The Econometrics of Social Insurance

New Zealand Treasury Invited Lecture
2001 Australasian Meeting of the Econometric Society
Auckland, New Zealand

John Rust
Yale University $\Longrightarrow$ *University of Maryland*

July 6, 2001

Overview

- The aging of the world and problems in the design of "first generation" social insurance systems make it likely that there will be significant changes in the structure of these systems in the coming decades.
- To date, almost all changes in social insurance policy amount to "policy experiments" without any formal *ex ante* predictions from economists about their welfare, behavioral, or fiscal impacts.
- Societies have created social insurance institutions through combination of trial and error (a "groping process") and via imitation, after judging the pros and cons of various innovations to these systems adopted by other countries.

- Automakers find it faster and more cost-effective to "crash test" new car designs using computer models before they actually build a real vehicle. The same approach should be used when contemplating changing social insurance institutions: we should "crash test" a proposed change using computer models instead of using humans as live gineau pigs.
- I call this approach **computational mechanism design**
- **Claim:** *It is much more cost-effective to "crash test" a proposed change to a social insurance system using simulation models with millions of computerized "crash dummies" than it is to conduct unguided policy experiments using millions of real human beings. But can computer models provide sufficiently realistic and reliable predictors of actual outcomes?*

- It is easy to point out a number of *naive* albeit well-intentioned actual and/or proposed changes to social insurance institutions which might not have been undertaken or would have been substantially changed if policy makers had had better *ex ante* guidance about their actual consequences (e.g. the aborted retrenchment in the disability insurance program started by Reagan and aborted by Bush, the Clinton health care proposals, and the Clinton/Kennedy patient bill of rights legislation).
- Would better simulation/forecasting models reduce the likelihood of misguided policy changes?

- The degree to which policy makers should rely on computer models and simulations depends on:
  a) the degree of reliability/credibility in their predictions,
  b) the size of the up-front development costs and ongoing maintenance costs.
- Until very recently in the U.S. govt. agencies such as the SSA and the CBO have made due with seriously flawed, outdated models for policy analysis. This was not so bad prior to the 1990s when significant or "radical" changes to the structure of the Social Security (e.g. individual accounts) were not being contemplated. Now that these agencies are facing increasing pressure to provide guidance to Congress about the distributional and fiscal consequences of such reforms, they have begun to realize that their existing models are inadequate and have started to invest millions to create better ones.

Why have economists failed to provide formal guidance to policy makers?

Some hypotheses:

1. Governments have ignored economists and have not invested adequate resources to develop good policy forecasting models.
2. Economists have tried to develop good policy models before but have not succeeded. Good policy models are possible, but insufficient talent has been applied to this problem.
3. There are deep conflicts in the economics profession about the possibility and value of developing adequate formal policy forecasting models. The lack of consensus in the economics profession is the main reason why formal models are not being developed and used by policymakers.

Problems with econometric policy models

The first generation of "microsimulation models" (largely due to the work of Guy Orcutt, now embodied by the *Dynasim* model currently used by the Urban Institute or the Corsim model used by the U.S. Social Security Administration) are now regarded as *ad hoc* and unreliable "black boxes". Most of these models are based on simple accounting identities and a few behavioral relationships that were estimated via reduced-form methods using data sets that are now very out of date. While the models have a relatively realistic treatment of certain aspects of social insurance rules and are potentially good for analyzing distributional consequences of policy changes, they are based on static models that fail to incorporate the important effects of expectations and uncertainty.

The current generation of *dynamic structural models* use dynamic programming to solve individuals' life cycle optimization problems. These models are based on a internally consistent utility maximization framework and can account for dynamics, expectations, and uncertainty. However these models are much more complicated to solve and to estimate structurally, and currently I am not aware of anyone who has formally estimated and tested the full life cycle model.

I argue that we are on the verge of being able to formulate, estimate, test, and simulate realistic versions of the life cycle model. I believe these models will provide good approximations to observed behavior and will provide a reliable and cost-effective basis for policy forecasting and "computational mechanism design".

Counterpoint:

- **The Feldstein/Gruber "back of the envelope" view:** Dynamic structural models are too complicated. They are the modern "black boxes". All of the essential intuition can be gained from much simpler models using "back of the envelope" calculations. There are sharply diminishing returns to developing more complicated and realistic econometric models, and the predictions of these models will not have any credibility because most policy makers won't understand them, and most economists are incapable of estimating and using them.

- **The Ashenfelter/Card/LaLonde "antistructuralist" experimentalist view:** classical controlled experiments have demonstrated that the predicted effects of policy interventions from econometric models have been far from the "actual" effect of the policy intervention, computed by taking "differences in differences" in outcomes between the treatment and control groups. These experiments "prove" that predictions of structural econometric models are usually wrong, and therefore they should not be used for policy evaluation.

- The "structuralist/experimentalist" debate in econometrics began with Robert LaLonde's 1986 *AER* article on "Evaluating the Econometric Evaluations of Training Programs with Experimental Data".
- He used data from the National Supported Work Demonstration (NSW) where low income individuals seeking job training were randomly assigned to "treatment" and "control" groups.
- The experiment cost about $9,100 per participant (1986 dollars) and lasted 44 months, with members of the treatment group getting training and a guaranteed job for 9 to 18 months and the members of the control group getting nothing.
- **Finding:** members of the treatment group experienced slightly higher mean earnings ($851 per year) than the control group in the first year after leaving the program.

- LaLonde compared the experimental results with the predicted effects from Heckman-style wage regressions accounting for self-selection in application for participation in the training program.
- He found that "many of the econometric procedures and comparison groups used to evaluate employment and training programs would not have yielded accurate or precise impacts of the National Supported Work Program. The econometric estimates often differ significantly from the experimental results. Moreover, even when the econometric estimates pass conventional specification tests, they still fail to replicate the experimentally determined results." (LaLonde *AER* 1986, p. 617).

- Heckman, Hotz and Dabos (HHD,1987) *Evaluation Review* reinvestigated LaLonde's results using the NSW data. They concluded that:
- "Using several simple strategies for testing the appropriateness of alternative formulations of such estimators, we show that a number of nonexperimental estimators used in these studies can be rejected. Although we eliminate a number of nonexperimental estimators by such tests, were are able to find estimators that are not rejected by these tests. Esitmators not rejected by such tests yield net impact estimates that lead to the same inference aobut the impact of the program as the experimental estimates. The empirical results from our limited study provide tangible evidence that the recent denunciation of nonexperimental methods for evaluating manpower training effects is premature." (p. 395).

- "The real lesson from [LaLonde's] work is that invalid models produce wildly discordant estimates — not that nonexperimental methods are infeasible or that there is no objective way to choose among competing models." (p. 421).
- The HHD article also pointed out a number of severe, inherent limitations in exclusively relying on experimental methods for policy analysis:

  1. Experiments are almost always extremely costly and take a long time to complete. This severely limits the number of different "treatments" than can be considered.
  2. The experimentalists adopt the naive assumption that the everyone receives the same "treatment" and the "treatment effect" is the same for everyone.

- In practice, the treatment leads to a distribution of outcomes and one can't experimentally control for the level of effort by participants and the training agency, so in practice levels of treatment vary randomly in a way that can't be fully controlled for experimentally.
- More importantly, the "real world" is not a controlled experiment. Self-selection and "cream skimming" in training programs is endemic. Experimental predictions that attempt to eliminate the effects of self-selection and cream-skimming through classical randomization have little relevance to the question of how a real training program would operate and what its effects will be.

- "Even if all of these problems could be solved — or safely ignored — experimental data are of limited value and must be supplemented with nonexperimental analyses. Participation in a training program entails a multistage process of application, selection, continuation in the program, and placement. It is of interest to the know the effect of the training conditional on completing each stage of this process. In order to address this issue experimentally, it is necessary to randomize at each stage — something rarely done in social experiments because of the difficulty of doing so. Typically randomization occurs at only one stage in the process. ... Randomization can feasibly be used to answer only a limited subset of the interesting questions."

- Despite the congency and clarity of Heckman's arguments, the LaLonde "antistructuralist" an "experimentalist" view won out in the U.S. Congress.
- **Since that time the Congress has mandated that if government funding is used to predict the impacts of various policy changes, then classical experimental methods with random assignment of subjects to treatment and control groups** *must be used.*

Other Problems with the dynamic, structural approach to policy forecasting.

- **The Curse of Dimensionality** This problem, noted by Richard Bellman in 1957, is that the as the level of detail and "realism" of a life-cycle, dynamic programming model increases (as quantified by the number of "state variables" entering the model) the computational requirements necessary to the solve the model increase exponentially. This problem implies that it is only feasible to solve highly simplified, unrealistic versions of the life cycle problem that can't be taken "seriously" for use in policy forecasting.

- Rust (*Econometrica* 1997) proved that in certain classes of dynamic programming problems (discrete choice DP problems), randomization can be used to break the curse of dimensionality.
- Rust, Traub and Woźniakowski (*Econometrica* forthcoming) proved that with additional special structure, it is possible to break the curse of dimensionality using *deterministic methods* and that there exist methods for solving extremely high dimensional problems with essentially the same degree of difficulty as solving a one dimensional problem. These problems are said to be *strongly tractable* and from a computational standpoint their *effective dimensionality* is 1.

- Technologically, *Moore's Law* has lead to improvements in the speed of computer hardware at an exponential rate. My laptop processes at a faster rate than the first generation supercomputers I used to use in the mid 1980s.
- Via a combination of better hardware (including massive parallel processing) and better algorithms we have been able to steadily increase the level of realism or "virtual reality" in the life cycle models that we are able to solve numerically.
- With more significant investment in this area, it will soon be possible to develop life cycle models at a sufficient level of detail to be of considerable use in policy forecasting. It will possible to formulate and solve fairly detailed life cycle models that include education, marriage/divorce, fertility, housing decisions, career decisions, detailed models of health care, private pensions, and purchase of an array of different private insurance contracts and financial assets.

The identification problem: In my (1994) *Handbook of Econometrics* chapter "Structural Estimation of Markov Decision Prcoesses" I proved that dynamic programming models are non-parametrically unidentified. That is, it is always possible to find some pair of preferences and beliefs that succeed in "rationalizing" any observed behavior pattern. A corollary to this result is that there is generally an equivalence class with uncountably many pairs of preferences and beliefs that succeed in rationalizing the data. If different pairs of preferences and beliefs produce different predicted responses to policy interventions, then it will be no objective basis for determining which prediction is more likely to occur since the various models in the equivalence class succeed in explaining the historical data equally well.

- In practice, econometricians don't search over infinite families of preferences and beliefs, but instead (due to the identification problem) we work with simple, parsimoniously parameterized families.
- Generically, parametric classes of preferences and beliefs are identified.
- I realize that models are just approximations to reality, and view a parametric model as a parsimonious way of summarizing behavior and predicting behavioral responses to policy interventions.
- The real test of a parametric model is whether it does a better job than any other comparable parsimonious model (or other *ad hoc* methods of prediction) in summarizing behavior and predicting behavioral responses.

- I view classical controlled experiments as an ideal way to test the predicted behavioral responses from dynamic structural models. Although uncontrolled policy experiments also provide a "testing ground" for structural models, in a classical controlled experiment we are much more certain that changes in behavioral responses between the treatment and the control groups is due to the treatment and not due to some other change in the environment that was not controlled for.
- If the *ex ante* predictions from dynamic structural models closely mimic the actual behavioral responses we observe in controlled experiments and also in uncontrolled policy experiments, it provides policy makers with greater confidence that the model can be relied on to predict the impacts of other policy changes.

- *Thus the real test of a structural model is how well it forecasts behavioral responses in controlled experiments. In this sense structural estimation and experimental methods ought to go hand in hand.* However in practice, structuralists and experimentalists have viewed each other as arch enemies.
- I discuss a recent new controlled experiment that the U.S. Social Security is undertaking in response to the 1999 *Ticket to Work Act and the Work Incentives Improvement Act.* (TWWIIA) If subsequent appropriations enable the TWWIIA project to be implemented according to current plans, **it will constitute the largest controlled policy experiment since the Negative Income Tax experiments in the late 1970s.**

- I am one of the academic advisors to the TWWIIA project. I view this "demonstration project" as a great scientific opportunity to rigorously test the dynamic structural models of retirement and social insurance that I have been developing over the past decade.
- I will use the TWWIIA project to demonstrate the inherent limitations of an exclusive reliance of experimental methods for policy analysis, and how a combination of structural modeling and experimental methods can be complimentary activities and enable policy makers to analyze the impacts of a much broader range of policy interventions much more rapidly and cost-effectively. These experiments could be the entree for the use of more advanced life cycle policy models by the Social Security Administration.

Outline for Rest of Talk

0. Overview
1. Review of the U.S. Disability Insurance and Old Age Insurance Program
2. Summary of the TWWIIA Demonstration Project and its intellectual origins
3. Review of micro panel data on retirement and disability behavior from the Health and Retirement Survey (HRS)

4. Introduce a life cycle model of labor supply, consumption, and savings with an integrated treatment of Social Security.
5. Illustration the value of the life cycle model for policy simulations:
   a) assessing the winners and losers in privitization,
   b) predicting the impact of "induced entry" in the TWWIIA project,
   c) predicting the impact of the elimination of the "earnings test."
6. Discussion of related models: Benitez-Silva (a solution to the "annuity puzzle") and Heyma (a structural model of the Dutch social insurance system).
7. Conclusions

Origins of the 1999 Ticket to Work and Work Incentives Improvement Act (TWWIIA)

- 1996 National Academy of Social Insurance (NASI) panel headed by Jerry Mashaw of Yale Law School.
- Two economists on the panel suggested some new policies to help create better incentives for disabled individuals to return to work.
- Currently despite reasonably strong incentives including a 9 month trial work period where a DI beneficiary can return to work without losing any benefits, fewer than 0.18 of 1 percent of DI beneficiaries voluntarily return to work an enventually have their benefits terminated after the trial work period ends.

- **Monroe Berkowitz,** an emeritus economist from Rutgers, proposed the use of **return to work tickets.** These are vouchers that can be used by DI beneficiaries to give to a vocational rehabilitation agency of their choice. The agency would provide vocational training and rehabilitation services at no cost to the DI beneficiary. They would be paid by SSA *only if their efforts to rehabilitate the person was successful.* In that case the agency would be paid a fraction of the DI benefits SSA would have otherwise paid to the DI beneficiary had they not left the roles.
- **Richard Burkhauser,** a labor economist at Cornell, proposed the use of a **disabled worker tax credit.** This is a tax credit similar to the existing *Earned Income Tax Credit* (EITC) that would be payable to individuals with significant disabilities who return to work.

Skeptics of the Validity of the Life Cycle Model

- **Bernheim (1992)** "Models households' optimal saving and consumption choices as a function of family size, education, earnings, age, social security, pensions and other factors. He then compares household's actual saving with the simulation results. His primary finding, summarized in a "baby boomer retirement index" is that boomers' retirement saving averages only about one-third of that needed to maintain preretirement living standards in retirement." (quoted Engen, Gale, Uccello, *Brookings Papers* 1999, p. 124).

- **Bernheim and Scholz 1993** Compares actual wealth accumulation from their life cycle model with actual wealth accumulation in the 1983-86 Survey of Consumer Finances. They conclude that "many Americans, particular those without a college education, save too little."
- **Bernheim, 1989** Compares predicted rates of wealth decumulation after retirement from a deterministic life cycle model with no bequests to the actual rates of wealth decumulation for individuals in the 1969-1979 Retirement Health Survey. He concludes that "We show that the life cycle model has strong implications about how rates of accumulation and depletion will respond tothe imposition of nondiscretionary annuities. Implementation of these tests produces results which are unfavorable to the pure life cycle hypothesis." (p. 238).

Recent Work Supportive of the Validity of the Life Cycle Model

- **Engen, Gale and Uccello, 1999 "The Adequacy of Household Saving"**
- "Our study differs from previous work in that it uses a stochastic life-cycle model to generate optimal wealth accumulation benchmarks. Because of uncertainty of earnings, the model generates a distribution of optimal wealth-earnings ratios among households that are observationally equivalent. This distribution implies that some households have very low wealth earnings ratios are nonetheless saving optimally for retirement."

- "Applying the model to data from the HRS and SCF suggests, in the base specification, that more than half of households exceeed the median wealth-earnings ratios from the simulation. In addition, households at the 75th and 95th percentiles of the wealth-earnings distribution exceed the models' wealth benchmarks. There is some mixed evidence of inadequate saving among households with low wealth-earnings ratios. Our results appear, at least at first glance, to be significantly more optimistic than the interpretations provided in previous research. However a careful interpretation of previous work indicates that earlier results are in fact largely consistent with ours." (p. 142).

NIH Panel Evaluation of Proposal by Rust, Buchinsky and Benitez-Silva, "Dynamic Structural Models of Retirement and Disability"

- **Critique 1:** "There is clearly some merit in thinking ambitiously about the dynamic modeling of retirement. Reduced form models in this area have gone as far as they can, and that is not far enough. But I remain unconvinced that this is the framework that we would be using to do that thinking. The complicated modeling here does not demonstrably add value to simpler efforts being undertaken elsewhere to understand retirement decisions. And this revision, if anything, represents a step backward in making the case for this approach."

- **Critique 2:** "This is a revised proposal, and the researchers are remarkably unresponsive to the criticisms in the previous round. There were three main criticisms that I found troubling. The first is that this very complicated approach was technically impressive but had found little adherence in either the research or policy communities. The revised proposal does nothing to refute this, beyond simply asserting that it is not true! But, unless I am mistaken, it remains true that no one beyond this set of investigators has used this framework to model retirement decisions. This may be because it is so difficult, and that is not a reason to deny funding to the research. But it means that the general benefits to the research community have been thus far somewhat limited."

- **Critique 3:** "The research agenda is founded upon a sound idea, and it important to purse the idea as far as the state of the art, data limitations, and computational constraints will allow. There are few economists who have the expertise to make significant advances in this technically demanding area, and these investigators are clearly among the leader in the field. The most disappointing part of the proposal is the insufficient emphasis on the extensive information available in the data sets concerning expectations and preferences. This information is clearly relevant for the applications's research questions, and it does not get nearly the attention it deserves."

- **Critique 4:** "Fundamentally, the assessment of this proposal depends on whether one believes the dynamic programming model is a valuable way view the world. This proposal is a a state-of-the-art application of dynamic programming. Still, many economists do not view these models highly as a whole."

- **Panel Recommendation: REJECT**
  "This is an amended application to explore dynamic modeling of retirement. Reduced form models in this area have gone as far as they can. The research agenda is founded on a sound idea, and it is important to pursue the idea is far as the state of the art, data limitations, and computational constraints will permit. The investigators are clearly among the leaders in the field, and the topic is significant. However, the investigators have been largely unresponsive. This revision fails to make a case for the added value of dynamic modeling. There is insufficient emphasis on the extensive information in the data set concerning expectations and preferences, and this information is critically relevant for the application's research questions."

Economists on the NIH review panel:

- David Card, University of California at Berkeley
- David Cutler, Harvard University
- Jonathan Gruber, MIT
- Robert Moffitt, Johns Hopkins University