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

John Rust
Georgetown University

December, 2013

† Preliminary: This version of this review is not for general distribution and quotation.

‡ Department of Economics, Georgetown University, Washington, DC 20550, email: `jr1393@georgetown.edu`

1 Introduction

Ken 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. Ken's 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 Tom 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 figures such as Koopmans, Marschak, Haavelmo and others in the early days of the Cowles Foundation at Yale. In fact the very motto and *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 and honors a famous essay by Koopmans, "Measurement without Theory" (1947) that commented on the 1946 book by Burns and Mitchell, *Measuring Business Cycles.* Koopmans criticized their "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? Wolpin's monograph does an admirable job of illustrating the benefits of using theories and models to guide empirical research, but he does not explain with equal vigor why structural econometrics and the overall mission of the Cowles Foundation should still so be controversial more than six decades after the "Koopmans critique".[1]

[1] According to Wikipedia, Koopmans convinced the Cowles family to move it from Chicago to Yale in "response to rising hostile opposition to the Cowles Commission by the department of economics at University of Chicago during the 1950s".

To better understand what all the controversy is about, I refer readers instead to 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) provides a very vigorous and skeptical counterpoint to Wolpin's book. The difference in outlook can be summarized in the title of Manski's first chapter "Policy Analysis with Incredible Certitude" which provides a dramatic contrast to Wolpin's chapter titled "*Ex Ante* Policy Evaluation: The Role of Theory".

My own philosophy of science is much closer to that of Ken Wolpin and the Cowles Foundation than the agnostic, skeptical mindset of Chuck Manski. However Manski raises serious, well considered, and legitimate concerns about the potential downsides and limitations of the structural, model/theory driven approach to empirical work. Wolpin acknowledges these concerns on page 2 of his book "In one view, theory is seen as unnecessary or even detrimental to inferential data analysis." Ideally it should be a two way street where data are used to develop and test theory (leading to better theories and rejection of bad ones) just as theory guides the variables we measure, the econometric methods we use, and how we interpret our empirical findings.

Manski's work reflects his long standing concern that *a priori* assumptions and theories (and overly simplified models) have had an excessive impact on economic science, and that unjustified assumptions have often dominated, distorted, distracted, and sometimes completely mislead us. Whether we have been deliberately or unintentionally mislead by bad theories, Manski levels well targeted criticisms of researchers who use the cloak of science to mask the true uncertainty in their conclusions and predictions to the public and policy makers — the ultimate consumers of our scientific findings. As he notes in the Introduction to his book "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).

I believe that the *Manski critique* has validity, since in my own personal experience I have encountered well-intentioned practitioners of structural econometrics who are overly enamored with their theories/models, and they sometimes seem unable to appreciate the distinction between the predictions of their models and reality. As Ken notes, there is a danger that structural researchers may fail to "let the data speak for themselves" (p. 2), or worse, they may intentionally fail to report evidence that is inconsistent with the predictions of their favorite theory/model.

On the other hand, it is easier to criticize, question, and be skeptical/agnostic of the role of theory and assumptions in empirical work. It is much harder to confront these challenges, take a stand before incredulous and sometimes hostile anonymous (and non-anonymous) reviewers, and publish work that has a practical impact. While it is certainly fair to question the validity of the *assumptions* that Ken and his coauthors have made in their work, I have absolutely never questioned Ken's honesty or integrity as a scientist and a human being. Thus, I was quite frankly taken aback by the charges leveled by Paul Frijters (2013) in his review of Ken's book, which harshly criticized the seminal Keane and Wolpin (1997) paper and their subsequent (2010) paper. Frijters attacked what he interpreted as a key finding of both of these studies, "Why does the Keane and Wolpin (1997) model invariably tell you life is largely predetermined at conception (fixed types)? Because, mainly for analytical convenience, they assume that there are no persistent shocks later in life. Where does that particular rabbit go into the hat? It is in the assumption of independent identically distributed shocks every period. That technical assumption, coupled with persistence in the data, is what forces the users of this model to, time and again, conclude that life was largely predetermined at the first moment you came into the data." (p. 431).

If Frijters had bothered to even casually read Keane and Wolpin (1997) he would have realized that what he wrote above is completely wrong. There are *many* sources of persistence in outcomes in Keane and Wolpin's model, and many of them are via observable "state variables" included in their models such as measures of human capital. How Frijters managed to misinterpret their models as implying that a person's life is "largely predetermined at conception" can only be evidence that he did not carefully read Wolpin's book and/or he fails to understand his models. In fact, the conclusion of Keane and Wolpin (1997) takes pains to state that "It is important to consider carefully the exact meaning of this finding. First, it does not mean that lifetime utility is for the most part predestined regardless of one's behavior. Second, it does not mean that most of the welfare variation is genetically determined through exogenous endowments, so that inequality is intractable and cannot be significantly altered by policy." (p. 515).

Frijters did correctly quote Wolpin's summary of one of the conclusions of his 2010 paper with Keane, namely that "unobserved heterogeneity is the most important of the initial conditions in accounting for the variance of behaviors" (p. 97). However if Frijters would have bothered to read and quote a following sentence, he would have seen that Keane and Wolpin found a substantial role for time-invariant unobserved heterogeneity, but *at age 14, not at birth!* "Whatever the process by which these unmeasured preferences and endowments are formed by age 14, they are critical in determining completed schooling levels." (p. 97). This conclusion is entirely consistent with findings by James J. Heckman and many others using entirely different data and a different methodological approach. For example Cunha and Heckman (2007) state "It is well documented that people have diverse abilities, that these abilities account for a substantial portion of the variation across people in socioeconomic success, and that persistent and substantial ability gaps across children from various socioeconomic groups emerge before they start school." (p. 31)

Perhaps some of Frijters' confused and distorted reading of Wolpin's work is forgivable, but the number of patently wrong assertions in his review makes me wonder if he if even read Wolpin's book, much less the substantial research that it summarizes. But Frijters crossed the line when he impugned Wolpin's honesty and academic integrity by suggesting that Keane and Wolpin had intentionally distorted and provided a misleading interpretation of their empirical findings, as if they had some agenda and were, for some incomprehensible reason, plotting to take advantage of poorly trained policy makers to obfuscate their true findings and give them bad advice.

> "Now, I must confess that I have always found it intellectually disconcerting of Keane and Wolpin to not mention the importance of that assumption whenever they tell policy makers and others about their findings, and am bothered by seeing it again in this book. You see, almost no-one they talk to is as versed in their models as themselves, and hence if they don't religiously remind their audience about the technical rabbits they have put into their hats to get their amazing results, their audience's only choice is to take it or leave it on faith. And it is not good enough to say that criticizing their own model is a job for others because the few that understand how these models really work are usually co-authors or PhD students of the originators of these models and hence have a vested interest." (p. 431)

So let me make a clear distinction between the Manski critique which is valid, principled, and reflective of a deep understanding of the limits and problems of inference, and the unfounded allegations of Frijters which border on *ad hominem* attacks that are completely beyond the pale. I took the space to address the blatant ignorance and hostility evident in Fritjer's review to provide a concrete illustration of some of the unnecessary obstacles and criticisms that structural econometricians such as Wolpin are forced to deal with in this profession. Though Manski may have a legitimate difference of opinion about the validity of the *assumptions* Wolpin is willing to make in his empirical work, I cannot imagine that Manski would ever question Wolpin's motivations or professional integrity, or even include Wolpin in the group of researchers "who regularly assert certitude about the consequences of alternative decisions."

My main reservation about Manski's approach and the growing literature on *partial identification* that his work has spawned, is that it can err in the other direction by exaggerating the degree of incertitude in empirical work. His extreme aversion to assumptions (and apparently to the empirical application of economic theory) greatly weakens the empirical conclusions he can reach, and as a consequence, so too the policy implications that can be drawn from his style of empirical research. Manski advocates the use of only the most minimal possible assumptions to produce only *bounds* on values of "parameters of interest" but these bounds are often sufficiently wide that the only conclusion that emerges from many of these types of empirical studies is the helpless conclusion that we can conclude very little. These sorts of conclusions are often not too helpful to real world decision makers who do ultimately have to take a stand, take a chance, and make a decision based on limited information and considerable unresolved uncertainty. Occasionally they may even listen to the partially-informed guesses of economists who are willing to make assumptions when necessary in order to provide some meaningful advice.

For example, Manski (2013b) analyzes a static model of labor supply to address a classic question, "how do taxes affect labor supply?". Manski's main conclusion from his empirical analysis is that "Considering the classical static model, Section 2 showed that basic revealed preference analysis has little power to predict labor supply under proposed policies. Importantly, it does not predict whether increasing tax rates reduces or increases work effort." (p. 32). On the other hand Wolpin introduces a non-parametric matching estimator of the impact of taxes on mean hours worked in chapter 2 of his book. He notes that his estimator "illustrates the feasibility of performing *ex ante* policy evalauation without having to introduce functional form assumptions." though he also notes that "Although nonparametric, the method is not assumption-free. Indeed, the method requires an explicit characterization of a behavioral model and a number of key assumptions." (p. 12). Though it is hard to assess the relative strength of the assumptions imposed by Wolpin and Manski, it seems to me that Wolpin is able to get far more mileage from the weak assumptions he imposes than Manski was able to obtain in his analysis. At times, it seems as if Manski's goal is to show that researchers can conclude very little about any given question, whereas Wolpin's goal is to show that, actually yes we can.[2]

Even though assumptions must typically be imposed to obtain meaningful conclusions, it does not follow that all assumptions are "empirically unfounded". In fact many assumptions that are imposed in structural modeling are motivated from personal experience, or introspection, and thus have at least some empirical basis. Assumptions that are not easily justified can sometimes be varied in more or less systematic ways to judge the senstivity of the empirical findings and model predictions to assumptions that different reasonable individuals might disagree with. Thus, while it is certainly fair game to *question* assumptions made in empirical work, I would be wary of *demonizing* the act of making assumptions, especially when assumptions are clearly necessary to obtain meaningful results. As long as researchers are not consciously misrepresenting their assumptions, and make an honest and transparent attempt to assess the sensitivity of their key empirical conclusions to various modeling assumptions, I would have to say that I am far less worried about the ill effects of assumptions on empirical research than Manski appears to be.

Though there are few equations in Manski's book (compared to many in Wolpin's book), one of the equations from Manski's book that I like the best is this one

$$assumptions + data \longrightarrow conclusions \qquad (1)$$

because it elegantly illustrates the nature of the problem of inference. However when I showed this equation to a colleague, he suggested that the equation really ought to be of the form

$$assumptions * data \longrightarrow conclusions \qquad (2)$$

[2] Wolpin admits that he is not aware of any empirical applications of the nonparametric matching estimator he proposed to the question of measuring the effect of taxes on labor supply. I should note that I have severe reservations about the wisdom of trying to measure the effect of taxes on labor supply using the static textbook model of labor supply. There are numerous questionable and very strong implicit assumptions involved in using a static framework to answer questions that in my opinion can only be credibly analyzed in a dynamic context. In addition, the model assumes that workers are paid an hourly wage and they have full choice over their hours of work. For many, if not most, this is a very poor assumption. I could go on and list many other unrealistic aspects of the static textbook model of labor supply that in my opinion makes it completely unsuitable for use as a credible framework for assessing the affect of taxes on labor supply. To be charitable to Wolpin, I regard his discussion of this model in chapter 2 of his book to be a pedagogical illustration and not a serious recommendation about how anyone should actually go about studying this question. For an empirical analysis based on a dynamic structural model (though from an aggregate, representative consumer persective) see Prescott (2004). Prescott finds that "I have estimated the elasticity of labor supply and have found it to be large, nearly 3 when the fraction of time allocated to the market is in the neighborhood of the current U.S. level." and that "that virtually all the large differences between the U.S. labor supply and those of Germany and France are due to differences in tax systems." (p. 8). An elasticity of this size is "off the charts" compared to the range of elasticities estimated in microeconometric studies, see e.g. Chetty *et. al.* (2011). This huge level of disagreement about the labor supply elasticity between various micro and macro studies may in fact be consistent with the agnostic, skeptical viewpoint of Manski.

(i.e. change the addition operator to multiplication) since the latter variant of the "Manski equation" reflects the fact that without any assumptions we can't reach any conclusions, whereas Manski's version will. Equation (2) reflects a more powerful interaction between data and assumptions than Manski's equation (1) allows, and in particular (2) schematically illustrates a more powerful role for data in enabling us to weaken the number assumptions necessary to reach the same conclusions. The question is whether it is possible to reconcile or move to some level of compromise in the diametrically opposite viewpoints of Wolpin and Manski. I believe it is possible, and it is possible for researchers to convey more honestly and accurately the extent to which assumptions do drive or affect their conclusions. I believe Wolpin has attempted to do some of this in his book and in his research, though perhaps not in a sufficiently formal or comprehensive way to satisfy a critic such as Manski.

Another counterpoint to Wolpin's book — an alternative approach that Ken does an admirable job of addressing in his book and in his research — is the view that can be loosely attributed to literature on "treatment effects" and randomized experimentation (see, e.g. Imbens and Angrist (1994) and Banerjee and Duflo (2009)) that an integration of theory and empirical work is unnecessary because the main interesting questions in economics involve inferring/estimating an average treatment effect (e.g. the average effect of some policy policy change on some outcome of interest) via randomized controlled experiments (RCEs, or which Wolpin abbreviates as RCTs, reflecting the epedemiological term randomized controlled trials). The use of RCEs and the treatment effects perspective for the analysis of policy changes has had a profound impact on empirical work in economics, especially in development economics. As Wolpin notes "The absence of theory in inferential empirical work is pervasive. For example, of all the papers in the January 2008 maiden issue of the new American Economics Association journal *Applied Economics*, all of which were inferential, none contained an explicit model of 'man's economic behavior"' (p. 2).

In the remainder of this review, I would like to address my own view about the limits to inference that were not adequately expressed in Wolpin's books. Then I will discuss the role of experimentation, and how it can *complement* (as opposed to replace) the enterprise of structural modeling and inference. I then provide some concluding remarks. My main message is though there is ample room for getting far more knowledge from limited data (and even more when we have access to "big data") by optimally combining inductive and deductive approaches to inference and learning, it is important to recognize that there are a number of inherent *limits to inference* that may be insuperable. These limits were not adequately addressed in Wolpin's book, and motivated the title of this review.

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 structural method of inference that uses a model. 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.[3]

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 pretty 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 model in his lexicon, unless the regression model were somehow derived from a deeper theory, rather than simply posited as a relationship between a dependent variable y and some vector of independent variables x. For example Wolpin might classify a linear demand curve as a "model" if it had been derived from a utiity function via an application of Roy's Identity, or if a linear demand curve is assumed as a "primitive" but the analyst derives the implied indirect utility function via application of duality theory (e.g. Hausman, 1981).

But by failing to give a sufficiently clear and precise definition, Wolpin leaves himself open to criticism that he has a very narrow view of what "model" is. For example Frijters (2013) noted in his review of Wolpin's book, "From his examples, it is clear that what Ken 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

[3]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.

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, Ken 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 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 want to restrict attention to 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). Lucas 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. Not only did it largely uncut the credibility of the practitioners of these large scale forecasting models, it also provided an important impetus for the developments of both structural macro and microeconometric methods. The first dynamic versions of these models appeared in the late 1970s, shortly after Lucas's paper was published.

Looking back more than three 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 successful use of structural models for policy evaluation, there are still relatively few clearcut successes where structural models have had a clearly measurable positive *practical* impact on 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 in Mexico. Wolpin also cites the work of Lumsdaine, Stock and Wise (1992) who showed that structural retirement models provided much more accurate forecasts of changes in a company's retirement plan (the adoption of a temporary "retirement window" incentive plan) than reduced form models.

Besides the examples that Wolpin discussed, I might mention other examples in industrial organization, including Cho and Rust (2010) where we used a dynamic structural model to analyze the car replacement decisions of a large rental car company. We found that the company had adopted a suboptimal replacement policy and our model predicted the company could significantly increase its profits by keeping its rental cars longer and providing discounts to consumers to induce them to rent the older rental cars in its fleet. These forecasts convinced the company to undertake a controlled experiment to test the predictions of their model, and the results of the controlled experiment validated the predictions of the econometric model.

Another example 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 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 after they had reached the commission ceiling. "For instance, in a salary + commission scheme such as ours, sales-agents who achieve the quota required for earning the commission in the current compensation cycle may have a perverse incentive to postpone additional effort to the future." (p. 213). 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 these examples can be regarded as practical successes that validate the the structural approach to estimation and policy evaluation that Lucas envisioned in his 1976 paper. But three decades after the Lucas critique, the structural estimation industry would be in a much stronger position if there were a considerably larger number of clear successes than the handful discussed above.

Structural econometricians confront a number of challenges that have made faster progress in this area very difficult. Some of these challenges are actually fundamental limits to inference. Wolpin's book does not adequately discuss these challenges and inherent limits, and for young people who are considering whether to do structural econometrics, it is important to have a clearer picture of what the risks, rewards, and limits are. Ken has obviously been a very successful adviser, and has 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 will never be able to provide satisfactory answers, 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 requires an important *assumption* that there "deep policy invariant objects/parameters" of any system that can be recovered via structural methods of inference, and that once these policy invariant are inferred it is possible to predict how the system 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, and 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."

In my own work, I have found in a number of circumstances that certain parameters representing "stigma" are required in order to enable a structural model to fit the data. For example in my studies of the decision to apply for disability benefits with Benitez-Silva and Buchinsky (2003), we find puzzlingly low take up rates for disability benefits by low income 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, I find this level of ignorance to be implausible. So an alternative way to explain the low take up rate is to include parameters that reflect disutility or stigma for being on the disability rolls. However these stigma parameters do not seem to be policy-invariant preference parameters. It appears that 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.[4]

Wolpin acknowledges that stigma parameters have played a role in his own work on welfare participa-

[4] 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, *Aid for Families with Dependent Children* (AFDC) that is 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).

tion, 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 offers very little discussion or evidence of concern that some or all of these parameters may not be structural, i.e. invariant to policy. If they are not policy-invariant, then there is an important unresolved question as to how we would predict how policy changes will affect these policy-dependent utility parameters. Wolpin simply 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). This comment seems to suggest that he views stigma as relatively unimportant.

I disagree. 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 the large scale macromodeling industry in the late 1970s (i.e. it basically destroyed it).

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 computating power, the degree of precision in the inferences we can make and the size of the economic models we can solve will be limited. Though we will be able to say more with more data and greater computer power, the quality/reliability of the conclusions we can reach using more data and bigger computers to solve and estimate ever more complex and realistic models will grow far more slowly than the (exponential) rate of growth in data and computer power.

It is sometimes possible to break the curse of dimensionality though not without a cost. Rust (1997) introduced a random multigrid algorithm that can solve discrete choice dynamic programming problems in polynomial time, but at the cost of using a randomized algorithm that results in a solution with stochastic error (though the error can be made arbitrarily small by increasing the number of random draws used in this algorithm).[5] Barron (1989) showed that the curse of dimensionality of non-parametric regression can be broken for certain classes of multivariate functions that have "special structure" but doing this requires finding a global minimum of a nonlinear least squares problem, and the time to find this global minimum can increase exponentially fast as the number of variables increase.

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.

2.3 The Identification Problem

The most daunting limit to knowledge that structural econometricians face is the *identifcation problem* which is the problem of trying to infer the *structure* — set of underlying *primitives* that imply a probability distribution for the observable variables. 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 outside the domain of the identification analysis (i.e. they are treated as assumptions that cannot be altered, tested, or questioned). But it may not always be possible to infer the underlying structure, even with very strong maintained assumptions and unlimited data.

For example in single agent dynamic programming models a commonly imposed maintained hypothesis is that agents are expected utility maximizers, have rational expectations, and seek to maximize a discounted sum of utility (i.e. they have a time separable utility function). The *structure* under this main-

[5] 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.

tained hypothesis consists of the objects $\{\beta, u, p, q\}$ where β is the agent's discount factor, u is the the agent's utility function, p is a Markov transition probability kernel representing the agent's beliefs, and q is a transition probability for *unobservables* (variables that the agent observes and affect the agent's utility but are not observed by the econometrician).

The observables in this model are the observed state of the agent x, and the observed decision d. In addition most structural econometric model allow for *unobserved state variables* ε as well, and additional maintained assumptions must be imposed concerning how observed state and decision variables relate to the unobserved state variables in the model. A commonly imposed additional set of maintained assumptions are *additive separability* (i.e. the unobserved state variables enter the utility function u in an additive form), and *conditional independence* (the unobserved state variables are *IID* shocks that do not directly affect the probability distribution of observed state variables directly but only indirectly through the effect the unobservables have on the contemporaneous decision.

The *reduced form* corresponding to the structure $\{\beta, u, p, q\}$ consists of the implied probabilty distribution (or stochastic process in a dynamic model) of the observed state and decision variables (x, d). In a *dynamic discrete choice model* (which are the type of models that Wolpin and others, including myself, typically work with), the reduced form corresponds to the *conditional choice probability* $P(d|x)$, which provides the probability that an agent in observed state x will choose discrete alternative d. It is the probability the agent, whose choice is given by an *optimal decision rule* $d = \delta(x, \varepsilon)$, will find it optimal to choose alternative d in state x, after integrating out the effect of the unobserved states ε.

Dynamic programming implies that the decision rule δ and the choice probability $P(d|x)$ are implicit functions of the underlying structure $\{\beta, u, p, q\}$. Thus, the content of the theory can be expressed as a mapping $P = \Lambda(\beta, u, p, q)$ where P is the conditional choice probability. With enough data P can be estimated non-parametrically, that is, without imposing any of the maintained assumptions discussed above. The structure is *nonparametrically identified* if we can invert the mapping Λ to uniquely uncover the true underlying structure $\{\beta, u, p, q\}$ from the reduced form (choice probability P). Of course, since a positive affine transformation of utility has no effect on the decision rule of an expected utility maximizer, we cannot expect to identify the utility function uniquely but only up to an equivalence class consisting of all positive affine transformations of a given "true" utility function u^*.

Unfortunately, Rust (1994) and Magnac and Thesmar (2002) proved that this structure 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 ε can affect the observed state variables x. These very strong maintained assumptions automatically imply that the p component of the structure is nonparametrically identified, i.e. under rational expectations we can estimate an agent's beliefs over the observed state parameters p using only the observed data (x,d). This means that the unknown elements of the structure are reduced to (β,u,q).

However even if we were to fix q (say to assume it is a Type 3 Extreme value distribution, which implies that $P(d|x)$ takes the form of a *multinomial logit model*) it is still not possible to identify preferences u up to positive affine transformation of a single utility function u, nor is it possible to identify the agent's discount factor β. Instead the "identified set" of structural objects includes all discount factors β in the $[0,1]$, and a much wider class of preferences that include utility functions that are not monotonic affine transformation of a single underlying "true" utility function u^*. Thus, it is possible to "rationalize" any conditional choice probability $P(d|x)$ as resulting from an optimal decision rule and we can rationalize it in an infinite number of ways, including explaining agent's choices in terms of a static model ($\beta = 0$) or in a dynamic model with any $\beta \in (0,1)$.

Lack of identification of a structural model means that policy evaluation and forecasting is problematic. Suppose there are mutiple structures that map into the same reduced form. This means these alternative structures are *observationally equivalent.* 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 behavior responses and changes in agent welfare, which one of them do we believe?

I can describe the problem more precisely in the context of the dynamic discrete choice model. Let π_s denote a vector of "policy parameters" representing policies in effect under a *status quo* regime. Suppose this policy affects agents by changing agents' beliefs about the evolution of observable variables, so we write p (the decision-dependent transition probability for observed state variables) as an implicit function of π_s, say $p(\pi_s)$. Thus π_s, together with the structure of the problem results in the observed behavior under the *status quo,* which is captured by the conditional choice probability P_s. Now consider some hypothetical new policy π_n. If we knew the true structure $\{\beta,u,p(\pi_s),q\}$ we could solve the agent's dynamic programming problem to predict both behavior P_n and welfare under the new policy regime π_n as $P_n = \Lambda(\beta,u,p(\pi_n),q)$. However suppose there is an alternative observationally equivalent structure

$\{\beta', u', p(\pi_s), q'\}$, i.e. $P_s = \Lambda(\beta, u, p(\pi_s), q) = \Lambda(\beta', u', p(\pi_s), q')$. However it is entirely possible that the structural model predicts that the two structures will no longer be obsevationally equivalent under the new policy π_n. That is, $\Lambda(\beta, u, p(\pi_n), q) = P_n \neq P'_n = \Lambda(\beta', u', p(\pi_n), q')$. In this case, we have no way of telling what the true structure is from our "in sample" data under the *status quo* regime, and if we picked the wrong structure (β', u', q') it would result in incorrect forecasts of the behavioral and welfare effects of the policy change.

On the other hand, if we could do a controlled experiment on the agent, and if there were only the two observationally equivalent structures (β, u, q) and (β', u', q'), the experiment would reveal that the latter structure was the wrong one and the former was the right one, and in this way *additional data generated from an experiment* can 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 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 a series of experiments may not be enough to identify the true underlying structure.

Wolpin does not devote much space in his book to a discussion of the identification problem, which is fine with me since I find most theoretical analyses of identification to be profoundly boring, arid exercises. The problems I have discussed above are logical possibilities, but is there any concrete evidence that unidentified, or poorly identified structural models have resulted in misleading policy forecasts? I do not have any specific examples to back up the concerns raised above, other than some compelling auction design failures that resulted from invalid "maintained assumptions" such as the absence of collusion.[6] However it is evident to me that most applied econometricians are profoundly concerned about the identification of their models: for further discussion see, e.g. Nevo and Whinston (2010).

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*

[6] For example Klemperer (2004) notes that "many auctions — including some designed with the help of leading academic economists — have worked very badly" (p. 102) He concludes that "the most important features of an auction are its robustness against collusion and its attractiveness to potential bidders. Failure to attend to these issues can lead to disaster." (p. 122).

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 flexibiity in fitting a model to the data, and if 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 mypoic 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 optimzers.[7]

The growing interest in behavioral economics is also evidence that many other economists have similar

[7] 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."

opinions. However if structural econometricians are so good in rationalizing everyones' behavior using highly complex dynamic programming models, behavioral economists are very naive if they think it will it be easy to identify individuals who are not behaving 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 by 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 challege, 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 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 allowed 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. 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 there the criterion can also be locally flat at the maximum (at least for certain parameters) in which case 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 can say very little in general from a highly abstract, mathematical vantage point.

In my comments on Keane's article (Rust 2009) I quoted from a paper by Heckman and Navarro (2006) that complains that my views 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 acknowleding that most of the interesting economic models are non-parametrically unidentified necessarily implies that structural estimation is a nihilistic, meaningless exercise. I think that what Heckman and many other econometricians tend to lose sight of the fact that models are necessarily highly oversimplified approximations to reality and can never be correct. Of course can still have an identification problem for misspecified models (no model may fit the data perfectly but several different theories fit 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.

I sometimes wish that economics could be like physics, where the standard model of physics can be condensed to a single *parametric* equation with 19 free parameters. The question of non-parametric identification does not even enter the physicist's lexicon. Though their theory is parametric, elegant, and admirably compact, it is also amazingly powerful. The parametric standard model of physics has been able to correctly predict *ex ante* the existence of a large range of interesting phenomena including most recently, the existence of the Higgs Boson, whose existence was postulated via a theoretical model by Nobel Prize winner Peter Higgs and others in the mid 1960s. Though the standard model is by no means a perfect model that explains all phenomena, the power of physics to combine a strong parametric theory with well designed experiments that can verify (or refute) it should be highly instructive to economists.

Economists may not be as talented as physicists are, and cannot match their impressive theoretical and empirical accomplishments, but I suspect this is partly due to the fact that the "elementary particles" of economics (e.g. individual human beings) are simply vastly more complex and inherently less predictable (in a probabilistic sense) than the elementary particles of physics. As a result economists tend to have far many more models, too few predictions, and too little ability to conduct the definitive tests to help confirm various key theories. While we may not be able to aspire to the same tight, testable predictions from our theories, to me physics provides an example of the huge payoff to theoretical *parametric* modeling combined with well focused data gathering and experimentation.

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 provded 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 many papers 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 "exogenous state variable" in the model), so in effect there is a curse of dimeensionality 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, beliefs, and the (state-dependent) *equilibrium selection rule* in static and dynamic games (see, e.g. Aguirrebiria 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 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.

Though Wolpin does not devote any space to the structural estimation of dynamic games in his book, 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 we are using game theory to find the solution to a 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 choosing one of the many possible Nash equilibria in this game is just 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 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 real agents are strictly behaving according to the concept of Nash equilibrium, it seems reasonable to suppose that interacting adaptive, intelligent agents might converge to something close to a Nash equilibrium in a sufficiently stable environment.

However the dynamics of interacting, co-adapting intelligent agents can be highly complex 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 steady state/equilibrium situation. If this is correct, there is a high level of interdeterminancy in these complex systems which makes makes policy forecasting all the more difficult. 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 Lack of Good Data and Trust of Policymakers

In view of the problems discussed above, perhaps it is not surprising that there has not been a huge demand so far by policymakers (governments and firms) for structural econometric models. Having lived "inside the Beltway" in Washington DC for over a decade, it is all too apparent to me that major policy decisions are made largely based on intuition/gut instinct, and to the extent there is outside consultation and advice, it is usually with lobbyists, political consultants and other power brokers. Thus formal econometric models of any sort (structural or non-structural) play very little role in how most public policy decisions are actually made.

Though the lack of good data and the trust of policymakers may not be an inherent or insurmountable limit to inference, it is a very daunting practical obstacle. This is because structural econometrics is probably more dependent on having large quantities good data than other areas of econometrics (because the more detailed nature of structure models drives us towards attempting to measure the states, decisions, and even beliefs of individual agents as accurately and completely as possible). Since good data are very costly to collect and provide, unless structural econometricians can make a compelling argument in terms of benefits from data collection and improved policy analysis sufficiently outweighing the costs, we will not be able get the sort of data we need to estimate models that provide better approximations to reality. We will be in a catch-22.

Specifically, due to the lack of data, it is harder for us to develop good (realistic, trustworthy) models that are credible to policy makers. This tends to have a feedback effect that further reduces the lack of credibility these models have among policy makers, since structural econometricians have limited opportunities to actually practice any real policy forecasting and evaluation with their models. Frankly, much of the talk of policy evaluation is academic fantasy that is confined to academic journals. When it comes to very large scale, important policy decisions such as evaluating changes in Social Security or Disability Insurance reform, I would be the first to admit that the sort of structural models I have developed are not sufficiently realistic and well tested to be ready for “prime time.”

It seems reasonable that with a sufficient investment in data and research, the level of realism and reliability of these models could be improved by an order of magntitude, to the point where I think they could be valuable tools for policymaking. However it seems unlikely that these investments will be made given the level of controversy within the economics profession surrounding structural models. For the forseeable future the academics who have influence in policy making at the highest levels of government will be high profile “gurus” who rely primarily on economic intuition and may not have any special appreciation for structural econometric models or bother to incorporate the insights or predictions from these models when they whisper their advice into the ears of power.

So I see the internal controversies over structural models within the economics profession as a major reason why structural approaches to policy forecasting have little credibility among high level policymakers. Randomized experiments currently seem to have far more credibility, and indeed many new proposed government policy changes are mandated to be evaluated by controlled experiments (also called *demonstration projects* in the U.S.) instead of structural econometric models. The reason I devoted the space to Frijters' review of Wolpin's book was to illustrate just how poorly other *academics* understand and regard structural econometric work. If these models are still so controversial in academia, then it is clear that there is quite a long distance to go before these models could be explained and clearly understood by the public, the press, and public officials. The complexity and number of assumptions involved in doing structural work comes hand in hand with the risk that the methods and results will be misinterpreted and misunderstood, and end up having little credibility — *especially* when it comes to high stake policy decisions.

Based my own limited experience in Washington, I have given up on any hope that structural models will have any use for policy making for important public policy questions during my lifetime. Instead, I have turned my own focus towards *little league* by which I mean to try to have concrete successes in using structural models to do policy forecasting and evaluation for in smaller, more tractable and well-defined contexts in the private sector. One of the problems in using structural models for policymaking in a public context is that nearly any interesting policy change will have winners and losers. Without any clear metric for aggregating individual preferences into a social welfare function, it is difficult to provide any convincing evidence that the structural models helped identify policy changes that resulted in a clearly measurable improvement in social welfare. However in the case of private firms the objective function is much more clearly defined and easier to measure — profits. Thus, if an analysis of data under the *status quo* using structural models followed by policy analyis that results in recommended counterfactual policies that increase the firm's expected profits, this is something that captures firms' interest and is a situation where it is possible to document the benefits to the structural modeling approach much more convincingly.

However it is still a struggle for me personally to have any degree of credibility and get my foot in the door even in the private sector. While this may just be a reflection of my own limited abilities, it does seem clear that many private firms are run by executives who, similar to politicians, operate more on the basis of intuition and gut instinct than any formal science. Surprisingly, even the use of controlled experiments is something that is not frequently done by the businesses and businessmen I have interacted with. Thus, many businessman have a predisposition towards mistrust and suspicion of academics, and it takes quite a bit of convincing and hand-holding to get them to release any propriety data from their business, since these data can often be of substantial value to competitors.

However some businessman can understand and relate to sufficiently realistic models of their operations, especially when these models can be simulated and shown to do a good job of replicating their behavior/strategies under their *status quo* operating policy and when the results are presented in an intuitive manner. When we can show counterfactual simulations of these models that predict alternative operating policies that result in higher expected profits, they can be motivated to undertake controlled experiments to test whether the model's prediction is correct. However once again this is more likely to happen if the alternative policy can be explained in a fashion that makes intuitive sense, because the underlying mathematics and econometric/computer modeling are things that many businessmen will not be able to appreciate in any detail.

Indeed we are starting to see practical applications of relatively sophisticated mathematical and statistical models in related domains. For example the (2003) book *Moneyball* by Michael Lewis described the success of *sabremetrics*, which is the empirical analysis of the game of baseball that results in policy advice to baseball owners on how to cost-effectively assemble a winning team of baseball players. Nate Silver's success in predicting election outcomes and other phenomena described in his (2012) book *The Signal and the Noise* has also attracted considerable attention from politicians interested in advice on where to spend the marginal campaign dollar to turn a close election in their favor. These are not exactly examples of "structural econometric models" but they close couisins, because they are based on a combination of extensive data collection, model building and creative theorizing. The attention this work is getting in the popular press may help make it easier for structural modelers to attract the attention and gain the trust of policymakers in both the private and public sectors. Each additional success in structural policy forecasting helps convince these policymakers to provide us with more/better data, and more importantly, engage in a two way dialog that in my experience greatly improves the quality of our empirical work by helping academics to better understand and model the situation "on the ground."

3 Combining Structural Estimation and Experimentation

It should be evident from the preceding discussion that there are huge synergies between structural estimation and experimentation. I have discussed the important contribution by Todd and Wolpin (2003) which estimated a family-level structural model of fertility and schooling choice using the control group in the PROGRESA experiment, and showed that it could make reasonably good out-of-sample predictions of the behavioral change caused by the school atttendance subsidies using the treatment group. Further, 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 is likely it would 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, there has been a widely perceived 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)

The complementarities between structural econometrics and experimentation of all types is becoming more widely appreciated in fields outside industrial organization. I already discussed the huge payoff that El Gamal and Grether (1995) obtained from modeling and using sophisticated econometric techniques to analysis the data generated from their laboratory experiment on how people make inferences which showed that not all people are Bayesian decision makers.

Even in development, a field that is widely perceived to be dominated by experimental work and hostile to structural econometrics, there are signs that attitudes are changing. Besides the Todd and Wolpin (2003) study, I would also point out the paper by Kaboski and Townsend (2011) which was awarded

Frisch Medal by the Econometric Society in 2012 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). The Econometric Society noted that

> "Simulation of the model successfully matches the qualitative features of the post-program data and provides evidence of the role of credit constraints in household consumption decisions. The structural approach taken in the paper also allows for a cost benefit analysis of the microfinance program as compared to direct income transfers to households and shows that the microfinance program costs 30% more than a direct transfer program that would achieve the same average utility gain. The paper is noteworthy for its combination of rigorous theory and careful econometrics to produce important insights into a major development policy."

It is important to realize that not all experimentalists take such a narrow and dogmatic approach to inference reflected in the Angrist and Pischke (2010) survey paper. For example 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. Our view is that economists' insights can and should guide policy-making (see also Banerjee, 2002). They are sometimes well placed to propose or identify programs that are likely to make big differences. Perhaps even more importantly, they are often in a position to midwife the process of policy discovery, based on the interplay of theory and experimental research. 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, I see encouraging signs of change and methodological agreement in this very important area. 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.

4 Conclusion

Ken Wolpin's book is an excellent illustration of the limits to inference *without* theory. The main point of my review, perhaps obvious, is there are also limits to inference *with* theory. I don't think that Ken 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 *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 interpret the myriad of data around us, and I couldn't agree more. We would also both probably agree that if we were to exclude anything, it 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.[8] He would also say 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 — before we know whether any of them are good or bad — makes about as much sense as throwing away data because they might be difficult to analyze.

It is just common sense that we can make much more progress from combining inductive *and* deductive modes of inference. 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.* The aversion to structural estimation is still very much prevalent in the profession, and is reflected in the hostility and ignorance in another published review of Wolpin's book.

Given this degree of hostility, why do I emphasize the limits of inference *with* theory? This may reflect my own attempt to understand why so many influential empirical people in the profession are so disturbed by the use of theory in empirical work. I also see a great deal of indifference by theorists towards empirical work, as if empiricists and theorists ought to be inhabiting parallel universes. Structural econometricians like Wolpin ought to be the glue that can bridge the gap between the "pure empiricists" and the "pure theorists." Instead purists on both sides seem to regard the structural community as oddball renegade half-breeds who are vaguely threatening in some way.

The other reason I emphasize the limits to inference is to make clear to the people who dislike theory and structural econometrics that we are not elitists who feel smug and superior by the greater set of tools we bring to bear compared to empiricists who do not want to use models for whatever reason. I have tried to be as transparent as possible about the huge obstacles facing the structural estimation industry, and perhaps to explain why there haven't been bigger, more obvious successes that we can point to so far. While one explanation might obviously be a lack of talent and creativity on our part, I have tried to show that there may be strict inherent and *a priori* limits to what we can learn from inductive inference, just as Gödel proved there are inherent limits to what we can learn from deductive inference in his celebrated *Incompleteness Theorem.*

However even though we know there are limits to deductive inference (i.e. Gödel showed that there are truths that cannot be proved in any formal reasoning system that is at least as complex as formal arithmetic), this does not mean huge strides can be made via deductive modes of inference. Fermat's Last

[8]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.

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.

Similarly, though there are many daunting, perhaps insuperable challenges to inductive inference, even when complemented and augmented with theory, there is still ample room for great progress to be made. I have used the parametric "standard model of physics" as 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. 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 the safety of 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 could record, millisecond by millisecond, the forces acting on the car frame itself and the shocks experienced by the crash dummies inside the car as it crashed. However over time engineers developed increasingly realistic finite element models of cars and crash dummies. This allowed engineers to crash these cars, virtually, inside the supercomputer. With significant investments and constant refinement of the models, the virtual crashes began to predict crumpling in the car frame and forces on the virtual sensors in the virtual dummies that were virtually indistinguishable from from the data generated from test crashes of actual cars with actual test dummies and sensors. 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.

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. Three decades after the *Lucas critique* economic policy making is still in the dark ages where our leaders do 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.[9]

In neuroscience there is growing evidence that the human brain has an amazing innate, subconscious ability to model and simulate reality. Indeed 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 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)

They 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 (and correct) prediction that it would return every 76 years.

[9]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 spent to produce an obviously malfunctioning website. A well functioning website is key to the success of the program since attracting younger, healthier and more Internet saavy 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.

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 astronomical observations: They could have been produced by three different comets, each travelling 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)"

They note 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, in the form of an abstract theory, generates hypotheses about the candidate causal models that can apply in a given situation." and this "explains how people's inferences about the structure of specific causal systems can be correct, even given very little data." (p. 662).

Obviously millions of years of evolution has lead humans to have incredibly powerful internal *but subconscious* modeling abilities. In effect, we are *all* master "structural model builders" — even Joshua Angrist! Further we regularly use our models to conduct "policy evaluation" via internal simulations of our mental models. What is a dream if not an incredibly realistic, counterfactual simulated reality? 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) and that "In fact, 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 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 explanation 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, subsconcious 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: Moore's Law implies a continuous time rate of improvement in computing power of 46% per year!

I think the historical perspective on economic modeling the wonderful book by Mary Morgan (2009) is also very helpful, and is consistent with the view that interplay between model building and empirical analysis has had and will continue to have hugely transformative effects on the way we do economics:

> "The comparison between astronomical models and economic models that has woven its way through this chapter is not just an heuristic comparison which helps us see how economists use models, but keeps reminding us that the modelling style of reasoning has an illustrious history. Indeed, the scientific revolution of the sixteenth and seventeenth centuries was not just one of content, but of styles of reasoning. Modelling has been portrayed as the working method of Galileo no less, and continues to be prevalent in modern natural sciences. Despite this ancestry, economists are not fully sure that the method has a credible scientific respectability. Models are relatively small and simple compared to the economic world, they are made of different materials, and cannot well be applied directly to that world. Even so, like those models of the universe of earlier days, economic models may still capture the heart of the problems that economists seek to understand. Modelling is not an easy way to find truths about the economy, but rather a practical form of reasoning for economists, a method of exploration, of enquiry, into both their ideas and

their world. That is the thesis of this book."

I conclude this review by suggesting that anyone who has a good understanding of the history of science would agree that Ken Wolpin's ideas are uncontroversial and even almost obviously correct. Denying any role for theory in inference is an untenable, indefensible position. Nevertheless it is a position that still holds great sway in the economics profession fully six decades after the *Koopmans critique.* I guess it just shows that economists are slow learners.

I have absolutely no problem with people exploiting their comparative advantages in science, and in my comments on Keane's article, 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. And if you look closely at economists who are as skeptical of the value of economic modeling as Charles Manski or Joshua Angrist are, 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.

But to those remaining skeptics and haters or structural modeling, I would say, 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] Newell, Allen and Herbert A. Simon (1972) *Human Problem Solving.*

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

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

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

[5] 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.

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

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

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

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

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

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

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

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

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

[15] Mary S. Morgan (2009) *The World in the Model* Cambridge University Press.

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

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

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

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

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

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

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

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

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

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

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

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

[28] Hausman, J. (1981) "Exact Consumer's Surplus and Deadweight Loss" *American Economic Review* **71-4** 662–676.

[29] 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.

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

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

[32] 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.

[33] 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.

[34] 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.

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

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

[37] Prescott, E.C. (2004) "Why Do Americans Work So Much More Than Europeans?" *Federal Reserve Bank of Minneapolis Quarterly Review* **28-1** 2–13.

[38] Chetty, R. Guren, A., Manoli, D., and A. Weber (2011) "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins" *American Economic Review Papers and Proceedings* **101** 471–475.

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

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

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

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

[43] 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.

[44] 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.

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

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

[47] 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.

[48] 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.

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

[50] 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.

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

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

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

[54] 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.

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

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

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

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

[59] 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.

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

[61] 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.

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