

ISSN 1471-0498



DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES

**The future of macroeconomics: Macro theory and models at
the Bank of England**

David F. Hendry and John N. J. Meullbauer

Number 832
September, 2017

Manor Road Building, Oxford OX1 3UQ

The future of macroeconomics: macro theory and models at the Bank of England

David F. Hendry* and John N. J. Muellbauer**

Abstract: The adoption as policy models by central banks of representative agent New Keynesian dynamic stochastic general equilibrium models has been widely criticised, including for their simplistic micro-foundations. At the Bank of England, the previous generation of policy models is seen in its 1999 medium-term macro model (MTMM). Instead of improving that model to correct its considerable flaws, many shared by other non-DSGE policy models such as the Federal Reserve's FRB/US, it was replaced in 2004 by the DSGE-based BEQM. Though this clearly failed during and after the global financial crisis, it was replaced in 2011 by the DSGE COMPASS, complemented by a 'suite of models'. We provide a general critique of DSGE models for explaining, forecasting and policy analyses at central banks, and suggest new directions for improving current empirical macroeconomic models based on empirical modelling broadly consistent with better theory, rather than seeking to impose simplistic and unrealistic theory.

Keywords: DSGE, central banks, macroeconomic policy models, finance and the real economy, financial crisis, consumption, credit constraints, household portfolios, asset prices

JEL classification: E17, E21, E44, E51, E52, E58, G01.

* Economics Department and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, e-mail: david.hendry@nuffield.ox.ac.uk

** Nuffield College and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, e-mail: john.muellbauer@nuffield.ox.ac.uk

Financial support from the Robertson Foundation (award 9907422), Institute for New Economic Thinking (grant 20029822), and Statistics Norway, through Research Council of Norway Grant 236935, are all gratefully acknowledged. We are grateful for helpful comments to a referee, Kevin Hoover, Simon Hayes at the Bank of England, Ken Mayhew, and David Vines, but take responsibility for errors and omissions.

I. Introduction

After the global financial crisis, quantitative models used by central banks, as well as prevailing fashions in macroeconomic theory and the training of graduate students in economics, came in for heavy criticisms, some mentioned in Blanchard (2018, this issue). This paper is about policy modelling at central banks rather than the syllabus of first-year graduate courses in macroeconomics, though there are implications for the training of bank economists. The paper begins by reviewing some of the above criticisms, particularly of the New Keynesian dynamic stochastic general equilibrium (DSGE) models used at central

banks. To help appreciate the contrasts between this approach and the policy models that preceded it, section II critically analyses the Bank of England's 1999 policy model, an example of what Wren-Lewis (2018, this issue) calls a 'structural econometric model' (SEM). Section III provides a general critique of New Keynesian DSGE models. Sections IV and V explain, and drawing on section III, critically analyse the Bank of England's 2004 and 2011 applications of the New Keynesian DSGE approach. Section VI suggests directions for improvement in policy modelling and section VII concludes. Thus, sections II, IV, and V, analysing specific practice at the Bank of England, illustrate the more general critiques of section III and the proposals for improved modelling in section VI.

In the most fundamental and scathing of post-crisis critiques, Buiter (2009) said:

The Monetary Policy Committee of the Bank of England I was privileged to be a 'founder' external member of during the years 1997–2000 contained, like its successor vintages of external and executive members, quite a strong representation of academic economists and other professional economists with serious technical training and backgrounds. This turned out to be a severe handicap when the central bank had to switch gears and change from being an inflation-targeting central bank under conditions of orderly financial markets to a financial stability-oriented central bank under conditions of widespread market illiquidity and funding illiquidity. Indeed, the typical graduate macroeconomics and monetary economics training received at Anglo-American universities during the past 30 years or so, may have set back by decades serious investigations of aggregate economic behaviour and economic policy-relevant understanding. It was a privately and socially costly waste of time and other resources.

Most mainstream macroeconomic theoretical innovations since the 1970s (the New Classical rational expectations revolution¹ associated with such names as Robert E. Lucas Jr., Edward Prescott, Thomas Sargent, Robert Barro etc., and the New Keynesian theorizing of Michael Woodford and many others) have turned out to be self-referential, inward-looking distractions at best. Research tended to be motivated by the internal logic, intellectual sunk capital and aesthetic puzzles of established research programmes rather than by a powerful desire to understand how the economy works—let alone how the economy works during times of stress and financial instability. So the economics profession was caught unprepared when the crisis struck.

He went on to criticize the complete markets paradigm of the established research programme: 'Both the New Classical and New Keynesian complete markets macroeconomic theories not only did not allow questions about insolvency and illiquidity to be answered. They did not allow such questions to be asked.'

¹ See Wren-Lewis (2018) and Hoover (1994) for an illuminating intellectual history of this revolution.

Muellbauer (2010) noted that: ‘the recent generation of DSGE models failed to incorporate many of the liquidity and financial accelerator mechanisms revealed in the global financial crisis’. He argued:

Underlying conceptual reasons for the failure of central bank models of the DSGE type include their typical assumptions about representative agents, perfect information, zero transactions costs, and of efficient markets. For most of these models, with the notable exception of Bernanke *et al.* (1999), and others who also incorporate a financial accelerator for firms, e.g. Christiano *et al.* (2003), Christiano *et al.* (2009), it is as if the information economics revolution, for which George Akerlof, Michael Spence and Joe Stiglitz shared the Nobel Prize in 2001, had not occurred. The combination of assumptions, when coupled with the trivialisation of risk and uncertainty in these supposedly stochastic models, and the linearisation techniques used in their solution, render money, credit and asset prices largely irrelevant. The calibration/estimation methods currently used to apply these models to the data typically ignore inconvenient truths.

Caballero (2010) criticized the ‘pretence of knowledge syndrome’:

the current core of macroeconomics—by which I mainly mean the so-called dynamic stochastic general equilibrium approach—has become so mesmerized with its own internal logic that it has begun to confuse the precision it has achieved about its own world with the precision that it has about the real one. This is dangerous for both methodological and policy reasons.

Romer (2016) shares Cabellero’s perspective and is critical of the incredible identifying assumptions and ‘pretence of knowledge’ in both Bayesian estimation, e.g. Smets and Wouters (2007), and the calibration of parameters in DSGE and real business cycle (RBC) models.

Stiglitz (2018, this issue) uses insights from the information economics revolution, to which he made such major contributions, to criticize many aspects of recently fashionable macroeconomics.

As Reis (2018, this issue) notes, academic research since 2010 *has* moved on; for example, there are DSGE models incorporating financial frictions and banking, and some models that do not linearize around a steady state, so permitting more serious fluctuations and the potential of multiple solutions. The heterogeneous agent research agenda, incorporating liquidity constraints and uninsurable idiosyncratic income risk, of Kaplan *et al.* (2016) and others, has abandoned the complete markets paradigm. Moreover, microeconomic evidence with macroeconomic implications is being studied more intensively and such research is being published. Lindé (2018, this issue) argues that ‘improved DSGE models—modified to take the lessons of the recent crisis into account—will remain as a workhorse in many policy institutions for a long time to come’.

Blanchard (2018) points to a number of failings of DSGE models and recommends greater openness to more eclectic approaches. He acknowledges the usefulness of theoretical models with stylized simplifications² to explore important macroeconomic issues, but questions whether even the newer generation of DSGE models are suitable as central bank policy models. He recommends the co-existence of this kind of research with work on policy models such as the FRB/US model, in which economic theory is used more loosely and the data are allowed to speak more freely than in typical DSGEs. This is an example of what Wren-Lewis terms a ‘structural econometric model’³ (SEM). Up to 2003, the Bank of England maintained such a model and the UK Treasury continues to do so. The Bank of Canada and the Netherlands central bank have implemented new SEMs since 2010 and the European Central Bank (ECB) is actively developing SEMs for five major Eurozone economies in its ECB-MC project. Blanchard also usefully distinguishes policy models or SEMs from forecasting models, judged purely on how well they forecast.

As the above outline of the paper explained, our discussion of practice at the Bank of England illustrates more general themes of criticism and proposals for repair work. We share Wren-Lewis’s regret that the Bank of England and other central banks, with the exception of the US Federal Reserve, abandoned their SEMs before the financial crisis, instead of improving such models by addressing weaknesses such as those discussed in section II below. There are many parallels with developments at other central banks, see the general critique of New Keynesian DSGEs in section III, and connections with the wider macroeconomic debates addressed in the collection of papers in this issue. In our view, the Bank of England quarterly model (BEQM) introduced in 2004, based on the DSGE approach, was a lurch in the wrong direction—see section IV. Despite the failure, not unique to the Bank of England, of BEQM during and after the financial crisis, in 2011 a new DSGE model, the ‘central organising model for projection analysis and scenario simulation’ (COMPASS), was introduced by the Bank. Section V explains how the ‘CAST’ plan (‘COMPASS and suite transition’) has been used for modelling, forecasting, and policy analysis, and provides a critique.

II. The medium-term macro model of 1999 (MTMM)

We begin with the Bank’s 1999 MTMM, updated in 2000, and describe its broad structure, namely IS–LM theory plus a wage-price sector, broadly consistent with New Keynesian theory. As already noted, MTMM was a ‘structural econometric model’, incorporating a good deal of economic logic. The core macro-econometric model consisted of about 20 equations determining key endogenous variables. There were a further 90 or so identities defining relationships between variables, and about 30 exogenous variables whose paths had to be set. GDP was determined in the short term by the components of aggregate demand—private consumption, investment (including inventory investment), government consumption, and net exports. In the longer term, GDP was determined by supply-side factors, which determined potential output. Domestic firms were modelled as producing a single composite good using

² He usefully distinguishes foundational models making a deep theoretical point, DSGE models designed to explore macro implications of a set of distortions, and ‘toy models’ such as RBC or simple IS–LM models.

³ We follow his terminology for now but return to the proper meaning of ‘structural’ in section III.

an aggregate production function of the Cobb–Douglas form. So output was determined in the long run by the evolution of the capital stock, labour supply, and total factor productivity (TFP). These variables were assumed to be unaffected by the price level or the inflation rate. Credit stocks and flows played no role in the model and broad money (M4) was represented by a demand equation that essentially played no further role. The financial sector was not explicitly represented, though there were equations for short and long interest rates.

Price level dynamics and the adjustment of actual output towards potential were broadly determined by the interaction between aggregate demand and supply, augmented by explicit relationships for aspects of wage- and price-setting. Firms set domestic output prices as a cyclically varying mark-up over unit labour costs. Firms were also assumed to set the level of employment, with unemployment defined by subtracting employment from the labour force. Real wages were determined by bargaining in an imperfectly competitive labour market. Inflation expectations had an explicit role in wage determination. Price responses were sluggish, so there was slow adjustment towards both real and nominal equilibria.

However, there were some serious problems, some of which also arise in FRB/US and in other SEMs mentioned above, thus raising generic issues of contemporary relevance. Our discussion is therefore of more than historic relevance. The consumption function had an equilibrium-correction form with the long-run solution given by net worth, labour income, and a real interest rate. Net worth treated cash and other liquid assets, stock market and pension wealth minus household debt, and housing wealth as equally spendable. These are highly implausible restrictions, contradicted by empirical evidence: see sections III and VI below. Short-run dynamics plausibly allowed for negative cash flow effects of higher base interest rates on spending, negative effects from a rise in the unemployment rate—interpreted as a kind of income uncertainty effect—and positive effects from the *current* growth rate of real gross housing wealth, though this is likely to be fairly endogenous. There was no explicit consideration of income growth expectations, in marked contrast to the Federal Reserve’s 1996 FRB/US model: see Brayton *et al.* (1997). Even worse, there was no allowance for the radical liberalization of credit conditions after 1979, although, since house prices are correlated with credit conditions, the growth rate of housing wealth should have picked up some of the short-term consequences.

Moreover, the house price equation itself was grossly mis-specified, taking no account of shifts in mortgage market credit conditions, mortgage rates, and housing supply, as reflected in the stock of housing. The long-run solution depended only on earnings and the 10-year index-linked bond yield, though the latter was not significant ($t=1.2$) and the speed of adjustment was an incredibly slow 3.4 per cent per quarter ($t=2.0$), so the long-run solution was poorly determined.

With these defects, MTMM would not have captured well the credit-fuelled house price boom of the 2000s, nor the related rise in the consumption/income ratio. It would have largely missed the rising vulnerability of consumption to higher consumer debt relative to income. Nor would it have handled well the dramatic credit crunch beginning in 2008 and its

large joint impact on consumption together with the fall in house prices. With the omission of nominal mortgage rate effects from the house price equation and its slow speed of adjustment, it would not have captured adequately the relatively rapid feed-through of monetary policy to UK aggregate demand. Net worth did rise in the period before 2008 and fall afterwards, permitting some linkages in the right direction between asset prices and the real economy, but understating the size and dynamics of the relationship.

The other private-sector demand component, business investment, was also poorly determined, with t-ratios all below 2, and a weak long-run solution. Housing investment was assumed to be proportional to business investment, a heroic assumption. Aggregate business investment is notoriously difficult to model, with expectations of future profits, liquidity, credit constraints for smaller companies, and risk premia, all potentially playing a role. Moreover, with lumpy adjustment costs at the micro level, there may be non-linearities in the response of aggregate investment to drivers: see Bertola and Caballero (1990).

One place where expectations were formally embedded is the wage equation. Staggered wage-setting with some forward-looking elements and considerable nominal inertia make sense, as does the long-run solution, bringing in indirect employment taxes, the price level, productivity, unemployment, and slow structural changes (e.g. the fall in unionization) picked up by the Hodrick–Prescott filter. The treatment of expectations allowed both model-consistent and more backward-looking alternatives, and the coefficients were freely estimated and well determined. The diagnostics for the wage equation look reasonable. The import and export price equations were well determined and had backward-looking behaviour, but with plausible exchange rate transmission. However, the equation for the GDP deflator had a weak long-run solution, just tied to unit labour costs. It is remarkable that international prices and the real exchange rate played no role, nor did property prices, which should feed through slowly into rental costs: see Aron and Muellbauer (2013a) for US evidence. One suspects that the rise in UK inflation after the exchange rate depreciation of 2008–9 would not have been well captured by MTMM.

The other weakness in the price sector was the exchange rate equation itself, driven by the arbitrage condition defined by uncovered interest parity (UIP). There is strong empirical evidence against UIP. Evidence tends to suggest that for small deviations from long-run equilibrium in which purchasing power parity plays an important role, the exchange rate is not far from a random walk, but for large deviations, equilibrium correction is important: see Taylor *et al.* (2001).

As this discussion has illustrated, there was plenty of scope for model improvement within the same broad structure. However, the fashion among macroeconomists and many other central banks had shifted away from this kind of model towards New Keynesian DSGE models.

III. Misplaced origins of DSGEs at central banks

(i) What is ‘structural’?

Concerns about the non-constancy of parameters in models, under changed states of nature, have been a contentious issue in empirical econometrics since Robbins (1932), and discussed (*inter alia*) also by Frisch (1938) and Keynes (1939). Haavelmo’s (1944) classic article on the probability approach in econometrics is the first systematic definition of ‘structural’. Haavelmo contrasts the potential lack of autonomy of empirical regularities in an econometric model with the laws of thermodynamics, friction, and so forth, which are autonomous or ‘structural’ because they ‘describe the functioning of some parts of the mechanism *irrespective* of what happens in some *other* parts’, see Hoover (1994). Simon (1952, 1953) explicitly links the structural (or invariant)/non-structural distinction in terms of the representation of causality. It is the omission of *any* causal relation related to those remaining, not only those deriving from the expectations of economic agents about the functioning of the system that produces non-invariance.

Wren-Lewis (2018) and Hoover (1994) trace how the New Classical Revolution in macroeconomics gained dominance, and why SEMs were displaced as policy models. The Lucas (1976) critique of then current large econometric policy models was the key. Because the parameters of those models were not ‘structural’, i.e. not policy-invariant, they would necessarily change whenever policy (the rules of the game) was changed. However, there was a subtle shift in the meaning of ‘structural’:

Given that *the structure of an econometric model consists of optimal decision rules for economic agents*, and that optimal decision rules vary systematically with changes in the *structure* of series relevant to the decision maker, it follows that any change in policy will systematically alter the *structure* of econometric models. Lucas (1976, p.41).

He called into question the prevailing large-scale econometric models that lacked microfoundations of optimal decision rules of economic agents.

Lucas’s critique can be challenged on five fronts. First, one can question the use of the word ‘optimal’ in defining ‘structure’. Many economists and other social scientists question the information-processing ability of all individuals to make optimal decisions, see Simon (1972) and Kahneman (2003). There is now greater recognition of widespread cognitive biases and the use of heuristic decision rules, see Tversky and Kahneman (1974), and of sheer ignorance, see the evidence on financial illiteracy, Lusardi (2016).

Second, one can challenge Lucas on the nature of econometric models, seen as the aggregation of individualistic decision rules, denying collective decision-making, market externalities, and game-theoretic interactions between agents or groups of agents.

Third, one can challenge the assumption by Lucas of rational or model-consistent expectations, implying the ability of agents to rapidly learn about changes in the economic

environment. As we discuss further below, the rational expectations hypothesis runs into fundamental problems when structural breaks shift the underlying probability distributions of the variables so that the ‘law of iterated expectations’ fails. Indeed, the evidence of the intermittent failures by the most proficient professional forecasters, which are not necessarily due to the Lucas critique, suggests that agents are often not able to extract the ‘structure of series relevant to them’ (see, for example, Ericsson and Irons, 1995). For example, it is doubtful that UK households in 2017 have much idea about the Bank of England reaction function on quantitative easing, while forecasters have, for years, predicted some normalization of UK interest rates and been systematically wrong. Moreover, changes to the parameters of zero-mean variables are difficult to detect, so that changes in policy rules do not necessarily induce the forecast failure claimed by Lucas (see Hendry and Doornik, 1997).

Fourth, one can challenge the claim that *any* change in policy will systematically alter models. Most obviously, if the relevant exogenous policy variable is included in the model, changes in the variable need not change the structure of the model. Sims (1980, p.12) agrees and gives an example of a shift in an exogenous tax rate, previously also subject to shifts. By ‘policy change’ Lucas means a shift in a policy reaction function which would need to be included in the model. Under rational or model consistent expectations, this is liable to result in complex model changes. Even more important than shifts in policy reaction functions, are structural shifts in the economy (globalization, financial innovation, financial regulation and deregulation, the collapse of Bretton Woods, the formation of OPEC cartel, the decline in trade union power, new technologies). These are *not* mostly about changes in reaction functions by forward-looking policy-makers. Most of these changes are effectively exogenous—at least not the product of purposive policy decisions. Paradoxically, this makes it easier to incorporate them in models by conditioning on them, while Lucas’s rational expectations story of policy rule changes is somewhat harder to deal with.

Finally, one can challenge the claim that the Lucas critique has ubiquitous relevance⁴ by checking the empirical evidence of its relevance. There are many tests for the critique as discussed by Favero and Hendry (1992).

The SEMs of the day were attacked not only because they were potentially not ‘structural’, but because of the accusation that they contained ‘incredible restrictions’, partly because, if agents had rational expectations, individual behavioural equations would reflect complex, system-wide properties of the data. As Wren-Lewis notes, the proposition by Sims (1980) that vector autoregressive regressions (VARs) should be used as a way of modelling economic relationships without imposing theoretical restrictions took hold:

Both these developments allowed a bifurcation in academic work, and a degree of methodological purity in either case. Macroeconomists could focus on building models without worrying about the empirical fit of each equation they used, and

⁴ Sims (1980, pp. 13–14) appears to sympathize with this view, arguing that if policy shifts were not rare, agents would marginalize with respect to them, so that they would be part of the dynamics in a vector autoregressive system.

applied time series econometricians could allow the data to speak without the use of possibly dubious theoretical restrictions.

Unfortunately, because of the need to counter ‘the curse of dimensionality’—the explosive increase in the numbers of parameters in a VAR as the number of variables and lags increases—in practice VARs imposed new restrictions. These are on the number of variables and lags it is feasible to include,⁵ and/or via the application of Bayesian priors and/or exclusion restrictions common in ‘structural’ VARs. Moreover, a-theoretical VARs made it harder than in SEMs to formulate a constructive response to the Lucas critique by including relevant modifications of strategic equations. In this respect, the explicit treatment of income expectations in FRB/US, and that model’s flexibility in comparing different formulations of the expectations process was surely preferable to pretending that VARs were somehow immune to the critique. It is also a good example of the fourth of the challenges to the Lucas critique listed above.

Macroeconomists turned to DSGE models claiming to be microfounded in individual optimization and incorporating rational expectations. However, the use of the word ‘structural’ to describe an equation or a subset of a model has been hijacked by followers of Lucas (1976) away from its original meaning of invariant to shifts elsewhere in the system to mean ‘microfounded’. But rationalization in terms of optimizing behaviour does *not* guarantee invariance, nor is it true that invariance always fails in models where such microfoundations are lacking. Central to the claim to be microfounded was a set of further simplifying assumptions, in particular ‘representative agents’ and linear budget constraints, needed to obtain tractable models with what were seen to be desirable steady-state properties. If economies did not deviate too far from such steady-state paths, one could justify linear or log-linear approximations around such paths with a lack of concern for uncertainty even with locally non-linear relationships.

(ii) Microfoundations built on sand

The first element of our view that New Keynesian DSGE models had the wrong microfoundations concerns the representative agent assumption. As is well known, the conditions for exact aggregation of demand functions so that aggregate behaviour corresponds to that of an average household, are very restrictive: demands are linear functions of income and wealth, with households sharing the same marginal propensities. With optimizing behaviour under linear budget constraints, preferences need to have a very specific form, see Deaton and Muellbauer (1980, ch. 6), though, as Muellbauer (1976) showed, there is a more general notion of a representative household than ‘average’, consistent with slightly more general but still very restrictive preferences.⁶ With heterogeneous credit constraints across households, even such restrictions on preferences

⁵ In the context of US inflation models, Aron and Muellbauer (2013a) find that standard methods of choosing lag lengths systematically discard important longer-run information, and that this is true for all information sets considered.

⁶ In this generalization, the distributions of income and wealth, as well as averages, affect aggregate behaviour, see discussion in Deaton and Muellbauer (1980, ch. 6.2).

would be of no help in obtaining exact aggregation. Kirman (1992), Carroll (2000), Hoover (2001),⁷ and Stiglitz (2018) are among many criticizing the representative agent assumptions of RBC and New Keynesian macroeconomics.

Instead of representative agent economics, stochastic aggregation theory⁸ suggests we can often still make good progress with aggregate data even if behaviour at the micro-level looks different from an aggregate model. An excellent example is Houthakker (1956), who showed that a Leontief production function, with no substitution, and Pareto distribution of the parameters at the micro-level, implied substitution at the macro-level as if it arose from an aggregate Cobb–Douglas technology. The functional form at the micro-level could hardly differ more radically from that at the macro-level. In the literature on lumpy adjustment costs, micro behaviour switches discretely from no adjustment to adjustment when some micro-thresholds are reached. In the aggregate, however, behaviour is smooth, as explained by Bertola and Caballero (1990). A recent applied example of stochastic aggregation comes from models of aggregate mortgage delinquency and foreclosure rates (Aron and Muellbauer, 2016). A key driver is the proportion of mortgages with negative equity: if the distribution of mortgage debt to equity is fairly stable, a shift in the ratio of average debt to average equity shifts non-linearly the fraction of borrowers with negative equity.⁹ Since bad loans restrict the ability of banks to extend new credit, negative equity is an important non-linear element in the business cycle feedback loop.

The second element of our critique of the microfoundations of New Keynesian DSGE models is of their adoption of the complete markets paradigm, implicitly denying the asymmetric information revolution of the 1970s, see the quotations from Buiter (2009) and Muellbauer (2010) in section I. In these DSGEs, households discount temporary fluctuations in income to maintain spending in the face of shocks, thus providing a stabilizing anchor to the economy, in turn justifying the rational expectation that shocks will prove temporary.

This old-fashioned textbook view of consumption behaviour was challenged by Deaton (1991), Carroll (1992, 1997, 2001), and Aiyagari (1994). Given uninsurable individual income risk and liquidity constraints, the result of asymmetric information, they show that households engage in buffer-stock behaviour to ameliorate income risk and discount expected future income at higher rates than assumed by the textbook model. Moreover, given heterogeneous income processes, heterogeneous liquidity constraints, and heterogeneous

⁷ Hoover says of representative agent models: ‘While they appear to use the mathematics of microeconomics, the subjects to which they apply that microeconomics are aggregates that do not belong to any agent. There is no agent who maximizes a utility function that represents the whole economy subject to a budget constraint that takes GDP as its limiting quantity. This is the simulacrum of microeconomics, not the genuine article.’

⁸ In contrast with exact linear aggregation, in stochastic aggregation, assumptions on the joint distributions of the data allow aggregate behaviour to be represented by parameters of the distributions, such as means, variances, and covariances.

⁹ The distribution of debt/equity is approximated by logistic function defined on a cubic in the debt/equity ratio. The mean debt to mean equity ratio, with slight trend adjustment, is then used to generate estimates of the UK proportion of borrowers with negative equity, consistent with cross-section snapshots. Negative equity is highly significant in explaining aggregate defaults.

asset ownership, there will be considerable heterogeneity in the discount rates used by different households. On average, discount rates applied to expected incomes will be far higher than those of the textbook model.

This has profound implications, as the important paper by Kaplan *et al.* (2016) demonstrates. They contrast two general equilibrium models: a representative agent New Keynesian (RANK) model and a heterogeneous agents New Keynesian (HANK) model, and show that the monetary policy channel works quite differently in the latter. An important feature of their model, which is shared with their earlier papers on fiscal policy in the context of wealthy ‘hand-to-mouth’ consumers (Kaplan *et al.*, 2014; Kaplan and Violante, 2014), is that consumers own not only buffer stocks in the form of liquid assets but also illiquid assets, typically earning higher long-run returns. However, there are lumpy transactions costs in trading in and out of such assets and households face borrowing limits.

To keep the HANK model tractable, Kaplan *et al.* (2016) adopt a highly simplified view of housing. A heterogeneous agent model which incorporates somewhat more realistic features of housing and credit markets with important consumption and monetary transmission implications has been developed by Hedlund *et al.* (2016). Both papers imply that since heterogeneous households, facing idiosyncratic micro uncertainty and radical macro uncertainty, discount income expectations with much higher weights on near-term expectations, then aggregate behaviour cannot be adequately approximated by RANK-style models.

(iii) The omission of shifting credit constraints, household balance sheets, and asset prices

The asymmetric information revolution of the 1970s provided microfoundations for the application of credit constraints by the banking system; see Stiglitz (2018) for further discussion. In many countries, shifts in these constraints were among the most important structural changes in the economy; see the example of the US discussed below. Thus, a third criticism of New Keynesian DSGE models, linking closely with the previous, is the omission of debt and household balance sheets, including housing, crucial for understanding, together with shifts in credit availability, consumption, and macroeconomic fluctuations. The US Federal Reserve did not abandon its large non-DSGE econometric policy model FRB/US, but it too was defective in that it also relied on the representative agent permanent income hypothesis, which ignored shifts in credit constraints and mistakenly lumped all elements of household balance sheets, debt, liquid assets, illiquid financial assets (including pension assets), and housing wealth into a single net worth measure of wealth. Because housing is a consumption good as well as an asset, consumption responds differently to a rise in housing wealth than to an increase in financial wealth; see Aron *et al.* (2012). Second, different assets have different degrees of ‘spendability’. It is indisputable that cash is more spendable than pension or stock market wealth, the latter subject to asset price uncertainty and access restrictions or trading costs. This suggests estimating separate marginal propensities to spend out of liquid and illiquid financial assets. Third, the marginal effect of debt on spending is

unlikely just to be minus that of either illiquid financial or housing wealth. The reason is that debt is not subject to price uncertainty and it has long-term servicing and default risk implications, with typically highly adverse consequences.

There is now strong micro evidence that the effect of housing wealth on consumption, where it exists, is much more of a collateral effect than a wealth effect, see Browning et al (2013), Mian et al (2013), Windsor et al (2015), Mian and Sufi (2016) and Burrows (2017). The economics of this are further discussed in section VI below. As mortgage credit constraints vary over time, this contradicts the time-invariant housing wealth effect embodied in FRB/US.

The importance of debt was highlighted in the debt-deflation theory of the Great Depression of Fisher (1933).¹⁰ Briefly summarized, his story is that when credit availability expands, it raises spending, debt, and asset prices; irrational exuberance raises prices to vulnerable levels, given leverage; negative shocks can then cause falls in asset prices, increased bad debt, a credit crunch, and a rise in unemployment.

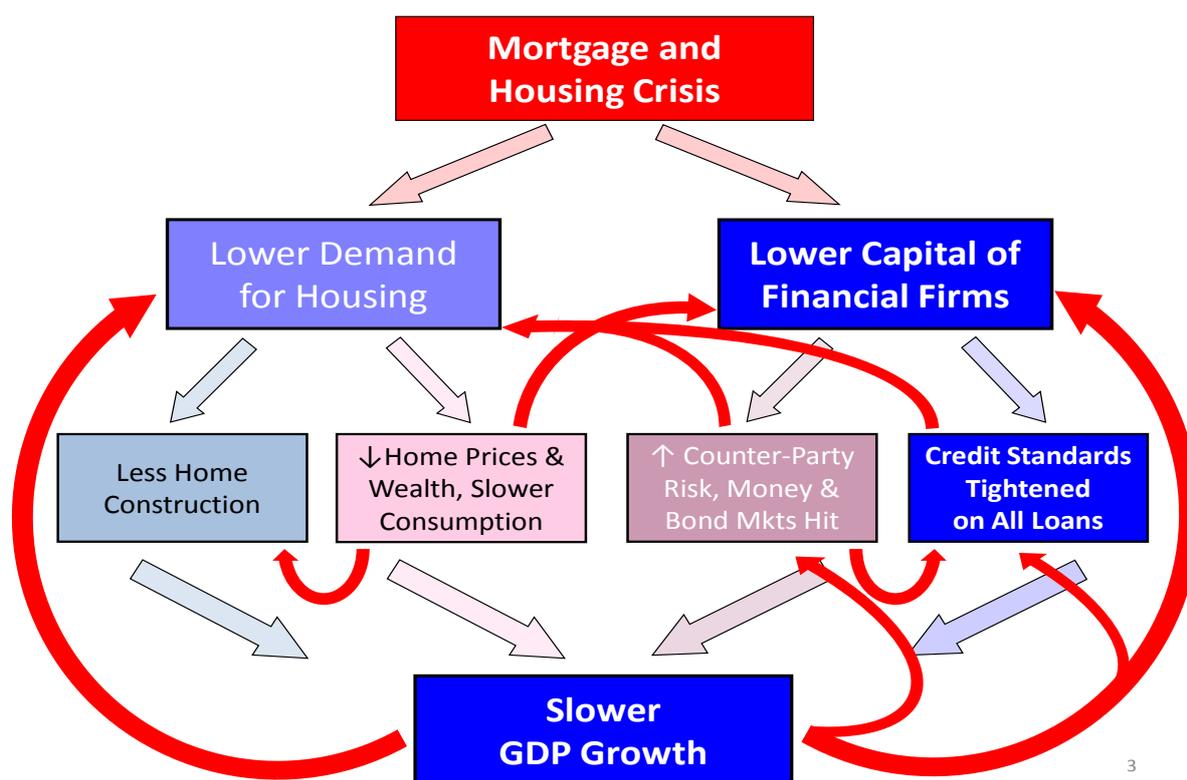
Of structural changes, the evolution and revolution of credit market architecture is often the single most important. In the US, credit card ownership and instalment credit spread between the 1960s and the 2000s. The government-sponsored enterprises—Fannie Mae and Freddie Mac—were recast in the 1970s to underwrite mortgages. Interest rate ceilings were lifted in the early 1980s. Falling IT costs transformed payment and credit screening systems in the 1980s and 1990s. More revolutionary was the expansion of sub-prime mortgages in the 2000s—driven by the rise of private label securitization backed by credit default obligations (CDOs) and swaps. The 2000 Commodity Futures Modernization Act (CFMA) made derivatives enforceable throughout the US with priority ahead of claims by others, e.g. workers, in bankruptcy. This permitted derivative enhancements for private label mortgage-backed securities (PMBS) so that they could be sold on as highly rated investment grade securities. A second regulatory change was the deregulation of banks and investment banks. In particular, the 2004 Securities and Exchange Commission (SEC) decision to ease capital requirements on investment banks increased leverage to what turned out to be dangerous levels and further boosted PMBS; see Duca *et al.* (2016). Similar measures to lower required capital on investment grade PMBS increased leverage at commercial banks also. These changes occurred in the political context of pressure to extend credit to poor.

(iv) The missing financial accelerator

¹⁰ In recent years, several empirical contributions have recognized the importance of the mechanisms described by Fisher (1933). Mian and Sufi (2014) have provided extensive microeconomic evidence for the role of credit shifts in the US sub-prime crisis and the constraining effect of high household debt levels. Focusing on macro-data, Turner (2015) analyses the role of debt internationally with more general mechanisms, as well as in explaining the poor recovery from the global financial crisis. Jordà *et al.* (2016) have drawn attention to the increasing role of real estate collateral in bank lending in most advanced countries and in financial crises.

In the 1980s and early 1990s, major credit market liberalization had occurred in Norway, Finland, Sweden, and the UK, causing credit, house price, and consumption booms which were followed by busts—precursors of the US sub-prime crisis. In the financial accelerator feedback loops that operated in the US sub-prime crisis, falls in house prices increased bad loans and impaired the ability of banks to extend credit. As a result, household spending and residential investment fell, increasing unemployment and reducing incomes, feeding back further on to lower asset prices and credit supply. The following diagram, due to John Duca (see Duca and Muellbauer, 2013), illustrates the feedback loops in the US sub-prime crisis.

Figure 1: The financial accelerator in the US sub-prime crisis



These feedback loops involve non-linearities and amplification. For example, falls in house prices, driving up the incidence of negative equity, cause, via bad loans, a sharper contraction in credit availability than rising house prices cause an expansion of credit availability. Moreover, a contraction in credit availability itself feeds back on to lower house prices. The combination of lower credit availability, which lowers the spendability of housing collateral, even at given house prices, and lower house prices, had a multiplicative effect in lowering consumption in the US sub-prime crisis; see Duca and Muellbauer (2013).

Such mechanisms were entirely missing in New Keynesian DSGE models, and hardly represented in those DSGE models, such as Bernanke *et al.* (1999) and Christiano *et al.* (2003), which incorporated a financial accelerator only for firms. Iacoviello (2005) and the estimated DSGE model of Iacoviello and Neri (2010) did introduce housing into DSGE. They assume two representative households, patient and impatient, present in a fixed proportion. Patient households apply a loan-to-value constraint when offering mortgage loans to the

impatient households, a kind of financial friction. But because of the assumption of infinitely lived or dynastic households, saving for a down-payment, one of the most important saving motives in industrial countries, is omitted. In their closed economy model, without banks and foreclosures, and assuming a frictionless and efficient housing market, transmission and amplification of monetary or other shocks via housing is *extremely* limited. For example, their model implies that aggregate home equity withdrawal, the excess of households' mortgage borrowing over acquisitions of housing, is always negative. In practice, US home equity withdrawal was *strongly positive* for much of the period from 2001 to 2006, and in the peak quarters was of the order of 10 per cent of that quarter's household income. However, this fact, and the realized foreclosures, were not in the set of salient data chosen by Iacoviello and Neri for their model calibration. Indeed, for their calibrated model, they compare the correlation between consumption growth and house price growth with and without the financial friction. Without the friction, the correlation is 0.099, the result of the common influence of the shocks¹¹ on house prices and consumption. With the friction, the correlation rises to 0.123. One would be tempted from this to conclude, but quite wrongly, that financial frictions have little impact on the macroeconomy. This is the opposite of what Figure 1 above implies.

(v) Rational expectations

The world is usually in disequilibrium: economies are wide-sense non-stationary from evolution and sudden, often unanticipated, shifts both affecting key variables directly and many more indirectly. Technology, globalization, both in trade and in finance, trade union power, credit conditions, monetary and fiscal policy rules and other legislation, social mores, skills, wars, resource and financial crises, climate, demography, health and longevity, and income and asset distributions all change over time. These, and other innovations keep perturbing the economic system in ways that not even rational individuals can foresee (Uber and Airbnb are two recent examples). DSGEs therefore suffer a double flaw: they are making incorrect assumptions about the behaviour of the agents in their model, and are also deriving false implications therefrom by using mathematics that is invalid when applied to real economies, as we now discuss. For formal proofs, see Hendry and Mizon (2014), which we now summarize.

Structural changes are a key source of forecasting error, as noted above. However, the mathematical basis of DSGEs fails when events suddenly shift the underlying distributions of relevant variables. The 'law of iterated expectations' becomes invalid because an essential, but usually unstated, assumption in its derivation is that the distributions involved stay the same over time. Economic analyses with conditional expectations ('rational expectations') and inter-temporal derivations then also fail, so DSGEs become unreliable when they are most needed.

¹¹ The major shock driving real house prices is a 'preference' shock, which Romer (2016) ironically terms a 'caloric' shock in contrast to the 'phlogiston' of productivity shocks, the major driver of real residential investment in their model.

To explain the law of iterated expectations, consider a very simple example – flipping a coin. The conditional probability of getting a head tomorrow is 50%. The law of iterated expectations says that one’s current expectation of tomorrow’s probability is just tomorrow’s expectation, i.e. 50%. In short, nothing unusual happens when forming expectations of future expectations. The key step in proving the law is forming the joint distribution from the product of the conditional and marginal distributions, and then integrating to deliver the expectation.

The critical point is that none of these distributions is indexed by time. This implicitly requires them to be constant. The law of iterated expectations need not hold when the distributions shift. To return to the simple example, the expectation today of tomorrow’s probability of a head will not be 50% if the coin is changed from a fair coin to a trick coin that has, say, a 60% probability of a head.

To explain more formally, failures of the intertemporal law of iterated expectations arise because, by definition:

$$E_{f_t}[E_{f_t}[y_{t+1} | y_t]] = [\int_{y_t} \int_{y_{t+1}} y_{t+1} f_t(y_{t+1} | y_t) dy_{t+1}] f_t(y_t) dy_t$$

which is not equal to $E_{f_{t+1}}[y_{t+1}]$ when $f_{t+1}(\cdot)$ is not equal to $f_t(\cdot)$. Here $f_t(\cdot)$ is the density of the distribution at t and E_{f_t} is the expectation using that distribution.

Conversely, for the law of iterated expectations to hold over time, all of the distributions involved must remain the same, yet shifts in endowments, asset allocations, policy variables, longevity, etc., are occurring intermittently. Worse, the largest shifts, as in a financial crisis, cause the greatest failures of models based on falsely assuming that the law of iterated expectations holds. When sensible economic agents realize that inter-temporal planning inevitably forces them to revise after changes, then DSGE models now additionally impose the wrong behaviour for the agents. The key implication is that models based on the law of iterated expectations are intrinsically non-structural, and hence are not a viable basis for economic policy: they are susceptible to the Lucas critique for all policy changes that shift the distributions of either policy instruments or targets, yet most policy interventions change means or variances of relevant variables.

Rational expectations (RE) are asserted to correspond to the conditional expectation of the future, given all available information formed from the actual current distribution of variables, so are in fact irrational when distributional shifts occur. The failure to realize this major problem for RE derives in part from the sloppy mathematical derivations that are all too common, as if forgetting that expectations are integrals over the associated distributions, as explained above.

Building on the well-known theorem that in constant distributions, the correct conditional expectation would be the minimum mean-square error (MMSE) predictor, by not indexing the expectations operator for the distribution over which the integral is being calculated, the MMSE result is supposedly ‘proved’ in many settings where the calculus is completely

inapplicable. This includes most DSGE derivations, but especially applies to so-called New-Keynesian Phillips curves (NKPCs) where the expression:

$$y_{t+1} = E[y_{t+1}/I_t] + v_{t+1}$$

is written, such that taking conditional expectations of both sides apparently proves that:

$$E[v_{t+1}/I_t] = 0,$$

and hence that agents can form an unbiased predictor of y_{t+1} at time t . Using an undated and unindexed conditional expectations operator makes it seem that the error at $t+1$ is unbiasedly predicted. At time t , agents cannot predict the random component v_{t+1} in advance, but that does *not* establish that the expectation is unbiased for y_{t+1} because the underlying distribution can shift. The apparent proof relies on not writing E with a subscript for the distribution integrated over $f_t(\cdot)$, to clarify that *today* we expect a zero-mean error. Written correctly:

$$y_{t+1} = E_{f_t}[y_{t+1}/I_t] + v_{t+1},$$

and hence that:

$$E_{f_t}[v_{t+1}|I_t] = 0.$$

Unfortunately, an E with the subscript $E_{f_{t+1}}$ is needed to prove that expectations are actually unbiased, which would be a crystal-ball predictor, already knowing the entire distribution of y_{t+1} at time t . Castle *et al.* (2014) show that the NKPC is an artefact of not handling shifts, which are then proxied by substituting the actual future value y_{t+1} for $E[y_{t+1}/I_t]$. The jumps in future values of variables induced by RE are implausible outside of ‘fully efficient’ speculative markets, which probably do not exist, noting that an efficient market implies that future price changes are unpredictable, but the converse does not follow: that future price changes are unpredictable does not entail that a market is efficient—it could be completely erratic.

The NKPC effectively takes the forward second difference of the log price level. High degrees of differencing make data that are non-stationary or subject to shifts look stationary. Omitted variables or shifts in an inflation model push the estimated dynamics towards finding such a differenced specification. Inflation, as seen in the dynamics of an aggregate price index, is almost certainly influenced by relative price adjustment, missing in simplifying assumptions of one good or symmetric goods in New Keynesian models. Evidence of such relative price adjustment and of shifts in inflation dynamics in the US comes from Aron and Muellbauer (2013a).

Another example of the consequences of specification errors comes from Leeper *et al.* (2017) who use Bayesian methods to estimate a DSGE model broadly of the Smets and Wouters (2007) type for US quarterly data from 1955 to 2014. The utility function incorporates external habits and, with relatively diffuse priors, the data suggest values of the habit parameter close to 1, implying that forward-looking households derive utility from changes

rather than levels of consumption and leisure. This is close to second differencing the data for forward-looking households. As the data set excludes not only credit, asset prices, and shifts in credit constraints and other regime shifts, but also oil prices, the estimates are pushed towards second differencing to cope with these omissions. The estimated degree of habit formation is a symptom of mis-specification.

(vi) Misunderstanding forecast failure

A failure to understand forecasting theory also seems partly responsible for the false belief that ‘theory-based models’ are essential. A typical assertion comes from Coletti *et al.* (1996) at the Bank of Canada:

The inability of relatively unstructured, estimated models to predict well for any length of time outside their estimation period seemed to indicate that small-sample econometric problems were perhaps more fundamental than had been appreciated and that too much attention had been paid to capturing the idiosyncrasies of particular samples.

The empirically relevant theory in Clements and Hendry (1995) allows for the forecasting model to be mis-specified for the data generation process, with parameters estimated from inaccurate observations, on an integrated-cointegrated system, intermittently altering unexpectedly from structural breaks. That theory helps explain the prevalence of forecast failure, can account for the results of forecasting competitions, and explains some of the good performance of so-called ‘consensus’ forecasts. However, it also demonstrates that in a stationary world, forecast failure cannot be shown to be necessarily due to ‘poor econometric methods’, ‘incorrect estimation’, or ‘data-based model selection’, although such mistakes will increase interval forecasts: shifts relative to the prevailing situation are needed for forecast failure. Claiming that forecasts are based on an ‘economic theory based model’, howsoever ‘structured’, will not by itself counter any of the causes of forecast failure and associated policy failure unless it can anticipate and model shifts.

Intercept corrections (ICs) in equilibrium correction models discussed by Clements and Hendry (1996) are a way of getting any forecast ‘back on track’ *after* a location shift, irrespective of its source, and assuming it is a location shift (so relatively permanent). ICs are not a magic bullet—‘forecast’ failure will still occur—unless they represent crystal-ball information about the shift that is going to occur (e.g. knowing a dock strike is imminent). But they can help avoid systematic failure after shifts, if appropriately implemented—as can other devices. After forecast failure, there are still non-trivial issues about *how* to do intercept correction, e.g., how to separate temporary from permanent components of shifts. There are parallels here with the difficulty forecasters have had in deciding whether and to what degree the UK’s capacity output had permanently declined.

In DSGEs an approximate equivalent to intercept correction would be to add the previous period's residual to every equation, which would reduce systematic mis-forecasting resulting from a structural break or from the omission of non-stationary variables. Even a poor model that, in Charles Goodhart's words, 'excludes everything I am interested in' (cited in Buiter, 2009) can thus be protected from the worst consequences of forecast failure, supporting Blanchard's distinction between policy models and forecasting models.

(vii) The lack of flexibility of DSGEs

A number of authors in this issue, including Lindé, Reis, McKibbin and Stoeckel, and Wren-Lewis, suggest that DSGEs are a flexible tool. Indeed, one can adapt a closed economy DSGE model for an open economy or add a terms-of-trade shock to a model (as does the Bank of Canada's TOTEM). One can add a fixed fraction of myopic households or a risk premium. One can add a housing sector to a model which previously did not include one. But as our discussion above of Iacoviello and Neri (2010) indicates, that particular effort probably *subtracted* from understanding because of the efficient market, rational expectations, representative agent framework where the consumption Euler equation is the crucial link between the present and the future.

The consumption Euler equation is thus the key mechanism for the operation of model consistent-expectations. This makes it the main *straitjacket* of the representative agent DSGE approach. In section VI we advocate its replacement by a solved-out 'credit-augmented' consumption function, incorporating the discounted present value of future incomes, using an average discount rate far higher than in standard textbook permanent income models. Such a replacement has fundamental implications as explicit expectations mechanisms are then needed for the other behavioural equations, also of the solved-out form. This allows a much more modular approach, as for example in FRB/US, allowing heterogeneity in expectations between households and firms. Within DSGE models resting on an aggregate or sub-aggregate Euler equation, this kind of modularity is hard to achieve. The Bank of England's BEQM and its successor COMPASS struggled with these issues, as we now explain.

IV. The Bank of England Quarterly Model of 2004 (BEQM)

In his review of modelling for the Bank of England, Pagan (2003) claimed a trade-off between theory-consistency and empirical relevance; see Wren-Lewis (2018) for a picture and discussion. He notes defects in the ability of MTMM to track the falling saving ratio of the 1980s and in the wage-price system to match the inflation data. But he notes that perhaps the biggest problem with MTMM was not being 'state of the art', presumably in theory-consistency. Bluntly, MTMM was no longer fashionable as it was not sufficiently in line with the New Keynesian perspective on the 'science of monetary policy'; see Clarida *et al.* (1999) for a summary of the contemporary view. The 'optimal' point on the trade-off between theory and evidence posited by Pagan is chosen by a tangent to the highest envelope, determined by

the preferences of the policy-makers. There are some qualifications and difficulties with Pagan's concept.

- (i) A unique, universally agreed theory (an 'agreed-upon conception of the way in which the economy is thought to function' in Pagan's own words) is essential if the 'theory consistency' of models is to be meaningful enough to even draw the axes on the diagram, as otherwise a different trade-off measure is needed for every theory class. But theories conflict. For example, textbook theory, which assumes efficient and close-to-complete markets, well-informed homogeneous agents, little uncertainty, no credit or liquidity constraints, and a stable economy, contrasts with theory that takes account of the asymmetric information revolution of the 1970s. Many economists, including Stiglitz (2018) as well as ourselves, argue that relevant theory should incorporate credit and liquidity constraints, incomplete markets with incomplete insurance and high levels of individual and macro uncertainty, as discussed in sections V and VI. If the data fit better an empirical approximation to the second theory than an empirical approximation to textbook theory, what meaning can be given to 'theory consistency'?
- (ii) Theories evolve, delivering new insights, thereby requiring a redefinition of models being 'consistent' each time. Pagan, indeed, acknowledges that the frontier evolves.¹²
- (iii) Even a single fixed theory does not imply a unique model: it matters how theory models are implemented empirically, how unobservables in the theory are measured, and how expectations are treated.
- (iv) Empirical coherence is not unique either: alternative data measures exist, revisions occur, estimation uncertainty is rarely accurately calculated, and the selection criteria for empirical models differ.
- (v) Most fundamentally, the diagram suggests theory consistency is bound to *reduce*, rather than improve, empirical coherence and suggests non-evidential preferences could legitimize falsified theory, making macroeconomics in danger of being unscientific, when the 'microfoundations' aim was precisely the opposite. In other words, the trade-off precludes achieving both theory consistency and empirical relevance, implying that empirically relevant macroeconomic theories cannot be constructed.¹³

BEQM, see Bank of England (2004) and Harrison *et al.* (2005), was claimed to be an advance on MTMM. Its core is a DSGE model in which rational expectations play a major role. This generates long-run steady-state solutions for the key endogenous variables, obtained by solving the core model forward. In the long run, all variables in the model settle

¹² He also says: 'At any point in time, there will be a frontier of "best-practice" models that shows the combinations of empirical and theoretical coherence that are attainable. There is no precise way of determining this frontier but sometimes opinions form about what is on and off the frontier.'

¹³ The notion of a trade-off between theory and empirical coherence is hard to sustain in physics, taking Newton's Laws as an example.

on paths that grow consistently with each other in a sustainable equilibrium. Short-run dynamics are handled by embedding these long-run solutions in equilibrium-correction equations in which short-run frictions can explain temporary deviations from the long-run solutions. It is essential to the logic of the approach that only *stationary* variables can represent such frictions, which do not affect the long-run paths.

The household sector is central to the whole model. A Blanchard–Yaari model of inter-temporal choice aggregates over different households to a representative agent form of a life-cycle consumption function driven by permanent labour and transfer income and net financial wealth. The equilibrium-correction form of the consumption function is fairly similar to that in MTMM, with four key differences in the long-run solution. First, the measure of consumption excludes housing. Second, income is replaced by permanent income. Third, net financial wealth replaces net worth including housing. Finally, theory implies a long-term interest rate response of consumption which is calibrated, not estimated as in MTMM.

Given the slightly different definition of consumption, the equation standard errors are not exactly comparable, but 0.01 in BEQM compared to 0.006 in MTMM suggests a considerable worsening of the match of the consumption function to the data. The slower speed of adjustment, 0.12 in BEQM vs 0.17 in MTMM, also suggests an even less well-functioning long-run solution. This is confirmed by the t-ratio of 1.5 on net financial assets in BEQM vs 3.6 on net worth in MTMM. Simulations suggest that the response of consumption to the short-term interest rate, mainly imposed by assumption rather than tested, was a good deal stronger than in BEQM; see Bank of England (2004). Though there is a treatment of expectations consistent with simple textbook theory, the indications are that compared to MTMM, BEQM is even less data-coherent. Also, as in MTMM, BEQM’s assumption that liquid assets minus debt, and pension and stock market wealth are equally spendable and that shifts in credit conditions are irrelevant is empirically untenable. And it seems that BEQM’s assumption of a zero marginal propensity to spend out of housing wealth does even more violence to the data over this period than MTMM’s assumption that the marginal propensity to spend out of housing wealth is the same as for other assets.

Poor though the fit of MTMM’s investment equation was, BEQM’s is even worse. Further, according to the logic of the model, the gaps between key variables and their long-run solutions are supposed to be stationary. However, augmented Dickey–Fuller (ADF) tests show evidence of non-stationarity for consumption, investment, and exports; see Harrison *et al.* (2005, p.116). These results for the three major components of aggregate demand do not instil confidence in the approach. The worse fit for these equations compared to MTMM is repeated in many other components of BEQM, for example in the wage equation. Simulations in Bank of England (2004) suggest that the response of CPI inflation to a rise in interest rates is much more rapid in BEQM than in MTMM, suggesting a larger role for forward-looking expectations and less nominal inertia in BEQM. The circumstantial evidence from the estimated equations is that the data do not support such a rapid response.

V. The Bank's COMPASS model and CAST: its structure and a critique

(i) The structure of COMPASS

If BEQM was a lurch further away from data coherence than MTMM, the introduction in 2011 of the Central Organising Model for Projection Analysis and Scenario Simulation, COMPASS, could be regarded as another milestone on this road. However, there was an explicit recognition that the core model was incomplete and that more weight would be put on the 'suite of models' for complementary input into the policy-making process. BEQM was felt to be too complicated and cumbersome for practical advice to the Monetary Policy Committee (MPC), so a simpler DSGE model was sought with more informal methods of introducing practical frictions than the core/non-core form of BEQM. It is interesting to note that the worse fit of BEQM compared to MTMM, and the fundamental tests rejecting the stationarity of the deviations between key endogenous variables and their long-run solutions, do not seem to have been an important factor in the search for a new model.¹⁴ However, BEQM's 'breakdown' during the financial crisis may have been a consideration.

The basic idea was to supplement the new DSGE model by stationary 'wedges' representing frictional deviations from the fundamental model using, in a fairly informal way, some information from the Bank's 'suite of models'. It could be said that for true believers, this had the advantage of protecting the core model from potentially contradictory empirical evidence. It seems that, unlike in BEQM, no formal tests of the stationarity of the 'wedges' have been carried out, perhaps because various de-trending procedures were thought to have removed the issue, or because of restrictions imposed in the 'suite'. However, the Bank has published model details and forecast evaluation data and encouraged external evaluation and critique, e.g. in the Stockton Review of 2013, and by a further external review in 2016 of the Bank's forecast infrastructure, the review of CAST (COMPASS and Suite Transition) to which the authors of the present paper contributed. The Bank also established in 2014 its internal Independent Evaluation Office, see Ashley and Paterson (2016), which examined forecast performance in some detail; IEO (2015).

We now describe in a little more detail how COMPASS, in conjunction with the Bank's suite of models, has been used for modelling, forecasting, and policy analysis. In 2011, Bank staff introduced this new forecasting platform to assist the MPC in producing its quarterly economic projections. The set-up, which is still in use, comprises a DSGE model supported by a suite of other models. The latter serve three purposes: they provide forecast profiles for sectors of the economy that are absent from COMPASS ('missing channels'); they provide complementary sectoral analysis as a cross-check on the plausibility of COMPASS profiles;

¹⁴ Apart from coherence with theory and data, three other criteria were listed (Burgess *et al.*, 2013, p. 5): 'Tractability. The central organizing model should be easy to use, easy to understand, reliable and robust. Flexibility. It should be possible to examine easily the implications of alternative economic assumptions (e.g. the implications of different parameter values) on the behaviour of the central organizing model. Comprehensiveness. The central organizing model should provide adequate coverage of the key economic mechanisms.'

and they inform the calibration of COMPASS responses on occasions when the unadulterated profiles seem implausible.

Burgess *et al.* (2013) explain how the Bank's forecasting platform operates, drawing on the Chari *et al.* (2007) business cycle accounting (BCA) framework as a primary justification for integrating 'off-model' information into the model. They argue that it is possible to represent a wide variety of models in terms of a very simple 'prototype' RBC model that contains stochastic disturbances to the equilibrium conditions (known as 'wedges'). However, the 'wide variety of models' referred to are all within the DSGE class with various added frictions, such as adjustment costs, labour hoarding, search and match labour markets, and even simple forms of financial frictions. In this way of thinking, externally computed 'shocks' derived from the suite of models could be interpreted as determining temporary stochastic disturbances to the underlying equilibrium conditions implied by COMPASS.

COMPASS contains 13 key domestic economic variables, including output, consumption, investment, government spending, exports, and imports. Employment is measured by total hours of work, and prices include export and import prices, CPI, nominal wages, the short-term interest rate and the sterling exchange rate. World prices and world output are also modelled, making 15 endogenous variables in all. These are subject to 18 'structural' shocks. An important early aim was to use shock decompositions to help the MPC better understand the nature of the shocks hitting the economy, though this has proved more difficult and less useful than anticipated. One innovation relative to BEQM was to introduce a time-varying risk premium shock. Note that money, credit, house prices, and other asset prices or balance sheets play no role. So the risk premium shock is one of the few ways in which any connection with the global financial crisis could be made.

Any macroeconomic policy model for an open economy needs to embody a production function, and relationships which allocate output between domestic and foreign purchasers, and demand between imports and domestic production. The simple assumptions built into COMPASS do not seem particularly controversial on this score, though the lack of variable labour and capital utilization, and hence of short-run labour hoarding, make it harder to interpret productivity data within COMPASS. The random walk assumption for total factor productivity, one might have thought, should increase the degree of uncertainty faced by economic agents, including policy-makers. This makes it particularly odd that precautionary, buffer-stock behaviour and the shorter-time horizons induced by uncertainty play no role, as further discussed below.

Apart from the production function, budget constraints, and accounting identities, the behavioural relationships in COMPASS are Euler equations for consumption, and first-order conditions for optimal investment and employment, and pricing equations embodying some stickiness combined with an important role for expectations. The data are detrended and Bayesian estimation is used to obtain structural parameter estimates. Iterative forward solution methods are used to obtain forecasts.

(ii) Forecasting failures

The failures during the global financial crisis of the DSGE-based model BEQM, adopted by the Bank in 2004, have been well documented, as the *Daily Telegraph* reported in August 2010: <http://www.telegraph.co.uk/finance/economics/7935732/Bank-of-England-overhauls-forecast-model-after-errors.html>. However, many other central banks, the International Monetary Fund (IMF), and the Organization for Economic Cooperation and Development (OECD) reported similar failures during the global financial crisis.

That article quotes Michael Saunders, now on the MPC: '(BEQM) misses the role of credit and financial conditions'; 'Liquidity strains and money market dysfunctions have no role in BEQM'; 'The need for a major overhaul is quite large.' The article reports on the large sums the Bank was spending on the new model, which came into operation in 2011. But, as noted above, COMPASS has a similar DSGE structure to BEQM, though it is simpler and easier to operate. Its 15 endogenous variables also exclude credit and financial conditions.

Fawcett *et al.* (2015a) examined what the forecast performance of COMPASS would have been, working with real-time data, had it been in existence in 2007: it would also have failed very badly over the financial crisis.¹⁵ Common sense about the economy, using a wider set of information, as embodied in private-sector expectations and by the MPC, forecasted better than COMPASS. In their Bank Underground blog, Fawcett *et al.* (2015b) write:

forecasts of UK GDP growth from standard DSGE models fare particularly badly during and after financial crises, and possibly in the presence of structural breaks more broadly. An obvious conclusion to draw is that these models might perform better if they contained meaningful financial frictions.

In defence, it is argued that COMPASS was meant to be complemented by ancillary models and by MPC judgement. However, it is questionable whether the way in which information from the suite of models has been combined with COMPASS has adequately compensated for the failings of COMPASS both as a forecasting platform and for understanding the economy. As ex-MPC member David Miles says (private communication):

In my time (early 2009 to autumn 2015) the MPC was pretty well aware of the enormous limitations and weaknesses of COMPASS. I think the 'suite of models' was a way of trying to do an emergency repair job on a machine that was not on its own of great use—particularly in 2009 and 2010 when credit conditions were at times dire. I think most of us learned much more from talking to companies up and down the country than from what the models were producing.

¹⁵ They also compare GDP and inflation forecasts for 2001–13 produced by COMPASS with MPC forecasts incorporating judgement as published in the *Inflation Report* and forecasts from a statistical suite of forecasting models including VARs. There is no outright winner for both GDP and inflation over the entire period and all horizons. COMPASS forecasts are better for inflation beyond a year ahead, but worse for GDP at all horizons and for short-term inflation.

Fawcett *et al.* (2015b) argue:

It should also be noted that forecast accuracy in and of itself is not the only relevant criterion for assessing the utility of a structural model like COMPASS, which also has a role to play, for example, in providing scenario analysis.

However, if the scenario analysis is actively misleading, a model could actually be damaging understanding of the economy and policy capability, as investigated by Castle, Hendry and Martinez (2017).

(iii) Problems with business cycle accounting interpretations

In the business-cycle accounting (BCA) framework underlying COMPASS and the suite, identifying shocks and their dynamic reactions in the model is key to interpretability. We are critical of evaluating models by impulse response functions (IRFs) showing reactions to shocks: these show the dynamic properties of a model, but not the data generating process *unless the model is complete and correct*. There is also an identification issue as changing the intercept or the error by the same amount in the equation being shocked cannot be distinguished: but unless super exogeneity holds, the effect will not be the same. Pagan (2016) shows the further problem that integrated, in other words, non-stationary, measurement errors are generated by current methods of calculating IRFs for changes in variables.

More fundamentally, our criticisms of the New Keynesian DSGE model in section III invalidate the BCA framework. The information revolution implies that the omission of shifts in credit constraints, balance sheets, and asset prices from DSGE models was a capital mistake as these are non-stationary variables and not ‘stationary wedges’. To put it another way, the assumption made in the BCA framework that the economy is never very far from its steady-state trend was simply wrong. These models were unable to capture the *possibility* of the kind of financial accelerator that operated in the US sub-prime crisis and the resulting financial crisis. They ignored the shock amplifying, propagating, and potentially destabilizing processes in the financial accelerator. Shifts in the credit market architecture are only one example of structural breaks and evolutions which make it hard to sustain the possibility of ‘rational’ or model consistent expectations.

Deep in the justification for why the economy is never far from its steady-state trend is the representative agent, rational (model consistent) expectations, permanent income view of household behaviour shared by COMPASS and BEQM. In that view, as discussed in section III, households discount temporary fluctuations in income to maintain spending in the face of such shocks, thus providing a stabilizing anchor to the economy, in turn justifying the rational expectation that shocks will prove temporary. To be fair, COMPASS assumes that only 70 per cent of households behave in this way and that 30 per cent are rule-of-thumb households who just spend current income. However, the fact that the mass of households follow the permanent income view still plays a major stabilizing role in the model.

While Deaton and Carroll emphasize idiosyncratic income uncertainty at the micro level, there is also often huge uncertainty at the macro level. Nothing makes this more obvious than the large forecasting errors made by the Bank, despite access to the best information in the economy and the most powerful information-processing capability in the UK. This macro uncertainty reinforces the Deaton–Carroll argument that, on average, households discount the future much more heavily (have shorter horizons) than the textbook permanent income hypothesis which underlies COMPASS. This undercuts the assumption that the economy is never far from its steady-state trend and the shock-decompositions of the BCA approach.

As David Miles says:

While I was on the MPC I always found the ‘shock decompositions’ derived from the Bank’s DSGE model rather hard to interpret—being both somewhat arbitrary and difficult to rationalize. The failure to model the credit side of the economy and its time-varying interaction with spending was a weakness, albeit one shared with many macroeconomic models. The Bank’s models also proved of limited value in understanding the biggest issue post financial crisis—namely whether the awful path of UK labour productivity reflects permanent changes or is significantly transitory.

(iv) UK-specific failures to capture the financial crisis

Section III explained the feedback loops operating in the US sub-prime crisis. In the UK, similar feedback loops operated but with important differences. In the absence of a pre-crisis construction boom, the halving of residential investment had a less severe effect on GDP. Importantly, the prevalence of floating rate mortgages meant that the cuts in interest rates fed through quickly to improve the cash flows of many mortgage borrowers and stabilized house prices, limiting losses on the lenders’ mortgage books. As Aron and Muellbauer (2016) show, rates of mortgage repossessions in the UK were of the order of 10 per cent of the US rates, and government policy also helped by improving income support for those in payment difficulties and encouraging forbearance. This meant that part of the feedback loop operating in the UK via the household sector was substantially less severe than in the US. The UK’s severe credit crunch owed more to losses on commercial property and developer loans, fines for financial misdemeanour, and losses on overseas activities, such as Royal Bank of Scotland’s disastrous takeover of ABN-Amro in 2007, which outweighed losses on the mortgage books of most domestic UK commercial banks. The Bank of England and the government eventually realized that the credit crunch was severely handicapping a UK recovery, despite quantitative easing (QE), and low rates, so the Funding for Lending Scheme and other measures were introduced.¹⁶

¹⁶ Judging from speeches by MPC members at the time, Adam Posen deserves particularly ‘credit’ for understanding the sensitivity of the economy to the credit crunch despite its omission from models, and proposing policies to ease it.

(v) Problems with the Bank's 'suite of models'

Unfortunately, the Bank has for some time had somewhat of a blind spot concerning the role of credit and the housing market in the UK economy, despite its historical interest in home equity withdrawal (HEW), e.g., Davey (2001) and Reinold (2011), and the maintenance of quarterly historical data on HEW back to 1970. HEW is about the housing collateral channel by which house prices, together with credit conditions, influence consumption. This channel is entirely absent from COMPASS and only plays a small role in the suite, to be discussed further below. The theory and UK evidence for such a channel was presented in Muellbauer (2008); see further discussion in section VI, and a comparison with the US and Japan provided in Aron *et al.* (2012).

Balance sheet effects on consumption via the stock market and housing in the UK are clearly important; see US and international evidence surveyed in Cooper and Dynan (2016). The Bank's suite does now include a three-equation estimated model for consumption, unsecured debt, and household broad money; see Cloyne *et al.* (2015). This imposes the restriction that broad money, stock market and pension wealth have the same marginal propensity to spend, and minus that on debt, though the short-term dynamics includes separate housing and financial wealth terms.¹⁷ Mortgage debt appears in another part of the system and is driven by the Fernandez-Corugedo and Muellbauer (2006) index of credit conditions, gross housing wealth, and the mortgage rate. In turn, changes in mortgage debt have a short-term effect on housing wealth, determined in the long run by income and the mortgage interest rate. There are therefore some transitory linkages from credit conditions and interest rates via housing wealth to consumption, and there is a direct effect from interest rates in the consumption equation but no time variation in the size of the effect.¹⁸

Balance sheets had previously been modelled in the Financial Stability Review; see Benito *et al.* (2001). Balance sheets are forecast by cumulating sectoral deficits/surpluses that are generated by the main macro forecast. Butt and Pugh (2014) provide commendable research on credit spreads for households and companies and these spreads apparently now play a role in modelling financial flows. A much more comprehensive non-DSGE model integrating the real side and financial flows and balance sheets with stock-flow consistency has recently been produced by Burgess *et al.* (2016). With more emphasis on the financial side and a short sample for 1998–2014, the real economy relationships such as consumption and investment functions take a fairly rudimentary form. This should nevertheless be a useful prelude to

¹⁷ Footnote 26 in Cloyne *et al.* (2015) explains that the housing wealth effect is excluded from the long-run solution for consumption (as in BEQM).

¹⁸ Cloyne *et al.* (2016) find evidence for a negative effect of changes in mortgage rates on spending using micro-data, first explained in a theory model by Jackman and Sutton (1982). With micro-data, it can be hard to establish the time-varying character of effects, but with a coherent model and aggregate time-series, Muellbauer (2008) and Aron *et al.* (2012) find that with higher debt relative to income, the cash-flow effect is larger, but partly offset by easier credit access. The implication, missing in Cloyne *et al.* (2015), is that in the post-2007 credit crunch, with high inherited debt levels, the cash-flow effect would have been particularly powerful.

greater future integration between the real and the financial sides of a policy model with richer behavioural relationships.

Recently, the Bank has included Geoff Meen's house price model for the Office for Budget Responsibility (OBR) in the suite, which does embody equilibrium correction and has a mortgage rationing effect for the post-2007 period.¹⁹ However, while the main economic forecasts feed into balance sheet forecasts, the latter apparently play relatively little role in the main economic forecasts, though there have been periods when extreme discrepancies between the two have led to shifts in the forecasts of the former.²⁰

There is a disconnect between COMPASS and the suite on one side and the financial stability responsibilities of the Bank on the other. The term 'financial stability' does not occur in Burgess *et al.* (2013), nor is there any reference to the important insights in Aikman *et al.* (2009) and the RAMSI model of Burrows *et al.* (2012). It seems as though the linkage is almost entirely one way, from the 'real economy' to finance. The contradictory lesson from the global financial crisis apparently remains to be learned.

While it is on the demand side and in the weak links between the real economy and finance where arguably the most severe problems arise (see the views of David Miles quoted above), there have also been issues in modelling pricing behaviour as the failure to forecast the inflationary implications of sterling's depreciation in the financial crisis makes clear. The prevalent strategy of linearizing a DSGE around a supposed steady state, entails that individual equation specifications are not rigorously investigated, thereby missing potentially important shifts and non-linearities, as with the role of a non-linear response of wages to price inflation leading to wage-price spirals in Castle and Hendry (2014).

VI. Towards better macroeconomic policy models

To improve policy models, central banks need research that merges theory-driven and data-driven approaches, rather than treating them as adversaries. In micro-econometrics, great strides have been made in evidence-based research. In macro, the methodological blinkers discussed above have hampered evidence-based progress in developing policy models. Blanchard (2018) writes:

For these models, capturing actual dynamics is clearly essential. So is having enough theoretical structure that the model can be used to trace the effects of shocks and policies. But the theoretical structure must by necessity be looser than for DSGE: aggregation and heterogeneity lead to much more complex aggregate

¹⁹ The quarterly speed of adjustment of under 6 per cent suggests either very slow adaption of house prices or some omitted variables.

²⁰ However, according to Burgess *et al.* (2013), p. 47: 'The profiles from the balance sheet model are used regularly as an input to MPC discussions, and judgements about financial conditions and the nominal environment are often fed back into the projections in COMPASS.'

dynamics than a tight theoretical model can hope to capture.²¹ Old-fashioned policy models started from theory as motivation, and then let the data speak, equation by equation. Some new-fashioned models start from a DSGE structure, and then let the data determine the richer dynamics.

It is possible that, in future, the generation of vast amounts of micro-data from administrative sources rather than surveys subject to selection bias and large measurement errors, may allow quantitative models for the whole economy to be constructed. Ideally, such macro-models would be based on statistically tested models of micro-behaviour, aggregated up from micro-data on millions of households and many thousands of firms. Testing should establish whether such models best assume full information optimizing behaviour at the micro-level or heuristic behaviour rules adopted in agent-based modelling approaches (Haldane and Turrell, 2018, this issue). In the absence of such data, there is an important place for policy-relevant models using aggregate data, general enough to be consistent with plausible micro-behaviour and with plausible assumptions about information and market structure. Such models should be able to encompass insights from multiple stylized models, and use aggregate time series data to learn about the relevance of these insights.

Given the fundamental role of structural breaks in forecast failure, more attention should be paid to disequilibrium states, especially when forecasting, as after a location shift, all equilibrium-correction models naturally converge back to their previous equilibrium means, leading to systematic forecast failure. However, there are methods for robustifying forecasts after location shifts are detected using intercept shifts that help avoid systematic forecast failure (see Clements and Hendry, 1996) and Castle *et al.*, 2015).

These methods also suggest a possible way forward for replacing rational or model-consistent expectations assumed in DSGE models by hypothesizing that agents use such robust methods, at least in the absence of good survey data on expectations. It is sometimes argued in favour of rational expectations that there is something wrong with an account of economic policy in which the authorities can consistently exploit an informational advantage over the public without the public ever catching on, and where what the model predicts is consistently at variance with what it assumes that the public expects. Robust forecasting methods avoid such objections as well as mitigating the failure of the rational expectations hypothesis in the face of structural breaks and radical uncertainty.

This remainder of this section begins by discussing the encompassing approach for evidence-based model selection. It continues by discussing the pros and cons of system estimation, proposing sub-system estimation as the most sensible option. Examples of sub-systems are discussed, with applications to the household sector. A UK illustration is given of how the specification of the consumption function can be improved. It is followed by highlighting some of the insights for economic policy derived from this kind of evidence-based research.

²¹ He is therefore critical of the FRB/US model's derivation of the lag structure from optimization by representative agents facing polynomial costs of adjustment.

(i) Encompassing

Encompassing aims to reconcile the plethora of empirical models that often exist to ‘explain’ some economic phenomenon. If there are several different empirical models, even if each appears viable when evaluated against its own information set, all but one must be either incomplete or incorrect, and all may be.

A natural way to reduce the set of ‘locally admissible’ models is to test whether any one model can account for the results found by all the others, thereby making them inferentially redundant. In a Monte Carlo setting, knowing the data generating process (DGP) would enable a researcher to derive the properties of all models thereof (called population encompassing), and hence account for their simulation outcomes. Encompassing asks whether any given empirical model can mimic this property to account for the results of other empirical models of the same dependent variables. Many of those models will be mutually non-nested, in that they use different sets of explanatory variables, and hence could capture different features of the DGP. Knowing the DGP would explain such outcomes, but it is manifestly more demanding to ask of another empirical model to statistically encompass rival contenders; see Bontemps and Mizon (2008) for an excellent overview.

Consider models M1 and M2 of a variable denoted y , with its (unknown) DGP denoted M0. To implement a progressive research strategy, encompassing must satisfy the three properties that define a partial ordering across models:

- (i) reflexive: a model explains itself;
- (ii) anti-symmetric: if M1 explains M2, then M2 cannot both be different and explain M1;
- (iii) transitive: if M1 explains M2 and M2 explains M3, then M1 explains M3.

Creating a ‘super model’ M4, say, that nests all possible contenders would achieve (i)–(iii), but need not provide a useful or interpretable explanation. Thus, the concept of parsimonious encompassing is used in practice, which requires a small model to account for the results of a larger model within which it is nested. This still requires (i)–(iii), but also that M1 (say) explains M4.

Encompassing tests are based on two implications of the encompassing principle. First, if one model encompasses a rival, then it ought to be able to explain the predictions of that rival model. Alternatively, as the encompassed model does not capture all the features of the DGP which are captured by the encompassing model, the first model ought to be able to predict some mis-specifications of the rival model. The latter is a powerful criterion when structural breaks occur, demanding of M1 that it can account for the shifts in M2.

In the context of model selection, encompassing tests of any simplified model against the initial model provide a key statistical control, revealing if reductions have lost relevant information. This enables tight significance levels to be used for individual selection

decisions to retain or omit variables, thereby avoiding ‘overfitting’ without losing valuable information; see Hendry and Doornik (2014).

(ii) System specification and sub-system estimation

Developing better models entails a better approach, using theory more loosely but broadly consistent with microfoundations of the new generation of heterogeneous agent models with idiosyncratic uninsurable household income uncertainty and liquidity constraints. (Dis-)equilibrium correction models with interpretable long-run conditional solutions are invaluable. Although system *specification* is needed, system *estimation* is not required. Sub-system estimation is better, as system estimation overstates the benefits of imposing restrictions assuming the model is correct versus the costs of using a mis-specified model. Instead, we stress the importance of testing, especially for structural change. Indeed, a ‘correct’ theory can be rejected by data evidence simply because of an overly restrictive empirical implementation; see Hendry and Mizon (2011).

System estimation might be adopted (a) because of efficiency gains from ‘full-information’ methods, or (b) to impose cross-equation restrictions on parameters believed to be the same, or (c) for deriving an estimate of a missing or ‘latent’ variable, see below. On the first, Hendry (1976) provides a complete analysis, and finds such gains are small and dependent on the correct specification of the system. On the second, efficiency gains can result when the parameters involved are indeed identical, and when empirical identification is difficult. However, these gains can largely be achieved in estimating a sub-system of closely related equations, for example for the household sector or for the non-financial corporate sector, rather than the system as a whole where the imposition of erroneous restrictions can result in inconsistent estimates.

What is important is system *specification*, and one of the major flaws of models like BEQM or COMPASS was to omit key sectors, possibly in a misguided attempt to keep the system tractable for joint estimation. Rather, there are major advantages to adopting the approach proposed in Hendry and Doornik (2014) and Hendry and Johansen (2015) for each equation, then checking the resulting system for coherence and completeness. In their method, all available theory information is retained unaffected by selecting over additional candidate variables, longer lags, and non-linear functions of all variables, as well as tackling outliers and location shifts. When the theory model is correct, identical distributions of parameter estimates are obtained as from direct fitting of that model to the data, but when the theory specification is incomplete, incorrect, or changes over time, selecting from the additional features can deliver an improved model; see <http://voxeu.org/article/data-mining-more-variables-observations> for an explanation and example.

An example of sub-systems estimation applied to the household sector is the LIVES structure—a ‘latent interactive variable equation system’; see Duca and Muellbauer (2013). Where data on hard-to-measure phenomena such as credit conditions are not available, latent variables can be used to link with observable data, e.g. on credit spreads, other measures of

loan conditions, and bank lending survey data. State-space methods or spline functions can be used to update the latent variable or variables, responding quickly to evolving data and incorporating new information, so adapting to structural changes. Coherent multi-equation models for major sectors of the economy have the advantage that sensible economic interpretations can be made at the sectoral level. The coherence of an integrated approach means that one is not relying just on one piece of data or one equation for inference. An example of such a model is the six-equation system for the German household sector in Geiger *et al.* (2016) and a similar model for France in Chauvin and Muellbauer (2017). As noted above, sub-system estimation is preferable to the attempt to estimate all the equations in a large system simultaneously. To learn from data, one needs reasonably general formulations allowing hypothesis tests for reductions to more parsimonious equations—far more feasible at the sub-system level.

The sub-system LIVES approach focuses on financial innovations/regulatory shifts as key sources of breaks for the household sector equations. The latent variables pick up everything, subject to smoothing, not explained by the variables in the system. The flexible use of dummies or the state-space alternative to represent the latent variables should allow the system to respond quickly to structural breaks as new data arrive. The latent variables resemble intercept corrections in individual equations but with cross-equation restrictions and economic interpretations. This is co-breaking, see Hendry and Massmann (2007), not just in the intercepts but including interactions with selected regressors.

The task of developing LIVES sub-systems is not trivial. Given the cross-equation restrictions in this set-up, automatic model selection methods to reduce general specifications to parsimonious ones, such as those available in the Doornik and Hendry (2013) *Autometrics* software, cannot (yet) be applied. The German and French six-equation models consist of equations for consumption, liquid assets, unsecured debt, mortgage debt, house prices, and permanent income. There are two main latent variables, credit conditions for unsecured and for mortgage debt. One of the biggest problems concerns the role of demographic variables. These are typically I(2) or I(1) variables²² so the risk of spurious regression is large and empirical identification relative to the latent variables is not trivial. Long samples, cross-country consistency, judgements on what parameter magnitudes are sensible given ranges of variation in the data, and qualitative conclusions from economic theory and country-specific institutional knowledge are all helpful in arriving at parsimonious models with strong economic interpretations.

For France, the latent variable measuring mortgage credit conditions tracks closely, since 1990, minus the ratio of non-performing bank loans to total bank loans to the private sector. This implies an important and necessarily non-linear linkage²³ between the household sector

²² I(2) indicates that twice-differencing is needed to convert such a variable to stationarity, while I(1) indicates that once-differencing is enough.

²³ In Duca *et al.* (2016), one of the main drivers of US house prices is a measure of the median loan-to-value ratio for first-time home buyers, a good proxy for credit conditions. The model for the latter includes a non-linear measure of lagged falls in house prices in the previous 2 years.

and the banking system. It also supports the economic interpretation given to this latent variable.

(iii) Improving the UK consumption function: economic and policy insights

In the UK, the down-payment constraint and access to home equity loans has varied over time, with a relaxation in the 1980s, tightening in the early 1990s, followed by a relaxation which ended when the credit crunch began in 2007–8. In turn, the funding for lending scheme, help-to-buy and the recovery in commercial bank balance sheets have led to a more recent relaxation. These shifts have altered the relationship between house prices and consumption. They have also altered the short-term impacts of changes in nominal mortgage rates on aggregate spending. Joint modelling of consumption and household balance sheets would help understand these phenomena, debt dynamics, and the implications of debt for consumption, and should be central to the information base for policy formation, within a larger model where the feedback loops can be more fully explored.

A graphical illustration below, based on a 2012 update²⁴ of the UK consumption function explained in Aron *et al.* (2012), gives an indication of how these phenomena played out for consumption. The long-run ‘credit-augmented’ solution for the log consumption to income ratio generalizes in several respects the textbook permanent income form, which depends on the log ratio of permanent to current income and on the net worth to income ratio. First, it splits net worth into three categories, net liquid assets defined as liquid assets minus debt, illiquid financial assets (stock market and pension wealth), and housing wealth. Second, shifting mortgage credit conditions, measured by a credit-conditions index, has an intercept effect capturing the shifting implications for saving behaviour of variations in the down-payment constraint, and interacts with housing wealth, capturing how access to home equity finance varies with access to credit. Third, the discount rate used to discount expected income growth is higher than the real interest rate to account for income uncertainty and liquidity constraints. Moreover, the coefficient on the log ratio of permanent to current income is freely estimated,²⁵ instead of imposing a coefficient close to one implied by the textbook permanent income model.

To incorporate shifts in credit constraints such as the down-payment constraint for a mortgage, the disaggregation of balance sheets, a role for house prices, income uncertainty, interest rates, and demography, the long-run version²⁶ of the credit-augmented generalized aggregate consumption function is:

²⁴ The update was prepared for a presentation by Muellbauer for a December 2012 Bank of England Monetary Policy Roundtable. Up to 2001, the credit conditions index comes mainly from Fernandez-Corugedo and Muellbauer (2006) with three smooth transition dummies capturing later shifts, estimated just from the consumption equation, rather than from a full LIVES system for consumption and household balance sheets.

²⁵ It also introduces an interaction effect with the credit conditions index, as easier credit access should allow households to be more forward looking.

²⁶ The dynamic version includes partial adjustment, and changes in the unemployment rate, income, and interest rates.

$$\ln(c_t/y_t) = \alpha_{0t} + \alpha_{1t}r_t + \alpha_{2t}\theta_t + \alpha_{3t}\ln(y_t^p/y_t) + \gamma_1NLA_{t-1}/y_t + \gamma_2IFA_{t-1}/y_t$$

$$+ \gamma_3\ln(hp_{t-1}/y_{t-1}) + \gamma_4HA_{t-1}/y_t + \gamma_5demog_t$$

Here c is consumption, y is income, r is a real interest rate, θ is an indicator of income uncertainty, y^p/y is the ratio of permanent to current income, NLA is liquid assets minus debt,²⁷ IFA is illiquid financial assets, hp is an index of house prices, HA is gross housing wealth, and $demog$ captures the effect of demography on consumption. Key coefficients can be time varying because of shifts in credit conditions.

The intercept α_{0t} increases with greater availability of non-housing loans and of mortgages, as the need to save for a down-payment is reduced. The coefficient measuring the sensitivity of down-payment requirements to house prices relative to income, γ_{3t} , should become less negative as access to mortgages rises. If access to home equity loans increases, the coefficient, γ_{4t} , measuring the marginal propensity to spend out of housing wealth, should increase. One might also anticipate that expectations of future income growth, captured in α_{3t} , would have a larger effect on consumption when credit constraints ease. It is also possible that α_{1t} , the sensitivity of consumption to the real interest rate, might be affected by credit conditions. This is the functional form underlying the following graphical decompositions.

Figure 2: Long-run decomposition of log consumption/income into effects of credit conditions and their interaction with housing wealth/income and into log permanent/current income

²⁷It is possible to disaggregate net worth into four main elements, with a separate coefficient on debt. However, relative to a common alternative restriction, the assumption that mortgage debt can just be netted off gross housing wealth, assuming that the coefficient on debt is minus that on liquid assets is better supported by the data.

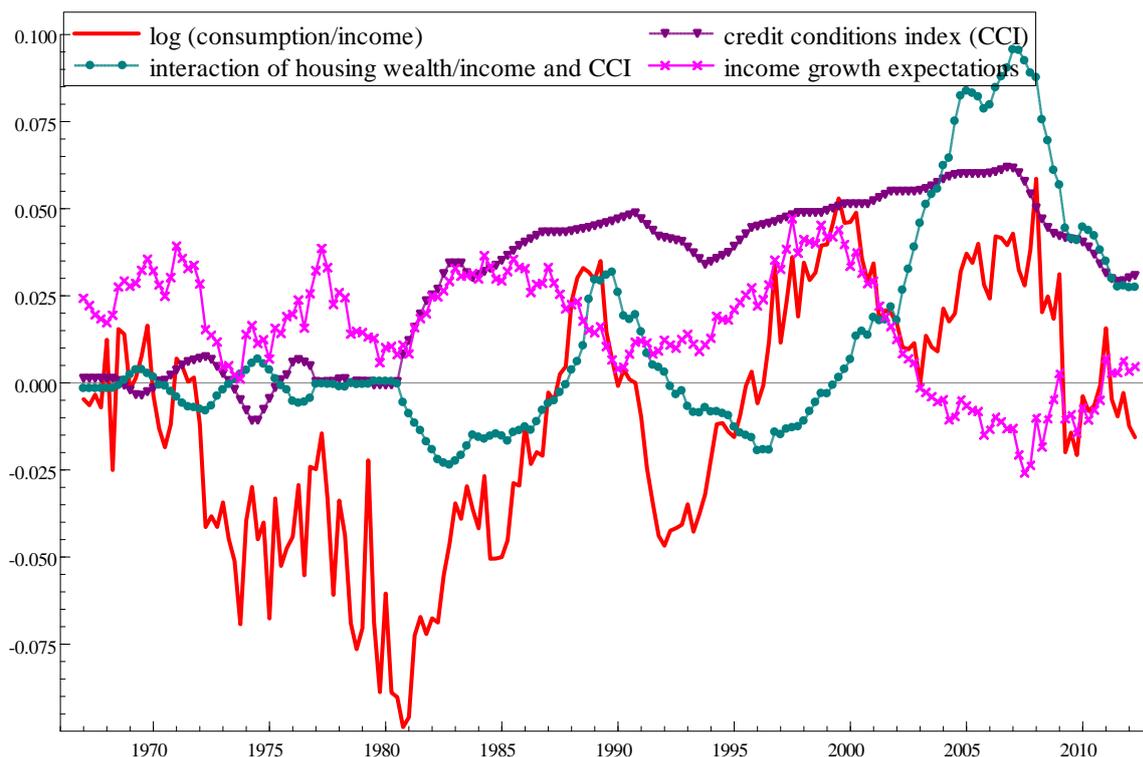
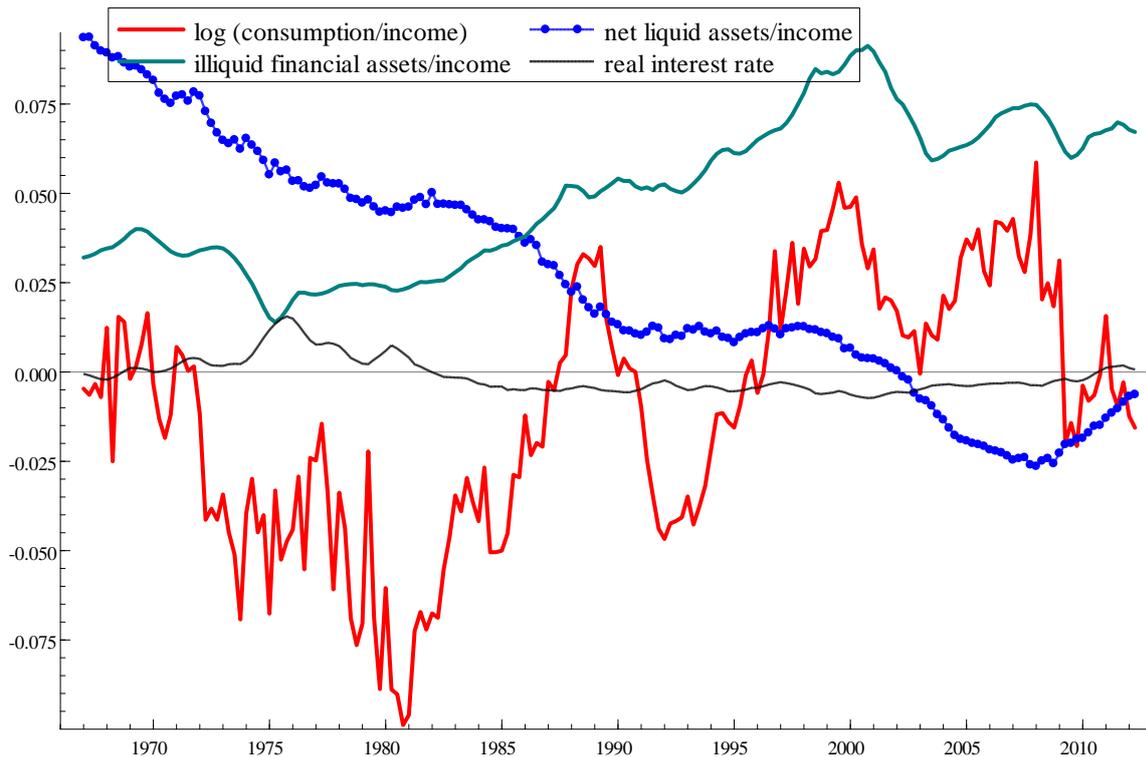


Figure 2 shows a rise from 1980 to 2007 in the log ratio of consumption to income, measured by non-property disposable household income. Some of this is attributable to the rise in the credit conditions index (CCI)—note the jump after 1979 induced by credit market liberalization under Mrs Thatcher. But the interaction effect between CCI and housing wealth/income has even sharper effects and captures much of the time variation. Note that in the 1970s, when credit was heavily rationed, this effect was essentially negligible. In the early 1990s, much of the fall in consumption relative to income is explained by the credit crunch and the decline in house prices, and the combination was even more pronounced after 2008 in the global financial crisis, when the credit crunch was even more severe. However, income growth expectations measured by estimates of log permanent to current income from a forecasting model, also explain some of the variation in log consumption/income. For example, in the first half of the 2000s, high levels of current income relative to permanent income (low log permanent/current income) offset some of the rise in consumption relative to income that would have been induced by the house price boom. And in the global financial crisis, when current income fell, permanent income does provide some stabilization for the consumption/income ratio. The remainder of the long-run effects are shown in Figure 3.

Figure 3: Long-run decomposition of log consumption/income into effects of net liquid assets/income, illiquid financial assets/income and the real interest rate



The rise in illiquid financial wealth relative to income makes a substantial contribution to the rise in consumption relative to income from 1980 to 2000, and explains some of the decline in the aftermath of the collapse of the dotcom stock market boom in the early 2000s. Crucially important, the other major story told by Figure 3 is the effect of the long-term build-up in debt implied by the decline in liquid assets minus debt, relative to income.²⁸ This is the pay-back for the credit liberalization and boom in asset prices which boosted consumption but built up debt burdens, and illustrates the vulnerability of the household sector to high debt levels when asset prices fall and access to new credit contracts. The estimated long-run effect of real interest rates is relatively small.

The two figures show only long-run effects. The estimated short-run dynamics also reveal two further important effects. One is a highly significant negative effect of the change in the unemployment rate on consumption, also a feature of the 1999 MTMM consumption function. The second is a negative effect of the change in the nominal borrowing rate, a mix of the mortgage rate and base rate, capturing the asymmetric short-run effect on borrowers compared with savers. This is weighted by the debt/income ratio and also includes an interaction with CCI: for given CCI, a higher debt/income ratio implies a more negative effect of higher interest rates on spending. But with easier access to credit, refinancing is easier when interest rates rise, so softening the impact. The implication is that in 2008–10,

²⁸ The estimated model imposes the restriction, empirically supported, that the coefficient on debt equals minus the coefficient on liquid assets. For a range of countries, we find aggregate housing ‘wealth’ effects of zero or negative before mortgage credit liberalization, the marginal propensity to consume for net liquid assets between 0.08 and 0.16, for illiquid financial assets between 0.015 and 0.03, and speeds of adjustment typically 0.4–0.7 for aggregate consumption, somewhat lower for non-durables.

when debt/income reached record levels, but access to new credit was very constrained, the impact of lower interest rates on aggregate UK consumption was particularly strong, evidence for monetary policy effectiveness. The 1999 MTMM consumption function also included a negative effect from the rise in the nominal borrowing rate on consumption, but with no account taken of the time variations in the effect.

Of course, such partial equilibrium decompositions tell only part of the story of policy transmission. For example, higher consumption feeds into higher output, higher asset prices, and lower unemployment, adding to the direct channel of transmission. Integration into a larger model of the economy is necessary to capture the full feedback loops, many of which are missing in COMPASS.

(iv) Wider economic and policy insights

There are close parallels between consumption behaviour in the US and the UK, though with mainly fixed-rate mortgages, the response of consumption and house prices to lower interest rates in the US is necessarily slower than in the UK.²⁹ The above generalization of the textbook permanent income model is an example of the encompassing approach as it encompasses the textbook model as a special case. It is also an example of the looser, and in our view more relevant, application of theory. In contrast to the FRB/US consumption function which incorporates no shifts in credit constraints and aggregates the household balance sheet into a single net worth concept, contradicted by micro evidence, see section III (iii), it no longer corresponds to a representative agent optimizing model. The claimed microfoundations of the FRB/US consumption function do not save it from parameter instability: the estimated speed of adjustment for data up to 2009 of 0.19 falls to 0.10 for recent data.³⁰ This is clear evidence against treating the FRB/US consumption function as a ‘structural’ equation in the classical sense of invariant to shifts in the economic environment.

Because of its omissions, the FRB/US model failed to give proper warning of risks faced by the US economy after 2007. At the Jackson Hole conference in 2007, Mishkin (2008) reported the results of FRB/US simulations of a 20 per cent decline in real house prices spread over 2007–8. The standard version of model simulated GDP lower than the baseline by 0.25 per cent in early 2009 and consumption lower by only 0.6 per cent in late 2009 and 2010. The simulations suggested a rapid recovery of residential investment given the lowering of the policy rate in response to the slowing economy. FRB/US failed to include a plausible model of house prices and so also missed the feedback from the credit crunch back on to house prices modelled in Duca *et al.* (2011, 2016). Consistent with this time series evidence, Favara and Imbs (2015) provide strong micro-evidence for the link between credit supply and house prices in the US.

²⁹ Duca and Muellbauer’s (2013) version of the US consumption function uses a latent variable to model the shifting marginal propensity to spend out of housing assets as access to home equity finance has varied, capturing the multiplicative effects of the credit crunch and falling house prices discussed after Figure 1 above.

³⁰ Even at 0.19, Mishkin (2007, pp. 391–2) expresses unease about the slow transmission of wealth effects on consumption.

Among the findings of the LIVES models for the household sectors of Germany and France discussed above are that major shifts took place in French credit conditions compared to Germany's, and these help explain the radically different patterns of house price developments in the two countries. Ignoring post-1980 shifts has catastrophic effects on the French consumption³¹ and house price equations and rather less serious ones for Germany. In contrast to the UK, higher house prices relative to income in Germany and France tend to reduce aggregate consumption, other things being equal. The interpretation is that conservative lending practices and high down-payment constraints force many younger households to save for a housing down-payment. Moreover, many renters may be more cautious anticipating that higher rents will follow higher house prices. However, as the French mortgage market has liberalized, so these negative effects of higher house prices relative to income have softened. These findings suggest that monetary transmission in Germany is very different from the US and UK, and somewhat different from France.

The three-equation LIVES model for the household sector in Canada of Muellbauer *et al.* (2015) reveals striking differences between Canada and its neighbour in linkages between house prices and aggregate consumption. The ATM-like role of home equity in the US is far smaller in Canada, highlighting the importance of institutional differences. In Canada, unlike in many US states, mortgage contracts are 'full recourse', meaning that defaulters' other assets and income flows can be legally drawn upon by mortgage lenders, and pursued for years. There is no tax relief for mortgage interest payments. The banking sector is concentrated and with compulsory mortgage insurance and federal oversight, lending standards are high, with almost no sub-prime lending.

In Japan, too, high down-payment constraints and conservative mortgage lending practices are factors implying a small negative effect from higher house prices on consumption, given income and other asset prices; see Aron *et al.* (2012). The radically different structure of Japanese household balance sheets compared to those of the US implies that one-size-fits-all ideas derived from the US about monetary policy transmission are wrong, as explained in Muellbauer and Murata (2011). No G7 economy has such a high ratio of liquid assets (mainly bank and post office saving deposits) to income, or such a low rate of stock-market participation (though Germany is not far behind). There is strong empirical evidence that lower real interest rates on deposits *reduce* aggregate consumption in Japan, given income and other asset prices. This explains why aggregate consumption in Japan has failed to respond to low and now negative policy interest rates or, indeed, to forward guidance on higher future inflation: many pre- and post-retirement households can hardly be enthused by the promise that the real value of their liquid assets will be further eroded in the future with the real income stream remaining negative for longer. Lower policy rates transmit to the real economy in other ways, such as the real exchange rate, the stock market, and higher investment including in residential housing, but overall monetary transmission in Japan is

³¹ For South Africa, the omission of credit conditions has similarly catastrophic effects for models of consumption and household debt, Aron and Muellbauer (2013b).

almost certainly weaker than in the US. This suggests that the ‘zero lower bound’ on nominal interest rates is not the issue for monetary policy in Japan, paralleling scepticism by Stiglitz (2018) on discussions of the zero lower bound in standard models.

These insights have applications to the Chinese economy. While mortgage markets have developed a great deal—household debt/GDP rose from 34.5 per cent in January 2014 to 44.4 per cent in January 2017—down-payment constraints are still far more stringent than in the US. The easy assumption that higher house prices in themselves will fuel higher aggregate consumption in China is certainly wrong. The easier credit flows that drive up house prices stimulate residential construction which creates higher employment and incomes. And though easier credit to households can temporarily stimulate consumption, given house prices, this is offset by the increased saving for down-payments by many younger households induced by higher house prices relative to incomes. When the credit expansion comes to an end, the hit to residential construction will affect employment and income, while still high house prices and high levels of debt constrain consumption. These are major problems for the hoped-for transition of the Chinese economy away from investment and export-led growth to consumption-led growth.

With countries going through major demographic transitions, empirical evidence on the implications is of great relevance. Empirical evidence from the LIVES models for Germany and France suggest that demographic effects on aggregate consumption *conditional on household portfolios disaggregated into the main asset and debt classes* are small. However, these portfolios, including the important debt components, themselves appear to be quite sensitive to demographic change, implying a slow but important feed-through of demography to household saving rates. Further research on these lines should illuminate the role of demography in the secular stagnation feared by some economists.

Empirical insights not available from micro-cross-sections or short panels include that, contrary to simple textbook models, there is a major role in Germany and France for nominal mortgage interest rates in driving house prices and the mortgage stock. The role of real rates as embodied in ‘user cost of housing’ increases with leverage. Another is that increased access to unsecured credit reduces demand for liquid assets—ignored by previous research on household demand for money. Finally, evidence for the buffer stock role of both liquid assets and unsecured debt comes from the negative reaction of the former and the positive reaction of the latter to a rise in the unemployment rate.

The careful distinction between the demand for and the supply of credit in LIVES models helps understand the paradox of the frequently found positive correlation between economic growth and credit growth and the negative one with debt levels. The evidence uncovered that, in aggregate, debt has far more negative effects on consumption than stock market or housing wealth have positive effects, has sobering implications for the extended use of monetary policy, including large-scale purchases of government bonds. In the short run, monetary policy, and, of course, provision of liquidity to banks and other institutions under liquidity stress, are important policy levers. But if extended periods of low interest rates and low

returns on safe assets drive up debt/income and prices of risky assets to levels beyond what would be sustainable under moderately higher interest rates, the boost to spending will be reversed later.

VII. Conclusions

Blanchard (2018) argues that DSGEs are unlikely to serve as useful macro-econometric policy models and recommends their co-existence with policy models where economic theory is applied more loosely and empirical evidence is used more intensively and more creatively. Wren-Lewis (2018) calls these ‘structural econometric models’ (SEMs), though see our discussion of the term ‘structural’ in section III. He regrets the abandonment, in preference to further testing and model development, by central banks such the Bank of England of their SEMs, exemplified by the Bank of England’s 1999 MTMM, discussed in section II. In section III we provided a general critique of the use at central banks of New Keynesian DSGE models as policy models.

We traced much of the problem back to the Lucas critique. Followers of Lucas effectively hijacked the term ‘structural’ from its earlier usage by Haavelmo (1944), the Cowles Commission, and Simon (1952, 1953) as referring to the invariance of a model or sub-model to shifts in the economic environment to now mean ‘microfounded’. The pincer movement of the Lucas critique and Sims’s (1980) advocacy of VARs supposedly free of the ‘incredible restrictions’ imposed in SEMs led to the intellectual fashion turning against SEMs. However, ‘microfounded’ was based on the illusion that macro data can be treated as the result of optimizing behaviour of representative agents. Moreover, the simplified textbook assumptions of linear budget constraints and close to complete markets contradict the fundamental insights of the asymmetric information revolution of the 1970s.

These imply buffer-stock behaviour of households faced with liquidity constraints and undiversifiable individual uncertainty, contradicting the textbook permanent income hypothesis and the adequacy of an aggregate Euler equation for consumption underlying RBC and New Keynesian DSGE models; see Muellbauer (2010) on the extensive empirical evidence against aggregate consumption Euler equations. The information revolution implies that the omission of shifts in credit constraints, balance sheets, and asset prices from DSGE models was a capital error. As a result, these models were unable to capture the *possibility* of the kind of financial accelerator that operated in the US sub-prime crisis and the resulting financial crisis. Shifts in the credit market architecture are only one example of structural breaks and evolutions, implying that the notion that the economy follows a stable long-run trend is highly questionable, despite heroic attempts by policy-makers to stabilize the path. Uncertainty then becomes radical. Structural breaks also make it hard to sustain the possibility of ‘rational’ or model consistent expectations. The law of iterated expectations necessary to derive DSGE models then breaks down. To illustrate, the uncertainty caused by Brexit emphasises the fundamental role of macroeconomic uncertainty, and should the UK exit the EU and the Single Market without a viable trade deal, many distributions will shift

radically, so the mathematics of DSGE models will again fail when good policy models are most needed.

The experience at the Bank of England illustrates these more general themes. Although the Bank of England's 1999 MTMM was mis-specified as explained in section II, BEQM discussed in section IV was even worse, and COMPASS whose structure is the subject of section V, worse still, though the idea was that by supplementing it with a suite of models, it could still be a useful policy guide.

The poor forecasting performance of both BEQM and COMPASS during the financial crisis was discussed. Unfortunately, key omissions in the suite of models contributed to the inability of the suite to properly correct for the omissions in COMPASS.

Such omissions, shared with the Bank's old MTMM, the current Treasury model, FRB/US and models at the Bank of Canada, and (to a degree) at the Netherlands central bank, include mis-specifications of the consumption function. The two most important are the omission of shifts in credit availability and the assumption that all assets and liabilities can be aggregated into a single net worth measure to drive consumption. The latter is patently absurd. It seems incredible that pension wealth is as spendable as cash-in-hand (though recent changes in UK pension legislation may have made pension pots more spendable for some than before). Moreover, inter-temporal theory, even in the absence of credit constraints, implies that housing, a consumption good as well as an asset, has a different relationship with consumption than cash or stock market wealth. When there are credit constraints in the form of down-payment constraints for a mortgage, saving for a down-payment becomes an important component of aggregate saving. And access to home equity loans, far from universal or constant over time, has implications for the relationship between house prices and aggregate consumption, as confirmed by micro evidence discussed in section III (iii).

Section VI discussed improvements in research methodology in the form of the encompassing approach to take into account the implications of different theories to better use evidence in the design of macro policy models. We recommended sub-system estimation in preference to complete systems estimation. The latent interactive variable equation system, LIVES, for a multi-equation approach to jointly model consumption together with main elements of balance sheet and house prices for the household sector is an example of such a sub-system. It enables the extraction of hard-to-observe directly shifting credit conditions as latent variables. An illustration was provided of a better UK consumption function, able, for example, to capture the interaction of the credit crunch and falling house prices in the financial crisis. Some important empirical insights from this kind of evidence-based research to policy debates were discussed. The relevance of this kind of research raises questions about whether the current training of graduate students in economics is adequate for those wishing to work in central banks.

To conclude, a generic issue in debates about policy models concerns their theoretical foundations. A major problem with the claim of 'theory consistency' is the question of 'which theory?' For example, textbook theory, which assumes efficient and close-to-

complete markets, well-informed relatively homogeneous agents, little uncertainty, no credit or liquidity constraints, and a stable economy, contrasts with theory that takes account of the asymmetric information revolution of the 1970s and early 1980s associated with Nobel prize winners Stiglitz, Akerlof, and Spence. Relevant theory must incorporate credit and liquidity constraints, incomplete markets with incomplete insurance and high levels of individual and macro uncertainty. In our view, approximate consistency with relevant theory, as illustrated above, is preferable to closer consistency with highly stylized ‘textbook’ theory.

References

- Aikman, D., Piergiorgio, A., Bruno, E., Gai, P., Kapadia, S., Martin, E., Mora, N., Sterne, G., and Willison, M. (2009), ‘Funding Liquidity Risk in a Quantitative Model of Systemic Stability’, Working Paper No. 372, Bank of England.
- Aiyagari, S. R. (1994), ‘Uninsured Idiosyncratic Risk and Aggregate Saving’, *Quarterly Journal of Economics*, **109**(3), 659–84.
- Aron, J., and Muellbauer, J. (2013a), ‘New Methods for Forecasting Inflation, Applied to the USA’, *Oxford Bulletin of Economics and Statistics*, **75**, 637–61.
- Aron, J., and Muellbauer, J. (2013b), ‘Wealth, Credit Conditions and Consumption: Evidence from South Africa’, *Review of Income and Wealth*, **59**(S1), 161–96.
- Aron, J. and Muellbauer, J. (2016), ‘Modelling and Forecasting Mortgage Delinquency and Foreclosure in the UK’, *Journal of Urban Economics*, **94**, 32–53, and Centre for Economic Policy Research
http://www.cepr.org/active/publications/discussion_papers/dp.php?dpno=11236
- Aron, J., Duca, J., Muellbauer, J., Murata, K., and Murphy, A. (2012), ‘Credit, housing collateral, and consumption: Evidence from Japan, the UK, and the US’, *Review of Income and Wealth*, **58**(3), 397–423.
- Ashley, S., and Patterson, L. (2016), ‘The Bank of England’s Independent Valuation Office’, The Bank of England’s Independent Evaluation Office.
- Bank of England (1999), *Economic Models at the Bank of England*, and September 2000 update.
- Bank of England (2004), ‘The New Bank of England Quarterly Model’, *Bank of England Quarterly Bulletin*, Summer, 188–93.
- Benito, A., Whitley, J., and Young, G. (2001), ‘Analysing Corporate and Household Sector Balance Sheets’, *Financial Stability Review*, 160–74.

- Bernanke, B., Gertler, M., and Gilchrist, S. (1999), 'The Financial Accelerator in a Quantitative Business Cycle Framework', ch. 21 (1341–93) in J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics: Volume 1*, Amsterdam, Elsevier.
- Bertola, G., and Caballero, R. J. (1990), 'Kinked Adjustment Costs and Aggregate Dynamics', *NBER Macroeconomics Annual*, **5**, 237–96.
- Blanchard, O. (2018), 'On the Future of Macroeconomic Models', *Oxford Review of Economic Policy*, **34**(1–2).
- Bontemps, C., and Mizon, G. E. (2008), 'Encompassing: Concepts and Implementation', *Oxford Bulletin of Economics and Statistics*, **70**, 721–75.
- Brayton, F., Levin, A., Lyon, R., and Williams, J. C. (1997), 'The Evolution of Macro Models at the Federal Reserve Board', *Carnegie-Rochester Conference Series on Public Policy*, **47**, 43–81.
- Browning, M., Gørtz, M. and Leth-Petersen, S. (2013), 'Housing Wealth and Consumption: A Micro Panel Study.' *Economic Journal*, 123: 401–428.
- Buiter, W. (2009), 'The Unfortunate Uselessness of Most 'State of the Art' Academic Monetary Economics', VOXEU, March, <http://voxeu.org/article/macroeconomics-crisis-irrelevance>
- Burgess, S., Burrows, O., Godin, A., Kinsella, S. and Millard, S. (2016), 'A Dynamic Model of Financial Balances for the United Kingdom', Bank of England Working Paper 614, Bank of England.
- Burgess, S., Fernandez-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K., and Waldron, M. (2013), 'The Bank of England's Forecasting Platform: COMPASS, MAPS, EASE and the Suite of Models', Bank of England Working Paper 471, Bank of England.
- Burrows, O., Learnmouth, D., and McKeown, J. (2012), 'RAMSI: A Top-down Stress-testing Model', Bank of England Financial Stability Paper 17.
- Burrows, V. (2017), 'The Impact of House Prices on Consumption in the UK: a New Perspective'. *Economica*. doi:10.1111/ecca.12237.
- Butt, N., and Pugh, A. (2014), 'Credit Spreads: Capturing Credit Conditions Facing Households and Firms', *Bank of England Quarterly Bulletin*, **54**(2), 137–48.
- Caballero, R. J. (2010), 'Macroeconomics after the Crisis: Time to Deal with the Pretence-of-Knowledge Syndrome', National Bureau of Economic Research Working Paper 16429.
- Carroll, C. D., (1992), 'The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence', *Brookings Papers on Economic Activity*, **23**, 61–156.

- Carroll, C. D. (1997), 'Buffer Stock Saving and the Life Cycle/Permanent Income Hypothesis', *Quarterly Journal of Economics*, **112**(1), 1–56.
- Carroll, C. D. (2000), 'Requiem for the Representative Consumer? Aggregate Implications of Microeconomic Consumption Behavior', *American Economic Review Papers and Proceedings*, **90**(2), 110–15.
- Carroll, C. D. (2001), 'A Theory of the Consumption Function, With and Without Liquidity Constraints', *Journal of Economic Perspectives*, **15**(3), 23–46.
- Castle, J. L., and Hendry, D. F. (2014), 'Semi-automatic Non-linear Model Selection', in N. Haldrup, M. Meitz, and P. Saikkonen (eds), *Essays in Nonlinear Time Series Econometrics*, Oxford, Oxford University Press, 163–97.
- Castle, J. L., Clements, M. P., and Hendry, D. F. (2015), 'Robust Approaches to Forecasting', *International Journal of Forecasting*, **31**, 99–112.
- Castle, J.L., Hendry, D.F. and Martinez, A. B. (2017) 'Evaluating Forecasts, Narratives and Policy using a Test of Invariance', forthcoming *Econometrics*.
- Castle, J. L., Doornik, J. A., Hendry, D. F., and Nymoen, R. (2014), 'Mis-specification Testing: Non-invariance of Expectations Models of Inflation', *Econometric Reviews*, **33**, 553–74.
- Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2007), 'Business Cycle Accounting', *Econometrica*, **75**(3), 781–836.
- Chauvin, V., and Muellbauer, J. (2017), 'Consumption, Household Portfolios and the Housing Market in France', ms.
- Christiano, L. R., Motto, R., and Rostagno, M. (2003), 'The Great Depression and the Friedman–Schwartz Hypothesis', *Journal of Money, Credit and Banking*, **35**(6), 1119–97.
- Christiano, L. R., Trabandt, M., and Walentin, K. (2009), 'Introducing Financial Frictions and Unemployment into a Small Open Economy Model', Sveriges Riksbank Working Paper 214.
- Clarida, R., Galí, J., and Gertler, M. (1999), 'The Science of Monetary Policy: A New Keynesian Perspective', *Journal of Economic Literature*, **37**, 1661–707.
- Clements, M. P., and Hendry, D. F. (1995), 'Forecasting in Cointegrated Systems', *Journal of Applied Econometrics*, **10**(2), 127–46.
- Clements, M. P., and Hendry, D. F. (1996), 'Intercept Corrections and Structural Change', *Journal of Applied Econometrics*, **11**(5), 475–94.
- Clements, M. P., and Hendry, D. F. (1999), *Forecasting Non-stationary Economic Time Series*, Cambridge, MA, MIT Press.

- Cloyne, J., Clodomi, F., and Surico, P. (2016), 'Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism', Staff Working Paper No. 589, Bank of England.
- Cloyne, J., Thomas, R., Tuckett, A., and Wills, S. (2015), 'A Sectoral Framework for Analyzing Money, Credit and Unconventional Monetary Policy', Bank of England Working Paper 556, Bank of England.
- Coletti, D., Hunt, B., Rose, D., and Tetlow, R. J. (1996), 'The Bank of Canada's New Quarterly Projection Model, Part 3. The Dynamic Model: QPM', Technical Report 75, Bank of Canada, Ottawa.
- Cooper, D., and Dynan, K. (2016), 'Wealth Effects and Macroeconomic Dynamics', *Journal of Economic Surveys*, **30**, 34–55, doi:10.1111/joes.12090, published online in 2014.
- Davey, M. (2001), 'Mortgage Equity Withdrawal and Consumption', *Bank of England Quarterly Bulletin*, Spring, 100–3.
- Deaton, A. S. (1991), 'Saving and Liquidity Constraints', *Econometrica*, **59**, 1221–48.
- Deaton, A. S., and Muellbauer, J. N. J. (1980), *Economics and Consumer Behaviour*, Cambridge, Cambridge University Press
- Doornik, J. A., and Hendry, D. F. (2013), *Empirical Econometric Modelling using PcGive: Volume I*, 7th edn, London, Timberlake Consultants Press.
- Duca, J. V., and Muellbauer, J. (2013) 'Tobin LIVES: Integrating Evolving Credit Market Architecture into Flow of Funds Based Macro-Models', ECB Working Paper No. 1581, also in B. Winkler, A. van Riet, and P. Bull (eds), *A Flow-of-Funds Perspective on the Financial Crisis Volume II—Macroeconomic Imbalances and Risks to Financial Stability*, Palgrave Macmillan 2014.
- Duca, J. V., Muellbauer, J., and Murphy, A. (2010), 'Housing Markets and the Financial Crisis of 2007–2009: Lessons for the Future', *Journal of Financial Stability*, **6**, 203–17.
- Duca, J., Muellbauer, J., and Murphy, A. (2011). 'House Prices and Credit Constraints: Making Sense of the US Experience.' *The Economic Journal* 121 (552): 533-551.
- Duca, J. V., Muellbauer, J., and Murphy, A. (2016), 'How Mortgage Finance Reform Could Affect Housing', *American Economic Review*, **106**(5), 620–4.
- Ericsson, N. R., and Irons, J. S. (1995), 'The Lucas Critique in Practice: Theory Without Measurement', in K. D. Hoover (ed.), *Macroeconometrics: Developments, Tensions, and Prospects*, Boston, MA, Kluwer Academic Publishers, 263–312.
- Favara, G., and Imbs, J. (2015), 'Credit Supply and the Price of Housing', *American Economic Review*, **105**(3), 958–92.

- Favero, C., and Hendry, D. F. (1992), 'Testing the Lucas Critique: A Review', *Econometric Reviews*, **11**, 265–306.
- Fawcett, N., Koerber, L., Masolo, R., and Waldron, M. (2015a), 'Evaluating UK Point and Density Forecasts from an Estimated DSGE Model: The Role of Off-model Information over the Financial Crisis', Bank of England Working Paper 538.
- Fawcett, N., Masolo, R., Koerber, L., and Waldron, M. (2015b), 'How did the Bank's Forecasts Perform Before, During and after the Crisis?', available at <https://bankunderground.co.uk/2015/11/20/how-did-the-banks-forecasts-perform-before-during-and-after-the-crisis/>
- Fernandez-Corugedo, E., and Muellbauer, J. (2006), 'Consumer Credit Conditions in the United Kingdom', Bank of England Working Paper 314.
- Fisher, I. (1933), 'The Debt-Deflation Theory of Great Depressions', *Econometrica*, **1**, 337–57.
- Frisch, R. (1938), 'Statistical versus Theoretical Relations in Economic Macrodynamics', mimeograph dated July 1938, League of Nations Memorandum.
- Geiger, F., Muellbauer, J., and Rupperecht, M. (2016), 'The Housing Market, Household Portfolios and the German Consumer', Working Paper Series No. 1904, European Central Bank.
- Haavelmo, T. (1944), 'The Probability Approach in Econometrics', *Econometrica*, **12**, 1–115.
- Haldane, A., and Turrell, A. (2018), 'An Interdisciplinary Model for Macroeconomics', *Oxford Review of Economic Policy*, **34**(1–2).
- Harrison, R., Nikolov, K., Quinn, M., Ramsay, G., Scott, A., and Thomas, R. (2005), 'The Bank of England Quarterly Model', London, Bank of England.
- Hedlund, A., Karahan, F., Mitman, K., and Ozkan, S. (2016), 'Monetary Policy, Heterogeneity, and the Housing Channel', Working Paper, University of Missouri.
- Hendry, D. F. (1976), 'The Structure of Simultaneous Equations Estimators', *Journal of Econometrics*, **4**, 51–88.
- Hendry, D. F., and Doornik, J. A. (1997), 'The Implications for Econometric Modelling of Forecast Failure', *Scottish Journal of Political Economy*, **44**, 437–61.
- Hendry, D. F., and Doornik, J.A. (2014). *Empirical Model Discovery and Theory Evaluation*, MIT Press.
- Hendry, D. F., and Johansen, S. (2015), 'Model Discovery and Trygve Haavelmo's Legacy', *Econometric Theory*, **31**, 93–114.
- Hendry, D. F., and Massmann, M. (2007), 'Co-breaking: Recent Advances and a Synopsis of

- the Literature’, *Journal of Business and Economic Statistics*, **25**, 33–51.
- Hendry, D. F., and Mizon, G. E. (2011), ‘Econometric Modelling of Time Series with Outlying Observations’, *Journal of Time Series Econometrics*, **3**(1).
- Hendry, D. F., and Mizon, G. E. (2014), ‘Unpredictability in Economic Analysis, Econometric Modelling and Forecasting’, *Journal of Econometrics*, **182**, 186–95; also see <http://www.voxeu.org/article/why-standard-macro-models-fail-crises>
- Hoover, K. D. (1994), ‘Econometrics as Observation: The Lucas Critique and the Nature of Econometric Inference’, *Journal of Economic Methodology*, **1**, 65–80.
- Hoover, K. D. (2001), ‘Does Macroeconomics Need Microfoundations?’, in K. D. Hoover *The Methodology of Empirical Macroeconomics*, Cambridge, Cambridge University Press, 57–89, reprinted in D. M. Hausman (ed), *The Philosophy of Economics: An Anthology*, 3rd edn, Cambridge, Cambridge University Press, 2007.
- Houthakker, H. S. (1956), ‘The Pareto Distribution and the Cobb–Douglas Production Function in Activity Analysis’, *Review of Economic Studies*, **23**, 27–31.
- Iacoviello, M. (2005), ‘House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle’, *American Economic Review*, **95**, 739–64.
- Iacoviello, M., and Neri, S. (2010), ‘Housing Market Spillovers: Evidence from an Estimated DSGE Model’, *American Economic Journal: Macroeconomics*, **2**, 125–64.
- IEO (2015), ‘Evaluating Forecast Performance’, Bank of England, Independent Evaluation Office, November, 1–91.
- Jackman, R., and Sutton, J. (1982), ‘Imperfect Capital Markets and the Monetarist Black Box: Liquidity Constraints, Inflation and the Asymmetric Effects of Interest Rate Policy’, *Economic Journal*, **92**, 108–28.
- Jordà, O., Schularick, M., and Taylor, A. M. (2016), ‘The Great Mortgaging: Housing Finance, Crises and Business Cycles’, *Economic Policy*, **31**(85), 107–52.
- Kahneman, D. (2003), ‘Maps of Bounded Rationality: Psychology for Behavioral Economics’, *American Economic Review*, **93**, 1449–75.
- Kaplan, G., and Violante, G. L. (2014), ‘A Model of the Consumption Response to Fiscal Stimulus Payments’, *Econometrica*, **82**, 1199–239.
- Kaplan, G., Moll, B., and Violante, G. L. (2016), ‘Monetary Policy According to HANK’, Working Paper 1602, Council on Economic Policies.
- Kaplan, G., Violante, G. L., and Weidner, J. (2014), ‘The Wealthy Hand-to-Mouth’, *Brookings Papers on Economic Activity*.

- Keynes, J. M. (1939), 'Professor Tinbergen's Method', *The Economic Journal*, **44**, 558–68.
- Kirman, A. (1992), 'Whom or What Does the Representative Individual Represent?', *Journal of Economic Perspectives*, **6**, 117–36.
- Leeper, E. M., Traum, N., and Walker, T. B. (2017), 'Clearing Up the Fiscal Multiplier Morass', *American Economic Review*, **107**(8), 2409–54.
- Lindé, J. (2018), 'DSGE Models: Still Useful in Policy Analysis?', *Oxford Review of Economic Policy*, **34**(1–2).
- Lucas, R. E. (1976), 'Econometric Policy Evaluation: A Critique', in K. Brunner and A. Meltzer (eds), *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conferences on Public Policy, Amsterdam, North-Holland, 19–46.
- Lusardi, A. (2016), 'Financial Literacy and Ignorance', available at <http://annalusardi.blogspot.co.uk/2016/>
- Mian, A., and Sufi, A. (2014), *House of Debt*, Chicago, IL, University of Chicago Press.
- Mian, A., Rao, K., and Sufi, A. (2013). 'Household balance sheets, consumption, and the economic slump'. *The Quarterly Journal of Economics*, 128(4):1687–1726.
- Mian, A., and Sufi, A. (2016). 'Who Bears the Cost of Recessions? The Role of House Prices and Household Debt', *NBER Working Paper* No. 22256.
- Mishkin, F. (2008). 'Housing and the Monetary Transmission Mechanism', in *Housing, Housing Finance and Monetary Policy*, Federal Reserve Bank of Kansas City, 359–413.
- Muellbauer, J. (1976), 'Community Preferences and the Representative Consumer', *Econometrica*, **44**, 979–99.
- Muellbauer, J. (2008), 'Housing, Credit and Consumer Expenditure', Proceedings, Economic Policy Symposium, Jackson Hole 2007, Federal Reserve Bank of Kansas City, 267–334.
- Muellbauer, J. (2010), 'Household Decisions, Credit Markets and the Macroeconomy: Implications for the Design of Central Bank Models', Bank for International Settlements Discussion Paper 306, March.
- Muellbauer, J., and Murata, K. (2011), 'Mistaken Monetary Policy Lessons from Japan?', available at <http://voxeu.org/article/mistaken-monetary-policy-lessons-japan>.
- Muellbauer, J., St-Amant, P., and Williams, D. (2015), 'Credit Conditions and Consumption, House Prices and Debt: What Makes Canada Different?', Bank of Canada Staff Working Paper 2015-40.

- Pagan, A. R. (2003), 'Report on Modelling and Forecasting at the Bank of England', *Bank of England Quarterly Bulletin*, Spring, available at <http://www.bankofengland.co.uk/archive/Documents/historicpubs/qb/2003/qb030106.pdf>
- Pagan, A.R. (2016), 'An Unintended Consequence of Using "Errors in Variables Shocks" in DSGE Models?', unpublished paper, ANU.
- Reinold, K. (2011), 'Housing Equity Withdrawal since the Financial Crisis', *Bank of England Quarterly Bulletin*, Q2.
- Reis, R. (2018), 'Is Something Really Wrong with Macroeconomics?', *Oxford Review of Economic Policy*, **34**(1–2).
- Robbins, L. (1932), *An Essay on the Nature and Significance of Economic Science*, London, Macmillan.
- Romer, P. (2016), 'The Trouble with Macroeconomics', unpublished paper, Stern School of Business, New York University.
- Simon, H. A. (1952), 'On the Definition of the Causal Relation', reprinted as ch. 3 in H. A. Simon, *Models of Man*, 1957, New York, JohnWiley.
- Simon, H. A. (1953), 'Causal Ordering and Identifiability', ch. 3 in W. C. Hood and T. C. Koopmans (eds), *Studies in Econometric Method*, Cowles Commission Monograph 14, New York, John Wiley.
- Simon, H. A. (1972), 'Theories of Bounded Rationality', in C. B. McGuire and R. Radner (eds), *Decision and Organization*, Amsterdam, North-Holland, 161–76.
- Sims, C. A. (1980), 'Macroeconomics and Reality', *Econometrica*, **48**, 1–48.
- Smets, F., and Wouters, R. (2007), 'Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach', *American Economic Review*, **97**, 586–606.
- Stiglitz, J. E. (2018), 'Where Modern Macroeconomics Went Wrong', *Oxford Review of Economic Policy*, **34**(1–2).
- Taylor, M. P., Peel, D. A., and Sarno, L. (2001), 'Nonlinear Mean-reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles', *International Economic Review*, **42**(4), 1015–42.
- Turner, A. (2015), *Between Debt and the Devil: Money, Credit, and Fixing Global Finance*, Princeton, NJ, Princeton University Press.
- Tversky, A., and Kahneman, D. (1974), 'Judgment under Uncertainty: Heuristics and Biases', *Science. American Association for the Advancement of Science*, **185**(4157), 1124–31.
- Windsor, C., Jääskelä, J. P. and Finlay, R. (2015), 'Housing Wealth Effects: Evidence from an Australian Panel', *Economica*, **82**: 552–577.

Wren-Lewis, S. (2018), 'Ending the Microfoundations Hegemony', *Oxford Review of Economic Policy*, **34**(1–2).