Cardiff Economics Working Papers

Working Paper No. E2016/5

# Comparing different data descriptors in Indirect Inference tests on DSGE models

*Patrick Minford, Michael Wickens and Yongdeng Xu*

May 2016

ISSN 1749–6101

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

# Comparing different data descriptors in Indirect Inference tests on DSGE models

Patrick Minford (Cardiff University and CEPR)*

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

Yongdeng Xu (Cardiff University)‡

January 2016

**Abstract**

Indirect inference testing can be carried out with a variety of auxiliary models. Asymptotically these different models make no difference. However, in small samples power can differ. We explore small sample power with three different auxiliary models: a VAR, average Impulse Response Functions and Moments. The latter corresponds to the Simulated Moments Method. We find that in a small macro model there is no difference in power. But in a large complex macro model the power with Moments rises more slowly with increasing misspecification than with the other two which remain similar.

**Keywords**: Indirect Inference, DGSE model, Auxiliary Models, Simulated Moments Method

**JEL Classification**: C12; C32; C52; E1

## 1 Introduction

When applying the Indirect Inference (II) test on DSGE models, many choices for the 'data descriptors', or the 'auxiliary model', are possible. A natural choice of auxiliary model is an unrestricted VAR, because a VAR is the reduced form of a DSGE model. One is then comparing the reduced form as restricted by the model and the unrestricted reduced form found in the data. The II evaluation criterion is then based on the differences between relevant VAR coefficients from simulated and actual data as represented by a Wald statistic. However, other data descriptors and auxiliary models might also be considered apart from VAR

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

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

‡Corresponding author: xuy16@cf.ac.uk;Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

coefficients. In particular impulse response functions and moments have been widely used in an informal way to evaluate the degree of 'matching' between DSGE models and the data; these could also be used in the formal testing procedures of II. One, the Simulated Method of Moments, is well-known as a form of II, though experience of its use has been limited.

In this paper, we compare the power of these two alternative ways of applying II with that of the VAR method which we have explored in previous papers (see Le et al, 2016). We evaluate the power of these different methods in small samples using Monte Carlo simulations (asymptotically they do not differ; their asymptotic equivalence in estimation is noted by Le et al, 2011). We find that in a small macro model there is no difference in power. But in a large complex macro model the power with Moments rises more slowly with increasing misspecification than with the other two which remain similar.

# 2 Properties of first order VAR

We first review the properties of first order VAR, as an auxiliary model

$$y_t = A_1 y_{t-1} + \varepsilon_t \tag{1}$$

where $\varepsilon_t$ is assumed to be $NID(0, \Sigma)$. The OLS estimates of $\hat{A}_1$ and $\hat{\Sigma}$ are:

$$\hat{A}_1 = (y_{t-1} y'_{t-1}) y_{t-1} {y_t}'$$

$$\hat{\Sigma} = \frac{\hat{\varepsilon}_t \hat{\varepsilon}'_t}{n-k} = \frac{(y_{t-1} - \hat{A}_1 y_{t-1})(y_{t-1} - \hat{A}_1 y_{t-1})'}{n-k} \tag{2}$$

The VAR model can be written as an infinite moving average process:

$$y_t = A_1 y_{t-1} + \varepsilon_t = A_1^{j+1} y_{t-j-1} + \sum_{i=0}^{j} A_1^i \varepsilon_{t-i}. \tag{3}$$

If all eigenvalues of have modulus less than 1, the sequence is absolutely summable (see Lutukepohl (2005) Appendix A, Section A.9.1). Hence,

$$y_t = \sum_{i=0}^{\propto} A_1^i \varepsilon_{t-i} \tag{4}$$

The error $\varepsilon_t$ is related to the structural innovations of the DSGE model $u_t$ as $\varepsilon_t$=$Bu_t$ , where $u_{it}$ is uncorrelated with $u_{jt}$ for $i \neq j$. We assume $B$ is known so that we can identify the structural errors causing the impulses. The impulse response function(IRF) to the shocks of the structural errors is then:

$$IRF(h) = \frac{dy_{t+h}}{du'_t} = {A_1}^{(h-1)} B, \qquad h = 0, 1, 2, ... \tag{5}$$

The average of IRF over $M$ periods is defined as

$$IRF_{Ave} = \frac{1}{M+1}\sum_{h=0}^{M} IRF(h) \quad . \tag{6}$$

We can derive the asymptotic second moments of the $y_t$ process:

$$\Gamma_y(h) = E(y_t-\mu)(y_{t-h}-\mu)'$$
$$= \lim_{n\to\propto}\sum_{i=0}^{n}\sum_{j=0}^{n} A_1^i E(\varepsilon_{t-i}\varepsilon'_{t-h-i})A_1^{j\,\prime} \tag{7}$$
$$= \sum_{i=0}^{\propto} A_1^{h+i}\Sigma A_1^{i\,\prime} \tag{8}$$

as $E(\varepsilon_t\varepsilon_s)=0$ for $s\neq t$ and $E(\varepsilon_t\varepsilon_t)=\Sigma$ for $t$.

The covariance matrix can be obtained by setting $h=0$,

$$\Gamma_0 = E(y_t-\mu)(y_t-\mu)'$$
$$= \sum_{i=0}^{\propto} A_1^i\Sigma A_1^{i\,\prime} \tag{9}$$

# 3 II test

The II test criterion is based on the difference between descriptors, the auxiliary model, from simulated data and actual data as represented by a Wald statistic, hence we call it an IIW (Indirect Inference Wald) test. If the DSGE model is correct (the null hypothesis) then the simulated data, and the data descriptors based on these data, will not be significantly different from those derived from the actual data. The simulated data from the DSGE model are obtained by bootstrapping the model using the structural shocks implied by the given (or previously estimated) model and computed from the historical data. The test then compares the data descriptors estimated on the actual data with the distribution of data descriptors derived from multiple independent sets of the simulated data. We then use a Wald statistic based on the difference between $a_T$, the estimates of the data descriptors derived from actual data, and $\overline{a_S(\theta_0)}$, the mean of their distribution based on the simulated data, which is given by:

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

where $W(\theta_0)$ is the inverse of the variance-covariance matrix of the distribution of simulated estimates $a_S$ and $\theta_0$ is the vector of parameters of the DSGE model on the null hypothesis that it is true.

Appendix shows the steps involved in finding the Wald statistic. A detailed description of the IIW test can also be found in Le et al. (2016).

When one compares the IIW tests, one finds:

1) With VAR coefficients as the data descriptors, the test uses the estimated VAR coefficients, as given in equation (2).

2) With IRF functions as the descriptors, the test uses the estimated IRF functions, as given in equation (5), which reveals that the IRF function is a nonlinear combination of VAR coefficients and the error covariance matrix (which is identified by the B matrix). If we considered the IRF over 4 years(16 periods) and take its average, then this average of IRF has 9 elements for a 3 variable VAR (1) model. This number equals the number of VAR coefficients. So the test utilises a comparable number of descriptors. We here take averages of IRFs for different shock/variable combinations.

3) With the simulated Moments (SM) as the data descriptors, the test uses the simulated moments of the data. Consider the covariance matrix, and use its lower triangular elements. For a 3-variable VAR model, we have 3(3+1)/2=6 elements to compare. The first order autocorrelation coefficients are added as additional moments. This brings the number of elements in the Wald statistic again to 9. From the theoretical moments derived above, we know that the data covariance is a nonlinear combination of VAR coefficients and the error covariance matrix. Again the number of descriptors is comparable with the number of VAR descriptors.

We know from Le et al (2016) that the DSGE models we are examining are over-identified, so that the addition of more VAR coefficients (e.g. by raising the order of the VAR) increases the power of the test, because more nonlinear combinations of the DSGE structural coefficients need to be matched. Analogously, adding more elements to the IRF descriptors (e.g. by taking averages over shorter periods) or to the moment descriptors (e.g. by taking lagged cross-moments) should do the same. Le et al (2016) noted that increasing the power in this way also reduced the chances of finding a tractable model that would pass the test, so that there was a trade-off for users between power and tractability.

# 4 Monte Carlo Experiments

We now perform some experiments comparing the power of IIW tests under these three methods in small samples, using Monte Carlo experiments on two major DSGE models. One is the three equation New Keynesian model due to Clarida, Gali and Gertler (1999). The other is the original version of Smets and Wouters model (2007); both are used with US data. The sample size is chosen as 200, which is typical for macro data.

We design Monte Carlo simulation following the same approach as Le et al (2016). Specifically, we generate the falseness by introducing a rising degree of numerical mis-specification for the model parameters. Thus we construct a model whose parameters were moved x% away from their true values in both directions (+/- alternation); similarly the higher moments of the error processes (standard deviation) are altered by the same x%. [1] We create 1000 samples from the model that is assumed to be true: then we obtain from these samples

[1] See Le et al (2016) section 4.1 for full details of the experiments.

the distribution of the Wald statistic by bootstrapping (the bootstrap number is 500) when the model is true. We use this distribution to assess how many times the x% False model is rejected with 95% confidence; notice that this fixes the size of the test throughout at 5%. The Monte Carlo simulation results are presented in Tables 1.

Table 1: Rejection Rates at 95% level (falseness is given by +/- alternatation)

| Three equation New Keynesian model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type of IIW | 0 | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
| VAR Coeffs | 0.05 | 0.314 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Average IRFs | 0.05 | 0.410 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.316 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| The Smets-Wouters model | | | | | | | | |
| Type of IIW | 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 |
| Average IRFs | 0.05 | 0.140 | 0.852 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.114 | 0.326 | 0.665 | 0.913 | 0.997 | 1.000 | 1.000 |

What is rather remarkable about these comparisons is how similar the power is across all three methods for the 3-equation model, which has a rather simple structure. For the Smets-Wouters model, it emerges that the moments IIW has substantially smaller power than the other two.

To check if our results are stable across different model mis-specification, we redo the Monte Carlo experiment by constructing a false model under two alternative arbitrary falseness criteria:

1) model parameters were moved x% away from their true values in -/+ alternation; this is the oposite ordering of +/- we used before.

2) model parameters were moved x% away from their true values in +/- randomly; here we randomise the + and the - instead of keeping a fixed sequence.

The Monte Carlo simulation results are presented in tables 2 and 3 respectively.

Table 2: Rejection Rates at 95% level (falseness is given by -/+ alternatation)

| Three equation New Keynesian model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type of IIW | 0 | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
| VAR Coeffs | 0.05 | 0.230 | 0.942 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Average IRFs | 0.05 | 0.414 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.374 | 1.000 | 1.000 | 1.000 | 0.970 | 1.000 | 1.000 |
| The Smets-Wouters model | | | | | | | | |
| Type of IIW | 0 | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
| VAR Coeffs | 0.05 | 0.112 | 0.862 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 |
| Average IRFs | 0.05 | 0.119 | 0.850 | 0.989 | 1.000 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.064 | 0.112 | 0.154 | 0.440 | 0.970 | 1.000 | 1.000 |

Table 3: Rejection Rates at 95% level (falseness is given by+/- randomly)

| Three equation New Keynesian model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type of IIW | 0 | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
| VAR Coeffs | 0.05 | 0.354 | 0.964 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Average IRFs | 0.05 | 0.446 | 0.996 | 0.985 | 0.998 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.352 | 1.000 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 |
| The Smets-Wouters model | | | | | | | | |
| Type of IIW | 0 | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
| VAR Coeffs | 0.05 | 0.220 | 0.993 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Average IRFs | 0.05 | 0.147 | 0.876 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.075 | 0.302 | 0.479 | 0.596 | 0.759 | 0.934 | 0.981 |

For the three-equation New Keynesian model, the Monte Carlo results are very stable across different experiments. The power is similar across all three methods, and essentially no different from those for the original falseness criterion. For the Smets-Wouters model, it is again clear that the Moments method has less power than the other two which remain quite similar.

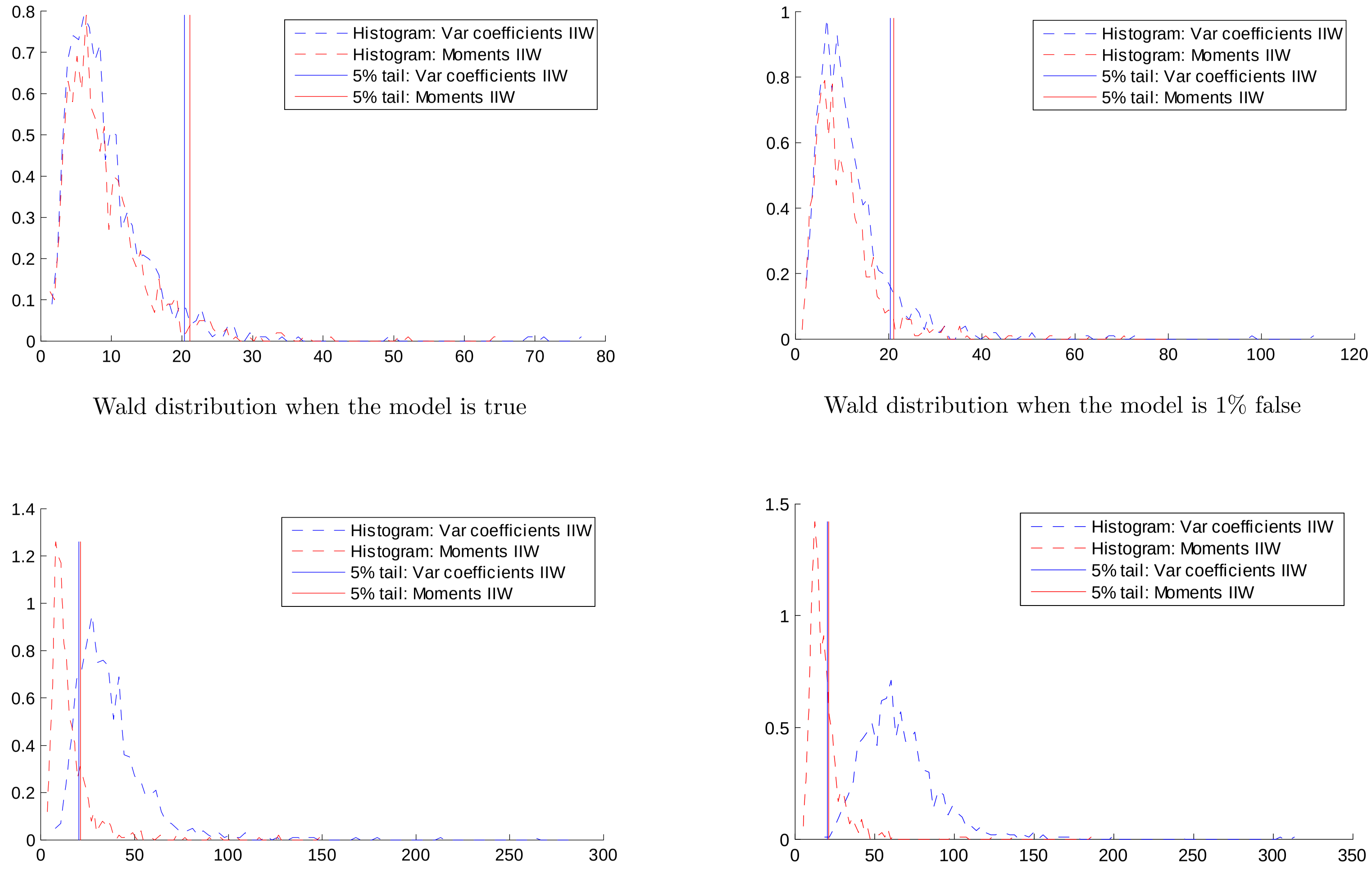
We also tried different measures of moments (for example, the first order covariance which mirrors the lag structure of the VAR coefficients; here we use the cross-moments at one lag to mirror the VAR coefficients: this implies that the auto-covariance is along the diagonal. The results are presented in table 4. Again, the results are similar.

Table 4: Rejection Rates at 95% level (falseness is given by+/- randomly)

| The Smets-Wouters model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Type of IIW | 0 | 1% | 3% | 5% | 7% | 10% | 15% | 20% |
| VAR Coeffs | 0.05 | 0.187 | 0.615 | 0.884 | 1.000 | 1.000 | 1.000 | 1.000 |
| Average IRFs | 0.05 | 0.137 | 0.422 | 0.878 | 1.000 | 1.000 | 1.000 | 1.000 |
| Moments | 0.05 | 0.110 | 0.178 | 0.243 | 0.955 | 0.999 | 1.000 | 1.000 |

It is not surprising that tests based on IRFs and VAR coefficients give similar power: the IRFs are linear combinations of the VAR coefficients as shown above by equation (5).

However what is striking is the systematically lower power of tests based on Moments for a large structural model like Smets-Wouters. Why might this arise? We may note from equation (9) that the moments are linear combinations of squared VAR coefficients. In the Wald test these numbers are in turn used as squared deviations from their model-simulated mean values. The distribution of the Wald for these values may be rather different in small samples from its distribution for VAR coefficients. So we plot the Wald distributions in Figure 1 (based on the results from experiment 1) .




Wald distribution when the model is true

Wald distribution when the model is 1% false

Wald distribution when the model is 3% false

Wald distribution when the model is 5% false

Figure 1: Wald distribution of Var coefficients and moments IIW.

What we find is rather striking. For the true distribution there is no difference in the Wald distributions. However, as the degree of falseness increases the Wald distribution for the moments does not shift to the right at all quickly, unlike that for the VAR coefficients. It seems that in a highly restricted model like Smets and Wouters false parameters do not shift the simulated moments nearly as much as they shift the simulated VAR coefficients; yet in a much simpler model with few restrictions they shift them by similar amounts. We have no explanation for this small sample finding.

In deciding whether to use VAR coefficients, IRFs, users must consider the trade-off between power and tractability we mentioned earlier and as discussed in Le et al (2016). By tractability we mean the chances of finding a model that addresses their main concerns and passes the test. Le et al (2016) give the example of a policymaker concerned to improve the stabilising performance of monetary policy for the economy; the object is to find a model that can evaluate different policies accurately. If the power of the test is set very high, then no model will be found. If the power is lowered and the test focused on the data features relevant to stability, a model may be found and a possible range established for its parameters outside which it would bound to be rejected. Then the policymaker can have confidence in the evaluation made by models in this range; supposing all are acceptable for a proposed policy improvement, then this can confidently be adopted.

What we can see from this comparative work, however, is that the use of simulated moments, rather than VAR coefficients or IRFs, could involve a substantial loss of power when using a complex model.

# 5 Conclusions

Indirect inference testing can be carried out with a variety of auxiliary models. Asymptotically these different models make no difference. However, in small samples power can differ. We explore small sample power with three different auxiliary models: a VAR, average Impulse Response Functions and Moments. The latter corresponds to the Simulated Moments Method. We find that in a small macro model there is no difference in power. But in a large complex macro model the power with Moments rises more slowly with increasing misspecification than with the other two which remain similar.

The object of high power is for users such as policymakers to have some certainty about how wrong their model could be and so calculate the robustness of their policy proposals. The greater the power the less the range of uncertainty. Our findings suggest that VAR coefficients and average IRFs are more or less interchangeable for this purpose; but that Moments give less power in testing large complex macro models and accordingly create a higher range of uncertainty.

## References


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

[2] Le, V.P.M., 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. Journal of Economic Dynamics and Control 35(12), 2078–2104.

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

[4] Lutkepohl, H. 2005.New Introduction to Multiple Time Series Analysis. Springer, Berlin and Heidelberg.

[5] McCallum, B.T., 1976. Rational expectations and the natural rate hypothesis: some consistent estimates. Econometrica, 44, 4–52.

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

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


## Appendix: Steps in deriving the Wald statistic

The following steps summarise our implementation of the Wald test by bootstrapping:

*Step 1: Estimate the errors of the economic model conditional on the observed data and $\theta_0$.*

Estimate the structural errors $\varepsilon_t$ of the DSGE macroeconomic model, $x_t(\theta_0)$, given the stated values $\theta_0$ and the observed data. The number of independent structural errors is taken to be less than or equal to the number of endogenous variables. The errors are not assumed to be normally distributed. Where the equations contain no expectations the errors can simply be backed out of the equation and the data. Where there are expectations estimation is required for the expectations; here we carry this out using the robust instrumental variables methods of McCallum (1976) and Wickens (1982), with the lagged endogenous data as instruments — thus effectively we use the auxiliary model $VAR$. An alternative method for expectations estimation is the 'exact' method; here we use the model itself to project the expectations and because these depend on the extracted residuals there is iteration between the two elements until convergence.

*Step 2: Derive the simulated data*

On the null hypothesis the $\{\varepsilon_t\}_{t=1}^T$ are the structural errors. The simulated disturbances are drawn from these errors. In some DSGE models, including the SW model, many of the structural errors are assumed to be generated by autoregressive processes rather than being serially independent. If they are, then under our method we need to estimate them. We derive the simulated data by drawing the bootstrapped disturbances by time vector to preserve any simultaneity between them, and solving the resulting model using Dynare (Juillard, 2001). To obtain the $N$ bootstrapped simulations we repeat this, drawing each sample independently.

*Step 3: Compute the Wald statistic*

We estimate the auxiliary model — a VAR(1) — using both the actual data and the $N$ samples of simulated data to obtain estimates $a_T$ and $a_S(\theta_0)$ of the vector $\alpha$. The distribution of $a_T - \overline{a_S(\theta_0)}$ and its covariance matrix $W(\theta_0)^{-1}$ are estimated by bootstrapping $a_S(\theta_0)$. The bootstrapping proceeds by drawing $N$ bootstrap samples of the structural model, and estimating the auxiliary VAR on each, thus obtaining $N$ values of $a_S(\theta_0)$; we obtain the covariance of the simulated variables directly from the bootstrap samples. The resulting set of $a_k$ vectors ($k = 1, ...., N$) represents the sampling variation implied by the structural model from which estimates of its mean, covariance matrix and confidence bounds may be calculated directly. Thus, the estimate of $W(\theta_0)^{-1}$ is

$$W(\theta_0)^{-1} = \frac{1}{N}\Sigma_{k=1}^N (a_k - \overline{a_k})'(a_k - \overline{a_k})$$

where $\overline{a_k} = \frac{1}{N}\Sigma_{k=1}^N a_k$. We then calculate the Wald statistic for the data sample; we estimate the bootstrap distribution of the Wald from the $N$ bootstrap samples.The IIW statistics are given by

$$IIW = (a_T - \bar{a}_s(\theta_o))' W(a_s(\theta_o))^{-1}(a_T - \bar{a}_s(\theta_o)) \tag{10}$$

We note that the auxiliary model used is a VAR(1) and is for a limited number of key variables: the major macro quantities which include GDP, consumption, investment, inflation and interest rates. By raising the lag order of the VAR and increasing the number of variables, the stringency of the overall test of the model is increased. If we find that the structural model is already rejected by a VAR(1), we do not proceed to a more stringent test based on a higher order VAR.