DISCUSSION PAPER SERIES

No. 7621

SOME PROBLEMS IN THE TESTING OF DSGE MODELS

Vo Phuong Mai Le, Patrick Minford and Michael R. Wickens

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP7621.asp**

ISSN 0265-8003

SOME PROBLEMS IN THE TESTING OF DSGE MODELS

Vo Phuong Mai Le, Cardiff University
Patrick Minford, Cardiff University and CEPR
Michael R. Wickens, Cardiff University, University of York and CEPR

Discussion Paper No. 7621
January 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 7621

January 2010

ABSTRACT

Some Problems in the Testing of DSGE Models*

We review the methods used in many papers to evaluate DSGE models by comparing their simulated moments and other features with data equivalents. We note that they select, scale and characterise the shocks without reference to the data; crucially they fail to use the joint distribution of the features under comparison. We illustrate this point by recomputing an assessment of a two-country model in a recent paper; we find that the paper's conclusions are essentially reversed.

JEL Classification: C12, C32, C52 and E1
Keywords: anomaly, boostrap, dsge, indirect inference, puzzle, us-eu model, var and wald statistic

Vo Phuong Mai Le
Cardiff Business School
Cardiff University
Colum Rd
Cardiff
CF10 3EU
Email: LeVP@cardiff.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=164365

Patrick Minford
Cardiff Business School
Cardiff University
Colum Rd
Cardiff
CF10 3EU
Email: minfordp@cf.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100320

Michael R. Wickens
Department of Economics
and Related Studies
University of York
Heslington
York YO10 5DD
Email: mrw4@york.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100329

* We thank Arnab Bhattarachee, Huw Dixon, George Evans, Paul Levine, David Meenagh, Jo Pearlman, Christian Thoenissen, Jiang Yue and participants at the CDMA Conference in St. Andrew's 2009, for helpful comments. This work was supported by the UK's Economic and Social Research Council under grants RES-165-25-0020 and PTA-026-27-1623.

Submitted 18 December 2009

1 Evaluating DSGE models—some problems with existing procedures

A popular way to evaluate macro DSGE models is to ask whether they can replicate 'facts' such as a correlation between two variables, like consumption and the real exchange rate—generally a reduced form relationship that is reliably supposed to be observed in various available samples of data. Thus a model, when subjected to the shocks that under its theory are supposed to strike it, should generate under such stochastic simulation a relationship on average close to the one in the data, while also similarly generating other relationships close to those found in business cycle data.

The question we address in this paper is whether this procedure, as widely carried out at present, is being correctly implemented from the viewpoint of statistical testing. We highlight three aspects of this procedure:

1) It is usual to limit the shocks applied to the DSGE model to a few: for example productivity, money supply and government spending shocks. Yet consider a DSGE model of say 5 endogenous variables and 5 equations. When confronted with the data this model's implied residuals can be estimated together with the residuals' time-series properties. These residuals are the 'shocks' it implies. Furthermore, unless the model includes all 5 of these shocks, 'stochastic singularity' occurs, whereby at least one variable is a deterministic combination of the others—necessarily this will be false unless the 'shock' in one equation is zero. While one or more shocks might be treated as non-stochastic (eg as 'measurement errors'), this must be justified on the grounds that such were 'one-off' in nature; then the data needs to be adjusted for their effect before the comparison of the stochastic simulations with the 'facts'.

2) One or more of these shocks may also be scaled to fit some basic variances, such as that of GDP; also their time-series properties may be imposed by the modeller. Yet when we know as above the 'true' shocks conditional on the model being tested (the null), any arbitrary choice of scale and time-series properties will violate the model and data. For example it is possible that the model greatly over- or under-predicts the data moments: by scaling the errors this discrepancy can be removed. Yet in effect one is multiplying the true errors by a set of impact parameters that now enter the model in addition to existing parameters; these parameters also impact on other variables.

3) There is no consideration of the joint distributions of the chosen descriptors. It is usual to show the distributions of the descriptors individually. Yet the test is whether as a group they are consistent with the model; for this it is necessary to use their joint distribution since the model will imply covariances between these descriptors. Thus for example the Fisher equation in a DSGE model will imply that the persistence of inflation and interest rates, two favourite descriptors, will be highly correlated: thus in samples created by the DSGE model from its shocks where inflation is persistent, so will interest rates be; and vice versa.

2 The problems in practice in a well-known two-country DSGE model

We now illustrate the problems we have highlighted by reworking a well-known paper modelling a two-country world of the US and EU, Chari, Kehoe and McGrattan (2002), CKM; their work is fairly representative—see also for example Henriksen, Kydland and Sustek (2008), and Kollmann (2009).

In their paper CKM calibrate a model of the US and Europe in which wage and price rigidity, adjustment costs in investment, habit persistence in consumption, and incomplete asset markets are all assumed at various stages. CKM's aim is primarily to replicate the autocorrelation and volatility of prices, the nominal and the real exchange rate in the data but they are also concerned to replicate some business cycle moments—variability of consumption/investment/net exports relative to GDP, their autocorrelations and their cross-correlations. Their conclusion is that they can match the business cycle behaviour reasonably but are left with a minor anomaly that the model cannot quite generate enough pesistence in the real exchange rate and a key anomaly, that the real exchange rate is highly positively correlated with relative consumption in the model but hardly at all in the data.

Our aim is to examine these claims by applying the methods of CKM to an essentially similar model Le et al. (2009) constructed based on the work of Smets and Wouters (2003, 2007; SW). CKM emphasise that their claims are robust to a wide range of model specifications, including adding a variety of shocks, varying the monetary rule between a money supply rule and a Taylor Rule, including complete international risk sharing as opposed to uncovered interest parity in bonds. It is therefore natural to make use of the work Le et al. did on SW's model where they addressed the full range of issues noted above; they replaced assumed shocks with actual and estimated shocks backed out of the model, they included all observable shocks, unscaled; and they computed the distributions (both individual and joint) of the descriptors or 'facts' under the null hypothesis of the model itself. We are able, using their methods, largely to replicate the nature of the CKM results as reported, though we find that their 'minor anomaly' is no longer an anomaly. Thus allowing for the 'true' shocks makes some difference. However, what we will show is that when joint distributions are allowed for, the implications of these results are quite importantly different.

2.1 Replicating CKM's findings

Let us first show how our procedures replicate or alter CKM's findings: a) that the model matches business cycle behaviour b) that it just fails to match the real exchange rate's persistence c) that it fails massively to match the absence of correlation between the real exchange rate and relative consumption.

a) We find that the SW model we have used to represent CKM's broadly replicates their results on business cycle variables. Thus the Table below finds as they do in Table 6 of their paper that the SW model matches a set of business cycle moments. For all but two of the moments shown the model mean bootstrap estimate is within two (model-generated) standard deviations of the data values. This is actually rather better than CKM find for their model, so if anything it strengthens their claim that models of this general type do well in matching the business cycle individual measures.

	Data	Model bootstrap Estimate	(Model bootstrap Stdev)
Standard deviations Relative to GDP			
US			
Consumption	1.1324	1.0411	(0.1414)
Investment	4.1514	2.4789*	(0.4564)
Net exports	3.3390	3.7587	(1.1972)
EU			
Consumption	1.2313	1.0692	(0.0881)
Investment	3.2080	3.9733	(0.5338)
Autocorrelations			
US			
GDP	0.9362	0.9331	(0.0340)
Consumption	0.9484	0.9400	(0.0345)
Investment	0.9697	0.9379	(0.0290)
Net exports	0.9356	0.9132	(0.0460)
EU			
GDP	0.9502	0.8724	(0.0598)
Consumption	0.9651	0.8643	(0.0624)
Investment	0.9638	0.9269	(0.0353)
Cross-Correlations			
between foreign and domestic			
GDP	0.4449	0.1282	(0.3551)
Consumption	0.0261	0.0096	(0.3749)
Investment	−0.2342	0.0611	(0.3339)
between net exports and US GDP	−0.1043	−0.4916	(0.2842)
between real exchange rate and			
US GDP	−0.2312	−0.3687	(0.3185)
Net exports	−0.1653	0.9786*	(0.0177)

* outside 2-standard-deviation bounds

Table 1: Business cycle statistics for the model

With b) we find in fact that the model can comfortably match the persistence of the real exchange rate, taken on its own, at all lags; thus CKM's minor anomaly is as they imply not at all serious. They get close and we find that when full stochastic simulation of all shocks is carried out, they would have succeeded. The reason appears to be simpy a failure to generate the model's bounds using all the shocks coming from the data. This is the main alteration we find due to using the true shocks implied by the model.

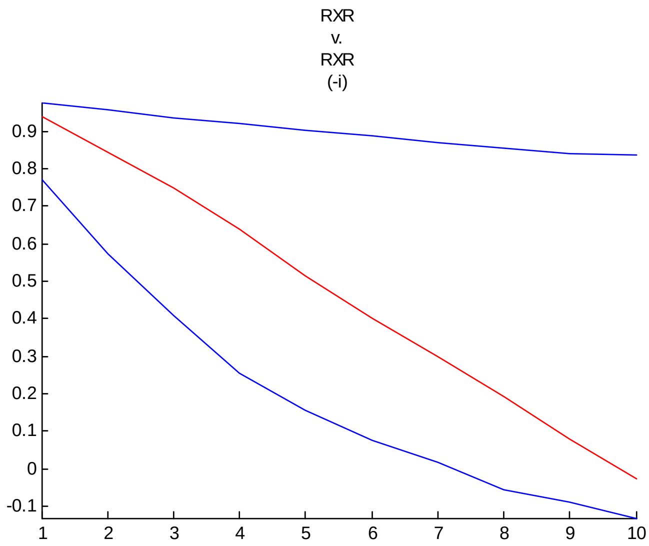

Figure 1: RXR auto-correlation

Finally, we consider c) CKM's 'key anomaly'. Here as with a) we replicate CKM's results with the SW model. It too generates powerful positive correlations between the real exchange rate and relative consumption at all leads and lags. The lower 95% bound of these does indeed lie well above the data cross-correlation.

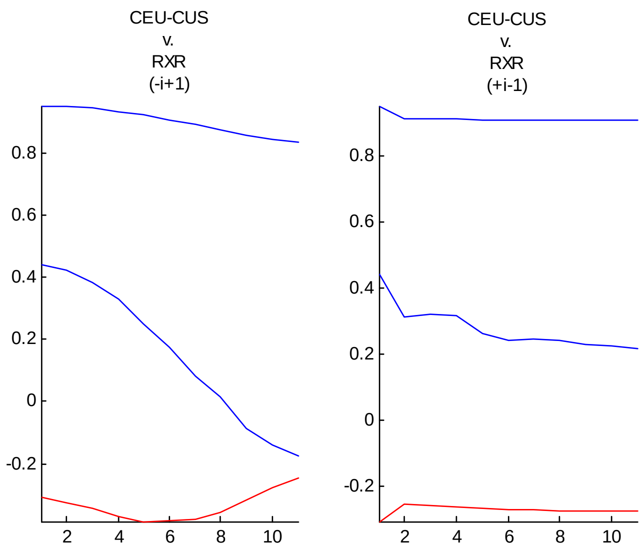

Figure 2: Correlation of $C^{EU} - C^{US}$ with RXR

Thus we can see that the SW version of CKM does replicate qualitatively the results they report in their work, with the marginal exception of the real exchange rate's persistence where we suggest their 'minor anomaly' was due to not including the full range of shocks backed out of the data and the model.

2.2 Augmenting and reinterpreting CKM's findings

CKM's conclusion from these results is that even though the model matches business cycle data it has a particular flaw in being unable to match the real exchange rate/relative consumption behaviour. Thus they suggest future research should look for a mechanism that remedies this flaw (possibly in the international risk-sharing area)—such as later for example Kollmann (2009).

However, now consider the joint distribution of the business cycle features in Table 1. In contrast to CKM's claims, while the SW model of the US and the EU matches these business cycle moments individually as shown, it does not does not match the *joint* behaviour of real business cycle variables at all well. The Wald statistic for the joint distribution of all these descriptors is 100 (ie the joint data values fall in the 100th percentile of the bootstrapped distribution), indicating strong joint rejection. Even if we exclude the two features (asterisked) that individually do not fit from this distribution, the Wald statistic remains at 100. The normalised Mahalanobis distances are respectively 1593 and 54 (against a 95% value of 1.65).

This strong rejection is general across whatever measures one likes to use of the model's fit to business cycle data. We report in Le et al. (2009) a wide variety of measures that this model fails to fit jointly, including data variances, VAR parameters and Impulse Response Functions.

Since these features are generally within their 95% bounds individually, the reason for this rejection seems to be that the joint distributions are far more demanding than the totality of the single distributions because these chosen features are highly correlated according to the DSGE model being tested.

Now if we turn to the major anomaly highlighted by CKM we see it in a rather different light. It now appears that this particular correlation rejects the model in common with general business cycle behaviour. Notice that the joint behaviour of variables in the model in general sets up tight restrictions on the joint distribution of data features. Whenever several variables are interacting, their joint behaviour according to the model follows particular patterns and these are in general not found in the data. Thus the failure to match the complex joint corrrelation between two countries' consumption and the real exchange rate is an unsurprising implication of a failure to match consumption's joint behaviour with other business cycle variables.

3 Conclusions

Our aim here has been to highlight the pitfalls of current procedures used to assess DSGE model performance. We have argued that the model's shocks should be estimated from the data, not imposed, and that all the shocks should be used; but more importantly that model properties should be assessed against the data using their joint distributions which generally pose more stringent requirements than their single distributions viewed collectively. We illustrated this in a DSGE model of the US and EU characterised by a high degree of nominal rigidity; we found for such a model that while indeed data features could be matched individually they rejected the model jointly, in particular consumption's behaviour jointly with other business cycle variables including the real exchange rate. The general point is not that this model failed in some unique respect but rather that it failed generally by imposing filigree restrictions on joint variable behaviour that were just not found in the data.

References

[1] Chari, V., P. J. Kehoe and E. McGrattan (2002) 'Can sticky price models generate volatile and persistent real exchange rates?' Staff Report 277 from Federal Reserve Bank of Minneapolis, published in Review of Economic Studies, Vol. 69, No. 3, July 2002, pp. 533-563.

[2] Henriksen, E., Finn E. Kydland, and Roman Sustek (2008) 'The High Cross-Country Correlations of Prices and Interest Rates', mimeo, Bank of England

[3] Kollmann, Robert (2009) 'Household Heterogeneity and the Real Exchange Rate: Still a Puzzle', CEPR Discussion Paper No 7301, C.E.P.R., London.

[4] Le, M., D. Meenagh, P. Minford and M. Wickens (2009) 'Two Orthogonal Continents: Testing a Two-country DSGE Model of the US and EU Using Indirect Inference', Cardiff University Economics Working Paper http://www.cf.ac.uk/carbs/econ/workingpapers/papers/E2009_3.pdf, forthcoming in Open Economies Review.

[5] Smets, F. & R. Wouters, 2003. 'An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area', Journal of the European Economic Association, MIT Press, vol. 1(5), pages 1123-1175, 09.

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