UNIVERSITY *of York*

Discussion Papers in Economics

No. 14/17

How did we get to where we are now? Reflections on 50 years of macroeconomic and financial econometrics

Michael Wickens

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

How did we get to where we are now? Reflections on 50 years of macroeconomic and financial econometrics

Michael Wickens
Cardiff Business School and University of York

Special Lecture
Money, Macro, Finance Conference, Durham
17-19 September 2014

Abstract

This lecture is about how best to evaluate economic theories in macroeconomics and finance, and the lessons that can be learned from the past use and misuse of evidence. It is argued that all macro/finance models are 'false' so should not be judged solely on the realism of their assumptions. The role of theory is to explain the data, They should therefore be judged by their ability to do this. Data mining will often improve the statistical properties of a model but it does not improve economic understanding. These propositions are illustrated with examples from the last fifty years of macro and financial econometrics.

JEL: B1, C1, E1, G1

Keywords: Theory and evidence in economics, DSGE modelling, time series modelling, asset price modelling

1. Introduction

Instead of a 'heads down' discussion of a particular piece of economics this lecture tries to give a 'heads up' view of what we have done and are doing in macroeconomics and finance and what we can infer from this. More specifically, the aim is to provide a review of how I perceive the relation between theory and evidence has evolved in macroeconomics and finance (asset pricing) over the last 50 years.

As a student at LSE in the 1960s my thinking about these things was dominated by two contradictory views on knowledge: deductive and inductive inference. Popper's lectures claimed that knowledge could only be obtained through deductive inference, i.e. through testing whether theories are false, as opposed to the logical positive view of testing whether they are true. This type of evidential support for theories was central to Popper's thinking. He viewed theory as metaphysics if it is not testable. Archibald and Lipsey (1963), economics colleagues of Popper at LSE, wrote a remarkable first-year economics text book in which they put Popper's methodology into practice by emphasising the testable predictions of each theory. The most recent update of this text, and now in its 12th edition, is Lipsey and Chrystal (2011).

My problem was that I was reading statistics. This is usually interpreted as based on inductive inference, which has been described as drawing general conclusions from specific instances. I have been wrestling with the dilemma of theory versus evidential-based knowledge ever since. Through econometrics, initially I focused on what the data had to tell us, but over the years I discovered that a satisfactory answer to this question did not lie simply in modelling the data, or in 'data-mining' but in what the data had to tell us about our economic theories of how the data were generated. This can be summed up as: "You can torture the data as you wish but what interpretation do you give to the results?"

Two particular areas have interested me. One is macroeconomics and the other is finance, particularly asset pricing. The recent financial crisis has prompted widespread criticism of our current theories in both areas. Krugman (2009), writing in the New York Times claimed that the macroeconomics of the last 30 years is spectacularly useless at best and positively harmful at worst. He asserted that we are living through the dark age of macroeconomics in which the hard-won wisdom of the ancients has been lost and that economists, as a group, have mistaken beauty clad in impressive-looking mathematics, for truth. Views of this sort are now institutionalised by the New Institute for Economic Thinking whose stated aim is to reform the undergraduate curriculum. Particular targets appear to be contemporary theories of macroeconomics and finance and the use of mathematics in economics. A return to older ways of analysis seems to be preferred by these critics, with Keynes being cited prominently and with approval. In my view these are not sensible conclusions to draw from the crisis.

The emphasis in these criticisms seems to be on the realism of the assumptions, particularly the assumptions of rational expectations and efficient markets. Friedman dismissed this type of criticism many years ago, arguing that a theory should be judged not on its assumptions, but on how accurate its predictions are. He also prized simple over complicated theories. It is worth recalling that Muth's (1960) concept of rational expectations was an optimal forecast which is obtained by using the correct (time series) model. At the time expectations were nearly always based on partial or adaptive expectations models which have the uncomfortable property of making forecast errors persistent unless the correct model is used, which in the case of AE is an ARIMA(0,1,1). No wonder rational expectations has been preferred as a forecasting device. Nonetheless, the description 'rational' is too strong; consistent expectations - i.e. consistent with the model - is a more accurate and less contentious description. For an assessment of this and other criticisms of DSGE models see Wickens (2009)

Another illustration of the drawbacks of focusing on the accuracy of assumptions was the famous capital debate between Cambridge (England) and Cambridge (Massachusetts) at the Econometric Society World Congress in 1970. Joan Robinson attacked Robert Solow for the use in growth theory of the concept of an aggregate capital stock. She was, of course, correct about the problems of aggregation, but we can now see how useful the concept of an aggregate capital stock has been in macroeconomics.

I take the view that all of our economic theories are 'wrong' and our models are misspecified. They are deliberate simplifications, but nonetheless they may be useful. Box and Draper (1987), known for their pioneering work on time series modelling, went even further arguing that "all models are wrong; the practical question is how wrong do they have to be to not be useful." If all of our models are misspecified, there is no point in testing them in a Popperian sense for their falseness. On the other hand, if a theory cannot 'explain' the data it is not likely to be of much use except as a pedagogic device. My reconciliation is to judge (compare) theories by how well (closely) they explain the data - even though they are all 'wrong'. This entails going back to the theory drawing board if the evidence doesn't support the theory rather than fixing up the model by introducing ad hoc assumptions such as improving its dynamic performance by introducing serially correlated disturbances. Although they might be interpreted as the combined effect of variables whose individual influence is too small to be included separately, when it comes to judging a theory, the error terms are better regarded as representing areas of ignorance than a device for improving the model's fit.

Why have theories at all if they are all wrong? This is where the flaw in inductive inference arises through trying to infer general laws from particular instances or evidence. The role of theory is to try to give a logical reason for the observed results; i.e. to provide that general law. We then look for ways of checking how well this theory fits the data, both old and new. This approach comfortably embraces the 'as if' approach of Friedman who prized simple theories that could explain much, and also Popper's deductive approach to knowledge, as it acknowledges the primacy of the predictive ability of theories.

The question of the truth of a theory has a long history in philosophy. Socrates was concerned with definitions and the logical flaws in his interlocators' attempts to provide them. Socrates might well have asked how to define a true theory. Aristotle rejected Plato's theories of forms - for example, the ideal of truth - and preferred to search for explanations to which now we might add 'of the data'. David Hume is well-known as an empiricist, stressing the importance of experience over introspection in the acquisition of knowledge. He was particularly interested in what observation had to tell us about causality. But he warned that just because the evidence pointed to a particular sequence of events, this may not always hold. This is consistent with Popper's view that we cannot infer the truth of a theory from the evidence. This very brief excursion into the theory of knowledge is consistent with the view that theories should be judged not by their truth but by their ability to explain the data. We will see that it also matters how these data are represented.

To a large extent the methods of analysis and the statistical procedures used have reflected the computing power available at the time. We resorted to asymptotic distribution theory because we couldn't work out the small sample distributions analytically. Increasingly these days, we just bootstrap the small sample distribution using simulation methods.[1]

The aim of this lecture is to offer support for the position that I have stated through a review of developments in macroeconometrics and financial econometrics over the last 50 years and, hence, in the process, explain how we got to where we are now. In section 2 I illustrate with a few examples how macroeconomics developed in response to empirical findings and how theoretical developments challenged the interpretation of evidence. In section 3 I discuss the challenge that time series analysis posed for macroeconometrics and whether economic theory can be ignored entirely when carrying out time series analysis of economic data. In section 4 I bring the discussion up to date by discussing the merits and problems of the DSGE approach to macro modelling. Section 5 discusses the conflict between economic theories of asset pricing and the data-based analysis prevalent in finance. I offer some conclusions in section 6. Clearly, these are very personal views, but I hope that they might clarify the eternal issue theory versus evidence.

2. Macroeconomic models

Macroeconometric modelling is vastly different now from what it was 50 years ago. In large part this is due to the empirical failures of the models, not the plausibility of their assumptions. This goes back to the primitive two-equation income expenditure model of the 1940s in which the government expenditure multiplier is $\frac{1}{1-mpc}$ which led to the development by Laurence Klein of large-scale macroeconometric models (for which he was awarded the Nobel Prize in 1980) and subsequently to the Brookings model, and to project Link which aimed to combine national models into a world model. The philosophy behind Klein's models was "if the model is misspecified then add another equation". This led to models with hundreds of equations. In the early 1970's, when economists first became interested in using these macro models for optimal control, emminent control theorists asked why economists required such large models when they had guided rockets to the moon using only very small models.

The philosophy behind the Brookings model was to distribute the equations to different researchers and then assemble them in a single model, see Duesenbury et al. (1964). No-one knew what the properties of the final model would be as the equations were not constructed on a common methodology. This is also true of the increasingly popular agent-based models, which aim to combine macroeconomic models with models of other aspects of the economy such as demography, the environment and behavioural finance. Over time failing components of these macroeconometric models were replaced by new equations, constructed on a new basis. Even today models of this sort are used by governments to provide forecasts and for policy analysis.

This procrustean approach to macroeconomic modellling was challenged by a number of undermining analyses that achieved considerable prominence and led to the increased use of small-scale models. One of the first such studies was by Anderson and Jordan (1968) who found that in a simple regression of output on distributed lags of the money supply and government expenditures, it was money that was significant and not government expenditures. This was a direct challenge to Keynesian-style fiscal policy and ushered in monetarism. The focus then changed to finding a stable money demand function so that monetary policy could be used to control nominal GDP, see Hendry and Ericsson (1991). This raised the problem of which measure of money should be used: narrow or broad money? If the latter, how to control it? Also, given broad money is largely credit, which interest rate(s) to use as measures of opportunity cost?

In the UK monetarism was treated in the 1970s as an irrelevant foreign idea. This overlooked the fact that with the breakdown of the Bretton Woods system the UK had lost the nominal anchor provided by targeting the US dollar and needed to find a new one. Not only did the UK have a floating, instead of a fixed, exchange rate, it also abandoned capital controls. Theories like the Mundell-Fleming model and the monetary model of the exchange rate, which had very different implications for the effectiveness of monetary and fiscal policy compared with the old Keynesian-style models, began to dominate thinking, see Mundell (1963), Fleming (1962) and Mussa (1976, 1982). Even now, despite the Lucas critique, we find models - especially time series models - estimated through various different monetary, exchange rate and capital control regimes as though this were unimportant.

One of the most influential papers was that by Phillips (1957) which, using some innovative but odd empirical methods - see Desai (1975) - suggested that there was a stable trade-off between inflation and unemployment, and that by controlling unemployment, inflation could be also be controlled. As empirical evidence mounted against the model, new versions of the model were proposed - such as the natural rate hypothesis and the expectations augmented Phillips curve - only to be challenged by the data in their turn. The conflict between theory and evidence on the Phillips curve continues to this day with its successor, the New Keynesian inflation equation, still a source of controversy both on theoretical and empirical grounds; the Calvo model is clearly not intended to be taken literally and the fit of the model is not good, see Smith and Wickens (2007). Given its central place in inflation policy, the extremely shaky empirical foundations of the modern Phillips relation should be more of a source of concern.

As most economic decisions are inter-temporal and forward-looking, modelling expectations is an important aspect of modern macroeconomic and financial models. It is also a source of considerable current controversy. Prior to the work of Sargent and Wallace (1976), macroeconomic models were largely based on the assumption that expectations are adaptive which has the uncomfortable implication that agents make systematic errors and repeat mistakes. For example, under adaptive expectations, agents do not revise their expectations even if the government announces a policy that involves increasing the money supply more than expected. Revisions to expectations would only be made after the increase in money supply has occurred, and even then agents would react only gradually, thereby incorporating the forecasting error into their expectations. This would be helpful to government as it would be able to maintain employment above its natural level indefinitely.

It was widely believed that the rational expectations revolution had gone too far when it challenged the bedrock relation of Keynesian economics: the dependence of consumption on income. Hall (1978) showed that rational expectations life-cycle theory implied that consumption would follow a random walk. This was generally interpreted as suggesting that consumption did not depend on income, a standard result supported by innumerable econometric studies with later studies that reinforcing this result, see Davidson, Hendry, Srba and Yeo (1978). This seemed to be proof positive that the RE assumption was flawed.

What is most interesting about this conflict between theory and evidence is that both can be correct. The key to the reconcilation lies in what Hall's result actually implies which is that the expected future change in consumption is unpredictable. It doesn't say that consumption does not depend on income, but instead that unanticipated future changes in wealth, income and interest rates will have no effect on current consumption. After such changes are observed, consumption may change. The key test of this theory is not, therefore, whether or not consumption depends on income, but whether income innovations affect current consumption. There are of course many difficulties in obtaining a clean test of the theory, and the theory depends on households not being credit constrained, nonetheless, it illustrates the importance of theory in interpreting evidence. We develop this argument in the next section.

3. Time series models

If economic theory is so problematic, can we just model the data using statistical methods such as time series analysis? This would have the considerable advantage of not requiring us to learn economics, which would become more like a branch of metaphysics. Sims's (1980) Econometric Society lecture seemed to offer this possibility. He started a debate that continues to this day. He argued that standard large-scale macroeconometric models embodied incredible identifying restrictions and that for business cycle analysis, forecasting and policy analysis these restrictions are neither essential nor innocuous. He proposed an alternative style of macroeconometric model, the VAR model. Ever since VAR models have been a standard tool in macro and financial econometrics. Here I discuss both their strengths and their weaknesses as tools of economics.

In effect, a statistical model of the data is given by the formula for their conditional mean. If the joint distribution of all of the observations $\{t = 1, ..., T\}$, i.e. of $z = \{z_1, z_2, ..., z_T\}$ is $D(z; \theta) = D(z_1, ..., z_T; \theta)$ then this can be written as product of their condition distributions $D(z_1, ..., z_T; \theta) = \prod_{t=1}^{T} D(z_t|z_{t-1}, ..., z_1; \phi_t)$ where $D(z_t|z_{t-1}, ..., z_1; \phi_t)$ is the conditional distribution of z_t given $z_{t-1}, ..., z_1$ and ϕ_t is a function of θ. Recall that if the z_t were independent then $D(z_t|z_{t-1}, ..., z_1) = D(z_t)$. If we make the additional assumption that the distribution is Normal and given by $D(z; \theta) \sim N(\mu, \Sigma)$ then the mean of the conditional distribution is *linear* and can be written $E(z_t|z_{t-1}, ..., z_1) = \sum_{s=1}^{t-1} A_s z_{t-s}$. If we define the deviation of z_t from the conditional mean as $e_t = z_t - E(z_t|z_{t-1}, ..., z_1)$ then we have the VAR

$$z_t = \sum_{s=1}^{t-1} A_s z_{t-s} + e_t.$$

Thus a VAR is a generic representation of the mean of the conditional distribution of the data given its past no matter what model is assumed to have generated this distribution. But notice that this VAR has $t - 1$ lags, implying that there are more parameters to estimate than there are total observations. It is therefore impossible to estimate the parameters of this VAR. In contradiction to what Sims says, therefore, the role of economics is to provide restrictions to this generic VAR. Going one step further, we can show that, after linearisation, virtually all macroeconomic and financial econometric models - even DSGE models - can be represented as a VAR (or a VARMA) with restrictions.

To illustrate, consider the structural econometric model (SEM); later we perform a similar analysis for DSGE models. The SEM can be written

$$B(L)y_t = C(L)x_t + e_t,$$

where y_t is a vector of endogenous variables, x_t is a vector of exogenous variables generated by

$$D(L)x_t = F(L)y_{t-1} + \varepsilon_t$$

with $D_0 = I$ and $z_{t-i} = L^i z_t$. The SEM can be rewritten as the VAR

$$\begin{aligned} \begin{bmatrix} y_t \\ x_t \end{bmatrix} &= G_0^{-1}G^*(L) \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + G_0^{-1} \begin{bmatrix} e_t \\ \varepsilon_t \end{bmatrix} \\ &= A(L) \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} G(L) &= \begin{bmatrix} B(L) & -C(L) \\ -F(L)L & D(L) \end{bmatrix} \\ &= G_0 - G^*(L)L \end{aligned}$$

In general both the VAR coefficients and the covariance of the VAR disturbances are restricted. In a typical VAR analysis we do not say which variables in a VAR are endogenous and which are exogenous. All variables are treated in the same way.

VARs are often used for impulse response analysis where we are interested in the dynamic responses of the variables to particular shocks. The problem is to identify what the shocks are. Sims suggested using a Choleski decomposition to do this, but although this doesn't restrict the covariance matrix of the shocks, the ordering of the variables affects the impulse response functions. To overcome this we need to introduce some theoretical restrictions. For the SEM these are embodied in G_0.

3.1. Non-stationarity

The discovery that many macroeconomic and financial time series are non-stationary had a further dramatic effect on modelling. It also made redundant much of the previous thirty years of estimation and testing theory in econometrics, which had implicitly assumed that economic time series were stationary. It explained why all those different simultaneous equation estimators gave similar results in practice: the long-run parameters were all super-consistent and therefore relatively unaffected by misspecification due to omitting stationary variables and to misspecified dynamics which, in effect, amounts to leaving out differences in non-stationary variables.

Non-stationarity has two main implications for modelling. First, it enables us to distinguish between temporary and permanent shocks; assuming that the model is stable, the former are just the structural equation disturbances and the latter are the shocks to the non-stationary exogenous variables and transmit non-stationarity to the endogenous variables. The temporary component can be interpreted as the business cycle effect and the permanent component as the growth effect. This indicates the need to distinguish between endogenous and exogenous variables. It is interesting to recall that Adelman and Adelman (1959) in their study of the causes of the business cycle concluded that it was not due to the internal dynamics of the SEM, or to its disturbances, and must therefore be due to the exogenous variables. Later we will see this again for DSGE models why they failed to forecast the recent recession.

The second effect of non-stationarity on modelling was the introduction of cointegration into VAR models. The super-consistency of long-run parameter estimates enabled the VAR to be specified in a way that separated the long and short runs. Moreover, it seemed that this required no additional restrictions on the VAR as the maximum likelihood estimator of Johansen (1988) would produce estimates of the long-run structural coefficients. Unfortunately, this is not the case. Not only are the long-run coefficients not economically identified - though they are identified statistically - nor are the impulse response functions uniquely defined, see Wickens (1996).

The reason for this is that when all of the variables z_t are $I(1)$ and the VAR

$$A(L)z_t = e_t$$

is written as the VECM

$$\Delta z_t = Az_{t-1} + A^*(L)\Delta z_{t-1} + e_t$$

then $A = \alpha\beta'$ is not of full rank as α and β are $n \times r$ matrices, where β are r cointegrating vectors, α is the loading matrix and $\beta' z_t \sim I(0)$ defines r cointegrating relations.

For the SEM above, the long-run model is

$$B(1)y_t - C(1)x_t = \bar{e}_t, \;\; \bar{e}_t \sim I(0)$$

if $B(1)$ has been normalised, the cointegrating vector is $\beta' = \begin{bmatrix} B(1) & -C(1) \end{bmatrix}$ and so is the matrix of long-run structural coefficients.

Unfortunately the Johansen estimator doesn't give an estimate of β. Instead it gives an estimate of $\beta^* = \beta H'$ where H is an unknown linear transformation of β chosen by the program for statistical and not economic reasons. We note that if H is a non-singular $r \times r$ matrix then $A = \alpha\beta' = \alpha H^{-1} H \beta' = \alpha^* \beta^{*\prime}$ and β^* is another set of cointegrating vectors that satisfy the model. Only if $r = 1$ can we give an economic interpretation to the cointegrating vector without introducing further information.

The problem for the economist is even worse than this as the impulse response functions are not identified either. These are obtained from the vector moving average representation of the VAR

$$\Delta z_t = R(L)e_t$$

This can be re-written as

$$\begin{aligned} \Delta z_t &= R(1)e_t + R^*(L)\Delta e_t \\ &= \gamma\theta' e_t + R^*(L)\Delta e_t \end{aligned}$$

where the $n \times n$ matrix $R(1)$ is of rank $n - r$ and so can be factorized into $R(1) = \gamma\theta'$where γ and θ are $n \times (n - r)$ matrices. Another factorization is $R(1) = \gamma Q^{-1} Q \theta'$where Q is $(n - r) \times (n - r)$ matrix. If we define $\Delta\tau_t = e_t$ then τ_t is $I(1)$ and we can write

$$\Delta z_t = \gamma\theta' \Delta\tau_t + R^*(L)\Delta e_t$$

implying that

$$z_t = z_0 + \gamma\theta'\tau_t + R^*(L)e_t$$

Thus we have decomposed z_t into $n - r$ stochastic trends $z^T = z_0 + \gamma\theta'\tau_t$ which is I(1), and its cycle $R^*(L)e_t$ which is I(0). The problem is that Q is not uniquely defined and neither, therefore, are θ or τ_t.

If the exogenous variables are generated by the process

$$D(L)\Delta x_t = F(L)\Delta y_{t-1} + \varepsilon_t$$

where $D(0) = I$ then it can be shown that the stochastic trends satisfy $\tau_t = Q\varepsilon_t$ and hence are an unknown transformation of the disturbances to the processs generating the exogenous variables.

A simple solution to this lack of identification of β and τ_t is to specify which variables are endogenous and which are exogenous, see Wickens and Motto (2000). This imposes zero restrictions on the loading matrix α for all rows corresponding to the exogenous variables. Countless numbers of papers have not taken these problems into account when they carry out a VAR analysis and so their authors and readers are puzzled when they try to give an economic interpretation of the results.

Sims suggests that a VAR may be used for policy analysis. We have seen that misleading results may be obtained from the impulse response functions. It can also be shown that using a VAR to examine the effects of a change in a policy reaction function may be misleading. This is, essentially, because it falls foul of the Lucas critique as the change in policy will alter the VAR parameters. Polito and Wickens (2013) explain how to avoid this problem and how to carry out optimal policy using a VAR. In brief, first a transformation of the VAR is required that is in order to condition on the policy instrument.

3.2. Model misspecification

Econometricians often speak about "the (time series) model that generates the data" as though there were a true time series generating process to be discovered. Misspecification tests that are performed on a model search for flaws in the specification. The problem with this approach is that data can normally be represented by more than one time series model. For example, if a $VAR(n)$ is 'misspecified' by the omission of p variables then this implies the remaining variables follow a $VARMA(n-p,p)$ - which could be approximated by a high order VAR. For the $n-p$ included variables the new model is just as much the 'true' data generating process as the orginal VAR. Logically, therefore, a univariate representation of a single variable is also a 'true' representation of the data generating process of that variable. However, the errors in the new models will be mixtures of the original VAR errors. Giving an economic interpretation to the resulting impulse response functions will be nearly impossible.

The only way that we can decide which is the 'true' time series representation is through using some economic theory. If we start with an economic model this enables us to derive the implied time series model and determine which variables to include. In the case of the SEM we can derive the final form or the final equation. In the next section we show how all DSGE models can be represented as a VAR.

The answer to the question posed at the start of this section seems therefore to be no. In order to interpret time series evidence of economic data we require economic theory. We have illustrated this by considering VAR analysis as recommended by Sims. We found that this may give misleading results for the analysis of business cycles and policy. Later we consider Sims's other claim that VARs are better for forecasting than macroeconometric models.

4. DSGE models

A different response is to employ not less but a lot more economic theory than before. The outcome was the DSGE approach to macroeconomics, which marked a root and branch rethink of how to do macroeconomics. The DSGE approach should be thought of as a methodology and not a set of models to be judged by the realism of their assumptions. Instead of putting together macroeconometric models equation by equation based on partial equilibrium foundations, and only finding out later what the resulting general equilibrium properties are, the DSGE methodology starts with a general equilibrium model, however small. The key feature of the methodology is the intertemporal decision of consuming today or saving to invest, and adding to later consumption. Equilibrium is defined not as flow equilibrium in terms of consumption, but as a stock equilibrium. This solves a major difficulty that had plagued Keynesian models in which equilibrium was defined as savings equals investment, a flow equilibrium. The switch of emphasis provides a natural link between macroeconomics and finance as it necessitates in the macro model the determination of stock (asset) prices and hence returns. In the process it provides a fundamentals explanation of risk different from the relative asset-pricing models commonly used in finance which are largely data-based.

Although DSGE models have certain theoretical advantages, their empirical properties are a continuing source of contention. These problems may be traced back to the early real business cycle (RBC) models and continue in the latest DSGE models despite the considerable technical sophistication and computing power used in their estimation. The RBC model was a bold attempt to meet Friedman's requirement that models should be simple but have powerful predictions as their aim was to explain the business cycle as being due to the transmission through the economy of a single shock, a productivity shock.

The main failures of the RBC model were that consumption was too smooth compared with the data, that the labour market adjusted too quickly with real wages being too flexible and employment too stable, and the real rate of return being too highly correlated with output, see for example King, Plosser and Rebelo (1988). Most subsequent DSGE models may be interpreted as attempting to address these problems. The new benchmark for DSGE models is the Smets-Wouters model (2003, 2007) which incorporates habit persistence, sticky prices and wages, mark-up effects and nominal shocks.[2] The latest DSGE models include financial frictions and fiscal and open economy effects but very few have well-defined asset market features. In terms of size, DSGE models are therefore becoming more like the earlier vintage of macroeconometric models. The main differences are that DSGE models have many cross-equation restrictions reflecting their general equilibrium underpinnings and they have forward-looking rational expectations.

It would not be too strong to say that most econometricians were scandalised by how empirical evidence was brought to bear on RBC models through the use of calibration rather than classical estimation. Subsequently, Sargent - see Evans and Honkapohja (2005) - explained why Lucas and Prescott chose calibration in preference to maximum likelihood estimation, namely, that "likelihood ratio tests were rejecting too many good models. The idea of calibration is to ignore some of the probabilistic implications of your model, but to retain others." Current practice favours the use of Bayesian estimation. This employs less strong priors than calibration but limits the influence of the data. A concern about Bayesian estimation is that although prior information, if correct, improves the accuracy of the estimates, if the model is misspecified this may be masked. We are, therefore, confronted once again with the issue of theory versus evidence: how should we interpret such estimates of DSGE models and how well do DSGE models explain the data?

4.1. The time series representation of DSGE models

To illustrate some of the problems of bringing evidence to bear on DSGE models consider the following basic centralised growth model. The model assumes that the economy is seeking to choose consumption C_t, labour N_t and capital K_t to maximise

$$E_t \sum_{s=0}^{\infty} \beta^s U(C_{t+s})$$

where $U(C_t) = \frac{{C_t}^{1-\sigma}}{1-\sigma}$ and $\beta = \frac{1}{1+\theta}$, subject to the economy's resource constraint which is derived from the national income identity, the production function and the capital accumulation equation and is

$$A_t K_t^{\alpha} N_t^{1-\alpha} = K_{t+1} + C_t - (1-\delta)K_t$$

We assume that labour is growing at the constant rate n, implying that $N_t = (1+n)^t N_0$. and that A_t, technological change, is a random walk with drift so that $A_t = (1+\mu)^t Z_t$ with $\ln Z_t = z_t$ and $\Delta z_t = e_t \sim i.i.d(0, \omega^2)$. The rate of balanced growth is $\eta \simeq n + \frac{\mu}{1-\alpha}$. In per capita deviations from their growth paths we have $y_t = \frac{Y_t}{(1+\eta)^t N_0}$, $k_t = \frac{K_t}{(1+\eta)^t N_0}$, $c_t = \frac{C_t}{(1+\eta)^t N_0}$ and $y_t = Z_t k_t^{\alpha}$. The solution consists of two equations, the economy's resource constraint and the consumption Euler equation

$$\begin{aligned} c_t + (1+\eta)k_{t+1} - (1-\delta)k_t &= Z_t k_t^{\alpha} \\ E_t\left[\beta[(1+\eta)\frac{c_{t+1}}{c_t}]^{-\sigma}(\alpha Z_{t+1}k_{t+1}^{\alpha-1} + 1 - \delta)\right] &= 1. \end{aligned}$$

Omitting intercepts, a log-linear approximation to these two equations is

$$\begin{aligned} \ln k_{t+1} &\simeq -\frac{\theta + \eta(\sigma - \alpha - 2) + (1-\alpha)\delta}{\alpha}\ln c_t + [1 + \theta + (\sigma - 1)\eta]\ln k_t \\ &\quad + \frac{\theta + \delta + \eta(\sigma - 1)}{\alpha} z_t \\ E_t \Delta \ln c_{t+1} &\simeq -(\eta + \frac{\delta + \theta}{\sigma})(1-\alpha)E_t \ln k_{t+1} + (\eta + \frac{\delta+\theta}{\sigma})z_t \end{aligned}$$

which can be written as the system

$$\begin{bmatrix} 1 + \theta + (\sigma - 1)\eta & -\frac{\theta+\eta(\sigma-\alpha-2)+(1-\alpha)\delta}{\alpha} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \ln k_t \\ \ln c_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (\eta + \frac{\delta+\theta}{\sigma})(1-\alpha) & 1 \end{bmatrix} E_t \begin{bmatrix} \ln k_{t+1} \\ \ln c_{t+1} \end{bmatrix} - \begin{bmatrix} \frac{\theta+\delta+\eta(\sigma-1)}{\alpha} \\ \eta + \frac{\delta+\theta}{\sigma} \end{bmatrix} z_t$$

or as

$$x_t = AE_t x_{t+1} + F z_t$$

where $x_t = [\ln k_t \quad \ln c_t]'$. It can be shown that if $\sigma \geq \alpha + 2$ then we have the saddlepath solution

$$x_t = \frac{1}{\eta_1}x_{t-1} + \frac{1}{\eta_1}\Sigma_{s=0}^{\infty}\eta_2^s G_2 E_t z_{t+s} - \frac{1}{\eta_1}F z_{t-1}$$

where $G_2 = (1-\eta_2)[adj(A) - I]F$, $\eta_1 > 1$ and $\eta_2 < 1$. Noting that $E_t z_{t+s} = z_t$ for $s \geq 0$, this simplifies to

$$x_t = \frac{1}{\eta_1} x_{t-1} + G_0 z_t + G_1 z_{t-1}$$

where $G_0 = \frac{1}{\eta_1}(adj(A) - I)F$ and $G_1 = -\frac{1}{\eta_1}F$ are functions of the model's structural parameters and hence are restricted by the model. As z_t is non-stationary due to the presence of the permanent productivity shock, x_t is also non-stationary. However, as $\Delta z_t = e_t$, the solution can be re-written as

$$\Delta x_t = \frac{1}{\eta_1} \Delta x_{t-1} + G_0 e_t + G_1 e_{t-1}$$

which is a VARMA in the change in x_t and not the level of x_t; i.e. it is a restricted VARIMA(1,1,1).

This solution is characteristic of all DSGE models. Higher order lags in the model will lead to a longer lag structure and permanent shocks will introduce unit roots. In this RBC model the exogenous variable is unobservable but more generally DSGE models will have observable exogenous variables which will need to be forecast. Suppose, for example, that in this model z_t were observable and generated by the autoregression $z_t = \rho z_{t-1} + \varepsilon_t$ then $E_t z_{t+s} = \rho^s z_t$ and the solution may be written

$$x_t = \frac{1}{\eta_1} x_{t-1} + \frac{1}{\eta_1}[\frac{\rho}{1-\rho\eta_2} G_2 - F] z_{t-1} + \frac{1}{\eta_1(1-\rho\eta_2)} G_2 \varepsilon_t.$$

The data may therefore be written as being generated by the restricted VAR

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} = \begin{bmatrix} \frac{1}{\eta_1} I & \frac{1}{\eta_1}[\frac{\rho}{1-\rho\eta_2} G_2 - F] \\ 0 & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{1}{\eta_1(1-\rho\eta_2)} G_2 \\ 1 \end{bmatrix} \varepsilon_t$$

Again, this is characteristic of all DSGE models.

More generally, we may write DSGE models as

$$\begin{bmatrix} x_{t+1} \\ E_t y_{t+1} \end{bmatrix} = A \begin{bmatrix} x_t \\ y_t \end{bmatrix} + C z_t$$

where x_t, is a vector of predetermined variables (stocks), y_t is a vector of "jump" variables (flows or asset prices) and z_t is a vector of exogenous variables (including policy variables) and structural disturbances. The solution is a forward-looking VARX

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = M \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + N \Sigma_{s=0}^{\infty} \Gamma_{yy}^{-s} P_y E_t z_{t+s} + J z_{t-1} + K \xi_t$$

where $\xi_t = x_t - E_{t-1} x_t$. If the exogenous variables may be represented by the VAR $z_{t+1} = R z_t + \varepsilon_{t+1}$ then the complete solution for the data is the VAR

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = F \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + G \begin{bmatrix} \xi_{xt} \\ \xi_{yt} \\ \varepsilon_t \end{bmatrix}.$$

where F and G are restricted.

An implication of these results is that, provided it is not incorrectly constrained, a DSGE model, having a similar dynamic structure, will behave very like a VAR estimated unrestrictedly. It is noticeable that starting with the Smets-Wouters models most DSGE models now include flexible lag structures and have highly serially correlated disturbances. This greatly improves the ability of the models to explain the data and makes them perform much more like a VAR. In other words, there now is a strong element of data-mining in DSGE modelling as there was in the earlier SEM. Del Negro and Schorfheide (2008) ask whether it makes a difference whether prices or wages are flexible. They conclude that it doesn't matter. But what they fail to comment on is that their results show that if wages are constrained to be flexible, the serial correlation in the disturbance term of the wage equation increases significantly. This suggests that there is considerable wage stickiness. Allowing the disturbances of the model to be serially correlated appears to be a data-mining device that hides model misspecification and is another example of the conflict between theory and evidence.

4.2. Trend removal

It is coventional in DSGE modelling to filter macroeconomic data before estimation using the HP filter in order to remove trends and render the variables stationary. The problem with doing this is that the HP over-differences the data and induces cycles which will affect the impulse response functions.

The aim of the HP filter of z_t, which is assumed to be I(1), is to decompose it into two components: a trend μ_t and a "cycle" or deviation from trend c_t. The HP filter, which is a two-sided symmetric moving average of the original data, estimates the trend by solving the following problem which is a trade-off between choosing a trend that minimises deviations from the data and fluctuations in the trend:

$$\min_{\mu_t} \sum_{t=1}^{T} [(z_t - \mu_t)^2 + \lambda(\Delta^2 \mu_{t+1})^2]$$

where λ is a pre-specified number. The resulting trend is

$$\begin{aligned}
\mu_t &= \theta(L)z_t = \textstyle\sum_{s=-n}^{n} \theta_s z_{t-s}, \quad \sum \theta_s = 1, \quad \theta_s = \theta_{-s} \\
\theta(L) &= \textstyle\sum_{s=-n}^{n} \theta_s L^s = \theta_0 + \sum_{s=1}^{n} \theta_s (L + L^{-1})
\end{aligned}$$

If $\lambda = 0$ then $\mu_t = z_t$, and μ_t tends to a linear trend as $\lambda \to \infty$. Thus by making λ smaller, μ_t is made to follow z_t more closely.

The trend μ_t is the solution of the MA

$$z_t = \lambda\mu_{t+2} - 4\lambda\mu_{t+1} + (1 + 6\lambda)\mu_t - 4\lambda\mu_{t-1} + \lambda\mu_{t-2}$$

which can be re-written as

$$c_t = z_t - \mu_t = \lambda(1 - L)^4 \mu_{t+2} = \lambda\Delta^4 \mu_{t+2}$$

or as

$$\Delta^4 z_t = \Delta^4 c_t + \lambda c_{t-2} = \varphi(L)c_t$$

Thus the trend is I(1) and the de-trended series is I(0) and is proportional to the 4^{th} difference of the trend. In other words, μ_t is a 4^{th} difference even though z_t may only be I(1). z_t is therefore over-differenced and induces cycles in the de-trended series even though none may exist.

An alternative to the HP filter is to use a polynomial trend in time which does not induce cycles. For the logarithm of UK GDP 1830-2009 the output gap measured by the HP filter and a third order polynomial trend are shown in Figure 1. The cycle as measured by the HP filter is much smaller.

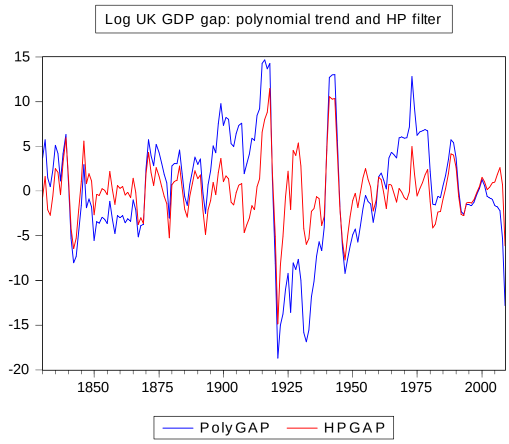

Figure 1. HP and polynomial UK output gap

According to the solution of the RBC model above the trend is a random walk with drift. The drift is the balanced growth path and can be removed by first differencing without distorting the cycle or the impulse response functions. This still leaves a non-stationary component, the random walk which implies that the VAR representation of the model can be rewritten as a cointegrated VAR in deviations from the balanced growth path.

More generally, non-stationarity in nominal variables may be attributed to real and nominal permanent shocks. If the real shock is an unobservable productivity shock and the nominal shock arises from observable money growth then the two shocks must be treated differently in the time series representation of a DSGE model. If some of the exogenous variables in the VAR representation are unobservable - as well as being non-stationary - then it will be necessary to first-difference the solution. Moreover, as growth in steady-state due to permanent productivity and monetary shocks implies that the main macro aggregates have both deterministic and stochastic trends, this also needs to be taken into account in a VAR analysis by including a time trend as well as dealing with the stochastic trend usually by using cointegration analysis.

4.3. Forecasting

The forecasting performance of DSGE models is also revealing. It is well known that time series models forecast at least as well as structural models. We can see from the above result that an unrestricted VAR is likely to forecast at least as well as a DSGE model. This is confirmed by the results of Gurkaynak, Kisacikoglu and Rossi (2013), Wickens (2013), Wieland and Wolters (2012). Figure 2, from Wieland and Wolters, shows the forecasting performances of an AR(4) (red line) and a BVAR(4) (blue line) compared with the data (black line). Figure 3, from Gurkaynak, Kisacikoglu and Rossi, compares forecasts from the Smets-Wouters model with the data.

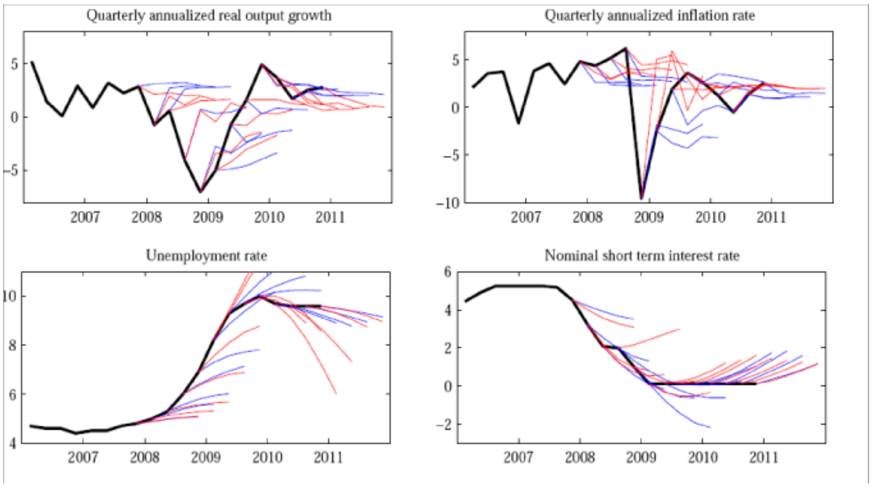

Figure 2. Forecasts of the US economy from AR and BVAR models

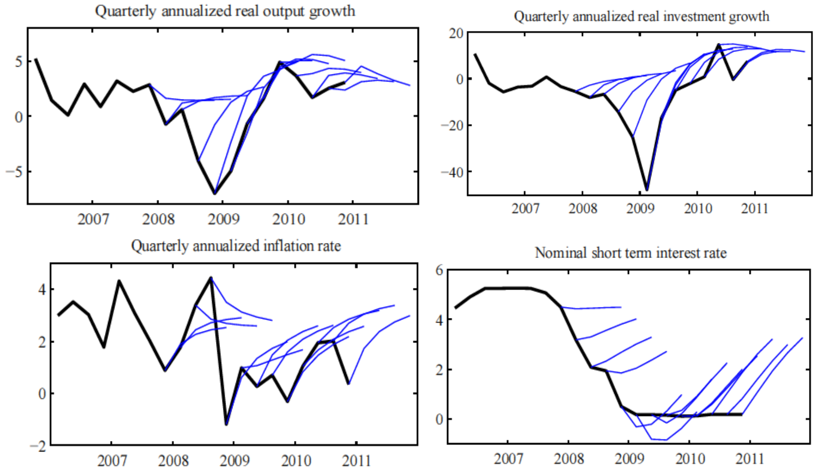

Figure 3. Forecasts of the US economy from the Smets-Wouters model

The results from all of the forecasts are very similar. They miss the recession as they are all consistenly mean-reverting. They forecast that growth will return to its long-run level and so also miss the depth of the recession. The inflation, unemployment and nominal interest rate forecasts flatline more, but also mean-revert. The growth and inflation forecasts explain the over-estimate of the interest rate forecasts. These findings suggest that, as implied by the theoretical results above, the time series and the DSGE models have similar internal dynamics. They also suggest that the failure to forecast the recession may be due to poor forecasts of the exogenous variables and not to the DSGE model itself. This supports the findings of Adelman and Adelman that the business cycle is due to the behaviour of the exogenous variables. It will be noted that the exogenous variables are not part of the DSGE model.

4.4. Identification problems of DSGE models

In order to be able to make inferences about a model from the data that model must be identified in order to ensure that other models would not give exactly the same empirical results. In SEM analysis the identification of the model was routine, but for DSGE models it is rare. Not all DSGE models are, however, identified. Canova and Sala (2009) give examples of models in which the reduced form properties of different DSGE models are hard to distinguish and argue that a weak form of observational equivalence between DSGE models is widespread. Schorfheide (2011) notes that the extreme nonlinearity of DSGE model solutions in their structural parameters makes checking identification difficult except by numerical methods.

To illustrate, consider the following simplified New Keynesian model:

$$\begin{aligned}
\pi_t &= \omega E_t \pi_{t+1} + \lambda y_t + e_{\pi t}, \qquad 0 < \omega < 1 \\
y_t &= E_t y_{t+1} - \frac{1}{\sigma}(r_t - E_t \pi_{t+1}) + e_{yt} \\
r_t &= \gamma \pi_t + \eta y_t + e_{rt} \\
e_{it} &= \rho_i e_{i,t-1} + \varepsilon_{it} \qquad (i = \pi, y, r)
\end{aligned}$$

The solution can be shown to be, see Le, Minford and Wickens (2013)

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \begin{bmatrix} 1+\frac{\eta}{\sigma}-\rho_\pi & \lambda & -\frac{\lambda}{\sigma} \\ -\frac{1}{\sigma}(\gamma-\rho_\pi) & 1-\omega\rho_y & -\frac{1}{\sigma}(1-\omega\rho_r) \\ \gamma-(\gamma-\frac{\eta}{\sigma})\rho_\pi & \lambda\gamma+\eta-\eta\omega\rho_y & 1-(1+\omega+\frac{\lambda}{\sigma})\rho_r+\omega\rho_r^2 \end{bmatrix}$$
$$\times \begin{bmatrix} \frac{e_{\pi t}}{1+\frac{\eta+\lambda\gamma}{\sigma}-[\frac{\lambda}{\sigma}+\omega(1+\frac{\eta}{\sigma})]\rho_\pi+\omega\rho_\pi^2} \\ \frac{e_{yt}}{1+\frac{\eta+\lambda\gamma}{\sigma}-[\frac{\lambda}{\sigma}+\omega(1+\frac{\eta}{\sigma})]\rho_y+\omega\rho_y^2} \\ \frac{e_{rt}}{1+\frac{\eta+\lambda\gamma}{\sigma}-[\frac{\lambda}{\sigma}+\omega(1+\frac{\eta}{\sigma})]\rho_r+\omega\rho_r^2} \end{bmatrix}$$

It is therefore identified through the covariance matrix. If, however, $\rho_i = 0$ for all i then the solution becomes

$$\begin{bmatrix} \pi_t \\ y_t \\ r_t \end{bmatrix} = \frac{1}{1+\frac{\eta+\lambda\gamma}{\sigma}} \begin{bmatrix} 1+\frac{\eta}{\sigma} & \lambda & -\frac{\lambda}{\sigma} \\ -\frac{1}{\sigma}\gamma & 1 & -\frac{1}{\sigma} \\ -\gamma & \lambda\gamma+\eta & 1 \end{bmatrix} \begin{bmatrix} e_{\pi t} \\ e_{yt} \\ e_{rt} \end{bmatrix}$$

This does not involve ω, which is not therefore identified; any non-negative value of ω less than unity would give the same solution. The other coefficients are over identified. This shows the important role here of ad hoc error dynamics in identifying ω. More generally, often there are difficulties in identifying the rate of time preference. Not surprisingly, this tends to be a calibrated parameter in most studies.

4.5. Bayesian estimation of DSGE models

At best Bayesian methods improve estimates by taking account of additional (prior) information; at worst they disguise a misspecified model. The current dominance of Bayesian estimation for DSGE models reflects both the concern that the estimates would look implausible using classical methods, and the vast improvement in numerical procedures. It is also an attractive choice where maximum likelihood estimation is impractical, perhaps due to the number of parameters to be estimated, or to the difficulties of deriving a likelihood function analytically. To set against this, it is not encourging that most posterior estimates of DSGE parameters - whether the mean, median or mode - are very close to the mean of their prior distribution. Are the priors well chosen, or are the priors so strong that they are overwhelming the sample information? Bayesian estimation for econometric models was first strongly advocated in the 1970s - see for example Zellner (1977) and Dreze and Richard (1983) - but the numerical difficulties in obtaining the estimates made this intractable. Having solved the numerical problems, the principal issue that remains is whether the priors are too strong and are disguising the model's misspecification.

To illustrate what Bayesian estimation does compare Bayesian with maximum likelihood estimation. In ML estimation we choose θ to maximise the log likelihood function $\ln L(x/\theta)$; i.e.

$$\arg\max_{\theta} \ln L(x/\theta)$$

The ML estimator $\widetilde{\theta}$ is obtained from

$$\frac{\partial \ln L(x/\theta)}{\partial \theta}|_{\theta=\widetilde{\theta}} = 0$$

If $p(\theta)$ is the prior distribution then the posterior distribution is

$$p(\theta|x) = \frac{L(x/\theta)p(\theta)}{p(x)} = \frac{p(x,\theta)}{p(x)}$$

where $p(x)$ is the marginal probability function of x. In earlier years a major problem for Bayesian estimation was in deriving $p(x) = \int p(x, \theta)d\theta$. Suppose, however that we estimate θ by the mode of the posterior distribution then we require

$$\arg\max_{\theta} p(\theta/x) \equiv \arg\max_{\theta} \ln p(\theta/x)$$

But $\ln p(\theta/x) = \ln L(x/\theta) + \ln p(\theta) - \ln p(x)$ hence $p(x)$ plays no role and can be ignored. The mode $\widehat{\theta}$ is then given by

$$[\frac{\partial \ln L(x/\theta)}{\partial\theta} + \frac{\partial \ln p(\theta)}{\partial\theta}]_{\theta=\widehat{\theta}} = 0$$

and can be obtained numerically from

$$\widehat{\theta} - \theta_0 = -[\frac{\partial^2 \ln L(x/\theta)}{\partial\theta\partial\theta'} + \frac{\partial^2 \ln p(\theta)}{\partial\theta\partial\theta'}]^{-1}[\frac{\partial \ln L(x/\theta)}{\partial\theta} + \frac{\partial \ln p(\theta)}{\partial\theta}]_{\theta=\theta_0}.$$

In effect, the information from the data and the prior are weighted inversely to their precision. It follows that if the log likelihood function $\ln L(x/\theta)$ is flat then $\frac{\partial \ln L(x/\theta)}{\partial\theta}$ is close to zero for a range of values of θ, and the Bayesian estimator is dominated by the prior. And if $p(\theta)$ is flat, or uninformative, then $\frac{\partial \ln p(\theta)}{\partial\theta} = 0$ and the data dominate. The choice of prior may therefore be critical, especially if the data do not discriminate well between alternative sets of parameter values of the DSGE model. There is, however, an argument for treating inappropriate priors on a par with misspecified DSGE models and judge both on how well the resulting estimated model explains the data. This requires a model evaluation procedure.

4.6. Evaluating DSGE models

I have asserted that all macroeconometric models are misspecified, perhaps deliberately in order to work with simple models. This raises the question about the DSGE model posed by Canova (1994): "given that it is false, how true is it?" Invoking Friedman's criterion for a good model, we can ask how well this false model explains the data. Our previous discussion has provided us with a means of answering this question because we have shown that a first-order approximation to DSGE models can be represented as a VAR (or a VARMA). By comparing the properties of a VAR estimated on actual data with one estimated on data simulated from a previously estimated DSGE model we can establish how well the DSGE model explains the data. We can even perform a formal statistical test of the differences between the two as explained in Le et al. (2011). It illustrates how advances in computing power have opened up new methods of statistical analysis. We can, for example, derive the small sample distribution of the test statistic using numerical simulation rather than resort to asymptotic distribution theory.

Indirect inference is particularly useful for Bayesian estimated DSGE models. It may be noted that the test could even be performed on a DSGE model obtained by indirect estimation, the SQML estimator. Here the parameters of the DSGE model are chosen to minimise differences between an auxiliary model (say a VAR) estimated on data simulated from the model and the actual data. Giraitis, Kapetanios, Theodoridis and Yates (2014) provide indirect estimates of the Smets-Wouters model based on an unrestricted auxiliary VAR model in which monetary shocks are arbitrarily identified by a Choleski decomposition. They find that with the exception of the inflation and wage equations their estimates of the impulse response functions from the DSGE model are similar to those of Smets and Wouters. However, Le at. al. find that the Smets-Wouters model is rejected against a model with more classical (flexible) prices and wages.

It has been assumed so far that the parameters of DSGE models are fixed. Allowing them to be time-varying could be interpreted either as a different form of data-mining and obtaining a better fit to the data, or as taking into account structural change in the model. Giraitis et al. examine the behaviour of the Smets-Wouters model when its parameters are time varying. They use indirect estimation but with a time-varying parameter VAR. They find considerable time-variation in most of the estimates and hence in the impulse response functions. They highlight in particular fluctuations in the parameters determining the speeds of response of prices and wages in different monetary regimes, with prices and wages showing much more flexibility when inflation is stable. They also find considerable time variation in the habit persistence parameter (which falls to near zero from around 1995), the labour supply elasticity, the response of investment to changes in the user cost of capital and the response to a government spending shock. These results suggest that there is in fact considerable change in behaviour that depends on the external environment of shocks facing the economy. It is what one would expect if there were changes to policy regimes. The results therefore pose a challenge both to the time series modelling of macroeconomic data and to providing a theoretical explanation for the changes.

Most DSGE models are non-linear. A first-order approximation - as made above - may not therefore be accurate enough. This would make a linear VAR solution also an approximation as a higher order approximation would be required. It may also cause a linear VAR to have time-varying coefficients as found by Giraitis et al. There is therefore the possibility that a DSGE model that appears to have time-varying parameters may just be highly non-linear and have constant parameters. Obtaining better numerical approximations to DSGE models is an active area of current research.

To what extent has employing economic theory in a more logically consistent way improved our understanding of the economy? The many critics of DSGE models might well agree with Krugman's view that it has been a backward step. This assessment seems to be more on the basis of the plausibility of their assumptions than on the explanatory power of these models. There are, as pointed out, many difficulties that remain to be addressed in the specification and implementation of DSGE models. Despite these limitations, my own view is that the DSGE approach to macro modelling has greatly improved our understanding of the interactions within the macroeconomy and provided a richer analytic framework due to the breadth of their predictions. This is, however, both a strength and a weakness as the wider the set of predictions, the more exacting the empirical standards they must pass.

5. Asset pricing models

The problem of theory versus data-mining is even more stark in asset pricing. Until recently asset pricing was the preserve of finance. The aim was to price one asset off another which is known as relative asset pricing. Even today, models of this sort are the basis of CAPM, equity and bond pricing. Although many relative pricing theories are based on the principle of no arbitrage, this is not true of all asset-pricing theories. It is, however, likely to be true of econometric studies that treat financial data as a suitable object for time series analysis without much accompanying theory. The strength of relative asset pricing theory is that it provides models that fit the data well. Its weakness is the absence of an explicit theory of risk. Instead, the expected excess return on an asset - the risk premium - is given an ex-post interpretation as being due to risk. For example, the implied risk premium in CAPM is proportional to the expected excess return on the market as in

$$E_t(r_{t+1} - f_t) = \beta E_t(r^M_{t+1} - f_t)$$

where r_{t+1} is the risky-rate of return on the asset, r^M_{t+1} is the market return and f_t is the risk-free rate.This does not involve any explicit theory of risk even though β may be interpreted as reflecting risk.

As previously noted, in DSGE macroeconomics,where equilibrium is defined in terms of stocks, the price of an asset is a key variable explained by the model. DSGE models therefore provide an alternative approach to asset pricing, known as general equilibrium models of asset pricing. Further, the theory models the risk premium explicitly as the conditional covariance between the asset price and the pricing kernel. In the case of consumption CAPM (C-CAPM) the pricing kernel is the inter-temporal marginal rate of substitution, $\frac{\beta U'(c_{t+1})}{U'(c_t)}$ and the risk premium is the expected excess return

$$E_t(r_{r+1} - f_t) = \frac{1}{E_t[\frac{\beta U'(c_{t+1})}{U'(c_t)}]} Cov_t[\frac{\beta U'(c_{t+1})}{U'(c_t)}, r_{t+1} - f_t]$$

It can be shown that CAPM can be given a general equilibrium gloss because as consumption is approximately proportional to wealth and the market return is approximately the return on wealth, the implied risk premium in CAPM is approximately the same as the risk premium derived from C-CAPM. It can also be shown that general equilibrium asset pricing is a special case of the stochastic discount factor approach to asset pricing, which invokes the principle of no-arbitrage but involves no explicit theory of risk. Only if the pricing kernel in the SDF approach can be given a general equilibrium interpretation can the SDF approach convey a theory of risk. This requires that the conditional expectation of the pricing kernel satisfies $E_t M_{t+1} = \frac{1}{1+f_t}$, where, for C-CAPM, $M_{t+1} = \frac{\beta U'(c_{t+1})}{U'(c_t)}$.

Financial data provide a useful alternative way of testing the DSGE model to macroeconomic data. A key feature of these data is that a theory must explain is the time-varying volatility of returns. This implies that DSGE models must have time-varying second moments if they are to explain both macro and financial data. When approximating a DSGE model more attention must therefore be paid to higher order terms. The convention in most macro applications of simply using a first-order linearisation of the DSGE model and assuming that the resulting disturbance terms have constant second moments is not appropriate for financial data.

5.1. Equity

Mehra and Prescott (1985) showed that the equity risk premium generated by C-CAPM had insufficient variability to explain the much greater volatility of excess returns. They called this the equity premium puzzzle. Subsequently, this has been a frequent finding in tests of general equilibrium models of asset pricing based on classical estimation and time-varying conditional covariances. It applies to equity, bonds and foreign exchange, see Smith and Wickens (2002).

There have been two types of response to the equity premium puzzle. One is to reformulate the DSGE model in order to increase volatility of the stochastic discount factor and hence the risk premium. Allowing the coefficient of risk aversion to be very large is one way to achieve this, but is not a plausible solution. Another is to alter the DSGE model by assuming habit peristence. A further suggestion is to assume time non-separable utility functions such as Epstein-Zin utility. Campbell (2002) and Campbell and Cochrane (1995) found support for the last two theoretical modifications using calibration and unconditional covariances. The weakness of this evidence is that it is based on calibration and the use of unconditional expectations which implies constant risk premia. Smith, Sorensen, Smith and Wickens (2006), assuming time varying conditional covariances, use maximum likelihood estimation but found that even with these theoretical modifications the equity premium puzzle remains.

A second approach, dominant in the finance industry, and much closer in spirit to data-mining, is to model excess returns with ad hoc factors without necessarily imposing a no-arbitrage condition on the model. Fama and French (1993, 1995, 1996) inspired a vast literature in response to their studies of CAPM where they found that the book-to-market ratio and the high-low spread dominated the excess return on the market in explaining the excess return to equity. In a similar vein, there have also been a large number of studies trying to forecast equity returns using an array of explanatory variables. A recent study adopting this approach based on both observed and unobserved factors by Morano (2013) finds that observable demand-side macro shocks account for much more than unobservable financial shocks in explaining short-term risk-factor fluctuations in stock returns. An alternative approach suggested by Pongrapeeporn, Smith and Wickens (2013) using panel data is to reformulate the Fama-Fench model within a no-arbitrage framework based on C-CAPM. The main findings are that consumption growth, inflation and the spread between high and low returns (HML) are all significantly-priced sources of risk but, due to the significance of HML, C-CAPM is rejected in favour of a more general SDF model.

5.2. Bonds

The dominant approach in bond pricing is relative asset pricing in which long rates are priced off the short rate. Essentially this is curve fitting with the imposition of a no-arbitrage condition and otherwise lacks economic fundamentals. Although the yield curve changes shape over time, a key stylised fact is that three factors derived solely from the term structure of interest rates explain at least 98% of the variation in all yields. In order of importance, the three factors are a shift, a slope and a curvature factor. The problem is to provide a theory that accounts for this finding.

The expectations hypothesis of the term structure, which is based on the assumption of risk-neutral investors, explains long yields as the average of current and expected future short rates. Although this may explain the shift factor as being due to a rise in current and future short rates, it does not explain the other two factors. In order to do so it is necessary to show why short rates combine to give three factors. The empirical findings provide contradictory evidence on the theory. Whereas Campbell and Shiller (1984) found that long rates under-reacted to short rates, Campbell and Shiller (1991) and Hardouvelis (1994) found that long rates over-reacted to short rates. However, Tzavalis and Wickens (1997, 1978) found that, once allowance is made for a term premium, long rates react to short rates as theory predicts.

The predominant current approach is the affine factor model of the term structure, examples of which are the Vasicek (1977) and Cox, Ingersoll and Ross (CIR,

1985) models. These models are based on the SDF approach. It is assumed that both yields and the stochastic discount factor are a linear function of unobserved factors that are generated by an autogressive or a VAR model with either constant or time-varying volatility. Neither the single factor Vasicek or CIR models are adequate. The risk premium for the Vasicek model is constant over time while that of the CIR model is proportional to the short rate and so fixes the shape of the yield curve for all time. The multi-factor affine CIR model with three factors appears to fit the data much more closely, see Dai and Singleton (2000). The three factors can be linear functions of the short rate and any other two yields. An extension of this approach is to add observable factors as in Ang and Piazzesi (2003). This requires modifying the equation for the stochastic discount factor. They claim that up to 85% of the forecast variance for long rates at short and medium maturities can be explained by two macro variables in which inflation is the dominant factor.

I have argued that the critical measure of the usefulness of a theory is how well it explains the data. The attraction of these models is that they fit the data very closely. They are, however, less attractive from an economic point of view as they do not provide an economic explanation of the data. Nonetheless, they provide a time series model of the data which economic theories can, perhaps, use as benchmark to measure themselves against.

Embedding the term structure in a DSGE model is a growing area of research. Balfoussia and Wickens (2007) applied C-CAPM to nominal bond yields, adding the constraint of no-arbitrage across the term structure. The pricing kernel for this model involves two macro variables: inflation and consumption. Like Ang and Piazzesi, they found that inflation was an important significant factor. Nonetheless, they observed the same problem as in the equity premium puzzle that the pricing kernel was not sufficiently volatile.

Rudebusch et al. (2007) and Dew-Becker (2013) embed the term structure within a New Keynesian DSGE model with a Taylor rule determining the short rate, and seek to determine the contribution of macro shocks to both macro variables and the term structure. Rudebusch et al calibrate their model, then take a third-order approximation to capture time-varying volatilty effects, and finally simulate the model for shocks to the Fed Funds rate, government expenditures and technology. They find that none of these shocks generate sufficiently large changes in the term premium. Dew-Becker uses a mixture of calibration and Bayesian estimation, and allows the shocks in the model to be autocorrelated. He finds that the model can generate a large and volatile term premium that is driven almost entirely by two factors: a negative response of interest rates to positive technology shocks and to variation in risk aversion.

These findings highlight the gap that remains between the empirical performance of asset-pricing models based on macroeconomic fundamentals and the data-based methods based on relative asset pricing. The challenge for economists is to develop macroeconomic models that close this gap. The pay-off is that we will then have a much better interpretation of risk and its sources.

6. Concluding remarks

This lecture has been about how best to evaluate economic theories and the lessons that can be learned from the past use and misuse of evidence. In the process I have examined the uneasy relationship between theory and evidence in macroeconomics and finance. I have taken a highly personal perspective based on

50 years of experience. I have argued that there is no such thing as a true macroeconomic theory and so judging macroeconomic theories by the validity of their assumptions is unlikely to be helpful. More useful criteria for judging macroeconomic theories are the insights that they provide on the behaviour of the economy, their usefulness for policy and their ability to explain macroeconomic and financial data. Data mining will often improve the statistical properties of a model but it does not improve economic understanding.

I have argued that it is not helpful to think in terms of a true data-generating process as there is usually more than way to represent data statistically. Further, interpreting the evidence from modelling macroeconomic and financial data using time series methods requires economic theory. On their own time series models tell us very little about how the data have been generated.

Over the last 50 years opinion has changed a lot about how best to formulate macroeconomic theories and to bring evidence to bear on them. Some of this is due to the substantial increase in computing power. But I prefer to think that the main influence has been the drive to develop a framework such as the DSGE approach to macro and financial modelling that can be extended to encompass in a logically consistent way features that are thought to be relevant to a problem and, as a result, provide an integrated explanation of the economy. Despite the increasing size of DSGE models, I doubt the usefulness of making models all-embracing as they become unwieldy and run the risk of misspecification in one part of the model affecting the results.

Finally, I mention two concerns. First, a gap exists in the way that financial economists in the industry and those working with general equilibrium theories think about explaining financial data. The onus is on the latter to produce better theories. Second, a gap has opened up between much of the current research in econometrics and the putative users in macroconomics and finance of these new methods. Econometric research has become increasingly inward-looking over the years and, as a result, its influence on macroeconomics has diminished. DSGE modelling has become steadily more technical using advanced numerical analysis. I hope that this too will not prove to be at the expense of its usefulness to macroeconomists.

References

Abhakorn, P., P.N. Smith and M.R. Wickens (2013). “What do the Fama-French factors add to C-CAPM?” *Journal of Empirical Finance.*

Adelman, F.L. and I. Adelman. (1959). "The dynamic properties of the Klein-Goldberger model". Econometrica, 27, 596-625.

Andersen, L. and J. Jordon (1968), "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization," *Federal Reserve Bank of St. Louis Review*, 50, 11-24.

Archibald, G.C. and R.G. Lipsey (1963), *An introduction to positive economics*, Weidenfeld and Nicolson.

Balfoussia, H. and M.R. Wickens (2007), “Macroeconomic Sources of Risk in the Term Structure”, *Journal of Money, Credit and Banking*, 39, 205-236.

Box, G.E.P. and N.R. Draper (1987), Empirical model-building and response surfaces, Wiley.

Canova, F. (1994). "Statistical inference in calibrated models". *Journal of Applied Econometrics*, 9, S123-144.

Canova, F. and Sala, L. (2009). "Back to Square One: Identification Issues in DSGE Models", *Journal of Monetary Economics*, 56(4), 431-449.

Campbell J.Y. (1987). "Stock returns and the term structure" *Journal of Financial Economics*, 18, 373-399.

Campbell, J.Y. (2002), "Consumption-based Asset Pricing," G. Constantinides, M. Harris and R. Stulz eds., *Handbook of the Economics of Finance*, Amsterdam: Elsevier.

Campbell, J.Y. and J.H. Cochrane (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy,* 107, 205-251.

Campbell J.Y. and R.J. Shiller (1984), "A simple account of the behaviour of long-term interest rates" *American Economic Review,* 7, 44-48

Campbell J.Y. and R.J. Shiller (1991). "Yield spreads and interest rate movements: a bird's eye view" *Review of Economic Studies*, 58, 495-514

Cox, J.C., J.E. Ingersoll and S.A. Ross (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-408.

Dai, Q. and K. Singleton (2000), "Specification analysis of affine term structure models," *Journal of Finance*, 55, 1943-1978.

Davidson, J.E.H., D.F. Hendry, F. Srba, and J.S. Yeo (1978). "Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom," *Economic Journal*, 88, 661-692.

Del Negro, M., Schorfheide, F. (2006). "How good is what you've got? DSGE-VAR as a toolkit for evaluating DSGE models." *Economic Review*, Federal Reserve Bank of Atlanta, issue Q2, 21–37.

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, 1191-1208.

Del Negro, M., F. Schorfheide, F. Smets and R. Wouters (2007). "On the fit of new Keynesian models." *Journal of Business and Economic Statistics* 25,123–143.

Desai, M. (1975). "The Phillips Curve: A Revisionist Interpretation." *Economica,* 42, 1-19.

Dew-Becker, I. (2013). "Bond pricing with a time-varying price of risk in an estimated medium-scale Bayesian DSGE model" mimeo 2013.

Dreze, J. and J-F. Richard (1983). "Bayesian Analysis of Simultaneous Equation Systems," The Handbook of Econometrics, (Chapter 9), edited by Z. Griliches and M.D. Intriligator, North Holland, 519-598.

Duesenberry, J.S. ed., (1965), *The Brookings Quarterly Econometric Model of the United States,* Brookings Institution, Amsterdam: North Holland.

Evans, R. and S. Honkapohja (2005) 'Interview with Thomas J. Sargent', *Macroeconomic Dynamics*, 9, 561–583.

Fama, E.F., and K.R. French (1993) "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, 33, 3-56.

Fama, E.F., and K.R. French (1995). "Size and book-to-market factors in earnings and returns," *Journal of Finance,* 50, 131-155.

Fama, E.F., and K.R. French (1996). "Multifactor explanations of asset pricing anomalies," *Journal of Finance* 51, 55-84.

Fleming, J.M. (1962), "Domestic Financial Policies Under Fixed and Under Floating Exchange Rates," *International Monetary Fund Staff Papers*, 9, 369-79.

Giraitis, L., G. Kapetanios, K. Theodoridis and T. Yates (2014) "Estimating time-varying DSGE models using minimum distance methods", Bank of England Working Paper no. 507.

Gurkaynak, R.S., B. Kisacikoglu and B. Rossi (2013). "Do DSGE models forecast more accurately out-of-sample than VAR models?" CEPR Discussion Paper DP9576.

Hall, R.E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy,* 86, 971-987.

Hardouvelis G.A. (1994). "The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle?" *Journal of Monetary Economics*, 33, 255-283.

Hendry, D.F. (1989). "Comment on intertemporal consumer behaviour under structural changes in income." *Econometric Reviews*, 8, 111-121.

Hendry, D.F. and N.R. Ericsson (1991). "Modeling the the demand for narrow money in the United Kingdom and United States," *European Economic Review*, 35, 883-886.

King, R.G., C.I. Plosser and S.T. Rebelo (1988), "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21 195-232.

Krugman, P., 2009. “How did economists get it so wrong?” The New York Times, Sept. 6.

Johansen, S. (1988). "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Le, V.P.M., D. Meenagh, P. Minford and M.R.Wickens (2011). “How much nominal rigidity is there in the US economy? Testing a New Keynesian DSGE Model using indirect inference”. *Journal of Economics Dynamics and Control*, 35, 2078-2104.

Le, V.P.M., D. Meenagh, P. Minford and M.R.Wickens (2012). ”Testing DSGE models by Indirect Inference and other methods: some Monte Carlo experiments”, Cardiff Economics Working paper 2012/15; also CEPR discussion paper 9056.

Lipsey, R.G. and A. Chrystal (2011), *Economics*, Oxford University Press

Lucas, R.E. (1976a), "Econometric Policy Evaluation: A Critique." In K. Brunner and A. H. Meltzer (eds.), *The Phillips Curve and Labor Markets,* Amsterdam: North-Holland, 19-46.

Mehra, R. and E.C. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics,* 15, 145-161.

Morana, C. (2013) "Insights on the global macro-finance interface: structural sources of risk factors fluctuations and the cross-section of expected stock returns", University of Milan Bicocca Department of Economics, Management and Statistics Working Paper No. 264.

Mundell, R.A. (1963), "Capital Mobility and Stabilization Policy Under Fixed and Flexible Exchange Rates," *Canadian Journal of Economics and Political Science,* 29, 475-85.

Mussa, M. (1976), "The Exchange Rate, the Balance of Payments, and Monetary and Fiscal Policy Under a Regime of Controlled Floating," *Scandinavian Journal of Economics,* 78, 229-248.

Mussa, M. (1982), "A Model of Exchange Rate Dynamics," *Journal of Political Economy,* 90, 74-104.

Muth, J.F. (1960), "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association*, 55, 299-306.

Phillips, A. W. (1958), "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957." *Economica,* 25, 283-299.

Polito, V. and M.R. Wickens (2012). “Optimal Monetary Policy Using an Unrestricted VAR”, *Journal of Applied Econometrics*, 27, 525-553.

Rudebusch, G.D. and E.T. Swanson (2008). "Examining the bond premium puzzle with a DSGE model." *Journal of Monetary Economics*, 55, S111-S126.

Rudebusch, G.D. and E.T. Swanson (2012). "The bond premium in a DSGE model with long-run real and nominal risks." *American Economic Journal: Macroeconomics*, 4, 105-143.

Sargent, T. and N. Wallace (1976) "Rational Expectations and the Theory of Economic Policy". *Journal of Monetary Economics,* 2, 169–183.

Schorfheide, F. (2011). "Estimation and evaluation of DSGE models: progress and challenges", Working paper 11-7, Federal Reserve Bank of Philadelphia.

Sims, C.A. (1980), "Macroeconomics and Reality," *Econometrica,* 48, 1-48.

Smets, F. and R. Wouters (2003). "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5),1123–1175.

Smets, F. and R. Wouters. (2007). "Shocks and frictions in US business cycles: a bayesian DSGE approach", *American Economic Review*, 97, 586-606.

Smith, P.N. and M.R. Wickens (2002), “Asset Pricing with Observable Stochastic Discount Factors”, *Journal of Economic Surveys*, 16, 397-446.

Smith, P.N. and M.R. Wickens (2007), "The New Consensus in Monetary Policy: Is the NKM Fit for the Purpose of Inflation Targeting," in P.Arestis ed., *The New Consensus in Monetary Policy*, New York: Palgrave Macmillan, 97-127.

Smith, P.N., S. Sorensen and M.R. Wickens (2006), "General Equilibrium Theories of the Equity Risk Premium: Estimates and Tests," *CEPR* Discussion Paper.

Tzavalis, E. and M.R. Wickens (1997). "Explaining the failures of term spread models of the rational expectations hypothesis of the term structure". *Journal of Money Credit and Banking*, 29, 364-380.

Tzavalis, E. and M.R. Wickens (1998). "A re examination of the rational expectations hypothesis of the term structure: reconciling the evidence from long run and short run tests", *International Journal of Finance and Economics*, 3, 229-240.

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

Wickens, M.R. (1996). "Interpreting cointegrating vectors and common stochastic trends", *Journal of Econometrics*, 74, 255 271.

Wickens, M.R. (2010). "What's wrong with modern macroeconomics? Why its critics have missed the point," *CESifo Economic Studies,* 56, 536-553.

Wickens, M.R. (2014). "How useful are DSGE macroeconomic models for forecasting?" *Open Economies Review*, 25, 171-193.

Wickens, M.R. and R. Motto (2000), "Estimating Shocks and Impulse Response Functions", *Journal of Applied Econometrics*, 16, 371-387.

Wieland, V. and M.H. Wolters. (2012). "Forecasting and policy making", in G. Elliott and A. Timmerman (eds.), *Handbook of Economic Forecasting*, Vol. 2, Elsevier

Zellner, A. (1977). Bayesian Analysis in Econometrics and Statistics: The Zellner View and Papers

Edward Elgar Pub.