

Minford, Patrick; Xu, Yongdeng

Working Paper

Indirect inference: A methodological essay on its role and applications

Cardiff Economics Working Papers, No. E2024/1

Provided in Cooperation with:

Cardiff Business School, Cardiff University

Suggested Citation: Minford, Patrick; Xu, Yongdeng (2024) : Indirect inference: A methodological essay on its role and applications, Cardiff Economics Working Papers, No. E2024/1, Cardiff University, Cardiff Business School, Cardiff

This Version is available at:

<https://hdl.handle.net/10419/284192>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Cardiff Economics Working Papers



Working Paper No. E2024/1

Indirect Inference- a methodological essay on its role
and applications

Patrick Minford and Yongdeng Xu

January 2024

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

This working paper is produced for discussion purpose. These working papers are expected to be published in due course, in revised form.

Cardiff Economics Working Papers are available online from:

<http://econpapers.repec.org/paper/cdfwpaper/> and

https://carbsecon.com/wp/index_abstracts.html

Enquiries: EconWP@cardiff.ac.uk

Indirect Inference- a methodological essay on its role and applications

Patrick Minford

(Cardiff Business School, Cardiff University, and CEPR)

Yongdeng Xu *

(Cardiff Business School, Cardiff University)

January 2024

Abstract

In this short paper we review the intellectual history of indirect inference as a methodology in its progress from an informal method for evaluating early models of representative agents to formally testing DSGE models of the economy; and we have considered the issues that can arise in carrying out these tests. We have noted that it is asymptotically equivalent to using FIML-i.e. in large samples; and that in small samples it is superior to FIML both in lowering bias and achieving good power. In application its power needs to be evaluated by Monte Carlo experiment for the particular context. Structural models need to be defined in terms of their scope of application and auxiliary models chosen suitably to test their applicability within this scope. Power can be set too high by using too many auxiliary model features to match; and it can be pushed too low by using too few. Excessively high shocks, such as wars and crises, may also limit a model's applicability by causing unusual behaviour that cannot be captured by the model. If so, these need to be excluded so that the model is evaluated for the 'normal times' in which it is applicable.

*Corresponding Author: Cardiff Business School, Colum Drive, Cardiff, CF10 3EU, UK.
Email: xuy16@cardiff.ac.uk

1 Introduction- the Lucas Critique and the emergence of indirect inference

Modern macroeconomics can be traced back to Lucas' critique of models in the aftermath of the general adoption of rational expectations. He pointed out that reduced form models were not causal and could not therefore be used for policy analysis. They consisted of correlations produced by agents' reactions to existing policies and other exogenous processes. The models that were causal (i.e. consisted of the relationships which generated these correlations) consisted of the decision rules of households and firms, the micro agents whose reactions created these correlations. These models were termed 'representative agent' models, and latterly dynamic Stochastic General Equilibrium (DSGE) models. In them households maximise their utility subject to their budget constraints and firms their profits subject to their production functions; and markets clear, setting supplies equal to demands. The (structural) parameters of these models are those of utility and technology- sometimes called 'deep structure'-; the equations determining agents' actions are the first order conditions of the maxima combined with market clearing conditions. The dynamics come from expectations and adjustment costs.

How should such models be estimated or tested? The standard methods of OLS or best of all FIML that were used with the old reduced form models could still in principle be applied. Thus one could solve these models for their reduced form, estimate that by FIML and extract the latent structural parameters; alternatively, one could search directly for the structural parameters whose solution maximised the data likelihood (i.e. produced errors with minimum joint variance). Assuming the models are identified, this would provide estimates that would be consistent asymptotically and so with large samples quite satisfactory.

Nevertheless, Lucas, Prescott, Sargent and others leading the rational expectations revolution were doubtful about this approach, fearing it would lead to the 'rejection of too many good models'¹. Instead they favoured simulating these DSGE models to generate data moments such as cross- and lag- correlations and compare these with the data moments for 'matching closeness'. This amounted to an informal Simulated Method of Moments, already in use formally as equivalent to FIML asymptotically. Later Anthony Smith (1993) formalised this process as 'Indirect Inference'; he and Gourieroux et al (1993) demonstrated that by choosing an auxiliary model to describe the data- essentially any descriptive model, including moments and scores- and matching the structural model's simulated values for this auxiliary model's parameters as closely as possible yielded estimates asymptotically equal to those from FIML, which they

¹In a recent interview Sargent remarked of the early days of testing DSGE models: "...my recollection is that Bob Lucas and Ed Prescott were initially very enthusiastic about rational expectations econometrics. After all, it simply involved imposing on ourselves the same high standards we had criticized the Keynesians for failing to live up to. But after about five years of doing likelihood ratio tests on rational expectations models, I recall Bob Lucas and Ed Prescott both telling me that those tests were rejecting too many good models." Tom Sargent, interviewed by Evans and Honkapohja (2005).

termed 'direct inference'. In practical terms, indirect inference could be used for nonlinear models where the likelihood function is intractable.

It might seem from this account that Lucas et al should not have been concerned about the use of FIML in estimating their DSGE models, provided they could be linearised or loglinearised, both of which were often possible. If not, they could have got the equivalent FIML estimate by using formal indirect inference. However, matters do not stop there because of the practical importance empirically, especially in macroeconomics, of small samples.

When only small samples can be used for estimation and testing, it is well known that FIML produces substantial small sample bias. The reason for this appears to lie in weak identification due to the reduced sample information-Canova and Sala (2009). A good fit to the limited data can be obtained with a variety of different combinations of structural parameters and error process parameters; in other words as the structural parameters move away from the true ones, the AR and other error coefficients can be moved to keep the model as close as before to predicting the data. Ironically, this implies that Lucas and co were quite wrong about FIML 'rejecting too many good models' in the relevant context of small macro data samples; the opposite is true, namely that FIML spuriously estimates models, good or bad, and has low power in rejecting them, because it cannot distinguish reliably between good and bad models.

In recent work using Monte Carlo experiments under small samples, several papers (Le et al, 2016; Meenagh et al, 2019) have confirmed that this is the case. FIML turns out to give highly biased estimates of DSGE model parameters, and to have low power in rejecting false parameters of well-specified models, and virtually no power in rejecting mis-specified models.

However, Lucas and co were also, it turns out, right to favour indirect inference, at least in its formal form. The same experiments reveal that formal indirect inference yields low estimation bias and high power in rejecting both false parameters in well-specified models and mis-specified models. The reason appears to be that the auxiliary model parameters reflect the reduced form behaviour of the model- for brevity we will make the default assumption in what follows that the auxiliary model is a VAR-, while any falsity in the structural model will be reflected in incorrect reduced form behaviour; hence the model simulated behaviour will differ from the data behaviour. This will be true even when the error parameters move to offset falsity in the structural ones; this offsetting movement will change the reduced form, even though it can equally well replicate the data. We can see the contrast with FIML's reliance on the data fit by noting that many different VARs can fit the same data approximately well, and FIML chooses one by specifying a set of structural and AR parameters implying a VAR that fits the data well but others will also fit adequately- it is this that undermines the FIML effectiveness. Under indirect inference the process is reversed: a VAR or other auxiliary model is chosen that fits the data closely, then the structural and AR model parameters are chosen so that their simulated behaviour is closest to that of the auxiliary model. Given the chosen data-based VAR, alternative false parameter sets will not match the VAR, as they will imply a different one. Plainly another VAR could be found to describe

the data behaviour which these alternative sets could match; but under indirect inference this further VAR search on the data is not allowed. The structural model must match the one VAR chosen to best describe the data.

To put this another way, FIML in effect chooses both the VAR parameter estimates and the structural parameter estimates simultaneously and consistently with each other to achieve the best fit to the data; this 'direct inference' permits a wide variety of estimates that will fit approximately equally well- the weak identification problem. Indirect inference independently first estimates the VAR parameters that best fit the data and only then estimates the structural parameters as those that when simulated yield the best matching reduced form VAR. Given the true structural parameters, the estimated VAR will be close to the reduced form implied by the true structure: then only the true structural parameter estimates will coincidentally yield the same implied VAR. However, false structural parameters will in general produce a different simulated VAR from the true one. Thus it is the requirement of a matching coincidence between the data-based VAR and the model-simulated VAR that gives indirect inference its power and low bias.

It follows that indirect inference is indeed the ideal method for estimating and testing DSGE models. We noted above that ironically FIML would reject far fewer models, good or bad; also that indirect inference has high power and low bias. Could it then have excessive power and so threaten to reject good models? This indeed, it turns out, is a potential problem. If the auxiliary model is large and detailed, then it may only be possible for models extremely close to the true model to pass the test. In this case good models that are sufficiently close to the truth to give useful policy guidance will be rejected. For this reason, it is important to gauge the test power implied by the auxiliary model for the modelling context involved, via Monte Carlo experiment. The papers cited above (Le et al, 2016; Meenagh et al, 2019) find that an auxiliary model needs to involve three variables' coefficients in a VAR (plus the VAR error variances)- that is about a dozen features in total- or equivalent numbers of moments or IRFs- to give enough power to test structural models; more than this will reject too many useful models close to the truth.

2 Determining the scope of models and their appropriate auxiliary model

A general issue that arises with structural models is to define their scope of application and a suitable auxiliary model to test them. Thus for example, we have RBC models intended to apply to an economy's real longer term behaviour, such as regional and sectoral growth, where the shocks are local supply and demand shocks; for these the appropriate auxiliary model will be sectoral/regional outputs and employment and/or their long term growth rates- Minford, Gai and Meenagh (2022) illustrates. Then there are macro business cycle models, whose focus is on the short term behaviour of output, prices and interest rates; for

these a typical auxiliary model is a VAR, designed to pick up the dynamics of these variables- these models are illustrated by Le et al (2011) and many of the models examined in Le et al (2016) and Meenagh et al (2019).

2.1 The case of trade models

Then again there are trade models- effectively a subclass of RBC models- whose aim is to model the long term evolution of trade and traded prices; here the appropriate auxiliary model is a set of cointegrating relationships mirroring these long term trends; also the models themselves are regarded as sets of cointegrating relationships. Thus we set up the CGE trade models as equilibrium relationships and so cointegrated,

$$Ax_t = Bz_t + u_t. \quad (1)$$

where A is the cointegrating matrix, x is the vector of endogenous variables, z is the vector of non-stationary exogenous variables, such as productivity and u is the vector of other shocks. In this co-integration model, z is a nonstationary I(1) process, defining the changing equilibrium trend. The other shock vector, u , must be stationary under the true model. For simplicity we model it as AR(1), so that

$$u_t = Pu_{t-1} + \eta_t \quad (2)$$

where P is the AR coefficients for each error along its diagonal and η is an *i.i.d* innovation term. Notice that the shock includes the whole current deviation of x from its equilibrium value, $A^{-1}Bz_t$, including the 'dynamic' effects in response to the shocks due to adjustment costs and expectations. It is the gradual disappearance of these effects that creates the autocorrelation. The reduced form of this model is a VARX(1), as we can show using the ABCD method of Fernandez-Villaverde et al (2007). Hence,

$$\begin{aligned} x_t - A^{-1}Bz_t &= A^{-1}u_t = A^{-1}Pu_{t-1} + A^{-1}\eta_t \\ &= A^{-1}P(Ax_{t-1} - Bz_{t-1}) + A^{-1}\eta_t \\ &= \Lambda(x_{t-1} - A^{-1}Bz_{t-1}) + v_t, \end{aligned} \quad (3)$$

where $\Lambda = A^{-1}PA$.

Thus x can be written either as a VARX, with z as its exogenous driving vector, x . From the VARX in (3), we can write x as a VECM. Subtracting $(x_{t-1} - A^{-1}Bz_{t-1})$ on both sides in (3), we have

$$x_t - A^{-1}Bz_t - (x_{t-1} - A^{-1}Bz_{t-1}) = (\Lambda - I)(x_{t-1} - A^{-1}Bz_{t-1}) + v_t \quad (4)$$

Then, it follows that

$$\Delta x_t = A^{-1}B\Delta z_{t-1} - (I - \Lambda)(x_{t-1} - A^{-1}Bz_{t-1}) + v_t \quad (5)$$

Thus, we can write x as a VECM, where the lagged deviation from its equilibrium acts on it, pushing it towards equilibrium. In the VECM, x changes

with the change its equilibrium value as well as adjusting in response to its lagged deviation from equilibrium. We can also note that the elements of x will be cointegrated in a variety of reduced form relationships with each other and with z , owing to their common trends in z . These relationships we treat as the auxiliary model. In these cases, as with the other models, the choice of auxiliary model can be tested for suitable power by Monte Carlo experiment- illustrated by Minford, Xu and Dong (2023).

2.2 Excessive model variance- how should we deal with it?

A particular problem that may occur in indirect inference testing is that the structural model shocks may exhibit massive variation, so that the simulation distribution of auxiliary parameters is huge, implying that it can 'match' any auxiliary model, in the sense that none can be rejected at the usual confidence level- destroying the power of the test. This happens occasionally, with models of high complexity, strong responses and data samples embracing episodes with very large shocks.

An example of this occurred in testing World Trade Models against global and country group data- Minford, Xu and Dong (2023) cited above. It turned out that in this model the simultaneity of all countries' net supplies being equilibrated by world prices in the global markets for traded goods created great volatility in some simulations, which seriously weakened the test power. The sample period of annual data ran from 1970 to 2019, during which there were major shocks to world trade, including China's accession to the WTO at the end of 2001 and the Global Financial Crisis in 2008. In order to create good power, the authors eliminated simulations with high volatility (the top 5%) and checked via Monte Carlo experiment that with many simulations of moderate volatility the test power was appropriate. The Table below is reproduced from the paper. The experiment treated the Classical model as the true one and generated (500) data samples from it, using the full model and bootstrapping the shocks across their full range; these would therefore have been the possible data samples from this model. To generate the Wald statistic, the model was simulated with the largest variance simulations (the top 5%) eliminated; by restricting the range in this way, the test had good power. The second column of Table 1 shows that when simulations are chosen in this way, the test power is quite good, with the full world model being rejected nearly 60% of the time when model coefficients are falsified by 7%. However, when we use the full range of simulation, the power would be low- as shown in third column of Table 1 where we redid the tests using the full range of simulations- the power is so low that the model with its parameters 20% falsified is only rejected 13.5% of the time. In effect false models cannot be distinguished from the true model because all can predict the data behaviour with high probability. To distinguish them, as we must, they must be tested with shocks that are within the normal size range, which is intended scope of the model. What this reveals is that to create power in testing the model, simulation variance must be limited. in line

Percent Misspecified	Rejection Rates at 95% Confidence Level	
	Excl. most volatile simulations	All simulations
True	5.00%	5.00%
1	5.45%	6.29%
3	13.8%	9.14%
5	31.6%	14.26%
7	57.42%	16.07%
10	67.66%	21.26%
15	68.25%	12.83%
20	82.05%	13.50%

Note: The second column replicates Table 1 from Minford, Xu, and Dong (2023), where the top 5% largest and smallest simulations have been removed. The third column replicates the same simulation but includes all simulations.

with the model's scope.

It is important, given the variable power of different indirect inference procedures that the one chosen for any context is evaluated in this way by Monte Carlo experiment. This allows us to gauge from the test the probable degree of accuracy of the models being tested..

The essential point about Indirect Inference is that it is a procedure for checking the match of a simulated model to estimated data behaviour that can take many forms in detail. As the object of the procedure is to test the truth of a model, the precise form the procedure should take is to be selected according to its test power; the form can vary in the type and extent of the detail used in the auxiliary model and in the range of simulated behaviour examined. This power should be 'goldilocks'- not so great that it rejects all tractable models but great enough to reject poor models. Since the context in which the test is used is that of small samples, this choice should be based on Monte Carlo experiment on the model being tested.

Limiting the number of auxiliary model elements to achieve goldilocks power is natural enough. There is no point in setting power so high that only the exact true model (i.e. actual reality) could pass. Restricting the variance of simulations might seem arbitrary, on the grounds that one is jettisoning the model shocks that are big, purely to provide discrimination between model fits. Yet such a discard is a familiar process in applied macroeconomics. We regularly exclude, or equivalently create dummies for, wartime and such unusual shocks as Covid or even the Great Financial Crisis on the grounds that usual behaviour would change in ways difficult to analyse. Indeed in the GFC one of us was asked at a business executives' forecast conference whether it was possible 'to see a bottom', and replied 'yes given the aim of the bailout to do what it takes to create a bottom'; essentially the bailout aimed to restore normal behaviour. Thus our models are intended to explain 'normal times'; and excluding extreme shocks is a way of testing them for such times. Essentially the

low power of Indirect Inference when very large shocks are included is confirming that models may behave unusually in response to such shocks, making it difficult to discriminate between different models and so also to reject any particular model.

In effect we are here determining the scope of the model as being over normally-sized shocks, and saying the model is inapplicable under very large shocks because these bring other elements into play. In this respect the determination is similar to defining a model as a short term business cycle model without applicability to long term growth or related trends. The auxiliary model for this trade model remains the trended cointegrating relationships found in the data which should be little affected by occasional large shocks.

It would seem from our results that the data sample trend relationships produced by the full model with the full set of shocks are dominated by the normal-sized shocks, so that the cointegrating parameters are modestly affected by the extreme shocks. However, as simulations are repeated many times with the full range of shocks, these modest variations between simulation estimates accumulate (eg sim 1 minus sim2, and sim3 minus sim2. both small, will if in the same direction add to a large sim3 minus sim1) creating large cross-simulation variance, which is the source of the loss of power. By restricting the shocks to normal size, we are discarding model behaviour that would be dominated by the extreme shocks, on the grounds that this behaviour reflects these shocks predominantly and not the model, witness the fact that false models reflect them similarly. Thus extreme shocks produce similarly extreme behaviour in true and false models, so giving poor information about the models. In short we are comparing our models on the basis of normal behaviour, sans extreme shocks.

3 Conclusions

In this short paper we have reviewed the intellectual history of indirect inference as a methodology in its progress from an informal method for evaluating early models of representative agents to formally testing DSGE models of the economy; and we have considered the issues that can arise in carrying out these tests. We have noted that it is asymptotically equivalent to using FIML-i.e. in large samples; and that in small samples it is superior to FIML both in lowering bias and achieving good power. In application its power needs to be evaluated by Monte Carlo experiment for the particular context. We have seen how structural models need to be defined in terms of their scope of application and auxiliary models chosen suitably to test their applicability within this scope. Power can be set too high by using too many auxiliary model features to match; and it can be pushed too low by using too few. Excessively high shocks, such as wars and crises, may also limit a model's applicability by causing unusual behaviour that cannot be captured by the model. If so, these need to be excluded so that the model is evaluated for the 'normal times' in which it is applicable.

References

- [1] Canova, F., & Sala, L. (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56, 431-449.
- [2] Evans, G., & Honkapohja, S. (2005). An interview with Thomas Sargent. *Macroeconomic Dynamics*, 9(4), 561-583.
- [3] Fernandez-Villaverde, J., Rubio-Ramirez, F., Sargent, T., & Watson, M. (2007). ABCs (and Ds) of Understanding VARs. *American Economic Review*, 97(3), 1021-1026.
- [4] Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(Supplement: Special Issue on Econometric Inference Using Simulation Techniques), S85-S118.
- [5] Guerron-Quintana, P., Inoue, A., & Kilian, L. (2017). Impulse Response Matching Estimators for DSGE Models. *Journal of Econometrics*, 196(1), 144-155.
- [6] 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.
- [7] Le, V.P.M., Meenagh, D., Minford, P., Wickens, M., & Xu, Y. (2016). Testing macro models by indirect inference: a survey for users. *Open Economies Review*, 27, 1-38.
- [8] Le, V.P.M., Meenagh, D., Minford, P., Wickens, M. (2017). A Monte Carlo procedure for checking identification in DSGE models. *Journal of Economic Dynamics and Control*, 76(C), 202-210.
- [9] Lucas, R.E. Jr. (1976). Econometric policy evaluation: A critique. In K. Brunner & A.H. Meltzer (Eds.), *The Phillips Curve and Labour Markets* (Carnegie Rochester Conference Series on Public Policy No. 1, pp. 1-18). Supplement to the *Journal of Monetary Economics*.
- [10] Meenagh, D., Minford, P., Wickens, M., & Xu, Y. (2019). Testing Macro Models by Indirect Inference: A Survey of recent findings. *Open Economies Review*, 30(3), 593-620.
- [11] Meenagh, D., Minford, P., & Xu, Y. (2022). Why does Indirect Inference estimation produce less small sample bias than maximum likelihood? A note. *Cardiff Economics Working Papers E2022/10*, Cardiff University, Cardiff Business School, Economics Section.
- [12] Minford, P., Gai, Y., & Meenagh, D. (2022). North and South: A Regional Model of the UK. *Open Economies Review*, 33, 565-616.

- [13] Minford, P., Ou, Z., & Zhu, Z. (2021). Can a small New Keynesian model of the world economy with risk-pooling match the facts? *International Journal of Finance & Economics*, 26(2), 1993-2021.
- [14] Minford, P., Wickens, M., & Xu, Y. (2016). Comparing different data descriptors in Indirect Inference tests on DSGE models. *Economics Letters*, 145(C), 157-161.
- [15] Minford, P., Wickens, M., & Xu, Y. (2018). Testing Part of a DSGE Model by Indirect Inference. *Oxford Bulletin of Economics and Statistics*, 81(1), 178-194.
- [16] Minford, P., Xu, Y., & Dong, X. (2023). Testing competing world trade models against the facts of world trade. *Journal of International Money and Finance*, 138(C).
- [17] Smith, A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8, 563-584.