Journal of Economic Methodology

ISSN: 1350-178X (Print) 1469-9427 (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

Modeling model selection in model pluralism

Till Grüne-Yanoff & Caterina Marchionni

To cite this article: Till Grüne-Yanoff & Caterina Marchionni (2018) Modeling model selection in model pluralism, Journal of Economic Methodology, 25:3, 265-275, DOI: 10.1080/1350178X.2018.1488572

To link to this article: https://doi.org/10.1080/1350178X.2018.1488572

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 22 Jul 2018.

Submit your article to this journal

Article views: 878

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=rjec20

JOURNAL OF ECONOMIC METHODOLOGY
2018, VOL. 25, NO. 3, 265–275
https://doi.org/10.1080/1350178X.2018.1488572

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Modeling model selection in model pluralism

Till Grüne-Yanoff[a] and Caterina Marchionni[b]

[a]Philosophy Unit, Royal Institute of Technology, Stockholm, Sweden; [b]Practical Philosophy, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland

ABSTRACT

In his recent book, Rodrik [(2015). *Economics rules. Why economics works, when it fails, and how to tell the difference*. Oxford University Press] proposes an account of model pluralism according to which multiple models of the same target are acceptable as long as one model is more useful for one purpose and another is more useful for another purpose. How, then, is the right model for the purpose selected? Rodrik roughly outlines a selection procedure, which we formalize to enhance understanding of his account of model pluralism and to advance the critical discussion.

ARTICLE HISTORY
Received 27 April 2017
Accepted 14 January 2018

KEYWORDS
Rodrik; model selection; pluralism

JEL CODE
B41

1. Introduction

A number of economists have recently addressed the methodological question of how unrealistic economic models can be used to justify explanations and policy decisions (Gilboa, Postlewaite, Samuelson, & Schmeidler, 2014; Rodrik, 2015; Sugden, 2000, 2009). The core idea is in the form of *model pluralism*, according to which multiple models of the same target T are acceptable as long as one model of T is more useful for one purpose P, and another model of T is more useful for another purpose P'. Consequently, confronting a model with empirical evidence no longer confirms or falsifies it in any universal sense, but rather facilitates the *selection* of the appropriate model for a specific target T for a particular purpose P. This raises several questions. What procedure should be used to select the right model for a specific purpose? What assumptions should such a procedure satisfy? Under what conditions, pertaining to real-world economics, is such a procedure likely to be successful? In his recent book (2015), Dani Rodrik offers an interesting account of how model pluralism coupled with a suitable selection procedure can deliver the right model for the purpose.

Our aim in this paper is to analyze Rodrik's selection procedure. In particular, we seek a deeper understanding of what is required for such a procedure to operate in such a way that model pluralism becomes a tenable position. To this end, we first identify 'gaps' in Rodrik's description of the selection procedure – assumptions that are necessary for selection to proceed, but which are not stated explicitly. Second, we fill these gaps with explicit assumptions that we believe are closest in spirit to Rodrik's account. Third, we question the conditions under which the completed procedure would be decidable, in other words, the procedure would select, for any given purpose, a single model from the multitude of candidate models. Decidability is a strong criterion for the selection procedure given that, in practice, modelers might be content with a small, manageable number of models from

CONTACT Till Grüne-Yanoff gryne@kth.se Philosophy Unit, Royal Institute of Technology, Teknikringen 76, 10044 Stockholm, Sweden

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

which to choose. However, 'small' and 'manageable' are vague concepts that seem inextricably context dependent. For the purposes of our paper, we idealize this pragmatic requirement as the stronger, but more precise 'decidable' and address it from a more pragmatic perspective at the end of Section 5.

In carrying out this analysis, we formalize the selection procedure that Rodrik describes only verbally. We believe that this formalization will help to make the procedure sufficiently precise to identify which conditions (ideal and otherwise) it should satisfy to ensure the selection of the right model for the purpose, and thus to advance understanding and critical discussion of Rodrik's account of model pluralism. The goal of our analysis is constructive. We consider Rodrik's suggestions to be very fruitful, and through our analysis we hope to make his account even more relevant and thereby to contribute to the progress of economic modeling.

The paper continues as follows. We summarize Rodrik's selection procedure in Section 2 and formalize it in Section 3. In Section 4, we consider assumptions that remain unspecified in Rodrik's description, but that need to be addressed to complete the model. Finally, in Section 5, we identify the conditions under which the procedure is likely to fail to select a single model.

2. Rodrik's selection procedure

Rodrik, in his account of economic modeling, breaks away from the idealized image of economic science as abiding by idealized models of theory testing (such as falsification). He offers valuable insights into how, when things go right, economists go about selecting the model that fits the purpose. In his description of the selection procedure, he also carves out a role for informal procedures of verification – whereby theoretical models are assessed, but might not yield the kind of assurance one expects from formal empirical tests. Commentators and practitioners seem to have paid little attention to such features. The aim of the latter is perhaps to hold up the assertion that economics endorses something akin to falsificationism, whereas the former, at least in some cases, press the charge that economics shields its models from empirical evidence altogether. In this sense, Rodrik offers a balanced and realistic picture of how economic models are used.

In the same spirit, Rodrik points out that no model fits all situations, and that different models are right depending on the purpose to which they are put. This is what we call *model pluralism*, a concept we find highly appealing. If it is to be useful, however, we need to find a way of selecting, from amongst a plurality of models, the one that is right for a specific purpose. According to Rodrik, the general feature of the right model for the purpose is that it 'highlights the dominant causal mechanism or channels at work' (p. 85) in the specific setting of interest. How, then, from amongst the plurality of models is it possible to identify the one that highlights the dominant causal mechanism? Rodrik (2015) describes a three-step procedure for selecting such a model: (1) selecting candidate models, (2) identifying the critical assumptions, and (3) carrying out verification procedures. In the following, we describe each of these steps in more detail, drawing on Rodrik's discussion.

(1) *Selecting the set of candidate economic models*. The set of candidate economic models encompasses all 'plausible' models of a given target and for a specific purpose. Like fables, such candidate models provide a narrative 'whose storyline revolves around clear cause-and-effect, if–then relationships' (p. 19). Furthermore, as with experiments, 'the value of models resides in being able to isolate and identify specific causal mechanisms, one at a time' (pp. 24 and 25). Beyond having a convincing storyline, they should also be simple, and they are typically required to be formulated mathematically to meet economists' standards of precision. Otherwise, few constraints are imposed on the initial set of candidate models. Yes, they should be 'plausible' (pp. 41, 46, 64, 96, 114, 121), 'reasonable' (p. 41), and 'make sense intuitively' (p. 99). However, though Rodrik acknowledges that economic models tend to be derived from first principles, he does not mention derivation from general economic theory as a constraint. In addition, even if the models are derived from first principles, the derivation does not impose any substantial constraints: 'General economic theories are no more than a

scaffolding for empirical contingencies. They are a way of organizing our thoughts, rather than stand-alone explanatory frameworks.' (p. 116).

(2) *Identifying critical assumptions*. In the second step, the candidate models are distinguished on the basis of their critical assumptions: 'A critical assumption is one that if modified in an arguably more realistic direction would produce a substantive difference in the conclusion produced by the model' (Rodrik, 2015, p. 27). Let us take Rodrik's own example of the perfectly competitive market model (Rodrik, 2015, pp. 27–28). When applied to the effects of taxation on cigarettes, the model predicts that raising the tax rate will increase the price of cigarettes. A tax will lead to an increase in price, for example, regardless of the exact number of tobacco firms on the market or of the firms' degree of rationality, at least within certain bounds. The model might make 'unrealistic' assumptions about the number of firms or their rationality in the sense that they diverge from the specific situation to which the model is supposed to be applied.[1]

One could be tempted to make the models more 'realistic' by changing the latter assumptions, so that they approximate the situations more closely, but such changes would not have any effect on the model outcome (i.e. the price of cigarettes) given a change in input (i.e. the tax). Thus, none of the particular assumptions that a model makes about these features (i.e. number of firms or rationality) is 'critical' – they could be different, but the model result would still predict the direction of change fairly accurately. Rodrik employs this notion of criticality not only to characterize assumptions, but also to compare models. According to him, models that share all their critical assumptions should not be considered different models (p. 96).

(3) *Carrying out verification procedures*. Once the critical assumptions have been identified, judgments have to be made concerning the extent to which they approximate the particular situation to which the model is supposed to apply. Such judgments, Rodrik claims, are to be made in light of empirical evidence. The procedure operative in economics is neither standard hypothetico-deductivism nor falsificationism, in spite of some economists seemingly endorsing them.[2] In Rodrik's view, verification only targets the empirical adequacy of *critical* assumptions and of their direct or indirect implications; it is only meant to judge whether a model is appropriate for a particular case, given a specific purpose; and it accepts or rejects a model not as true or false, but rather as good enough for the purpose. Accordingly, models are verified only in specific applications.

Model verification relies on 'some combination of four separate verification strategies' (Rodrik, 2015, pp. 93–112).[3] The first of these concerns whether the *critical assumptions* of a model sufficiently approximate reality in a particular context of application. Rodrik's illustration is as follows (p. 97). To discriminate between monopolistic competition and a perfectly competitive model of the market, it is necessary to establish whether or not firms have market power. This, in turn, requires the examination of prevailing conditions such as the number and size distribution of firms, and how easily new firms can enter the industry. The second strategy is to verify that the *mechanism* the model assumes to be operating does, in fact, operate in the particular case.

The third and fourth strategies involve the verification of the direct and incidental implications of critical assumptions, respectively. The former work for models that are not constructed specifically to explain or predict a given phenomenon, but are built from 'first principles': in such cases, the question of whether the results sufficiently approximate a particular variable of interest is not trivial (in contrast to models that were built solely to approximate such a variable), and thus is subject to *direct* verification. The aim in the fourth strategy is to verify whether the model allows the derivation of further implications (possibly with the help of further background assumptions) that approximate the intended real situation. For example, a model in which public spending is a buffer – a source of social insurance and a stabilizer for economies that might otherwise be subjected to extensive foreign shocks – is proposed as an explanation of the observed correlation between exposure to international trade and the size of a country's public sector. Being built to account for this correlation, the model cannot be directly verified by it. However, 'the size of the public sector, upon analysis [of the model, possibly involving additional assumptions about determinants of public sector size], would appear particularly sensitive to fluctuations in the economy, rather than exposure to trade per se' (p. 62). This further implication of critical assumptions could then be tested against the data, and thus allow for *incidental* verification.

3. A model of the selection procedure

To investigate the steps of the procedure in more detail, we provide a formal reconstruction of Rodrik's account. Our goal is to make his account more precise and to unearth some of its implicit assumptions. Reconstructing the procedure by means of a formal model also brings to the fore the conditions under which it might fail to yield a single model. Such an undertaking might be at odds with Rodrik's caveat that choosing the right model for the purpose is not purely a mechanical procedure, it is also a craft (p. 83). Although we appreciate this point, our aim in this paper is to investigate how far we can push the notion of a 'mechanical procedure' to learn more about the selection processes required for a tenable pluralist position.

Our formalization of Rodrik's account takes models to consist of an ordered tuple $<A_1, \ldots, A_n, T, D>$ of assumptions $A_1, A_2, \ldots, A_n$, theoretical base T, and derivational rules D.[4] All these elements are propositions. The theoretical base T consists of concepts, assumptions, and (sometimes empirically certified) principles from which models might be derived. In game theory, for example, T contains the concepts of player, strategy profile, and Nash equilibrium (amongst others) that provide constituents and solution approaches to all game-theoretic models. The only theoretical principle in these models, however, is the maximization of expected utility. Many more such principles often go into the models in other fields, hence the set T is typically larger. To give just one example, Krugman (1991) derives the spatial agglomeration of economic activity from the theoretical framework of monopolistic competition and new trade theory. Finally, not all models have an explicit theoretical base. All the assumptions in Schelling's (1978) model of segregation, for example, are tailor-made for the model in a way that proceeds autonomously from any particular economic (or sociological) theory. Hence, in this model T is empty.

Derivational rules D consist of logical rules such as modus ponens and mathematical techniques such as the identification of local maxima. They might also contain fallible inference rules such as statistical inferences drawn from simulation runs.

Constraints are imposed on the set of all logically possible models $\boldsymbol{M}$**.** The first constraint is that the models to be considered derive results that satisfy a relevant purpose. We model the *relevance* of M ($M \in \boldsymbol{M}$) for a given purpose as M implying at least one element of a set $\boldsymbol{R}$ of externally given input–output pairs $< I, O >$. I and O are ranges of inputs $i_1, i_2, \ldots, i_n$ and outputs $o_1, o_2, \ldots, o_n$. Inputs complement the assumptions either by specifying the parameters of the model or by introducing further elements. $\boldsymbol{R}$ is a representation of the model user's *purpose* for using a model, and thus must be distinguished from the model M. The basic idea is that the model user might want to know how specific instances $i_1, \ldots, i_n$ of I affect the variable O. The question might be only about a sign change in which case O only contains two instances o_1, o_2; it might be about a limited number of intervals $o_1, o_2, \ldots, o_n$; or it might be about a continuum of values, in which case $n \to \infty$. A model M is relevant for a purpose represented in $\boldsymbol{R}$ if M maps every input $i_1, i_2, \ldots, i_n$ from $\boldsymbol{R}$ to an output $o_1, o_2, \ldots, o_n$ from $\boldsymbol{R}$.

For example, if one's interest is in the effects on the market price of oil of introducing a price cap, then $\boldsymbol{R}$ contains all the pairs of real variables '< price cap, market price > within a specified range. Hence, the input is the addition of a price cap into the model, and the output is the consequence of this for the market price. Any model that implies an output o for every input i is relevant for such an $\boldsymbol{R}$. Models that either do not take i as an input (e.g. in which a price cap cannot be meaningfully introduced) or do not yield an output o for some i (e.g. in which introducing a price cap does not yield any result concerning production) are, therefore, not relevant according to our definition.

Rodrik also imposes the following constraints: *narrative relevance*, *simplicity*, *plausibility*, *reasonableness*, and *intuitiveness*, such that the model candidate set $\boldsymbol{M}_c$ is

$$M_c \subseteq M.$$

Rodrik does not characterize these notions any further. We will take them up again in the next section.

In the second step, the assumptions of each member M of M_c are checked for 'criticality.' An assumption $\hat{A}_k \neq A_k$ is critical when a model in which assumption A_k has been changed to a more realistic $\hat{A}_k$ produces, for at least one input, a different output than the unchanged model. To make this more precise, let $L_n\{A_1, \dots A_n, T, D, i_l\}$ be the logical closure of the model elements and an input i_l. We then check whether the change of A_k to $\hat{A}_k$ affects the model output for any of the relevant input–output pairs in $\boldsymbol{R}$ (this similarly holds for models with multiple critical assumptions). In other words, assumption $\hat{A}_k \neq A_k$ is critical if and only if for at least one $< i_l, o_l > \in R{:}L_n\{A_1, \dots, \widehat{A}_k, \dots., A_n, T, D, i_l\} \cap R \neq L_n\{A_1, \dots, A_k, \dots., A_n, T, i_l\} \cap R$. The intersection-with-R requirement here ensures that only the outputs of the two models are compared.[5]

The set of *critical candidate models*, $\boldsymbol{M}_{cc}$, is a refinement of M_c that contains only the models that do not share all critical assumptions with another member:

$$\boldsymbol{M}_{cc} = \{M_i \in \boldsymbol{M}_c | \text{ for all } M_j \in \boldsymbol{M}_c, i \neq j{:}\text{there is at least one } \widehat{A}_k \in M_i \text{ such that } \widehat{A}_k \notin M_j\}.$$

Members of $\boldsymbol{M}_{cc}$ are empirically verified in the third step. To avoid the complexities surrounding the issue of evidential support, here we simply assume that there is a set of propositions $\boldsymbol{E}_k$ that lend support to a given assumption $\hat{A}_k$. As mentioned above, Rodrik proposes that empirical verification obtains through four alternative routes, or a combination thereof.

Direct verification selects the models whose critical assumptions $\hat{A}$ are supported by the evidence:

$$\boldsymbol{M}_{vcc} = \{M_i \in \boldsymbol{M}_{cc} | \text{ for all } \widehat{A}_k \in M_i{:}\widehat{A}_k \in \boldsymbol{E}_k\}.$$

Mechanistic verification purports to show that the mechanism the model assumes to be operating to bring about output o_l from input i_l does, in fact, operate in the target. To model mechanistic verification, we introduce the concept of (a sequence of) intermediate results $S_0, \dots ,S_n$ that stand between input i_l and output o_l. In a mechanistic model, this sequence is interpreted as representing different stages of the target mechanism. Each S_i is (i) implied by the conjunction of critical assumptions $\{\hat{A}_k\}$ and its immediate antecedent S_{i-1} (where S_0 is input i_l: the first mechanistic stage) and (ii) in conjunction with the critical assumption it implies its successor S_{i+1} or output o_l (the last mechanistic stage). In the price cap example, the effect of the price cap on production could be considered an intermediate stage between the input (the introduction of the cap) and the output (its effect on the price of oil).

A model M is mechanistically verified when some of its intermediate results are supported by evidence:

$$\boldsymbol{M}_{mcc} = \{M_i \in \boldsymbol{M}_{cc} | \text{ for some } S_i \in M_i{:}S_i \in \boldsymbol{E}_k\}.$$

Direct implication verifies the model by its predictive capacity. Hence, it selects the models whose output, given a certain input, is supported by the evidence:

$$\boldsymbol{M}_{dcc} = \{M_i \in \boldsymbol{M}_{cc} | \text{ Ln}\{A_1, \dots, A_k, \dots., A_n, T, D, i_k\} \cap \boldsymbol{R} \in \boldsymbol{E}_k\}.$$

Incidental verification selects the models whose critical assumptions $\hat{A}_1, \dots ,\hat{A}_n$, when introduced into a different model with additional assumptions $A'_1, \dots .,A'_n, T', D'$ produce results R' that are supported by evidence[6]:

$$\boldsymbol{M}_{icc} = \{M_i \in \boldsymbol{M}_{cc} | \text{ Ln}\{\widehat{A}_1, \dots, \widehat{A}_n, A'_1, \dots, A'_n, T', D', i'_k\} \cap \boldsymbol{R}' \in \boldsymbol{E}_k\}.$$

We believe this is an adequate formal reconstruction of Rodrik's model selection procedure in the sense that it translates his informal descriptions into a formal system using set theory and propositional logic. However, apart from this desideratum of accuracy, we would also like our reconstructed procedure to be *decidable*, in other words, to give an unambiguous decision to select or not to select an arbitrary model M_i for a given R. As we reconstructed the procedure, to make it decidable we had to make various specifications for which we could not find a textual basis in Rodrik's descriptions. Hence, our formal reconstruction allows us to identify gaps in Rodrik's procedure, on which we focus on the next section.

4. Five procedural gaps

The formal reconstruction of Rodrik's selection procedure reveals the existence of five gaps, which need to be filled to give a decidable account of how model selection proceeds.

(i) *Inclusion criteria*. It remains unclear what criteria regulate the inclusion of a model as a candidate model in $\boldsymbol{M}_C$. Rodrik mentions a few criteria including being intuitive, reasonable, and plausible. However, he does not spell them out. This is the first gap in the procedure as the following specifications highlight. We propose distinguishing three kinds of criteria based on the model components to which they apply, namely the theoretical base $\boldsymbol{T}$, the derivation rules $\boldsymbol{D}$, and the assumptions $\boldsymbol{A}$. Rodrik does not use this terminology, which we propose merely to keep the three criteria apart conceptually:

Intuitiveness: T is a member of a theoretical basis $\boldsymbol{T}$, which is intuitive.

Reasonableness: D is a member of the set of reasonable derivation rules $\boldsymbol{D}$.

Plausibility: $A_1, \ldots, A_k, \ldots, A_n$ are members of the set of plausible assumptions $\mathbf{A}$.

Rodrik also requires *narrative relevance* as a constraint on models, but remains vague in its specification. Our proposal to fill this gap starts from our earlier specification of a model's relevance: a model M is relevant for a purpose only if M implies an element of $\boldsymbol{R}$ associated with this purpose. Note that such a condition would be trivially satisfied if these elements of $\boldsymbol{R}$ were already built into M. Such a model would not tell an interesting narrative – following Rodrik, to tell an interesting narrative requires a 'storyline that revolves around clear cause-and-effect, if–then relationships' (p. 19). We interpret Rodrik here as saying that the model must give different relevant outputs conditional on different inputs, and thus that the input–output results must not already be contained in the model assumptions:

Narrative relevance: $A_1, \ldots, A_k, \ldots, A_n \notin \boldsymbol{R}$. i.e. inferences are not trivial

(ii) *The dependence of critical assumptions on conclusions*. Rodrik's definition of a critical assumption as producing a substantive difference in the conclusion (when modified in a more realistic direction) is vacuous as long as one does not constrain the notion of *model conclusion*. Any assumption, of course, makes a difference to some conclusion, at the very least trivially to the deduction of the assumption itself. Consider, again, the earlier example of a perfectly competitive market model applied to the effects of taxation on cigarettes. There, the only critical assumptions concerned the slope of the demand and supply functions. By contrast, if our purpose were to find out about the effects of imposing *price controls* on the cigarette industry, assumptions about substitutes would be critical because the presence of substitutes affects the degree of competition in the industry. Hence, whether an assumption is critical depends on the purpose for which the model is used, which in turn determines what it means to modify the assumption in a more realistic direction. Rodrik, being aware of this purpose-dependence, explicitly states, 'what makes an assumption critical depends in part on what the model is used for' (p. 29). The nature and importance of this dependence remain vague, however, and require specification in our model.

To make the procedure decidable, some criterion on the basis of which purpose-dependence can be determined needs to be specified. In our reconstruction, we chose input–output pairs $< i_l, o_l >$ as such a criterion: input–output pairs specify the modeler's purpose as looking for a relation between i and o. Furthermore, we modeled purpose as addressing different kinds of contrasts. Let us go back to our example of the effect of introducing a price cap on oil. We could be interested in whether or not to introduce a price cap or whether to introduce a large or a small cap. Depending on the contrast, different model results become relevant. In the first case, we are simply interested in the sign of the effect, whereas in the second we are also interested in the quantitative difference. Consequently, different model assumptions turn out to be critical. The purpose of a model is thus specified by a set $\boldsymbol{R}$, which contains pairs $< i_l, o_l >$ that specify the input–output relations, as well as the contrast between them in which the model user is interested.

Our model clarifies the importance of this purpose-dependence. First, whether an assumption is critical or not depends on $\boldsymbol{R}$. Second, direct and mechanistic implications depend on the determination of critical assumptions and thus on $\boldsymbol{R}$. Third, the verification of direct and incidental implications directly depends on $\boldsymbol{R}$ or $\boldsymbol{R}'$. Consequently, any result from these steps in the selection procedure is valid only as long as neither the input–output relation nor the contrast in $\boldsymbol{R}$ changes.

(iii) *Critical assumptions and realisticness*. According to Rodrik, assumptions are critical if their modification *in a more realistic direction* produces a substantive difference in the conclusion. This implies that only a specific kind of (either possible or actual) modifications indicates whether or not an assumption is critical. Presumably, the rationale behind this restriction on modifications is that for the result of a model to be explanatory or predictive of a particular case, not all the model's assumptions need to be even 'approximately' true. If an assumption that agents have perfect or imperfect information does not make a difference to the result, then it is not critical and need not be satisfied in that particular case.

Some assumptions in economic models are introduced purely for reasons of mathematical tractability, however. These 'tractability assumptions' (Hindriks, 2006) such as whether the time is assumed to be continuous or discrete are either known to be false or have no clear empirical interpretation, and typically do not allow the kind of modifications Rodrik prescribes without compromising the tractability of the model. Yet, we still want to know whether the model's result is robust with respect to tractability assumptions. If it is not, we should not have much confidence in the result as it might just be an artifact of the particular set of tractability assumptions.

The need to establish which assumptions are critical in the broader sense of being necessary to the derivation of the result provides the rationale for recommending that the model's assumptions should be subject to a set of wider modifications than the one Rodrik has in mind. In other words, whereas some assumptions are to be modified towards greater realism, others should be replaced with other tractability assumptions, for which realisticness comparisons are not meaningful. Probing the robustness of a model result to this broad set of modifications is a practice that philosophers of economics have studied under the label of 'derivational robustness analysis' (e.g. Kuorikoski, Lehtinen, & Marchionni, 2010). It is argued that derivational robustness analysis plays a crucial role in economic modeling in that it allows the identification of which assumptions are crucial to the result independently of the possibility of comparing alternative assumptions with respect to their degree of realism.

The idea behind robustness analysis is that if a given result is found robust to different and independent tractability assumptions, it justifiably increases confidence in the result. Note that the justificatory role of robustness analysis depends on the different assumptions being independent of one another. How one should understand the requirement of independence is a matter of debate (see e.g. Lisciandra, 2017; Odenbaugh & Alexandrova, 2011), but for our purposes let us consider independent assumptions broadly as assumptions that do not share the same kind of error with respect to a particular result.[7] This broader set of modifications should act as the 'filter' (p. 26) through which the assumptions to be tested empirically are reduced to those that are really necessary to derive the result.

(iv) *Membership in* $\boldsymbol{M}_{cc}$. Two models M_i and M_j are treated as equivalent as long as all their critical assumptions coincide, even if they differ with respect to their non-critical assumptions. Given that the purpose of the criticality concept is to 'filter' the set of candidate models for application to a particular situation, this means that only one of these equivalent models should be included in the set $\boldsymbol{M}_{cc}$. The question of which one remains open, however. Rodrik does not tell us, and our specification in Section 3 also leaves it open. For a particular purpose defined by a given $\boldsymbol{R}$, it does not matter which one is chosen. However, if the purpose changes, this ambiguity might lead to unwanted path dependency as follows.

Let us assume that M_i and M_j differ with respect to two assumptions, A_{2i} and A_{3j}, whereas $A_{1i} = A_{1j}$. Because for purpose $\boldsymbol{R}$, only A_1 is critical, M_i and M_j are treated as equivalent. Only one of them – we arbitrarily chose M_i – is included in $\boldsymbol{M}_{cc}$, which is not a problem, as M_i needs to be verified only with respect to critical assumptions anyway (however, see below for problems with non-direct verifications). Now, let us suppose that the purpose changes from $\boldsymbol{R}$ to $\boldsymbol{R'}$. For $\boldsymbol{R'}$, A_1 and A_3 become critical. In practice, given that M_i has already been in use, it might be tempting to check only whether M_i and not M_j includes any additional critical assumption with respect to $\boldsymbol{R'}$. In this scenario, the criticality of A_3 would remain undetected. Our framework shows that whenever the purpose changes, one ought to revert to $\boldsymbol{M}_c$ and investigate all models therein for their critical assumptions with respect to $\boldsymbol{R}$, not just those that are already contained in $\boldsymbol{M}_{cc}$.

(v) *Verification procedures*. Of the four verification procedures, Rodrik describes both direct verification and the verification of direct implications sufficiently clearly for us to model them in a straightforward manner. This is not the case, however, with the two other procedures. The verification of mechanisms is presumably meant to provide empirical evidence of mechanisms – in other words, the causal pathway from an input to an output, as represented by a particular model. Following this understanding, in the previous section we formalized the procedure as providing evidence of intermediate results S_I, which are deduced from critical assumptions and previous intermediate results, and which in turn imply further results (including the final model output identified in $\boldsymbol{R}$). However, this interpretation gives rise to a number of ambiguities.

First, there is ambiguity with respect to the *individuation* of the intermediate results. It remains unclear whether one should consider only results that are explicitly derived in the model or whether one should also include the intermediate results of a theoretical refinement that might allow introducing additional mechanistic stages. Considering only intermediate results that are explicitly derived would bias the process of verification in favor of models without much structure: a model that simply determines output o as a function $f(i)$ of input i has no intermediate results and thus is less likely to be excluded by empirical evidence compared with a model with many intermediate results. This could perhaps be counteracted by more stringent demands on narrative relevance, but Rodrik remains vague on this point.

Second, there is ambiguity with respect to the role of non-critical assumptions. Verification supposedly only focuses on critical assumptions, which were identified through their impact on some input–output pairs of a given $\boldsymbol{R}$. Nevertheless, it is obvious and is often evidenced in economics that the same output can be produced from the same input in conjunction with the same critical assumptions through a number of substantially distinct mechanisms (for an example from behavioral policies, see Grüne-Yanoff, 2016). Two models M_i and M_j might be equivalent in their critical assumptions with respect to $\boldsymbol{R}$, for instance, but differ in their non-critical assumptions *Ai*1 and *Aj*1. In conjunction with the critical assumptions, these non-critical assumptions might end up generating different sequences of intermediate results in *Mi* and *Mj*. Because *Mi* and *Mj* differ in the sequences through which they connect inputs and outputs in $\boldsymbol{R}$, they could be interpreted as representing different mechanisms in spite of their being equivalent in their critical assumptions. Mechanistic verification, aimed at finding evidential support for the intermediate results of these sequences, thus compares and differentiates between models based on their non-critical assumptions. This need not be a problem in principle, but it goes against Rodrik's program of exclusively focusing on critical assumptions.

This insight furthermore spells trouble for the earlier stage of 'filtering' $\boldsymbol{M}_c$ into $\boldsymbol{M}_{cc}$. As discussed above, only *one* member of a set of equivalent models with respect to their critical assumptions is included in $\boldsymbol{M}_{cc}$. This reduces the relevance of mechanistic verification, as models with identical input–output relations, but different mechanisms could have already been excluded at this earlier stage.[8]

The verification of incidental implications purports to select a model by confronting its further implications with the evidence. Presumably, this does not mean *any* implication apart from the critical assumptions, the results in $\boldsymbol{R}$, or the intermediate results S_i (if it did, it would in effect mean that one evidenced the non-critical assumptions here). Rather, it seems that such 'incidental implications' arise when the question that motivated the model application in the first place is developed further. The initial stage in Rodrik's international trade example discussed above was to account for the correlation between exposure to international trade and the size of a country's public sector. A set of input–output pairs, $\boldsymbol{R}$ would be specified accordingly. In pointing out that the model implies the sensitivity of the size of the public sector to fluctuations in the economy, rather than to exposure to trade *per se*, Rodrik in effect refines and revises the original question to set $\boldsymbol{R'}$.

At least in some cases, the shift from $\boldsymbol{R}$ to $\boldsymbol{R'}$ requires the construction of a new model, as the previous one might not contain all the elements necessary to generate some result in $\boldsymbol{R'}$. Consequently, we chose to model the verification of incidental implications as the setting up of a new model, which consists of the critical assumptions of the original model in conjunction with new assumptions, and together allow the derivation of a revised result belonging to $\boldsymbol{R'}$.

However, we argued above that any change in $\boldsymbol{R}$ might lead to a change in the critical assumptions of any model in $\boldsymbol{M}_c$. This creates substantial ambiguity about which critical assumptions of the original model should be included in the revised model. In fact, given the change in $\boldsymbol{R}$, it might not even make sense to uphold the distinction between critical and non-critical assumptions. Again, this need not be a problem in principle, but it goes against Rodrik's program of exclusively focusing on critical assumptions.

5. When the procedure fails to select a single model

The basic idea of model pluralism is that multiple models of the same target *T* are acceptable as long as one model of *T* is more useful for one purpose, and another model of *T* is more useful for another purpose. Furthermore, it typically endorses the idea that different targets require different models, instead of attempting to subsume many targets under one unifying model. A necessary condition for the viability of model pluralism is the availability of a decidable selection procedure (Aydinonat 2018).[9] For a given target and a given purpose, this procedure should identify a single model (or at least a small subset of equivalent models) as the appropriate one. If such a selection procedure were not available, and an arbitrary large number of models had to be considered adequate for a given target and purpose, then model pluralism would not be a viable position.[10]

Rodrik endorses model pluralism. For him, successful selection leads to scientific progress and the ability to select the right model for the right (policy) problem. He considers progress as 'horizontal expansion' (p. 71). 'Newer models explaining aspects of social outcomes that were unaddressed earlier … Fresh models do not really replace older ones' (p. 67). 'Economic science advances by expanding its collection of useful cases' (p. 72). Notably, he does not specify any upper bounds on the cardinality of such a library. This raises the threat of an *embarrassment of riches*: when progress would turn into the production of non-processable white noise.

Rodrik sees his proposed selection procedure by offering a solution to such an over-rich diversity of models:

> This multiplicity can also be seen as problematic … yet models contain information about the circumstances in which they are relevant and applicable … This means that, in any specific setting, we can discriminate, *at least in principle*, between models that are helpful and models that are not. (Rodrik, 2015, pp. 72–73, our emphasis)

In previous sections, we have formalized Rodrik's model selection procedure and identified some gaps within it. Our assumption in this section is that these gaps can be filled, and a decidable procedure produced. On the basis of our reconstruction, we now discuss (i) when an embarrassment of riches occurs and (ii) the conditions under which the selection procedure fails to reduce this multiplicity to a manageable number.

Concerning (i). An embarrassment of riches will arise if the menu of candidate models, $\boldsymbol{M}_c$, is not sufficiently constrained. Given that Rodrik does not specify sufficient constraints, this is very much a live possibility – when plausibility or intuitiveness criteria are too permissive, for example. Concerning (ii), a selection framework will fail to reduce this multiplicity to a manageable number if the critical assumptions are underdetermined by the empirical evidence. This can occur in three different ways. First, the set $\boldsymbol{M}_{cc}$ remains too large because robustness analysis does not sufficiently filter the number of models. This might happen with economic models that are not derivationally robust: if in the extreme, all the assumptions in the models belonging to the set are crucial to the conclusions, then no model is discarded. Of course, the lack of robustness also raises doubts about the reliability of the models in the first place (cf. Cartwright, 2007; Grüne-Yanoff, 2011).

Second, empirical evidence might be scarce (e.g. $\boldsymbol{E}_k$ is smaller than the set of critical assumptions $\{\hat{A}_i\}$, or the set $\boldsymbol{R}$), because we are unable to experiment, or because of legal constraints on the data. Finally, there is a pragmatic limitation on how many models a diagnostic test for policy purposes can handle, even if the number of critical candidate models $\boldsymbol{M}_{cc}$ is not underdetermined by $\boldsymbol{E}_k$. The number of diagnostic tests is the product of the cardinality of $\boldsymbol{M}_{cc}$, the number of critical assumptions in each model M_i, and the cardinality of $\boldsymbol{E}_k$. If $\boldsymbol{E}_k$ and $\boldsymbol{M}_{cc}$ are large, the practical limit might soon be reached.[11]

6. Concluding remarks

Rodrik's diagnostic model test offers a novel and interesting alternative framework for choosing the right model for the purpose. By formally modeling this framework, we have revealed various gaps that needs to be specified further for the procedure to be decidable. We furthermore identified conditions under which this procedure would not yield a manageably small number of critical candidate models or that would require a test procedure with impractically many comparisons.

Notes

1. For a survey of different notions of 'realistic' and 'unrealistic,' see Mäki (2009).
2. The hypothetical-deductive model, for example, identifies *any* observable implication from (typically general) hypotheses, to either accept or reject this hypothesis as true or false through testing these implications.
3. Rodrik does not explicitly mention econometrics in his discussion of verification. Presumably, this is a simplification given that formal procedures for testing theoretical hypotheses against the data have their own (technical) challenges, as do other ways of testing theoretical hypotheses (e.g. experiments). We make the same simplification, though it would be an interesting project to reflect on how the selection procedure should be modified when such additional complications are taken into account.
4. Although we represent models with the help of set theory, we remain non-committal about the semantic view of theories holding that models are set-theoretic structures (Suppes, 2002).
5. Further complications arise if we allow for sets of assumptions, rather than individual assumptions, to be critical. A critical set of assumptions might contain (a mix of) assumptions that are jointly critical but not individually so, as well as assumptions that are both jointly and individually critical. We leave these complications aside.
6. Machlup (1955) and Nagel (1961) call this 'indirect confirmation'.
7. For a detailed account of error-independence, see Kuorikoski et al. (2010, 2012). For examples, in addition to these papers, see also Hands (2016).
8. Mechanistic verification is still relevant in terms of distinguishing between models with different input–output relations. However, these are the models that other verification procedures can deal with as well.
9. Model pluralism could also be interpreted as recommending the use of several models in the understanding of a specific target. This is the interpretation Aydinonat (2018) endorses. A more complicated version of our selection procedure could be employed to select the 'right set of models' for the purpose rather than 'the right model for the purpose.'
10. The availability of such a selection procedure is not sufficient to justify model pluralism. Even when the procedure successfully selects one model, or a manageable set of models, the doubt remains whether the model so selected *should* be legitimately trusted for the purposes of prediction, explanation, and intervention. This is a topic for another paper, however.
11. Rodrik's own example of a decision tree for navigating across potential models contains $4 \times 2 \times 3 = 24$ possible distinctions (p. 190), and this might well be close to the practical limit.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Till Grüne-Yanoff is the professor of philosophy at the Royal Institute of Technology (KTH) in Stockholm. His research focuses on the philosophy of science and on decision theory. In particular, he investigates the practice of modeling in science and engineering, develops formal models of preference consistency, and preference change and discusses the evaluation of evidence in policy decision-making. Till is the editor of the journal *Economics & Philosophy*.

Caterina Marchionni is University Researcher in Practical Philosophy, Faculty of Social Sciences, University of Helsinki. Her research mostly concerns explanation, modeling, and interdisciplinarity in economics and the social sciences more generally. She is an Associate Editor of the *Journal of Economic Methodology*.

References

Aydinonat, N. E. (2018). The diversity of models as a means to better explanations in economics. *Journal of Economic Methodology*, *25*. Advance online publication. doi:10.1080/1350178X.2018.1488478

Cartwright, N. (2007). The vanity of rigour in economics: Theoretical models and Galilean experiments. In N. Cartwright (Ed.), *Hunting causes and using them: Approaches in philosophy and economics* (pp. 217–261). Cambridge: Cambridge University Press.

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2014). Economic models as analogies. *Economic Journal*, *124*, F513–F533.

Grüne-Yanoff, T. (2011). Isolation is not characteristic of models. *International Studies in the Philosophy of Science*, *25*(2), 119–137.

Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, *32*(3), 463–483.

Hands, W. (2016). Derivational robustness, credible substitute systems and mathematical economic models: The case of stability analysis in Walrasian general equilibrium theory. *European Journal for Philosophy of Science*, *6*(1), 31–53.

Hindriks, F. (2006). Tractability assumptions and the Musgrave–Mäki typology. *Journal of Economic Methodology*, *13*, 401–423.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, *99*(3), 483–499.

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

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2012). Robustness analysis disclaimer: Please read the manual before use!. *Biology & Philosophy*, *27*(6), 891–902.

Lisciandra, C. (2017). Robustness analysis and tractability in modeling. *European Journal for Philosophy of Science*, *7*(1), 79–95.

Machlup, F. (1955). The problem of verification in economics. *Southern Economic Journal*, *22*, 1–21.

Mäki, U. (2009). Realistic realism about unrealistic models. In H. Kincaid & D. Ross (Eds.), *The handbook of philosophy of economics* (pp. 68–98). Oxford: Oxford University Press.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Hartcourt, Brace and World.

Odenbaugh, J., & Alexandrova, A. (2011). Buyer beware: Robustness analyses in economics and biology. *Biology & Philosophy*, *26*(5), 757–771.

Rodrik, D. (2015). *Economics Rules. The Rights and Wrongs of the Dismal Science*. New York: W.W. Norton.

Schelling, T. (1978). *Micromotives and macrobehavior*. New York, NY: W.W. Norton.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*, 1–31.

Sugden, R. (2009). Credible world, capacities and mechanisms. *Erkenntnis*, *70*, 3–27.

Suppes, P. (2002). *Representation and invariance of scientific structures*. Stanford, CA: CSLI.