Prediction versus Accommodation in Economics

Abstract

Should we insist on prediction, i.e. on correctly forecasting the future? Or can we rest content with accommodation, i.e. empirical success only with respect to the past? I apply general considerations about this issue to the case of economics. In particular, I examine various ways in which mere accommodation can be sufficient, in order to see whether those ways apply to economics. Two conclusions result. First, an *entanglement thesis*: the need for prediction is entangled with the methodological role of orthodox economic theory. Second, a *conditional predictivism*: if we are not committed to orthodox economic theory, then (often) we should demand prediction rather than accommodation – against most current practice.

1. Introduction

In order to confirm a hypothesis, should we insist on *prediction*, i.e. correctly forecasting the future? Or can we rest content with *accommodation*, i.e. empirical success only with respect to the past? In this paper, I examine this issue in the context of economics. I will arrive at two main conclusions. The first is an *entanglement thesis*: the relative epistemic merit of prediction is entangled with the methodological role of orthodox economic theory. The second is a *conditional predictivism,* which fleshes out this entanglement: if we are not committed to orthodox economic theory, then (often) we should demand prediction rather than accommodation – against most current practice.

To be clear from the start on terminology: first, in both philosophy and science, 'prediction' can refer to several different things: the empirical implications of a model or theory for past data, or for data not thought of by the modeler, or for any data not used in a model's construction. In this paper, I will always mean the attempt to predict future data, and in particular will never use it to denote the accommodation of past data. I will also use the terms 'prediction' and 'forecast' interchangeably.[1] Second, I will mean by 'orthodoxy' the neoclassical models that dominate mainstream economic theory, i.e. formal models that deduce the equilibrium outcomes of interactions between economically rational agents.

To be clear from the start also on the paper's scope: the aim is to bring to bear on economics general considerations about the relative epistemic merits of prediction and accommodation. I have in mind using prediction and accommodation for the purpose of confirmation, i.e. for testing a theory or model. This in turn has consequences for policy advice, since presumably such advice should be given on the basis of a correct model for the case at hand. I do not wish to imply the different claim that prediction is intrinsically a worthier goal than explanation. Finally, I will not discuss the fascinating but separate issue of what are the best ways actually to achieve successful predictions, nor how these ways relate to existing forecasting methods in economics and econometrics (Ericsson 2017, Tetlock and Gardner 2015, Northcott 2017).

There is a considerable philosophy of science literature about prediction versus accommodation. The context for most of it has been the scientific realism debate: do we need successful predictions to justify claims that our theories are true, or are mere successful accommodations sufficient? Does it matter if the 'predictions' are of past data, so long as that data was unknown to the scientist when formulating their theory, or was not used in the theory's construction? The literature's focus has been on famous theories from physics and chemistry. But this paper's focus is instead on economics and is primarily methodological: is prediction necessary for economics to be useful and to develop successfully?

I begin by reviewing attitudes to prediction in economics, before discussing the general issue of when prediction should be favored. I then discuss one by one particular circumstances in which accommodation is endorsed, and thus in which prediction need not be insisted upon, considering each time whether that circumstance applies to economics. In light of this, I argue for an entanglement thesis, i.e. that the need or otherwise for prediction is not

[1] There is no uniform usage of these two terms across sciences anyway. In different cases, 'prediction' may denote any of: in-sample consequences of a model; extrapolation to new subjects; deterministic future earthquake claims; probabilistic future climate claims. Conversely, 'forecast' may denote respectively: forecasts strictly of future, out-of-sample data; forecasts, based on past data, only for known subjects; probabilistic future earthquake claims; deterministic future weather claims.

independent of our commitment or otherwise to methodological orthodoxy. The paper concludes by detailing exactly when we should be committed to prediction.

2. Attitudes to prediction in economics

A long tradition has doubted that systematic predictive success in economics (and social science more generally) is possible. Among the reasons: that economic systems are *open*, i.e. are chronically subject to significant influences from non-economic factors that are inevitably unmodeled; that economic systems exhibit *reflexivity*, i.e. that models themselves may influence their subject matter, thus creating a moving target[2]; or simply that there are usually too many significant variables interacting too complexly (Taylor 1971, Giddens 1976, Hacking 1995, Lawson 1997, MacIntyre 2007). According to Ludwig von Mises, "[predicting the economic future is] beyond the power of mortal man" (quoted in Rosenberg 1993, 53). Other distinguished pessimists include Weber, Durkheim, Popper, Winch and, more recently, Daniel Hausman, Alexander Rosenberg, Daniel Little and Deirdre McCloskey. (Those more optimistic include Milton Friedman and Julian Reiss.)

Such pessimism seems to be borne out in practice. Take GDP, for example: forecasts more than 18 months ahead fail to beat the naïve benchmark of assuming the same growth rate as today, and this weak predictive performance has not improved in 50 years (Betz 2006). Unemployment and inflation predictions fare little better. The situation is arguably even worse with respect to many financial variables. For instance, most studies confirm that forecasts of exchange rate movements cannot outperform the naïve benchmark of a random walk, and the same is true of stock market predictions too.

In the face of this apparent impossibility, or at least great difficulty, of prediction, it has been argued that the goal of economics should instead be explanation.[3] This methodological view stretches back to Mill, who argued (1843) that the ever-changing mix of causes in uncontrolled, field cases makes accurate prediction infeasible. As a result, theory should

[2] Arguably, the impossibility of accurate prediction implied by some rational expectations models can be thought of as a version of reflexivity.

[3] W. G. Runciman: "The proper function of a social science … is not prediction but diagnosis" (1963, 17). Herbert Simon: "We should be wary of using prediction as a test of science, and especially of whether economics is a science, for an understanding of mechanisms does not guarantee predictability" (1989, 100).

instead state core causal tendencies, such as human agents' tendency to maximize their wealth. In any particular application, we may compose relevant tendencies in a deductive way and then add in as necessary local 'disturbing causes' – i.e. causal factors not captured by theory but that are also present. In this way, deductive theory is claimed to be *more* empirically fruitful than narrowly predictive alternatives because it offers generalizability – i.e. the prospect of empirical success in many applications by adding in different disturbing causes each time.

Moreover, such Mill-type theorizing also offers the promise of understanding and insight. Indeed, these achievements are claimed to be *superior* to mere predictive success, even when the latter is possible. The reason is that what is of greater interest than predictive success in any particular case is knowledge of those factors or theoretical structures that generalize. It is the task of science, as a pursuit of systematic knowledge, to discover and isolate the latter. We might be interested in a particular structure of choices and incentives, such as, for instance, the Prisoners' Dilemma, because we understand this structure, and the outcomes to be expected from it, in terms of the discipline's fundamental building block of agent rational choice, and because we think it crops up in many different places. Therefore, rather than get lost in local detail, it is more fruitful to focus on the Prisoners' Dilemma structure – even though it may predict accurately in hardly any particular cases. An analogy is with the methodological role of *mechanisms* in other sciences, such as neuroscience: explanation in neuroscience, many philosophers have persuasively argued, is via appeal to mechanisms that, analogously to the Prisoners' Dilemma, are well understood and generalizable (Machamer et al. 2000 and many others). Accordingly, just as in neuroscience, it is knowledge of mechanisms that enables true explanation in economics, and our efforts should be directed accordingly. Moreover, it is knowledge of mechanisms that provides the understanding *of* any empirical success that we do achieve – and also provides understanding even in the many cases when our empirical success is imperfect.[4]

[4] Throughout, following Mill and the great majority of economists, I take economic theory to have a causal interpretation. Meanwhile, again following the relevant literature, I understand a 'mechanism' to be, roughly, an arrangement of entities related in such a way as to reliably produce particular effects. Accordingly, economic theory and mechanisms are conceived as playing similar roles. The paper's main points, about prediction, are not sensitive to the distinction between them.

On this view, then, while generalizable predictive success would of course be highly desirable, because of such success's infeasibility it is unfortunately best to abandon precise prediction as an epistemic goal.[5] Economic practice has overwhelmingly followed Mill's prescription and accordingly has de-emphasized prediction: "Forecasting has been regarded as the orphan of economics. Those who could do economics, did it; those who couldn't do economics, forecasted." (Ericsson 2017, 539)

Yet there has also always been some implicit resistance to this anti-prediction view. Much of economists' business is applied work, such as government or business consultancy, often involving cost-benefit and impact analyses. Much of this work is in turn forward-looking – in other words, implicitly or explicitly, it concerns prediction. This is not surprising, given the close relation between prediction and interventions and decision-making.

Independently of this, there are also familiar pragmatic reasons for prioritizing prediction over accommodation. The Popperian virtue of falsifiability is much easier to apply to forward-looking prediction than to retrospective accommodation – as Popper himself emphasized. Similarly, the epistemic vice of confirmation bias, i.e. of disproportionately looking for or noticing supportive rather than disconfirming evidence regarding one's own beliefs or theories, tends to be checked better by prediction than accommodation. So where does the balance lie?

3. When should prediction be favored?

Consider the kinds of explananda often faced by economists: Why did GDP grow by only 0.3% this year? Why are wages in local restaurants lower than six months ago? Why did the dollar rise in value today? In such cases, there are typically many plausible after-the-fact explanations. Perhaps, for instance, the dollar rose because of increased expectation of a rise in interest rates, or because of good domestic manufacturing data, or because of political trouble in the country of a major foreign currency, or because of profit-taking by traders who

[5] There is a serious worry about whether the theory-first method in economics delivers enough empirical success – either prediction or accommodation – to warrant claims even of explanation (Alexandrova and Northcott 2009, Northcott and Alexandrova 2013). Moreover, the evidence for the Mill-type causal tendencies themselves is often shaky. In these respects, economics is crucially disanalogous to the exemplar of neuroscience. But the focus of this paper is elsewhere, so I will not discuss these familiar controversies here.

had previously bid it down, or because it is regressing to its long-run equilibrium level given fundamentals, or because of some other unknown reason. The point is that many different plausible explanations are compatible with the same headline evidence, namely that the dollar rose. This multiplicity is helped by framing the explanandum qualitatively – we need to explain merely why the dollar rose without necessarily specifying by exactly how much. In contrast, typically there are only a few successful predictions. That is, whereas typically there are many plausible explanations after the fact, there are only few successful predictions beforehand. This lies at the heart of the intuitive case for prediction: it is better at filtering competing hypotheses.[6]

Can this intuitive argument be made good on formally? Much work in philosophy of science has done just that. A core conclusion is the central importance of the *likelihood ratio*. In particular, for evidence E, hypothesis H and rival hypothesis H*, degree of confirmation is a positive function of the probability ratio $p(E|H) / p(E|H^*)$.[7] In words, the key issue is how much the evidence discriminates for our hypothesis over salient alternatives. The heart of the intuitive case above is that successful prediction discriminates more. In the case of accommodation, there are many alternative hypotheses H* that (given their assumptions) entail E just as our hypothesis H does, so $p(E|H) = p(E|H^*) = 1$, and H is not favored over the alternatives.[8] But in the case of prediction, if no one else predicts successfully then $p(E|H) > p(E|H^*)$, and so this time the evidence does favor H over its rivals.

This pro-prediction reasoning does not depend on knowing the mindset of a scientist, or on what the scientist might or might not have known when formulating their hypothesis, or on when relevant evidence was discovered, or on any other such historical factors. Nor does it rely on any a priori prejudice in favor of prediction, or on any appeal to prediction's

[6] In statistical terminology, an advantage of prediction over accommodation is that it guards against *overfitting*.

[7] Howson and Urbach (1993) and Worrall (2014), among others, give detailed Bayesian presentations. A similar conclusion can be reached in non-Bayesian ways too.

[8] The hypotheses *entail* E because, in economics, results are typically deduced from a model's assumptions. Thus, $p(E|H) = p(E|H^*) = 1$. Whether a model's assumptions themselves hold may of course be investigated in turn. Such supplementary investigation might reveal that $p(E^*|H) > p(E^*|H^*)$ with respect to new *supplementary* evidence E*. In this way, the evidential tie may be broken by supplementary investigation, which indeed is one route by which accommodation may be endorsed (section 6).

pragmatic advantages. Rather, it is based purely on a logic of confirmation that is itself neutral. How might this pro-prediction conclusion be blocked and thus accommodation defended? There are several ways. In the following sections, I consider each of them in turn. Which, if any, apply to economics?

4. First defense of accommodation: No plausible alternatives

Suppose there are *no plausible alternative* explanations. Then the mere fact of accommodation will be decisive. Take the German hyperinflation of 1923, for instance. This is explained by Germany's rapid monetary expansion (relative to real production) in that year. Because there is no plausible alternative explanation of the hyperinflation, retrospective evidence is sufficient in this case. Formally, if E is the accommodated evidence, then in such cases $p(E|H) > p(E|H^*)$ for all salient H*, and thus accommodation confirms H. Intuitively, if no one else can accommodate the past evidence then the fact that you can tells strongly in your favor. This is accepted by almost everybody.[9]

At first sight, this defense of accommodation will apply to economics only rarely. It is hard to prove this claim in a non-anecdotal way but, as we saw, it does seem that often in economics there are many plausible explanations after the fact. However, there is an important caveat: in economics, often the set of admissible hypotheses is severely restricted on *non-evidential* grounds. In particular, hypotheses must obey rational choice orthodoxy in order to be considered seriously. Because of this, the set of hypotheses consistent with the evidence may become much smaller – and small enough to vindicate accommodation after all.

Here is an example to illustrate, taken from Julian Reiss (2008, 106-122). In a famous article, Milton Friedman and Anna Schwartz argued that money is the main cause of fluctuations in nominal income (1963). There exists statistical evidence for this claim. Friedman and

---

[9] Logically speaking, of course, there are always many hypotheses that can accommodate any given body of evidence – this is the familiar problem of under-determination. But methodologically speaking, the key issue is whether these alternatives, in addition to being logically possible, are also plausible or to be taken seriously. If not, then they are not of practical interest. (See also footnote 13.) Formally, in the Bayesian calculus this is reflected in the priors. Non-Bayesian approaches incorporate the same point in other ways; some, for instance, only *define* confirmation contrastively, i.e. as being a matter of evidence telling in favor of one theory over a particular alternative (Sober 1999).

Schwartz offered one explanation for it. In a more recent paper, Jess Benhabib and Roger Farmer (2000) offer an alternative explanation. Both of these explanations are consistent with the aggregate data, so at first glance there might seem to be an evidential tie.[10] However, at one point Friedman and Schwartz's model assumes that agents are subject to money illusion, thereby offending rational choice orthodoxy. Their model is also rather informal by modern standards. Benhabib and Farmer's model, by contrast, satisfies orthodox desiderata much better: it is presented mathematically, it features agents maximizing formally specified utility functions given the structure of incentives facing them, and the outcomes are equilibria of those agents' utility maximizations. This in turn is why Benhabib and Farmer's paper is celebrated: it is the only postulated explanation of the aggregate data that also satisfies orthodox methodological desiderata. This is no mean intellectual achievement – so far no one else has emulated it. But note that the preference for Benhabib and Farmer's model is not motivated by its empirical superiority, because both of the candidate explanations accommodate the macroeconomic data equally well. Indeed, Benhabib and Farmer's model is arguably worse off with respect to accommodating other empirical data, given that more of its assumptions are highly idealized and thus false, such as that the economy comprises many identical representative households all of which rationally maximize the same utility function, and so on. Rather, Benhabib and Farmer's model is preferred because of a non-empirical constraint on admissible hypotheses.

If we accept this non-empirical admissibility constraint, there is (currently) no alternative to the Benhabib and Farmer model. Therefore, their model is favored merely by accommodating the aggregate money and nominal income data. Formally, for E = the money-nominal income data and H = the Benhabib-Farmer model, while there may be many inadmissible H^* (such as Friedman and Schwartz's model) for which $p(E|H) = p(E|H^*)$, for all admissible H^* we have $p(E|H) > p(E|H^*)$.

[10] Friedman and Schwartz do appeal to supplementary evidence in addition to the basic money-income aggregates, in effect utilizing the 'additional evidence' defense of accommodation discussed below (section 6).

I do not adjudicate here the controversial – but separate – issue of whether this admissibility constraint is itself defensible.[11] Rightly or wrongly, endorsement of the constraint is widespread in the profession, even among those sensitive to methodological reflection (Rodrik 2015). I also do not address here exactly how often there are no plausible alternatives even once given the admissibility constraint. No doubt this will vary case by case. The point here is rather the entanglement of prediction with methodological orthodoxy: the more that a commitment to orthodoxy whittles down the number of plausible alternatives, the more that accommodation will be sufficient for confirmation.

Conclusion: accommodation is vindicated only if we accept that we should be constrained by modelling orthodoxy (or at least this is the tendency).[12]

5. Second defense of accommodation: Calibration

Suppose that a hypothesis is already accepted by all and the remaining task is only to *calibrate* it. Then accommodation can again be epistemically sufficient. For example, suppose we wish to know the mass of a newly discovered moon of Jupiter. We can calculate this by observing closely the motion of nearby objects. The underlying theory of gravity is not in question; rather, it is just a matter of using it to identify by observation the correct value for the new moon's mass. That is, a theory may leave a parameter value still unidentified and past evidence then be used to identify it. Accommodation is perfectly sufficient for such calibration.

Formally, the admissible hypothesis space is now different calibrations of the same underlying theory. H and H* are particular such calibrations, and E is the evidence used to calibrate. The condition $p(E|H) > p(E|H^*)$ can then adjudicate the competition between different calibrations. When the inequality holds, it means that H is the particular calibration supported by E – and this inequality often will hold perfectly well even if E is past evidence,

[11] My own view is that it is not (Northcott forthcoming). I also do not discuss here whether it is defensible to focus purely on accommodation of the headline macroeconomic correlation rather than on the other empirical data contradicted by the orthodox model's assumptions.

[12] In principle, some commitment other than to orthodoxy could also severely constrain the range of admissible hypotheses, thereby creating its own opening for accommodation. But in practice, orthodoxy is the only current candidate for this role.

i.e. if it is a case merely of accommodation. But, as is widely noted in the literature, such calibration does not provide any support for the underlying theory relative to other underlying theories (Worrall 2014, 56-57; Howson and Urbach 1993, 410-411).

Some empirical methods in economics are exercises in just such calibration. Benhabib and Farmer's paper is in part an example (see below). Other examples include the measurement of monetary velocity by assuming Fisher's equation of exchange, and much work in the real business cycles and econophysics literatures. However, in all of these cases it is controversial, to say the least, whether the underlying theory being calibrated should indeed be accepted.

Ultimately, the calibration defense of accommodation is a subset of the no-plausible-alternatives defense. Before, accommodation could (sometimes) be endorsed if we accepted an orthodoxy constraint on models' admissibility. Now, in the calibration case, the constraint is in effect even stronger – the only admissible hypotheses are different calibrations of the same underlying model.

6. Third defense of accommodation: Additional evidence

There may be many different explanations for why, say, wages in local restaurants are lower than six months ago. Suppose that they all successfully accommodate the past data of the wage decrease. However, there may also be *additional* historical evidence available that favors one of these explanations over the others. If so, there is no need to rely on prediction to break the tie. Perhaps, for instance, it is discovered that an especially generous owner of a chain of local restaurants recently retired and that their successor offered only much lower wages to new staff, and further that this new evidence supports one explanation over the others. Formally, there would then be additional evidence E* that breaks the tie between H and H*, i.e. for which $p(E^*|H) > p(E^*|H^*)$.[13]

Usually, gathering additional evidence is indeed possible. It is what historians do all the time – not coincidentally, it is thus a central way in which accommodation may be endorsed,

[13] This can be seen as an instance of the general point that empirical equivalence does not imply epistemic parity between hypotheses once we consider asymmetric empirical support for each hypothesis's background assumptions (Laudan and Leplin 1991).

although I am not aware of any systematic attempt to connect this to the role of prediction in economics. The significant issue for us is how this interacts with the role of theory. Such additional evidence is typically *idiographic*. That is, it tends to highlight sui generis local causes. The explanations that result thus tend not to be those offered by theory, because theory by its very nature focuses instead on systematic causes that recur across contexts.

Turn now to economics. Orthodox economic models outline the actions that should be expected from economic agents in particular situations. They are usually given a causal interpretation and these causes, at least to some extent, are taken to apply across contexts. Accordingly, orthodox models typically do not capture every idiosyncratic local cause, and do not aim to. Indeed, as per Mill's original methodological program, such models' utility is in part precisely that they *don't* capture every local cause – it is this that enables them to be applied to many contexts, adding in different disturbing causes each time.

This mismatch between orthodox models and sui generis local explanations is not controversial. The point here is how it bears on the need for prediction. We saw previously that a way accommodation can be endorsed is by *excluding* unorthodox alternatives. But now accommodation can be endorsed only by *including* them, i.e. by appealing to explanations that incorporate sui generis – and thus usually unorthodox – local causes. So, the pattern is reversed. The same commitment to orthodoxy that favored the first two ways of supporting accommodation will now tend to rule out this third way of supporting it.

7. Fourth defense of accommodation: Experimental control

Suppose that our evidence E is obtained from a controlled experiment. Such experiments are situations deliberately engineered to decide between competing hypotheses H and H* by shielding the phenomenon at hand from confounding background factors. Similar remarks apply to trials: samples are divided randomly between treatment and control groups, again with the aim of ensuring that there are no systematic confounds, thus ensuring that no salient alternative to H can explain any difference between the results in the two groups. Accommodation in such circumstances is quite sufficient. Formally, for evidence E collected in a successful controlled experiment, regardless of whether it is prediction or accommodation, $p(E|H) > p(E|H^*)$. But the relevance of experiments to prediction in economics does not seem to have been much discussed.

In economics, as in many other social sciences, experiments are difficult to conduct. Compared to laboratory sciences, often there are insuperable practical and ethical barriers. How might we test *experimentally*, for example, Friedman and Schwartz's hypothesis that money causes nominal income? We can hardly commandeer two real economies, equalize all relevant background factors, shield these economies from any disturbing causes, and then run different monetary policies on each.

Nevertheless, economists have been able to utilize a range of methods to replicate the inferential conditions of experiments as often as possible. One such method is *natural experiments*, i.e. when processes outside the investigator's control happen to divide a sample into treatment and control groups in the same way as a random assignment would have. A second method is the *quasi-experiment*, i.e. (typically) when an investigator controls the assignment of subjects to a treatment group but using some criterion other than random assignment. A third method is the *field trial*, i.e. when a controlled experiment is carried out 'in the field', for instance when different villages are randomly assigned to either treatment or control groups for some social policy intervention (Banerjee and Duflo 2011). A fourth method, finally, is the *laboratory experiment* itself (Kagel and Roth 2016).

These different methods have various strengths and weaknesses. One shared difficulty remains range of applicability: practical and ethical limitations mean there is only a limited range of economic questions that experiments can usefully elucidate. A second shared difficulty is external validity: will results from one context extrapolate to that of another? In the case of laboratory experiments, is agent behavior in the laboratory a reliable guide to agent behavior in often very different and more complex circumstances in the field (Levitt and List 2007)?

In contrast to the previous three defenses of accommodation, the experimental defense holds regardless of whether we are committed to orthodoxy. Still, although there may thus be no connection in terms of epistemic logic, in practice there has been a positive correlation between heterodoxy and use of experimental approaches (section 9).

8. Accommodation and entanglement

The issue of prediction versus accommodation in economics is thus deeply entangled with methodological orthodoxy. If we insist on orthodoxy, it becomes more likely that two of the first three defenses of accommodation apply: the no-plausible-alternatives justification and the calibration justification. On the other hand, the contextual, historian's additional-evidence justification typically will not apply. If we do *not* insist on orthodoxy, on the other hand, the situation is reversed: now neither the no-plausible-alternatives nor the calibration justification is likely to apply, but the contextual, historian's justification may well do. What are some of the implications of this?

Several methods of accommodation are common in economics. Among them are regression analyses that retrospectively calibrate a model or that test a model's degree of fit with the data. The problem in the first case is that the particular model must already be assumed true but that this assumption will rarely be justified. The problem in the second case is that it must be assumed that no other model can achieve a similar degree of fit with the data, and this assumption too may be implausible without an orthodoxy constraint (or even with one) – and especially so because typically the fit with data in such cases is imperfect and so it is all the more difficult to rule out that alternatives may fit equally well. The conclusion is that such regressions are more likely to lose epistemic force without an additional commitment to orthodox models only.

Similar remarks apply to more sophisticated accommodation methods too. To illustrate, return to the example of Benhabib and Farmer's (2000) model of how money causes nominal income. Their empirical work proceeds in several steps. The first step is calibration – using past data to estimate the values of various model parameters. The second step is simulation – shocks, themselves calibrated from actual data, are fed into the calibrated model to generate simulated data. The third step is then an assessment of the degree of fit between the simulated data and the corresponding actual data. This assessment is itself a complex process, involving several regressions, but the eventual verdict is somewhat informal, the authors concluding (544): "The main finding from the comparison of these two sets of figures is the broad similarity in the qualitative and quantitative nature of the responses of the model economy with that of the data." Overall, the empirical method used is a blend of various forms of accommodation.

In their discussion, the criticisms that Benhabib and Farmer consider necessary to address are either internal technical ones, such as how to derive determinate empirical implications from a model with multiple equilibria, or else are empirical ones but only in a rather stylized and informal sense, such as the worry that the degree of increasing returns to scale that needs to be assumed is implausible. Their solution to the latter issue is to avoid the awkward empirical implication in the first place by supplementing the model with a representation of how agents form beliefs – without any detailed empirical justification for this new theoretical supplement.

Benhabib and Farmer thus implicitly assume that their model needs defense only from criticisms within orthodoxy. No defense is thought necessary against heterodox alternatives such as Friedman and Schwartz's original explanation – and so no argument is offered that their model fits the data better than do heterodox alternatives. Indeed, tellingly, there is not a single reference to Friedman and Schwartz in the entire paper, even though Friedman and Schwartz were addressing exactly the same economic explanandum, namely why changes in money supply cause changes in income. In effect, Benhabib and Farmer assume that because there are currently no orthodox alternatives to their model, therefore there are no alternatives at all. This is an example of the orthodox admissibility constraint in action. Not surprisingly, this constraint is paired with an empirical concern exclusively with accommodation rather than prediction.

This example is arguably typical, in two ways: first, accommodations of one kind or another are the pre-eminent empirical method; and second, only orthodox alternatives are considered.[14]

9. Implications for the heterodox

What if we are not committed to orthodoxy? Begin by briefly surveying how often this is actually the case. Following the Cowles Commission after the war, there was a strong norm across the profession that econometrics should aim at testing particular theoretical models rather than at discovering more fragmented or a-theoretical causal relations (Malinvaud 1988). Although later empirical methods deviated from many of the details of the Cowles Commission's approach, this focus on theory was retained. Such work is typically committed

[14] In this particular example, it is notable that the accommodation of both the headline aggregate data and the data justifying particular assumptions, is rather informal and casual. In other orthodox examples, accommodation of the headline data (although not of particular assumptions) is often more precise.

to orthodox modelling and also is a species of accommodation rather than prediction. But more recently, empirical work has less often been theory-based. Biddle and Hamermesh (2016) report that whereas in the 1970s all microeconomic empirical papers in the profession's five most prestigious journals exhibited a theoretical framework, in the 2000s there was a resurgence of a-theoretical studies.[15] Moreover, citation numbers suggest that the a-theoretical work is at least as influential. Angrist and Pischke (2010) also report the rise of a-theoretical practice in several subfields. Such a-theoretical work typically tests for causal relations between variables using one of the variety of experimental techniques described earlier, and is not committed to orthodoxy. Accordingly, overall the commitment to orthodoxy in empirical work in economics seems to be decreasing, albeit from a high base.[16]

This paper's analysis suggests that rejection of a commitment to orthodoxy tends to undermine the epistemic credentials of accommodation, leaving us with only three empirical methods, all currently minority ones: experiments, prediction and idiographic historical analysis. Of these three, experiments are often not available. So, if we give up on prediction too, then (often) economics can only be an idiographic, historical science. This would make it difficult for economics to advise policymakers or other decision-makers about interventions, except when experiments are available.[17]

There is a caveat: perhaps it is possible sometimes to justify policy-relevant predictions without any prior predictive success but instead purely by means of local knowledge, and in particular by detailed knowledge of current conditions and local causes. For example, perhaps local knowledge may justify the prediction that a particular free trade treaty will increase two countries' GDPs – because we can know that the classic factor model for the benefits of free trade will indeed operate and in addition know that this operation will not be

[15] It is true that there has always been empirical work in economics not committed to orthodoxy, in fields such as agricultural and labor economics and in activities such as national accounting and cost-benefit analyses. Nevertheless, empirical work not committed to orthodoxy is now at a minimum more prestigious (Cherrier 2016).

[16] In tandem, the reliance on experimental methods is increasing. I do not have figures for the proportions of empirical work that are calibration of an agreed model, idiographic historical analysis, or prediction.

[17] Arguably, giving up on prediction would also mean giving up on the important benefit of forcing economic modelling to become more empirically responsive. But that is a separate matter, touching again on the issue of whether we should indeed commit to orthodoxy.

outweighed or disturbed by other factors. In principle, this is a way to warrant predictions, and thus to offer policy usefulness, without actual predictive success. How often will it work in practice? Ultimately, that is an empirical question, but prima facie the record does not seem hopeful – systematic predictive success at the policy level via such methods, from either historians or economists, is hard to find, perhaps precisely because of the ubiquity of unpredicted disturbing causes. Accordingly, unless there is a record of actual predictive success, I think we should be suspicious of claims that some predictions are justified. Admittedly, the prospects might be brighter at a humbler level than national policy because the likely number of disturbing causes might be much lower. Perhaps, for example, local knowledge might indeed justify a prediction that decreasing a concert's ticket price will increase attendance. But even here our confidence would be much greater if there was a local history of successful predictions about attendance.

10. Conclusion

Predictive success is hard to achieve in economics and generally it has not been seen as a priority. But if there are many possible explanations of a given event, and if there is no supplementary investigation of local details, then predictive success is necessary for confirming models and, thus, for policy usefulness. An exception is if decisive experimental evidence is available – but usually it is not. Accordingly, prediction should be required uncomfortably often.

References

Alexandrova, A., and R. Northcott (2009). 'Progress in economics', in D. Ross and H. Kincaid (eds) *Oxford Handbook of Philosophy of Economics*, 306-337. Oxford.

Angrist, J., and S. Pischke (2010). 'The credibility revolution in empirical economics: How better research design is taking the con out of econometrics', *Journal of Economic Perspectives* 24, 3-30.

Banerjee, A., and E. Duflo (2011). *Poor Economics*. Penguin.

Betz, G. (2006). *Prediction or Prophecy?* (Wiesbaden: Deutscher Universitaets Verlag)

Benhabib, J., and R. Farmer (2000). 'The monetary transmission mechanism', *Review of Economic Dynamics* 3, 523–550.

Biddle, J., and D. Hamermesh (2016). 'Theory and measurement: emergence, consolidation and erosion of a consensus', NBER Working Paper No. 22253.

Cherrier, B. (2016). 'Is there really an empirical turn in economics?', Institute for New Economic Thinking blog 29th September 2016 **https://www.ineteconomics.org/perspectives/blog/is-there-really-an-empirical-turn-in-economics**

Ericsson, N. (2017). 'Economic forecasting in theory and practice: An interview with David F. Hendry', *International Journal of Forecasting* 33, 523-542.

Friedman, M., and A. Schwartz (1963). 'Money and business cycles', *Review of Economics and Statistics* 45, 32–64.

Giddens, A. (1976). *New Rules of Sociological Method: A Positive Critique of Interpretative Sociologies*. London: Hutchinson.

Hacking, I. (1995). 'The Looping Effect of Human Kinds', in *Causal Cognition an Interdisciplinary Approach*. Oxford: Oxford University Press.

Hausman, D. (2012). 'What Economics Can (and Can't) Do', *The New York Times*, 14th July 2012.

Howson, C., and P. Urbach (1993). *Scientific Reasoning: The Bayesian Approach* (2nd edn). Open Court.

Kagel, J., and A. Roth (eds) (2016). *The Handbook of Experimental Economics, Volume 2*. Princeton.

Laudan, L., and J. Leplin (1991). 'Empirical Equivalence and Underdetermination', *Journal of Philosophy* 88, 449-472.

Lawson, T. (1997). *Economics and Reality*. Routledge.

Levitt, S., and J. List (2007). 'What do laboratory experiments measuring social preferences reveal about the real world?', *Journal of Economic Perspectives* 21, 153-74.

Machamer, P., L. Darden, and C. Craver (2000). 'Thinking About Mechanisms', *Philosophy of Science* 67, 1-25.

MacIntyre, A. (2007). 'The Character of Generalizations in Social Science and their Lack of Predictive Power', chapter 8 in his *After Virtue* (3rd edn). Notre Dame.

Malinvaud, E. (1988). 'Econometric methodology at the Cowles Commission: rise and maturity', *Econometric Theory* 4, 187-209.

Mill, J. S. (1843). *A System of Logic*. London: Parker.

Northcott, R., and A. Alexandrova (2013). 'It's just a feeling: why economic models do not explain', *Journal of Economic Methodology* 20, 262-267.

Northcott, R. (2017). 'When are purely predictive models best?', *Disputatio* 9.47, 631-656.

Northcott, R. (forthcoming). 'Economic theory and empirical science', in C. Heilmann and J. Reiss (eds), *Routledge Handbook of Philosophy of Economics*. Routledge.

Reiss, J. (2008). *Error in Economics: Towards a More Evidence-Based Methodology*. Routledge.

Rodrik, D. (2015). *Economics Rules: the rights and wrongs of the dismal science*. Norton.

Rosenberg, A. (1993). *Economics: Mathematical Politics or Science of Diminishing Returns?* Princeton.

Runciman, W.G. (1963). *Social Science and Political Theory*. Cambridge.

Simon, H. (1989). 'The state of economic science', in W. Sichel (ed) *The State of Economic Science. Views of Six Nobel Laureates*, Kalamazoo, Michigan, W. E. Upjohn Institute for Employment Research.

Sober, E. (1999). 'Testability', *Proceedings and Addresses of the American Philosophical Association* 73, 47-76.

Taylor, C. (1971). 'Interpretation and the Sciences of Man', *Review of Metaphysics* 25, 3-51.

Tetlock, P. and D. Gardner (2015). *Superforecasting*. Random House.

Worrall, J. (2014). 'Prediction and accommodation revisited', *Studies in History and Philosophy of Science* 45, 54-61.

Acknowledgements
For useful comments I thank two anonymous referees and Julian Reiss, as well as audiences at the International Network for Economic Method and at the British Society for the Philosophy of Science.