likelihood ratio tests were rejecting “too many good models” and that “calibration was intended as a balanced approach to professing that your model, though not correct, is still worthy as a vehicle for quantitative policy analysis”. The clear implication is that DSGE models, although possibly mis-specified, may still be ‘good’ in some a priori sense even if rejected statistically, and that calibrating rather than estimating DSGE models using classical econometric methods was a way of imposing a priori knowledge of the model parameters.

If traditional macroeconometric models are not structural but, due to the flexibility this allows, particularly in their dynamic specification, they can be specified in such a way that they pass statistical tests, whereas DSGE models are structural, deliberately simple and, because they are usually rejected using classical inference, strong prior restrictions are imposed in their estimation. One reason why classical likelihood methods are likely to reject mis-specified models is that they take account of all of the model’s features and parameters, and not just its key features which may be well specified. For example, the RBC model’s long-run properties may be good but the short-run properties may be poor because there is little information on how to specify them. Consequently, failure to allow sufficient lags in the dynamics could easily result in rejection even though the long-run characteristics of the model may be acceptable. Likelihood methods deal with this by just choosing the parameter values to best fit the dynamics revealed in the data and may generate biased estimates in order to do so. This has led to the counter criticism that conventional econometric models and methods are likely to involve a strong element of data-mining. In effect, Lucas and Prescott were looking for a way to give more weight to some features of the model than others.

Subsequently a compromise was reached between the strong use of a priori knowledge in calibration and classical estimation through the use of Bayesian methods. If the mode of the posterior distribution is chosen as the point estimate of a parameter then, in effect, the Bayesian estimate is a weighted average of the prior and the likelihood functions where the weights reflect the precision of the two types of information; the tighter the prior or the the less informative the data likelihood, the more the posterior estimates are dominated by prior beliefs.

As a result of using Bayesian methods to estimate DSGE models, the models are rarely tested for mis-specification. This is despite the fact that the models are almost certainly to some degree incorrect. It is usually regarded as sufficient that the coefficient estimates are significantly different from zero as judged from their posterior distribution. It is common to find that the priors for the parameters are close to their posterior distributions. This could be interpreted in two ways. Either it implies that the priors are very strong relative to the data likelihood and so dominate the posterior thereby

undermining this check on model specification. Or, the data likelihood is small implying that the model would not fit the data using classical inference which suggests either that the model is mis-specified, or that these data are not very informative.

Indirect inference is ideally suited to testing a previously estimated model such as a Bayesian estimated DSGE model. As we know that a DSGE model is not a ‘true’ representation of an economy the test should be interpreted not as whether the model is true’ but whether it is ‘pseudo-true’ and hence provides a sufficiently good representation of the data as determined by the probability that the model is not rejected by the test. As useful as this is, it is still somewhat of a blunt instrument as, in the event that the model is rejected, it does not reveal which features of the model are causing the rejection. We have attempted to address this issue in Minford et al. (2019) by proposing an indirect inference test of parts of a model rather than the whole model. We found that the rejection of the whole Smets-Wouters model of the United States (Smets and Wouters 2003) was likely to be due to misspecification of the wage-price sector of the model.

### 2.3 The Auxiliary Model

The early, calibrated, RBC models were evaluated against the data by comparing data simulated from the calibrated model with observed data by the method of “matching the moments”. This consists of inspecting the variances, covariances and serial correlation of the two sets of data. Inspecting the moments is not a formal statistical test but it does illustrate the use of an auxiliary model to evaluate a model by comparing simulated and observed data. In order to obtain a formal statistical test it is necessary to derive a single statistic that summarises the model. Indirect inference provides such as test through a comparison of the properties of an auxiliary model that is estimated on the simulated and the observed data. There is more than one choice of which property to base the test on. Examples are (i) the scores, (ii) the estimated coefficients of the auxiliary model, and (iii) the impulse response functions.

A natural choice of the auxiliary model is the VAR solution of a DSGE model: a detailed example is provided in Appendix 1.

It can be shown that a linearised DSGE model of general form

$$E_t y_{t+1} = A y_t + C z_t$$

where the variables  $y_t$  are determined by the model and  $z_t$  are exogenous variables, including shocks, has the solution

$$y_t = P y_{t-1} + \sum_{s=0}^{\infty} Q_s E_t z_{t+s} + R z_{t-1} + S \xi_t$$

where  $\xi_t = y_t - E_{t-1} y_t$ . And if the exogenous variables are generated by the VAR

$$z_{t+1} = B z_t + \varepsilon_{t+1}$$

then  $E_t z_{t+s} = B^s z_t$ , ( $s > 0$ ) and so the solution is the VARX

$$y_t = P y_{t-1} + H z_t + R z_{t-1} + S \xi_t$$

and the complete data set  $(y_t, z_t)$  is the VAR

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = M \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + N \begin{bmatrix} \xi_t \\ \varepsilon_t \end{bmatrix}.$$

Hence, a VAR is a natural choice of auxiliary model.

We may note in passing that the form of the solution of DSGE models implies that their forecasting performance will be similar to that of a VAR unless there is better information available about future values of the exogenous variables, for example, where the exogenous variables are policy variables and there is a credible and accurate announcement of the future policy stance. Supporting evidence for this is provided by Wieland and Wolters (2012) and Wickens (2014).

We have based our indirect tests on a comparison of the VAR coefficients from simulated and observed data. Using impulse response functions would just involve using functions of these coefficients. We have also found that using VARs of low order in only a few variables provide an indirect test with good power even for a large model such as that of Smets and Wouters which has a reduced form solution that is a VAR of fourth order in seven variables. And we have found that using numerical estimates of the distribution of the test statistic gives tests of higher power than those based on the asymptotic distribution of the test statistic. This may reflect the small sample sizes available for estimating and testing DSGE models consisting of 200 or less observations.

Although a formal test of a DSGE model should clearly be carried out, matching the moments in the way used to evaluate RBC models does have some advantages. Whereas testing the whole model via a single statistic can be carried out in a way consistent with statistical theory, inspecting individual elements of the model can help reveal which parts of the model best capture the data and which don't. Thus, as discussed in more detail in Wickens (2012), the basic RBC model based on a single productivity shock such as that of King et al. (1988) produces simulated moments which show that consumption is not smoothed as much as the theory predicts, and that real wages are not as flexible as predicted. Much of the subsequent development of DSGE models was to produce models that fitted the data for these features. Another way of focussing on the aspects of the model that are inconsistent with the data, and do so in a formal statistical manner is, as noted earlier, to test parts of the model. .

### 3 Testing DSGE Models by Frequentist Methods-Indirect Inference and the Likelihood Ratio

In our previous survey we devoted much space to comparing the standard 'direct inference' test, the likelihood ratio test (LR), to the indirect inference test (IIW) in the context of small samples. (The two test procedures are explained in detail in Appendix 2). We showed two main things. First, that the LR test had no power against a mis-specified model, whereas the IIW test had very high power, effectively rejecting a mis-specified model 100% of the time. Secondly, in the situation where the model is well specified but the parameters are numerically inaccurate, we found that the IIW test had considerably more power, rejecting models whose parameters were randomly

wrong by 5–7% for 100% of the time, whereas the LR test rejects such models with only small frequency; to reject 100% of the time it needs parameters to be wrong by as much as 20%.

We explained these results by noting that the LR test, when faced with a numerically incorrect model, is similar in power to an indirect inference test where the distribution of the auxiliary model parameters - e.g. the VAR coefficients - is found from the data, as opposed to the restricted model. Thus, here, the IIW test is simulated from the structural model being tested, and the distribution of the VAR coefficients that this structural model implies is used in the test. Plainly this distribution closely reflects the features of the structural model. By contrast, the LR test takes the distribution implied by the sample data, as generated by the (unknown) true model.

The low power of the LR test when the model is mis-specified (but has unrestricted error dynamics) has a different cause. When one tests such a mis-specified model by LR, one first has to estimate it by ML. Thus one is asking: is this (mis-specified) structural model generating the data sample? To give the model the best chance of passing the test, first one re-estimates the model together with its unrestricted error dynamics. In finding a model with best fit, ML will use the estimates of the error processes thereby minimising the forecast error that the LR test is based on. As a result, finds a well-fitting model by substituting fitted error processes that are good at creating low forecast error. Thus it is hard to distinguish the mis-specified model from the true model using the LR test. Moreover, the data sample from the true model will also easily fit the mis-specified model. By contrast when one re-estimates the mis-specified model by II, there is simply no way the structural model can generate the same data behaviour as the true model. Whatever estimates are found will generate different reduced form behaviour. Hence the very high rejection rate. We show the key results from our previous paper on this and report some further findings.

### 3.1 Mis-Specification

First we show the experiment where we test a mis-specified model. The Table below shows the results when the true (Smets-Wouters) model is New Keynesian and the mis-specified is a New Classical version and for the reverse case where the true model is New Classical and the mis-specified model is New Keynesian. As can be seen, the power of the II test is close to 100%, whereas that of the LR test is zero (Table 1).

We now investigate a more subtle form of mis-specification. It takes the form of a failure to include in our model features from a more complex model that is in fact generating the data. *A fortiori* an LR test would again be unable to detect this more subtle form of mis-specification.

**Table 1** Power of the test to reject a false model

	Percentage Rejected	
	NK data NC Model	NC data NK Model
II	99.6%	77.6%
LR	0%	0%



We set up a Monte Carlo experiment in which the data generating process (DGP) is a more complex model. Our starting point will be, as in Le et al. (2011, 2016a), the Smets-Wouters model of the US where the data are from the early 1980s. Le et al. (2011, 2016b) found that this model when modified to allow for a competitive sector and for banking, can explain well the main US macro variables: output, inflation and interest rates. It is this model, and versions similar to it, that we have used in previous Monte Carlo experiments.

To this model we can add money and a regime shift contingent on the state of the economy, from the Taylor Rule to the zero bound, as in Le et al. (2016b). This makes the model's parameters state-contingent so that it has this form of nonlinearity. We then treat this nonlinear model as if it were the DGP generating the data. Using the indirect inference test with a VAR as the auxiliary model we estimate the power function for the falseness criterion described above in order to assess the sensitivity of the power function to the presence of greater nonlinearity in the true model than the 'assumed true' DSGE model we started with.

We looked at three very similar models, of varying complexity. All three are based on the Smets-Wouters model as modified in Le et al. (2011). Model 1 is that model exactly. Model 2 is, that model with the financial shock replaced by the Bernanke et al. (1999) model of banking (the 'financial accelerator'). Model 3 is the same model, together with an extension in which collateral is required and base money acts as cheap collateral, and the additional nonlinearity of the zero bound constraint, triggered whenever the Taylor Rule interest rate falls below a low threshold. These last two models are set out in Le et al. (2016b).

From the point of view of 'realism' and 'truth' we regard model 3 as the most realistic, model 2 as a linear approximation to it, and model 1 as a simpler approximation to model 2. We investigate whether in each case the simpler, less realistic, model can be treated as a valid approximation to the more realistic one.

We carried out the following three experiments with sample sizes between 75 and 200, and with 1000 sample replications. Our IIW test was based in all cases on the coefficients of a three variable VAR(1) in  $(y, \pi, R)$ ; as is the usual practice in applying these tests, it also includes the three variances making 12 parameters in all. Table 2 shows that in all cases there is an overwhelming probability of rejection, close to 100% and falling to 80% with the smallest sample size of 75 and the two models closest in complexity (models 1 and 2).

The models used for this experiment were those that Le et al. (2011, 2016a, b) where the parameters were estimated by indirect estimation using US data. In practice, when we carry out the IIW test on a model for a particular sample we re-estimate the model for each sample generated by the true model. As this is highly time-consuming, we did

**Table 2** Testing mis-specified models: percentage rejection rates using IIW

3 variable VAR(1)	T = 200	T = 125	T = 75
1) Generating data from model 2 as true data, testing model 1 by IIW	99.9	98.2	79.7
2) Generating data from model 3 as true data, testing model 2 by IIW	100	99.3	95.2
3) Generating data from model 3 as true data, testing model 1 by IIW	100	100	98.1



it selectively for two model pairs and different sample sizes: for a sample of 125, model 3 is the complex model and model 1 is the simpler model; with a sample of 75, model 2 is the complex model and model 1 is the simpler model.

Table 3 shows that the rejection rate for model 1 for the first pair is still 100%, even though there is some increase in closeness. Thus, even for a sample as small as 125, the rejection of mis-specification remains virtually 100%. In the second case, Table 4, where model 1 is tested using data from model 2 with a sample of only 75, it is somewhat harder to distinguish model 1 from model 2, the two closest models: the rejection rate falls to 68.6%. However, rejection is still overwhelmingly probable. As noted above, an LR test would have had zero power for these mis-specification tests; plainly, the IIW test provides considerable power against mis-specification.

To summarise, we have found that our IIW test can establish a) whether a model can predict relevant features of the data and, if so, b) the bounds within which we can be sure of its specification and parameter values. For the widely-used DSGE model examined here, we found that, if the model passes the test about the behaviour of three key macro variables, then its power largely guarantees that no other specification can be correct, and its parameter values lie within a 7–10% region of those estimated.

### 3.2 Parameter Inaccuracy within a Correct Specification

We now consider for a wide variety of VAR auxiliary models how the LR test performs, as compared with the IIW test for different levels of structural parameter inaccuracy. Our results, based on the simulations of the structural model being tested, are reported in Table 5.

Two things can be seen in the comparison of these two tests in the bottom two panels. First, in general, the power of the LR test does not reach 100% rejection until the model parameters reach 20% falsity regardless of the VAR order which reflects the detail with which data behaviour is described. The IIW test has high power even when the detail is kept rather low - e.g. with a VAR(1) in just three variables. Thus, we find that a structural model that is just 7% false is rejected with 98% frequency. Second, and by contrast, as the detail included in the VAR description rises with higher order, the IIW test acquires more power.

Turning to the top panel, where the II test is carried out using the distribution of VAR coefficients derived from the data sample VAR itself (i.e. unrestricted by the structural model being tested), two things emerge. First, as the size of the VAR increases, the information from the data sample is increasingly unable to provide well-defined

**Table 3** Transformed Wald for model 1 when tested on model 3 samples, T = 125

Transformed Wald	Min	Max	Mean	Rejection rate
				(Critical value = 1.645)
Re-estimated by II	1.686	79.314	21.739	100%
Original estimates	2.459	9.07E+15	2.57E+14	100%

**Table 4** Transformed Wald for model 1 when tested on model 2 samples, T = 75

Transformed Wald	Min	Max	Mean	Rejection rate
				(Critical value = 1.645)
Re-estimated by II	1.480	8.412	2.409	68.6%
Original estimates	1.291	11.818	2.909	79.7%

estimates of the distribution of the VAR coefficients. This is because the variation in too many coefficients has to be evaluated on too little data. Second, for lower order VARs, where evaluation is possible, the power is similar to that of the LR test. This is because, as noted in Appendix 2, these two tests are transforms of each other and therefore tend to produce similar test results.

This section has summarised the findings of our earlier survey paper, from which we concluded that, whatever VAR was used as the benchmark comparator model, LR tests would provide little or no power against mis-specification, and rather low power against numerical inaccuracy. It has also added a further result, namely, that more subtle aspects of mis-specification can be detected with high precision by indirect inference tests. Next we focus just on the IIW test, considering how best to use it and in what particular form.

**Table 5** Comparison of rejection rates at 95% level for Indirect Inference and Direct Inference

VAR — no of coefficients	TRUE	1%	3%	5%	7%	10%	15%	20%
<b>IIW TEST with unrestricted VAR</b>								
variable VAR(1) — 4	5.0	6.2	20.3	69.6	61.0	99.8	100.0	100.0
variable VAR(1) — 9	5.0	3.4	7.5	30.7	75.0	97.4	100.0	100.0
variable VAR(2) — 18	5.0	3.8	5.2	19.1	57.5	84.3	98.4	99.5
variable VAR(3) — 27	5.0	3.9	6.4	21.6	54.5	84.0	97.5	98.7
variable VAR(1) — 25	5.0	2.8	3.2	2.6	5.4	6.2	4.5	100.0
variable VAR(3) — 147	5.0	5.1	3.4	1.4	0.9	0.2	0.0	100.0
<b>IIW TEST with restricted VAR</b>								
variable VAR(1) — 4	5.0	9.8	37.7	80.8	96.8	100.0	100.0	100.0
variable VAR(1) — 9	5.0	9.5	36.1	71.0	98.1	100.0	100.0	100.0
variable VAR(2) — 18	5.0	8.3	35.5	80.9	96.9	100.0	100.0	100.0
variable VAR(3) — 27	5.0	9.2	32.9	78.0	95.1	100.0	100.0	100.0
variable VAR(1) — 25	5.0	17.8	85.5	99.8	100.0	100.0	100.0	100.0
variable VAR(3) — 147	5.0	77.6	99.2	100.0	100.0	100.0	100.0	100.0
<b>LIKELIHOOD RATIO TEST</b>								
variable VAR(1) — 4	5.0	12.0	28.3	45.9	63.4	83.2	97.0	99.7
variable VAR(1) — 9	5.0	9.4	21.8	37.5	58.9	84.0	99.0	100.0
variable VAR(2) — 18	5.0	8.9	20.7	36.8	57.6	82.9	98.7	100.0
variable VAR(3) — 27	5.0	8.9	20.4	36.7	56.7	82.2	98.7	100.0
variable VAR(3) — 147	5.0	5.7	10.6	23.6	46.3	83.2	99.6	100.0

## 4 Comparing Different Ways of Carrying out the IIW Test- Monte Carlo Experiments with Small Samples

### 4.1 Different Variables and Different Features

There are a variety of practical concerns about best to carry out indirect inference tests. One is whether it matters which features of the data are chosen as the ‘descriptors’ for the Wald test. As noted above above, a VAR is a natural choice of auxiliary model and the test can be based on its moments, its coefficients or its associated impulse response functions; the test results are largely the same. The key element in the power of the test is *how many* of these descriptors are used; this relates to the point made above that using more is equivalent to requiring the model to replicate more detailed features of the data, as with pixels in a photograph. Thus we showed above how increasing the number of VAR coefficients raised the power.

Another concern is the choice of variables used for the data description. For example, in evaluating the Smets-Wouters model we showed test results using the three main macro variables, output, inflation and interest rates. Would it be any different had we used three other variables, say consumption, investment and real wages? The answer is no, or there is a negligible difference, (Meenagh et al. 2018) as any three variables give much the same results. The reason is related to the last issue: provided the same amount of information about the data features is included, the test works similarly. One can think of each piece of information as being a nonlinear combination of the model’s structural coefficients; the question is whether this is matched by their effect on the data. The number of pieces of information gives the number of matches required by the test. What matters is the amount: more VAR coefficients for example, but not which VAR coefficients (Tables 6 and 7).

This may appear to be puzzling; the original idea in DSGE modelling of comparing the moments of data simulated from the model with those from the observed data stressed that it was good to *select* data features *of interest* to the user. We have found in our Monte Carlo work on the small sample properties of the IIW test that, provided the *number* of descriptive features is held constant, whichever variables are focused on, and however their behaviour is measured, whether by moments, IRFs or VAR coefficients, the IIW tests provide roughly the same power and estimates with much the same bias. Why is this?

**Table 6** Power of the II test across data descriptors

Degree of falseness	Rejection Rate							
	0%	1%	3%	5%	7%	10%	15%	20%
Var coeffs	0.05	0.128	0.866	0.997	1.000	1.000	1.000	1.000
IRF	0.05	0.140	0.852	0.998	1.000	1.000	1.000	1.000
Moment	0.05	0.114	0.326	0.665	0.913	0.997	1.000	1.000

Three variables are used in VAR are ( $y$ ,  $\pi$ ,  $r$ ), as in Le et al. (2011). Source: Minford et al. (2016)

**Table 7** Power of the II test across variable choice

Degree of falseness	Rejection Rate							
	0%	1%	3%	5%	7%	10%	15%	20%
$(y, \pi, r)$	0.05	0.128	0.866	0.997	1.000	1.000	1.000	1.000
$(c, i, l)$	0.05	0.094	0.561	0.923	0.986	1.000	1.000	1.000
$(q, w, r)$	0.05	0.072	0.276	0.771	0.984	1.000	1.000	1.000

VAR coefficients used as data descriptors, as in Le et al. (2011). Source: Meenagh et al. (2018)

An analogy would be with taking a fingerprint or footprint or skin sample from a person; given that person's DNA, each would be a test of whether that person was the source. Here we take a sample of behaviour and ask whether it comes from this DSGE model? It would seem that provided our data sample is of reasonable size and samples enough features, it can act as a rather accurate test of whether the DSGE model generated it.

For policymakers this is a reassuring conclusion. If they find a model that passes ANY of these tests, whichever they choose, they can establish from the test's power with certainty how false their model could be, in different respects that concern them. Usually they will be concerned with 'general falseness' where any or all of their parameters could be mis-estimated: they can then say within what bounds the parameters would lie for their model. They could use this to calculate robustness tests for their policy proposals.

## 4.2 Forecasting Tests

In effect the within-sample LR test assesses how accurate a model is for 'nowcasting' as it is based on the size of within-sample predictive errors. The next Table shows the power of out-of-sample forecasting LR tests for three variables jointly 4 and 8 quarters ahead, together with the IIW test. It can be seen that the out-of-sample LR test has weak power compared to the IIW test; the rejection rises close to 100% only with 20% or more falsity. Non-stationarity does not appear to affect these results (Table 8).

**Table 8** Rejection rates for wald and likelihood ratio for 3 variable VAR(1)

% mis-specified	II Wald in-sample	Joint 3:4Q	:8Q
True	5.0	5.0	5.0
	19.8	6.0	4.9
	52.1	9.4	5.2
	87.3	15.3	6.0
	99.4	22.9	6.6
	100.0	36.2	9.8
	100.0	73.8	29.5
	100.0	99.8	90.7

Source: Minford et al. 2015, Table 2, p 341

### 4.3 Tests of Parts of Models

The focus in our tests to this point has been on a full DSGE model. As noted earlier, if the full model fails the test then it is necessary to find out which parts of the model are causing the rejection in order to re-specify the model. Alternatively, the investigator may not have the time and resources to construct or investigate a full model and would be satisfied with using a few equations of a partially specified model. Minford et al. (2019) have proposed a limited information indirect inference test for testing a subset of equations of the full DSGE model. Like the indirect inference test for a full DSGE model, this limited information test of part of the model is very powerful.

The test exploits a useful way to carry out the limited information estimation of a part of a simultaneous equation system. The idea is to replace the structural equations that are not to be tested by their reduced form solution while keeping the structural equations that are to be tested and then to carry out the indirect inference test for the resulting completed model. In practice the unrestricted auxiliary model is used to replace the structural equations not tested. As a result the equations in the auxiliary model based on the simulated data from the completed model will be the same as those for the observed data. The test of the subset of interest is therefore based solely on the two sets of estimates of the other auxiliary equations. The simulated data from the completed model can also be used for the analysis of a partially specified model.

To illustrate the use of this test we examine the specification of two sectors of the SW model: the wage-price equations and the consumption-investment equations. Using an auxiliary VAR(1) model the limited information and full model tests of the wage-price sector have the power properties reported in Table 9.

The corresponding tests of the consumption-investment equations have the power properties reported in Table 10.

For both sectors the power of the limited information and the full information tests are similar. However, the power of the tests of the wage-price sector are much higher than those of the consumption-investment equations. The first implication is that there is little or no loss in power from using the limited information test. A second implication is that the behaviour of the Smets-Wouters model is much more sensitive to the specification of the wage-price sector than the consumption-investment equations. This suggests that the reason the original full Smets-Wouters model is rejected is that the wage-price sector is mis-specified; any misspecification of the consumption-investment equations appears not to affect the full model.

**Table 9** Rejection Rates at 95% level: falseness is given by  $\pm$  alternation, test 3 variables ( $y_t$ ,  $p_t$ ,  $rt$ ) VAR(1)

Falseness is given by $\pm$ x% alternation								
	0%	1%	3%	5%	7%	10%	15%	20%
Using VAR	0.050	0.076	0.345	0.923	1.000	1.000	1.000	1.000
Using Full Model	0.050	0.074	0.385	0.973	1.000	1.000	1.000	1.000

**Table 10** Rejection Rates at 95% level: falseness is given by  $\pm$  alternation, test 3 variables ( $y_t, p_t, r_t$ ) VAR(1)

Falseness is given by $\pm$ x% alternation								
	0%	1%	3%	5%	7%	10%	15%	20%
Using VAR	0.050	0.051	0.086	0.094	0.123	0.174	0.311	0.288
Using Full Model	0.050	0.055	0.057	0.056	0.056	0.069	0.103	0.140

## 5 Estimation in Small Samples

So far in this review we have considered only the properties of tests of models, especially their power in small samples, for models that have already been estimated. In the case of DSGE macroeconomic models this is typically by Bayesian methods for which the model is not subject to a test except, possibly, by inspection of the significance of the posterior estimates via their posterior distribution. We have previously noted some problems in interpreting such a check. If the prior dominates the likelihood function and hence the posterior then such a check will not reveal how well the model fits the observed data. We have also noted the claim that classical likelihood-based methods tend to reject too many “good” models, hence the use of Bayesian estimation to limit the role of observed data. One reason why likelihood methods may produce estimates not consistent with theory is that they attempt to find a model of best fit. If the model is mis-specified then it will produce biased estimates in order to maximise the fit. This may be why “good” models are often rejected, meaning the estimated parameters are not compatible with economic theory. Nonetheless, we have found that models estimated by maximum likelihood and tested using the LR test have low power compared with indirect inference. Specifying the model with unrestricted error dynamics is often found to be a way of improving the in-sample fit of models estimated by maximum likelihood, but it may also have the effect of lowering the power of the LR test, especially in small samples.

Indirect estimation is an alternative to both Bayesian and maximum likelihood estimation. We have found that it provides an estimator with smaller biases than maximum likelihood, typically around 1% for the average absolute bias across the DSGE parameters. This property comes from the high power of the test in rejecting false parameter values: the IIW tends to rise increasingly rapidly as the parameter estimates diverge from the true values. Regardless of whether the estimator is based on VAR coefficients, IRFs or moments, the small sample bias is small: it is essentially the same for the first two, with a slightly larger bias for moments. To illustrate, when one compares the indirect estimates based on these three descriptors using different three-variable sets, we find that the bias is small and hardly differs. These results are reported in Tables 11 and 12.

- $\alpha$  income share of capital
- $h$  external habit formation
- $t_p$  degree of price indexation
- $t_w$  degree of wage indexation

**Table 11** Bias of II estimates by using different data descriptors

Data Descriptor		VAR coeffs			IRF			Moments		
		II estimates			II estimates			II estimates		
True Values		Mean	Std	Bias	Mean	Std	Bias	Mean	Std	Bias
$\alpha$	0.19	0.189	0.020	0.32%	0.191	0.021	0.37%	0.197	0.018	1.50%
$h$	0.71	0.692	0.065	2.51%	0.684	0.069	3.61%	0.669	0.049	2.77%
$\iota_p$	0.22	0.221	0.023	0.45%	0.223	0.026	1.32%	0.229	0.021	4.08%
$\iota_w$	0.59	0.586	0.060	0.69%	0.584	0.068	0.96%	0.570	0.058	3.40%
$\xi_p$	0.65	0.664	0.068	2.08%	0.659	0.074	1.35%	0.674	0.063	3.71%
$\xi_w$	0.73	0.725	0.075	0.71%	0.731	0.083	0.17%	0.702	0.073	3.82%
$\varphi$	5.48	5.531	0.557	0.93%	5.555	0.608	1.36%	5.712	0.524	4.24%
$\Phi$	1.61	1.607	0.166	0.16%	1.635	0.182	1.53%	1.576	0.140	2.14%
$\psi$	0.54	0.540	0.057	0.00%	0.543	0.062	0.57%	0.561	0.051	3.88%
$r_{\Delta y}$	0.22	0.222	0.023	1.05%	0.222	0.024	1.09%	0.210	0.020	1.38%
$\rho$	0.81	0.787	0.047	2.80%	0.770	0.060	4.95%	0.842	0.058	3.91%
$r_\pi$	2.03	2.007	0.188	1.12%	1.992	0.204	1.86%	1.965	0.188	3.20%
$r_y$	0.08	0.079	0.009	1.25%	0.081	0.009	1.38%	0.082	0.008	2.96%
$\sigma_c$	1.39	1.358	0.137	2.29%	1.377	0.155	0.95%	1.356	0.132	2.45%
$\sigma_l$	1.92	1.930	0.208	0.54%	1.944	0.211	1.24%	1.981	0.191	1.18%
Average			0.113	1.13%		0.124	1.51%		0.106	3.04%

The true parameter values are from Smets and Wouters (2007), Table 4. Three variables used in VAR are ( $y$ ,  $\pi$ ,  $r$ ), as in Le et al. (2011) Std denotes standard deviation. Source: Meenagh et al. (2018).

- $\xi_p$  degree of price stickiness
- $\xi_w$  degree of wage stickiness
- $\varphi$  elasticity of the capital adjustment cost function
- $\Phi$  1 + the share of fixed costs in production
- $\psi$  elasticity of the capital utilization adjustment cost
- $r_{\Delta y}$  Taylor rule coefficient
- $\rho$  Taylor rule coefficient (interest rate smoothing)
- $r_\pi$  Taylor rule coefficient
- $r_y$  Taylor rule coefficient
- $\sigma_c$  elasticity of intertemporal substitution for labor
- $\sigma_l$  elasticity of labor supply to real wage

This bias can be driven as low as one wishes by increasing the number of features in the auxiliary model, as the power rises with the number of features until the point is reached where the full reduced form VAR is used (or its equivalent in moments or IRFs). This is because the indirect estimator and the indirect inference test are based on the same objective function: deviations of the estimates of the parameters of the auxiliary model from simulated and observed data. The bias reduction comes at the cost of the indirect inference test having extremely high power against even the slightest parameter falsity. It follows that if one has a tractable model that is not exactly the true model the parameter values it will estimate a) may not pass this powerful test -



**Table 12** Bias of II estimates by using different variable combinations

Var Comb	$(y, pi, r)$			$(c, i, l)$			$(g, w, r)$			
	II estimates			II estimates			II estimates			
True Values	Mean	Std	Bias	Mean	Std	Bias	Mean	Std	Bias	
$\alpha$	0.19	0.189	0.020	0.32%	0.188	0.021	1.2%	0.193	0.020	1.74%
$h$	0.71	0.692	0.065	2.51%	0.711	0.073	0.2%	0.674	0.057	5.08%
$\iota_p$	0.22	0.221	0.023	0.45%	0.215	0.025	2.3%	0.223	0.024	1.40%
$\iota_w$	0.59	0.586	0.060	0.69%	0.571	0.068	3.2%	0.580	0.063	1.70%
$\xi_p$	0.65	0.664	0.068	2.08%	0.635	0.073	2.3%	0.666	0.069	2.53%
$\xi_w$	0.73	0.725	0.075	0.71%	0.727	0.082	0.4%	0.719	0.080	1.49%
$\varphi$	5.48	5.531	0.557	0.93%	5.385	0.613	1.7%	5.575	0.579	1.73%
$\Phi$	1.61	1.607	0.166	0.16%	1.595	0.141	0.9%	1.583	0.172	1.69%
$\psi$	0.54	0.540	0.057	0.00%	0.524	0.062	3.0%	0.553	0.057	2.36%
$r_{\Delta y}$	0.22	0.222	0.023	1.05%	0.218	0.025	1.1%	0.217	0.024	1.49%
$\rho$	0.81	0.787	0.047	2.80%	0.797	0.088	1.6%	0.816	0.065	0.74%
$r_\pi$	2.03	2.007	0.188	1.12%	2.015	0.234	0.7%	1.983	0.205	2.32%
$r_y$	0.08	0.079	0.009	1.25%	0.078	0.009	2.3%	0.080	0.009	0.10%
$\sigma_c$	1.39	1.358	0.137	2.29%	1.370	0.156	1.4%	1.345	0.141	3.26%
$\sigma_l$	1.92	1.930	0.208	0.54%	1.876	0.218	2.3%	1.957	0.204	1.93%
Average			0.113	1.13%		0.123	1.70%		0.118	1.97%

The true parameter values are from Smets and Wouters (2007), Table 4. VAR coefficients are used as data descriptors, as in Le et al. (2011). Source: Meenagh et al. (2018).

as they are simply the values that get closest to passing it, b) may not pass the weaker-powered test either - as they were not selected as the values that get closest to passing this weaker test. From the user's viewpoint the key aim is to find a tractable model with parameter values that pass the test; having found such a model it is then possible to assess the robustness of the whole-model results to potential falsity, as described earlier. If, however one cannot find a model that passes the test, there is no way to make this assessment. Hence our view is that it is best to use in indirect estimation the same criterion as for the indirect test statistic that was given above.

It is notable that Dridi et al. (2007) propose a two-step procedure to achieve both objectives: estimation and evaluation of mis-specified DSGE models. In the first step the model is estimated using a well chosen set of moments; in the second step, the model is evaluated with chosen features of the data that the model tries to replicate. They derive the asymptotic distribution of the test statistic under the hypothesis that the DSGE model is mis-specified and therefore use the variance-covariance matrix from the unrestricted VAR. Hall et al. (2012), and Guerron-Quintana et al. (2017) use the IRFs as the data descriptor in the auxiliary model and discuss both the small sample and large sample properties of the indirect approach to estimation, but not as a method of testing a model. As we have seen above, Minford et al. (2016) compare the test with different data descriptors and find that mostly the properties are quite similar. The two ways of conducting indirect estimation are, of course, closely related.

## 6 Identification

There has been a persistent suspicion that DSGE model parameters cannot be uniquely recovered from the data. In other words, the model is not identified. Canova and Sala (2009) and Le et al. (2012) have provided examples of the lack of identification of DSGE models in the New Keynesian model, see also Appendix 1. In response to such concerns various efforts have been made to establish whether various prominent DSGE macro models are identified. One avenue has been to use the rank condition which tests whether, with no limits on data availability, a DSGE model's parameters can be uniquely discovered - Iskrev (2010), Komunjer and Ng (2011), Qu and Tkachenko (2012). Another avenue is to use indirect inference in the way suggested by Le et al. (2017). These two methods establish 'precise' identification - i.e. whether, in the presence of unlimited data, another set of parameters can be found that generate exactly the same reduced form as the DSGE model in question.

A related issue is that of 'weak identification'. By this is meant whether individual model parameter values can be distinguished with any confidence from other values from some false, competing, version of the model. As the data sample becomes smaller this becomes of increasing concern; when data is unlimited this reduces to the same as precise identification. In practice data is limited and so one would like to examine identification in the context of much, but not unlimited, data. This can be addressed by indirect inference by asking how much power the test has against false parameter values in relevant data sample sizes. Le et al. (2017) find that Smets-Wouters parameters are generally strongly identified according to this criterion but, in a small three equation New Keynesian model similar to that of Clarida et al. (1999), several parameters suffer from weak identification.

This analysis used the model's full reduced form which gives maximum discrimination against false parameter values. In practice we use, as we have seen, a low order VAR in only a few variables (or the equivalent in moments or IRFs). Does using such a low-order approximation to the reduced form create a weakness of identification?

We know that the Smets Wouters model is identified in the sense that its full reduced form is unique to it - see Le et al. (2017). We may, however, be concerned that as we reduce the number of variables and the order of the VAR that is approximating the reduced form, the amount of information included drops sufficiently to give very imprecise estimates of the DSGE model parameters. When the sample is also small, this problem could become more acute. We then have the problem of weak identification.

This problem can be investigated through Monte Carlo experiments. We falsify individual parameters of the Smets Wouters model progressively and check the rejection rate over many samples from the true model. If the power is poor so that the rejection rate barely rises with such falsity then we can regard the parameters as weakly identified as, plainly, we cannot distinguish well between possible false and true parameter values. The results of this experiment are reported in Table 13.

This exercise shows that rejection rates vary across parameters: while the rejection rates of three of the parameters rise only slowly, rejection rates for the others rise quite fast. This mirrors the findings of Le et al. (2017) for the full VAR reduced form of the Smets and Wouters model and a large sample. It seems that reducing the VAR to be of

**Table 13** Rejection rates for individual parameters falsified

	Degree of Falseness							
	1%	2%	3%	5%	10%	20%	50%	75%
$\alpha$	0.0700	0.0720	0.0720	0.0710	0.0810	0.1230	0.6020	0.9480
$h$	0.0720	0.0710	0.0760	0.0900	0.1950	0.6720	1.0000	1.0000
$t_p$	0.0700	0.0700	0.0700	0.0720	0.0730	0.0790	0.1000	0.1350
$t_w$	0.0690	0.0690	0.0690	0.0760	0.0830	0.1820	0.8490	0.9870
$\xi_p$	0.0670	0.0620	0.0650	0.0670	0.1440	0.6360	0.9950	0.9960
$\xi_w$	0.0710	0.0740	0.0770	0.0860	0.0970	0.4970	1.0000	1.0000
$\varphi$	0.0700	0.0700	0.0700	0.0710	0.0710	0.0700	0.0740	0.1410
$\Phi$	0.0700	0.0690	0.0720	0.0770	0.0930	0.2420	0.5230	0.8340
$\psi$	0.0700	0.0700	0.0710	0.0740	0.0720	0.0780	0.3270	0.9930
$r_{\Delta y}$	0.0690	0.0710	0.0690	0.0750	0.1260	0.5070	1.0000	1.0000
$\rho$	0.0870	0.1570	0.3590	0.8910	1.0000	1.0000	1.0000	1.0000
$r_\pi$	0.0750	0.0850	0.1160	0.2160	0.8710	1.0000	1.0000	1.0000
$r_y$	0.0710	0.0690	0.0680	0.0770	0.1380	0.5370	1.0000	1.0000
$\sigma_c$	0.0710	0.0720	0.0730	0.0750	0.1050	0.3140	1.0000	1.0000
$\sigma_I$	0.0700	0.0720	0.0730	0.0740	0.0770	0.0800	0.1060	0.1390

order 1 with only three variables and using a much smaller sample, does not create weak identification.

## 7 Conclusions

Indirect Inference is a method for testing and estimating models of any size, complexity or nonlinearity by comparing the behaviour of data simulated from the model with that of selected observed data where behaviour is summarised by a set of descriptive features, known as the auxiliary model. Its use in the testing and estimation of macro-economic and other structural models has increased in the past few years. Nonetheless, as it is a relatively novel procedure, many questions remain about its detailed application. As these questions go beyond the scope of the survey in Le et al. (2016a), we update that survey by providing answers to some of these additional questions.

The earlier survey discussed the power of indirect inference tests in small samples in comparison with tests based on the main frequentist testing alternative using the data likelihood. The survey suggested ways in which modellers could use this power to determine the robustness of their policy or other user results. In that survey it was assumed that the auxiliary model would be a VAR of low order in a few variables while the structural model would be a DSGE model of some sort.

In this update we have discussed the view that DSGE models are so unrealistic that it is pointless to test them. We have argued that the main purpose of testing is to establish whether, even if they are mis-specified, they are robust enough to be of practical use and they should be regarded as pseudo-true for the purposes of hypothesis testing. As

DSGE models are commonly estimated using Bayesian methods in which the priors often dominate the data likelihood, but not as a result tested, a check on the robustness of the estimates seems to us to be advisable.

Indirect inference requires the use of an auxiliary model. We have reported results from our previous work and from new results obtained for this paper which show how the choice of auxiliary model affects the power of the test. We have found that using a VAR of low order in only a few variables may be adequate for providing a powerful test of a large model such as that of Smets and Wouters (2007) even though its true reduced form is a much higher order VAR in more variables. Moreover, the selection of which variables to use in the auxiliary VAR model is found to make little difference. It also makes little difference to the test results whether one uses a VAR as the auxiliary model or moments (as in the Simulated Method of Moments) or impulse response functions as in some recent applied work.

We also showed how one can test parts of models using indirect inference. There is little or no loss of power compared to using the full model. We have shown how indirect inference tests can be used to check both the precise and weak identification of DSGE models, although their identification does not seem to be much of a problem in practice. Likelihood ratio tests often have low power within sample, we have shown that they also have low power in forecasting tests out of sample.

In summary, we continue to think, as suggested in the original survey of Le et al. (2016a), that indirect inference provides a powerful test of structural macro models, which enables policymakers and other users to assess rather accurately how robust their models are to possible errors of specification and estimation.

## Appendix 1: The auxiliary model: a VAR representation of a DSGE model

There are several ways of deriving a VAR representation of a DSGE model. We make use of the ABCD framework of Fernandez-Villaverde et al. (2007). We consider solely what these authors call the ‘square’ case, where the number of errors and the number of observable variables are the same. We also consider only DSGE models with no observable exogenous variables. Both the Smets-Wouters model (Smets and Wouters 2003, 2007) and the 3-equation model New Keynesian model used by Le et al. (2017) and Liu and Minford (2014) for their numerous IIW tests fit this framework. (Other classes of models, for example those with ‘news shocks’, require a different treatment which is beyond our scope here.)

To illustrate, consider the 3-equation New Keynesian model of Le et al. (2017):<sup>1</sup>

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

<sup>1</sup> Further lags in both endogenous variables and the errors could be added; but for our main treatment we suppress these. Our results can be extended to deal with them, without essential change.

This has the solution

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = KH \begin{bmatrix} e_{\pi t} \\ e_{yt} \\ e_{rt} \end{bmatrix} \quad (2)$$

where.

$$K = \begin{bmatrix} 1 + \frac{\eta}{\sigma} - \rho_\pi & \lambda & -\frac{\lambda}{\sigma} \\ -\frac{1}{\sigma}(\gamma - \rho_\pi) & 1 - \omega\rho_y & -\frac{1}{\sigma}(1 - \omega\rho_r) \\ \gamma - \left(\gamma - \frac{\eta}{\sigma}\right)\rho_\pi & \lambda\gamma + \eta - \eta\omega\rho_y & 1 - \left(1 + \omega + \frac{\lambda}{\sigma}\right)\rho_r + \omega\rho_r^2 \end{bmatrix},$$

$$H = \begin{bmatrix} H_{11} & 0 & 0 \\ 0 & H_{22} & 0 \\ 0 & 0 & H_{33} \end{bmatrix},$$

$$H_{11} = \frac{1}{1 + \frac{\eta + \lambda\gamma}{\sigma} - \left[\frac{\lambda}{\sigma} + \omega\left(1 + \frac{\eta}{\sigma}\right)\right]\rho_\pi + \omega\rho_\pi^2}$$

$$H_{22} = \frac{1}{1 + \frac{\eta + \lambda\gamma}{\sigma} - \left[\frac{\lambda}{\sigma} + \omega\left(1 + \frac{\eta}{\sigma}\right)\right]\rho_y + \omega\rho_y^2}$$

$$H_{33} = \frac{1}{1 + \frac{\eta + \lambda\gamma}{\sigma} - \left[\frac{\lambda}{\sigma} + \omega\left(1 + \frac{\eta}{\sigma}\right)\right]\rho_r + \omega\rho_r^2}.$$

or

$$z_t = \Phi e_t \quad (3)$$

$$e_t = Pe_{t-1} + \varepsilon_t \quad (4)$$

where  $z_t' = [\pi_t, y_t, r_t]$ ,  $e_t' = [e_{\pi t}, e_{yt}, e_{rt}]$ ,  $\Phi = K \times H$ . Thus the matrix  $\Phi$  is restricted, having 9 elements but consists of only 5 structural coefficients (the  $\rho_i$  can be recovered directly from the error processes), implying that the model is over-identified according to the order condition. The model is not identified, however, if the  $\rho_i = 0$  for all  $i$ .<sup>2</sup>

The solved structural model can be written in ABCD form as follows where  $y$  (replacing  $z$  above) is now the vector of endogenous variables and  $x$  (replacing  $e$  above) is the vector of error processes:

$$x_t = Ax_{t-1} + B\varepsilon_t$$

<sup>2</sup> Le et al. (2017), also establish that it is identified using the IIV test in unlimited-size sampling.

$$y_t = Cx_{t-1} + D\varepsilon_t$$

where  $A = P = \begin{bmatrix} \rho_\pi & 0 & 0 \\ 0 & \rho_y & 0 \\ 0 & 0 & \rho_r \end{bmatrix}; B = I; C = \Phi P; D = \Phi$ .

Note that  $y_t = \Phi x_t$  is the (solved) structural model. Hence  $x_t = \Phi^{-1}y_t$ . The VAR representation is<sup>3</sup>

$$y_t = \Phi P \Phi^{-1} y_{t-1} + \Phi \varepsilon_t = V y_{t-1} + \xi_t \tag{5}$$

We may also note that

$$y_t = \Phi \sum_{i=0}^{\infty} P^i \varepsilon_{t-1} = \sum_{i=0}^{\infty} P^i \xi_{t-i}$$

More generally, the solution of a linearised DSGE model (including the SW model and the 3-equation model) can be summarised by a state-space representation:<sup>4</sup>

$$x_t = Ax_{t-1} + B\varepsilon_t y_t = Cx_t$$

where  $x_t$  is an  $n \times 1$  vector of possible unobserved state variables,  $y_t$  is a  $k \times 1$  vector of variables observed by an econometrician, and  $\varepsilon_t$  is an  $m \times 1$  vector of economic shocks affecting both the state and the observable variables, i.e., shocks to preferences, technologies, agents' information sets, and economist's measurements. The shocks  $\varepsilon_t$  are Gaussian vector white noise satisfying  $E(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon_t') = I$ . The matrices  $A, B$  and  $C$  are functions of the underlying structural parameters of the DSGE model. Using the ABCD framework of Fernandez-Villaverde et al. (2007), the state-space representation can be written as the VAR

$$y_t = V y_{t-1} + \eta_t \tag{6}$$

where  $E(\eta_t \eta_t') = \Phi \Phi' = \Sigma$ .

We have assumed that the DSGE model includes no observable exogenous variables. If it does then the solution to the DSGE model contains exogenous variables as well as lagged endogenous variables: in general, lagged, current and expected future

<sup>3</sup> If the DSGE model also had one-period lags in one or more of the equations so that the solution became  $z_t = \Phi e_t + \Lambda z_{t-1}$  then we would obtain a VAR(2) as follows:

- (1)  $x_t = Ax_{t-1} + B\varepsilon_t$
- (2)  $y_t = Cx_{t-1} + D\varepsilon_t + \Lambda y_{t-1}$

Using  $x_{t-1} = \Phi^{-1}(y_{t-1} - \Lambda y_{t-2})$  we obtain

$$y_t = (\Phi P \Phi^{-1} + \Lambda) y_{t-1} - \Phi^{-1} \Lambda y_{t-2} + \Phi \varepsilon_t$$

<sup>4</sup> The solution of the model can be obtained by using either Blanchard and Kahn (1980) or Sims (2002) type of algorithms.

exogenous variables. If, however, the exogenous variables are assumed to be generated by a VAR process then the combined solution of both the endogenous and exogenous variables is a purely backward-looking model that can be represented as a VAR.<sup>5</sup>

## Appendix 2: The LR and the IIW test statistics- asymptotic comparisons

In indirect inference we do not impose the restrictions on the coefficients of the auxiliary model that are implied by the structural model. Instead, we estimate the auxiliary model on data simulated from the structural model and compare these estimates with those obtained from using the observed data. In both cases the auxiliary model is estimated without any coefficient restrictions. The restrictions imposed by the DSGE model are reflected in the simulated data and not through explicit restrictions on the auxiliary model.

Since both the LR test and the IIW test involve estimation of an unrestricted VAR, first we briefly review the maximum likelihood estimation (MLE) of a standard unrestricted VAR. Consider a randomly generated sample of  $y_t$  of size  $T$ . If  $\eta_t$  is assumed to be  $NID(0, \Sigma)$  then the log-likelihood function is

$$\ln L(V, \Sigma) = - \left[ \frac{Tn}{2} \ln(2\pi) + \frac{T}{2} \ln|\Sigma| + \frac{1}{2} \sum_{t=1}^T (y_t - Vy_{t-1})' \Sigma^{-1} (y_t - Vy_{t-1}) \right]$$

Maximising with respect to  $\Sigma^{-1}$  gives

$$\frac{\partial \ln L(V, \Sigma)}{\partial \Sigma^{-1}} = \frac{T}{2} \Sigma^{-1} \sum_{t=1}^T (y_t - Vy_{t-1})(y_t - Vy_{t-1})'$$

Setting this to zero and solving gives the MLE estimator of  $\Sigma$  as

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (y_t - Vy_{t-1})(y_t - Vy_{t-1})' \quad (7)$$

Substituting this back into the likelihood function gives the concentrated likelihood

$$\ln L(V, \hat{\Sigma}^{94}) = - \left[ \frac{Tn}{2} \ln(2\pi) + \frac{T}{2} \ln|\hat{\Sigma}^{94}| + \frac{Tn}{2} \right]$$

Maximising this with respect to  $V$  is identical to minimising  $\ln|\hat{\Sigma}^{94}|$  with respect to  $V$ . Thus

$$\frac{\partial \ln|\hat{\Sigma}^{94}|}{\partial V} = 2 \hat{\Sigma}^{-1} \sum_{t=1}^T (y_t - Vy_{t-1}) y_{t-1}' = 0$$

<sup>5</sup> For further discussion on the use of a VAR to represent a DSGE model, see for example Canova (2005), Dave and De Jong (2007), Del Negro and Schorfheide (2004, 2006) and Del Negro et al. (2007a, 2007b) (together with the comments by Christiano (2007), Gallant (2007), Sims (2007), Faust (2007) and Kilian (2007)), and Wickens (2014).



and hence the MLE of  $V$  is

$$\hat{V} = \left( \begin{matrix} T \\ t=1 \end{matrix} y_t y_t' \right) \left( \begin{matrix} T \\ t=1 \end{matrix} y_{t-1} y_{t-1}' \right)^{-1}$$

and can be calculated by applying OLS to each equation separately. The MLE of  $\Sigma$  becomes

$$\Sigma^{94} = \frac{1}{T} \sum_{t=1}^T \left( y_t - \hat{V} y_{t-1} \right) \left( y_t - \hat{V} y_{t-1} \right)' \tag{8}$$

In order to find the variance matrix of  $\hat{V}$  it is convenient to re-express the VAR. Denoting the  $T$  observations on the  $i^{th}$  element of  $y_t$  as the  $T \times 1$  vector  $y_i$  and of  $\eta_t$  as  $\eta_i$ , each equation of the VAR may be written as

$$y_i = Z v_i + \eta_i \tag{9}$$

where  $v_i'$  is the  $i^{th}$  row of  $V$  and  $Z$  is a  $T \times k$  matrix with  $t^{th}$  row  $y_{t-1}$ . The VAR may now be written in matrix form as

$$Y = X v + \eta \tag{10}$$

where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, X = \begin{bmatrix} Z & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & Z \end{bmatrix} = I_k \otimes Z, \dots \eta = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_T \end{bmatrix}, \dots v = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}$$

$\otimes$  denotes a Kronecker product. Hence  $\eta$  is  $N(0, \Omega)$  where  $\Omega = \Sigma \otimes I_T$ . Generalised least squares estimation gives the MLE of  $v$  as

$$v^{94} = \left( X' \Omega^{-1} X \right)^{-1} X' \Omega^{-1} Y = \left[ I_k \otimes \left( Z' Z \right)^{-1} Z' \right] Y = v + \left[ I_k \otimes \left( Z' Z \right)^{-1} Z' \right] \eta$$

In general  $v^{94}$  is a biased estimate of  $v$  as  $Z$  consists of lagged endogenous variables, but  $plim v^{94} = v$  and the limiting distribution of  $\sqrt{T}(v^{94} - v)$  is  $N(0, W)$  where

$$W = plim T \left[ I_k \otimes \left( Z' Z \right)^{-1} Z' \right] \left( \Sigma \otimes I_T \right) \left[ I_k \otimes \left( Z' Z \right)^{-1} Z' \right]' = \Sigma \otimes \left( plim T^{-1} Z' Z \right)^{-1}$$

### The LR test

The LR test for a DSGE model based on the observed data compares the likelihood function of the auxiliary VAR derived from the DSGE model with the likelihood function of the unrestricted VAR computed on the observed data. The former is based on the estimate of the variance matrix of the structural errors from the solution to the DSGE model. On the assumption that the auxiliary model is the solution to the DSGE

model and is a VAR, this is also the error variance matrix of a restricted version of the auxiliary VAR. The latter is based on the estimate of the error variance matrix of the unrestricted auxiliary VAR. As the auxiliary model is a VAR, the LR test is, in effect, based on the one-period ahead forecast error matrix. Thus, the logarithm of the likelihood ratio test is

$$LR = 2(\ln L_U - \ln L_R) = T(\ln|\Sigma_R| - \ln|\hat{\Sigma}|) \quad (11)$$

where  $L_R$  and  $L_U$  denote the likelihood values of the restricted and unrestricted VAR, respectively, and  $\Sigma_R$  and  $\hat{\Sigma}$  are the restricted and unrestricted error variance matrices. Note that, given estimates of the DSGE model, we can solve the model for  $v$ , and hence we can calculate  $\eta_t$  and  $\Sigma_R = T^{-1} \sum_{t=1}^T \eta_t \eta_t'$ . Note also that the LR test can be routinely transformed into a (direct inference) Wald test between the unrestricted and the restricted VAR coefficients,  $v$ .

To obtain the power function of the LR test we endow the structural model with false values of the structural coefficients and compare the restricted VAR with the unrestricted VAR on the observed data which are assumed to be generated by the true model. The implied false model has the VAR

$$y_t = V_F y_{t-1} + \eta_{Ft} \quad (12)$$

The forecast errors for the false model are

$$\eta_{Ft} = y_t - V_F y_{t-1} = \eta_t + (V - V_F) y_{t-1} = \eta_t + q_t$$

where  $q_t = Y_{t-1}(v - v_F)$ . If we let

$$\Sigma_F = \frac{1}{T} \sum_{t=1}^T \eta_{Ft} \eta_{Ft}' = \frac{1}{T} \sum_{t=1}^T (\eta_t + q_t)(\eta_t + q_t)'$$

then the LR test for the false model is given by:

$$LR_F = T[\ln|\Sigma_F| - \ln|\hat{\Sigma}|] \quad (13)$$

Thus the power of the test derives from the distance

$$\ln|\Sigma_F| - \ln|\Sigma_R|. \quad (14)$$

### The IIW test

In the IIW test we simulate data from the solution to the already estimated DSGE, randomly drawing the samples from the DSGE model's structural errors. We then estimate the auxiliary VAR using these simulated data. We repeat this many times to obtain the average estimate of the coefficients of the VAR which we take as the estimate of the unrestricted VAR. The simulated VAR may be written

$$y_{S,t} = V_S y_{S,t-1} + \eta_{St}$$

where  $y_{S,t}$  is the data simulated from the DSGE model and  $V_S$  is the (average estimate of  $\nu$ ) or, in the form of eqs. (9) and (10), as

$$y_{S,i} = Z_S v_{s,i} + \eta_{S,i} Y_S = X_S v_S + \eta_S$$

where  $E(\eta_{S,i} \eta'_{S,i}) = \Sigma_S$ . The IIW test statistic, which computes the distance of these estimates from the unrestricted estimates based on the observed data, is:

$$IIW = [\hat{v} - v_S]' W_S^{-1} [\hat{v} - v_S] \tag{15}$$

where  $W_S$  is the covariance matrix of the limiting distribution of  $v_S$ , and is given by

$$W_S = \Sigma_S \otimes (\text{plim} T^{-1} Z'_S Z_S)^{-1} \tag{16}$$

On the null hypothesis that the DSGE model — and hence the auxiliary VAR — are correct, the asymptotic distribution of the estimate of  $v_S$  is the same that of the MLE  $\hat{v}$ . Moreover, asymptotically, this IIW statistic will have the same distribution as  $[\hat{v} - v]' W^{-1} [\hat{v} - v]$  and hence will have the same critical values.<sup>6</sup> In general, the IIW statistic differs from a standard Wald statistic in indirect inference which is  $[\hat{v} - v_S]' W^{-1} [\hat{v} - v_S]$  where  $W$  is the covariance matrix of the unrestricted model; we refer to this as the unrestricted IIW statistic.

The power of the IIW test is calculated, like that for the power calculations for the LR test, by simulating the DSGE model using false values of its coefficients and now using these data to estimate the unrestricted VAR from eq. (12). The IIW statistic is then computed from

$$IIW = [\hat{v} - v_F]' W_F^{-1} [\hat{v} - v_F] \tag{17}$$

where  $v_F$  is the mean vector of coefficients and  $W_F$  is their variance matrix, which corresponds to  $W_S$ . Consider the decomposition

$$\hat{v} - v_F = (\hat{v} - v) + (v - v_F).$$

It follows that the IIW statistic can be decomposed as

$$[\hat{v} - v_F]' W_F^{-1} [\hat{v} - v_F] = \eta' \left[ I_k \otimes (Z'Z)^{-1} Z' \right]' W_F^{-1} \left[ I_k \otimes (Z'Z)^{-1} Z' \right] \eta + [v - v_F]' W_F^{-1} [v - v_F] \tag{18}$$

<sup>6</sup> The IIW test can also be carried out for a sub-set of  $\nu$ .

$$= \eta' \left[ \Sigma^{-1} \otimes \text{plim} T \left( Z'Z \right)^{-1} \right] \eta + [v-v_F]' W_F^{-1} [v-v_F] \tag{19}$$

where the last term is based on the difference between the true and the false values of the coefficients. Hence the power of the IIW test derives from the second term on the right-hand side of eq. (19).

**Comparing the power of the two tests**

We have seen that the LR test compares the one-step ahead forecast error matrix of the unrestricted VAR with that of the model-restricted VAR using the observed data, whereas the IIW test asks whether the distribution of the VAR coefficients based on the simulated data (the restricted model) covers the VAR coefficients based on the observed data (the unrestricted model). We have also found that on the null hypothesis that the DSGE model is true the limiting distributions of the two sets of estimates are the same. It follows from eq. (7) that, on the null hypothesis, the error variance matrix using simulated data is

$$\begin{aligned} \Sigma_S &= \frac{1}{T} \sum_{t=1}^T (y_{St} - V_S y_{S,t-1})(y_{St} - V_S y_{S,t-1})' = \frac{1}{T} \sum_{t=1}^T (y_t - V_S y_{t-1})(y_t - V_S y_{t-1})' + \Delta \\ &= \hat{\Sigma} + (\hat{V} - V_S) \frac{1}{T} \sum_{t=1}^T y_{t-1} y_{t-1}' (\hat{V} - V_S)' + \Delta \end{aligned}$$

where  $\hat{\Sigma}$  is the error variance matrix of the unrestricted VAR using the observed data and  $\Delta$  is  $O_p(T^{-\frac{1}{2}})$ .

Using the result that  $\text{vec}(AXB) = (B' \otimes A)\text{vec}(X)$ , and  $\text{vec}(V) = v$ , it can be shown that

$$\text{vec} \left[ (\hat{V} - V_S) \frac{1}{T} \sum_{t=1}^T y_{t-1} y_{t-1}' (\hat{V} - V_S)' \right] = v' \left( I \otimes \frac{1}{T} \sum_{t=1}^T y_{t-1} y_{t-1}' \right) v$$

Hence,

$$\begin{aligned} LR &= T(\ln|\Sigma_S| - \ln|\hat{\Sigma}|) = T \left[ \ln \left| 1 + \frac{(\hat{V} - V_S) \frac{1}{T} \sum_{t=1}^T y_{t-1} y_{t-1}' (\hat{V} - V_S)' + \Delta}{|\hat{\Sigma}|} \right| \right] \\ &= T \left[ \ln \left| 1 + (\hat{v} - v_S)' \left( \hat{\Sigma} \otimes \frac{1}{T} \sum_{t=1}^T y_{t-1} y_{t-1}' \right)^{-1} (\hat{v} - v_S) + \frac{\Delta}{|\hat{\Sigma}|} \right| \right] \rightarrow IIW + O_p(T^{-\frac{1}{2}}) \end{aligned}$$

In other words, on the null hypothesis that the DSGE model is the true model, the LR test based on observed data is asymptotically equivalent to using the IIW test, which is based on simulated data.

In the power calculations we use

$$LR = T(\ln|\Sigma_F| - \ln|\hat{\Sigma}|) = T(\ln|\Sigma_S| - \ln|\hat{\Sigma}|) + T(\ln|\Sigma_F| - \ln|\Sigma_S|)$$

The power of the test derives from the last term which reflects the difference between  $V_S$  and  $V_F$ . This makes  $\Delta$  of order  $O_p(1)$ , which does not vanish as  $T \rightarrow \infty$ , but causes the power of the test to tend to unity.

It is worth relating this finding to the work of Dridi et al. (2007) who propose a Wald II test that treats the model being investigated as mis-specified; they therefore use the variance-covariance matrix from the unrestricted VAR which is generated by the unknown true model. This II test is asymptotically equivalent to the LR test, as we have seen, and differs from the IIW test proposed here which is based on the restricted VAR generated by the DSGE model being investigated- in effect this IIW test treats the DSGE model as pseudo-true and therefore as the null. In what follows we systematically compare the small sample properties of these different tests; as we will see the IIW test has the greatest power. It is really irrelevant whether the DSGE model being tested is regarded as ‘mis-specified’ or not, since as we have already argued, this not the issue: the issue is whether such a model is sufficiently close to the data on the test chosen to be regarded as ‘pseudo-true’. For establishing this it is helpful to have a test with as much potential power as possible; as we will show below, that potential power can then be tailored flexibly to the user’s problem.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Bernanke BS, Gertler M, Gilchrist S (1999) The financial accelerator in a business cycle framework, handbook of macroeconomics, vol 1, edited by J.B. Taylor and M. Woodford, ch. 21, 1341-1393. Elsevier
- Blanchard OJ, Kahn CM (1980) The solution of linear difference models under rational expectations. *Econometrica* 48(5):1305–1311
- Canova F (2005) *Methods for applied macroeconomic research*. Princeton University Press, Princeton
- Canova F, Sala L (2009) Back to square one: identification issues in DSGE models. *J Monet Econ* 56:431–449
- Christiano LJ (2007) Comment on ‘on the fit of new Keynesian models. *J Bus Econ Stat* 25:143–151
- Clarida R, Gali J, Gertler M (1999) The science of monetary policy: a new Keynesian perspective. *J Econ Lit* 37(4):1661–1707
- Cox, D. 1961. Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*
- Cox, D. 1962. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, B24*, 406–424
- Dave C, De Jong DN (2007) *Structural Macroeconomics*. Princeton University Press
- Del Negro M, Schorfheide F (2004) Priors from general equilibrium models for VARs. *Int Econ Rev* 45:643–673
- Del Negro M, Schorfheide F (2006) How good is what you’ve got? DSGE-VAR as a toolkit for evaluating DSGE models. *Economic Review*, Federal Reserve Bank of Atlanta, issue Q2:21–37
- Del Negro M, Schorfheide F, Smets F, Wouters R (2007a) On the fit of new Keynesian models. *J Bus Econ Stat* 25:123–143
- Del Negro M, Schorfheide F, Smets F, Wouters R (2007b) Rejoinder to comments on ‘on the fit of new Keynesian models. *J Bus Econ Stat* 25:159–162

- Dridi R, Guay A, Renault E (2007) Indirect inference and calibration of dynamic stochastic general equilibrium models. *J Econ* 136:397–430
- Evans GW, Honkapohja S (2005) Interview with Thomas J. Sargent *Macroeconomic Dynamics* 9(4):561–583
- Faust J (2007) Comment on ‘on the fit of new Keynesian models. *J Bus Econ Stat* 25:154–156
- Fernandez-Villaverde J, Rubio-Ramirez F, Sargent T, Watson M (2007) ABCs (and ds) of understanding VARs. *Am Econ Rev* 97:1021–1026
- Friedman M (1953) The methodology of positive economics. In: *Essays in positive economics*. University of Chicago Press, Chicago, pp 3–43
- Gallant AR (2007) Comment on ‘on the fit of new Keynesian models. *J Bus Econ Stat* 25:151–152
- Guerron-Quintana P, Inoue A, Kilian L (2017) Impulse response matching estimators for DSGE models. *J Econ* 196(1):144–155
- Hall AR, Inoue A, Nason JM, Rossi B (2012) Information criteria for impulse response function matching estimation of DSGE models. *J Econ* 170:499–518
- Iskrev N (2010) Local Identification in DSGE Models. *J Monet Econ* 57(2):189–202
- Kilian L (2007) Comment on ‘on the fit of new Keynesian models. *J Bus Econ Stat* 25:156–159
- King RG, Plosser CI, Rebelo ST (1988) Production, growth and business cycles : I. The basic neoclassical model *Journal of Monetary Economics* 21(2–3):195–232
- Komunjer I, Ng S (2011) Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica* 79(6):1995–2032
- Le VPM, Meenagh D, Minford P, Wickens M (2011) How much nominal rigidity is there in the US economy — testing a new Keynesian model using indirect inference. *J Econ Dyn Control* 35(12):2078–2104
- Le, V.P.M. Minford, P. and Wickens, (2012). A Monte Carlo procedure for checking identification in DSGE models. Cardiff economics working papers
- Le VPM, Meenagh D, Minford P, Wickens M, Xu Y (2016a) Testing macro models by indirect inference: a survey for users. *Open Econ Rev* 27:1–38
- Le VPM, Meenagh D, Minford P (2016b) Monetarism rides again? US monetary policy in a world of quantitative easing. *J Int Financ Mark Inst Money* 44:85–102
- Le VPM, Meenagh D, Minford P, Wickens (2017) A Monte Carlo procedure for checking identification in DSGE models. *J Econ Dyn Control* 76(C):202–210
- Liu C, Minford P (2014) How important is the credit channel? An empirical study of the US banking crisis. *J Bank Financ* 41:119–134
- Meenagh, D., P. Minford, M. Wickens and Y. Xu (2018) Comparing different data descriptors in Indirect Inference tests on DSGE models, Working paper E2018/7, Cardiff Economics working papers, [http://carbsecon.com/wp/E2018\\_7.pdf](http://carbsecon.com/wp/E2018_7.pdf)
- Minford P, Xu Y, Zhou P (2015) How good are out of sample forecasting tests on DSGE models? *Italian Economic Journal* 1:333–351
- Minford P, Wickens M, Xu Y (2016) Comparing different data descriptors in indirect inference tests on DSGE models. *Econ Lett* 145:157–161
- Minford P, Wickens M, Xu Y (2019) Testing part of a DSGE model by indirect inference. *Oxf Bull Econ Stat* 81(1):178–194
- Qu Z, Tkachenko D (2012) Identification and frequency domain quasi-maximum likelihood estimation of linearised dynamic stochastic general equilibrium models. *Quant Econ* 3(1):95–132
- Sims C (2002) Solving linear rational expectations models. *Comput Econ* 20(–2):1–20
- Sims C (2007) Comment on ‘on the fit of new Keynesian models. *J Bus Econ Stat* 25:152–154
- Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the euro area. *J Eur Econ Assoc*, 1(5), p1123–1175, 1123
- Smets F, Wouters R (2007) Shocks and frictions in US business cycles: a Bayesian DSGE approach. *Am Econ Rev* 97(3):586–606
- Wickens, M.R. (2012). *Macroeconomic theory: a dynamic general equilibrium approach* (second edition). Princeton University Press
- Wickens M (2014) How useful are DSGE macroeconomic models for forecasting? *Open Econ Rev* 25:171–193
- Wieland, V. and M.H. Wolters. (2012). Forecasting and policy making, in G. Elliott and A. Timmerman (eds.), *Handbook of Economic Forecasting*, Vol. 2, Elsevier