The Limits of Inference *with* Theory:
A Review of Wolpin (2013)

John Rust
Georgetown University

April, 2014

1 Introduction

Kenneth Wolpin is a leading economist who has made pathbreaking contributions to economics. He is best known for his work on "structural microeconometrics" which attempts to tightly integrate theoretical models into empirical work. His work, and that of his numerous students and coauthors, has had a huge impact on applied micro. It parallels in many respects the revolutionary impact that Lars Hansen and Thomas Sargent have had on applied macro from their equally pathbreaking work on structural macroeconometrics.

His monograph, *The Limits of Inference without Theory,* evolved from two lectures he presented at the Cowles Foundation at Yale in 2010 in honor of Tjalling Koopmans. The antecedents of modern structural macro and micro econometrics (including the simultaneous equations model and some of the earliest work on endogeneity and instrumental variables) can be traced to Cowles, and particularly to Koopmans, and other founding fathers such as Jacob Marschak and Trygve Haavelmo. In fact the *raison d'être* of the Cowles Foundation is to promote a tighter integration between theory and measurement in economics. The title of Wolpin's monograph recalls an essay by Koopmans (1947) "Measurement without Theory" that reviewed Burns and Mitchell (1946) *Measuring Business Cycles.* Koopmans criticized the "decision not to use theories of man's economic behavior, even hypothetically" because the absence of theory "limits the value to economic science and to the maker of policies" and "greatly restricts the benefit that might be secured from the use of modern methods of statistical inference." (p. 172).

Though it is hard to disagree with the Cowles Foundation's mission to forge a tighter bond between theory and empirical work in economics, Wolpin concludes his monograph by stating that "The proper role of economic theory in empirical research has been, and remains a controversial issue." (p. 149). Why? Though he does an admirable job of illustrating the benefits of using theories and models to guide empirical research, Wolpin does not explain with equal vigor why structural econometrics and the central mission of the Cowles Foundation should still be so controversial more than six decades after the "Koopmans critique".

Disagreement over the role of theory in empirical work is long standing. According to Wikipedia, it was a major reason why Koopmans moved the Cowles commission to Yale in response to "rising hostile opposition . . . by the department of economics at University of Chicago

during the 1950s". The hostility is just as strong today, and not just at Chicago. To better understand the opposition to "modeling" I refer readers to the work of Charles F. Manski, an equally eminent economist and econometrician who has been extremely influential and whose work I also very much admire. Manski's most recent book *Public Policy in an Uncertain World: Analysis and Decisions* (2013a) (reviewed by John Geweke, 2014) provides a very vigorous and skeptical counterpoint to Wolpin's book. In the Introduction, Manski notes that "researchers regularly express certitude about the consequences of alternative decisions. Exact predictions of outcomes are common, and expressions of uncertainty are rare. Yet policy predictions often are fragile. Conclusions may rest on critical unsupported assumptions or on leaps of logic. Then the certitude of policy analysis is not credible." (p. 2-3).

Three decades prior to Manski, Edward Leamer disparaged the state of applied econometrics for many of the same reasons in his famous (1983) "let's take the con out of econometrics" paper. He advocated the use of "extreme bounds analysis" that is similar in some respects to the bounding approach Manski has advocated to map out what can learned about parameters of interest from data under minimal assumptions. Leamer (1978) also expressed deep skepticism about the validity of traditional classical econometric inference because it does not properly reflect the results of "specification searches." In their review of the current state of the empirical literature in applied micro Angrist and Pischke (2010) echo Leamer's concerns, particularly the "distressing lack of robustness to changes in key assumptions" (p. 3). However they argue there has been a "credibility revolution" in empirical micro research over the last two decades in which improvement has not come from better modeling, rather "Improvement has come mostly from better research designs, either by virtue of outright experimentation or through the well-founded and careful implementation of quasi-experimental methods." (p. 26).

Notice the huge difference in world views. The primary concern of Leamer, Manski, Pischke and Angrist is that we rely too much on assumptions that could be wrong, and this could result in incorrect empirical conclusions and policy decisions. Wolpin argues that assumptions and models could be right, or at least they may provide reasonable first approximations to reality. He provides convincing examples of how the use of theory in empirical work results in much greater insight and understanding of complex phenomena, and this results in *better* policies and decisions. The

other major reason why it is good to have a tighter integration of theorizing and empirical work is to test and improve our assumptions, models, and theories. Our empirical conclusions may be sensitive to assumptions, but by confronting data and theory we can develop better, more empirically relevant theories and we can discard or relax assumptions that are inconsistent with what we observe. How can anyone disagree with that?

My main criticism is that Wolpin could have done a better job of acknowledging the legitimate concerns of leading skeptics such as Angrist and Pischke, Leamer, and Manski. By overstressing the limits of inference *without* theory and failing to clearly explain the limits of inference *with* theory, some readers may conclude that Wolpin is selling a methodology that may have hidden flaws. These limits, not adequately expressed in Wolpin's book or in Manski's, are the main focus of section 2, though several of them have already been discussed by James J. Heckman (1992a,b), (2007), (2010). Insightful papers by Heckman and Navarro (2006), Heckman and Urzua (2009) and Heckman and Vytlacil (2007a,b) provide creative new ways to deal with some of these problems. Heckman is arguably the deepest thinker on these issues, and he is one of very few who have been able to bridge the huge gulf between the structural and non-structural branches of econometrics and make huge contributions to both sides. While my views are generally consonant with Heckman's, there are some areas of more significant disagreement, particularly with respect to the identification problem as a fundamental limit to inference and whether structural econometrics has been "empirically fruitful" (Heckman 1992, p. 883). Section 3 discusses the huge payoffs from the tight integration of theory and inference in physics, engineering, and neuroscience. This section makes clear that I am in overall agreement with Wolpin (2013): we face fewer limits and can learn much more when we do inference *with* theory than without. Section 4 discusses the role of experimentation, because it is widely and mistakenly seen as a *substitute* for theory and structural modeling. I am in complete agreement with both Heckman (2010) and Wolpin (2013) that experiments can be considerably more powerful when they are done as a *complement* to structural modeling and inference. Section 5 offers a few concluding remarks. I warn the reader at the outset that this is not a traditional review. I have relatively little to say about what is in this book because Wolpin already says it very well and I agree with most of it. This review is about the 900 pound gorilla that Wolpin chose not to talk about.

2   Models, Theories, and the Limits of Inference in Economics

Models play a key role in structural inference, yet the reader of Wolpin's book may be disappointed to find there is no general definition of what a "model" is, or anything close to a formal proof of the proposition implicit in the title of this book: namely, that methods of inference that fail to make use of a model or a theory will be somehow limited in what can be learned/inferred from any given data compared to methods of inference that use models and theories. Instead Wolpin makes this point by a series of examples. I think this is the best he could do, since I am not aware of any truly formal, reasonably general, and successful "theory of inference" wherein such a proposition could be formally proved.[1] Heckman (1992a) agrees that statistics and econometrics are very far from constituting an adequate theory or guide to empirical scientific *discovery* and an inadequate and incomplete theory of how individual scientists and the scientific community at large should optimally *learn from data* "Since we do not fully understand the process of discovery or the social nature of the knowledge achieved from this process ('agreement' in the scientific community), and the role of persuasion in forming consensus, it is not surprising that mathematically precise models of discovery are not available." (p. 883).

It is important to have some degree of agreement on what a model is, since different people have very different definitions, some more encompassing than others. For example Thomas Sargent defines it simply as "A model is a probability distribution over a sequence (of vectors), usually indexed by some fixed parameters." (private communication). This definition seems overly encompassing, and it would seem to include the "linear regression model" as a special case. I doubt that Wolpin would agree that the linear regression model would count as a model in his lexicon, unless the regression model were somehow derived from a deeper theory, rather than simply posited as a best linear predictor relationship between a dependent variable *y* and some vector of independent variables *x*. The linear regression has been a central model in economics,

---

[1]There have been a number of interesting attempts to construct formal theories of learning, inductive/deductive inference reasoning. A short survey includes a theory of inductive inference by Solomonoff (1964), Simon's work on modeling human problem solving and learning (Newell and Simon, 1972, Feigenbaum and Simon (1984)), decision theory, including Bayesian decision theory and recent work on decision making under uncertainty and "ambiguity" (i.e. where agents are not fully aware of the probability distributions governing uncertain payoff-relevant outcomes, see e.g. Gilboa and Schmeidler 1989, Einhorn and Hogarth 1986, Klibanoff *et. al.* 2005) and extensions to dynamic decision making under ambiguity (e.g. Hansen and Sargent 2008 book on "robust control"), the literature on machine learning and statistical learning theory, (e.g. Vapnik 1998, Mohri *et. al.* 2012), and recent work by economic theorists to model inductive inference (e.g. Gilboa *et. al.* 2013a) and "meta-models" of how and why economists construct models and use them to gain new knowledge (Gilboa *et. al.* 2013b). It is beyond the scope of this review to suggest how Wolpin's proposition might be stated and proved more formally, but these references, particularly the last, provide the beginning of a framework under which this might be done.

but mainly under the additional assumption that it reflects a *causal relationship* between x and y.[2]

By failing to give a sufficiently clear and precise definition, Wolpin leaves himself open to criticism that he has an overly narrow view of what a model is. For example Frijters (2013) in his review of Wolpin's book concludes "From his examples, it is clear that what Wolpin means by structural is the assumption that individual agents rationally maximise a discounted stream of utility functions themselves dependent on stable preference parameters, augmented by a whole set of issue-specific ancillary assumptions to make the estimation tractable. Reduced form is then primarily the absence of the requirement that particular choices are maximising a given function." (p. 430).

A careful reading of Wolpin's book reveals that his view of a model is *not* this narrow. Though it is true that his own models usually involve rational, optimizing agents, Wolpin includes a much wider class of theories in the class of structural models, including "behavioral" theories, models of agents who have "irrational" or subjective beliefs, or theories involving time-inconsistent or suboptimal decision making such as the work by Fang and Silverman 2009 on hyperbolic discounting (which Wolpin cites in chapter 3). Indeed Wolpin states early on in the book "The structural estimation approach requires that a researcher explicitly specify a model of economic behavior, that is, a theory." (p. 2) and then he quotes a more detailed definition of Marschak (1953) that a *structure* consists of "(1) a set of relations describing human behavior and institutions as well as technological laws and involving in general, nonobservable random disturbances and nonobservable random errors in measurement; (2) the joint probability probability distribution of these random quantities." Note there is no requirement of *rationality* or *optimization* in this definition.

The term *structure* or *structural model* has an additional meaning that many economists ascribe to, that requires the analyst to be able to specify and identity *deep parameters* that are *policy invariant*. Wolpin ascribes to this view too, since he uses this as an additional criterion to distinguish "quasi-structural models" from "structural models". In a quasi-structural model "The

---

[2]See Heckman and Pinto (2013) who contrast a "statistical" definition of causality with the Cowles/Haavelmo structural model-based definition of causality. They credit Haavelmo with providing the "first formal analysis of the distinction between causation and correlation." They point out the limits of statistical definitions of causality, and conclude that a purely statistical framework "cannot accommodate the fundamentally non-recursive simultaneous equations model. The hypothetical model readily accommodates an analysis of causality in the simultaneous equations model." (p. 45).

relationships that are estimated are viewed as approximations to those that are, or could be, derived from the theory … The parameters are functions of the underlying deep (policy-invariant) structural parameters in an unspecified way." (p. 3). This also corresponds to the view of Sims (1981) which he in turn credits back to Koopmans and others in early work at the Cowles Foundation: "A structure is defined (by me, following Hurwicz 1962 and Koopmans 1959) as something which remains fixed when we undertake a policy change, and the structure is identified if we can estimate it from the given data." (p. 12).

The reason why we are interested in doing inference with structural models is well understood: econometric policy evaluation and forecasting is either impossible or highly unreliable using non-structural or quasi-structural models. This is the point of the famous *Lucas critique* (1976). Robert Lucas, Jr. criticized the quasi-structural models at that time, such as the large scale macroeconomic forecasting models developed by Lawrence Klein and others, as being unreliable vehicles for policy forecasting. Lucas stated the key rationale for why structural models will provide a more reliable basis for policy forecasting quite simply: "given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models." (p. 41). Lucas acknowledged that he was not the first to make these observations, but his paper had a powerful impact.[3] Not only did it largely uncut the credibility of the practitioners of the large scale macro forecasting models, it also provided an important impetus for the development of both structural macro and microeconometric methods. The first dynamic structural econometric models appeared in the late 1970s, shortly after Lucas's paper was published.

Looking back nearly four decades after the Lucas critique paper, it is fair to ask whether structural models really have succeeded, and resulted in significantly more accurate and reliable policy forecasting and evaluation. I think the jury is still out on this, because even though Wolpin has offered some compelling examples of the use of structural models for policy evaluation,

---

[3]Lucas built on similar ideas expressed in Marschak (1953), and he in turn was preceded by Haavelmo and Frisch. Heckman (1992a) in his review of Mary Morgan's (1991) book *The History of Econometric Ideas* notes that "a major feature of Haavelmo's program was the careful definition of 'policy invariant' or 'autonomous' structural relationships that he defined to be the objects of interest to economists evaluating alternative policy interventions" (p. 879) and Frisch "attacked the equations in Tinbergen's model for their lack of 'autonomy' — what we now call lack of 'policy invariance.' … Another popular theme in modern macroeconometrics and 'deep structural' microeconometrics was already being played in 1938." (p. 878). However it is fair to say that Frisch's, Haavelmo's and Marschak's insights were ahead of their time, and the article by Lucas is more well known, perhaps because conditions were more ripe for these points at the time it appeared.

there are still relatively few clearcut successes where structural models have had a objectively measurable positive impact on *actual* policymaking.

I do give Wolpin huge credit for the successful application of his structural model with Petra Todd (2003) on fertility and school attendance of Mexican households and their demonstration that their model provided reasonably accurate out of sample forecasts of the effect of the Progresa school attendance subsidy on the treatment group (treated as a holdout sample), having estimated the model using only individuals in the control group.[4]

Wolpin also cites the work of Lumsdaine, Stock and Wise (1992) who showed that structural retirement models provided much more accurate forecasts of changes of a Fortune-500 company's retirement plan (the adoption of a temporary "retirement window" incentive plan) than reduced form models. But it not clear that their analysis changed the firm's retirement policy.

In my own work with Sungjin Cho (2010) we used structural econometric methods to uncover evidence of suboptimal decision making by a car rental company. Our model predicted the company could make significantly higher profits by adopting an alternative policy of keeping its rental cars longer and offering discounts to its customers to induce them to rent the older cars in its fleet. The company was convinced by the econometric exercise to conduct a controlled experiment, which validated the predictions of our model (in fact, profits increased by more than our model predicted). However the company did not adopt this alternative more profitable operating policy. This is a puzzle because company executives found both our intuitive economic arguments and econometric modeling to be sufficiently convincing to motivate them to do the experiment, and they told us they found the experiment to be a compelling proof that switching to a policy of keeping its rental cars longer would significantly increase profits.[5]

The most convincing example of a practical success from a structural approach to inference and policymaking that I am aware of is Misra and Nair (2011) who estimated a dynamic structural model of the sales effort of a sample of contact lens salesman. They showed that the company

---

[4]Angrist and Pischke (2010) and others might disagree that the ability of the Todd and Wolpin model to provide relatively accurate out of sample predictions is the main success, rather they would assign credit to the Progresa experiment itself. For example rather than acknowledging Todd and Wolpin (2003) they quote Paul Gertler "Progresa is why now thirty countries worldwide have conditional cash transfer programs." (p. 4).

[5]One possible explanation for the failure to change policy is that the company was acquired by a large conglomerate shortly after the experiment was completed and the new owners had little expertise or understanding of the rental business. The rental company executives were worried about being replaced, and that it would not be easy to explain the logic of how the policy of keeping rental cars longer than industry norm could increase profits and not jeopardize the firm's "high quality" reputation to their superiors in the acquiring conglomerate.

had adopted a suboptimal compensation plan consisting of salary, quota, and bonus that inefficiently motivated its sales force. Their structural model revealed that the company's combination of a sales quota and maximum commission ceiling introduced a particular inefficiency, namely the most productive sales people would slack off once they had reached the commission ceiling. Using the estimated structural model, they designed an improved incentive plan that reduced the sales quota and removed the commission ceiling. The company actually implemented their recommended alternative compensation scheme. "Agent behavior and output under the new compensation plan is found to change as predicted. The new plan resulted in a 9% improvement in overall revenues, which translates to about $12 million incremental revenues annually, indicating the success of the field-implementation. The results bear out the face validity of dynamic agency theory for real-world compensation design. More generally, our results fit into a growing literature that illustrates that dynamic programming-based solutions, when combined with structural empirical specifications of behavior, can help significantly improve marketing decision-making, and firms' profitability." (p. 211-212).

So this is an objectively verifiable practical success that validates the structural approach to estimation and policy evaluation that Lucas envisioned in his 1976 paper. But nearly four decades after the Lucas critique, the structural estimation industry would be in a much stronger position if we had a larger number of clear successes that we could point to. Heckman (1992a) offers a more critical assessment: "It is unfortunate that Morgan never adequately addresses why the Haavelmo-Cowles program has not been empirically fruitful. More was involved than the computational complexity of the econometric methods. Roy Epstein notes that by the late 1940s, the empirical returns from the Cowles program were perceived to be at best mixed even by its advocates." (p. 883). In his review of Haavelmo's legacy, Heckman (2007) notes "Haavelmo made basic contributions to econometric methodology and to the foundations of causality, prediction and policy analysis. His program for learning from data is less successful." (p. 42).

However I would take issue with Heckman's view that the structural approach to inference has not been empirically fruitful and his own writings send a mixed message on this. First, Heckman does acknowledge that structural macroeconometrics has been successful "As 'the' paradigm of 'scientific' work in econometrics, it and the important revision of it by Robert Lucas and Thomas

8

Sargent (1981) are successful as measured by frequency of citation to it in the official rhetoric of structural econometricians." Second, Heckman and Vytlacil (2007a) cite six different chapters in Volume 6B of the *Handbook of Econometrics* that survey a rapidly growing literature that uses "newly available microdata on families, individuals and firms to build microstructural models" (p. 4782). Third, Heckman suggests that though non-structural empirical work is *perceived* to be highly successful, some of this may be illusory: "In many influential circles, ambiguity disguised as simplicity or 'robustness' is a virtue. The less said about what is implicitly assumed about a statistical model generating data, the less many economists seem to think is being assumed. The new credo is to let sleeping dogs lie. Haavelmo himself knew that he was promising the intellectual equivalent of 'blood, sweat, toil, and tears' if economists took his program for empirical research seriously (1944, p. 114)." (p. 882).

It is true that structural econometricians attempt to confront a number of challenges that non-structural econometricians prefer to sweep under the rug, so it is unrealistic to expect that it will be as popular or that progress measured by the number of publications and practical successes will come as fast. One thing is clear from Wolpin's book is the vast majority of the work he has done did not involve taking convenient shortcuts: he faced significant challenges head-on and was clear about the modeling assumptions he made. While there are certainly many more non-structural empirical papers than structural ones, it is not obvious to me that Heckman, Manski, Angrist or Pischke could point to a substantially larger number of clear cut successes in practical policymaking that can be directly credited to specific reduced-form econometric studies.[6]

However it is important to consider whether some of challenges facing structural econometrics are actually fundamental limits to inference, and which are more akin to sociological/cultural obstacles such as ignorance and fear of modeling which Heckman (1992a) notes is widespread,

---

[6]I asked Heckman and Manski for their best examples of *practical success* in econometric policy making. Heckman did not reply, but Manski replied that "Your definition sets a high bar and I cannot easily think of any empirical social science work that truly achieves it." Angrist and Pischke (2010) cite Lalonde (1986) as a key success "A landmark here is Lalonde (1986), who compared the results from an econometric evaluation of the National Supported Work demonstration with those from a randomized trial. The econometric results typically differed quite a bit from those using random assignment. Lalonde argued that there is little reason to believe that statistical comparisons of alternative models (specification testing) would point a researcher in the right direction." (p. 5). Heckman, whose work/legacy was most directly called into question by Lalonde's results, acknowledges this as an "influential paper" but notes that "Heckman and V. Joseph Hotz (1989) cautioned that many applications of the structural approach by those comparing structural estimates with experimental estimates did not perform specification tests to see if the estimated structural models were concordant with the preprogram data. They show that when such tests are performed, the surviving structural models closely match the estimates produced from the experiment analyzed by LaLonde, findings duplicated for other experiments" (Heckman, 2010, p. 357). I agree that Lalonde's work was *academically influential* but it is not clear to me that his paper generated new insights that had practical payoffs for *policymakers,* or had any effect on how job training programs are structured/administered. LaLonde's work, similar to that of Leamer (1983), was influential in a mainly negative sense — it questioned the credibility of policy forecasts of structural econometric models without offering a more credible alternative other than the already well known approach of randomized experiments.

even among statisticians "Mechanism and explanation are avoided by modern statisticians who define parameters of interest to be determined by outcomes of randomized experiments rather than as the outcomes of scientific model building using controlled variation and logic to derive estimating equations." (p. 880). A clear discussion of what the risks, rewards, and limits are is especially important for young people who are considering whether to do structural econometrics. Wolpin has obviously been a very successful adviser, and has had considerable success on his own in getting his research published, so he is serving as a very effective role model for young people considering following his path. I have already discussed some of the professional risks and obstacles in my comments on the paper by Michael Keane's essay "Structural vs. Atheoretic Approaches to Econometrics" (Rust 2009) and won't repeat them here. Below I summarize the key *logical limits* to inference that explain why, despite all the talent and effort invested in the development of models and structural methods of inference, there will be challenging questions to which we may never be able to provide satisfactory answers to, and for the easier questions to which answers might be found, progress in finding credible answers is likely to be painstakingly slow.

2.1   Policy Invariant Objects or Parameters May Not Exist

Structural econometrics is based on a key *assumption* that there are "deep policy invariant objects/parameters" that are identified and can be recovered via structural methods of inference. Once these policy invariant are estimated/inferred it is possible to use them and the model to predict how the "system" (i.e. an economy, market, firm, or individual) will evolve under alternative policies and technologies. Economists typically assume that *preference parameters* and *technology parameters* are of this variety — they are the truly structural or deep policy-invariant parameters that structural econometricians are trying to uncover. But what if this is a fiction? What if there really are no fully policy, technology, or socially/culturally independent parameters or objects? Joseph Stiglitz, in his 2001 Nobel Prize lecture, made precisely this point "There were other deficiencies in the theory, some of which were closely connected. The standard theory assumed that technology and preferences were fixed. But changes in technology, R&D, are at the heart of capitalism. ... I similarly became increasingly convinced of the inappropriateness of the

assumption of fixed preferences."[7]

All the recent discussion of technologically induced "structural change" in the economy has to make the "true believers" of the Lucas critique rather nervous. In fact, few things seem truly "invariant" these days, other than Kurzweil's (1999) *Law of accelerating change.* This law states that the rate of technological progress is itself accelerating at an exponential rate. Kurzweil (2005) claims that we are rapidly approaching a *singularity* (to occur sometime around 2050) when the rate of change will approach infinity. Kurzweil believes that it is impossible to predict in any detail what things will be like beyond this singularity, and it will be increasingly difficult to forecast the future as the rate of change accelerates. So not only are there no clear structural, invariant parameters or objects, accelerating change is making the future inherently more unpredictable. Rapidly evolving technology and knowledge alters our behavior and institutions, and thus both the structure of individual preferences and decision making, and thus the economy as whole. This is a huge challenge to the Cowles Commission's approach to econometric policy evaluation and forecasting. For example, it calls into question the validity of most long term policy forecasts, such as projections of the Social Security Trust Fund that go out to 2070.

Examples of the lack of policy invariance of structural parameters arise in a number of contexts such as in models of the decision to apply for welfare or disability benefits. Moffitt (1983) was one of the first to show that *stigma* associated with applying for welfare could explain why only "69 percent of families eligible for AFDC (Aid to Families with Dependent Children) participated in the program" (p. 1023). Similarly, in my own work on decision to apply for disability insurance with Benitez-Silva and Buchinsky (2003), we find long delays and puzzlingly low application rates among individuals who experience a disabling condition. Though it is also possible to explain this low take up by assuming that these individuals are simply unaware of the option to apply for disability, we find this level of ignorance to be implausible. An alternative way to explain the low take up rate is to include parameters that reflect disutility or stigma associated with applying for benefits and being on the disability rolls. However stigma parameters are not

---

[7]James Heckman, in a comment on a version of this review, notes that there are structural models of the evolution of preferences, however from what we know of neuroscience and the development of the human brain, it is not entirely clear what the deeper "policy invariant" parameters of dynamically evolving preferences might be. Perhaps some of these parameters will be related to heredity and genetic structure of intelligence and cognition that affects how the development of the brain, preferences, and knowledge interact and coevolve with environmental experiences.

policy-invariant preference parameters: the government can and has used the media in what might be described as a propaganda effort to stigmatize/demonize individuals who apply for disability and welfare, such as during the Reagan and Clinton administrations.[8]

Wolpin acknowledges that stigma parameters have played a role in his own work on welfare participation, including his (2007) and (2010) studies with Michael Keane. Though he briefly mentions that their models include parameters that capture "direct utilities or disutilities for school, pregnancy and welfare participation" (p. 94) he does not point out that these parameters may not be structural, i.e. policy-invariant. Instead, he notes that "The effect of welfare participation of replacing the level of welfare stigma of black women with that of white women is relatively small, as is the effect on other outcomes as well." (p. 100). In private comments on this review Wolpin acknowledges that "One could rightly claim that some of our policy experiments, such as introducing time limits, are not credible if one believes that the degree of stigma associated with the new policy differs from that with the old policy."

I agree. I think structural econometricians need to think more deeply about whether they can justify whether *any* parameters of their models are really "structural" in the sense of being policy-invariant, and what do if it turns out they have no good justification for this. Otherwise I feel there is a ticking time bomb and some future Robert Lucas will come along and write a review titled "Structural Econometric Model Policy Evaluation: A Critique." This review will echo the same sorts of criticisms that Robert Lucas lodged against the large scale macro models in his 1976 paper, and it could have the same revolutionary effect on today's structural estimation industry that Lucas's paper had on the large scale macromodeling industry in the late 1970s (it basically destroyed it).

---

[8]The Clinton administration disallowed alcoholism as a disabling condition, and instituted a much tougher version of welfare, *Temporary Aid for Needy Families* (TANF) based in part on derogatory view of that previous program, AFDC, that it encouraged "welfare mothers" and higher rates of out-of-wedlock births and a culture of welfare dependency. Though the policy change was deemed successful in greatly reducing the number of poor people receiving benefits, it may have done this partly by increasing the level of social stigma, and thereby reduce the incentive to apply to the program. If so, it is hard to describe the stigma parameters representing the disutility of receiving welfare benefits (and typically necessary to enable structural models to fit the data), as structural or policy-invariant parameters. The Reagan administration suggested that many individuals receiving disability benefits were imposters and instituted a mass audit policy that resulted in huge numbers of terminations of disability benefits, and new applications for disability benefits also dramatically fell in the aftermath of this policy change as well. While some of the response may be a rational response to an expectation of lower disability award rates, the structural models I have estimated must rely on an increase in the stigma to being on disability to explain the full magnitude of the response to the Reagan "reforms".

2.2 The Curse of Dimensionality

Richard Bellman coined the term "curse of dimensionality" to refer to the exponential increase in the amount of computer time required to solve a dynamic programming model as the number of variables (or other measures of the "size" or "complexity" of the problem) increases. Subsequent work in computer science (e.g. Chow and Tsitskilis, 1989) established that the curse of dimensionality is an insuperable problem and not just a reflection of insufficient creativity in finding better algorithms to solve dynamic programming problems. The curse of dimensionality also appears in statistics: for example the rate of convergence of *any* nonparametric estimator of an unknown regression function is inversely proportional to the number of continuous variables in the regression function (see, e.g. Stone, 1989). Thus with limited data and computing power, the degree of precision in the inferences we can make and the size of the economic models we can solve will be limited. The curse of dimensionality forces us to work with fairly simple models because we can't solve bigger, more realistic ones. It also implies that it may be a very long time before we will have sufficient data and computer power to be able to provide more realistic and accurate structural models of highly complex interacting phenomena (e.g. the financial system) to have any confidence that the policy forecasts of structural models of complex systems have any degree of credibility. Though more data and greater computer power improve the realism of our models and the quality/reliability of the conclusions we can reach, the rate of progress will be far slower than the (exponential) rate at which data and computer power are increasing.

Sometimes we can break the curse of dimensionality though not without cost. Rust's (1997) random multigrid algorithm solves discrete choice dynamic programming problems in polynomial time, but the approximate solution has stochastic error that can be made arbitrarily small, but only at the cost of increasing the number of random draws and thus the computer time required by the algorithm.[9] Barron (1989) proved that the curse of dimensionality of non-parametric regression can be broken for certain classes of multivariate functions that have "special structure" but at the cost of finding a global minimum to a nonlinear least squares problem, and the time to do this increases exponentially fast as the number of variables increases.

---

[9]Rust, Traub and Woźniakowski (2003) showed that it is possible to break the curse of dimensionality for a class of contraction fixed point problems (which include Bellman equations for discrete choice dynamic programming problems as a special case) that satisfy stronger smoothness properties than the Lipschitz continuity assumptions Rust (1997) used in his analysis.

Whether the curse of dimensionality is a fundamental limit to knowledge and inference depends on whether we think that humans are really actually solving dynamic programs, or that markets and economies are actually in equilibrium. There may be a much more easily computable behavioral adaptive learning process that agents are using that can be modeled as solutions to simpler mathematical problems or "rules of thumb" that are not subject to the curse of dimensionality. After all, human beings do seem to be capable of amazingly complex calculations, and the failure of "artificial intelligence" so far may simply be an indication that we are still at a primitive stage of discovering and replicating the complex calculations and effective rules of thumb that nature has "discovered" over the course of millions of years of evolution. It is entirely possible that we are being boneheaded in viewing all consumers as solving complex dynamic programming problems and/or playing dynamic Markov-perfect Nash equilibrium strategies. We may find that there are simpler behavioral "agent-based" models that can generate rich, and highly realistic behavior but are not subject to the curse of dimensionality. However the problem is that there may be *too many* of these behavioral models that can explain the data we observe.

2.3   The Identification Problem

The most daunting limit to knowledge that structural econometricians face is the *identification problem.* This is the problem of trying to infer the *structure* — set of underlying *primitives* that imply a probability distribution for the observable variables.[10] Structural models depend on a number of *maintained assumptions* such as the assumption that agents are expected utility maximizers, or have rational expectations. The maintained assumptions are generally outside the domain of the identification analysis (i.e. they are treated as assumptions that cannot be altered, tested, or questioned) unless we are willing to impose other independent maintained assumptions (i.e. parametric assumptions on the functional form of payoffs and beliefs of agents of firms). The identification problem tells us that even in the presence of very strong maintained assumptions such as rational expectations and dynamic expected utility maximization, it may not always be

---

[10]We can discuss the identification problem without also assuming all of the model "primitives" are "structural" in the sense that all of the primitives must be policy-invariant as Lucas (1976) and Marschak (1953) envisioned.

possible to infer the underlying structure without making a number of additional very strong assumptions about the functional form of agents' preference and beliefs.

Unfortunately, Rust (1994) and Magnac and Thesmar (2002) proved that the underlying structure (agents' preferences, beliefs and discount factors) is non-parametrically unidentified, even in the presence of the *very strong maintained assumptions* that include 1) rational expectations, 2) expected utility maximization, 3) preferences that are additively separable over time and over the unobserved states, and 4) conditional independence that restricts the way unobserved variables can affect the observed state variables. These strong maintained assumptions automatically imply that we can uncover part of the structure — agents' beliefs — nonparametrically, and thus beliefs are identified if we are willing to impose the very strong assumption of rational expectations.

However even if we assume the distribution of unobserved state variables is known (for example if this distribution is a Type 3 Extreme value, which McFadden showed leads to choice probabilities that have the form of a *multinomial logit*) it is still not possible to identify preferences up to a positive affine transformation of of the agent's true utility function, nor is it possible to identify the agent's discount factor. Instead the "identified set" of structural objects includes all discount factors in the $[0, 1]$ interval, and a much wider class of preferences that include utility functions that are not monotonic affine transformation of the 'true" utility function. In fact, it is possible to rationalize *any* conditional choice probability as resulting from an optimal decision rule from a discrete choice dynamic programming problem for a broad class of utility functions, and we can rationalize it in an infinite number of ways, including via a static model or a dynamic model for any discount factor in the unit interval.

These "single agent" non-parametric, non-identification results can be viewed as a special case of a more general result of John Ledyard (1986) who showed that the maintained hypotheses of rationality and Bayesian-Nash equilibrium fail to place any testable restrictions on behavior if we are given sufficient freedom to choose agents' preferences: "What behavior can be explained as the Bayes equilibrium of some game? The main finding is — almost anything. Given any Bayesian (coordination) game with positive priors, and given any vector of nondominated strategies, there is an increasing transformation of each utility function such that the given vector of strategies is a Bayes (Nash) equilibrium of the transformed game. Any non-dominated behavior

can be rationalized as Bayes equilibrium behavior." (p. 59).

Lack of identification of a structural model means that policy evaluation and forecasting is problematic. Suppose there are multiple structures that map into the same reduced form. This means these alternative structures are *observationally equivalent* — at least under the particular *status quo* policy regime from which the structure has been identified. Now consider some hypothetical policy change, for which there is no historical antecedent (and thus no basis in the data to forecast how agents will respond to the policy change). If the two different, but observationally equivalent structures result in different forecasted behavioral responses and/or changes in agents' welfare, which one of them do we believe?

On the other hand, suppose that we could do a randomized controlled experiment, where subjects randomly assigned to the treatment group are subjected to the new policy, whereas those in the control group remain with the *status quo*. Also suppose there were only the two observationally equivalent structures and the experiment would reveal that one of these structures was wrong and the other was the right one. Then we can use *additional data generated from an experiment* to help us identify the correct structure. But what if there are many different structures (in the worst case infinitely many) in the "identified set"? In that case, even though a single experiment can help to eliminate some of the structures as not being in the identified set (since these structures would predict a response that is inconsistent with the experimental outcome), it is entirely possible that there are still many structures that will correctly predict the agent's behavior under the *status quo* and under the hypothetical new policy (i.e there are multiple structures that correctly predict behavior of the "control" and "treatment" groups). If this is the case, then even an infinite series of experiments may not be enough to identify the true underlying structure.

Though he does talk about non-parametric approaches to policy evaluation in static models in chapter 2, Wolpin does nearly all of his empirical work using dynamic models that depend on *parametric functional forms* for preferences, beliefs, technology, and so forth. I make parametric functional form assumptions in virtually all of my empirical work as well. The reason we do this is that the additional *a priori* restrictions provided by the parametric functional form assumptions are generally sufficient to identify the underlying structure. However the cost of this is that the parametric functional form assumptions restrict our flexibility in fitting a model to the data, and if

16

the parametric assumptions are incorrect — i.e. if the model is misspecified — then the resulting model will generally not be able to provide a perfect fit to the data, unlike the case when we do not impose any parametric restrictions on preferences or beliefs where we generally have sufficient flexibility to perfectly rationalize the data we observe.

I believe that most interesting economic models are either non-parametrically unidentified or at best partially identified. If we allow the huge freedom of an infinite dimensional structural "parameter space" and find that we can rationalize any behavior in many different ways, have we really learned anything? I think the answer is no: a theory that provides so much freedom that it can explain everything actually explains nothing. Theories are only (empirically) interesting when they have testable, (and therefore rejectable) predictions.

Structural econometricians (myself and Wolpin included) can be caricatured as repeatedly going around and looking for ways to rationalize this or that observed behavior as optimal according to sufficiently elaborate and complicated dynamic programming model. In fact, we have gotten so good at rationalizing virtually *any* behavior as being "optimal" for some set of underlying preferences and beliefs that it is not even clear how we would define what a "bad decision" is! However the experience of the last decade — particularly the bad decision making leading the Bush administration to invade Iraq, the clearly myopic behavior of so many people in the mortgage boom leading up to the financial crash in 2008, and the complete cluelessness of economists about all of the above — has convinced me that many people, firms, and governments are behaving far from optimally and economists are being foolish in insisting on continuing to model all of the above as perfectly informed, perfectly rational dynamic optimizers.[11]

The growing interest in behavioral economics is also evidence that many other economists have similar opinions. However if structural econometricians are so good in rationalizing everyones' behavior using highly complex dynamic programming models, behavioral economists are equally foolish if they think it will it be easy to identify individuals who are not behaving

---

[11]I am not the only one who has made a relatively harsh assessment of the cluelessness of academic economists about the financial crash of 2008. A report by Colander *et.al.* (2009) concludes that "The economics profession appears to have been unaware of the long build-up to the current worldwide financial crisis and to have significantly underestimated its dimensions once it started to unfold. In our view, this lack of understanding is due to a misallocation of research efforts in economics. We trace the deeper roots of this failure to the profession's focus on models that, by design, disregard key elements driving outcomes in real-world markets. The economics profession has failed in communicating the limitations, weaknesses, and even dangers of its preferred models to the public. This state of affairs makes clear the need for a major reorientation of focus in the research economists undertake, as well as for the establishment of an ethical code that would ask economists to understand and communicate the limitations and potential misuses of their models."

optimally. If we already have a severe identification problem under the very strong maintained hypothesis of rational, dynamic expected utility maximization, how can behavioral economists possibly think things will be easier for them to identify a model from a substantially larger class of theories (i.e. weakening the maintained hypotheses to allow for non-expected utility, irrationality, time-inconsistency, time-non-separability, etc. etc.)? While it is true that expected utility has been rejected in narrow, specific cases using cleverly designed laboratory experiments (Allais paradox), the behavioral economists have failed to develop a comparably systematic, computationally tractable, and empirically convincing theory of human behavior that can replace expected utility theory as a workhorse for modeling a huge range of behaviors in many different contexts.

That said, I am positive about efforts to go beyond rational expected utility theory and consider a much richer class of more realistic behavioral theories. It would be really cool if we could make inference about the fraction of any given population who are "rational optimizers" and the fractions who are using any of a myriad of other alternative possible "irrational" or suboptimal behavioral decision rules/strategies. I believe this is a very difficult challenge, but a profoundly important one to undertake, since I think it matters *immensely* for policy making if we conclude that large fractions of individuals, firms and governments are not behaving rationally. While I think the identification problem is a very serious limit to knowledge/inference, I do not believe things are entirely hopeless. If we are willing to supply some prior input and take a stand, I believe we can get interesting and meaningful results.

For example El Gamal and Grether (1995) conducted a structural econometric study of inferential decision making by laboratory subjects. They imposed some prior assumptions but allowed subjects to use any one of a class of different decision rules for classifying which of two possible bingo cages a sample of colored balls (drawn with replacement) was drawn from. One of the possible decision rules they allowed was, of course, *Bayes rule,* but their study allowed other "behavioral" decision rules such as those based on *representativeness* (i.e. choosing the bingo cage that most resembles the sample that was drawn, irregardless of the prior probability of drawing from either of the two bingo cages). Surprisingly, they found that not all subjects use Bayes rule, but they found that the greatest fraction of the subjects used this rule, with the second most common rule being representativeness. Their analysis would have been impossible if they al-

lowed subjects to use *any possible* decision rule, but they found that they could obtain interesting results by imposing some *a priori* restrictions on the class of possible rules subjects could use along with parametric assumption about the distribution of "random errors" that enabled them to derive non-degenerate likelihood function for the observations. Thus, imposing parametric restrictions made it possible for them to conduct an interesting and informative study of an aspect of human decision making and inference.

The conclusion to their study is instructive of where further progress can be made more generally in structural estimation "The response of economists and psychologists to the discovery of anomalous violations of standard models of statistical decision theory has mainly been to devise new theories that can accommodate those apparent violations of rationality. The enterprise of finding out what experimental subjects actually do (instead of focusing on what they do not do; i.e., violations of standard theory) has not progressed to the point that one would hope. As a first step in that direction, we propose a general estimation/classification approach to studying experimental data. The procedure is sufficiently general in that it can be applied to almost any problem. The only requirement is that the experimenter or scientist studying the experimental data can propose a class of decision rules (more generally likelihood functions) that the subjects are restricted to use." (p. 1144).

Thus, I do not believe that interesting progress can be made if we insist on being completely agnostic and unwilling to place any restrictions on the structure of our models (e.g. on preferences and beliefs). While it is possible to go some distance with "nonparametric" restrictions such as monotonicity and concavity (see, e.g. Matzkin 1991) it is extremely computationally intensive to solve models that have no parametric structure whatsoever. I believe that parametric restrictions are more flexible and informative and greatly facilitate computational modeling. Further, we have a great freedom in which functional forms we choose, so we can think of parametric models as "flexible functional forms" whose flexibility can be indexed by the amount of data we have.

It is important to note that even when we impose parametric functional form assumptions, the resulting model will not always be identified, especially in actual situations when we are estimating a model with only a finite number of observations. The estimation criterion can have multiple global maxima (in case the estimation criterion is maximum likelihood) or minima (if

the estimation criterion is a minimum distance type of estimator), and there can be situations where the criterion can also be locally flat at the maximum (at least for certain parameters). In such cases the structural parameter estimates are set-valued instead of point valued. We learn very practically in the process of estimating a parametric structural model just what we can and cannot identify, so in my view, the identification of the model is very much a computational, data driven analysis, and very little can be said about identification in general from a highly abstract, mathematical vantage point.[12]

Heckman and Navarro (2006) complain that my results on the non-parametric non-identification of discrete choice models "has fostered the widespread belief that dynamic discrete choice models are identified only by using arbitrary functional form and exclusion restrictions. The entire dynamic discrete choice project thus appears to be without empirical content and the evidence from it at the whim of investigator choice about function forms of estimating equations and application of ad hoc exclusion restrictions." (p. 342). I do not believe that honestly and transparently acknowledging the reality that these models are non-parametrically unidentified means that structural estimation is a nihilistic, meaningless exercise. I think that Heckman and many other econometricians seem to lose sight of the fact that models are necessarily oversimplified approximations to reality and as such can never be correct. There is still an identification problem even if we recognize models are misspecified (no model may fit the data perfectly but several different "wrong" theories may fit the data almost equally well). But "econometrics as a search for truth" may be too idealistic a goal, given the limits to inference that we face. It might be better cast as a "search for models that provide reasonably good approximations" to otherwise highly complex phenomena. When we find different models that fit the data nearly equally well, we have a problem, and the best we can do is acknowledge the problem, and to try to gather more data, including running controlled experiments that can enable us to rule out the models/theories that do not provide good out of sample predictions of the behavioral response of the treatment group.

---

[12]What can be said at a high degree of generality is to apply the Morse Theorem of differential topology to prove that in appropriate topological sense, "almost all" parametric models are identified, in the sense that by making small perturbations of any unidentified parametric model, there exist other nearby parametric models that are identified in the sense that these models will have a unique parameter that best-fits the data (e.g. a likelihood that has a unique maximizer). However while this may be technically true, this result is not practically helpful when we learn in practice that the likelihood function is nearly flat at the maximum, i.e. when there are many parameters in a large subset of the parameter space that nearly maximize the likelihood function.

2.4 Multiplicity and Indeterminacy of Equilibria

Besides rationality and optimization, another fundamental economic principle is *equilibrium* —
be it dynamic general equilibrium in markets, or various flavors of Nash equilibria in static and
dynamic games. Finding even a single equilibrium has proved to be a daunting computational
challenge in many economic models, and until recently economists seemed content with just
proving that *an equilibrium exists.* However a line of work that includes a huge line of research
on the Folk Theorem for repeated games suggests that many economic models of games and
other types of dynamic models of economies with heterogeneous agents (which can often be
cast as large dynamic games) could potentially have a vast number of equilibria. For example,
Iskhakov, Rust and Schjerning (2013) show that even a simple finite state model of Bertrand
pricing with leapfrogging investments can have hundreds of millions of equilibria when the firms
move simultaneously to choose prices and whether or not to upgrade their plant to a state of the
art production technology. The number of possible equilibria grows exponentially fast with the
number of possible values for the "state of the art" production cost (which serves as an "exoge-
nous state variable" in the model), so in effect there is a curse of dimensionality in the number of
equilibria as a function of the number of discrete points in the state space.

These are disturbing findings because economic theory does not explain how players can
coordinate on a particular equilibrium when there many possible equilibria. Economists like
to impose equilibrium selection rules that pick out a preferred equilibrium from the set of all
possible equilibria of an economy or a game, but there is little evidence that I am aware of that
the different players have common knowledge of a given equilibrium selection rule and are able
to coordinate in the very sophisticated manner that game theorists presume in their equilibrium
existence and selection arguments.

Though there are studies that claim that we can identify, *nonparametrically,* preferences, be-
liefs, and the (state-dependent) *equilibrium selection rule* in static and dynamic games (see, e.g.
Aguirregabiria and Mira, 2013), I am very skeptical about these conclusions. I have already
discussed the non-parametric non-identification result for single agent dynamic programming
models in the previous section, but these can be viewed as "games against nature" and thus are
a very prominent and simple special case of the general class of games that Aguirregabiria and

21

Mira are considering. The general results of Aguirregabiria and Mira cannot be correct if they do not even hold in the special case of single agent games against nature, and their result appears to contradict Ledyard's (1986) theorem discussed in the previous section.

Wolpin does not devote any space to the structural estimation of dynamic games in his book, but he has worked on this problem in recent work with Petra Todd (2013). This paper models the joint choice of effort by students and the teacher in a classroom as a coordination game. "With student fixed costs, however, there are up to 2^N equilibria, where N is the class size. This makes it computationally infeasible to determine the full set of equilibria, which requires checking whether each potential equilibrium is defection-proof." (p. 4). Todd and Wolpin show that under a further assumption that "the ratio of the xed-to-variable cost does not vary among students within a class. In that case, students can be ordered in terms of their propensity to choose minimum effort and there are at most $N + 1$ equilibria that need to be checked, with different equilibria corresponding to different numbers of students supplying minimum effort."

While structural estimation of dynamic games is certainly an active "frontier area" of work, there are considerably more challenges to doing structural inference in games than in single agent decision problems. The first problem is how to compute all the equilibria and select a given equilibrium of interest out of the set of all equilibria. The estimation algorithms that are typically used require nested numerical solution of equilibria for different parameter values over the course of searching for best fitting parameter values (say parameters that maximize a likelihood function when it is possible to create a likelihood function that describes the probability distribution for different observed equilibrium outcomes of the game). One issue that is far from clear is what happens if the set of equilibria vary with different values of the structural parameters. It is not clear that it is possible to select a given equilibrium out of the set of all equilibria in a manner that an implicit function theorem can be established to guarantee basic continuity and differentiability properties needed to establish asymptotic properties of the estimator. But even more problematic is the question of how to do policy evaluation if a counterfactual policy alters the set of equilibria in the game. Does the policy alter the equilibrium selection rule as well? If so, what theory do we rely on to predict which equilibrium is selected after the policy change?

When there are many equilibria in a game, there is a "meta coordination" problem that needs

to be solved as to how the players select one of the large number of possible equilibria. It seems ironic to claim that game theory and Nash equilibrium provides a "solution" to the coordination problem (effort levels in the classroom in the case of Todd and Wolpin, or investment sequencing in the case of Iskhakov, Rust and Schjerning) when the players' choice of one of the many possible Nash equilibria in this game is itself another coordination problem.

It is not clear to me that there is compelling evidence that agents actually behave according to the predictions of Nash equilibrium, *especially* in situations where there are many possible Nash equilibria, or where the equilibria involved mixed strategies, or where the computational burdens of finding an equilibrium are implausibly large. If there is doubt about whether agents are individually rational, then it seems to be quite a leap to expect that collections of agents should exhibit the much higher level of rationality required to find a Nash equilibrium outcome. The work on "Oblivious Equilibrium" (Weintraub *et. al.* 2008) and related strategies can be seen as an attempt to relax the need for expectations over very high dimensional configurations of future states and decisions to find computationally simpler ways to approximate Markov Perfect Equilibria in games with many agents. However in view of the mindless, lemming-like behavior by so many investors and home buyers leading up to the 2008 financial crisis, perhaps we should be thinking of empirically more realistic theories that might be characterized as "oblivious *dis*equilibrium."

I do not want to be entirely dismissive of Nash equilibrium and rationality, and the fact that finding equilibria is difficult for us as economists may just be an reflection that we are still at a relatively primitive state of development in our ability to solve models. The concept of Nash equilibrium and modern digital computers are still in their relative infancy, having been invented just over 60 years ago. I note that progress in related areas such as artificial intelligence has also been far slower and more difficult than was previously expected. Even if we don't believe that real agents are behaving according to complex Nash equilibrium strategies, it seems reasonable to suppose that interacting adaptive, intelligent agents might converge to something close to a Nash equilibrium in a sufficiently stable environment. There are numerous theories that show this convergence is possible, though there are also counterexamples where plausible learning rules fail to lead repeatedly interacting agents to converge to Nash equilibrium behavior.

The dynamics of interacting, co-adapting, co-evolving intelligent agents can be highly com-

plex and can have multiple steady state outcomes. Thus, it may be very difficult to predict *ex ante* which of these steady state outcomes or "equilibria" are likely to arrive if a system is subjected to a shock that knocks it out of some initial steady state/equilibrium situation. If this is correct, there is a high level of interdeterminacy in these complex systems which makes makes policy forecasting all the more difficult (and if we also believe in the Law of accelerating change, then even the concept of a "steady state" is a figment of our imagination). It is not at all clear that we have good solutions to these problems, so it makes sense to acknowledge that given our present state of knowledge policy forecasting is far from something we would describe as a well understood science.

2.5   Limits of Deductive versus Inductive Modes of Inference

So far I have focused primarily on the limits to *inductive inference* (in the sense of learning from data) but there are also strict limits to *deductive inference* (in the sense of proving theorems from axioms). Gödel's (1931) celebrated *Incompleteness Theorem* shows that there are true propositions (theorems) which have no *proof* (i.e. no constructive procedure for establishing their truth from a consistent set of axioms) in any formal reasoning system that is at least as complex as formal arithmetic. However even though we know there are limits to deductive inference, this does not mean it is impossible to make huge strides via deductive modes of inference. Fermat's Last Theorem is one such example of a famous unsolved problem that has been proven to be true. Perhaps someday the $P = NP$ problem will be solved as well (i.e. it will be proved to be true, that $N = NP$, or false, $P \neq NP$). This has substantial practical importance, since if $P = NP$, a huge class of mathematical problems currently believed to be *intractable* (such as the *Traveling Salesman Problem*) will be solvable in polynomial time and hence be regarded as relatively easy problems to solve, assuming that the brilliant algorithm that can solve all these problems in polynomial time could be discovered. So perhaps we should feel too bad that there are limits to inductive inference, because there may be many mathematical propositions such as the $P = NP$ problem that we may never be able to solve as well. But this does not mean it is impossible to make huge strides and learn a huge amount from empirical work, just as we have been able to make great strides in mathematics despite the limits imposed by Gödel's Theorem.

3 The Value of Inference *with* Theory: Insights from other sciences

Even though there are many daunting, and sometimes insuperable limits to both deductive and inductive models inference, there is nonetheless ample room for great progress to be made. The "standard model of physics" is an example of the fundamental insights and incredibly accurate predictions that can be made by theories that are complemented by very focused data gathering and theoretically-motivated experimentation. The standard theory of physics could be described as a "parametric model" because it consists of a just a few equations with 19 unknown parameters. These parameters have been precisely estimated, and this combination of inductive and deductive inference has resulted in striking discoveries, including most recently in the confirmation that the theoretically predicted "God particle" (the Higgs boson) does indeed exist.

However economists might dismiss the physics example on the grounds that economics is a not a "hard science" — they might claim that economics is actually a *harder science* because the elementary particles in our science, human beings, are vastly more complex than the elementary particles in physics. To address this, I discuss two further examples of the power of combining inductive and deductive modes of inference by discussing examples from two other sciences that have more in common with economics: engineering and neuroscience.

The engineering example illustrates how the ability to model something successfully — even something as mundane as cars — can have very powerful, practical payoffs. Prior to the advent of finite element models and supercomputers engineers tested new car designs by crashing full scale prototypes into brick walls at 60 miles per hour. Crash dummies inside these cars were wired with sensors that recorded, millisecond by millisecond, the forces acting on the car frame and the shocks experienced by the crash dummies during crash. Over time engineers developed increasingly realistic finite element models of cars and crash dummies. This allowed them to crash cars, virtually, inside the supercomputer. Eventually, the virtual crashes began to predict crash results that were virtually indistinguishable from data generated in actual crash tests. Needless to say, it is far easier and faster to conduct virtual crash tests inside the supercomputer, and this sped up the design cycle and helped reduce the cost of producing newer, better cars.

One important thing to realize from the auto crash example is that even when models are abstract and incomplete in many respects, they can still be tremendously useful approximations

to the world. The finite element crash dummies do not have virtual hearts or virtual brains: we do not need to model their preferences over consumption and leisure, or even have accurate models that endogenously predict their last second reactions to an impending car crash. Yet these models are sufficiently good approximations for the task at hand to revolutionize the design of automobiles.

A similar approach is used to design integrated circuits: before new microchips are actually produced, virtual versions of these chips are simulated on computers to assess how these complex dynamical systems will perform if they were produced. This enables engineers to rapidly optimize the design of new computer chips. In effect, current generation computers are being used to design and simulate ever faster and more powerful next generation computers, and the result is Moore's Law, the 46% growth in computer power over time.

Think of what might be achieved if we were to devote similar resources to how we model economic agents and what might be achieved if we were able to conduct virtual "crash tests" to assess the behavioral and welfare responses to significant new economic policy changes such as the Obama Administration's signature Affordable Care Act. Instead of doing any formal modeling, policy advice comes from gurus who whisper in the President's ear. The policies are enacted with little or no pre-testing or even model-based predictions of what the consequences will be. It is sad to realize that despite all the work by the Cowles Commission, nearly six decades after the Koopmans critique and four decades after the Lucas critique economic policy making is still in the dark ages where our leaders do most policy evaluation only in the *a posteriori*. In effect, for policy changes that are too big to evaluate using randomized experiments, the government concludes there is no other alternative than to throw up its hands and use the entire American population as crash dummies to determine whether new policies will be successes or failures. I guess the American government is consigned to learning the hard way — by trial and error.[13]

---

[13]The fiasco with the launch of `healthcare.gov` shows that even the simple task of creating a reliable website to implement the new law is apparently beyond the capacity of our government and policy makers. This sort of computer work is far from "rocket science" yet over $800 million was reported to have been spent by the Federal government alone, resulting in an obviously malfunctioning website in the crucial first months of the program. A well functioning website is key to the success of the program since attracting younger, healthier and more Internet savvy enrollees is critical to keeping health premiums low. A reliable website could have been developed at a small fraction of the $800 million that was spent. Had this same amount been invested in basic research to improve economic policy making — assuming the funds were allocated in competitive manner to competent researchers and not to cronies and political insiders — one can only imagine how such a massive investment would have improved the science of economic policymaking. We can only speculate about how a more effective policy might have been formulated had there been a greater investment in structural models and interest in policy forecasting using them by the government. However there are now more than a few structural econometric studies of the Affordable Care Act, including the work of Fang and Gavazza (2011) and Aizawa (2013) who uses a structural model to numerically characterize more efficient designs for a health insurance exchange.

But learning by trial and error is not always the smartest way to learn, especially when it is costly to conduct experiments, or when the cost of making an error is very high. In fact, humans are intelligent because we *don't* always learn by trial and error. Instead for most important actions, most of us *think things through and produce internal forecasts the consequences of taking various hypothetical actions* and this helps us avoid doing some obviously dumb things. In neuroscience there is growing evidence that the human brain has an amazing innate, subconscious ability to *model and simulate reality*. Many neuroscientists believe that one of the keys to human intelligence is precisely our incredibly powerful ability to generate and modify *internal mental models of the world*. Griffiths and Tenenbaum's (2009) survey of neuroscience (GT) experiments notes that "Inducing causal relationships from observations is a classic problem in scientific inference, statistics, and machine learning. It is also a central part of human learning, and a task that people perform remarkably well given its notorious difficulties. People can learn causal structure in various settings, from diverse forms of data: observations of the co-occurrence frequencies between causes and effects, interactions between physical objects, or patterns of spatial or temporal coincidence. These different modes of learning are typically thought of as distinct psychological processes and are rarely studied together, but at heart they present the same inductive challenge — identifying the unobservable mechanisms that generate observable relations between variables, objects, or events, given only sparse and limited data." (p. 661).

GT start their survey with a wonderful example of Sir Edmund Halley's discovery of the comet now known as *Halley's comet* and his remarkable correct prediction that it would return every 76 years. This prediction was made possible by Newton's theory of physics but it required further data gathering to determine whether the comet was following an elliptical or parabolic orbit "Halley's discovery is an example of causal induction: inferring causal structure from data. Explaining this discovery requires appealing to two factors: abstract prior knowledge, in the form of a causal theory, and statistical inference. The prior knowledge that guided Halley was the mathematical theory of physics laid out by Newton. This theory identified the entities and properties relevant to understanding a physical system, formalizing notions such as velocity and acceleration, and characterized the relations that can hold among these entities. Using this theory, Halley could generate a set of hypotheses about the causal structure responsible for his astronom-

ical observations: They could have been produced by three different comets, each traveling in a parabolic orbit, or by one comet, travelling in an elliptical orbit. Choosing between these hypotheses required the use of statistical inference." (p. 661).

GT make the important observation that "People can infer causal relationships from samples too small for any statistical test to produce significant results … and solve problems like inferring hidden causal structure … that still pose a major challenge for statisticians and computer scientists." They stress the importance of prior knowledge, which "in the form of an abstract theory, generates hypotheses about the candidate causal models that can apply in a given situation." and that our ability to create internal mental models "explains how people's inferences about the structure of specific causal systems can be correct, even given very little data." (p. 662).

So it seems that millions of years of evolution has enabled human beings to develop big brains with incredibly powerful internal *but subconscious* modeling abilities. Further we conduct internal "policy evaluations" via counterfactual simulations of our mental models. Though the purpose of dreams is not entirely clear, they are perfect examples of the human capacity to conduct very convincing (at least to us, while we are dreaming) counterfactual simulated realities. Eagleman (2011) also stresses the subconscious nature of our brain's powerful internal modeling and simulation capabilities and that these models might not be perfectly accurate or complete models to constitute sufficiently good approximations to reality to give humans substantial cognitive advantages over other creatures. For example in relation to visual processing he notes that "Only slowly did it become clear that the brain doesn't actually use a 3-D model — instead, it builds up something like a $2\frac{1}{2}$-D *sketch* at best. The brain doesn't need a full model of the world because it merely needs to figure out, on the fly, where to look and when." (p. 164) He emphasizes that "The brain generally does not need to know most things; it merely knows how to go out and retrieve the data. It computes on a *need-to-know basis.*" (p. 168). As a result "we are not conscious of much of anything until we ask ourselves about it. … So not only is our perception of the world a construction that does not accurately represent the outside, but we additionally have the false impression of a full, rich picture when in fact we see only what we need to know, and no more." (p. 171)

So it seems to me that what the neuroscientists are discovering about how the human brain

works is very hopeful evidence for the eventual success of structural modeling. Neuroscience is beginning to reveal that a key reason why we are as intelligent as we are is due to our unconscious, spontaneous ability to model the world. Though our internal mental models are in many respects very incomplete, oversimplified, and inaccurate models, when combined with our ability to go out and gather data necessary to confirm or disconfirm these mental models at will — in essence our ability to combine model building with experimentation — the combined ability turns out to be incredibly powerful and may be a key to human intelligence. Our creativity in generating new models and hypotheses that explain/predict what we observe, combined with our ability to discard the poor models is very akin to the interplay between deductive and inductive modes of inference in science, where we use data and experiments both to discard bad theories and to generate new better ones.

Taking modeling from the internal, subconcious domain to the conscious, formal and symbolic domain is only relatively recent in evolutionary history. It may have begun with the advent of spoken language, then writing, and development of symbolic reasoning systems (e.g. mathematics) and modern science. The result of this has been fundamentally transformative to human evolution, in effect vastly speeding up the rate at which natural evolution occurs. The "artificial brain" — the modern digital computer or "von Neumann machine" is itself a very recent development in evolutionary history — having arisen only about six decades ago. Therefore perhaps we cannot be too hard on ourselves for being relatively clumsy at *formal modeling* and being relatively primitive in our attempts to build our first artificial brains. But the rate of change in our abilities to do computations on artificial brains is breathtakingly rapid and I noted previously that Moore's Law implies a continuous time rate of improvement in computing power of 46% per year.

4   Combining Structural Estimation and Experimentation

It should be evident from the preceding discussion that there are huge synergies between structural estimation and experimentation and the work of Todd and Wolpin (2003) who showed that their structural model can provide relatively accurate predictions of the treatment effect in the Progresa experiment is one such example. The main advantage of structural models, that they

can be used to simulate counterfactuals rapidly and cheaply, complements the main weakness of the experimental approach, which is that whether done in the field or in the laboratory, experiments are much more time and resource-intensive. At the same time, the main advantage of experiments is that they come much closer to predicting what the true impact of a policy intervention really is, and this complements the main weakness of structural models, which is that they can be wrong and produce incorrect policy forecasts. But by working together, experiments can help structural econometricians develop better models and discard inappropriate assumptions, while at the same time, the structural models can help experimentalists design more well-focused and productive experiments.

Physics provides an ideal example of the huge progress that can be achieved by effective integration between theory and experimentation. I discussed the "standard model" of physics, where theory lead experiments by over five decades. Yet it wasn't until a huge number of (very expensive) but brilliantly done experimental atomic "crash tests" that physicists were able to confirm one of the last few unconfirmed predictions of the standard model of physics: the existence of the Higgs boson. In the more mundane world of economics, Wolpin illustrates the benefits of structural policy forecasting by comparing the relative cost-effectiveness of seven alternative educational subsidy policies in Table 2.5 of his book. It would be prohibitively costly to do this comparison by running seven separate randomized experiments. Thus, credible structural econometric models seem ideally suited to *complement* experimental approaches to research by increasing the rate of return of costly investments in data gathering and randomized experimentation. Unfortunately the degree of productive cooperation between theorists and experimentalists in economics is currently far less than what we see in physics.

Instead there wholly unnecessary conflict between structural econometricians and "experimentalists" — researchers who conduct and analyze experiments run either in the lab or in the field. A caricature of the extreme experimentalist position is that theory, modeling, and knowledge of a econometric technique is unnecessary because a clever experiment can always be designed (or an historical policy change can be exploited as a "quasi experiment") to test most interesting causal hypotheses and infer policy "treatment effects." This extreme view is reflected in a survey by Angrist and Pischke (2010), whose review appears to exclude any important role

for structural econometrics in the analysis of laboratory, field, or even quasi experiments: "The econometric methods that feature most prominently in quasi-experimental studies are instrumental variables, regression discontinuity methods, and differences-in-differences-style policy analysis. These econometric methods are not new, but their use has grown and become more self-conscious and sophisticated since the 1970s." (p. 12).

In their response, Nevo and Whinston (2010) commented that "While Angrist and Pischke extol the successes of empirical work that estimates treatment effects based on actual or quasi experiments, they are much less sanguine about structural analysis and hold industrial organization (or as they put it, industrial disorganization) up as an example where progress is less dramatic. Indeed, reading their article one comes away with the impression that there is only a single way to conduct credible empirical analysis. This seems to us a very narrow and dogmatic approach to empirical work; credible analysis can come in many guises, both structural and nonstructural, and for some questions structural analysis offers important advantages." (p. 70).

In fact there has been a rather severe backlash over the last decade against the increasingly atheoretic mindset towards inference that the quote from Angrist and Pischke's survey paper epitomizes. For example, development is one of the fields of economics where the experimental approach has had the greatest impact on the way empirical research is done, but Deaton (2009) states that "Project evaluation using randomized controlled trials is unlikely to discover the elusive keys to development, nor to be the basis for a cumulative research program that might progressively lead to a better understanding of development." (p. 3). In contrast, the most recent Frisch Medal (awarded every two years by the Econometric Society to the best empirical paper published in *Econometrica* over the previous five years), went to Kaboski and Townsend (2011) for being the first study to use "a structural model to understand, predict, and evaluate the impact of an exogenous microcredit intervention program" (p. 1357).

Heckman and Urzua (2009) note that "even perfectly executed randomizations do not answer all questions of economic interest. There are important examples where structural models produce more information about preferences than experiments. A valid instrument is not guaranteed to identify parameters of economic interest when responses to choices vary among individuals, and these variations influence choices taken. Different valid instruments answer different ques-

tions. The sign of the IV estimator can be different from that of the true causal effect." (p. 2). Further, Heckman (2010) argues that many experimentalists let the limitations of their methodological approach circumscribe the types of empirical questions they can address "the parameters of interest are defined as summaries of the outputs of experimental interventions. This is more than just a metaphorical usage. Rubin and Holland argue that causal effects are defined only if an experiment can be performed. This conflation of the separate tasks of defining causality and identifying causal parameters from data is a signature feature of the program evaluation approach. It is the consequence of the absence of clearly formulated economic models." (p. 358).

But rather than attacking each other, the structural and experimental schools of econometrics have much more to gain by working together (or at least trying to coexist) rather than trying to prove one approach to inference is inherently better than the other. Heckman (2010) also takes this point of view, and it is a view that seems to be increasingly endorsed by many leading economists. In particular, the complementarity between structural econometrics and experimentation is evident in the work of El Gamal and Grether (1995) who combined detailed models of decision making and sophisticated econometric techniques to analyze data generated from a laboratory experiment on how people make inferences. As I discussed in section 3, their work provided convincing evidence that not everyone uses Bayes Rule to make decisions.

A review by Banerjee and Duflo (2008) notes that "We thus fully concur with Heckman's (1992) main point: to be interesting, experiments need to be ambitious, and need to be informed by theory. This is also, conveniently, where they are likely to be the most useful for policymakers. … It is this process of creative experimentation, where policymakers and researchers work together to think out of the box and learn from successes and failures, that is the most valuable contribution of the recent surge in experimental work in economics." (p. 30).

Overall, despite a recent history of unnecessary, counterproductive conflicts, I see encouraging signs of change and methodological cooperation. If structural econometricians and experimentalists can avoid dogmatism, methodological narrowness, and extreme debating positions, then I am optimistic that there is plenty of opportunity for very productive collaborations between economists of both persuasions. Doing this can only benefit and improve the quality of both structural econometric and experimental research.

5 Conclusion: The Limits of Econometrics

Wolpin's book does a good job of pointing out the limits to inference *without* theory. My main critique of his book is that he did not do as good a job of pointing out that there are also limits to inference *with* theory. I don't think that Wolpin would disagree that there are limits to inference, both with and without theory. But I think he would say that ruling out theory in empirical work amounts to a unnecessary *self-imposed limit.* Why do that? He's telling us that it makes no sense to arbitrarily rule out the use of theory and models when we try to make sense of the complex world we live in and I couldn't agree more. Perhaps the only thing we would want to exclude would be *bad theory,* i.e. models and theories that are not consistent with what we observe or which do not really help improve our understanding of the world.[14] Wolpin is following in the footsteps of Koopmans, Marschak, Haavelmo and other founding fathers of the Cowles Foundation in telling us that combining empirics with theory can help us produce *better theory* which helps us achieve a better understanding of our complex world. But excluding *all* theories makes about as much sense as throwing away data because they might be difficult to analyze.

It is common sense that we can make much more progress from combining inductive *and* deductive modes of inference. This progress is not only manifested in an improved understanding of the world: structural econometrics has advanced the computational methods for solving and simulating models, and in this way it has also improved our understanding of *economic theory.* Yet it is clear that the structural approach to inference remains controversial in economics fully six decades after the *Koopmans critique* and nearly three decades after the *Lucas critique.*

Unfortunately there is no overarching theory of inductive inference that we can appeal to in order to resolve this controversy. Though theoretical econometricians may give untrained observers the impression that the question of the best way to do empirical work is a completely solved problem, I agree with Heckman (1992a) that "There is no 'correct' way to pick an empirical economic model and the problems of induction, inference, and model selection are very much open." (p. 882). For the foreseeable future, empirical work will remain a relatively informal, intuitive, and subjective procedure, because we lack objective standards by which empirical

---

[14]Though even this could be regarded as a dogmatic and narrow minded attitude by theorists who like to do theory using "for theory's sake" even if the models are not realistic or do not help us improve our understanding of the world.

work can be judged.

In view of our collective ignorance even about how we, as economists, learn from and reason with data and models, it is perhaps wise to be humble and not too judgemental about the "right" way to do empirical work. Many of the dogmas prevalent in the profession and in statistics can and should be questioned. But in the act of questioning and trying to understand how to do interesting, credible and policy-relevant empirical work, we should try to avoid replacing one dogma with another one. For example while I believe that Manski very productively criticizes some empirical work for being excessively sensitive to untested assumptions, it is not productive to go overboard and demonize the act of model building, because it is generally impossible to obtain interesting, meaningful empirical conclusions if we are unwilling to make any assumptions.

Similarly, I believe Edward Leamer's critique of the practice of specification searching was an extremely valuable insight, because he showed that textbook asymptotic econometric theory could be invalid due if we fail to account for the endogenous process by which we choose a model to estimate and test, and the selection bias implicit in which model and empirical results we choose to report in published work. Yet at the same time I believe it is wrong to use Leamer's critique as a basis for demonizing the process of specification searching. Certainly specification searching calls into question the relevance of traditional asymptotic econometric theory and makes the lives of theoretical econometricians so much more messy if they want to account for the *behavior of researchers* when trying to quantify modeling and estimation uncertainty.

But it is absolute foolishness to make researchers feel guilty about specification searching just because it does not conform to the overly simplified conceptions of inference in econometric textbooks. I believe specification searching is at the heart of the creative learning process that good empirical economists use to reject bad models and discover better ones. Though Leamer analyzed specification searching from a Bayesian perspective, it does not adequately capture the creative process of model *discovery* because Bayes rule does not tell us anything about where models come from (i.e. how we come up with a particular parametric model, likelihood function, and prior distribution over its parameters).

While it is certainly right to try to question the validity of models and test them by the most rigorous available methods, I think the nihilistic view that all models are approximations that will

always be rejected given sufficient data and the conclusion that therefore none of our incorrect models are any good is also an unproductive view that does little to help science advance. Hendry and Johansen (2012) quote Haavelmo's 1989 Nobel Prize lecture which seems to indicate that he had become resigned to this view: "The basis of econometrics, the economic theories that we had been led to believe in by our forefathers, were perhaps not good enough. It is quite obvious that if the theories we build to simulate actual economic life are not sufficiently realistic, that is, if the data we get to work on in practice are not produced the way that economic theories suggest, then it is rather meaningless to confront actual observations with relations that describe something else." (p. 1). However rather than conceding defeat, Hendry and Johansen (2012) offer a more pragmatic yet very ambitious attempt to formalize the process of model discovery and selection "the earlier logic of scientific discovery in Popper (1959) (a serendipitous mis-translation of Poppers 1935 Logik der Forschung) suggests a more productive methodology, namely guiding empirical discovery by the best available theoretical ideas, but always being prepared to have those ideas rejected by sufficiently strong counter evidence." (p. 2).

Heckman (1992a) reflected the professional disdain for "structural elitism" in his caricature of the Cowles Commission's program for structural estimation: "creative empirical work along the lines of Tinbergen's early paradigm continues to be done, but it is not called rigorous econometrics any more. Many formal econometricians trained in the Cowles paradigm, or in one of its mutations, sneer at such work because it is not 'done right' i.e., within the Haavelmo-Cowles paradigm. Many serious empirical scholars no longer look to econometricians for guidance on their problems. In the current environment, cleverness, the distance between assumptions and conclusions, and the proximity of one's work to the ideas in a recent paper in the *Annals of Statistics,* are more often the measures of successful econometric work rather than its utility in organizing or explaining economic data. . . . The Haavelmo program as interpreted by the Cowles Commission scholars refocused econometrics away from the act of empirical discovery and toward a sterile program of hypothesis testing and rigid imposition of a priori theory onto the data. Exploratory empirical work on new data sets was dismissed as 'nonrigorous' or 'nonstructural' because models arising from such activity could not be justified within the Haavelmo-Cowles paradigm." (p. 883-884).

However Heckman's caricature does not describe the attitude of Wolpin or most of the leading applied structural econometricians that I know. Not only does it fail to describe how they go about their research, it strikes me that Heckman is demonizing a fictitious "straw man". I have been on the faculty at Yale and in my opinion it is one of the few departments where there is an excellent balance of methodological approaches, and a high degree of cooperation and mutual respect that are missing in many other departments, including Chicago. Given this degree of hostility, am I only further roiling the waters by choosing to focus this review on the limits of inference *with* theory?

I believe that we will start to make faster progress on an intractable debate once both sides back away from rigid polemics and acknowledge the limitations inherent in their preferred approaches to inference and empirical work. We need to be painfully aware of what these challenges and limitations are in order to have some hope of dealing with them. I believe that when difficult challenges are posed, often the best young and creative minds have an amazing ability to solve these challenges, and this is the spirit I have offered this review. I also want to make clear that though I disagree with some of Heckman's views, I believe he has made fundamental contributions by bridging the gap and finding solutions to the problems confronting structural estimation. Overall I feel my own general "philosophy of science" is very close to Heckman's and I have been incredibly influenced by his vastly deeper appreciation of these difficult issues.

The influential work by Mary Morgan on models in economics suggests that anyone who has a good understanding of the history of science would agree that Wolpin's ideas are uncontroversial and even almost obviously correct. Denying any role for theory in inference is an untenable, indefensible position, or as Heckman and Urzua (2009) state it, "No one trained in economics can doubt the value of credible, explicit economic models in interpreting economic data." (p. 2). Unfortunately the indefensible position that doubts and denies the role of theory in inference still holds great sway in the economics profession fully six decades after the *Koopmans critique.*

While I am critical of those who dismiss the use of theory in empirical work as a *basic philosophy of science,* when it comes to *individual researchers* I have absolutely no problem with the idea that many economists do best by specializing and exploiting their comparative advantage. This may mean that many economists are functionally specialized as pure experimentalists,

and others as pure theorists. Perhaps only relatively few economists will try to master both. In my comments on Keane's article (Rust, 2009), I noted that there seem to be two flavors of economists, *statistical modelers* and *economic modelers.* I think it would be equally indefensible to claim there is any one "right" way to go about modeling things.

If you look closely at economists who are skeptical of the value of economic modeling such as Charles Manski or Joshua Angrist, you will see that they are actually masters of statistical modeling and their incredible success in the profession owes at least partly to their success in this style of modeling and reasoning. Wolpin is obviously equally successful and influential as a master of economic modeling. Though we call the type of empirical work that Wolpin does "structural modeling" and neither Manski or Angrist would probably describe their empirical work as "structural" I think we could agree that all three of them are using models of some sort and ultimately the test of which approach is "best" will be determined by which of their models provides the most insight and understanding, and is most useful for policy making. I think the jury is still out as to whether statistical models or economic models will prove to be more useful and insightful. But if we can at least agree that there is a benefit to using some type of model, perhaps we are making progress.

To those remaining skeptics and haters or structural modeling, the main message of Wolpin's book is clear: be not fearful of the unknown, but go boldly into that brave new world — or at least, try not to stand in the way of progress.

References

[1] Abhijit V. Banerjee and Esther Duflo (2009) "The Experimental Approach to Development Economics" *Annual Review of Economics* **1** 151–178.

[2] Aguirregabiria, A. and P. Mira (2013) "Identification of Games of Incomplete Information with Multiple Equilibria and Common Unobserved Heterogeneity" manuscript, Department of Economics, University of Toronto.

[3] Aizawa, Naoki (2013) "Health Insurance Exchange Design in an Empirical Equilibrium Labor Market Model" manuscript, University of Pennsylvania.

[4] Angrist, J.D. and J. Pischke (2010) "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics" *Journal of Economic Perspectives* **24-2** 3–30.

[5] Barron, A. (1994) "Approximation and Estimation Bounds for Artificial Neural Networks" *Machine Learning* **14** 115–133.

[6] Benitez-Silva, H. Buchinsky, M. and J. Rust (2003) "Induced Entry Effects of a $1 for $2 offset in SSDI Benefits" manuscript, Gergetown University.

[7] Cho, S. and J. Rust (2010) "The Flat Rental Puzzle" *Review of Economics Studies* **77** 560–594.

[8] Chow, C.S. and J.N. Tsitsiklis (1989) "The Complexity of Dynamic Programming" *Journal of Complexity* **5** 466–488.

[9] Colander, D. Follmer, H. Haas, A. Goldberg, M.D. Juselius, K. Kirman, A. Lux, T. and B. Sloth (2009) "The Financial Crisis and the Systemic Failure of Academic Economics" University of Copenhagen Department of Economics Discussion Paper No. 09-03.

[10] Cunha, Flavio and James J. Heckman (2007) "The Technology of Skill Formation" *American Economic Review* **97-2** 31–47.

[11] Deaton, A.S. (2009) "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development" NBER Working Paper 14690.

[12] Eagleman, D. A. (2011) *Incognito: The Secret Lives of the Brain* Pantheon Books, New York.

[13] Einhorn, H.J. and R. M. Hogarth (1986) "Decision Making Under Ambiguity" *Journal of Business* **59-4** S225–S250.

[14] El-Gamal, Mahmoud and David Grether (1995) "Are People Bayesian? Uncovering Behavioral Strategies" *Journal of the American Statistical Association* **90-423** 1137–1145.

[15] Fang, H. and D. Silverman (2009) "Time Inconsistency and Welfare Program Participation: Evidence From the NLSY" *International Economic Review* **50-4** 1043–1076.

[16] Fang, H., and A. Gavazza (2011) "Dynamic Ineciencies in an Employer-Provided Health Insurance System: Theory and Evidence" *American Economic Review* **101-7** 3047–3077.

[17] Feigenbaum, E. A., and H.A. Simon (1984) "EPAM-like models of recognition and learning" *Cognitive Science* **8** 305–336.

[18] Frijters, P. (2013) "Review of the *Limits of Inference without Theory* by Kenneth I. Wolpin" *Economic Record* 429–432.

[19] Geweke, J. (2014) "Review of Manski's 'Public Policy in an Uncertain World'" forthcoming, *Journal of Economic Literature.*

[20] Gilboa, I, Postlewaite, A., Samuelson, L. and D. Schmeidler (2013b) "Economic Models as Analogies" *Economic Journal* forthcoming.

[21] Gilboa, I, Samuelson, L. and D. Schmeidler (2013a) "Dynamics of inductive inference in a unified framework" *Journal of Economic Theory* **148** 1399–1432.

[22] Gilboa, I. and D. Schmeidler (1989) "Maxmin Expected Utility with a Non-Unique Prior" *Journal of Mathematical Economics* **18** 141–153.

[23] Gödel, Kurt (1931) "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I" *Monatshefte fř Mathematik und Physik* (see also the english translation "On formally undecidable propositions of Principia Mathematica and related systems I" in Solomon Feferman, (ed) (1986) *Kurt Gödel Collected works, Vol. I*. Oxford University Press: 144–195.

[24] Griffiths, T.L. and J. B. Tenebaum (2009) "Theory-based Causal Induction" *Psychological Review* **116-4** 661–716.

[25] Hansen, L. and T.J. Sargent (2008) *Robustness* Princeton University Press.

[26] Heckman, James J. (1992a) "Randomization and social policy evaluation" in Charles Manski and I. Garfinkel (eds.) *Evaluating Welfare and Training Programs* Cambridge, MA: Harvard University Press.

[27] Heckman, James J. (1992b) "Haavelmo and the Birth of Modern Econometrics: A Review of the History of Econometric Ideas by Mary Morgan" *Journal of Economic Literature* **30-2** 876–886.

[28] Heckman, James J. (2007) "Haavelmo's Legacy" manuscript, University of Chicago.

[29] Heckman, J. J. and S. Navarro (2006) "Dynamic discrete choice and dynamic treatment effects" *Journal of Econometrics/* **136** 341–396.

[30] Heckman, J. J. and Rodrigo Pinto (2013) "Causal Analysis After Haavelmo" manuscript, Department of Economics, University of Chicago.

[31] Heckman, James J. and Sergio Urzua (2009) "Comparing IV With Structural Models: What Simple IV Can and Cannot Identify" NBER Working Paper 14706.

[32] Heckman, James J. and Edward Vytlacil (2007a) "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models, and Econometric Policy Evaluation" Chapter 70 in James J. Heckman and Edward E. Leamer (eds.) *Handbook of Econometrics, Volume 6B* Elsevier, North Holland.

[33] Heckman, James J. and Edward Vytlacil (2007b) "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments" Chapter 71 in James J. Heckman and Edward E. Leamer (eds.) *Handbook of Econometrics, Volume 6B* Elsevier, North Holland.

[34] Hendry, D. and S. Johansen (2012) "Model Discovery and Tyrgve Haavelmo's Legacy" University of Oxford Discussion Paper 598.

[35] Hurwicz, L. (1962) "On the Structural Form of Interdependent Systems" in E. Nagel *et. al.* (eds) *Logic, Methodology, and the Philosophy of Science* Stanford University Press.

[36] Imbens, G. (2009) "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)" NBER working paper 14896.

[37] Imbens, G. and J. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects" *Econometrica* **62** 467–475.

[38] Iskhakov, F. Schjerning, B. and J. Rust (2013) "Recursive Lexicographical Search: Finding All Markov Perfect Equilibria of Finite State Dynamic Directional Games" manuscript, University of Copenhagen.

[39] Kaboski, J.P. and R.M. Townsend (2011) "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative" *Econometrica* **79-5** 1357–1406.

[40] Keane, M. and K.I. Wolpin (1997) "The Career Decisions of Young Men" *Journal of Political Economy* **105-3** 473–522.

[41] Keane, M. and K. I. Wolpin (2010) "The role of labor and marriage markets, preference heterogeneity, the the welfare system in the life cycle decisions of black, Hispanic and white women" *International Economic Review* **51-3** 851–892.

[42] Klemperer, P. (2004) "What Really Matters in Auction Design" chapter 3 of *Auctions: Theory and Practice* Princeton University Press.

[43] Klibanoff, P. Marinacci, M. and S. Mukerji (2005) "A Smooth Model of Decision Making under Ambiguity" *Econometrica* **73-6** 1849–1892.

[44] Koopmans, T. C. (1947) "Measurement without Theory" *Review of Economics and Statistics* **29-3** 161–172.

[45] Koopmans, T. C. and A. F. Bausch (1959) "Selected Topics in Economics involving Mathematical Reasoning" *SIAM Review* **1** 138–148.

[46] Kurzweil, Ray (1999) *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* Viking Press.

[47] Kurzweil, Ray (2005) *The Singularity is Near* Penguin Books, New York.

[48] LaLonde, Robert J. (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data" *American Economic Review* **76-4** 604–620.

[49] Leamer, E. (1982) *Specification Searches: Ad Hoc Inference with Non Experimental Data* Wiley, New York.

[50] Leamer, E. (1983) "Let's Take the Con Out of Econometrics" *American Economic Review* **73-1** 31–43.

[51] Ledyard, John O. (1986) "On the Scope of the Hypothesis of Bayesian Equilibrium" *Journal of Economic Theory* **39-1** 59–82.

[52] Lewis, M. (2003) *Moneyball* W.W. Norton and Company.

[53] Lucas, R.E. Jr. (1976) "Econometric Policy Evaluation: A Critique" *Carnegie-Rochester Conference Series on Public Policy* **1** 19–46.

[54] Lumsdaine, R. Stock, J.H. and D. A. Wise (1992) "Pension plan provisions and retirement: Men and women, Medicare, and Models" in D. A. Wise (ed.) *Studies in the Economics of Aging* 183–220. University of Chicago Press.

[55] Magnac, T. and D. Thesmar (2002) "Identifying Dynamic Discrete Decision Processes" *Econometrica* **70-2** 801–816.

[56] Manski, C. F. (2013a) *Public Policy in an Uncertain World: Analysis and Decisions* Harvard University Press.

[57] Manski, C. F. (2013b) "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy" forthcoming, *Quantitative Economics.*

[58] Moffitt, Robert A. (1983) "An Economic Model of Welfare Stigma" *American Economic Review* **73-5** 1023–1035.

[59] Morgan, Mary S. (1991) *The History of Econometric Ideas* Cambridge University Press.

[60] Morgan, Mary S. (2012) *The World in the Model How Economists Work and Think* Cambridge University Press.

[61] Matzkin, R. (1991) "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models" *Econometrica* **59-5** 1315–1327.

[62] Misra, S. and H. S. Nair (2011) "A structural model of sales-force compensation dynamics: Estimation and field implementation" *Quantitative Marketing and Economics* **9** 211–257.

[63] Mohri, M. Rostamizadeh, A. and A. Talwalkar (2012) *Foundations of Machine Learning* The MIT Press.

[64] Nevo, A. and M. D. Whinston (2010) "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference" *Journal of Economic Perspectives* **24-2** 69–82.

[65] Newell, Allen and Herbert A. Simon (1972) *Human Problem Solving.*

[66] Rust, J. (1994) "Structural Estimation of Markov Decision Processes" in R. Engel and D. McFadden (eds.) *Handbook of Econometrics* volume 4, Amsterdam, North Holland, 3081–3083.

[67] Rust, J. (1997) "Using Randomization to Break the Curse of Dimensionality" *Econometrica* **65-3** 487–516.

[68] Rust, J. (2009) "Comments on Michael Keane's 'Structural vs Atheoretic Approaches to Econometrics'" *Journal of Econometrics*

[69] Rust, J. and C. Phelan (1997) "How Social Security and Medicare Affect Retirement Behavior in a World with Incomplete Markets" *Econometrica* **65-4** 781–832.

[70] Rust, J. Traub, J.F. and H. Woźniakowski (2003) "Is There a Curse of Dimensionality for Contraction Fixed Points in the Worst Case?" *Econometrica* **70-1** 285–329.

[71] Schorfheide, F. and Wolpin, K.I. (2013) "To Hold Out or Not to Hold Out"

[72] Silver, N. (2012) *The Signal and the Noise* Penguin Press.

[73] Sims, C. A. (1980) "Macroeconomics and Reality" *Econometrica* **48-1** 1–48.

[74] Solomonoff, R. J. (1964) "A Formal Theory of Inductive Inference: Part I" *Information and Control* **7** 1–22.

[75] Stone, C.J. (1980) "Optimal Rates of Convergence for Nonparametric Estimators" *Annals of Statistics* **8-6** 1348-1360.

[76] Sugden, R. (2000) "Credible Worlds: The Status of Theoretical Models in Economics" *Journal of Economic Methodology* **7-1** 1–31.

[77] Todd, P. and K.I. Wolpin (2006) "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility" *American Economic Review* **96-5** 1384–1417.

[78] Todd, P. and K. I. Wolpin (2013) "Estimating a Coordination Game in the Classroom" manuscript, Department of Economics, University of Pennsylvania.

[79] Vapnik, V. A. (1998) *Statistical Learning Theory* Wiley, New York.

[80] Weintraub, G. Y. Benkard, L.C. and B. van Roy (2008) "Markov Perfect Industry Dynamics with Many Firms" *Econometrica* **76-6** 1375–1411.

[81] Wolpin, Kenneth I. (2013) *The Limits of Inference without Theory* MIT Press, Cambridge.