Model-Based Knowledge and Credible Policy Analysis

Hsiang-Ke Chao and David Teira

May 2015

"Credible economic policy depends on going with the evidence."

Gordon Brown (then the British Chancellor of the Exchequer)[1]

1. Introduction

The analysis of public policies and the subsequent decisions about them are usually based on models. On the one hand, there is an analytical model about the reasons and consequences of adopting a policy. On the other hand, the actual implementation of such a policy in a given country becomes a model for other countries under a similar situation. Many monetary authorities, for example, regarded the quantitative easing monetary policy promoted by the Federal Reserve as having effectively helped the U.S. economy following the global financial crisis. As a result, others such as the Bank of Japan and the European Central Bank followed the footsteps of Ben Bernanke to implement their own large-scale stimulation plans. Policy models operate on two different levels: first, the policy makers relied on an economic model to articulate and implement a policy; the implemented policy then became in itself a model (a paradigm) for policy makers in other places, who assumed that the adoption of such a policy would achieve a similar, if not the same, outcome.

[1] House of Commons: Treasury Committee (2006), p. EV53.

In the literature of methodology and philosophy of science, standard benchmarks for the assessment of policies are conceptual dichotomies such as *effectiveness* (the effect of a policy under real-world conditions) versus *efficacy* (the effect of a policy under ideal circumstances) and *internal validity* (impact of a policy in the study population) versus *external validity* (generalizability of the findings in a study population to other target populations). In this chapter, we elaborate on the concept of *credibility* as a yardstick for the assessment of model-based policies. The concept of *credibility* has been studied in the context of model specification in empirical economics, and philosophers of science have even more recently been investigating whether scientific models provide us with "credible worlds" (references). Section 2 provides an overview of this debate. Our goal is to further distinguish between *epistemic* and *strategic* credibility and show how this distinction helps us to understand the relation between models and the assessment of public policies.

The underlying intuition behind the concept of epistemic credibility is that a public policy model will be accepted as being credible if it succeeds at helping us to infer the relevant causal relations for policy intervention from the available evidence. Econometricians disagree about how to make this inference. As we shall see in Section 3, empiricist econometricians only consider credible evidence-based models: the less a priori knowledge they require, the more credible a model will be. This is the so-called *reduced-form* approach in which the relations between variables are represented as the response of the variables of concern (endogenous variables) to the target variables (exogenous variables). Structural econometricians instead defend theories as a source of credibility for models, inasmuch as it allows them to better grasp causal mechanisms. We will show in Section 3 that the in reduced-form approach, commonly adopted by field trialists, credibility goes hand in hand with internal validity. By contrast, in the structural approach, credibility depends mostly on extrapolation: from sample to population (external validity) and from population to population. A public policy model is credible if the same causal mechanisms that ground its internal validity allow researchers and policy makers, under certain assumptions, to ground their generalizability and extrapolability. Here, we draw from recent contributions by Nancy Cartwright (Cartwright 2007a, 2007b; Cartwright and Hardie 2008) and Daniel Steel (2008) to show that, in this regard, the structural approach yields more solid extrapolations than the reduced-form approach.

While economists have extensively discussed the epistemic credibility of public policy models, they have also elaborated a *strategic* approach to their credibility under the rubric of the Lucas Critique. According to the latter, a credible public policy model should take into account the rational behavior of economic agents and their optimal strategic response to the government's intervention, as theorized in the model. Agents can react to public policy interventions in a *performative* or *counter-performative* manner (MacKenzie 2006). Section 4 presents that they can make the world more like its depiction in the model or the opposite. Section 5 further explicates the relationship between strategic credibility and epistemic credibility. In principle, both should be taken into account if we want to deem a public policy model credible. In the best possible scenario, we can reduce the former to the latter when the agents' reactions are incorporated into the set of causal variables considered in the model. However, even when we can unify our assessment of the model's credibility, we still see that there are philosophical reasons for suspecting that such credibility is bound to be transient.

2. Models and Economic Analysis of Credible Public Policy

There are three meanings of a credible public policy in economics. The first one is closely associated with new classical macroeconomics and the rational expectations school: a policy is credible if it is *believable*. Fellner (1976, 1979) first coined the term to present the idea that the U.S. aggregate demand policy in the second half of the 1970s was unsustainable and thus *unbelievable* to the public (McCallum 1984). The term later evolved to mean *believed* in the sense that a policy is credible when it is believed by the public that the policy is actually conducted in the way if it was announced.[2] From the viewpoint of the new classical, rational expectations macroeconomics, the question about credibility is concerned with how the expectations of an announced policy will be carried out subject to the states of policymakers' intentions and the states of the economy. To some extent, credibility is considered as a "mantra of policy" (Rudebusch 1996) for policy makers such as central bankers, because credible monetary policies are better implemented and more effective.

The second meaning of *credible public policy* appears in the empirical microeconomics literature in which a better policy decision can be made if it is based on the evidence produced by an experimental design, be it randomized field trials or natural experiments. This approach is empiricist in spirit, arguing that randomization and better experimental designs - instead of abstract economic concepts or analytical structure - make evidence credible, and hence also the policy, based on this evidence. Among academics, the experimental approach has become the "new orthodoxy" since the 1980s (Manski and Garfinkel 1992, p. 12) and is currently accepted by state agencies, such as

[2] See also, inter alia, Stokey (1991), and Drazen and Masson (1994).

the Behavioral Insights Teams of the UK government, as a useful tool for analyzing a policy's effectiveness. Empirical microeconomics has experienced a "credibility revolution", leading to a "consequent increase in policy relevance and scientific impact" (Angrist and Pischke 2010, p. 4).

The third meaning of *credible public policy* hinges on the policy's causal role. A policy is credible if it encompasses a true cause of its proposed goal: small class size *improves* students' achievement in reading (Cartwright and Hardie 2012); building a new railroad station *reduces* a region's poverty (Deaton, 2010); deworming *increases* the attendance rate of pupils (Miguel and Kremer, 2004). In this case, policy analysis runs parallel to causal analysis, in which the central issues are identifying counterfactual states and understanding causes as an effective means of manipulation and control.

These three definitions of a credible public policy are interrelated, since they are linked with different methodological conceptions of econometric models and causality. For new classical macroeconomists and design-based empirical microeconomists, the structural modeling advocated by the Cowles Commission in the 1950s is unacceptable, because they consider that a priori theories do not provide a credible basis for model identification. This was originally pointed out by Liu (1963) and Sims (1980) for defending a data-based approach to econometrics.[3] Going a step further, design-based empirical microeconomists claim today that we should rely more on experimental design and adopt randomized controlled trials (RCTs) as the gold standard for causal analysis, as it happens in other fields.

David Hendry (2005, p. 67), for example, develops this view about credibility into a full-fledged empiricist approach. For him, credible econometric *models* represent true

[3] See Chao and Huang (2011) for the history of Liu's and Sims' econometric approaches.

data-generating processes. Hendry points out two key dimensions of the credibility of the evidence (generated by the model): "persuasiveness" and "verisimilitude". Persuasiveness is related to "whether or not scholars will deem the evidence credible relative to their belief system"; verisimilitude refers to "whether or not they should do so" (Hendry 2005, p. 67). Hendry does not continue to explicate these two notions, but he is clearly concerned more with verisimilitude, which is somewhat similar to the concept of truthlikeness developed by Popper in his *Conjectures and Refutations* (Popper 1963). Hendry is concerned with empirical truth. If a model is derived from *a priori* theory, then its credibility depends on "the prior credibility of the theoretical model". However, the quick progress of economic theory discards many of these models, making the evidence they yield not credible anymore (Hendry 2005, p. 68).[4] In this vein, he sides with empirical econometricians such as Sims and Angrist, who deny the truth value of theory-based information.[5]

A priori theoretical information, in contrast, is considered essential by the econometric methodology of the Cowles Commission, as represented by Lawrence Klein's large-scale macroeconometric models and more recently by Charles F. Manski. Without this a priori information, model identification becomes unfeasible, especially when data alone cannot determine the causal order. Manski (2007, p. 1) refers to this as the "reflection problem": the observation of the almost simultaneous movements of a person and their reflections in the mirror does not tell you which induces which. One needs to understand optics and human behavior to reach the correct conclusion. Thus, the

[4] In this sense, Hendry subscribes to??? the position of pessimistic induction in the philosophy of science. See Chao (2009, 2014) for Hendry's empirical methodology of econometrics.
[5] Note that Sims (2010) forcibly denies both Angrist and Pischke's (2010) conception about macroeconomics and the similarity between their research design methodology and his vector-autoregressive modeling approach.

dispute centers on whether a priori information plays any epistemological role in building analytical tools for policy analysis.

By way of illustration, consider the following account for the methodology of policy evaluation. In their introductory chapter to an edited volume on the evaluation of social programs, Manski and Garfinkel (1992) identify two approaches within the field with two terms borrowed from econometrics: *structural* and *reduced-form* models. Structural models consist of a system of equations representing various types of functions and relations among economic factors. The structural model approach is intended to explain endogenous variables: how they are determined or *caused* by exogenous variables whose value is predetermined outside the model. By solving the value of endogenous variables, the structural equation can be expressed as the *reduced-form* model, in which each endogenous variable is represented as an equation of predetermined and exogenous variables. The ordinary least square method can then be applied to estimate the coefficients in a reduced-form equation: for each endogenous variable we can establish which exogenous variables would affect its value. Although econometrically speaking, the reduced-form and the structural models are two sides of the same coin under certain conditions, nowadays it seems customary to call a regression equation a *reduced-form model*, leaving aside the structural model from which the reduced-form model derives. Accordingly, Manski and Garfinkel denote *structural evaluation* as an approach to use models in which the social process, i.e., structure, is involved, whereas *reduced-form evaluation* only compares the outcome of a program.

The Manski-Garfinkel volume addresses both experimental and non-experimental methods as tools for evaluating welfare and training programs, but favors the latter, and in particular the structural econometric approach. This is because experimental methods, such as RCTs and natural experiments, are reduced-form evaluations and are black boxes about causal processes - that is, they model the "effects of causes without modeling the causes of effects" (Heckman 2005). Thus, insofar as public policy is concerned, Manski and Garfinkel point out that there is a "problem of extrapolation".[6] For reduced-form evaluations, the problem of extrapolation is whether the outcome of a social experiment or a pilot program would still hold true if the actual program were implemented. If so, then one can extrapolate from the experiment to the real world. However, as it has been often argued, there is a significant gap between ideal randomized controlled trials and real experiments, due to various "threats" such as selection bias. Manski and Garfinkel also note that reduced-form evaluation in fact needs specific structural assumptions on individual and organizational behaviors (p. 17).

Structural evaluations also suffer from the problme of extrapolations, since the understanding of social processes is usually conditional on strong assumptions. In structural evaluations one can always wonder whether the observed program can be moved to another location and implemented into another social process. If the structure or social processes hold true or approximately true elsewhere, then extrapolating from the observational data is possible. Moreover, there is disagreement on the credibility of prior information, but supporters of this approach are confident in the possibility of improving the credibility of their evaluations with further research on the structures analyzed. For Manski, policy analysis has to start with theory/prior information and then you let it set up some criteria to tell you what to look for.

[6] Manski and Garfinkel's notion of extrapolation originates from its usage in the econometric literature, which understands extrapolation to be an ex-post test on how the model fits the observed data.

As we have seen, Manski and Garfinkel's distinction between structural and reduced-form evaluation and their credibility in policy analysis go hand in hand with the econometric debate on the credibility of models presented above. In both cases, the issue at stake is the number of theoretical assumptions that a model requires in order to be credible.

3. Epistemic Credibility

The above illustration highlights what we call the *epistemic credibility* of a public policy model: a model will be epistemically credible if it succeeds at grasping the relevant causal levers for a policy intervention and achieves its intended goals in a given experimental set-up. In this sense, the *internal validity* of the model matters.[7] The question of *external validity* arises only when a policy based on the evidence gathered from a restricted area is extended to a larger area. This is the type of extrapolation (from a sample to a population) that usually occurs in a health or social policy, when a large-scale policy is implemented after the pilot program is evaluated. A policy could also be "borrowed" from its actual implementation on another population in a different time and place: here, external validity is about population-to-population extrapolations. Economic policy, especially macroeconomics, deals with this latter type of external validity, which is the case for public policy at the macro level.

The difference between these two approaches to external validity is the homogeneity of the population. Pilot programs and their subsequent large-scale implementation are generally conducted on the same population. Macroeconomic policies, instead, are externally valid when they work on a different country or region. In

[7] For this view, see Imbens (2010, p. 417).

this regard, policy makers are naturally concerned only with whether a policy works in the setting of their own population. From successful policies in other countries, policy makers wish to derive the general causal principles that they can apply to their own country. For instance, by adopting the Fed's QE policy, Japan's central bankers would only be interested in evaluating whether the policy can help cure the Japanese economy. Does this generalization-instantiation process produce a credibly policy?

Consider the following example (Deaton 2010). Suppose that a government receives the recommendation for building railway stations (*R*) in order to reduce poverty (*P*). The model for this recommendation can be written as a linear, reduced-form equation $P=\gamma+\theta R+\nu$, where γ is a constant and ν is a classical error term. The evaluation of the effect of such a policy depends on the magnitude of θ. The reduced-form approach evaluates such a program relying solely on this equation, by estimating θ. Field experimentalists would measure the treatment effect as some cities are regarded as "treated" with a railway station, while others are not. The problem of extrapolation is whether θ would be (approximately) the same if the policy is implemented somewhere else or at a different scale (e.g., nationwide). There is a heterogeneity problem if θ varies across cities, and "it is precisely the variation in θ that encapsulates the poverty reduction *mechanisms* that ought to be the main object of our enquiry" (Deaton 2010, p. 429; our emphasis). Without such a mechanism, this equation cannot even be regarded as a model (*ibid*.).

There hence might be a tradeoff between the credibility of a public policy model and its external validity. If a model is more credible when it requires fewer theoretical assumptions, then its credibility will not help in deciding about the possibility of using it elsewhere. In other words, perhaps credibility stems from the internal validity of the model under consideration. However, as we saw in the introduction, often it is the implementation of a model in a given setting that forces us to wonder about its extrapolation to a different scenario. Is there some sort of credibility here?

As we just noted, policy makers have been mostly concerned so far with this second kind of extrapolation: from successful policies implemented in another country, they wish to derive causal principles applicable to their own country. This is what the Japanese central bankers probably had in mind when they adopted the Fed's QE policy: they just wanted a cure for the Japanese economy. Is this a credible policy for Japan in any sense?

In his much appraised book, Steel (2008) deals with the problem of extrapolation in biology and social science. His account is a mechanism-based extrapolation: the knowledge of the mechanism is essential to understand how the result derived from an apparatus is sustained in another population. While discussing little about econometrics, Steel takes Manski and Garfinkel's volume as an example of his mechanism-based extrapolation in social science. In this sense, structural evaluation, the structural model, and the mechanism approach to extrapolation are all regarded as being equivalent, if not identical, to such an extent that Steel's account could shed some interesting light on econometrics, or vice versa.[8] At the outset, for the structural approach that rejects the idea of treating the economy as a black box, the mechanism is essential not only to the models' formulation and to understanding the economy, but also to the credibility of policy.

[8] Steel does discuss extrapolation in the context of applied economics. See his research (2013) on Donohue and Levitt's (2001, 2004) study on the causal relation between legalized abortion and crime rate.

Apparently inspired by the mechanistic approach of Machamer, Darden and Craver's (2000) seminal article, in which a mechanism can be represented abstractly as a schema, which can be traced by "bottoming out" from the lowest level of a mechanism, and Darden and Craver's (2002) paper in which they propose methods of schematic instantiation and forward chaining/backtracking, Steel proposes "comparative process tracing" that involves two steps: First, know the mechanism of the source or model by process tracing or other experimental means. Second, compare stages of the model and target mechanisms and look for what are most likely to "differ significantly" (p. 89) - that is, comparative process tracing is based upon generalizations like "Features A, B, and C of carcinogenic mechanisms in rodents usually resemble those in humans, while features X, Y, and Z often differ significantly". If there is greater similarity between two mechanisms at these key stages, then there is a stronger basis for extrapolation.

By way of illustration, we take Steel's example of the carcinogenic effect of aflatoxin B_1 (p. 91). In order to study the carcinogenic effects of aflatoxin B_1 in humans, the laboratory experiment is sensitive to what research model is selected. It is found that aflatoxin B_1 causes liver cancer in rats, but not in mice. Hence, a laboratory result of the carcinogenic effects of aflatoxin B_1 by using mice cannot be extrapolated to humans. It requires knowing the results from existing research studies on the functions and effects of DNA and the phases of the metabolisms of humans and rodents to conclude that the rat is a better model than the mouse. Steel argues that comparative process tracing can break the "extrapolator's circle", yet how much do we need to know about the inferential target, before we can actually extrapolate? In comparative process tracing, we do not need to have complete knowledge of both mechanisms - just those features where mechanisms differ significantly. In other words, knowledge of the mechanisms of the model and target is a prerequisite. However, according to comparative process tracing, we do not need to have complete knowledge of both mechanisms, but rather we only require the knowledge of those stages in which mechanisms differ significantly.

Steel is nonetheless pessimistic about social extrapolation. On the one hand, "it is unclear that comparative process tracing can facilitate extrapolation to new location of larger scales", since the original mechanism might be constrained by entirely local features (Steel 2008, p. 166). On the other hand, there are "structure-altering" interventions in which the policy itself changes the targeted mechanism. Both issues, yet, are well known in econometrics. The first one is Deaton's heterogeneity problem illustrated above: the responses to a particular policy or program are not the same across different places, though econometric analysis can estimate how much they differ.

Steel's second concern echoes the famous *Lucas critique*, to which we now return below: policy implementation changes the structure of relations between *ex ante* and *ex post* variables, undermining the model on which we ground our policy assessment. The structure it captures is not the same after the intervention is implemented. A moderate answer to the Lucas critique is to ignore it, since empirical evidence suggests that structure-altering policies are rare. A more radical answer would be to give up the structural approach and use reduced-form alternatives such as VAR models. Here, we would lose the possibility of tracing the comparative process and sorting out our extrapolation problem.[9]

[9] A more serious problem for Steel's account is whether comparative process tracing can really solve the extrapolator's circle. Comparative process tracing requires sufficient knowledge of the mechanism of the target, but how much information is sufficient for the researchers to decide what works for the model works for the target is perhaps non-consensusal. Furthermore, the model's external validity depends on its

Summing up our discussion above, comparative process tracing allows the extrapolation of structural models, provided we could quantify the contextual variation of the source mechanism, on the one hand; while there is evidence about the post-implementation integrity of the structure, on the other. However, there might not be consensus about how much information we need about the target mechanism in order to break the extrapolator's circle. Perhaps rather than extrapolation we should be performing what Steel calls a *simple induction*: "the causal generalization true of the base population also holds approximately in related populations, unless there is a specific reason to think otherwise" (Steel 2008, p. 80). Simple induction is thus a reasoning process based solely on similarity or analogy.[10] Of course, we find here the dilemma of the reduced-form approach: without a precise target mechanism, shared common features do not guarantee that what happens in one population would actually happen in a different one.

An alternative approach to judge extrapolation is Nancy Cartwright's. Interestingly, Steel (2008, p. 83) regards Cartwright's account as a simple induction. He especially finds a passage in Cartwright (1995, p. 180) that is similar to the definition of simple induction: "I have claimed that in the central uses of the concept [of tendency or capacity], we assume that within the specified domain tendencies when properly triggered always 'contribute' their characteristic behaviours unless there is a reason why not" (Cartwright 1995, p. 180). Is Cartwright's methodology a simple induction? Despite the seemingly similarity, Cartwright commits to simple induction neither in the book

resemblance to the target. As the example of the carcinogenic effect of aflatoxin B_1 shows, to decide rats are better models than mice, researchers need to study humans, but once there are such studies on humans, the rodent study seems unnecessary, except for the ethical reason that we cannot conduct experiments on humans (Howick, Glasziou, and Aronson, 2013). The task is indeed to find satisfactory models rather than to find out the nature of the target. In that case, given that extrapolation is hindered by various threats and structure-altering changes, social scientists may just go ahead to uncover the mechanism of the target to ensure that the policy works.

[10] See Hesse (1966) for analogy and Giere (1988) for similarity accounts of models.

(1989) on capacities nor in the more recent work on the theory of evidence. In fact, she recently argues against the idea that similarity is useful for policy extrapolation, because the notion of similarity is vague, demanding, wrong, and wasteful if we rely on it to claim a policy is externally valid (Carwright and Hardie, 2012, I.B.6.3). Cartwright also explicitly points out the importance of the notion of a mechanism for policy effectiveness (Cartwright and Stegenga 2011; Cartwright and Hardie, 2012).

Cartwright's attempts at emphasizing the notion of capacity are explicated in her 1989 book. A capacity - as in Millian tendency claims - is what a properly triggered factor would produce *if unimpeded*. Capacity claims hold only under circumstances in which disturbing factors do not perturb the effects. There are three key elements associated with capacities (Cartwright 1998, p. 45): *Potentiality* - what a factor can do in an ideal situation; *causality* - the results a factor can produce; and *stability* - the causal power of a factor must persist across some variations of circumstances. The capacities of a factor are not always realized though, depending on some circumstances - aspirin does not always cure one's headache, for example. If we want to know whether a policy works here, it is better to find out whether, on the one hand, there is a proper trigger and, on the other hand, whether the concurring factors facilitate or impede the capacity. Thus, capacity claims are local and case-based (Cartwright 1989, pp. 2-3).

The notion of capacities can help us to understand extrapolation (Reiss 2010).[11] The strategy Cartwright (1989) advocates in association with capacities is bootstrapping. As Glymour (1980) originally puts it, we use theory to deduce hypotheses from evidence in such a way that, with help from the theory, evidence is used to measure the values of the quantities in one of several specific assumptions. Cartwright suggests instead that

[11] However, the question for social science is whether capacities are easy to come by. See Reiss (2008).

given general background assumptions and situation-specific information, the hypothesis, rather than an instance of the hypothesis, can be deduced from evidence (Cartwright 1998, p. 147). In econometrics, Cartwright considers the structural approach as an example of the bootstrap method: we use data to measure structural parameters so that the structural model is regarded as a hypothesis of the macroeconomic theory that directs the selection of variables in the model.

Cartwright also helps us in the discussion of reduced-form models for policy assessment. For a randomized trial, Cartwright believes it to be a "narrow-clinching method". It is deductive and will clinch the conclusions if the premises are true, but these premises, the underlying assumptions, are usually not true - hence, the narrowness (Cartwright, 2007b, p. 12). In randomized trials, external validity is contingent on the assumption that both model and target are similar. Moreover, the sample size of a usual RCT is often not very large, and not all types of policy are suitable for a field trial: e.g., a macroeconomic policy, since no one can randomly assign a QE policy that pumps money into one country's financial system and a tightening of a monetary policy in another country. This is why randomized trials only allow us to perform narrow generalizations. They will be even narrower if they are to ground extrapolations between different populations.

When it comes to extrapolating between different populations, Steel and Cartwright show us that we are safer when we have a structural model than if we only have a reduced-form one. Our safety is, of course, conditional on a number of assumptions that allow us to break the extrapolator's circle. If there is the knowledge of a mechanism or capacity postulated, then we are allowed to ground our policy recommendations, explaining how they should work in each circumstance. With reduced-form models, there is no direct connection between the credibility stemming from their internal validity and the extrapolability of their conclusions. These only will hold under narrow additional assumptions.

4. Strategic Credibility

Let us now complicate our analysis a bit more. The previous sections primarily assessed the credibility of a public policy model by economists and policy makers according to their methodological views regarding the epistemic weight of theoretical and empirical knowledge. Now we should, however, take into account how credible a policy is once we take into account the self-interests of the economic agents who will be ultimately affected its consequences. A policy will be *strategically credible* to the extent that the model properly captures the agents' incentives to act for or against it, since their action is necessary for the accomplishment of the intervention.

A standard case in point is the Phillips curve, in at least one of its versions (Forder 2014).[12] Here, the curve shows a negative relation between inflation and unemployment in the short run.[13] Hence, policy-makers could regard this negative relation as a tradeoff between unemployment and inflation rates: they could maintain a lower unemployment rate so long as they are willing to accept the price of a higher inflation rate. However, assuming that the relation holds, it creates a strategic interaction when private economic agents negotiate their wages. One key factor in their decision to set a given wage level is

[12] See also Hoover (forthcoming) for the genesis of the version of the price-inflation Phillips curve.
[13] Friedman (1967) and Phelps (1968) argue that, in the long run, the inflation-unemployment tradeoff does not apply, because the agents are fully aware of aggregate prices and inflation such that price and wage decisions are consistent.

whether the government is committed to maintain price stability. According to Kydland and Prescott (1977), the government cannot make a credible commitment to keep prices stable: if wages are set on the basis of prices remaining stable, the policy-maker has an incentive to use the inflation-unemployment tradeoff relation indicated by the Phillips curve and create surprise inflation in order to lower unemployment. If private agents understand the structure of the economy and the goals of the policy-maker, then they should not negotiate their wages under the assumption that prices will remain stable, because if they do, inflation will rise thanks to government intervention.

The discussion of epistemic credibility assumes that the causal structures analyzed in the model are somehow invariant under the government's intervention, but the Phillips curve example suggests otherwise: theoretically, the agents may prevent the government from using inflation in order to reduce unemployment bargaining for higher wages, which discount a rise in inflation. A model for public policy intervention is *strategically credible* if it captures the agents' incentives to act for or against it, since the model targets regularities that may change if the agents react to the purported intervention.

It is a contentious matter whether this is something more than a theoretical possibility in the case of the Phillips curve (Forder 2014), but economic models sometimes seem to have *performative effects* that go beyond their actual epistemic content. Following Donald MacKenzie (2006), we speak of *performativity* when the practical use of an aspect of economics makes economic processes more like their depiction in economics. For instance, between 1976 and 1987, the Black-Scholes-Merton option-pricing model provided an excellent description of actual market prices - making it, in the words of Stephen Ross, "the most successful theory not only in finance, but in all of economics" (MacKenzie 2006, p. 177). However, in many of these markets the traders carried with them sheets displaying arrays of Black-Scholes prices for the stock under exchange in order to assess their opportunities for arbitrage. These sheets were sold, among others, by Fischer Black himself. No wonder that the fit was so good: the economic agents were adopting an economic model as a rule for action.

According to Guala (forthcoming), economic models may become *coordination devices* for economic agents if they decide to adjust their beliefs accordingly. The strategic credibility of an economic model may thus depend on it allowing economic agents to solve a coordination problem, aligning their decisions according to the model. The Black-Merton-Scholes model somehow provided incentives for traders to use it in order to set option prices, and this is what explains its actual performativity. The Phillips curve provided incentives for private economic agents to coordinate themselves in challenging some of its assumptions when they bargained for prices, making it *counter-performative* - economic reality became more and more unlike the theoretical assertation of the Phillips curve. The lack of invariance under government intervention makes public-policy models susceptible to performative and counter-performative effects: their goals will only be accomplished if economic agents cooperate according to the model. This is what makes their strategic credibility relevant.

5. Epistemic Versus Strategic Credibility

When the (causal) regularities captured by public policy models are invariant under government interventions, their credibility can be assessed on purely epistemic grounds. We now discuss here a second scenario: when there is no invariance and the intervention may trigger some performative or counter-performative effect, we need to take into account the model's strategic credibility as well. The question we want to address now is: to what extent can we "reduce" strategic to epistemic credibility? Field trialists have already faced this challenge. For instance, Chassang *et al*. (2012) discuss experiments in which the participants consider the intervention a priori ineffective and therefore do not bother with the implementation. If the effect size is small, then it will not allow us to estimate what the effect would have been like with fully committed participants. Their paper presents instead a principal-agent trial design, in which the incentives of the participants are explicitly taken into account in the model, reducing its strategic credibility to one more dimension of its external validity (its epistemic credibility in our terms). In the best possible scenario then, we could have a unified assessment of the model's credibility, and strategic interactions would pose no particular threat for the well-advised policy-maker. How far does this unified credibility reach? When and for how long can a model remain credible? The answers depend, in part, on our take on the stability of causal structures, including here their invariance under interventions.

Nancy Cartwright's appraisal of social capacities, for instance, illustrates an *agnostic take* on this problem: they may exist, but their existence should be illustrated on a case-by-case basis (Reiss 2008). It also depends on our understanding of performativity, or on the possibility of *containing* the performative effects of an intervention within the incentive structure articulated in a model. We will see an atheist take on the stability of social capacities (they cannot exist for long) discussing a recent paper by Alexander Rosenberg (2012).

Starting first with Cartwright, we notice that the distinction between epistemic credibility and strategic credibility is implicit in her approach. According to Reiss (2008), for policy interventions, causal capacities should be stable at two different levels: they should be stable under a certain range of causal interferences - this is captured by the external/internal validity assessment - and they should be autonomous under some range of interventions - the strategic dimension. For Cartwright, the distinction does not play a major role though. Interventions are just one particular type of causal interferences; in terms of her recent work on the assessment of evidence-based policies, the agents' reaction to a policy would be just one more *support* factor in the *causal cake* grounding each policy intervention.

Cartwright's approach thus captures the spirit of the Lucas critique: a good public policy model is one in which the strategic interactions are taken into account in the causal structure of the intervention, so that they become part of the assessment of its internal and external validity. However, for Catwright, the stability of a causal regularity on which an intervention is based cannot be taken for granted once and for all. For every public policy model, you need to verify, on the one hand, that the necessary support factors are present in every given context. In addition, you need to check that no competing causal cake is at work in that the same context may challenge the stability of the intervening capacity.

According to Cartwright and Hardie (2012), there are no "unambiguous rules for predicting the results of social policies". Hence, even if we succeed at reducing strategic crediblity to epistemic credibility, the model's credibility will have to be reassessed for every contextual use. Cartwright remains agnostic as to the credibility of a model: it must be checked on a case-by-case basis, but in principle there is no reason for a model not to remain credible inasmuch as the capacities are properly grasped. A more radical take on this problem is Alexander Rosenberg's (2012) Darwinian view. Here, strategic considerations are going to ultimately overflow the credibility of every public policy model.

Rosenberg defends that every social regularity is ultimately bound to be spatio-temporally restricted, precisely for the impossibility of containing once and for all the underlying strategic interactions. Putting it shortly, Rosenberg's argument is as follows. Let us assume a generalized Darwinian approach to social phenomena, in which all their significant features will be understood as *adaptations* endowed with evolutionary functions. Social phenomena emerge from the interaction of competing Darwinian agents (be they individuals, groups, institutions, etc.). If we adopt a game-theoretic perspective, then any regularity will be just a *local equilibrium* in these interactions. As such, argues Rosenberg, the competing parties will sooner or later face a prisoner's dilemma and be tempted to exploit the regularity to their own particular advantage. At this point, the regularity will collapse. Using Rosenberg's example, one established regularity in the social sciences is that no two democracies have ever gone to war. However, Rosenberg (2012, p. 11) argues, "nothing is forever. We can be confident that somewhere or some when, some democracy is going to find a way to exploit this regularity by attacking some completely unsuspecting fellow-democracy, lulled into a false sense of the permanence of peace among democracies".

If Rosenberg is correct, then every good causal model in the social sciences is open to an iterative Lucas critique: someone will try to exploit the regularity it captures, making it collapse. In our terms, no model can aspire to enduring strategic credibility:

whenever social scientists succeed at incorporating the invariance under interventions into their models, sooner or later the concerned agents will react, putting such invariance at risk.

Summing up the above discussion, the difference between Rosenberg and Cartwright is that the latter, skeptical as she is regarding the existence of stable social capacities, remains *agnostic* about the possibility of extending the unified credibility of a public policy model whenever the occasion arises: it remains to be seen in each particular case. Rosenberg is an atheist in this regard: even if a public policy model answers the Lucas critique and incorporates the incentives of the concerned Darwinian agents, then the latter will ultimately challenge the model's assumptions. No model can remain strategically credible forever. Of course, Cartwright and Rosenberg are just two pessimist accounts and on a more spirited approach, the credibility of a model could have a longer expected life. The point Carwright and Rosenberg illustrate is that perhaps this pessimism is justified: as a matter of fact, the social sciences have not given us much evidence of the existence of stable social capacities so far, and we have principled Darwinian reasons to suspect that, even if such evidence arose, it might not last long.

6. Concluding Remarks

We have examined herein the different ways of understanding the credibility of a public policy model among economists. Credibility depends, for some, on the internal or external validity of a model and how it grasps the causal efficacy of an intervention. This is, in our terms, *epistemic credibility*. Some other economists, following Lucas, considered a public policy model credible if it took into account the incentives of the concerned agents to act for or against the intervention. This is what we call *strategic credibility*.

We have made two claims. For policy assessment purposes, credibility depends mostly on extrapolation: from sample to population (external validity) and from population to population. Such extrapolations have a better chance of success if we draw on a causal mechanism, as structural models do. The sort of empiricism about causal interventions promoted, among others, by field trialists in economics has a lower chance of being epistemically credible.

We then have considered the possibility of reducing strategic credibility to epistemic credibility. A public policy model will answer the Lucas critique if it can incorporate the agents' reaction into the causal structure of the model, showing that it is invariant under interventions - that is, when the strategic credibility of the model can be assessed in terms of its internal validity and external validity. We have here presented two pessimist arguments showing that even when such a reduction is possible, the epistemic credibility of a model is at most spatio-temporally restricted (using Rosenberg's expression): the stability of the causal set-up should be re-checked in every new implementation of a policy; and, if we adopt a Darwinian outlook, then even when the stability is proven, it may collapse when a third party finds a way to exploit it for its own benefit.

We have thus seen how the economists' discourse about the credibility of public policy models hinges on a number of methodological assumptions (e.g., being an empiricist or not) that can be interpreted according to different philosophical approaches (e.g., our take on causality). Perhaps a different set of meta-theoretic choices would have yielded a different analysis of credibility, but this is probably a warning about the implicit disunity of the concept. As the reader may have suspected from the beginning, there are good reasons to disagree about what we consider a credible policy intervention.

Acknowledgements

Chao's research is sponsored by the Ministry of Science and Technology of Taiwan under grant NSC 102-2628-H-007 -003 -MY3. Teira's research has been funded by research grants FFI2011-28835 and FFI2014-57258-P

References

Angrist, Joshua D. and Jörn-Steffen 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.

Cartwright, Nancy. 1989. *Nature's capacities and their measurement*. New York: Oxford University Press.

Cartwright, Nancy. 1995. Reply to Eells, Humphreys and Morrison. *Philosophy and Phenomenological Research* 55: 177-187.

Cartwright, Nancy. 1998. Capacities. *The Handbook of Economic Methodology,* ed. John B. Davis, D. Wade Hands, and Uskali Mäki, 45-48. Aldershot: Edward Elgar.

Cartwright, Nancy. 2007a. *Hunting causes and using them*. Cambridge: Cambridge University Press.

Cartwright, Nancy. 2007b. Are RCTs the gold standard? *Biosocieties* 2: 11-20.

Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-based policy: a practical guide to doing it better.* Oxford: Oxford University Press.

Cartwright, Nancy and Stegenga, Jacob. 2011. A theory of evidence for evidence-based policy. *Evidence, Inference and Enquiry*, eds. Philip Dawid, William Twining, and Mimi Vasilaki. Oxford University Press: British Academy.

Chao, Hsiang-Ke. 2009. *Representation and structure in economics: the methodology of econometric models of the consumption function*. London and New York: Routledge.

Chao, Hsiang-Ke and Chao-Hsi Huang. 2011. Ta-Chung Liu's exploratory econometrics. *History of Political Economy* 43: 140-165.

Chassang, Sylvain, Gerard Padró I Miquel, and Erik Snowberg. 2012. Selective trials: a principal-agent approach to randomized controlled experiments. *American Economic Review* 102:1279-1309.

Darden, Lindley and Carl Craver. 2002. Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences* 33: 1-28.

Deaton, Angus. 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* 48: 424-455.

Donohue, John, and Steven Levitt. 2001. The impact of legalized abortion on crime. *Quarterly Journal of Economics* 116: 379-420.

Donohue, John, and Steven Levitt. 2004. Further evidence that legalized abortion lowered crime: a reply to joyce. *Journal of Human Resources* 39: 29-49.

Drazen, Allan and Paul R. Masson. 1994. Credibility of policies versus credibility of policymakers. *Quarterly Journal of Economics* 109: 735-754.

Fellner, William J. 1976. *Towards a reconstruction of macroeconomics: problems of theory and policy*. Washington, DC: American Enterprise Institute.

Fellner, William J. 1979. The credibility effect and rational expectations: implications of the gramlich study. *Brookings Papers on Economic Activity* 1: 167-78.

Forder, James. 2014. *Macroeconomics and the Phillips curve myth*. Oxford: Oxford University Press.

Friedman, Milton. 1968. The role of monetary policy. *American Economic Review* 58: 1-17.

Giere, Ronald N. 1988. *Explaining science: a cognitive approach*. Chicago: University of Chicago Press.

Glymour, Clark. 1980. *Theory and evidence*. Princeton: Princeton University Press.

Guala, Francesco. Forthcoming. Performativity rationalized. *In Enacting the Dismal Science: New Perspectives on the Performativity of Economics*, eds. Boldyrev, I. and Svetlova, E. London: Palgrave-MacMillan.

Heckman, James J. 2005. The scientific model of causality. *Sociological Methodology* 35:1-97.

Hendry, David F. 2005. Bridging the Gap: Linking Economics and Econometrics. *New Trends in Macroeconomics,* ed. Claude Diebolt and Catherine Kyrtsou, 53-77. Berlin: Springer.

Hesse, Mary B. 1966. *Models and analogies in science*. Notre Dame University Press.

Hoover, Kevin D. Forthcoming. The Genesis of Samuelson and Solow's Price-Inflation Phillips Curve. *History of Economics Review*.

House of Commons: Treasury Committee. 2006. *The 2006 Budget, Fourth Report of Session 2005–06* Volume II, London: The Stationery Office Limited.

Howick, Jeremy, Paul Glasziou, and Jeffrey K. Aronson. 2013. Problems with using mechanisms to solve the problem of extrapolation. *Theoretical Medicine and Bioethics* 34: 275–291.

Imbens, Guido W. 2010. Better late than nothing. *Journal of Economic Literature* 48: 399-423.

Liu, Ta-Chung. 1963. Structural estimation and forecasting: a critique of the cowles commission method. *Tsing Hua Journal of Chinese Studies* 4: 152–171.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67:1–25.

MacKenzie, Donald A. 2006. *An engine, not a camera: how financial models shape markets, inside technology*. Cambridge, Mass: MIT Press.

Manski, Charles F. 2007. *Identification for prediction and decision*. Cambridge, MA: Cambridge University Press.

Manski, Charles F. and Irwin Garfinkel. 1992. *Evaluating welfare and training programs*. Cambridge, MA: Harvard University Press.

Miguel, Edward and Michael Kremer. 2004. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72: 159-217.

McCallum, Bennett T. 1984. Credibility and monetary policy. *Prices Stability and Public Policy: A Symposium Sponsored by the Federal Reserve Bank of Kansas* City: 105-135.

Phelps, Edmund S. 1967. Phillips curves, expectations of inflation and optimal unemployment over time. *Economica* 34: 254-281.

Popper, Karl R. 1963. *Conjectures and refutations: the growth of scientific knowledge*. London: Routledge.

Reiss, Julian. 2008. Social capacities. *Nancy cartwright's philosophy of science*, ed. Stephen Hartmann and Luc Bovens, 265-288. London: Routledge.

Reiss, Julian. 2010. Review of across boundaries. *Economics and Philosophy* 26: 382-390.

Rosenberg, Alex. 2012. Why do spatiotemporally restricted regularities explain in the social sciences? *The British Journal for the Philosophy of Science* 63: 1-26.

Rudebusch, Glenn D. 1996. Is opportunistic monetary policy credible? FRBSF Economic Letter archive: http://www.frbsf.org/economic-research/publications/economic-letter/1996/october/is-opportunistic-monetary-policy-credible/. Accessed 4 Oct 1996.

Steel, Daniel. 2008. *Across the boundaries: extrapolation in biology and social science*. New York: Oxford University Press.

Steel, Daniel. 2013. Mechanisms and extrapolation in the abortion-crime controversy. *Mechanism and causality in biology and economics*, ed. Hsaing-Ke Chao, Szu-Ting Chen, and Roberta L. Millstein, 185-206. Dordrecht: Springer.

Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica* 48: 1-48.

Sims, Christopher A. 2010. But economics is not an experimental science. *Journal of Economic Perspectives* 24: 59–68.

Stokey, Nancy L. 1991. Credible public policy. *Journal of Economic Dynamics and Control* 15: 627-656.