# Cardiff Economics Working Papers

Working Paper No. E2018/14

# Testing DSGE Models by indirect inference: a survey of recent findings

*David Meenagh, Patrick Minford, Michael Wickens and Yongdeng Xu*

May 2018

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

# Testing DSGE Models by indirect inference: a survey of recent findings

David Meenagh [*]
Cardiff University, UK

Patrick Minford[†]
Cardiff University, UK and CEPR

Michael Wickens[‡]
Cardiff University, University of York and CEPR

Yongdeng Xu[§]
Cardiff University, UK

March 2018

**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 model tractability; 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 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

## 1 Introduction

Indirect Inference is a method for testing and estimating models of any size, complexity or nonlinearity, by comparing their simulation behaviour with the behaviour of selected data. This data behaviour is summarised by a set of descriptive features, known as the auxiliary model; the structural model being investigated is then simulated many times to generate many samples for the same period as the data and the auxiliary model is then also estimated on each sample. The data-based estimates occur with some probability within this model-generated distribution and the test rejects the model if this probability lies below the test threshold. The structural model is estimated by finding the set of coefficients for which the data behaviour gets closest to the model-simulated behaviour.

The use of indirect inference in the testing and estimation of macroeconomic and other structural models has increased in the past few years and has posed many questions about its detailed application as a relatively novel procedure. In the survey by Le et al (2016) some of these questions were answered; however since then the need for answers to further questions has become apparent. In this further survey we attempt to provide answers to at least some of these questions.

The earlier survey discussed the power of the procedure in small samples, by comparison with the main frequentist testing alternative based on data likelihood, and 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, whether of small size as in Clarida et al (1999), or large as in Smets and

[*]Meenaghd@cf.ac.uk; Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

[†]Patrick.minford@btinternet.com; Cardiff Business School, Cardiff University, Colum Drive, Cardiff, CF10 3EU, UK.

[‡]Cardiff Business School, Cardiff University, Colum Drive, Cardiff, CF10 3EU, UK.

[§]Xuy16@cardiff.ac.uk; Cardiff Business School, Cardiff University, Colum Drive, Cardiff, CF10 3EU, UK.

Wouters (2007). Since then the method has spread to microeconomic and trade models, where the features of the structural model are quite different; we do not consider these here but the issues we consider and the answers we give should be easy to extend to them.

Within macroeconomic DSGE models, which we focus on here, a number of questions have surfaced. Some econometricians have suggested such models are too absurdly unrealistic to be treated as potentially 'true' under the test; instead they should be regarded as misspecified and evaluated accordingly. Then others have asked whether using VARs of low order in only a few variables can be adequate for a large model such as Smets and Wouters given that its true reduced form is a VAR of fourth order in seven variables; or would the SW parameters be very weakly identified; also whether it would matter which variables were selected for the low order/low number of variables auxiliary model? Yet again, why use a VAR as the auxiliary model instead of moments (as in the Simulated Method of Moments) or Impulse Response Functions (IRFs) as in some applied work on matching of structural models to the facts? There has also been interest in the relationship of these small sample properties to asymptotic properties of the method. These are some of the questions we address in this new survey. In some cases they have been answered in published papers, in which cases 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.

We begin with the philosophical question of whether DSGE models can be treated as 'true' for testing purposes; in some recent years a number of econometricians have dismissed DSGE models as so 'unrealistic' that one must consider them as inherently 'mis-specified'. We regard this as essentially a philosophical misunderstanding in the sense that models are not intended by construction to be 'realistic' but rather to embody economic decision-making in a logical set-up which could capture sufficient elements of economic behaviour to pass empirical tests. Such models could be termed 'pseudo-true', that is not close representations of 'reality' but rather abstract approximations designed to match the key data behaviour according to frequentist tests- much in the sense of Friedman (1953). Having renamed models in this way we can proceed in the usual manner to test them; we also show that the Indirect Inference test we use on them which tests them by treating them for the test purposes as being true, is also the most powerful way to test them for mis-specification. We set out the full philosophical argument in Appendix A.

In practice all tests are carried out on DSGE models necessarily in small samples of 200 or less observations. Hence our key concern in this paper is to review the performance of Indirect Inference in small samples. In the next section we compare it with the main alternative frequentist method, the Likelihood Ratio test, in such small samples- asymptotically the tests cannot be distinguished under the usual assumption that the model is true as Appendix C explains; we describe these two available frequentist tests — the LR and the indirect inference IIW test — and investigate the power of the IIW test to detect parameter falsity and mis-specification and ask how it can establish the extent of a model's 'truth; this section largely recaps material from our earlier survey. In section 3, we review a variety of issues in carrying out the test: such as how many and which variables it should use, and whether moments, IRFs or a VAR are preferable auxiliary models. In section 4 we review the relationship between indirect inference and identification. In section 5 we conclude with a review of our main findings.

# 2 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, to the Indirect Inference test, the IIW, in the context of small samples. What we showed was two main things: that the LR test had no power at all 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 100% of the time models whose parameters were randomly wrong by 5-7%, whereas the LR test rejects such models with only small frequency and 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- eg the VAR coefficients- is found from the data, as opposed to from the restricted model. Thus the IIW test itself is simulated from the structural model being tested; and it is the distribution of the VAR coefficients that this structural model implies, that is used in the test. Plainly this distribution is tightly related to the features of the structural model. The LR test by contrast takes the distribution implied by the sample data, as generated by the (unknown) true model.

When one considers the low power of the LR test against mis-specification, the reason is different. When one tests a mis-specified model by LR, one first has to estimate it by ML. Thus one is asking: is this structural model (mis-specified in fact) the model generating the data sample? To give the model the best chance of passing the test, one reestimates the model first. However ML finds a well-fitting model by substituting fitted error processes that are good at creating low forecast error. Thus the mis-specified model is hard to distinguish from the true model by LR. The data sample from the true model will also easily fit the mis-specified model.

By contrast when one reestimates 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 a different reduced form behaviour. Hence the very high rejection rate.

In what follows we show the key results on this from our previous paper and add some further ones.

## 2.1 Mis-specification

First we show the experiment where we test a mis-specified model: the example in the Table below shows where the true (Smets-Wouters) model is New Keynesian and the mis-specified is a New Classical version; and vice versa where the true model is New Classical and the mis-specified is New Keynesian. As can be seen the power of the II test is close to 100%, whereas that of the LR test is simply zero.

| Percentage Rejected | | |
|---|---|---|
| | NK data<br>NC model | NC data<br>NK model |
| II | 99.6% | 77.6% |
| LR | 0% | 0% |

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

We now investigate a more subtle form of mis-specification: here it takes the form of a failure to include in our model features from a more complex model that is treated as the DGP that in fact generates the data. A fortiori LR testing would be again quite unable to detect this more subtle form of mis-specification.

We set up a Monte Carlo experiment in which the DGP generating the data is such a more complex model. Our starting point will be as in Le et al (2011, 2016a) the well-known Smets-Wouters model on US data 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 the main US macro variables, output, inflation and interest rates well. 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 procedure with a VAR as the auxiliary model we estimate the power function for the falseness criterion we described above in order to assess the sensitivity of this 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, Gertler and Gilchrist (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 $(y, \pi, R)$ VAR(1) (including the three variances, as is the usual practice in applying these tests; so 12 values in all). Table 6 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).

| 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 |

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

The models we 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, in practice, we re-estimate the model. We therefore carried out the test on this basis, re-estimating the tested model on each sample generated by the true model. As this is highly time-consuming, we did this 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 7 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 8, 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.

| 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 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 | 1.480 | 8.412 | 2.409 | 68.6% |
| Original estimates | 1.291 | 11.818 | 2.909 | 79.7% |

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

As noted above a Likelihood Ratio test would have given no power for these mis-specification tests. However, plainly the IIW test manages to provide considerable power against mis-specification, as we found above.

What we find therefore is that our IIW test can establish for users a) whether they have a model that can predict relevant features of data behaviour and if so b) the bounds within which they can be sure of its specification and parameter values. With the widely-used DSGE model examined here, we found that if the model passes the test on the behaviour of three key macro variables, the power of the test largely guarantees that no other specification can be correct and that its parameter values lie within a 7-10% region of the estimated ones.

## 2.2 Parameter inaccuracy within a correct specification

Second, we show the results for numerical inaccuracy.

| VAR — no of coeffs | TRUE | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
|---|---|---|---|---|---|---|---|---|
| **IIW TEST with unrestricted VAR** | | | | | | | | |
| 2 variable VAR(1) — 4 | 5.0 | 6.2 | 20.3 | 69.6 | 61.0 | 99.8 | 100.0 | 100.0 |
| 3 variable VAR(1) — 9 | 5.0 | 3.4 | 7.5 | 30.7 | 75.0 | 97.4 | 100.0 | 100.0 |
| 3 variable VAR(2) — 18 | 5.0 | 3.8 | 5.2 | 19.1 | 57.5 | 84.3 | 98.4 | 99.5 |
| 3 variable VAR(3) — 27 | 5.0 | 3.9 | 6.4 | 21.6 | 54.5 | 84.0 | 97.5 | 98.7 |
| 5 variable VAR(1) — 25 | 5.0 | 2.8 | 3.2 | 2.6 | 5.4 | 6.2 | 4.5 | 100.0 |
| 7 variable VAR(3) — 147 | 5.0 | 5.1 | 3.4 | 1.4 | 0.9 | 0.2 | 0.0 | 100.0 |
| **IIW TEST with restricted VAR** | | | | | | | | |
| 2 variable VAR(1) — 4 | 5.0 | 9.8 | 37.7 | 80.8 | 96.8 | 100.0 | 100.0 | 100.0 |
| 3 variable VAR(1) — 9 | 5.0 | 9.5 | 36.1 | 71.0 | 98.1 | 100.0 | 100.0 | 100.0 |
| 3 variable VAR(2) — 18 | 5.0 | 8.3 | 35.5 | 80.9 | 96.9 | 100.0 | 100.0 | 100.0 |
| 3 variable VAR(3) — 27 | 5.0 | 9.2 | 32.9 | 78.0 | 95.1 | 100.0 | 100.0 | 100.0 |
| 5 variable VAR(1) — 25 | 5.0 | 17.8 | 85.5 | 99.8 | 100.0 | 100.0 | 100.0 | 100.0 |
| 7 variable VAR(3) — 147 | 5.0 | 77.6 | 99.2 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| **LIKELIHOOD RATIO TEST** | | | | | | | | |
| 2 variable VAR(1) — 4 | 5.0 | 12.0 | 28.3 | 45.9 | 63.4 | 83.2 | 97.0 | 99.7 |
| 3 variable VAR(1) — 9 | 5.0 | 9.4 | 21.8 | 37.5 | 58.9 | 84.0 | 99.0 | 100.0 |
| 3 variable VAR(2) — 18 | 5.0 | 8.9 | 20.7 | 36.8 | 57.6 | 82.9 | 98.7 | 100.0 |
| 3 variable VAR(3) — 27 | 5.0 | 8.9 | 20.4 | 36.7 | 56.7 | 82.2 | 98.7 | 100.0 |
| 5 variable VAR(1) — 25 | 5.0 | 8.9 | 22.4 | 44.3 | 68.6 | 89.6 | 99.6 | 100.0 |
| 7 variable VAR(3) — 147 | 5.0 | 5.7 | 10.6 | 23.6 | 46.3 | 83.2 | 99.6 | 100.0 |

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

We show in this Table for a wide variety of VAR auxiliary models how the LR test performs, as compared with the IIW test based on the simulations of the structural model being tested. 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. This last measures the detail with which data behaviour is described. Second, by contrast, as the detail included in the VAR description rises with higher order, the IIW test acquires more and more power. But it has high power already when the detail is kept rather low- eg with a VAR1 in just three variables, where a structural model that is just 7% false is rejected with 98% frequency.

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 (ie. unrestricted by the structural model being tested), one can see two things. First, as the size of the VAR increases, the information from the data sample becomes increasingly unable to give well-defined estimates of the VAR coefficient distribution, as too many coefficients' variation 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 the case because these two tests are transforms of each other, as noted in Appendix C, and therefore tend to produce similar test results.

This section has summarised the findings of our earlier survey paper, from which we concluded that LR testing would provide no power against mis-specification and rather low power against numerical inaccuracy, whatever VAR was used as the benchmark comparator model for the Likelihood Ratio. It has also elaborated how more subtle aspects of mis-specification are detected with high precision by Indirect Inference.

In the rest of this paper we move on from any comparison with the LR test and consider IIW tests on their own, in different forms. This is to establish how best one should use IIW tests and in what particular forms.

# 3 Comparing different ways of carrying out the IIW test- Monte Carlo experiments with small samples

## 3.1 Different variables and different features

Indirect Inference is not well known among economists as an empirical tool; 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. Hence there are a variety of practical concerns about its nature and reliability as a testing procedure.

One is whether it matters which features of the data are chosen as the 'descriptors' for the Wald test. It does not appear to matter as noted above: Minford, Wickens and Xu (2016a) showed that one may use Moments, Impulse Response Functions or VAR coefficients and 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 with the choice of variables used for the data description. For example, in evaluating the Smets-Wouters model we showed the 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 not or negligibly (Meenagh, Minford, Wickens and Xu, 2018). 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 being a nonlinear combination of the model's structural coefficients; the question is whether this is matched by the data value. 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.

Table 6: Power of the II test across data descriptors

| | Rejection Rate | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Degree of falseness | 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 |

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

Source: Minford, Wickens and Xu (2016a)

Table 7: Power of the II test across variable choice

| | Rejection Rate | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Degree of falseness | 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 |

Notes: VAR coefficients used as data descriptors, as in Le et al. (2011). Source: Meenagh, MInford, Wickens and Xu (2017)

This may appear to be puzzling: the original ideas in DSGE modelling of comparing data behaviour, such as moments, as found the data with those found in model simulations, stressed that it was good to *select* data features *of interest* to the user. However what has been found in this Monte Carlo work on small sample properties is that all these tests, provided the *number* of descriptive features is held constant, provide roughly the same power and the estimates much the same bias, whichever variables are focused on, and however their behaviour is measured, whether by moments, IRFs or VAR coefficients. What is going on?

Notice that if they have the same power these tests are evaluating the whole model and not just its ability in some selected variables' behaviour. 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 such general falseness would lie for their model and calculate robustness tests accordingly on their policy proposals.

### 3.2 Forecasting tests

Out-of-sample forecasting tests are similar to in-sample tests through the Likelihood Ratio which depends on how accurate 'nowcasting' is (i.e. the size of current predictive errors). However their power is lower. The next Table shows the power of forecasting tests for three variables jointly 4 and 8 quarters ahead, side by side with the IIW test. It can be seen that it is weak; the rejection rises close to 100% only with 20% or more falsity. Non-stationarity does not appear to affect these results.

Source: Minford, Xu and Zhou, 2015, Table 2, p 341.

### 3.3 Tests of parts of models

The focus in our tests tothis point has been on a full DSGE model. because this is what policymakers need to answer policy questions and whose reliability must be gauged in answering those questions. Sometimes however a model simply cannot be found that will pass the test set by the policymaker; or alternatively the investigator has not got the time and resources to investigate a full model but is concerned only with a few equations of a partial model. In this case a powerful Indirect Inference test for any part of a model can be carried out under Limited Information (Minford, Wickens and Xu, 2016b) whereby the rest of the model is simulated using the data VAR (from the unknown true model) while the part under examination is simulated as it is specified. This test mimics closely what one would obtain had one been able to simulate the rest of the model as it truly was (which of course one cannot know) jointly with the part being examined.

We show a fairly typical experiment with each of two parts of the SW model: the wage-price equations and the consumption-investment equations

Wage-price equations tested on auxiliary model with wages and prices VAR1

Consumption-investment equations tested on auxiliary model with consumption and investment VAR1

What we see here in the first line of each table is the power of the test for the model part using a VAR1 on all variables other than the two being explained by the equations being tested. Notice that they are being tested on whether they can match a VAR1 in these two variables.

There is strong power in the wage-price test; it seems that these two equations have a big effect on the model performance. For consumption-investment test there is less power; it seems that they are not so important to the model behaviour.

In the second line of each table one sees what the same test would have done if instead of using the VAR1 on all other variables, we had used the true model, as in the full Wald test. In both cases we find that the Limited Information test behaves rather similarly to the Full test, had we had access to the Full true model. This is reassuring as we can use the Limited Information test with some confidence when we really do not know the rest of the model- or indeed wish to test it.

## 4 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. However, tests are of little use unless one can find a model which is close enough to the truth to pass the test. Tests without estimation are like a opera tenor without top notes- of little practical operatic use. However slight reflection reveals that effective estimation is closely related to high test power: an estimator that uses a powerful test as a guide to choosing parameter values will be rejecting values that are not close to the truth. It should have low small sample bias.

When looking for a tractable model it is thus necessary to have a good estimator in the usual situation of a small sample; the key feature in this situation is low small sample bias. As is well-known the maximum likelihood estimator suffers from large small sample bias- this is directly related to its low power as a test since it it is hard to distinguish between models where different combinations of parameters, including error parameters, can give similar forecasts. The Indirect Inference estimator however has low small sample bias, 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. What we find is that 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. When one compares estimation using different three-variable sets, again we find that the bias is small and hardly differs.

$\alpha$: income share of capital
$h$: external habit formation
$\iota_p$: degree of price indexation
$\iota_w$: degree of wage indexation
$\xi_p$: 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, since the power rises with this number until the point where the full reduced form VAR is used (or its equivalent in moments or IRFs). However, this bias reduction comes at the cost of massive 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- as they are simply the values that get closest to passing it b) may not pass the weaker-powered test either-since 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. However if 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 as the estimator the same test statistic as is used in the testing process.

It is notable that Dridi et al. (2007) propose a two-step procedure to achieve both objectives: estimation and evaluation of misspecified 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 misspecified 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 II as an estimation approach, but not as a method of testing a model (Minford, Wickens and Xu, 2016 as we have seen above compare the test with different data descriptors and find that mostly the properties are quite similar). However, the two are closely related, as noted above.

# 5 Identification

A further question concerns identification. There has been a persistent suspicion that DSGE models' parameters cannot be uniquely recovered from the data- a fairly recent contribution is Canova and Sala (2009). 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 Ing (2011), Qu and Tkachenko (2012). Another avenue is to use indirect inference in a way suggested by Le et al (2017) for the same purpose. These two methods establish 'precise' identification- i.e. whether another set of parameters can be found that generate exactly the same reduced form as the DSGE model in question in the presence of unlimited data.

A related question is that of 'weak identification': by this is meant whether individual model parameter values can be distinguished from other values (from some false, competing, model version) with any confidence. Plainly as the data sample becomes smaller this becomes an increasing concern. When data is unlimited this boils down to the same as precise identification. But in practice data is limited and one would like to know about such weakness 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 that in the small three equation New Keynesian model along the lines of Clarida et al (1999) several parameters suffer from weak identification.

This analysis was done using 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?

## 5.1 Weak 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). However we may be concerned that as we reduced the number of variables and the VAR order of the (approximate) reduced form the amount of information included drops sufficiently to give highly imprecise estimates of the DSGE model parameters. When the sample is also small, this problem could become more acute. This is the problem of 'weak identification'.

We can investigate this problem through Monte Carlo experiment. We now falsify individual parameters of the SW 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 the falsity we can regard the parameter as weakly identified, since plainly we cannot distinguish well between different possible paremeter values if false cannot be distinguished from true ones.

What we can see from this exercise is that while three parameters' rejection rates rise only slowly, those of the others rise quite fast; thus the stregth of identification varies across parameters but never deserts any of them. This rather mirrors what Le et al (2017) found, using the full VAR reduced form of Smets and Wouters' model on a large sample. It seems that reducing the VAR to only three variables and order 1, together with a much smaller sample, does not create weak identification.

# 6 Conclusions

Indirect Inference is a method for testing and estimating models of any size, complexity or nonlinearity, by comparing their simulation behaviour with the behaviour of selected data. Its use in the testing and estimation of macroeconomic and other structural models has increased in the past few years and has posed many questions about its detailed application as a relatively novel procedure, that go beyond the scope of the survey in Le et al (2016a). In this latest survey we have provided answers to some of these questions.

The earlier survey discussed the power of the procedure in small samples, by comparison with the main frequentist testing alternative based on data likelihood, and 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. Within such macroeconomic DSGE models, we have argued that it is a misconception to think of a DSGE model as too 'unrealistic' to test: DSGE models can be effective models and the most efficient way to test them is 'as if true' (even if they are in fact misspecified). We have also shown that using VARs of low order in only a few variables can be adequate to provide a powerful test of a large model such as Smets and Wouters (2007) even though its true reduced form is a much higher order VAR in more variables; and that which variables are selected for this task makes little difference. It also makes little difference to the test results whether one uses a VAR as the auxiliary model instead of moments (as in the Simulated Method of Moments) or Impulse Response Functions (IRFs) as in some recent applied work.

We also showed how by using indirect inference one can test parts of models powerfully; and check too for both precise and weak identification- DSGE models do not seem to suffer much from these problems. Just as Likelihood Ratio test provide low power within sample, so too do forecasting tests out of sample.

In general we continue to think as suggested by 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.

# References


[1] Bernanke, B. S., 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.

[2] Blanchard, O. J., and C. M. Kahn. 1980. The solution of linear difference models under rational expectations. Econometrica 48 (5): 1305–11.

[3] Canova, F., 2005. Methods for Applied Macroeconomic Research, Princeton University Press, Princeton.

[4] Canova, F., Sala, L., 2009. Back to square one: Identification issues in DSGE models. Journal of Monetary Economics 56, 431–449.

[5] Christiano, L. J., 2007. Comment on 'On the fit of new Keynesian models'.Journal of Business & Economic Statistics, 25,143-151.

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

[7] Cox, D. 1961. Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statististics and Probablity

[8] Cox, D. 1962. Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society, B24, 406-424.

[9] Dave,C., De Jong, D.N., 2007. Structural Macroeconomics. Princeton University Press.

[10] Del Negro, M., Schorfheide, F., 2004. Priors from General Equilibrium Models for VARs. International Economic Review, 45, 643–673.

[11] 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.

[12] Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007a. On the fit of new Keynesian models. Journal of Business and Economic Statistics 25,123–143.

[13] Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007b. Rejoinder to Comments on 'On the fit of new Keynesian models'. Journal of Business and Economic Statistics 25,159–162.

[14] Dridi, R., Guay, A. and Renault, E. (2007) Indirect inference and calibration of dynamic stochastic general equilibrium models, Journal of Econometrics 136, 397–430.

[15] Evans, G. W. and S. Honkapohja (2005). "Interview with Thomas J. Sargent".." Macroeconomic Dynamics, 9(4): 561–83.

[16] Fernandez-Villaverde, J., Rubio-Ramirez, F., Sargent, T, and Watson, M. ,2007 'ABCs (and Ds) of Understanding VARs', American Economic Review, pp 1021-1026.

[17] Faust, J., 2007. Comment on 'On the fit of new Keynesian models. Journal of Business and Economic Statistics 25,154–156.

[18] Friedman, M. (1953). "The Methodology of Positive Economics".." In Essays in Positive Economics. Chicago: University of Chicago Press: 3–43.

[19] Gallant, A.R., 2007. Comment on 'On the fit of new Keynesian models'. Journal of Business and Economic Statistics 25,151–152.

[20] Guerron-Quintana, P., Inoue, A. and Kilian, L.(2017). "Impulse response matching estimators for DSGE models." Journal of Econometrics 196(1): 144-155.

[21] Hall, A.R., A. Inoue, J.M. Nason, and B. Rossi (2012), "Information Criteria for Impulse Response Function Matching Estimation of DSGE Models," Journal of Econometrics, 170, 499-518.

[22] Iskrev, N. (2010). "Local Identification in DSGE Models", Journal of Monetary Economics, 57(2), pp. 189–202.

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

[24] Kilian, L., 2007. Comment on 'On the fit of new Keynesian models'. Journal of Business and Economic Statistics 25,156–159.

[25] Le, V.P.M., D. Meenagh, P. Minford, M. Wickens, 2011. How much nominal rigidity is there in the US economy — testing a New Keynesian model using indirect inference. Journal of Economic Dynamics and Control 35(12), 2078–2104.

[26] Le, V.P.M., Meenagh, D., Minford, P.,Wickens, M., and Xu, Y.,2016(a). Testing macro models by indirect inference: a survey for users. Open Economies Review 27, 1-38.

[27] Le, V.P.M., D. Meenagh, P. Minford 2016(b). Monetarism rides again? US monetary policy in a world of Quantitative Easing. Journal of International Financial Markets, Institutions and Money.

[28] Le, V.P.M, Meenagh, D., Minford, P., Wickens, 2017. A Monte Carlo procedure for checking identification in DSGE models Journal of Economic Dynamics and Control, 2017, 76, (C), 202-210

[29] Liu, C. and Minford, P., 2014, How important is the credit channel? An empirical study of the US banking crisis. Journal of Banking and Finance, 41, 119-134.

[30] 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

[31] Minford, P., M. Wickens, and Y. Xu, 2016a. Comparing different data descriptors in Indirect Inference tests on DSGE models, Economics Letters, 145:157-161.

[32] Minford, P., Wickens, M., and Y. Xu, 2016b. Testing part of a DSGE model by indirect inference. Working paper E2016/12, Cardiff Economics working papers, http://carbsecon.com/wp/E2016_12.pdf, forthcoming Oxford Bulletin of Economics and Statistics.

[33] Minford, P., Xu, Y., Zhou, P., 2015. How Good are Out of Sample Forecasting Tests on DSGE Models?, Italian Economic Journal, 1:333–351.

[34] Qu, Z. and Tkachenko, D. (2012). "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearised Dynamic Stochastic General Equilibirum Models", Quantitative Economics, 3(1), pp. 95–132.

[35] Sims, C., 2002. Solving linear rational expectations models. Computational Economics 20(-2): 1-20.

[36] Sims, C., 2007. Comment on 'On the fit of new Keynesian models', Journal of Business and Economic Statistics 25,152–154.

[37] Smets, F., Wouters, R., 2003. An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. Journal of the European Economic Association, 1(5), p1123–1175.

[38] Smets, F., Wouters, R., 2007. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. American Economic Review 97, 586–606.

[39] Wickens, M., 2014. How useful are DSGE macroeconomic models for forecasting? Open Economies Review, 25, 171–193.

[40] Wright, S., 2015. The predictable and the unpredictable economy. Presentation at European Monetary Forum Conference, Erasmus University, Rotterdam, March 20-21, 2015.

## 7 Appendix A:

## 8 An inferential framework for testing DSGE models- models as 'pseudo-true'

In this paper we suggest a frequentist econometric approach to testing DSGE models under the classical null hypothesis that they are true. This approach is an alternative to other ways in which DSGE models are evaluated today, often under the assumption that they are (highly) mis-specified. We show, using both asymptotic analytic and numerical methods, that the approach has considerable advantages for users of DSGE models[1]. Thus our aim is to allow DSGE models to be used and tested in the same way as traditional 'econometric' models of the economy once were. DSGE models are the workhorse of modern policymakers, especially in central banks; and yet they are rarely tested with traditional econometric tools, so that a divide has in practice opened up between policymakers and the econometrics community. It is this divide we wish to address and try to bridge in this paper with the approach to testing DSGE models that we propose.

Much of classical econometrics was developed with a view to estimating and testing macroeconomic models. It was an integral component of the aim of making economics a science. But since the advent of DSGE macroeconomic modelling, formal tests have rarely been carried out. Calibration and Bayesian estimation of DSGE models have replaced the classical estimation of traditional macroeconometric models, arguably therefore undermining the status of macroeconomics as a science.

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. Moreover, using best-fit classical time series methods of estimation of macroeconometric models with their flexible dynamics has come to be viewed as data-mining and to have undermined the credibility of tests of these models.

DSGE modelling is, however, not without its problems. Lucas and Prescott soon found that when tested using classical methods, DSGE models were invariably rejected by classical likelihood tests. They therefore proposed the use of calibration rather than classical estimation, and their tests consisted of an informal comparison of moments simulated from the calibrated model with those observed in actual data, rather than formal statistical tests.

[1] Le et al, 2016a, gives an earlier practical guide to these methods. Here we provide an extended fundamental examination of the issues that have been raised in connection with them.

Bayesian estimation is now widely used instead of calibration or classical estimation. Its attraction is that it is a compromise between using strong priors, as in calibration, and diffuse priors, which would give the same result as classical estimation. In practice, however, the Bayesian posterior estimates are often found to be little different from their prior values but considerably different from their classical estimates, thereby providing prima facie evidence that the prior beliefs are not supported by the data, and that the model may be misspecified. If the mode of the posterior distribution is used as the point estimate then Bayesian estimation is, in effect, a weighted average of the prior values and the maximum likelihood estimates, where the weights are inversely proportional to the strength of the prior beliefs and the precision of the maximum likelihood estimates. The stronger the priors, therefore, the more likely that the posterior estimates will be close to their prior values, and the more like calibration would Bayesian estimation become.

The focus in DSGE models on structural modelling (estimating deep structural rather than "reduced-form" parameters) has resulted in the models being smaller and simpler than traditional macroeconometric models, especially in their dynamic specification. Friedman (1953) regarded the use of simple rather than complicated models as an advantage, but it makes it more likely that DSGE models are misspecified. This is one reason why DSGE models fit the data less well and are frequently rejected. The structural disturbances of DSGE models, which are the residuals between the data and the model and can be regarded as exogenous variables, are commonly found to have autocorrelated errors which are often regared as a signal of mis-specifcation of the model. The rejection of DSGE models using conventional testing procedures, the arbitrary weight given to prior distributions, and the presence of highly serially correlated structural disturbances, have all seemed to undermine the high ideals originally envisaged for the DSGE approach to macroeconomic modelling. This has led some econometricians to regard DSGE models as mis-specified and of little empirical relevance.

These arguments reveal a fundamental methodological divide between traditional macroeconometric modelling and DSGE macro modelling. 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. DSGE models are also a useful theoretical policy tool and have become the workhorse of modern macroeconomics.

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. In addition to devising suitable tests, and because, being deliberate simplications of reality, all macroeconomic models are "false" - both DSGE models and conventional macroeconometric models - we might, nonetheless, wish to know the "extent of their falseness" in order to be able to judge how useful they might still be. This has been expressed by the question "how true is your false model?" In order to answer this question we require an inferential framework that can gauge the degree of falsity of macroeconomic models. Traditional statistical tests adopt the null hypothesis that a theory is true; the power of a test is the probability of rejecting the theory if it is false. This framework does not fit easily if one starts from the premiss that the theory is false and we seek to find how true or false it is. However an alternative to the traditional approach is the null hypothesis that a model is "pseudo-true". The idea, which was developed from testing non-nested hypotheses - Cox (1961, 1962) - is to test an approximation to the "true" - if it exists - but unknown, and probably highly complex, model using the estimates of the parameters of the approximating model. (If estimated by maximum likelihood these are called quasi-maximum likelihood estimates.) In other words, we may treat DSGE models as deliberately simplified representations or approximations of the economy for which it is appropriate to apply a pseudo-true inferential framework rather than classical statistical inference. The same argument can be applied to traditional macroeconometric models. The difference is that instead of testing DSGE models directly we will use indirect inference.

This has been the focus of work of ours and coauthors in the past decade and a half in which we have used indirect inference to test prominent DSGE models estimated by others but not tested by them. Indirect inference can be seen as an example of pseudo-true inference. It involves approximating the DSGE model by an auxiliary model based on its solution, and conducting inference on this. This auxiliary model will also be a pseudo-true representation of the economy. The idea is to simulate the DSGE model and to base a test of the model on a formal comparison of estimates of the auxiliary model derived from the simulated and actual data. This is, in effect, a generalisation and formalisation of the original method used to judge the performance of calibrated DSGE models through a comparison of the moments of simulated and actual data.

This idea is, in effect, a return to Friedman's (Friedman, 1953) 'as if' methodology in which a model is treated as if it is true and is tested on that basis even though it is known to be strictly and literally untrue. The 'as if true' assumption asserts that the model has a data generating mechanism that is a close approximation to the true model, so close that statistical testing will not be able to distinguish between the two. Such a model is 'pseudo-true' in our use of Cox's definition. Strictly, if the structural model is untrue, then so is its reduced form; however, both are hypothesised to be pseudo-true.

It is helpful to illustrate these ideas using Friedman's own favourite example of a pseudo-true model: perfect competition. This, Friedman says without fear of contradiction, cannot truly exist any more than the speed of a falling object can be accurately calculated as if it is in a vacuum- the 'gravity model'. But, Friedman goes on, perfect competition is an excellent model of a highly competitive market. Aspects that are poorly modelled, due to the frictions created by such things as temporary monopoly rents, can be replaced by error terms. These can be modelled as univariate time-series processes which may be autocorrelated because such frictions may persist for some time. Thus the structural model would consist of the systematic demand and supply equations - first order conditions - and the market-clearing condition, together with the structural errors; while the reduced form model can be obtained, for example, as a VAR solution of the structural model with its own reduced form error processes derived from the structural error processes. We assert that the DGP of each model is a close approximation to the true DGP of the corresponding structural and reduced form model: they are both 'pseudo-true'. As we cannot know what the true models are, in practice we cannot check whether any candidate model is true. But we can use normal statistical methods to test whether any candidate pseudo-true model has a DGP that conforms to the actual data where the test is defined in terms of properties of the data relevant to the user. If it passes our test at the chosen confidence level then we treat it as pseudo-true and hence as if it is true.

We show in this paper that an Indirect Inference Wald (IIW) test that focuses on the parameters of the auxiliary model performs better than a maximum likelihood test which is, in effect, based on predictions from the auxiliary model. We then ask how concerned users should be about the possible mis-specification of their pseudo-true model. To investigate this we generate data from a DSGE model constructed to be more complex than the DSGE model from which we form the pseudo-true model used in our test. We find that in a typical small sample the IIW test will reject any mis-specified pseudo-true model with a probability of virtually 100%. This shows that if a pseudo-true model passes the test, it provides a sufficiently good representation of the generated data to be regarded as if it were the true model. More generally, it implies that the non-rejection of a pseudo-true model is a useful guide to the validity of the DSGE model it is based on. By calculating the power of the test as parameters are moved further away from their estimated values it is possible to establish bounds for their possible numerical falsity.

This implies that a DSGE model can be tested using classical statistical inference as if it were a true representation of the economy even though the economy's "reality" is unknown. The test of the model is whether it is pseudo-true and hence a valid statistical representation of the relevant data properties. In effect as we have said this returns us to Friedman's original methodology whereby a model is a deliberate simplification of the economy's complex reality, and in which we should test the model as if it is true in order to see whether it can get 'close' to those aspects of reality relevant for the model's user. Under this interpretation traditional macroeconometric models may also be regarded as being only pseudo-true. What distinguishes DSGE models from traditional models is their interpretation as being structural.

# 9 Appendix B: 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. (2013) 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. (2013):[2]

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

This has the solution

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

where

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

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

$$\begin{aligned}
H_{11} &= \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma}-[\frac{\lambda}{\sigma}+\omega(1+\frac{\eta}{\sigma})]\rho_\pi+\omega\rho_\pi^2} \\
H_{22} &= \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma}-[\frac{\lambda}{\sigma}+\omega(1+\frac{\eta}{\sigma})]\rho_y+\omega\rho_y^2} \\
H_{33} &= \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma}-[\frac{\lambda}{\sigma}+\omega(1+\frac{\eta}{\sigma})]\rho_r+\omega\rho_r^2}.
\end{aligned}$$

or

$$z_t = \Phi e_t \tag{3}$$

$$e_t = P e_{t-1} + \varepsilon_t \tag{4}$$

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 $i$.[3]

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:

$$\begin{aligned}
(1) \quad x_t &= Ax_{t-1} + B\varepsilon_t \\
(2) \quad y_t &= Cx_{t-1} + D\varepsilon_t
\end{aligned}$$

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 [4]

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

[2] 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.

[3] Le et al., 2013, also establish that it is identified using the IIW test in unlimited-size sampling.

[4] 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
$y_t = (\Phi P \Phi^{-1} + \Lambda) y_{t-1} - \Phi^{-1}\Lambda y_{t-2} + \Phi\varepsilon_t$

We may also note that

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

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:[5]

$$\begin{aligned} x_t &= Ax_{t-1} + B\varepsilon_t \\ y_t &= Cx_t \end{aligned}$$

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

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

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 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.[6]

# 10 Appendix C: 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

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

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

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

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

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

[5] The solution of the model can be obtained by using either Blanchard and Kahn (1980) or Sims (2002) type of algorithms.

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

Substituting this back into the likelihood function gives the concentrated likelihood

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

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

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

and hence the MLE of $V$ is

$$\widehat{V} = (\textstyle\sum_{t=1}^{T} y_t y_{t-1}^{'})(\sum_{t=1}^{T} y_{t-1}y_{t-1}^{'})^{-1}$$

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

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

In order to find the variance matrix of $\widehat{V}$ 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

$$y_i = Zv_i + \eta_i \tag{9}$$

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

$$Y = Xv + \eta \tag{10}$$

where

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

$\otimes$ 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

$$\begin{aligned} \hat{v} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \\ &= [I_k \otimes (Z'Z)^{-1}Z']Y \\ &= v + [I_k \otimes (Z'Z)^{-1}Z']\eta \end{aligned}$$

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

$$\begin{aligned} W &= plim\ T[I_k \otimes (Z'Z)^{-1}Z'](\Sigma \otimes I_T)[I_k \otimes (Z'Z)^{-1}Z']' \\ &= \Sigma \otimes (plim\ T^{-1}Z'Z)^{-1} \end{aligned}$$

### 10.0.1 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

$$\begin{aligned} LR &= 2(\ln L_U - \ln L_R) \\ &= T\left(\ln|\Sigma_R| - \ln\left|\widehat{\Sigma}\right|\right) \end{aligned} \tag{11}$$

where $L_R$ and $L_U$ denote the likelihood values of the restricted and unrestricted VAR, respectively, and $\Sigma_R$ and $\widehat{\Sigma}$ 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

$$y_t = V_F y_{t-1} + \eta_{Ft} \tag{12}$$

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:

$$LR_F = T[\ln|\Sigma_F| - \ln\left|\widehat{\Sigma}\right|] \tag{13}$$

Thus the power of the test derives from the distance

$$\ln|\Sigma_{_F}| - \ln|\Sigma_{_R}|\,. \tag{14}$$

### 10.0.2 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 $v$) or, in the form of equations (9) and (10), as

$$\begin{aligned} y_{S,i} &= Z_S v_{_{S,i}}, + \eta_{S,i} \\ Y_S &= X_S v_S + \eta_S \end{aligned}$$

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:

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

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

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

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 $\widehat{v}$. Moreover, asymptotically, this IIW statistic will have the same distribution as $[\widehat{v}-v]'W^{-1}[\widehat{v}-v]$ and hence will have the same critical values.[7] In general,

[7] The IIW test can also be carried out for a sub-set of $v$.

the IIW statistic differs from a standard Wald statistic in indirect inference which is $[\widehat{v}-v_S]'W^{-1}[\widehat{v}-v_S]$ 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 equation (12). The IIW statistic is then computed from

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

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

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

It follows that the IIW statistic can be decomposed as

$$\begin{aligned}
& [\widehat{v}-v_F]'W_F^{-1}[\widehat{v}-v_F] & (18)\\
= \; & \eta'[I_k\otimes(Z'Z)^{-1}Z']'W_F^{-1}[I_k\otimes(Z'Z)^{-1}Z']\eta \\
& +[v-v_F]'W_F^{-1}[v-v_F] \\
= \; & \eta'[\Sigma^{-1}\otimes plim\ T(Z'Z)^{-1}]\eta+[v-v_F]'W_F^{-1}[v-v_F] & (19)
\end{aligned}$$

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 equation (19).

### 10.0.3 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 equation (7) that, on the null hypothesis, the error variance matrix using simulated data is

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

where $\widehat{\Sigma}$ 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 $vec(AXB)=(B'\otimes A)vec(X)$, and $vec(V')=v$, it can be shown that

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

Hence,

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

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

$$\begin{aligned} LR &= T\left(\ln|\Sigma_F| - \ln\left|\widehat{\Sigma}\right|\right) \\ &= T\left(\ln|\Sigma_S| - \ln\left|\widehat{\Sigma}\right|\right) + T\left(\ln|\Sigma_F| - \ln|\Sigma_S|\right) \end{aligned}$$

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 $T \to \infty$, 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.

| % Misspecified | II Wald in-sample | Joint 3:4Q | :8Q |
|---|---|---|---|
| True | 5.0 | 5.0 | 5.0 |
| 1 | 19.8 | 6.0 | 4.9 |
| 3 | 52.1 | 9.4 | 5.2 |
| 5 | 87.3 | 15.3 | 6.0 |
| 7 | 99.4 | 22.9 | 6.6 |
| 10 | 100.0 | 36.2 | 9.8 |
| 15 | 100.0 | 73.8 | 29.5 |
| 20 | 100.0 | 99.8 | 90.7 |

Table 8: Rejection Rates for Wald and Likelihood Ratio for 3 Variable VAR(1)

Table 9: Rejection Rates at 95% level: falseness is given by +/- alternation, test 3 variables $(y_t, p_t, r_t)$ VAR(1).

| Falseness is given by +/- 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 +/- alternation, test 3 variables $(y_t, p_t, r_t)$ VAR(1).

| Falseness is given by +/- 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 |

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

| Data Descriptor | | VAR coeffs II estimates | | | IRF II estimates | | | Moments 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% |

Notes: 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 sdandard deviation. Source: Meenagh, Minford, Wickens and Xu (2017)

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

| Var Comb | | $(y, pi, r)$ | | | $(c, i, l)$ | | | $(q, w, r)$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 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% |

Notes: 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, Minford, Wickens and Xu (2017)

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 |
| $\iota_p$ | 0.0700 | 0.0700 | 0.0700 | 0.0720 | 0.0730 | 0.0790 | 0.1000 | 0.1350 |
| $\iota_w$ | 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 |
| $\sigma_l$ | 0.0700 | 0.0720 | 0.0730 | 0.0740 | 0.0770 | 0.0800 | 0.1060 | 0.1390 |