

# CAMA

Centre for Applied Macroeconomic Analysis

---

## Checking if the straitjacket fits

---

### CAMA Working Paper 81/2019 November 2019

**Adrian Pagan**

University of Sydney

Centre for Applied Macroeconomic Analysis, ANU

**Michael Wickens**

Cardiff Business School, Cardiff University

University of York

CEPR and

CESifo

### Abstract

Pesaran and Smith (2011) concluded that DSGE models were sometimes a straitjacket which hampered the ability to match certain features of the data. In this paper we look at how one might assess the fit of these models using a variety of measures, rather than what seems to be an increasingly common device - the Marginal Data Density. We apply these in the context of models by Christiano et.al (2014) and Ireland (2004), finding they fail to make a match by a large margin. Against this, there is a strong argument for having a straitjacket as it enforces some desirable behaviour on models and makes researchers think about how to account for any non-stationarity in the data. We illustrate this with examples drawn from the SVAR literature and also more eclectic models such as Holston et al (2017) for extracting an estimate of the real natural rate.

## Keywords

## JEL Classification

## Address for correspondence:

(E) [cama.admin@anu.edu.au](mailto:cama.admin@anu.edu.au)

ISSN 2206-0332

[The Centre for Applied Macroeconomic Analysis](#) in the Crawford School of Public Policy has been established to build strong links between professional macroeconomists. It provides a forum for quality macroeconomic research and discussion of policy issues between academia, government and the private sector.

**The Crawford School of Public Policy** is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

# Checking if the Straitjacket Fits

Adrian Pagan\* and Michael Wickens†

November 7, 2019

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>A Measure of Fit</b>	<b>5</b>
<b>3</b>	<b>General and Specific Discrepancy Measures</b>	<b>6</b>
3.1	General Tests : Moments Matter . . . . .	6
3.2	Specific Measures . . . . .	7
<b>4</b>	<b>System Wide Discrepancy Measures</b>	<b>8</b>
4.1	Targeting Component Parts of the System . . . . .	8
4.2	Targeting VAR Dynamics . . . . .	9
4.3	Targeting the VAR Covariance Matrix: Agnostic Shocks . . .	10
<b>5</b>	<b>Is the Straitjacket Fit for Purpose?</b>	<b>14</b>
<b>6</b>	<b>Can we Modify the Straitjacket, not the Model?</b>	<b>15</b>
<b>7</b>	<b>When is a Straitjacket a Useful Thing?</b>	<b>17</b>
<b>8</b>	<b>Conclusions</b>	<b>24</b>

---

\*Emeritus Professor, University of Sydney and Honorary Professor, Australian National University (adrian.pagan@sydney.edu.au)

†Professor of Economics, Cardiff Business School, Cardiff University; Emeritus Professor of Economics, University of York; CEPR and CESifo; YO10 5DD, UK; mike.wickens@york.ac.uk

### Abstract

Pesaran and Smith (2011) concluded that DSGE models were sometimes a straitjacket which hampered the ability to match certain features of the data. In this paper we look at how one might assess the fit of these models using a variety of measures, rather than what seems to be an increasingly common device - the Marginal Data Density. We apply these in the context of models by Christiano et.al (2014) and Ireland (2004), finding they fail to make a match by a large margin. Against this, there is a strong argument for having a straitjacket as it enforces some desirable behaviour on models and makes researchers think about how to account for any non-stationarity in the data. We illustrate this with examples drawn from the SVAR literature and also more eclectic models such as Holston et al (2017) for extracting an estimate of the real natural rate.

## 1 Introduction

Pesaran and Smith (2011. p14) conclude their critique of the state of DSGE models around 2010 with the words "We have argued that macroeconomic modelling would benefit from a more flexible approach which does not require narrow adherence to one particular theoretical framework. In the process one would need to be more explicit about the trade-offs between consistency with theory, adequately representing the data and relevance for particular purposes". They also say (p. 12) "We have argued that theory, while essential, should be regarded as a flexible framework rather than a straitjacket, because features that the theory abstracts from may be important in practice". Straitjackets come in different sizes. The objective is to keep them as small as possible but maintain the patient in a secure way. Because there are side-effects of being compressed in one, there is a need to ensure that they fit well. So it should be with DSGE models (and any other macroeconomic model whose use is being). In this paper we ask whether that is being done and how it might be done better.

Section 2 of the paper begins with the question of whether there is a single criterion to use to check fit. This may seem an odd question to ask, but it stems from the widespread adoption of Bayesian methods to estimate macroeconomic models, and what seems to have become a single criterion,

namely the value of the marginal data density. In fact, there are many proposals in the Bayesian literature about model-checking, and the need for it, but these seem to have been little used in macroeconomics. Section 2 discusses these. That literature emphasizes the need for *discrepancy measures* to characterize the difference between the data and model outcomes. These can be as simple as just checking the correspondence of moments. As we show, some well-known DSGE models fail to produce moments that agree with those computed from the data. These are generic tests, but there can be more *targeted* discrepancy measures looking at items such as the cycles likely to be generated by the model, and also whether the assumption of model-consistent expectations is in agreement with the data.

Section 3 considers measures specifically oriented towards discrepancies between systems. Pesaran and Smith highlighted the need to have flexible dynamics and so it is natural to ask if the dynamics present in the data are being reproduced by the model. A simple way of doing this is through a comparison with a VAR. This was certainly a focus of attention in early DSGE model work but seems to have lapsed. One problem may have been that researchers were fitting high-order VARs to the data which, as they involved a large number of parameters, were imprecisely estimated, and so may produce a test with low power. However, the indirect estimation literature shows that low-order VARs can also produce tests with good power. We carry out such tests for a number of models in the literature. In another vein, there has recently been a suggestion by den Haan and Drechsel (2018) that testing DSGE models can be done by introducing "agnostic shocks". We look at that work and how it relates to testing via a VAR.

Section 4 asks whether the assumptions embodied in the DSGE model set up and used in estimation - namely that shocks are uncorrelated and innovations have no serial correlation - remain valid for the empirical shocks calculated after estimation. Unlike SVAR models, DSGE models are heavily over identified, so the proclaimed assumptions need not hold for the estimated model shocks - see Liu et al (2018). In particular, the empirical shocks may be correlated, and this can affect standard SVAR procedures such as variance and variable decompositions which rely on shocks being uncorrelated, i.e. what is true of an exactly identified SVAR need not be true of an over-identified DSGE model. One case where these shocks *must be correlated* is when there are more shocks in the DSGE model than observed variables in the data. Ravenna (2007) noted that one could not recover the true shocks in that case. This has recently been further analyzed by Pagan and Robinson

(2019) and Canova and Ferroni (2019). We show that this is an issue for two well-known DSGE models.

Section 5 turns to the question of what happens if the straitjacket doesn't fit. An alternative to formulating a new model is to try to adjust the existing model. An approach often seen is to allow for stochastically time-varying behaviour in the coefficients of the VAR - i.e. a TVP-VAR - or to allow for regime switching, as with Markov Switching models. Basically, this involves adding extra shocks to the model that are associated with the coefficients. This makes it virtually inevitable that we end up with excess shocks. The consequences of this are the same as noted above. It is shown that one would now get an identity connecting the model and coefficient shocks, but with time-varying weights so that at any point in time the identity holds. This has the consequence that one cannot vary the model shocks without affecting the other shocks; this will affect computation and the validity of impulse responses.

Section 6 considers when the theoretical restrictions entailed by the DSGE straitjacket may be useful. Obviously such models are important for clear thinking about relationships in the economy, and are often good for thinking about policy issues, although one needs to keep them relatively small for that purpose. DSGE models are not alone in this as the need to keep the story simple was often given as a rationale for thinking in terms of the IS-LM model. DSGE models can, however, have other uses. Pesaran and Smith argued that any co-integrating information evident in the data needed to be used better. An important advantage of the DSGE model is that it often implies co-integration, and so can be made to yield a variety of co-integrating vectors. In addition, DSGE models provide an argument for why the integration and co-integration seen in data exists, namely that it is due to permanent shocks which are latent in the model. This feature has important implications for modelling. One is that any VECM model in just observables will fail to capture the error correction terms involving a latent  $I(1)$  variable, and this mis-specification may affect impulse response estimates. Pagan and Robinson (2019) argued that one should add an  $I(1)$  latent variable into observed VECMs to enable the missing error correction terms to be captured. No economic assumption is involved here; it is just a statistical feature. This is equally true of DSGE models, where shocks are presumed to have an AR structure. Pagan and Robinson (2018) showed that this gave far better estimates of impulse responses. A by-product of this same feature is that it points to problems in estimating SVARs with the

levels of  $I(1)$  variables rather than their growth rates (as in DSGE models), which is something frequently done in the literature. Section 7 concludes.

## 2 A Measure of Fit

Is there just one measure of fit? The answer to this from a frequentist perspective would most clearly have been no; the likelihood could be maximized by fitting more and more parameters so that wasn't a satisfactory index. Adjustments could be made to it to reflect the number of parameters being estimated, and many penalty functions have been proposed. Today, Bayesian estimation is the most common estimation method used for the parameters of DSGE models. In reading these papers, one sometimes gets the impression that there is just one criterion to judge models, namely, the marginal data density (MDD), as increasingly one sees just this being reported. Thus Christiano et al (2014) tabulate it for various DSGE models in which different shocks have been suppressed, as do den Haan and Drechsel (2018). The problem with the MDD serving as a single measure is that it was conceived of as a way of *comparing models*. The basis of this test is the Bayes factor, which say what the odds are for one model compared to another. It does not really address whether any of the models fit the data. One needs to look at that question first. Conn et al (2017) look at what is available for checking Bayesian model fit when estimating ecological models. They say (p. 527) that "The implicit requirement that one conduct model checking exercises is not often adhered to when reporting results of Bayesian analysis", and (p.527) that "Perhaps there is a mistaken belief among authors and reviewers that convergence to a stationary distribution, combined with a lack of prior sensitivity, implies that a model fits the data". This seems to be very descriptive of a lot of modern Bayesian work with macroeconomic models. In a Bayesian context a model is essentially a set of structural relations which involve parameters, some auxiliary statistical assumptions needed to form a likelihood, and some priors. It may be that any if one of these is incorrect this could cause a very poor model fit.

Conn et al then give an answer to the question posed in this section by tabulating a variety of methods that Bayesians could use to check models. One is prior predictive tests. This compares the value of some discrepancy measure obtained from data simulated from the structural model - where parameter values are drawn from the prior distribution - with the value of

the measure obtained directly from observed data. This is probably not a good measure of fit but it is useful to have some idea of what the prior implies in terms of some measure that is informative. More focussed on fit would be *posterior predictive tests* where the prior is replaced in the computation by the posterior. One problem that can arise when comparing the discrepancy measures found from simulations of the DSGE model and the data is that the test is often conservative, i.e. tends to accept the model since (p. 520) "data are used twice: once to approximate the posterior distribution and to simulate the reference distribution for the discrepancy measure, and a second time to calculate the tail probability". This problem is easily solved in the frequentist case by formulating the discrepancy function as a moment condition and then adjoining the moment conditions for the scores and the p-value. This, of course, leaves open what sort of discrepancy measures might be used. We now turn to this issue.

### 3 General and Specific Discrepancy Measures

#### 3.1 General Tests : Moments Matter

A test quantity, or discrepancy measure,  $D(y, \theta)$  is a quantity dependent on variables and model parameters that can be used to compare actual data with the model-implied values found from simulations. A p-value for any test statistic can be formed from the number of values of  $D$  from simulations that are less than the  $D$  computed from the data. Perhaps the simplest discrepancy measures are those for the moments of selected variables e.g. GDP growth, inflation. Because many DSGE models work with mean-corrected data it will only be the variance and higher moments that provide measures. It may well be that the model fails to fit the first moment but this is hidden.

To give an example we look at the model in Christiano et al.(2014) (CMR) Table 1 shows the standard deviations of variables from their fitted model together with those of the data. Because the key result in CMR was that risk shocks were crucially important for outcomes in the macro economy, we also show the standard deviations when all risk shocks are suppressed. The standard deviations are in %. The p-values are found by simulating the CMR model, that is the DSGE model is assumed to be the DGP.



Variable	Model, All Shocks	Model, no risk	Data (p value)
gdp	1.22	.836	.7 (0.00)
cons	.925	.653	.5 (0.0)
inv	3.796	2.073	2.1 (0.00)
credit	2.043	1.086	1.2 (.001)
premium	.319	.096	.17 (.005)
spread	1.085	.697	.38 (0.0)

There are large discrepancies between the data and the fitted model. Taking out the risk shock results in a much closer match. If a model that includes risk shocks fails to match the data by a large margin, while one that doesn't have them gives a far better match, this must cast some doubt on the importance of risks modelled like this in actual economies. Note there are other moments such as skewness that the CMR model fails. This seems to be a common issue with models featuring risk shocks. The latest version of the ECB's New Area Wide Model - Coenen et al (2018) - has the same feature of excess volatility, i.e. the variance of the variables exceed what is in the data, although not by as much as in CMR.

### 3.2 Specific Measures

These discrepancy measures are targeted in such a way as to have particular appeal. For example, if one were interested in business or financial cycles then it can be instructive to look at items such as the number of booms/busts, bear and bull markets in the actual and simulated data using well-defined rules for these events. An example of such a rule would be finding the location of turning points in data defining expansions and recessions in a similar way to that which underlies the NBER business cycle dates, as was done in Pagan and Robinson (2014). If one does not wish to use complex rules then even simple ones such as the number of negative growth rates in the data versus that implied by the model can be extremely informative. Applying that test to the CMR model we find that it produces *three times as many negative growth rates* for per capita GDP as seen in the data.

## 4 System Wide Discrepancy Measures

Early on, fitted DSGE models were often checked by asking if they could match a VAR fitted to data. This required the constructed DSGE model to capture both the VAR dynamics and the covariance matrix of its residuals. This is rarely done today, but seems an obvious systems test. Hence, we turn to investigating its use below. Before doing so we look at using some measures that either test a crucial assumption in the model or which are phrased in such a way as to appeal to policy makers.

### 4.1 Targeting Component Parts of the System

These aim to focus upon some aspects of the system. In recent Bayesian work priors have often been constructed that focus not on the parameters of a model but on outcomes of it which are combinations of the parameters and variables. Christiano et al (2011) called these *endogenous priors*. A better description would be that by Andrieu and Beněs (2013), who termed them *system priors*. Using a sacrifice ratio to check the model is one example. A particularly important system measure would seem to be whether the model-consistent expectations embedded in many DSGE models agree with the data. This discrepancy measure would (say) take what the model says unanticipated inflation is and ask whether it really is independent of the information set used in the model. Such a test is possible because DSGE models are over-identified, often by a large degree. Applying this test to the Ireland (2004) model discussed later we regress the expected value of inflation one period ahead, obtained from the estimated model, against inflation and lagged inflation. With HAC standard errors the p-value of the test that there is no explanation is .06, so a marginal rejection. However, if we ask about the expected value of the growth in GDP one period ahead, the regression of the unanticipated component on current GDP growth and inflation gives an F test of 36 (using HAC standard errors) which has a p-value of zero. This anticipates results concerning the dynamics in this model given later.

A well-known issue in Bayesian estimation concerns the relative contributions of the prior and the data. This is particularly important because there seems to be an increasing tendency to use "priors" simply as a way of penalizing the likelihood of the data by trying many priors before producing the quoted results. The situation is reminiscent of the common frequentist practice of performing a large number of statistical tests, something that was

criticized by Bayesians. Full disclosure of what is being done might enable the reader to form some idea of the influence of the prior, but this is rarely reported.

As has been observed by Koop et. al. (2013), it is not sufficient to just compare prior and posterior. Priors can greatly influence the interpretation of the model, including its dynamics. For example, Del Negro and Schorfheide (2008) find that time series data cannot distinguish between price and wage rigidities in their New Keynesian model. They use priors involving different degrees of rigidity of prices and wages. An interesting feature of their results which is not commented on is that, when the prior on wages imposes high rigidity, the disturbances in the wage equation are highly autocorrelated, but when the prior imposes low rigidity the disturbances have low autocorrelation. Clearly the data is saying that wages are sticky while the choice of priors is determining where this shows up in the estimates and, hence, in the interpretation of what the stickiness is due to: structural factors in wage determination or persistent shocks? This illustrates a possible effect of the straightjacket imposed by the priors and what a more flexible approach to the dynamics might reveal.

## 4.2 Targeting VAR Dynamics

As mentioned earlier an obvious systems test involves checking the VAR found from simulating the model with that from the data. It may be that the solution to the DSGE model is not a VAR but a VARMA process, so the VAR is an incorrect representation. But that does not prohibit a test under the assumptions of the DSGE model, i.e. when the model is taken to be the null hypothesis. The situation is like indirect estimation. There an auxiliary model is used to fit the data. The mapping between that and the DSGE model parameters can enable one to recover consistent estimates of the latter. There are conditions in the estimation theory about whether the mapping is invertible, but no such constraint operates if all we want to do is test a DSGE model. Canova (1994) suggested this and Meenagh et al. (2019) have presented extensive analysis of its use, particularly in the context of the Smets and Wouters (2007) model, which they reject.

We apply the test to Ireland's (2004) model. This involves using his parameter estimates to simulate a long set of data from the model, fitting a VAR(1) to the three variables used in estimation - GDP growth, inflation and an interest rate - and then comparing it to the same VAR(1) when

estimated from the data. It is immediately apparent that the GDP growth rate VAR equation is very different when based on the data from that implied by Ireland's model. The lagged inflation rate in this equation has a value in the large set of simulated data of  $-.776$  (basically the pseudo-true value) versus  $.06$  when using actual data. One finds that the probability of getting a value like  $.06$  from simulations of the VAR implied by the DSGE model is zero, so the model is emphatically rejected as failing to match the dynamics seen in the data. Perhaps the story underlying the model is good but, as Martin Luther King once said, "They can talk the talk, but can they walk the walk".

The CMR model was also investigated along these lines. Here we have an issue of there being twelve observables, so any VAR in these has a large number of parameters. One possible test of whether the dynamics in the data are being captured properly by the CMR model is to look at the eigenvalues of the VAR(1) matrix. One could look at the maximum or (say) the ratio of the maximum to the 9th largest to capture the diversity. Here we used the sum of the modulus of the eigenvalues. This is  $4.72$  in the data with a p-value of  $.3$ . Other possible ways of using eigenvalues mentioned above give the same result. So from this viewpoint the CMR model seems to give a good match to the dynamics in the data.

### **4.3 Targeting the VAR Covariance Matrix: Agnostic Shocks**

In addition to looking at a model's ability to replicate the dynamics found in the data we can examine its ability to replicate the covariance matrix of the VAR. Here we find that the log of the determinant of the innovations using the data is  $-135$ , and this has a p-value of zero when compared to possible simulated model outcomes, which gives an emphatic rejection of the CMR model. We might have expected this, given its failure to match the standard deviations of the variables. However, because the latter depend upon both the dynamics and the covariance matrix of the innovations, it is useful to know exactly where the mismatch occurs..

A recent proposal has been to evaluate DSGE models by introducing into them what are called "*agnostic shocks*", and then seeing if this looser structure has a better fit. Basically, the method involves replacing some model shocks with agnostic shocks in such a way as to eliminate some of

the restrictions imposed by the DSGE model. The test looks at whether eliminating these restrictions proves significant. To illustrate this we consider the standard three variable New Keynesian model in  $\pi_t, y_t$  and  $r_t$ , where  $y_t$  is log output,  $\pi_t$  is inflation and  $r_t$  is an interest rate. Ireland (2004) derived such a model using theory to obtain parameter restrictions. In the following we have initially departed from Ireland's original derivation by setting one of his four model shocks - that for preferences - to zero, leaving only three shocks - technology, the mark up and a monetary shock. We also assume initially that technology is a stationary AR(1) process and not a unit root process as he does. We return to these assumptions later. All data were mean-corrected in estimation and so steady-state terms are omitted.

With these modifications Ireland's model is

$$\pi_t = \delta E_t(\pi_{t+1}) + \psi y_t - \psi z_t + e_t \quad (1)$$

$$y_t = E_t(y_{t+1}) - (r_t - E_t\pi_{t+1}) + (1 - \rho_z)z_t \quad (2)$$

$$r_t = \rho_\pi \pi_t + \rho_x y_t - \rho_x z_t + \varepsilon_t^m \quad (3)$$

$$z_t = \rho_z z_{t-1} + \varepsilon_t^z$$

$$e_t = \rho_e e_{t-1} + \varepsilon_t^e,$$

where  $z_t$  is the technology shock,  $e_t$  is the mark-up shock and  $\varepsilon_t$  is the monetary policy shock.

Letting  $\zeta_t = \begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix}$  this has the form

$$G\zeta_t = ME_t(\zeta_{t+1}) + F\xi_t, \quad (4)$$

where  $\xi_t = \begin{bmatrix} e_t \\ z_t \\ \varepsilon_t^m \end{bmatrix}$ . The shocks are assumed to follow the VAR form

$$\xi_t = \Phi\xi_{t-1} + \varepsilon_t, \quad (5)$$

where  $\varepsilon_t = \begin{bmatrix} \varepsilon_t^e \\ \varepsilon_t^z \\ \varepsilon_t^m \end{bmatrix}$ . Hence from (4)

$$\begin{aligned} \zeta_t &= G^{-1}ME_t(\zeta_{t+1}) + G^{-1}F\xi_t \\ &= \Gamma E_t(\zeta_{t+1}) + H\xi_t \\ \therefore \zeta_t &= \Gamma^j E_t(\zeta_{t+j}) + H + \Gamma H\Phi + \Gamma^2 H\Phi^2 + \dots + \Gamma^{j-1} H\Phi^{j-1} \xi_t. \end{aligned}$$

If the eigenvalues of  $\Gamma$  and  $\Phi$  are less than unity - the usual situation in such NK models - this becomes

$$\zeta_t = P\xi_t.$$

Then from (5)

$$\begin{aligned} P^{-1}\zeta_t &= \Phi P^{-1}\zeta_{t-1} + \varepsilon_t \\ \zeta_t &= P\Phi P^{-1}\zeta_{t-1} + P\varepsilon_t \end{aligned}$$

is a VAR(1) with errors  $v_t = P\varepsilon_t$ . This solution method of undetermined coefficients is the type of approach used in Binder and Pesaran (1995).

It is now clear that, potentially, the VAR(1) dynamics contain all of the DSGE parameters, except the standard deviations of  $\varepsilon_t$ . This means that the DSGE model parameters should be identifiable from a VAR in  $\zeta_t$  since the covariance matrix of  $v_t = P\varepsilon_t$  has six elements in it, and we only have 3 standard deviations of shocks to estimate.

Now consider replacing the technology shock with an agnostic shock  $\eta_t$ . To see the impact of this we need to express the system (1)-(3) in terms of the observable variables  $y_t, \pi_t$  and  $r_t$  and then replace the technology shock with the agnostic shock. Since the former appears in all equations this results in:<sup>1</sup>

$$\begin{aligned} \pi_t &= \beta E_t(\pi_{t+1}) + \psi y_t + \phi_1 \eta_{1t} + e_t \\ y_t &= E_t(y_{t+1}) - (r_t - E_t \pi_{t+1}) + \phi_2 \eta_{1t} \\ r_t &= \rho_\pi \pi_t + \rho_x y_t + \phi_3 \eta_{1t} + \varepsilon_t^m \\ \eta_{1t} &= \rho_\eta \eta_{1,t-1} + \varepsilon_t^\eta \\ e_t &= \rho_e e_{t-1} + \varepsilon_t^e. \end{aligned}$$

In the Den Haan and Drechsel paper the agnostic innovation  $\varepsilon_t^\eta$  seems to have a unit variance. As before, this system has a VAR(1) solution but now there are coefficients in the  $F$  matrix of (4) - specifically the  $\phi_j$  - that do not appear in the  $G$  matrix above, and hence in the dynamics. Call the errors for the new VAR equations  $\tilde{v}_t$ . The parameters  $\phi_j$  appear only in the  $cov(\tilde{v}_t)$ . This has six elements and so we can identify the two standard errors

---

<sup>1</sup>Den Haan and Drechsel say in footnote 31 "... one could choose to leave the agnostic disturbance out of some equations". Later they say "Imposing such restrictions moves us away from being fully agnostic, but there may be cases where this flexibility of the restricted formulation is very useful".

for the DSGE shocks as well as the three  $\phi_j$ . All parameters can therefore be estimated. They would produce an estimated  $\tilde{V} = cov(\tilde{v})$  which can be compared to  $\hat{V} = cov(v)$  from the DSGE model. Introducing agnostic shocks therefore provides a way of performing such a test and may be based on the difference in the likelihoods.

It is not entirely clear how one proceeds if a model shock appears in only a single structural equation. The reason is that replacing  $e_t$  with  $\phi_4\eta_{2t}$  imposes no restrictions since  $var(\eta_{2t}) = 1$  and so  $\phi_4 = std(e_t)$ . One might add  $\eta_{2t}$  to all three equations, in which case three extra parameters would need to be estimated, making a total of six  $\phi_j$ 's and one standard deviation for  $\varepsilon_t^m$ . This would give more parameters than there are elements in the VAR covariance matrix making the test impossible. If we only put  $\eta_{2t}$  into two equations  $\tilde{V}$  would just be the unrestricted VAR covariance matrix. It might therefore be better just to test an unrestricted covariance VAR matrix against that implied by the model which we do later.

Returning to Ireland's original model, consider what happens when the technology shock has a unit root, i.e.  $\rho_z = 1$ . As before, the model is set up in terms of an "output gap"  $\tilde{y}_t = y_t - z_t$  and would then be

$$\pi_t = \delta E_t(\pi_{t+1}) + \psi \tilde{y}_t + e_t \quad (6)$$

$$\tilde{y}_t = E_t(\tilde{y}_{t+1}) - (r_t - E_t\pi_{t+1}) \quad (7)$$

$$r_t = \rho_\pi \pi_t + \rho_x \tilde{y}_t + \varepsilon_t^m \quad (8)$$

$$z_t = z_{t-1} + \varepsilon_t^z$$

$$e_t = \rho_e e_{t-1} + \varepsilon_t^e.$$

The observables are now  $\Delta y_t, \pi_t$  and  $r_t$ . We need therefore to find the system representation in terms of these three variables in order to decide how agnostic shocks are to be introduced and what their impact is.

Because the solution to the system in  $\tilde{y}_t, \pi_t, r_t$  is a VAR(1) it follows that

$$E_t \tilde{y}_{t+1} = a_1 \tilde{y}_t + a_2 \pi_t + \alpha_3 r_t \quad (9)$$

$$E_t \pi_{t+1} = b_1 \tilde{y}_t + b_2 \pi_t + b_3 r_t \quad (10)$$

Substituting these into (7) and re-arranging we get

$$\tilde{y}_t = c_1 \pi_t + c_2 r_t \quad (11)$$

As  $\Delta \tilde{y}_t = \Delta y_t + \varepsilon_t^z$  we obtain

$$\Delta y_t = c_1 \Delta \pi_t + c_2 \Delta r_t - \varepsilon_t^z.$$

Given equations (11), (10), (6) and (8)

$$\begin{aligned}\pi_t &= d_1 r_t + d_2 e_t \\ r_t &= f_1 \pi_t + f_2 \varepsilon_t^m,\end{aligned}$$

so replacing the technology shock  $e_t^z$  with an agnostic one has no effect as it only enters the structural equation for  $\Delta y_t$ .<sup>2</sup> This is an instance where using indirect inference on the covariance matrix of the VAR seems to be a good way to assess the fit of the DSGE model.

## 5 Is the Staitjacket Fit for Purpose?

Most DSGE models are heavily over-identified. The CMR model has 348 moment restrictions but only 36 parameters are estimated. Therefore not all moment conditions *may be* satisfied. Some of these moment conditions are that the shocks are uncorrelated and that the innovations in the shocks have no serial correlation. Liu et al (2018) found a similar situation in a larger DSGE model - the Multi Sector Model of Rees et.al. (2016).<sup>3</sup> Table 2 shows that the CMR model also has correlations between the innovations of a number of its empirical shocks. Moreover, some of the "innovations" are far from being white noise. Given that DSGE models are commonly used to provide impulse responses to uncorrelated shocks this raises the issue of the fitness of the model to do so and how it might be achieved.

Table 2	
<i>Correlation Between Shocks</i>	<i>Value</i>
Stationary tech, govt expend	.55
$I(1)$ tech growth, stationary tech	-.5
$I(1)$ tech growth, govt expend	-.39
<i>AR(1) Coefficient for Innovations</i>	<i>Value</i>
investment Good shock	.44
Composite commodity shock	-.3

<sup>2</sup>This remains true even if  $\tilde{y}_{t-1}$ ,  $\pi_{t-1}$  and  $r_{t-1}$  appear in the NK equations of Ireland's model.

<sup>3</sup>As they note even if there is exact identification the use of Bayesian rather than MLE estimates will mean that the shocks will be correlated.



Many DSGE models have *more shocks than observables*, e.g. the CMR model has 12 observables and 20 shocks, while the Ireland (2004) NK model has four shocks - preference, technology, mark-up and monetary shocks - but only three observable variables - GDP growth, inflation and an interest rate. In the above analysis we dropped the preference shock. In such situations empirical shocks *must be correlated* - see Ravenna (2007), Pagan and Robinson (2019) and Canova and Ferroni (2019). For Ireland's model we find a correlation between the mark-up and technology innovations of 0.7, implying that one cannot separate them, and that the variance decomposition Ireland performs is invalid. Because the preference and the monetary shocks seem to be uncorrelated with the other shocks, presumably we can find the fraction of the variance attributable to them. Suppose that another observable variable is added so the model now has four variables to match the four shocks. The obvious variable to add is per capita hours. Ireland's equation (6) shows that the extra relation in the model is  $h_t = \tilde{y}_t$ , where  $h_t$  is log of hours. We now find that the cost push and technology shocks are uncorrelated. It follows that the techniques that work when there are at least as many observables as shocks fail to apply when this is not the case. A significant qualification therefore seems to be needed for simple estimated NK models which use only three observed variables. In order to be able to separate mark-up and technology shocks information about per capita hours should be exploited.

## 6 Can we Modify the Straitjacket, not the Model?

It is tempting to think that one might be able to modify the straitjacket to get a better fit with the data. A little tuck here and a little tuck there might do the job. Two of these strategies seem to be in evidence. One is to say that the DSGE model should not be expected to explain the data because it has "measurement error". Fundamentally, the problem in DSGE models is that there are no residuals to assess the fit of the model to the data.

There are two philosophies to deal with this. One is "theory ahead of data", which maintains we shouldn't be surprised if the DSGE model fails to fit, and any such gap is of no interest. The other is that these measurement errors play the role of residuals and so shed light on how useful the model is for providing an account of what is observed. If one estimates the coefficients

using the Bayes posterior mode (or MLE) then the computed filtered shocks from the DSGE model must add up to the observed data. Any difficulties with the model therefore just carry over into computed shocks. Introducing measurement errors into DSGE models means that the model shocks no longer fully explain the data. Basically, they reflect the gap between what the model says the variables would be and what the data says. If the total number of model plus measurement error shocks are no more than the number of observable variables then the measurement error shocks represent residuals. However, the standard way in which measurement errors have been introduced produces more shocks than observables, which means that the shocks have to be correlated. As the shocks are no longer independent it is now unclear what a residual is. There are many models in the literature with measurement errors like this, e.g., the New Area Wide Model II (Coenen et al. 2018).

Another tuck that has been applied is to allow the model to have time-varying parameters (TVP). They are generally assumed to evolve as a unit root process. To see what problems may arise with this strategy consider the following TVP model which captures the most common approach. For simplicity we just have a single time-varying coefficient,  $\alpha_t$ .

$$\begin{aligned} z_t &= b_t \alpha_t + e_t \\ \alpha_t &= \alpha_{t-1} + v_t \\ std(v_t) &= \sigma_v, std(e_t) = \sigma_e, b_t = z_{t-1}. \end{aligned}$$

A first difficulty with this model is that the process for  $z_t$  is not stationary. If the sample size is long enough then  $z_t$  will always explode. By making  $\sigma_v$  small this explosive property can be delayed. With Bayesian estimation the prior on  $\sigma_v$  is often very tight. This is most likely the reason that  $z_t$  is not found to explode. But is that really prior information or is it simply a device to prevent the inevitable consequences of the model for  $z_t$  emerging? It would be far better to change the model and to make  $\alpha_t$  a stationary process. In that case we can allow more variation in the coefficients.

Ignoring this difficulty with the typical specification used in the literature, we look at another issue, namely, it is effectively adding extra shocks to the system and so works like measurement error. One would expect therefore that, when there are excess shocks present, they will end up being correlated. To see this more formally cast the above model into state-space form, as

follows:

$$\begin{aligned}
\psi_t &= M\psi_{t-1} + Cu_t \\
z_t &= D_t\psi_t \\
u'_t &= [v_t \ e_t], \psi_{1t} = \alpha_t, \psi_{2t} = e_t \\
M &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, D_t = [z_{t-1} \ 1], C = \begin{bmatrix} \sigma_v & 0 \\ 0 & \sigma_e \end{bmatrix}.
\end{aligned}$$

Because  $D_t = z_{t-1}$  is fixed at time  $t$  it can be conditioned on when deriving the Kalman filter whose equations will be

$$E_t\psi_t = \Phi_t E_{t-1}\psi_{t-1} + K_t z_t \quad (12)$$

$$K_t = [MP_{t-1|t-1}\Psi'_t + CC'D'_t][\Psi_t P_{t-1|t-1}\Psi'_t + \Lambda_t \Lambda'_t]^{-1} \quad (13)$$

$$P_{t|t} = P_{t|t-1} - K_t[\Psi_t P_{t-1|t-1}\Psi'_t + \Lambda_t \Lambda'_t]K'_t$$

$$P_{t+1|t} = MP_{t|t}M' + CC'$$

$$\Psi_t = D_t M, \Lambda_t = D_t C, \Phi_t = M - K_t \Psi_t, \quad (14)$$

where  $K_t$  is the gain of the Kalman filter.

Due to the form of  $M$  we find that  $\Phi_t = \begin{bmatrix} \phi_{11,t} & 0 \\ \phi_{21,t} & 0 \end{bmatrix}$  and so

$$E_t\psi_{1t} = \phi_{11,t}E_t\psi_{1t-1} + K_{1t}z_t$$

$$E_t\psi_{2t} = \phi_{21,t}E_t\psi_{1t-1} + K_{2t}z_t.$$

Hence, at any point in time  $t$  there is a relation

$$b(t)E_t\psi_{1t} + E_t\psi_{2t} = a(t)z_t,$$

making the two shocks perfectly correlated at that point. In fact,  $a(t) = 1$  in this case. Consequently, any impulse responses to  $e_t$  computed at a particular  $t$  must also involve movement in  $E_t\alpha_t$ . What, therefore, are the impulse responses a response to? Is it a shock to  $e_t$  or to  $v_t$ ? To summarize, the introduction of a stochastic process for the time variation in the coefficients introduces excess shocks and so makes it hard to assess impulse responses.

## 7 When is a Straitjacket a Useful Thing?

The purpose of a straitjacket is to constrain the patient, i.e. it enforces some discipline. We have identified problems that might arise in the analysis and

use of DSGE models. They also have some attractive features that arise from the straightjacket they provide, such as the ruling out of free lunches. A possible danger with models that are more loosely based in theory is that they may not do that. A classic example arises with an SVAR model of the form

$$\begin{aligned}\Delta y_t &= a_{11}^1 \Delta y_{t-1} + a_{12}^1 \pi_{t-1} + a_{12}^1 \pi_{t-1} + \varepsilon_{1t} \\ \pi_t &= a_{21}^0 \Delta y_t + a_{21}^1 \Delta y_{t-1} + a_{22}^1 \pi_{t-1} + a_{23}^1 r_{t-1} + \varepsilon_{2t} \\ r_t &= a_{31}^0 \Delta y_t + a_{32}^0 \pi_t + a_{31}^1 \Delta y_{t-1} + a_{32}^1 \pi_{t-1} + a_{33}^1 r_{t-1} + \varepsilon_{3t}.\end{aligned}$$

The monetary shock  $\varepsilon_{3t}$  in this model has a permanent effect on the level of output  $y_t$ . This does not seem to be a reasonable outcome; it would seem better to have a zero effect of the monetary shock on the long-run level of output. To achieve this it is necessary to ensure that the cumulated impulses of  $\Delta y_t$  to  $\varepsilon_{3t}$  tend to zero by imposing this long-run restriction. A DSGE model always does this as it specifies that the only shock that can affect output in the long run is the technology shock. Thereby it eliminates the free lunch. SVARs generally involve loose theory, but this should not mean no theory.

A DSGE model also forces one to think seriously about how to handle non-stationary data. A trend can be either deterministic or stochastic. In the latter case it is necessary to have a shock in the model that has a unit root. This shock is then used to render  $I(1)$  variables stationary. Thereafter the model in  $I(0)$  variables can be solved and estimated. The observable data will often be growth rates in the  $I(1)$  variables. To avoid issues arising from the variables not having a common deterministic trend it is standard practice in DSGE model estimation to de-mean the growth rates. This implies that the DSGE model prediction of co-trending behaviour is not correct but that analysis will proceed with variables in which their deterministic trends have been removed.

Compare this to working with SVARs. Although it seems sensible to transform the data into growth rates as in DSGE models, often the levels of the variables are used. From a theoretical perspective this seems odd. For example, it is hard to conceive of interest rate decisions made by looking at the level of GDP, which is why one mostly sees an output gap in that equation. In Ireland's model both  $y_t - z_t$  and  $\Delta y_t$  were present.

There are also more serious difficulties in working with SVARs using levels of variables rather than growth rates. To see this we look at a DSGE model

used by Ravenna (2007) and Poskitt and Yao (2017). This is a two variable RBC model with GDP growth ( $\Delta y_t$ ) and hours ( $h_t$ ). The labour-augmenting technology shock  $z_t$  is a unit root process and the second shock is a stationary labor supply shock. These have innovations  $\varepsilon_t^z$  and  $\varepsilon_t^h$  respectively. The log level of output  $y_t$  is made stationary by forming  $\tilde{y}_t = y_t - z_t$ . The solution for the variables  $\tilde{y}_t$  and  $h_t$  is then a VAR(1)

$$\begin{aligned}\Delta \tilde{y}_t &= b_{11}^1 \tilde{y}_{t-1} + b_{12}^1 h_{t-1} + \gamma_{11} \varepsilon_{1t}^z + \gamma_{12} \varepsilon_t^h \\ h_t &= b_{21}^1 \tilde{y}_{t-1} + b_{22}^1 h_{t-1} + \gamma_{21} \varepsilon_{1t}^z + \gamma_{22} \varepsilon_t^h.\end{aligned}$$

$\tilde{y}_t$  is unobservable but

$$\Delta \tilde{y}_t = \Delta \tilde{y}_t + \varepsilon_t^z,$$

so that

$$\Delta y_t = b_{11}^1 \tilde{y}_{t-1} + b_{12}^1 h_{t-1} + (1 + \gamma_{11}) \varepsilon_{1t}^z + \gamma_{12} \varepsilon_t^h.$$

To compute the impulse responses we need estimates of  $\gamma_{ij}$  and of the VAR coefficients  $b_{ij}^1$ . A long-run restriction that  $\varepsilon_t^h$  has no impact on  $y_t$  enables the  $\gamma_{ij}$  to be estimated, as this produces an extra restriction when there are then three parameters to estimate from the covariance matrix of the VAR(1) residuals. But to get the residuals we need a consistent estimator of  $\beta_{ij}^1$ . Many papers argue that the levels of the series  $y_t$  and  $h_t$  can be used rather than  $\Delta y_t$  and  $h_t$ . Its essence is that we can write

$$y_t = \delta y_{t-1} + b_{11}^1 \tilde{y}_{t-1} + b_{12}^1 h_{t-1} + (1 + \gamma_{11}) \varepsilon_{1t}^z + \gamma_{12} \varepsilon_t^h,$$

and get a super-consistent estimator of  $\delta = 1$  as all the other variables in the system are  $I(0)$ . Hence, working with the levels  $y_t$  and  $h_t$  could be justified if all that was needed was an estimate of  $\delta$ . But, for the purpose of computing impulse responses, we also need to be able to estimate  $b_{11}^1$  and  $b_{12}^1$  consistently. The problem in doing this is that  $\tilde{y}_{t-1}$  is not an observed variable; the regression would be of  $y_t$  against  $y_{t-1}$ , and  $h_{t-1}$  and would not include  $\tilde{y}_{t-1}$ . If this is the source of the non-stationarity, we would therefore not get a correct estimate of impulse responses by working with the levels of observed data. Figure 1 shows this for the response of hours worked to a technology shock.

There are very large differences between the impulse responses based on levels data and the correct ones, even though the estimate of  $\delta = .9999$ . In this graph we have set the initial impulse response to the correct values, i.e.

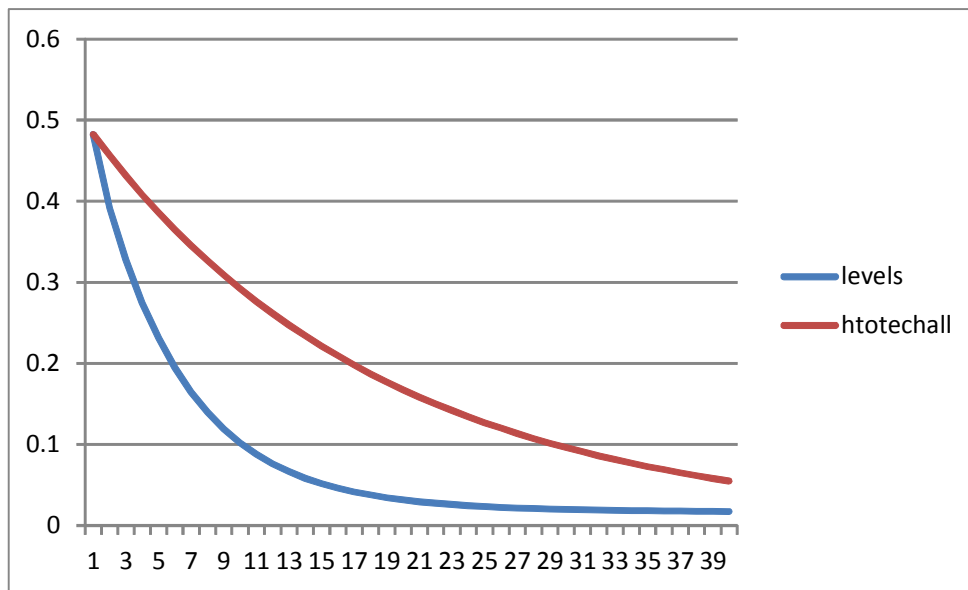


Figure 1: Comparison of the True Impulse Responses from the DSGE Model ("all") and the Responses Found from Using Variables in Levels.

it is assumed that the true  $\gamma_{ij}$  are known. This discrepancy is also true if one fitted a VAR in  $\Delta y_t$  and  $h_t$ . The purpose of the figure is to show the bias coming from using levels. Pagan and Robinson (2019) introduce a unit root latent variable  $a_t$  into the system, just as in a DSGE model, and then estimate a VECM in  $\Delta y_t$ ,  $h_t$  and  $\Delta a_t$ . They shows that accounting for the latent EC term in this VECM is very important in obtaining correct impulse responses. One might argue that the non-stationarity in  $y_t$  could come from some  $I(1)$  exogenous variable that is observed, for example foreign output in a small open economy. The DSGE model framework has the good feature of pointing to the fact that one has to propose a way of generating a unit root process for  $y_t$ ; SVAR exercises are often careless about this.

A further example is the very influential natural or "neutral" rate paper of Holston et al (2017). This consists of the following equations

$$r_t^* = g_t + z_t \quad (15)$$

$$\tilde{y}_t = \alpha_1 \tilde{y}_{t-1} + \alpha_2 \tilde{y}_{t-2} + \frac{\alpha_r}{2} \sum_{j=1}^2 (r_{t-j} - r_{t-j}^*) + \varepsilon_{1t} \quad (16)$$

$$\pi_t = \beta \pi_{t-1} + (1 - \beta) \frac{1}{3} \sum_{j=2}^4 \pi_{t-j} + b \tilde{y}_{t-1} + \varepsilon_{2t} \quad (17)$$

$$\Delta z_t = \varepsilon_{3t} \quad (18)$$

$$\Delta y_t^* = g_{t-1} + \varepsilon_{4t} \quad (19)$$

$$\Delta g_t = \varepsilon_{5t} \quad (20)$$

$$\tilde{y}_t = y_t - y_t^*. \quad (21)$$

where  $r_t^*$  is the time varying natural rate of interest,  $y_t$  is the log level of GDP,  $g_t$  is the permanent component of growth,  $z_t$  "other" influences on the natural rate,  $\pi_t$  is inflation, and  $y_t^*$  is the potential level of output. The first comment to make about this model is that there are only two observables - inflation and GDP - but there are five shocks. So the empirical shocks must be correlated. The real interest rate  $r_t$  is not treated as an endogenous observable. If one had written out a DSGE model in Dynare then of course there would be an interest rate rule since every DSGE model has to have an equation for each endogenous variable.

Substituting (15) into (16) one obtains

$$\tilde{y}_t = \alpha_1 \tilde{y}_{t-1} + \alpha_2 \tilde{y}_{t-2} + \frac{\alpha_r}{2} \sum_{j=1}^2 (r_{t-j} - g_{t-j} - z_{t-j}) + \varepsilon_{1t}.$$

Because  $\tilde{y}_t$  is being treated as  $I(0)$  under the model above this must mean that  $\sum_{j=1}^2 (r_{t-j} - g_{t-j} - z_{t-j})$  is  $I(0)$ , i.e.  $r_t - g_t - z_t$  is  $I(0)$ . From equation (15) this means that  $r_t - r_t^*$  is  $I(0)$ . Under the model  $r_t^*$  is  $I(1)$  so  $r_t$  must be  $I(1)$ . This means that both the nominal interest rate and inflation are  $I(1)$  and there is no co-integration between them. Turning to the inflation equation (17) it can be written as

$$\pi_t = \pi_{t-1} - (1 - \beta) \frac{1}{3} \sum_{j=1}^3 \Delta_j \pi_{t-1} + b \tilde{y}_{t-1} + \varepsilon_{2t}.$$

Since  $(1 - \beta) \frac{1}{3} \sum_{j=1}^3 \Delta_j \pi_{t-1} + b \tilde{y}_{t-1} + \varepsilon_{2t}$  is  $I(0)$ , inflation must be  $I(1)$ . The variable  $y_t^*$  and  $y_t$  are of course  $I(2)$ . As is well known, unless the real interest rate responds sufficiently to inflation this is an unstable economy. From a theoretical perspective, the lack of an interest rate rule in the system to ensure this doesn't happen seems an issue.

Jorda and Taylor (2019) have modified the above system in a number of ways. First, they set  $\varepsilon_t^4 = 0$ . Second, the Phillips curve and IS curves are modified, (18) and (20) are retained, and some extra equations are added:

$$\begin{aligned} \tilde{y}_t &= .65 \tilde{y}_{t-1} + \alpha_r (r_{t-1} - r_{t-1}^*) + \varepsilon_{1t} \\ \pi_t &= \beta \pi_{t-1} + (1 - \beta) \pi_{t|t-1}^* + b \tilde{y}_{t-1} + \varepsilon_{2t} \\ \Delta \pi_{t|t-1}^* &= \varepsilon_{6t} \\ i_t^b &= .65 (r_t^* + \pi_{t|t-1}^*) + .35 i_{t-1}^b + \varepsilon_{7t} \\ i_t^B &= .65 (r_t^* + \pi_{t|t-1}^* + \eta_t^B) + .35 i_{t-1}^B + \varepsilon_{8t} \\ \Delta \eta_t^B &= \varepsilon_{9t} \\ r_t &= i_t^b - \pi_{t|t-1}^* \end{aligned}$$

There are now four observables but eight shocks, so the latter must be correlated. Taking the conditional expectation of inflation from the Phillips curve and re-arranging it we get

$$\pi_{t|t-1}^* = \pi_{t-1} + \frac{b}{\beta} \tilde{y}_{t-1}$$



implying that  $\pi_t$  and  $\pi_{t|t-1}^*$  cointegrate. Because in the system above

$$r_t = .35r_{t-1} + .65r_t^* + \varepsilon_{7t}, \quad (22)$$

$r_t$  is  $I(1)$ , and there is no cointegration between  $i_t^b$  and  $\pi_{t|t-1}^*$ . This also implies that the real short term interest rate  $r_t$  doesn't have a variance, although Jorda and Taylor compute one for their "variance decomposition". Inflation doesn't get stabilized in this model either, even though there is a response of the nominal short-term bill rate to inflation, but it is not sufficient.<sup>4</sup> Finally, subtracting the bill rate from the bond rate we get

$$i_t^B - i_t^b = .65\eta_t^B + .35(i_{t-1}^B - i_{t-1}^b) + \varepsilon_{8t} - \varepsilon_{7t}$$

and, since  $\eta_t^B$  is  $I(1)$ , then  $i_t^B - i_t^b$  is  $I(1)$ , i.e. there is no co-integration between the two interest rates. These features seem inconsistent with the data. For example, the spread between the long and short rate using their data over 1955 to 2016 never has a serial correlation coefficient above .55 and has an ADF test of -6.9. Hence it is hard to see these rates as not being cointegrated if they are taken to be  $I(1)$ .

Regarding the variance decomposition they write  $r_t = r_t - r_t^* + r_t^*$  and then record the  $var(r_t - r_t^*)/var(r_t^*)$ .<sup>5</sup> We note from (22) that  $r_t - r_t^*$  is  $I(0)$  and  $r_t^*$  is  $I(1)$  so that, conditioning on the initial values,  $var(r_t^*)$  rises with  $t$  and  $var(r_t - r_t^*)$  is a constant, so the ratio goes to zero as the sample size rises i.e. eventually everything will seem to be driven by  $r_t^*$ . Furthermore,  $r_t - r_t^*$  depends only on  $\varepsilon_{7t}$ , while  $r_t^*$  depends on  $\varepsilon_{5t}$  and  $\varepsilon_{3t}$ . So  $r_t - r_t^*$  and  $r_t^*$  are not independent after estimation due to the shocks being correlated and we cannot attribute separate effects to them. Because the shocks are correlated in the data,  $r_t^*$  is not an exogenous variable, even though that is assumed in the model.

The equations for  $\tilde{y}_t$  and  $g_t$  imply that

$$\begin{aligned} \Delta^2 y_t &= \Delta^2 y_t^* + \Delta^2 \tilde{y}_t \\ &= \varepsilon_{5t} + \frac{\alpha_r(\Delta^2 r_{t-1} - \Delta^2 r_{t-1}^* + \Delta \varepsilon_{1t})}{(1 - .65L)}, \end{aligned}$$

---

<sup>4</sup>This is also true of the re-configuration offered in Fiorentini et al (2018).

<sup>5</sup>Because they deal with a number of countries there is a further term that is the difference between the country  $r_t^*$  and the average of these, called the global rate. But the point is the same.

making  $\Delta y_t$  a unit root process. There is, of course, is no serial correlation in  $\Delta y_t$  in the data. Finally, the Kalman filter will produce estimates of  $r_t^*$  by weighting  $\Delta^2 y_t, \pi_t, i_t^b$  and  $i_t^B$  and their lags. The first shows no sustained downward movements. The second has been stable since the 1990s. So the decline in  $r_t^*$  from the 1990s must come from the declines in the nominal rates. Hence the conclusion has to be that the estimated neutral real interest rate is declining because nominal interest rates have declined.

## 8 Conclusions

In this paper we have examined some potential conflicts between the straight-jacket of theory (especially as applied to DSGE models) and the empirical support for these models. Our analysis - both theoretical and empirical - provides support for Pesaran and Smith's argument that "theory, while essential, should be regarded as a flexible framework rather than a straight-jacket, because features that the theory abstracts from may be important in practice". We discuss how best to measure the overall fit of a model. Given the difficulties in providing such a measure, we consider a variety of measures that focus on specific aspects of the model, such as its moments, its business cycle properties, its dynamics, the restrictions implied for the covariance matrix of VAR representations of the model, and the shock structure of the model. We illustrate these issues using various New Keynesian models and provide new evidence on them. Our findings suggest that the Bayesian estimates of these models found in the literature fail on most of these criteria. We also find that models based loosely on theory such as many SVARs are vulnerable to producing theoretically implausible results.

All of this indicates that Pesaran and Smith are correct in that, while theory is essential, in order to match the data a more flexible framework will usually be required. The challenge is to incorporate this flexibility into the theory in such a way as to be compatible with the data. The danger is that strong priors may be used in relation to the extra parameters that the extra flexibility requires and these restrictions are being imposed on the estimates and not tested in relation to the data. In this paper we have suggested what sort of tests might be used.

## 9 References

Andrle, M. and J. Beněš (2013). "System priors: Formulating priors about DSGE model's properties", *IMF Working Paper No. 13/257*.

Binder, M. and M. H. Pesaran (1995), "Multivariate Rational Expectations Models and Macroeconomic Modelling: A Review and Some New Results", in M.H. Pesaran and M. Wickens (eds.) *Handbook of Applied Econometrics: Macroeconomics*, Basil Blackwell, Oxford.

Canova, F. (1994), "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, vol. 9(S), 123-144.

Canova, F. and F Ferroni (2019) "Mind the Gap! Stylized Dynamic Facts and Structural Models", CEPR Working Paper 13948

Conn,P.B. D. S. Johnson. P J. Williams, S. R. Melin, and M B. Hooten (2018) "A guide to Bayesian model checking for ecologists", *Ecological Monographs*, 88, 526-542.

Christiano, L.J. M. Trabandt, and K. Walentin (2011). "Introducing financial frictions and unemployment into a small open economy model", *Journal of Economic Dynamics and Control*, 35, 1999–2041.

Christiano, L.J., Motto, and M. Rostagno (2014), "Risk Shocks", *American Economic Review* 2014, 104, 27-65

Coenen,G. P. Karadi, S. Schmidt and A. Warne (2018), "The New Area-Wide Model II: an extended version of the ECB's micro-founded model for forecasting and policy analysis with a financial sector", *Working Paper No 2200*, European Central Bank

Den Haan, W.J., and T. Drechsel (2018), "Agnostic Structural Disturbances (ASDs): Detecting and Reducing Misspecification in Empirical Macroeconomic Models".Discussion Paper 1826, Centre for Macroeconomics (CFM). London School of Economics

Del Negro, M. and F. Schorfheide (2008), "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *Journal of Monetary Economics*, 55, 191-1208.

Fiorentini. G., A. Galesi. G. Pérez-Quirós and E. Sentana (2018), "The Rise and Fall of the Natural Interest Rate", *Working Paper 18-29*. Rimini Center for Economic Analysis.

Holston, K, T. Laubach, and J. Williams (2017), "Measuring the Natural Rate of Interest: International Trends and Determinants," *Journal of International Economics* 108, S39–S75.

- Ireland P.N (2004), "Technology shocks in the New Keynesian Model", *Review of Economics and Statistics*,86(4),923-936.
- Jordà,O. and A. M. Taylor (2019), "Riders on the Storm," *Federal Reserve Bank of San Francisco Working Paper 2019-20*.
- Koop, G., M.H. Pesaran and R.P. Smith (2013), "On Identification of Bayesian DSGE Models", *Journal of Business and Economic Statistics*, 31, 300-314.
- Liu, X., A. R. Pagan and T. Robinson (2018), "Critically Assessing Estimated DSGE Models: A Case Study of a Multi-sector Model", *Economic Record*, 94, 349-371.
- Meenagh, D. P. Minford, M. Wickens and Y.Xu (2019), "Testing DSGE Models by Indirect Inference: a Survey of Recent Findings", *Open Economies Review*,
- Pagan, A.R. and T. Robinson (2014) "Methods for Assessing the Impact of Financial Effects on Business Cycles in Macro-econometric Models" *Journal of Macroeconomics* , 41, 94 -106
- Pagan A and T. Robinson T (2019), "Implications of Partial Information for Applied Macroeconomic Modelling", Melbourne Institute Working Paper No. 12/19
- Pesaran, M.H. and R.P. Smith (2011), "Beyond the DSGE Straitjacket", *Manchester School*, 79, 5-16
- Poskitt, D. and Yao, W. (2017), "Vector Autoregressions and Macroeconomic Modelling: an Error Taxonomy", *Journal of Business and Economics Statistics*, 35, 3, 407-419.
- Ravenna, F. (2007), "Vector Autoregressions and Reduced Form Representations of DSGE Models", *Journal of Monetary Economics*, 54, 7, 2048-2064.
- Rees, D.M., P. Smith and J. Hall (2016), "A Multi-sector Model of the Australian Economy", *Economic Record*, 92, 298, 374-408.
- Smets, F. and R. Wouters (2007), "Shocks and Frictions in US Business Cycles: a Bayesian DSGE Approach", *American Economic Review*, 97, 586-606.