Model selection in macroeconomics: DSGE and ad hocness

Jaakko Kuorikoski & Aki Lehtinen

Routledge
Taylor & Francis Group

Check for updates

Model selection in macroeconomics: DSGE and ad hocness

Jaakko Kuorikoski[a] and Aki Lehtinen[b]

[a]New Social Research, University of Tampere, Tampere, Finland; [b]Practical Philosophy, University of Helsinki, Helsinki, Finland

ABSTRACT
We investigate the applicability of Rodrik's accounts of model selection and horizontal progress to macroeconomic DSGE modelling in both academic and policy-oriented modelling contexts. We argue that the key step of identifying critical assumptions is complicated by the interconnectedness of the common structural core of DSGE models and by the ad hoc modifications introduced to model various rigidities and other market imperfections. We then outline alternative ways in which macroeconomic modelling could become more horizontally progressive.

Introduction

Dani Rodrik (2015) has written an exceptional book. We believe it will help economists and non-economists to communicate more effectively with each other, and economists to appreciate better what they are doing. His notion of *critical assumptions* facilitates the articulation of a point that economists have probably always been aware of: they do care about the realisticness of at least some assumptions, even though Friedman seemed to say otherwise.

Our purpose in this paper is to investigate the applicability of Rodrik's account of model selection to the dominant modelling platform in macroeconomics, namely, dynamic stochastic general equilibrium (DSGE) models. Given that macroeconomics is one of the major branches of economics, especially in terms of application, it should be covered in any general account of the use of economic knowledge. Yet, although Rodrik outlines the recent twists and turns in macroeconomics at the level of general theory and outlook, he does not really examine the practice of applied macroeconomic modelling, merely suggesting that his account of model selection (and possibly of diagnostics) ought also to be applicable to macroeconomics.

According to Rodrik, economic knowledge accumulates 'horizontally' in producing ever more partial, context-sensitive and even mutually inconsistent models of new aspects of economic phenomena. This is to be contrasted with the lay idea of the 'vertical' growth of scientific knowledge, meaning that a science progresses through the revision and refining of a single core theory, from which specific applications are straightforwardly derived as special cases. The vertical conception also holds that a more accurate and truthlike theory automatically leads to improved empirical applicability. In contrast, the implication behind a horizontally growing toolbox of partly disjoint models is that the crucial skill of applying economic knowledge to explain or influence the real world lies in the judicious selection of the right model(s) for the particular epistemic task at hand. This is as much an art as it is 'science'. In Rodrik's view, there is nothing wrong with such a 'science' as long as the models wear their conditions for applicability on their sleeves.

We examine how the key step of identifying critical assumptions is complicated by the interconnectedness of the common structural core of DSGE models, and the myriad of more or less ad hoc

---

modifications introduced to model various rigidities and other market imperfections. We also examine the implications of these difficulties for the applicability of Rodrik's scheme to model selection in macroeconomics. There is definitely an art to macroeconomics, but it is the art of ad hoc modification rather than of outright model selection. We conclude our paper with a brief exploration of the alternative ways of adopting Rodrik's philosophy of economics into macroeconomics.

What is model selection?

Let us start with a description of Rodrik's ideas about selecting models. He tells the following story about how economics ought to be, and partly is, conducted. Economists consider a menu of models such that the question to be studied determines which assumptions are to be considered 'critical'. Critical assumptions are those that ought to approximate reality at least reasonably well (p. 29). The economist's task is to select the right model, the one that best fits the setting (p. 11). Theoretical progress occurs via developing new models that consider aspects of social phenomena not covered in earlier models rather than replacing older models with new, more 'realistic' ones (p. 64). The judicious selection of models hinges on two ideas. First, the critical assumptions must fit the setting, in other words, they must be realistic enough in the context of the study; second, the direct and incidental implications of the model must be consistent with the observed outcomes (p. 94). Empirical evidence thus has a role in model selection, but Rodrik is aware that some models are very difficult to test, and that many empirical tests are brittle (p. 65). Consequently, the empirical results are often weakened or overturned by subsequent empirical analyses, hence 'the profession's progression of favoured models tends to follow fad and fashion, or changing tastes about what is an appropriate modelling strategy, instead of evidence per se' (p. 65). Nevertheless, the horizontal expansion of models is frequently based on empirical observations that contradict existing models (p. 70). Finally, what makes economists engage in model selection is either a change in circumstances or a change in the setting on which they focus (p. 6).

In principle, if model selection is interpreted as simply selecting individual models from a given menu, one might think that it would be more applicable to central bankers or advisors to governments than to academic researchers. However, given that even many macroeconomists outside academia modify and combine existing models to create new ones (cf. Boumans, 1999), we adopt a flexible interpretation of model 'selection' including the whole business of selection, modification and application. If Rodrik's model-selection story is to be understood as a methodological guideline, one might ask what exactly is it that it rules out. One of the clearest morals of the story is that economists should be careful with benchmark models that presume well-functioning markets: 'common blueprints are out; model selection is in' (p. 166). One size does not fit many, and models often need to be modified with case-specific market imperfections (p. 61). He also notes that model selection is not a onetime affair, but a continuous process of modification and accommodation to changing epistemic needs.

How exactly, then, are models to be selected? How should one weigh the different models when making judgments about, say, appropriate monetary policy? It is easy to understand why Rodrik does not really have much to say about this, other than noting that economics may also advance through better methods of model selection, by 'improving the match between model and real-world setting' (p. 183). When it comes to macroeconomics, it is easy to say with hindsight that 'there was too much Fama, too little Shiller' (p. 159). Willem Buiter's (2009) lament that during the recent turmoil central banks could not find any useful support from highly trained young economists is well known. There was very little demand for dynamic programmers in central banks when financial instability became the main concern. All of a sudden central banks were implementing policies that could only be justified with recourse to economic theories developed decades ago. Such events indicate that there has been a deeper problem with the kind of macroeconomics that economic scientists have developed, and its use by the central banks: if economists are not actively developing models that central banks can use, choosing from *such* a menu is not a good idea (Solow & Touffut, 2012).

Nevertheless, Bernanke[1] argued that the financial crisis was more of a failure of economic engineering and management, than of economic science.

Let us now consider the development of macroeconomics since the emergence of real business cycle (RBC) economics circa 1982 (Kydland & Prescott, 1982). It was clear from the outset that the RBC models would not be able to track empirical data particularly well. Consequently, they were amended by increasing the number of shocks and were later transformed into NK-DSGE models by introducing various frictions. New Keynesian models such as these have since been changed into a form in which they can be estimated by first calibrating a few parameters. The latest incarnations take into account the lessons learned from the financial crisis in explicitly modelling the financial sector. Nevertheless, there is a fundamental sense in which the core model has remained the same: what some call the 'Ramsey model' of intertemporal optimization has been a key ingredient of all these models. The core model has thus been extensively supplemented with various components so as to obtain a better fit with the data.

Although both theoretical and applied macroeconomics certainly involve model selection, it does not mean that the development described above can be characterized as horizontal expansion. This is understandable given that macro modelling seems less suitable for progressing via horizontal expansion than other fields of economics simply because macro models, at least those used for forecasting, simply cannot ignore some parts of the economy to allow one to concentrate one's modelling efforts on something else. This is especially evident in large-scale, country- or region-specific models used by central banks, the aim of which is simultaneously to include as many sectors, market imperfections and other amendments as possible while retaining the essential core elements. This is vertical expansion in that new modifications and 'improvements' are added to the generally accepted but in itself simplified and idealized core model to improve its realisticness and empirical accuracy.

Differing from Rodrik's horizontal expansion, model development in macroeconomics appears, at least at the outset, to proceed along the lines proposed by Ylikoski and Aydinonat (2014), who emphasize the notion of a *model family* (see also Balzer & Dreier, 1999). According to Aydinonat and Ylikoski, a model family is formed around a set of critical assumptions (the central mechanism), which remains the same within the family, and testing the robustness of the main results is essential to the empirical interpretability of the model family as a whole (see also Kuorikoski, Lehtinen, & Marchionni, 2010). Although clearly constituting a family of models in some sense, DSGE modelling also differs from Aydinonat & Ylikoski's account in crucial respects. DSGE models are also built to capture new macroeconomic mechanisms, and this brings in an aspect of horizontal development. In addition, robustness analysis is problematic within the DSGE family, as shown below. However, insofar as the new model is built on top of a large number of familiar theory-based elements, it brings vertical rather than horizontal progress.[2] A common blueprint, a model template into which more context-specific additions are plugged into, is definitely in, and advocates of the blueprint credit it with turning macroeconomics back into a normal science. Whether what has transpired has been a desirable development from the perspective of Rodrik's account of model selection depends, in part, on the stance one takes on the necessity of microfoundations and intertemporal maximization.

Selecting the 'right' DSGE model

According to Rodrik, the function of an economic model is to capture the central causal mechanism(s) at work in the particular empirical context of application. DSGE models are indeed tailored to fit particular contexts: big models used for forecasting and for running alternative policy scenarios are built to be country- or region-specific, whereas theoretical models are built to highlight only some macroeconomic relations of interest. Such tailoring is achieved by introducing modifications to the core 'microfounded' growth model, and these modifications may or may not have an underlying justification in terms of individual maximization.

Let us consider next the questions that arise about how model selection is supposed to work when considering the modifications required for tailoring the core DSGE model. Models should be selected so that their critical assumptions fit the situation at hand. According to Rodrik, critical assumptions would produce a substantively different result if they were altered to be more realistic (p. 94). The notion of a critical assumption is thus multi-functional. As a concept, it purports simultaneously to define what is crucial (or what is the key causal mechanism), what the modellers' intended target is in some context, and what changes in assumptions change the results that can be derived. The two main problems in applying the concept to macroeconomics are that it is difficult to tell what would count as a critical assumption, and that the different functions of the concept come apart.

First, many of the most central assumptions, such as intertemporal optimization, never change in DSGE models: even if the modifications concern the behavioural assumptions, the core optimization model is never abandoned. In other words, altering this assumption to make it more realistic is only possible if the whole DSGE framework is abandoned, and even then it can only be changed along with so many others that it becomes difficult to pin the changes down to this specific assumption.

Second and more generally, identifying what assumptions are truly critical is much more difficult in the case of DSGE models than in simpler applied microeconomic models. The structural core is interconnected with the 'ad hoc' modifications and with various other assumptions that are practically necessary for tractability reasons. Furthermore, all DSGE models share some assumptions simply because it has been mathematically too difficult to modify them, or because the modifications require resorting to even more problematic ad hoc fixes elsewhere in the model.

Consider, for example, two consequences of the ubiquitous Dixit-Stiglitz aggregator: the assumption that the consumer's preferences are homothetic and the assumption that firms engage in monopolistic competition. It is eminently reasonable to assume that the unemployed have a radically different consumption basket from the well-to-do, and that such differences also have consequences in terms of economic fluctuations. If one wishes to study the consequences of non-homothetic preferences on macroeconomic development with a macro-model one first has to get rid of the representative consumer, and then give different consumers different non-homothetic preferences. This used to be too difficult to be a feasible modelling option in the DSGE framework, and DSGE modellers thus typically treated homothetic preferences as an auxiliary assumption. One could thus conclude that homothetic preferences were not a critical assumption. Since Iacoviello (2005), economists have found ways of modelling such heterogeneity, and it is now known to affect various results. The status of this assumption has thus changed from being an auxiliary to a critical unrealistic assumption. The Dixit-Stiglitz aggregator became popular in macro models primarily because it provided a tractable way of modelling *some form* of less than perfect competition. The fact that it implies specifically monopolistic competition is not really considered an additional asset. Ideally, macroeconomists would like to employ a fully realistic array of market forms, but thus far including oligopolies and monopolies together with monopolistic competition has not been tractable. We will thus only know whether the assumption of universal monopolistic competition is critical if this tractability problem is solved.

Furthermore, theoretical and methodological criticism of the DSGE framework has put the causal interpretability of its microeconomic core in doubt (see e.g. Colander, Howitt, Kirman, Leijonhufvud, & Mehrling, 2008; Kirman, 2010a, 2010b): It has been claimed that the use of representative-agent constructs is not theoretically justified because results within general equilibrium theory show that the rationality properties of individuals do not aggregate nicely. In effect, a consequence of using such constructs in DSGE models is that economic phenomena resulting from coordination problems among heterogeneous agents cannot arise by assumption. One cannot safely draw causal conclusions about alternative policies with such models, because the reactions of the representative agent to shocks or parameter changes during economic crises may not coincide with the aggregate reactions of the actual agents. These arguments have been used to imply that the DSGE model

cannot be interpreted as a causal model of the market economy based on assumptions about the market mechanism. Such a criticism would therefore also invalidate the rationale for model selection. We do not further dwell on these arguments here, however.

Another set of contentious methodological issues relates to the empirical assessment of models. Rodrik admits that economic models cannot really be tested against the specific empirical features they are built to capture, and that checking the secondary implications of the modelled mechanism is therefore important in model selection. Although the official empirical DSGE methodology of model 'calibration' was originally intended to allow for the estimation of crucial model parameters (such as risk preferences or the subjective discount rates of consumers) from multiple independent sources of data, in practice it has often meant curve-fitting the model to available macroeconomic time-series or long-run averages (such as aggregate labour). The use of DSGE models in central banks roughly coincides with the development of estimable versions, starting with Smets and Wouters (2003), and current large-scale policy models are parametrized through a combination of calibration and Bayesian estimation.

The use of DSGE models in empirical work is routinely justified on the grounds that they allow modellers to impose probabilistic restrictions on the data, making identification and estimation of the structural model easier. With this in mind, let us consider Jon Faust's (2009) efforts to see how the Swedish central bank DSGE model (Ramses) behaves with respect to the short-term interest rate and consumption growth in a Bayesian estimation. The prior derived from the theoretical DSGE model, which is used in estimating Ramses, starts from a strongly negative contemporaneous correlation between these variables. A correlation of about –0.6 is most likely in the prior, and values near zero are considered quite improbable. Yet, it is a common finding that there is little systematic relation between these variables.[3] The Bayesian estimation approach combines the model, the prior belief and the data to form a new assessment of all aspects of the model. The posterior still fairly strongly favours a negative correlation with the most likely value of around –0.4, and once again values near zero are very implausible. Thus, the estimation of the Ramses model was based on a strong prior belief that consumption is quite sensitive to interest rates, and this prior belief continues to be reflected in the posterior. If this is how DSGE models help solve the problem of identification, we fail to see the benefit in imposing such probabilistic restrictions on the data which seem to systematically lead one astray (cf. Canova & Sala, 2009). More generally, some studies that do *not* start from DSGE-based priors even fail to find the technological shocks that are alleged to be driving the business cycles (Juselius & Franchi, 2007).

DSGE model selection in context

Despite the problems, there seems to be some justification for using these models. As Rodrik states, general equilibrium effects (i.e. consequences of the simultaneous interaction of all important markets) are among the core economic phenomena that tend to be deeply unintuitive and hence are especially in need of explicit modelling. DSGE has become the standard platform on which to conduct policy analysis in a general equilibrium setting. The models are also clearly superior to previous macro models in being able to incorporate imperfect competition, nominal rigidities and the non-neutrality of money. They now also have some claim to predictive success in that they seem to predict some key variables better than fully a-theoretical VAR models or expert judgement (Del Negro & Schorfheide, 2013; Smets & Wouters, 2003). All in all, proponents of DSGE models claim that they provide the best all-around performance and therefore constitute the preferred platform for discussing, debating and evaluating more specific macroeconomic questions (Del Negro & Schorfheide, 2013). Their use in empirical work is often justified by the claim that they are the only kind of models that have the necessary resources to provide coherent theory-based stories about what happens in the economy. As Tony Yates writes in his blog on macroeconomics and public policy:

Microfounded models are models which tell an explicit story about what the people, firms, and large agents in a model do, and why. What do they want to achieve, what constraints do they face in going about it? My own position is that these are the ONLY models that have anything genuinely economic to say about anything.[4]

The claim is thus that DSGE models are indeed able to provide, or at the very least have the resources to provide information on alternative causal hypotheses, and that they do this by telling a theoretically coherent story of individual behaviour. An interesting question then arises concerning what such claims amount to and what is their basis, given that DSGE models are based on the crucial assumption of intertemporal optimization, which is routinely falsified in empirical work. Before we explore this central issue, let us see how Rodrik's views on model selection fit current mainstream macroeconomics.

DSGE modelling involves significant amounts of seemingly ad hoc parameter adjustments, fiddling with the structural assumptions and multiple runs in producing what are judged to be reasonable predictions. As such, the models are more like platforms for integrating various kinds of data with expert judgement than alternative models capturing single or a few key causal mechanisms. To show how applied macro modelling differs from Rodrik's view of model selection, we next briefly explore the way in which DSGE models are used in policy-oriented research in the Bank of Finland as well as in more academic macroeconomic research. Although sharing the core defining DSGE features, models built for policy use tend to be more complex and estimated on a richer data set than 'academic' DSGEs built to answer specific theoretical questions. This is to be expected given that forecasting and the evaluation of policy counterfactuals require the inclusion of all relevant sectors of the economy with empirically plausible parametrization, whereas models aimed at answering theoretical questions should be as simple as possible so as to remain theoretically tractable.[5]

The central workhorse of the Bank of Finland is the Aino (2.0) model. Aino is BoF's main integrating macroeconomic forecasting model and the platform of choice for running policy scenarios. It encompasses 39 parameters and takes in data from 24 variables using combined methodology including calibration and Bayesian estimation. The model is country-specific in that it emphasizes features important to a small open economy, as well as being estimated with country-specific data (1995–2014). Although some specificities are built into the model (such as the structure of the banking sector and the economy as part of a currency union but having substantial trade with countries outside the union), many country-specific features are, in fact, modelling outcomes attributable to empirical estimation (such as the very rigid wage setting and the importance of technology-driven productivity shocks, reflecting the disproportionate importance of the telecoms industry during the period of estimation). We talked with a modeller and a policy advisor working in the BoF to get a better idea of the modelling practice.

There is an element of 'model selection' in this practice, in that for each modelling exercise the model has to be reprogrammed accordingly. As an example, let the Bank be commissioned to study the impulse response of an export shock. The 'modellers' now write a first version of the suitably modified program and study the initial results. The initial results are always problematic, at least with respect to some key variables.[6] The modellers therefore usually have to fiddle with the lags and expectations so as to find a better fit with the existing data. Although philosophers might find this curious, the practice is not perceived as problematic because the main culprit in these empirical discrepancies is well known: the purely forward-looking nature of the model (the rational expectation hypothesis). It is therefore natural to fix these problems by fiddling with the expectations (see also Alvarez-Lois, Harrison, Piscitelli, & Scott, 2008).

When 'the computer guys' have produced their first model version, they meet with economic advisors who know the details of what is going on in the Finnish and international economy. Further model development is then a dialogical process between the advisors and the modellers. It is far from evident how exactly the collapse of Nokia and the temporary Russian import restrictions affected the Finnish economy, for example. Obviously, there is no specific variable for Nokia's

profits or turnover in the model, but somehow such factors have to be included, and the modellers cannot usually decide how this ought to be done without substantial topic-specific knowledge.

In sum, even though big policy models are always customized for the specific country of application and for each given forecasting or policy-scenario evaluation, they always include all the main components (all the markets and their specific characteristics). In fact, many of the modifications are made to correct for the empirically implausible implications of the core 'blueprint'.

DSGE models come in many shapes and sizes, and those used in academic theory-driven research are usually considerably smaller. As such, they focus only on selected key mechanisms or effects at any one time. Such theoretical modelling is therefore likely to adhere more closely to Rodrik's account of model selection. Nevertheless, although simpler, these models are still computational and therefore have to be parametrized according to some set of data.

We interviewed[7] an academic economist about his DSGE-based research on the effects of a trade agreement or the effect of fiscal devaluation in a two-region (north–south) single-currency economy. Given that this is more theoretical research, the plausible results of such modelling exercises are not point predictions, but the signs and plausible magnitudes of the specific shocks in question. Nevertheless, because the results are based on a computational model, the parameter values have to come from somewhere, and there is considerable pragmatism in the way in which the models are parametrized: in addition to using straightforward empirical (macro) estimates, parameter values may be based on empirical micro studies, what is commonly used, what feels right, and what simply works in terms of accommodating the predicted dependencies to their best empirical estimates.

Some form of stability analysis is usually carried out, but there is no 'official' established methodology for this. Although there is therefore a substantial element of 'ad hocness' in this empirical calibration, one should keep in mind that the objective is not to provide point predictions, but to offer empirically grounded yet ultimately theoretical arguments. Whether various ad hoc assumptions, either in the model construction or in the calibration, are legitimate or not should be evaluated in light of *this* epistemic role.

Although theoretical DSGE models are constructed (or 'selected') to answer a particular question, the answers are always partly driven by the shared core model, as well as by the specific additions used in the exercise at hand (such as the two-region common currency model). As computational models, it also tends to be difficult to separate analytically the importance of the core and of the case-specific assumptions, and to derive theoretically isolated dependencies responding to the 'key mechanisms'. The response to criticisms targeted at the shared theoretical core of the DSGE model is swift and unambiguous: it is completely useless to criticize the model platform itself. DSGE is taken as the only reasonable choice, the best available tool for reasoning about interrelated multiple markets in a quasi-quantitative way. The common blueprint is seen as a necessary condition for a meaningful and theoretically rigorous macroeconomic discussion.

Microfoundations and ad hocness

DSGE models are certainly not loved by everybody. Olivier Blanchard (2016) bluntly states in a recent policy brief that they are based on seriously flawed assumptions, empirically estimated by means of unconvincing methods (with alternative methods having dramatic effects on the policy implications), and that the normative implications that are crucial for policy advice are consequently implausible. Having said that, he nevertheless defends their use, because even though they are flawed, he firmly believes that macroeconomics needs a shared modelling platform, a common blueprint, and that the basic DSGE modelling choices are 'obviously' the right ones.

How, then, can the 'seriously flawed' assumptions of the core New Keynesian model be the 'obviously' right ones? Most of the recalcitrant empirical shortcomings of DSGE models are direct consequences of the core assumptions of intertemporal maximization and rational expectations: shocks do not persist as long as observed, they do not result in hump-shaped responses, inflation and output are affected too rapidly by monetary shocks, and there is too little inflation inertia. Although these

empirical failings may be and are ameliorated by the addition of various frictions and other 'propagation mechanisms' to the models, the forward-looking nature of the core price-adjustment equation (the new Keynesian Phillips curve) in particular makes it hard for even the more developed models to capture central stylized facts about price dynamics – as is readily admitted by policy modellers.

The main function of DSGE models is to provide a general equilibrium platform for modelling the whole economy, thus allowing theoretical interpretation of the observed and predicted data. Central banks do not exclusively rely on DSGE models for forecasting, and the ECB and the Federal Reserve Bank of New York, for example, use them mostly for integrating data and expert judgement: they provide the 'coherent theoretical interpretation' of the forecasts and policy scenarios arrived at by multi-model and multi-method means (Alessi, Ghysels, Onorante, Peach, & Potter, 2014). Explicit microfoundations are considered necessary for providing both the coherent theoretical interpretation and the underlying structure necessary for conducting counterfactual policy analysis. These two functions of the theoretical core model are automatically equated because they were implicitly equated in the Lucas critique: only models and parameters that can be derived from the maximizing behaviour of individuals with rational expectations can possibly be structural (invariant under policy interventions).[8] As Smets, Christoffel, Coenen, Motto, and Rostagno (2010, p. 52) put it, 'The general equilibrium structure lends itself to telling economically coherent stories and structuring forecast-related discussions around it … They give a better feel for which parameters are likely to be policy invariant and which ones are not.' This makes it understandable why, despite their 'serious flaws', the chosen core assumptions are 'obviously the right ones'. As Blanchard states: 'Starting from explicit microfoundations is clearly essential; where else to start from? Ad hoc equations will not do for that purpose.' Ad hoc equations not derived from microfoundations are therefore not bad only in virtue of being epistemically dubious epicycles, but also by threatening the structural interpretability of the model.

The differences between macroeconomists and philosophers in their understanding of the term 'ad hoc' should be noted. The understanding of the term among mainstream philosophers is roughly the following: something is ad hoc if it is specifically designed for some particular purpose. [9] Thus an element of a theory is said to be ad hoc if it is introduced solely so that the theory will entail certain statements (accommodate certain evidence), and there is no further reason for including the ad hoc element. However, the term 'ad hoc' is more often used in macroeconomics to describe any model elements that are not backed up with an explicit maximization story. The main problem, as conceived of by most macroeconomists, is that the raw RBC model does not fit the data, but the new 'ad hoc' Keynesian additions cannot usually be justified on microfoundational grounds. New Keynesian DSGE models are thus typically internally inconsistent because they include a rational representative individual (in the sense that it optimizes intertemporally and employs rational expectations) who is simultaneously supposed to exhibit habit persistence or set prices as if they could never be reset again, for example. The add-ons are thus theoretically incompatible with the foundational assumptions.

The motivation for developing microfounded models is to avoid the Lucas critique by developing models in which the policy-invariant parts can be clearly distinguished from parts that change with policy, and those who vigorously demand microfoundations think that applying microeconomic theory to the representative agent provides a satisfactory microfounded model. Microfoundations are based on the notion that the behaviour of individuals is targeted at maximizing some objective function, hence one can avoid imposing 'ad hoc' restrictions on macroeconomic variables. As Faust (2009) notes, however, there is a weak and strong form of the microfoundation requirement. The former merely requires consistency of macro models with microtheory, whereas the latter also requires consistency with microdata. Wade Hands (1988) argued long ago that RBC economists were using the notion of 'ad hoc' in the sense now associated with Lakatos[10]: an assumption or hypothesis is ad hoc if it does not follow the dominant heuristics in a given field of research. Although most current DSGE modellers seem to hold on to this idea, is it justified to posit that mere

inconsistency with microtheory is always ad hoc? Note that this question can be asked independently of the validity of the Lucas critique.

We surmise that there are two distinct underlying reasons for thinking that assumptions of non-rational behaviour are ad hoc. First, as Harsanyi (1966) and Boudon (1998) argue, for example, one needs the theory of fully rational behaviour as a benchmark even for irrational behaviour, because one can only understand the latter as a deviation from the former. Thus, to understand irrational behaviour one needs to specify rational behaviour, which is taken as inherently understandable.[11] Although rarely explicated, we surmise that many economists hold a similar view. This is how we interpret the common plea for microfounded models emphasizing 'coherence', 'discipline', 'coherent stories' or 'precision'. Consider, for example, how Martin Eichenbaum describes the issue 'When I read Keynes, I have no idea what he means. What does it mean to understand? ... what the end of a complicated econometric procedure means in terms of agents' motivations. What do they think they are responding to? Using optimization models is just a way of being precise' (Eichenbaum, 2000).

Second, assuming irrational behaviour could be taken to be ad hoc because there are always several different ways in which one can be irrational (Wickens, 2010). McCallum (2000, p. 118) puts the point as follows: rational expectations are used because 'it is almost certainly unwise for policy to be conducted under the presumption that any *particular pattern* of expectational errors will prevail in the future' [emphasis ours]. This is particularly evident if the policy purports to take advantage of such errors. Rational expectations are thus taken not to be ad hoc because they provide a single theory-based solution to the problem of specifying expectations.

In contrast, non-rational microfoundations could be seen as offering an embarrassment of riches, making rigorous model comparison difficult if not impossible. One might argue that one needs fully rational agents to be able to see how deviations from rationality affect the economy. If fully rational models are to function in a successful way as such a baseline, they must allow comparison between different models: then one could at least hope to isolate the different factors affecting the economy.

Consider, for example, some standard ways of creating inertia in the models, such as assuming that part of the population is credit-constrained or exhibits habit persistence. Whatever proportion of constrained people one assumes, the assumption is to some extent ad hoc. It would perhaps not be ad hoc if it were possible to measure the actual proportion of credit-constrained individuals or the degree of habit persistence. However, even if there were available data on parameters that describe these issues, it would be difficult to interpret them with a model that assumes fully rational individuals (Hansen & Heckman, 1996). Thus, even though Chari, Kehoe, and McGrattan (2009) claim that macro models should be consistent with microdata, in practice they cannot be fitted to all relevant macro- and microdata simultaneously. This is why modellers typically test their models only with respect to a set of carefully chosen micro-parameters (Wren-Lewis, 2011).

Supplementing such considerations, Paul Romer (2015, forthcoming) notes that DSGE modellers impose various unobservable restrictions on their estimations. This critique meshes with Popper's insistence that ad hoc hypotheses are not independently testable (see Bamford, 1993). Summers (1991) argued that over-identifying restrictions make it very hard to determine whether the theory failed due to its logical structure or to particular auxiliary assumptions. Such problems are particularly acute because one kind of ad hoc fix is typically not sufficient to make the RBC model consistent with data, and because typically the models have to assume the existence of several different exogenous unexplained shocks (cf. Wren-Lewis, 2011). Finally, some argue that the dynamics of DSGE models are largely driven by ad hoc fixes (De Grauwe, 2011). The Smets–Wouters model, for example, struggles to distinguish between the elasticity of intertemporal substitution, habit persistence and the autocorrelation coefficient of the preference shock (Tovar, 2009, p. 15).

These problems could be taken to imply that the fully rational model fails to provide a useful benchmark after all: a benchmark must assist modellers in evaluating what depends on what, but if this is impossible, as critics argue, the justification for using 'microfounded' models must come from somewhere else. If people are not, in fact, fully rational, strong microfoundations requirements will never be fulfilled – not even in a possible future when all the modelling problems concerning

microfoundations will have been solved. The possibility that economists ultimately cling to 'micro-foundations' mostly for the Lakatosian reason is still very much alive.

Macroeconomic model selection reconsidered

In sum, there seem to be good reasons why modelling in macroeconomics is not practised in the way Rodrik prescribes. The whole point of macroeconomics, be it theoretical or policy oriented, is to take into account the general equilibrium effects across multiple interdependent markets. However, as macroeconometricians have become painfully aware, weeding these interdependencies out from macroeconomic time-series is next to impossible without theoretically backed constraints. Macroeconomics is practised for policy-advice purposes, and predicting the consequences of policy interventions is impossible without knowledge of the underlying causal structure. A sensible approach to finding such a theoretically backed structure would be to try to explicitly model the behaviour of the agents making up the economy – by providing microfoundations. In combination, the necessity of the common microfoundational core and the impossibility of abstracting or idealizing away parts of the economy (because of general equilibrium effects) mean that modelling proceeds in more of a vertical than a horizontal manner.

Nevertheless, microfoundations have to come from somewhere and the recalcitrant empirical short-comings implied by the RE assumptions, which in turn need to be corrected with apparently ad hoc modifications, cast some doubt on the appropriateness of those that are chosen. Kevin Hoover (2006) argues that Michael Woodford's strategy in Interest and Prices (2003) was to explore the current theoretical toolkit on the premise that these tools would be useful for policy *in the future*. As far as Woodford was concerned, the representative-agent model was only the starting point for a series of fuller and richer models that would eventually provide the basis for an adequate macromodel. Therefore, the current generation of models also deserves some credence. Hoover refers to this appeal to future vertical progress as 'eschatological justification' (see also Hoover, 2015), thereby challenging Woodford as follows:

> If a central banker asks for advice today, would Woodford give it? And on what basis? If the practical payoff of his research is only in its utility for some future analysis, then he should decline to give advice on the basis of Interest and Prices and related research (Hoover, 2006).

The impossibility of carrying out the microfoundational research programme may be one reason why Solow (2010) refers to the demand for microfoundations as 'generally phony'. Wren-Lewis also notes that it is impossible to tell whether some standard assumptions, such as 'all firms are identical', are critical to the model's results. This is often a judgment which cannot be shown to be correct because abandoning the assumption would prove intractable (2011, p. 142).

Should macroeconomists, then, be more open to alternative modelling platforms – to a more horizontal vision of macroeconomics? We see the following possible ways of proceeding. First, one could accept the demand for microfoundations and try to improve the models by improving the microfoundations. The problem is, of course, that the menu of options is severely limited. Although the idea that rational action is somehow explanatorily privileged, and that deviations from rationality are somehow automatically ad hoc, is fundamentally flawed, the worry that any alternative microfoundations would just introduce another set of ad hocness into the models should still be taken seriously. Even if assumptions concerning the behaviour of consumers and firms were more firmly rooted in empirical research in areas such as behavioural economics, they should still be implemented in a model, thus requiring at least some 'ad hoc' idealizations and abstractions in turn. As things stand, the prospects for truly behavioural macroeconomics are uncertain.

Second, one could take the stance that the DSGE framework somehow manages to capture structural macro-level dependencies regardless of the apparent falsity of the microfoundations. Something like this could be argued for on the basis of Don Ross's ontology of economics (Ross, 2014, chapter 5). DSGE models could be taken to capture patterns of cross-market dependencies, and any form of mechanistic justification for these structures in terms of individual behaviour would

just be a red herring. Such a stance would entail extreme pragmatism related to the myriad ad hoc assumptions involved in DSGE modelling: any modification improving the empirical performance of the model (Calvo-pricing being a prime example) should be acceptable regardless of whether or not it can be rationalized in terms of microfoundations.

Third, one could forgo the hope that any structure could be derived from economic theory and concentrate on more humbly empirical methods of causal inference. This is the stance taken by Kevin Hoover (2015), for example, and would also mean abandoning the whole DSGE programme as ill-founded theoretical hubris.

The crisis of 2009 has generated not just methodological critique but also new approaches that take some of it into account (Reis, 2018). Another noteworthy development is the emergence of an explicit model-comparison framework, the idea being to create a library of macro models that are easily accessible and that could thus be compared with little effort (see e.g. Wieland, 2012). We assume that such model comparisons would be a welcome addition to Rodrik's model-selection account because they would seem to allow an empirical comparison that is not restricted to 'micro-founded' models. As a matter of fact, central banks have always used several models simultaneously as tools for better judgment-making: only microfoundations fundamentalists believe in removing the role of judgment entirely (see e.g. Colander, 2011; du Plessis, 2010). Finally, although it has long been acknowledged to some extent, there is now more widespread acceptance among macroeconomists that several different kinds of models are needed. Central banks will continue to use whatever they find useful for their purposes, including DSGE models.

Notes

1. 'Implications of the Financial Crisis for Economics', Speech at the Conference Co-sponsored by the Center for Economic Policy Studies and the Bendheim Center for Finance, Princeton University, Princeton, New Jersey, 24 September 2010. Stiglitz (2011) disagrees with Bernanke.
2. Thus, when Bernanke and Gertler (1989), for example, formalized the financial accelerator in a general equilibrium framework they started a model family because several later papers used their framework. Similarly, when Iacov-iello (2005) built a model with 'heterogeneous' consumers with respect to borrowing constraints (and thus pro-pensities to consume) on top of Bernanke and Gertler's model, he started another model family. Yet, neither of these examples count as horizontal development because the models were constructed within a strictly theory-driven framework.
3. This is regularly mentioned in discussions on macroeconomic methodology (e.g. Gerlach, 2017). See e.g. (Campbell & Mankiw, 1989) for a review of the econometric results.
4. https://longandvariable.wordpress.com/.
5. Justiniano, Primiceri, and Tambalotti (2017) raise concerns about this apparently growing gap between theoretical and applied DSGE modelling.
6. This was especially true of the older, non-linearized Aino 1.0 model, which routinely produced 'completely crazy' initial results. The linearized and estimated 2.0 behaves much better. Whether this is attributable to structural improvements or to more efficient empirical accommodation due to linearization is a key question that is difficult to answer.
7. 13 September 2015, Helsinki. Notes available on request.
8. This is also given as an answer to the question of why no-one in the private sector uses DSGE models for forecasting: if you cannot do policy, you do not need to be Lucas-proof.
9. See e.g. Grünbaum (1976) for a classic, and Votsis (2016) for more recent discussion. We also note that some recent discussions of ad hoc hypotheses in philosophy emphasise the fact that all efforts to explicate them have failed, and they have not been able to come up with a clear notion of what they are (Hunt, 2012).
10. Lakatos in fact distinguished between three senses of ad hoc.
11. Although this is not the place to enter into this discussion, we believe this idea is fundamentally flawed (see Yli-koski & Kuorikoski, 2016).

Disclosure statement

Notes on contributors

Jaakko Kuorikoski is an associate professor of New Social Research at the University of Tampere. His main areas of specializations are philosophy of social sciences and philosophy of economics. He has published on scientific explanation, modeling, simulation and causal reasoning. See https://jaakkokuorikoski.wordpress.com/.

Aki Lehtinen is a university researcher at the Finnish Centre of Excellence in the Philosophy of the Social Sciences (TINT), University of Helsinki. He has written on modelling and robustness, game theory, and other miscellaneous topics in the philosophy of economics. He has also constructed computer simulation models of strategic voting and published the results in economics and political science journals. See http://www.mv.helsinki.fi/home/alehtine/akilehtinen.htm for further information.

References

Alessi, L., Ghysels, E., Onorante, L., Peach, R., & Potter, S. (2014). Central Bank Macroeconomic Forecasting during the Global Financial Crisis: The European Central Bank and Federal Reserve Bank of New York Experiences. Federal Reserve Bank of New York Staff Reports 680.

Alvarez-Lois, P., Harrison, R., Piscitelli, L., & Scott, A. (2008). On the application and use of DSGE models. *Journal of Economic Dynamics and Control, 32,* 2428–2452.

Balzer, W., & Dreier, V. (1999). The structure of the spatial theory of elections. *The British Journal for the Philosophy of Science, 50*(26), 613–638.

Bamford, G. (1993). Popper's explications of Ad hocness: Circularity, empirical content, and scientific practice. *The British Journal for the Philosophy of Science, 44*(2), 335–355.

Bernanke, B., & Gertler, M. (1989). Agency costs, Net worth, and business fluctuations. *The American Economic Review, 79* (1), 14–31.

Blanchard, O. (2016). *Do DSGE models have a future? PIIE Policy Brief.*

Boudon, R. (1998). Limitations of rational choice theory. *American Journal of Sociology, 104*(3), 817–828.

Boumans, M. (1999). Built-in justification. In M. S. Morgan, & M. Morrison (Eds.), *Models as mediators* (pp. 66–96). Cambridge: Cambridge University Press.

Buiter, W. (2009). The Unfortunate Uselessness of most "State of the Art" Academic Monetary Economics.

Campbell, J. Y., & Mankiw, N. G. (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual, 4,* 185–216.

Canova, F., & Sala, L. (2009). Back to square One: Identification issues in DSGE models. *Journal of Monetary Economics, 56,* 431–449.

Chari, V. V., Kehoe, P. J., & McGrattan, E. R. (2009). New Keynesian models: Not Yet useful for policy analysis. *American Economic Journal: Macroeconomics, 1*(1), 242–266.

Colander, D. (2011). The Keynesian method, complexity, and the training of economists. In A. Arnon, J. Weinblatt, & W. Young (Eds.), *Perspectives on Keynesian economics* (pp. 183–206). Berlin: Springer.

Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., & Mehrling, P. (2008). Beyond DSGE models: Toward an empirically based macroeconomics. *American Economic Review, 98*(2), 236–240.

De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic Theory, 47*(2), 423–457.

Del Negro, M., & Schorfheide, F. (2013). DSGE Model-Based forecasting. In G. Elliott, & A. Timmermann (Eds.), *Handbook of economic forecasting, volume 2, part A* (pp. 57–140). North Holland: Elsevier.

du Plessis, S. (2010). Implications for models in monetary policy. *Journal of Economic Methodology, 17*(4), 429–444.

Eichenbaum, M. (2000). IS-LM: General discussion. In R. E. Backhouse, & A. Salanti (Eds.), *Macroeconomics and the real world, Vol. 2* (pp. 83–93). Oxford: Oxford University Press.

Faust, J. (2009). The New macro models: Washing our hands and watching for icebergs. *Economic Review, 1,* 45–65.

Gerlach, S. (2017). DSGE models in monetary committees. In R. S. Gürkaynak, & C. Tille (Eds.), *DSGE models in the conduct of policy: Use as intended* (pp. 31–37). CEPR Press. http://voxeu.org/

Grünbaum, A. (1976). Ad Hoc auxiliary hypotheses and falsificationism. *The British Journal for the Philosophy of Science, 27* (4), 329–362.

Hands, D. W. (1988). Ad hocness in economics and the popperian tradition. In N. B. De Marchi (Ed.), *The popperian legacy in economics* (pp. 121–138). Cambridge: Cambridge University Press.

Hansen, L. P., & Heckman, J. J. (1996). The empirical foundations of calibration. *Journal of Economic Perspectives, 10*(1), 87–104.

Harsanyi, J. C. (1966). A general theory of rational behavior in game situations. *Econometrica, 34*(3), 613–634.

Hoover, K. D. (2006). A neowicksellian in a New classical world: The methodology of michael woodford's interest and prices. *Journal of the History of Economic Thought, 28*(2), 143–149.

Hoover, K. D. (2015). Reductionism in economics: Intentionality and eschatological justification in the microfoundations of macroeconomics. *Philosophy of Science, 82*(4), 689–711.

Hunt, J. C. (2012). On Ad Hoc hypotheses. *Philosophy of Science, 79*(1), 1–14.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review, 95*(3), 739–764.

Juselius, K., & Franchi, M. (2007). Taking a DSGE model to the data meaningfully. *Economics, 1*(4), 1–38.

Justiniano, A., Primiceri, G., & Tambalotti, A. (2017). Empirical DSGE models: From the great moderation, to the great recession and beyond. In R. S. Gürkaynak, & C. Tille (Eds.), *DSGE models in the conduct of policy: Use as intended* (pp. 49–60). CEPR Press. http://voxeu.org/.

Kirman, A. (2010a). *Complex economics; individual and collective rationality.* London: Routledge.

Kirman, A. P. (2010b). The economic crisis is a crisis for economic theory. *History CESinfo Economic Studies, 56*(4), 498–535.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science, 61*(3), 541–567.

Kydland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica, 50*(6), 1345–1370.

McCallum, B. T. (2000). Recent developments in monetary policy analysis: The roles of theory and evidence. In R. E. Backhouse, & A. Salanti (Eds.), *Macroeconomics in the real world, Vol. 1 (Vol. 6)* (pp. 115–139). Oxford: Oxford University Press.

Reis, R. (2018). Is something really wrong with macroeconomics? *Oxford Review of Economic Policy, 34*(1-2), 132–155.

Rodrik, D. (2015). *Economics rules: Why economics works, when it fails, and how to tell the difference.* Oxford: Oxford University Press.

Romer, P. (forthcoming). The Trouble with Macroeconomics. *The American Economist.*

Romer, P. M. (2015). Mathiness in the theory of economic growth. *American Economic Review, 105*(5), 89–93.

Ross, D. (2014). *Philosophy of economics.* Basingstoke: Palgrave Macmillan.

Smets, F., Christoffel, K., Coenen, G., Motto, R., & Rostagno, M. (2010). DSGE models and their use at the ECB. *SERIES, 1*(1), 51–65.

Smets, F., & Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association, 1*(5), 1123–1175.

Solow, R. M. (2010). Building a science of economics for the real world. *Prepared statement for the Committee on science and technology subcommittee on investigations and oversight.* https://science.house.gov/

Solow, R. M., & Touffut, J. (2012). Introduction: Passing the smell test. In R. M. Solow, & J. Touffut (Eds.), *What's right with macroeconomics?* (pp. 1–7). Cheltenham: Edward Elgar.

Stiglitz, J. E. (2011). Rethinking macroeconomics: What failed, and how to repair It. *Journal of the European Economic Association, 9*(4), 591–645.

Summers, L. (1991). The scientific illusion in empirical macroeconomics. *The Scandinavian Journal of Economics, 93*(2), 129–148.

Tovar, C. E. (2009). DSGE models and Central Banks. *Economics, 3,* 1–31.

Votsis, I. (2016). Ad Hoc hypotheses and the monsters within. In V. C. Müller (Ed.), *Fundamental issues in artificial intelligence* (pp. 299–313). Berlin: Springer.

Wickens, M. (2010). What's wrong with modern macroeconomics? Why its critics have missed the point. *CESifo Economic Studies, 56*(4), 536–553.

Wieland, V. (2012). Model comparison and robustness: A proposal for policy analysis after the financial crisis. In R. M. Solow, & J. Touffut (Eds.), *What's right with macroeconomics?* (pp. 33–67). Cheltenham: Edward Elgar.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy.* Princeton: Princeton University Press.

Wren-Lewis, S. (2011). Internal consistency, price rigidity and the microfoundations of macroeconomics. *Journal of Economic Methodology, 18*(2), 129–146.

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology, 21*(1), 19–36.

Ylikoski, P., & Kuorikoski, J. (2016). Self-Interest, norms, and explanation. In M. Risjord (Ed.), *Normativity and naturalism in the philosophy of the social sciences* (pp. 212–229). New York: Routledge.