Philosophy of Science (2023), **90**, 836–854
doi:10.1017/psa.2023.69

ARTICLE

Multi-Model Reasoning in Economics: The Case of COMPASS

Jennifer S. Jhun

Duke University, Durham, NC, USA

(Received 01 July 2021; revised 11 March 2023; accepted 28 March 2023; first published online 13 April 2023)

Abstract

Economists often consult multiple models in order to combat model uncertainty in the face of misspecification. By examining modeling practices at the Bank of England, this paper identifies an important, but underappreciated modeling procedure. Sometimes an idealized model is manipulated to reproduce the results from another distinct auxiliary model, ones which it could not produce on its own. However, this procedure does not involve making the original model "more realistic," insofar as this means adding in additional causal factors. This suggests that there are ways to make models more representationally adequate that do not involve de-idealization in the straightforward sense.

1. Introduction

In practice, economists constantly face model uncertainty—there is always some amount of misspecification, in that any particular model will fail to capture all the details of a target system at hand. After all, no model is an exact duplicate of its target; idealization and abstraction are commonplace, inevitable, and in plenty of circumstances useful. Even so, a particular model may fail to capture even all the relevant ones. What to do about misspecification is a pressing issue that needs resolution in practice; policymakers rely on models to make decisions—important ones, like how to set interest rates. But it's unlikely that economists will rely on anything but coarse models.

What does a practicing economist at a place like a central bank do? Given that *any* given model may be lacking, one might consult *multiple* models.[1] Now, the use of multiple models in order to grapple with a complex target is not new. Approaches that rely on robustness analysis do this.[2] Other projects require stitching together an

[1] See Binder et al (2019, 716) for this sentiment.

[2] What exactly that thing is varies from author to author. See authors such as Odenbaugh and Alexandrova, (2011), Levins (1966), Kuorikoski et al. (2010), Kuorikoski and Marchionni (2016), Parker (2013), Lloyd (2010, 2015), and Wimsatt (1981). For a case from economics, see Lisciandra and Korbmacher (2021). See Woodward (2006) for a more detailed taxonomy of different kinds of robustness.

explanation using different models that capture different aspects of the target, or even stitching together the models themselves. Yet other projects involve formulating a model that renders another candidate model redundant, such that it can provide all the same information (and perhaps then some). But the activity I focus on in this paper is different from any of these. It involves a rather sophisticated process aimed at reproducing the results of one model in another, different model. I hazard that this practice is not at all unusual; instead of thinking of such maneuvers as a last-ditch tactic, we should think of them as part and parcel of the scientist's repertoire of model management strategies.

The particular activity I have in mind is pulled from the Bank of England's stock of modeling tactics, and in what follows I will be leaning heavily on Burgess et al's (2013) detailed account of the modeling platform used there.[3,4] To set the stage, Section 2 offers a simplified account of how the Bank's main model, COMPASS, incorporates information about the financial sector, which is missing from the model itself. Perhaps rather unexpectedly, the next move is *not* to explicitly add the financial sector into COMPASS in order to achieve increased realism. Rather, information about how the economy would behave in response to an exogenous disturbance (called a *shock*) is first determined in an auxiliary model, analyzed outside both the auxiliary and the main models, and finally incorporated into the main model. The upshot of this process—the relation that the economist must establish between the auxiliary and main models, as well as the inferences that can be made with COMPASS afterwards—are novel epistemic products that cannot be read off of either model.

This glimpse into practice reveals interesting methodological strategies that economists actually engage in, and I situate this example within philosophy of science in two ways. First (Section 3), this highlights a usage of multiple models that has perhaps been underappreciated in philosophy of science in general. Second (Section 4), my considerations imply that getting models to appropriately represent their targets does not necessarily involve "de-idealization" aimed at increasing the realism of those models. Section 5 concludes.

2. COMPASS and the modeling environment

Before I proceed to the example, here is a little bit of background on the goings-on at the Bank of England. Every quarter the Bank of England's Monetary Policy Committee (MPC) publishes the *Inflation Report,* since late 2019 known as the *Monetary Policy Report*. It is a product of a collaborative effort between the MPC and staff economists at the Bank, documenting the policy decisions of the MPC along with a discussion of why the MPC has taken these particular decisions. The most recent monetary policy report available at the time of this writing is from February 2021. At that time, the MPC voted unanimously to maintain a Bank Rate of 0.1% (the inflation target is 2%), amongst other actions.

Perhaps surprisingly, it is a very readable document—this is motivated by the Bank's wish for transparency to the public. Not only does it take stock of the status quo, it includes some background about why the economy has reached its current

[3] The practice I detail here is not representative of the whole gamut of practices involved with using models to represent, though I do find it to be illuminating.

[4] The COMPASS platform is, ultimately, a coordinative device for expert judgments. See Kuorikoski and Lehtinen (2018).

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

state and reports how the MPC expects ways that this state might evolve. Counterfactual and conditional speculation is plentiful; the document is peppered with "what if" musings; for example, "If concerns recede more quickly as those most at risk from Covid are vaccinated, consumer spending could increase more rapidly. Alternatively, some people could remain cautious about social activities until vaccination is much more widespread and Covid cases have declined markedly, which could pose a downside risk to consumer spending" (Bank of England 2021, 7–8).

Since 2011, the Bank of England has used the **C**entral **O**rganising **M**odel for **P**rojection **A**nalysis and **S**cenario **S**imulation, or COMPASS, as its core organizing model. It is a medium-scale, open-economy, New Keynesian, dynamic stochastic general equilibrium (DSGE) model, and resembles core forecasting models in use at other central banks.[5] As a DSGE model, the model is of an economy that evolves over time, buffeted by exogenous shocks from without, and is microfounded—i.e. the macroeconomic entities behave as if they were individual agents, governed by microeconomic theory. As a New Keynesian model, prices and wages are sticky, and monetary policy can have real (as opposed to nominal) effects in the short and medium but not long run. Things, once disturbed, take a while to settle down rather than adjusting instantaneously.

COMPASS's predecessors include the Bank of England Quarterly Model (BEQM) from 2003 to about 2011, and prior to that in the 1990s and the early 2000s the Medium Term Macroeconometric Model (MTMM). Today's COMPASS is much more theoretically constrained than BEQM and is also simpler, containing only fifteen observable variables, which makes it much more manageable and helps the bank staff communicate more easily to the MPC. It is also easier to spot shortcomings of the model, which appear more readily to modelers. So it is *deliberately* simple.

But COMPASS is not alone. It is accompanied by a suite of models. And this is motivated by concerns about model misspecification—no model captures *all* the causal detail of the target system, and may be omitting key relevant ones. Economists therefore have to supplement their use of COMPASS with some of the models in the suite, over fifty of them which have a variety of jobs. The suite includes:

> (a) Models which articulate economic shocks and channels that are omitted from COMPASS.
> (b) Models which expand the scope of the forecast, by producing forecasts for additional variables not included in COMPASS.
> (c) Models which generate alternative forecasts for variables which are in COMPASS.
>
> (Burgess et al. 2013, 10)

In this paper I consider the first case (a), where the suite models are filling in gaps in the COMPASS model. COMPASS itself, like its doomed predecessor BEQM, lacks an

[5] In addition, they favor the New Keynesian DSGE setup because it "[incorporates] a well-understood baseline description of some key elements of the monetary transmission mechanism that policymakers agree are important . . . New Keynesian DSGE models also incorporate the notion that expectations and the stabilising role of monetary policy are important for understanding the economy . . . " (7). Some of these reasons thus are not merely theoretical, but sociological in nature.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

explicit financial sector; there is no banking entity, nor does it explicitly model financial channels.[6] It is also missing a number of other sectors that may be important, such as the energy sector, and makes a number of simplifying assumptions such as lumping together different kinds of taxes and positing a single interest rate (Burgess et al 2013, 40). And "capturing the effects of these missing channels in a coherent way is an important part of dealing with misspecification" (63). Furthermore, COMPASS and the suite of models are accompanied by a MATLAB toolkit called MAPS (**M**odel **A**nalysis and **P**rojection **S**ystem) and a user interface called EASE (**E**conomic **A**nalysis and **S**imulation **E**nvironment) that exists as a desktop application. The information technology (IT) infrastructure in which the models reside introduces new functionalities that were previously unavailable, allowing for model analysis and construction, in addition to supporting the multiple models in the suite that this new generation of forecasts must rely on due to the simplified nature of COMPASS. One can implement (certain kinds of) models in MAPS, particularly those in linear state space. It can help analyze data by providing tools for transforming data (e.g. detrending data or even incorporating data incompatible with the assumptions within a model in question), estimating equations (including with Bayesian methods), and performing simulations. It can help analyze a model itself, by providing functionality for investigating the mechanistic structure of the model, such as its impulse response functions (55).

To build an extended COMPASS model with a financial sector would make it more complex than it might be worth.[7] If it is important to account for the financial sector, one may have to supplement the use of COMPASS by reaching for an additional model (or perhaps multiple models) that would help analyze and incorporate information about financial frictions. One such model is the Gertler-Karadi (2011) model, a quantitative macroeconomic model that explicitly includes a formal banking sector that intermediates between household savings and firms. It is like COMPASS in that it too is a New Keynesian DSGE model, though it omits some details that COMPASS does include.[8]

The conceptual advantage of the Gertler-Karadi model is that it allows the modeler to explicitly introduce credit spreads. Different investments with the same maturity may have different credit qualities. One may have a better yield than another. If there is a credit spread, this means that the official interest rate the Bank sets and the effective marginal interest rate that households and firms face come apart. An increase in the credit spread indicates a credit crunch. If households and firms face a higher interest rate, it will put downward pressure on consumption and investment, pulling down GDP. That is, a widening credit spread manifests as a demand shock, negatively affecting aggregate demand and thus GDP.

---

[6] Not *every* macroeconometric model lacks the financial sector. Other models such as the MPS (an acronym for MIT, the University of Pennsylvania, and the Social Science Research Council), the predecessor to today's FRB/US model currently at use at the Federal Reserve, does feature distinct sub-blocks, one of which is the financial sector.

[7] Burgess et al. 2013 admit that while it seems more sensible to keep COMPASS simple than take on the extra cost of complexity that would come along with adding a financial sector, this could be something that changed in the future as the discipline develops more tools (7, fn. 15).

[8] E.g. "nominal wage rigidities; international trade in imports and exports; and the presence of 'rule of thumb' households" (87, fn. 152).

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

Burgess et al. (2013) assess how credit spreads could come about and affect other macroeconomic variables in the Gertler-Karadi model.[9] First, the economist computes trajectory paths for variables such as consumption or investment after banks are hit by shocks in the Gertler-Karadi model. (This is done in a way consistent with the COMPASS model, so that behavior from both models can be aligned.[10]) These trajectory paths are then imposed on the COMPASS model itself.

In order to incorporate information from the Gertler-Karadi model, the behavior it models needs to be appropriately reproduced in the COMPASS model, which MAPS aids in doing. Suppose the economist has observed, in the Gertler-Karadi model, how certain inputs (shocks) result in certain outcomes (e.g., yield an increase in credit spreads and thus affect consumption and GDP). The next step is to select shocks to COMPASS that would appropriately mimic this behavior. That is, they wish to reproduce the consumption behavior in COMPASS seen in Gertler-Karadi, though COMPASS lacks the vocabulary of credit spreads. Instead, Burgess et al. take the domestic risk premium ($\hat{\varepsilon}_t^B$) as the appropriate shock in COMPASS that will mimic the increase in effective real interest rates that households face due to a credit crunch, which decreases consumption.[11] This risk premium variable can be found in the COMPASS model in the consumption Euler equation (Burgess et al. 2013, 18), which is reproduced below.

$$\begin{aligned} c_t = {} & \frac{1}{1+\psi_C+\varepsilon_\beta(1-\psi_C)\varepsilon_C^{-1}}\left[\mathbb{E}_t c_{t+1} + \psi_C c_{t-1}\right] \\ & - \frac{1}{(1+\psi_C)\varepsilon_C+\varepsilon_\beta(1-\psi_C)}\left[r_t - \mathbb{E}_t\pi_{t+1}^Z + \hat{\varepsilon}_t^B - \mathbb{E}_t\gamma_{t+1}^Z\right] \\ & + (1-\omega_0)\frac{wL}{C}\left[w_t + l_t - \frac{\mathbb{E}_t w_{t+1} + \mathbb{E}_t l_{t+1} + \psi_C(w_{t-1}+l_{t-1})}{1+\psi_C+\varepsilon_\beta(1-\psi_C)\varepsilon_C^{-1}}\right] \end{aligned} \tag{1}$$

Equation 1 tells us how certain factors affect present consumption. The first line of the equation is the contribution of expected future consumption and lagged (past) consumption to current consumption. The second captures the effect of the real interest rate $r_t - \mathbb{E}_t\pi_{t+1}^Z$ (adjusted for the risk premium shock $\hat{\varepsilon}_t^B$). The last line of the equation marks the influence of contemporaneous labor income. What's relevant to us is the risk premium shock term $\hat{\varepsilon}_t^B$ in the second line; notice that it enters into the equation the same way as the real interest rate. The domestic risk premium shock will, in the short term, increase how much it costs to borrow given a particular real interest rate. If that cost is high enough, then in response households will postpone expenditures (consumption and investment spending). An increase in the risk

[9] Burgess et al. are using the term "shocks" as an umbrella term for exogenous disturbances to the economy, since the term usually is defined as an *unanticipated* disturbance.

[10] Even before this, Burgess et al. add one more tweak—adjusting the Gertler-Karadi model itself by replacing its monetary policy reaction function with COMPASS's reaction function to make diagnosing differences between the two models easier.

[11] In their justification of the domestic risk premium, Burgess et al. appeal to work by Cúrdia and Woodford (2009), who introduce financial frictions into the economy and offer an extended New Keynesian DSGE model to accommodate these.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

premium is a negative demand shock; so, domestic demand will decrease and GDP will fall.[12] It is this variable that Burgess et al. elect as the relevant shock in COMPASS that will capture the effects of financial shocks in Gertler-Karadi, so changes in the domestic risk premium proxy the effects of changes in credit spreads.

So: the Gertler-Karadi model tells us something about how something like consumption behaves given a financial disturbance. Burgess et al. would like to replicate those effects in COMPASS. They determine that the effects of financial shocks can be accounted for via changes in the domestic risk premium in COMPASS. They need to determine the values of the domestic risk premium that will drive that behavior. That means solving the following problem. In general, the future projection of the behavior of a variable is given by its past behavior and some combination of anticipated and unanticipated disturbances or shocks. The task, then, is to do a bit of reverse engineering: the problem is to figure out which combination of shocks will produce the projected behavior of the endogenous variables the economist would like to see. That is, the profile of values for the risk premium that will drive the behavior of consumption in a particular way.

Expert judgment is applied to figure out what the disturbances that will impact our target variable (for our case, $\hat{\varepsilon}_t^B$) should be. There is no particular procedure to determine which shocks are the appropriate ones to impose on COMPASS, though MAPS helps translate those in an algorithmic way once they are selected. It is only after the application of judgment about which disturbances to employ that the economist seeks suitable quantitative values for those shocks, which the MAPS programming helps generate by providing an inversion algorithm. In sum, *given* what the economist thinks the endogenous variable trajectory should look like, she extracts what exogenous variables *would* give rise to it.[13] The MAPS algorithm helps codify her judgments and deliver the quantitative values for the shocks that deliver the outcome of interest. This exercise takes place "off-model"—it is not a bit of reasoning one conducts using either COMPASS or the Gertler-Karadi model, but in the IT infrastructure. To supplement this exercise, other models in the suite, some of which document empirical data, can be used "to cross-check the projections in COMPASS, expand the forecast to cover more variables, and challenge the key judgements in the forecast" (Burgess et al. 2013, ii).[14] And because even if financial

[12] These returns, however, may also be subject to shocks.

[13] Even this step where the problem is "inverted" in order to reverse engineer the right solution is not free of judgment. The system of equations that characterizes the system is inverted under an appropriate *identification scheme* to deliver unique values for those exogenous shocks. There are two potential problems that may arise when it comes to the inversion. One is over-identification: there are more shocks than targets (conditioning paths). The other problem is under-identification, where there are more targets or conditioning paths than instruments (shocks).

[14] The actual procedure is even more complicated than the summary I have given. In Burgess et al.'s actual considerations they use two suite models—not just the Gertler-Karadi model, but another by Barnett and Thomas (2014). The latter is not a New Keynesian model, but rather a structural vector auto-regression model that can be estimated from data. Both predict that GDP will decrease in response to financial shocks that result in credit spread increases (though at different rates), and disagree about the behavior of inflation. One could reconcile these differences in any number of ways. For instance, the concept of lending is understood differently in each model (91), and diagnosing this difference explains the discrepancy. Or, for another, the uncertainty associated with the increase in inflation that Barnett and Thomas find is high enough to allow for the possibility that the inflation response was actually negative (91).

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

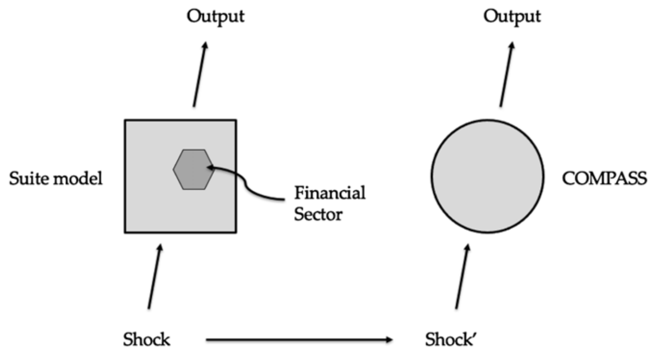

Figure 1. Schematic illustration of the modeling process.

frictions are captured in COMPASS, "the model will still be misspecified in some way," the economist might consult a class of models in the suite that offer alternative forecasts (47). Some of these models are statistical forecasting models of varying levels of complexity; others are suites of models designated to help study different sectors or particular quantities (such as the labor market, or inflation). To take a simple example, one might conduct a cross-check by comparing one's forecast with the weighted average of the statistical forecasts of many models (5).

Figure 1 above illustrates the difference between modeling in the suite model and modeling using COMPASS. Notice that the financial sector does not appear in the COMPASS model. Rather, the economist tries to identify a shock in the COMPASS model that brings about some relevant effect of interest. A shock that ultimately produces an output in the auxiliary model that has a financial sector will have to be translated into shocks recovered by the MAPS algorithm (which I have distinguished as *Shock'* on the right). Changes to the domestic risk premium affects consumption in COMPASS to reproduce the outcome of interest (say, the level of GDP). The result is that now COMPASS can provide a conception of how the economy would behave when hit by financial shocks. That is, COMPASS can be made to behave *as if* it had a financial sector. Whatever the effects on consumption in Gertler-Karadi due to financial shocks are, they can be imposed on COMPASS. But it will be generated by a set of shocks (which the economist must uncover) running through a different causal path than in its Gertler-Karadi counterpart (recall that it, but not COMPASS, includes a banking sector to account for financial frictions).

The upshot is a methodological one. Economists in practice often do not use models straight out of the box for forecasting purposes. To use COMPASS, the economist appealed to a structurally different model and coordinated the two in order to figure out which and how particular economic disturbances would lead to particular outcomes.

3. Model pluralism

Using multiple models is central in economic practice; it certainly is at central banks. There are a number of ways that philosophers of science have explored multi-model usage under the banner of "model pluralism." This literature attends to the fact that any particular model may be representationally and explanatorily inadequate and

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

endorses the use of a diversity of models in order to construct satisfactory economic explanations.[15] Rodrik (2015) notes that "The diversity of models in economics is the necessary counterpart to the flexibility of the social world. Different social settings require different models. Economists are unlikely ever to uncover universal, general-purpose models (5).[16]

In this section, I consider three ways of exploiting model pluralism. The first is what I call a triangulation project, while the second two are integration projects that involve constructing inter-model relations. The triangulation and integration approaches are distinguished by the ways they treat and reach conclusions about the system of interest. The target of agreement is different; triangulation approaches try to agree on something shared in common between models or methods, while integration approaches try to achieve agreement across models or methods offering up different—perhaps even conflicting—bits of information. The three strategies I consider are (a) robustness analysis, (b) model concatenation, and (c) making a model redundant relative to another model. None of these really captures what is going on in our case study set out in the previous section.

Consider robustness analysis. The idea is that if multiple models agree on something, that shared component has some kind of epistemic promise. What promise this is depends on which philosopher one is reading. Levins (1966) states, "If these models, despite their different assumptions, lead to similar results, we have what we can call a robust theorem that is relatively free of the details of the model" (20). There are a couple of noteworthy features of robustness analysis conceived in this way. The first is that this exercise is meant to clear away epistemic distractions or irrelevancies, pruning away unnecessary details of the model. The second is that the end product has some degree of generality. The robust theorem can be usefully applied across a wide range of scenarios. Of course, not all accounts of robustness are quite like this. Wimsatt (1981) focuses on separating the real from the illusory; this construal is neutral with respect to whether that common factor must take the form of a robust theorem. Robustness analysis is one way of guarding against the possibility that idealizing maneuvers will contaminate model results. That is, it provides "evidence that the result is not an artifact of particular idealizing assumptions" (Kuorikoski et al. 2010, 543). For example one might, by using multiple models, find that they all agree on a prediction.[17,18]

[15] A number of commentaries that explicitly address this issue appear in a 2018 special issue of *Journal of Economic Methodology* devoted to Dani Rodrik's *Economics Rules*.

[16] I take model pluralism quite generally. In their (2018) article, Grune-Yanoff and Marchionni mean something rather specific by model pluralism, "according to which multiple (2018) 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′*" (265).

[17] Parker (2013), in the context of climate modeling, remarks that with respect to robust predictions: "though different models and methods are used, similar predictions are produced" (219). Lloyd (2015) extends this further to argue that it is not only "the agreement or convergence of the empirically correct outcomes or predictions/retrodictions of a group of models" that matters, but also "the independent empirical support for the variety of assumptions and features of a span of models that all share a common 'causal core'" (58).

[18] See, for three different kinds of idealization, Weisberg (2007). What we have document here is probably closest captured by his notion of "multiple-models idealization," which involves incompatible models that have various trade-offs all together to give an accurate account of the way the world is. Yet, Weisberg's paradigmatic example of someone who endorses this strategy is Levins.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

Praising the virtues of robustness takes as granted that something *shared* between successful modeling attempts is the thing that is (potentially) *epistemically good* (perhaps in virtue of thinking that robustness is confirmatory). However, there is another way of using multiple models. Sometimes models are doing different things from one another, and despite this, economists are interested in getting them to cooperate *not* because they all share anything particular in common, but because they can illuminate different things about a larger, complex system. That is, in addition to the "triangulation" approach like a robustness analysis, economists will also engage in "integration" projects.[19]

Sometimes the integrative project involves combining information from two different models. This can sometimes be done by concatenating models. For instance, MacLeod and Nagatsu (2016) take two models in a "coupled-model" approach, where the models are "mathematically or otherwise combined such that information calculated or generated by one model is used as input to another to form an overall model-based solution to particular problems. Such a framework preserves for the most part the structure of those models" (423). Their particular case study joins an ecological model with an economic one by way of an optimization equation that brings together variables from both models, though each model remains intact. Ultimately, it is one model with two parts. Presumably, one would need this model if the system involves both ecological and economic factors that interact. But the relationship between COMPASS and the Gertler-Karadi model is different, as it is not the result of just *stitching the models together*, in the sense of concatenating them (even via intermediaries).[20]

There are other ways of thinking about integration; perhaps one model can be subsumed into another, in some sense. Reduction is one such avenue. For example, one theory reduces to another if it is deductively implied by the other.[21] We might thus think that a model of the reducing theory can capture everything a model of the reduced theory does. A thermodynamic model, the thought goes, should be formulable as a statistical mechanical model. Perhaps it would be epistemically better if we did so, because statistical mechanics is the more "fundamental" theory. Whatever information that one might find in the thermodynamic model, one can find in the statistical mechanical model.[22] But it should be clear that the relation between models that concerns us here is not one of reduction in either the epistemic or the ontic sense. These models yield different insights because they disagree on the economic structure they pick out; the set of equations that characterize one model will be different from that which characterizes the other, and neither set is

[19] See Kuorikoski and Marchionni (2016) for an explicit connection between triangulation and robustness (258).

[20] Several authors address *multi-scale* modeling frameworks which combine micro-, meso-, and macro-scale models of a phenomenon together. For nanocracks see Winsberg (2006a) and Bursten (2018), and for fluid flow see Darrigol (2013).

[21] For an overview of this discussion in physics, see Batterman (2020). Here, reduction (posed as a contrast to emergence) quite often takes center stage as a candidate inter-theoretic relation.

[22] Some usual reductionist suspects include Earman et al. (2002) or Rosenberg (2006). But as several philosophers across different sciences have noted, e.g. Batterman (2013, 2018), Bursten (2018), and Wilson (2006, 2017) in the physical sciences, Hoover (2015) in economics, and Parker (2009) in climate science, the reductive project as traditionally conceived looks rather unpromising.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

deductively implied by the other. Nor are the entities in one model constituted by entities in the other. The relationship between the Gertler-Karadi model and the COMPASS model is not at all mereological in the way that thermodynamic systems are made up of the stuff that statistical mechanical theories posit.

Nor can I appeal to something weaker like Hendry and Mizon's (1993) notion of "model encompassing." In this case, some econometric model of choice and a rival model are tested to see if the latter captures information that the former does not about the data-generating process. This again requires a choice between two models, with one of these models being made redundant because it can be nested in the other. But once again, sectors such as the financial sector in the Gertler-Karadi model do not correspond to anything in particular in the COMPASS model just as it is. Nor does one model capture all the information that the other does not, so one cannot be subsumed by the other. In general, it seems that the strategy is not making one of these models *redundant* at all.

That multiple models are often used in practice is not new from the perspective of philosophy of economics. Aydinonat (2018) suggests that using multiple models allows the scientist to produce *better* explanations than she could have produced without, because it allows her to produce a menu of "how-possibly" explanations for a phenomenon. By considering multiple possible models of a phenomenon, "The outcome of this complex and painful process of identifying, describing, measuring, exploring, conjecturing, verifying, and stitching is usually a patchwork of the most likely causal scenarios" (246). For example, Aydinonat considers using models of inequality to probe why it is that the wage-gap between skilled and unskilled workers increased during the 1980s. One might use a basic supply-demand model in order to assess how jockeying forces—the demand and supply of skilled and unskilled labor—might lead to changes in relative wages between the two groups. But another model like the Heckscher-Ohlin model addresses other factors such as globalization, technological changes, and institutional factors that might be explanatory power as well. For instance, immigration will change the relative supply of unskilled workers to skilled workers, and this may influence the wage-gap (244).

My focus is on fleshing out the inter-model relations that are constructed between the models themselves, and the kinds of adjustments that an economist might make in order to carry out explanatory work at all. So: one deploys multiple models in order to make up for their shortcomings.[23] More specifically, in the example of COMPASS, the baseline model does not yield all the information the economist needs (neither does the auxiliary model, for that matter). Therefore, one has to supplement the main model by appealing to information from outside that model—i.e., to the auxiliary (or auxiliaries) like the Gertler-Karadi model. In our case study, information about a causal path in the auxiliary model is, with some effort, imposed onto and reproduced in the main model. That is, a portion of COMPASS is recast to play the part of the financial sector—to proxy something that it does not explicitly represent—in order to simulate the effects of a financial shock. And importantly, this requires finding a set of shocks compatible with COMPASS, which will be different from those fed into the Gertler-Karadi model. The aforementioned ways of trying to connect different models

[23] For cases from other sciences, see Parker (2010, 2011) and Morrison (2011).

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

together already available in the philosophy of science (and economics) literature do not account for this activity.[24]

This case study demonstrates that (1) scientists often impose information onto a model from without, notably from other models which may differ or *very well disagree* in the central processes they capture, (2) in order to accommodate that information, models can be made to proxy things that they do not explicitly represent, and (3) it is the totality of the modeling process, and its various ingredients in concert, that tells us something about our target system.[25] The story I've told is, I think, a fairly general one about how scientists manage the models of complex systems. What motivates these activities are not ever-increasing gains in realism, in the sense of merely adding in what was missing in the first place, but something more subtle and rather involved.

The use of these sorts of strategies is not limited to economics.[26] In Winsberg's (2001) discussion of work done on the convective properties of certain kinds of stars (which require keeping track of fluid flows), we find similar maneuvers:

> . . . it is very important for researchers to keep track of the movement of kinetic energy in the system being simulated. But in any simulation of a fluid system that employs a discreet space-time grid, there is always the need to consider the impact of motions that occur "inside" the cells of the grid. In order to avoid losing energy to these local motions, simulations often keep track of a variable called "sub-grid-scale kinetic energy" that is calculated for each cell. But theory does not dictate what mathematical function should be used to "couple" this variable to the rest of the system. Therefore, the particular choice of how to calculate the interaction between this variable and the others of the simulation needs to be justified on some other basis. (S448)

This is not just adding those omitted lower-level processes back into the higher-level simulation model. To represent adequately, the original simulation has to *accommodate* information about these processes at smaller scales, which if anything, would require their own models at their characteristic scales. The information from the smaller-scale processes captured by the sub-grid-scale kinetic energy variable has

[24] One exception seems to be Hoyningen-Huene's (2022) in-depth discussion of Friedman's (1953) remarks on modelling a target "as if" it were something else using descriptively false assumptions that are not meant to be de-idealized. Hoyningen-Huene notes that Friedman, too, was interested in cases where some factor is substituted for another, which "has (approximately the same effect..but is easier to handle scientifically...The claimed effect equivalence...is preliminarily justified by a theory that connects [them]" (Hoyningen-Huene 2022, 158–159). In this paper, I have emphasized the details of how that reasoning process looks in a contemporary case, treating the whole enterprise as leading to an epistemic product and as an epistemic object in its own right.

[25] Or even, with respect to (2), to explicitly include things that do not exist. Winsberg (2010, 2006b) has also considered examples from fluid dynamics that include the incorporation of clearly unphysical elements like "artificial viscosity" which capture information from lower scales that could not be explicitly incorporated into one's model.

[26] As a further note, in general, the discussion of inter-model or inter-theory relations has proceeded within different subfields of philosophy of science in rather independent ways. Separately from the literature focusing on the physical sciences, Boumans (1999, 2005) has considered how one might forge such relationships in model construction, examining in detail three business cycle models from the history of economics.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

to be related (in a way, by the scientist's own hand) to the simulation at the larger scale in order to get it to generate realistic values for other variables of interest.

We can find this practice of sub-grid parameterization in climate science as well. Climate modeling also involves discretizing the differential equations describing the behavior of the system of interest to manageable algebraic ones, divvying up the modeled system into a grid of a certain resolution. Sub-grid modeling then involves "replacing missing processes—ones that are too small-scale or complex to be physically represented in the discretized model—by a more simple mathematical description" (Winsberg 2018, 48). Elsewhere, Winsberg and Harvard (2022) consider how one might use a less complex climate model over one that is more complicated, namely by imposing information onto the simpler model to get it to reproduce what one would have acquired with the more complex model:

> In complex climate models, a value of 'equilibrium climate sensitivity' (ECS) (a crucial feature of the climate system) is not assumed, but rather depends on the model's estimations of numerous mechanisms ... *Less complex models simply put a value of ECS into a model by hand* [emphasis mine]. The more complex models are obviously more 'microcausal' than the less complex ones. However, as long as the ECS value a model assumes is accurate, a less complex model could in principle deliver many of the same answers regarding the causal effect of a carbon pathway as a complex one. (3)

Possible parallels also arise in other areas, such as integrative systems biology. MacLeod and Nersessian (2013) describe a practice called "offlining" which "involves decoupling a variable from the dynamics of the system, because of a lack of information about how it connected to the system or because of the complexity of connecting it into the system, and instead feeding in its values by hand in order to get the model to fit the data" (544). Information is imposed onto a model that lacks a key variable in order to reproduce the relevant behavior, i.e. get it to behave as if it had incorporated that variable.

4. Idealization and de-idealization

Once we take (1)-(3) to summarize important ways of building inter-model relations, we notice that these complicate what it is for something to represent adequately (or for one thing to represent better than another). We tend to think of a model as representing successfully if, for instance, it represents accurately. The target is often too complex. But if we can include more of the model properties that map onto worldly properties, the better. The more "realistic" we can make a model, the better it is.[27]

[27] Authors such as Rice (2019) have pointed out that a model may not be decomposable anyway "into the contributions made by the features that are relevant and irrelevant to the target phenomenon" (180) where the accurate parts of the model correspond to relevant features of the target and the inaccurate stuff can be discarded. See Cartwright (1983, 1999), Batterman and Rice (2014), and Pincock (2022) on idealizations as essential features of models or theories.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

The idea is that the aspirational ideal—the complete model of the intended target—is the standard against which we assess the epistemic benefit of different maneuvers in the modeling process. A model is epistemically good insofar as it approximates the ideal. This way of thinking about modeling is wrongheaded. Sometimes models are not simple *just because* they are easier to handle. We might actually prefer it to the more complicated one; often *we* get some epistemic benefit from their simplicity. For example, a simpler model helps make salient what is *relevant*. So one would not want to include *all* the causal variables that might be at play at some given moment, but only the ones that are interesting given a certain time-scale and scope of the economy of interest. Furthermore, a simpler model's shortcomings may be easier to diagnose, as is the case with COMPASS.

Even if one doesn't buy into the story of the aspirational ideal, one might think something even weaker: that an adequate model should at least capture the *relevant* causal factors at play and discard the others. But notably, COMPASS is not made more applicable by explicitly incorporating more causal detail that was missing before. It *still* does not explicitly represent all the relevant elements of the economy (e.g. by positing a distinct equation for every mechanism to be found in the economy). Rather, the COMPASS model is, with the help of other models (and then some), *made to behave as if* it had a financial sector, rather than being *modified to have* the financial sector. Even in its final form, there is no explicit financial sector in the COMPASS itself. (And we see that there are similar moves made in other sciences.)

What this points to is that getting a model to be appropriately informative does not always require what gets called de-idealization, which is either the correction (or elimination) of those deliberately distorted features of the model (e.g. setting things to zero or infinity) or by explicitly inserting in additional information that had been initially omitted (e.g. neglected but actually present causal factors).[28] Philosophers such as McMullin (1985) took de-idealization to involve "eliminating simplifying assumptions " (261), and in a similar vein Nowak (1992) has suggests that we must "concretize" the idealized models.[29] But de-idealization conceived as mere correction would be an inadequate characterization of what goes into the enterprise of making a model applicable, in getting it in position to be able to contribute to explanatory questions. Increasing realism is often not the end goal.

Knuuttila and Morgan (2019) provide a more nuanced account of what "de-idealization" might amount to, documenting in detail four different strategies: "recomposing, reformulating, concretizing, and situating" (657). The first is exemplified by the removal of *ceteris paribus* conditions (649). The second helps "[make models] workable instruments" by altering certain formal constraints on the model's structure, e.g. its mathematical form (651). The third involves investigating "how such conceptual abstractions about the system, or the elements in it, are made concrete," e.g. interpreting utility as quality of life (652). The last involves engineering a model in order for it to be "made usable for specific situations in the world" (652) rather than be a general template, which can sometimes involve drastic (and even deep conceptual) alterations. They argue, more generally, against

[28] I have not distinguished, as Knuuttila and Morgan (2019) do, idealization from abstraction.

[29] See also Cartwright (1989, 1999).

the idea that models are first reached by idealization and that de-idealization ought to be conceived as reversing the steps taken in the idealizing process (641).[30]

All the above are involved in the case of using COMPASS. But let's consider the last kind, situating, which clearly applies to our case study. In their example, a "very simple iconic general model" like the supply-demand model can be situated "to be appropriate for categories or kinds of things in the economic world." One might wish to adapt such a supply-demand model to explain phenomena such as monopolistic competition as opposed to perfect competition (643). In our case, the baseline COMPASS model is situated to apply to an economy with a financial sector. The kind of "de-idealization" involved in getting COMPASS to adequately represent the economy involved getting it to behave a certain way, with the help of *another* model as a corrective. Insofar as there were changes to the COMPASS model itself, many of those changes were specifically aimed at getting COMPASS to accommodate information from the Gertler-Karadi model. I agree with Morgan and Knuuttila that an activity like correcting COMPASS does not seem to be a project of "reversing" idealizing moves in order to recover a more accurate model. Shoring up COMPASS in large part involves trying to align it in certain respects to another model, rather than trying to recover the ideal, complete model from which we initially fell short.

I have now laid out some of the peculiarities in the way economists engage with the COMPASS platform. We ought to pay attention to the coordinating strategies when multiple models—or more broadly, multiple sources of information—are in use that bring different kinds of information together, adjusting and sometimes transforming them. The final product, the forecasts it enables, and the causal knowledge that results, are epistemic products unavailable prior to exercising these coordinating strategies. Such strategies help us with a better grasp of the economy, but not by writing down a model that reads as more accurate than before.

5. Conclusion

This investigation suggests that the relationship manufactured by this imposition of information from one model to another is a *sui generis*, irreducible one. But there are a few loose ends. The first is a disclaimer. There is nothing in what I have said so far that indicates that a modeling platform with COMPASS at its center will be, in the end, an adequate one. Hendry et al. (2018), who believe both BEQM and COMPASS to be poor forecasting tools, are skeptical: "Unfortunately, key omissions in the suite of models contributed to the inability of the suite to properly correct for the omissions in COMPASS" (40). If the core organizing model is woefully flawed, then the narratives that we construct on the basis of that platform will surely also be flawed. It may very well be the case that a DSGE model is entirely unsuitable. But if this is the case, the problem is not one that might not be solved via robustness analysis of the kind that

[30] There is a separate line of thought that Carrillo and Knuuttila (2021) discuss, which concerns whether or not to think of idealization as an epistemic detriment or benefit. In the former, "The hope is that the advancement of science and the availability of better modeling methods could eventually deliver more accurate representations of worldly target systems" (52). However, "In contrast, the epistemic benefit account of idealization emphasizes the fact that in scientific modeling a detailed depiction is not often sought for" (52). See Batterman (2009) and Batterman and Rice (2014) for examples of the "benefit" account.

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

involves even extensive cross-checking mentioned in Section 2. Hendry et al. (2018) point to deeper conceptual difficulties. In such models one defect is that "all assets and liabilities can be aggregated into a single net worth measure to drive consumption . . . It seems incredible that pension wealth is as spendable as cash-in-hand" (323). For another, COMPASS uses a standard-looking Euler consumption equation to represent consumer behavior, and there is reason to think that it does not actually hold.[31] But worse, "these models were unable to capture the *possibility* of the kind of financial accelerator that operated in the US sub-prime crisis and the resulting financial crisis" (322).

Burgess et al. confess that the most recent developments were "an evolution instead of a revolution" (6). What Hendry et al. (2018) are suggesting is that a more radical move is needed.[32] Even if the best way forward is a different platform, Burgess et al. also note that the design of the central organizing model will have to balance a number of trade-offs along a number of dimensions. Harrison et al. (2005, 12) distinguish between "empirical" from "theoretical coherence," where: "For example, vector autoregression (VAR) models provide a good fit to the data ('empirical coherence') and dynamic stochastic general equilibrium (DSGE) models dictate that the model's predictions accord with the underlying theoretical assumptions ('theoretical coherence')." Not only do considerations of what causal processes should be considered as central drive the model, but considerations such as transparency, comprehensiveness, and tractability also motivate some of the design choices of the platform.

Second, I have also not discussed the infrastructure (or its history) in much detail; the MAPS algorithm that is used to back out shocks for a variable's trajectory is located in the infrastructure programming, not in any model. Explaining how the economy behaves in response to shocks makes reference to information and resources not limited to COMPASS. All of these things matter to the final projected forecast, even if they may not be obvious from looking at COMPASS and are not chronicled in the final *Monetary Policy Report*. The IT infrastructure, of which the MAPS algorithm is a part, is crucial to the economists who work at the Bank. So it seems plausible that infrastructure is a necessary ingredient in the process of making a model a workable, applicable instrument. This is in addition to the other tasks that the infrastructure helps the economist perform, such as providing a shared platform for a number of economists to perform different kinds of activities. Various models and modeling practices of economists have been documented in detail.[33] But there is usually little analysis of the infrastructure itself that accompanies modeling practices, though it is sometimes acknowledged.[34] That infrastructure is worth attending to is

[31] See Muellbauer (2010).

[32] Any framework that prioritizes attaining coherence, such as the COMPASS platform, is subject to the concern that in principle the whole project may be misguided. I also think that the final arbiter is probably whether or not the use of a particular platform is considered to be successful, which is a pragmatic consideration.

[33] See Morgan (2013) and Morrison and Morgan (1999) for detailed analyses of models in economics (and in the latter case, other sciences as well).

[34] Boumans (2004) documents that data filtering procedures encoded in computer packages in the late 20th century. He also notes that data filters qua instruments are also prone to producing artifacts, in the way any ordinary instrument would.

not often discussed in recent mainstream philosophy of science, with the exception of the literature on experiment (including that on instruments and even laboratories).[35] Much discussion can, however, be found in history of science, sociology of science, and science and technology studies.[36]

Lastly, I have not given an account of what expertise—which is required in order to make judgment calls, for instance about which models to privilege, which auxiliary models to use, what alterations to make and shocks to pick, and so on—actually amounts to, and when it has been exercised well. But I think that this means giving a full-bodied account of *bon sens,* which, unfortunately, I could not possibly do here.

Acknowledgments. I am especially grateful for feedback on earlier versions from Robert Batterman, Julia Bursten, Paul Hoyningen-Huene, Benjamin Eva, Kevin Hoover, Mary Morgan, James Nguyen, Elizabeth O'Neill, Meghan Page, Aleta Quinn, Alex Rosenberg, Walter Sinnott-Armstrong, Mark Wilson, and the anonymous referees.

References

Ackermann, Robert. 1989. "The New Experimentalism." *The British Journal for the Philosophy of Science* 40(2): 185–90. https://doi.org/10.1093/bjps/40.2.185

Aydinonat, N. Emrah. 2018. "The diversity of models as a means to better explanations in economics." *Journal of Economic Methodology* 25(3): 237–51. https://doi.org/10.1080/1350178X.2018.1488478

Bank of England. Monetary Policy Report. 2021. Accessed at: https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-report/2021/february/monetary-policy-report-february-2021.pdf?la=en&hash=3638A7091B34164428A54277B55BD6901709AA44 (May 2021)

Barnett, Alina and Ryland Thomas. 2014. "Has weak lending and activity in the UK been driven by credit supply shocks?" *The Manchester School* 82: 60–89. https://doi.org/10.1111/manc.12071

Batterman, Robert. 2009. "Idealization and Modeling." *Synthese* 169 (3): 427–46. https://doi.org/10.1007/s11229-008-9436-1

Batterman, Robert. 2013. "The Tyranny of Scales." In *The Oxford Handbook of Philosophy of Physics*, edited by Robert Batterman, 255–86, New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195392043.013.0008

Batterman, Robert. 2018. "Autonomy of Theories: An Explanatory Problem." *Noûs* 52 (4): 858–73. https://doi.org/10.1111/nous.12191

Batterman, Robert. 2020. "Intertheory Relations in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford: Stanford University Press. https://plato.stanford.edu/archives/fall2020/entries/physics-interrelate/.

Batterman, Robert, and Collin Rice. 2014. "Minimal Model Explanations." *Philosophy of Science* 81(3): 349–76. https://doi.org/10.1086/676677

Binder, Michael, Philipp Lieberknecht, Jorge Quintana, and Volker Wieland. 2019. "Model Uncertainty in Macroeconomics: On the Implications of Financial Frictions." In *The Oxford Handbook of the Economics of Central Banking*, edited by David G. Mayes, Pierre L. Siklos, and Jan-Egbert Sturm, 679–736. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190626198.013.23

Boumans, Marcel. 1999. "Built-In Justification." In *Models as Mediators: Perspectives on Natural and Social Science*, edited by Margaret Morrison and Mary S. Morgan, 66–96. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511660108.005

Boumans, Marcel. 2004. "The Reliability of an Instrument." *Social Epistemology* 18(2–3): 215–46. https://doi.org/10.1080/0269172042000249309

Boumans, Marcel. 2005. *How Economists Model the World into Numbers*. London and New York: Routledge. https://doi.org/10.4324/9780203324073

[35] Quite a few examples in philosophy of science can be found in the 80s and 90s. See, for instance, Hacking (1981), Cartwright (1989), Galison (1987), Ackermann (1989), and Mayo (1994).

[36] For work on infrastructure, see Pitt (1996) and Star and Ruhleder (1996).

https://doi.org/10.1017/psa.2023.69 Published online by Cambridge University Press

Burgess, Stephen, Emilio Fernandez-Corugedo, Charlotta Groth, Richard Harrison, Francesca Monti, Konstantinos Theodoridis, and Matt Waldron. 2013. "The Bank of England's Forecasting Platform: COMPASS, MAPS, EASE and the Suite of Models: Bank of England Working Papers 471," Bank of England. https://www.bankofengland.co.uk/working-paper/2013/the-boes-forecasting-platform-compass-maps-ease-and-the-suite-of-models

Bursten, Julia R.S. 2018. "Conceptual Strategies and Inter-Theory Relations: The ase of Nanoscale Cracks." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 62: 158–65. https://doi.org/10.1016/j.shpsb.2017.09.001

Carrillo, Natalia and Tarja Knuuttila. 2021. "An Artifactual Perspective on Idealization: Constance Capacitance and the Hodgkin and Huxley Model." In *Models and Idealizations in Science: Fictional and Artefactual Approaches*, edited by Alejandro Cassini and Juan Redmond, 51–70. Cham: Springer. https://doi.org/10.1007/978-3-030-65802-1_2

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. New York: Oxford University Press. https://doi.org/10.1093/0198247044.001.0001

Cartwright, Nancy. 1989. *Nature's Capacities and Their Measurement*. New York: Oxford University Press. https://doi.org/10.1093/0198235070.001.0001

Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139167093

Darrigol, Olivier. 2013. "For a Philosophy of Hydrodynamics." In *The Oxford Handbook of Philosophy of Physics*, edited by Robert Batterman, 12–42. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195392043.013.0002

Earman, John, John Roberts, and Sheldon Smith. 2002. "Ceteris Paribus Lost." *Erkenntnis* 57: 281–301. https://doi.org/10.1007/978-94-017-1009-1_1

Friedman, Milton. 1953. "The Methodology of Positive Economics," in Essays in Positive Economics, 2–43. Chicago: University of Chicago Press.

Galison, Peter. 1987. *How Experiments End*. Chicago: University of Chicago Press.

Gertler, Mark and Peter Karadi. 2011. "A Model of Unconventional Monetary Policy." *Journal of Monetary Economics* 58(1): 17–34. https://doi.org/10.1016/j.jmoneco.2010.10.004

Grüne-Yanoff, Till and Caterina Marchionni. 2018. "Modeling Model Selection in Model Pluralism." *Journal of Economic Methodology* 25(3): 265–75. https://doi.org/10.1080/1350178X.2018.1488572

Hacking, Ian. 1981. "Do We See Through a Microscope?" *Pacific Philosophical Quarterly* 62(4): 305–22. https://doi.org/10.1111/j.1468-0114.1981.tb00070.x

Harrison, Richard, Kalin Nikolov, Meghan Quinn, Gareth Ramsay, Alasdair Scott, and Ryland Thomas. 2005. *The Bank of England Quarterly Model*. London: Bank of England.

Hendry, David F. and Grayham Mizon. 1993. "Evaluating Dynamic Econometric Models by Encompassing the VAR." In *Models, methods and applications of econometrics*, edited by Greer Phillips and Peter C.B. Phillips, 272–300. Oxford: Basil Blackwell.

Hendry, David F. and John N.J. Muellbauer. 2018. "The Future of Macroeconomics: Macro Theory and Models at the Bank of England." *Oxford Review of Economic Policy* 34(1–2): 287–328. https://doi.org/10.1093/oxrep/grx055

Hoover, Kevin D. 2015. "Reductionism in Economics: Intentionality and Eschatological Justification in the Microfoundations of Macroeconomics." *Philosophy of Science* 82(4): 689–711. https://doi.org/10.1086/682917

Hoyningen-Huene, Paul. 2022. "Revisiting Friedman's 'On the methodology of positive economics'("F53")." *Methodus* 10.2(2022): 146–82.

Knuuttila, Tarja, and Mary Morgan. 2019. "Deidealization: No Easy Reversals." *Philosophy of Science* 86(4): 641–61. https://doi.org/10.1086/704975

Kuorikoski, Jaakko, and Aki Lehtinen. 2018. "Model Selection in Macroeconomics: DSGE and Ad Hocness." *Journal of Economic Methodology* 25(3): 252–64. https://doi.org/10.1080/1350178X.2018.1488563

Kuorikoski, Jaakko, Aki Lehtinen and Caterina Marchionni. 2010. "Economic Modelling as Robustness Analysis." *The British Journal for the Philosophy of Science* 61(3): 541–67. https://doi.org/10.1093/bjps/axp049

Kuorikoski, Jaakko, and Caterina Marchionni. 2016. "Evidential Diversity and the Triangulation of Phenomena." *Philosophy of Science* 83(2): 227–47. https://doi.org/10.1086/684960

Levins, Richard. 1966. "The Strategy of Model Building in Population Biology. *American Scientist* 54(4): 421–31.

Lisciandra, Chiara, and Johannes Korbmacher. 2021. "Multiple Models, One Explanation." *Journal of Economic Methodology* 28(2): 186–206. https://doi.org/10.1080/1350178X.2021.1892800

Lloyd, Elisabeth. A. 2010. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77: 971–84. https://doi.org/10.1086/657427

Lloyd, Elisabeth. A. 2015. "Model Robustness as a Confirmatory Virtue: The Case of Climate Science." *Studies in History and Philosophy of Science Part A* 49: 58–68. https://doi.org/10.1016/j.shpsa.2014.12.002

MacLeod, Miles and Michiru Nagatsu. 2016. "Model Coupling in Resource Economics: Conditions for Effective Interdisciplinary Collaboration." *Philosophy of Science* 83(3): 412–33. https://doi.org/10.1086/685745

MacLeod, Miles, and Nersessian, Nancy J. 2013. "Building Simulations from the Ground Up: Modeling and Theory in Systems Biology." *Philosophy of Science* 80 (4): 533–56. https://doi.org/10.1086/673209

Mayo, Deborah G. 1994. "The New Experimentalism, Topical Hypotheses, and Learning from Error." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* vol. 1994: 270–79. https://doi.org/10.1086/psaprocbienmeetp.1994.1.193032

McMullin, Ernan. 1985. "Galilean Idealization." *Studies in the History and Philosophy of Science* 16(3): 247–73. https://doi.org/10.1016/0039-3681(85)90003-2

Morgan, Mary S. 2013. *The World in a Model: How Economists Work and Think*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139026185

Morrison, Margaret. 2011. "One Phenomenon, Many Models: Inconsistency and Complementarity." *Studies in History and Philosophy of Science Part A* 42(2): 342–51. https://doi.org/10.1016/j.shpsa.2010.11.042

Morrison, Margaret and Mary S. Morgan. 1999. "*Models as Mediators*." *Perspectives on Natural and Social Science*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511660108

Muellbauer, John. March 2010. "Household Decisions, Credit Markets and the Macroeconomy: Implications for the Design of Central Bank Models." Bank for International Settlements Discussion Paper 306. https://www.bis.org/publ/work306.pdf

Nowak, Leszek. 1992. "The Idealizational Approach to Science: A Survey." In *Idealization III: Approximation and Truth*, edited by Jerzy Brzeziński and Leszek Nowak, 9–63. Amsterdam: Rodopi. https://doi.org/10.1163/9789004457249_002

Odenbaugh, Jay. and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology and Philosophy* 26(5): 757–71. https://doi.org/10.1007/s10539-011-9278-y

Parker, Wendy S. 2009. "Confirmation and Adequacy for Purpose in Climate Modelling." *Aristotelian Society Supplementary Volume* 83(1): 233–49. https://doi.org/10.1111/j.1467-8349.2009.00180.x

Parker, Wendy S. 2010. "Whose probabilities? Predicting climate change with ensembles of models." *Philosophy of Science* 77(5): 985–97.

Parker, Wendy S. 2011. "When climate models agree: The significance of robust model predictions." *Philosophy of Science* 78(4): 579–600.

Parker, Wendy S. 2013. "Ensemble Modeling, Uncertainty and Robust Predictions." *Wiley Interdisciplinary Reviews: Climate Change* 4(3): 213–23. https://doi.org/10.1002/wcc.220

Pincock, Christopher. 2022. "Concrete Scale Models, Essential Idealization, and Causal Explanation." *The British Journal for the Philosophy of Science* 73(2): 299–323. https://doi.org/10.1093/bjps/axz019

Pitt, Joseph. C. 1996. "Philosophical Methodology, Technologies, and the Transformation of Knowledge." *Society for Philosophy and Technology Quarterly Electronic Journal* 1(3/4): 146–58. https://doi.org/10.1007/978-94-007-0820-4_6

Rice, Collin. 2019. "Models Don't Decompose That Way: A Holistic View of Idealized models." *The British Journal for the Philosophy of Science* 70: 1790–208. https://doi.org/10.1093/bjps/axx045

Rodrik, Dani. 2015. *Economics Rules: The Rights and Wrongs of the Dismal Science*. New York: Oxford University Press.

Rosenberg, Alex. 2006. *Darwinian Reductionism: Or, How to Stop Worrying and Love Molecular Biology*. Chicago: University of Chicago Press.

Star, Susan L. and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7(1): 111–34. https://doi.org/10.1287/isre.7.1.111

Weisberg, Michael. 2007. "Three Kinds of Idealization." *The Journal of Philosophy* 104(12): 639–59. http://doi.org/10.5840/jphil20071041240

Wilson, Mark. 2006. *Wandering Significance: An Essay on Conceptual Behavior*. New York: Oxford University Press.

Wilson, Mark. 2017. *Physics Avoidance: And Other Essays in Conceptual Strategy*. New York: Oxford University Press.

Wimsatt, William C. 1981. "Robustness, Reliability, and Overdetermination." In *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*, edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William Wimsatt, 61–78. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2759-5_2

Winsberg, Eric. 2001. "Simulations, Models, and Theories: Complex Physical Systems and Their Representations." *Philosophy of science* 68(S3): S442–S454. https://doi.org/10.1086/392927

Winsberg, Eric. 2006a. "Handshaking Your Way to the Top: Simulation at the Nanoscale." *Philosophy of Science* 73(5): 582–94. https://doi.org/10.1086/518328

Winsberg, Eric. 2006b. "Models of Success versus the Success of Models: Reliability without Truth." *Synthese* 152: 1–19. https://doi.org/10.1007/s11229-004-5404-6

Winsberg, Eric. 2010. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226902050.001.0001

Winsberg, Eric. 2018. *Philosophy and Climate Science*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108164290

Winsberg, Eric, and Stephanie Harvard. 2022. "Purposes and Duties in Scientific Modelling." *Journal of Epidemiology and Community Health* 76(5): 512–17. http://dx.doi.org/10.1136/jech-2021-217666

Woodward, James. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13(2): 219–40. https://doi.org/10.1080/13501780600733376

Cite this article: Jhun, Jennifer S. 2023. "Multi-Model Reasoning in Economics: The Case of COMPASS." *Philosophy of Science* 90 (4):836–854. https://doi.org/10.1017/psa.2023.69