May 6, 2001

Yale University

John Rust

Auckland, New Zealand

2001 Australian Meeting of the Econometric Society

New Zealand Treasury Invited Lecture

The Econometrics of Social Insurance


Innovations to these systems adopted by other countries.

After judging the pros and cons of various combinations of trial and error (a "groping process") and via society's have created social insurance institutions through

or fiscal impacts.

Predictions from economists about their welfare, behavioral, amount to "policy experiments without any formal experimentation. "

To date, almost all changes in social insurance policy systems in the coming decades.

There will be significant changes in the structure of these generation's social insurance systems make it likely that

The aging of the world and problems in the design of "first
actual outcomes?
models provide sufficiently realistic and reliable predictors of using millions of real human beings. But can computer
simulation models with millions of computerized “crash
proposed change to a social insurance system using

Claim: It is much more cost-effective to “crash test” a
computation mechanism design

I call this approach golden pigs.
using computer models instead of using humans as live
institutions: we should “crash test” a proposed change
used when contemplating changing social insurance
actually build a real vehicle. The same approach should be
test” new car designs using computer models before they
Automakers find it faster and more cost-effective to “crash
likelihood of misguided policy changes?

Would better simulation/forecasting models reduce the

Clinton/Kennedy patient bill of rights legislation.

by Bush, the Clinton health care proposals, and the
disability insurance program started by Reagan and abortion
dactual consequences (e.g. the aborted reternchment in the
policy makers had had better ex ante guidance about their
undertaken or would have been substantially changed if
insurance institutions which might not have been
well-intentioned actual and/or proposed changes to social

It is easy to point out a number of naive albeit
to invest millions to create better ones.

that their existing models are inadequate and have started
consequences of such reforms, they have begun to realize
how Congress about the disfavored and fiscal
agencies are facing increasing pressure to provide guidance
accounts (were not being contemplated. Now that these
structure of the Social Security (e.g., Individual
prior to the 1990s when significant or “radical” changes to
outdated models for policy anlaysis. This was not so bad
SSA and the CBO have made due with seriously flawed,
until very recently in the U.S. govt. agencies such as the
maintenance costs.

(b) the size of the up-front development costs and ongoing
(a) the degree of reliability/credibility in their predictions,
models and simulations depends on:
The degree to which policy makers should rely on computer
are not being developed and used by policymakers. Economics profession is the main reason why formal models are not being developed and used by policymakers. The lack of consensus in the forecasting models and value of developing adequate formal models is a possible solution. There are deep conflicts in the economics profession about the possibility and value of developing adequate formal models.

2. Economists have tried to develop good policy models before forecasting models.

1. Governments have ignored economists and have not invested adequate resources to develop good policy.

Some hypotheses:

Why have economists failed to provide formal guidance to policy makers?
Important effects of expectations and uncertainty are based on static models that fail to incorporate the analytically distributable consequences of policy changes, they are partly good for aspects of social insurance rules and are potentially good for the models have a relatively realistic treatment of certain methods using data sets that are now very out of date. While behavioral relationships that were estimated via reduced-form models are based on simple accounting identities and a few regarded as ad hoc and unreliable "black boxes". Most of these model used by the U.S. Social Security Administration) are now model currently used by the Urban Institute or the Corism to the work of Guy Orcutt, now embodied by the Dynamic

Problems with econometric policy models
Forecastsing and computational mechanisms design provide a reliable and cost-effective basis for policy approximations to observed behavior and will good approximations to observe these models. I believe these models will provide of the life cycle model. I argue that we are on the verge of being able to formally estimated and tested the full life cycle model, structurally, and currently I am not aware of anyone who has are much more complicated to solve and to estimate dynamics, expectations, and uncertainty. However, these models consistent utility maximization framework and can account for optimization problems. These models are based on a internally dynamic programming to solve individuals' life cycle. The current generation of dynamic structural models use
of estimating and using them, and most economists are incapable of understanding them, and most policy makers will not have any credibility because most econometric models, and the predictions of these models, are sharply(diminished) from much simpler models using "back of the envelope" calculations. There are gains from developing more complicated and realistic returns to developing more complicated and realistic models. All of the essential intuition can be gleaned from much simpler models, using "back of the modern "black boxes". They are too complicated. They are

The Feldstein/Grubber "back of the envelope" view: Counterpoint:
Policy evaluation

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

•

to 18 months and the members of the control group.

The experiment cost about $9100 per participant (1986 dollars) and lasted 44 months, with members of the demonstration (NSW) where low income individuals seeking job training were randomly assigned to „treatment“ and „control“ groups.

He used data from the National Supported Work Programs with Experimental Data „Evaluating the Economic Evaluations of Training„ began with Robert LaLonde’s 1986 AER article on "The \textit{structuralist/experimentalist} debate in econometrics"
Results (Lalonde AER 1986, p. 617). They still fail to replicate the experimental determinants econometric estimates pass conventional specification tests, however, even when the experimental results. Moreover, the economic estimates often differ significantly from precise impacts of the National Supported Work Program. Training programs would not have yielded accurate or comparable groups used to evaluate employment and He found that "many of the econometric procedures and
in the training program.
accounting for self-selection in application for participation predicted effects from Heckman-style wage regressions
Lalonde compared the experimental results with the
•
manpower training effects is premature:” (p. 395).

denunciation of nonexperimental methods for evaluating
limited study provide tangible evidence that the recent
experimental estimates. The empirical results from our
inference about the impact of the program as the
tests yield net impact estimates that lead to the same
rejected by these tests. Estimators not rejected by such
rejected by these tests. Estimators not rejected by such
tests, we are able to find estimators that are not
we eliminate a number of nonexperimental estimators by
estimators used in these studies can be rejected. Although
estimators, we show that a number of nonexperimental
appropriateness of alternative formulations of such

• Using several simple strategies for testing the

concluded that:

reinvestigated Lalonde's results using the NSV data. They
Heckman, Hotz and Dabos (HHD, 1987) Evaluation Review

•
"treatment effect" is the same for everyone.
the everyone receives the same "treatment" and the

2. The experimentalists adopt the naive assumption that
different "treatments" than can be considered.

1. Experiments are almost always extremely costly and take

methods for policy analysts:

inherent limitations in exclusively relying on experimental

The HHD article also pointed out a number of severe,

(421).

objective way to choose among competing models." (p.
nonexperimental methods are infeasible or that there is no
models produce wildly discordant estimates — not that
"The real lesson from [Lalonde’s] work is that invalidid
would operate and what its effects will be. Relevant to the question of how a real training program
cream-skimming through classical randomization have little
attempt to eliminate the effects of self-selection and
programs is endemic. EXPERIMENTAL predictions that
experiment. Self-selection and „cream skimming“ in training
more importantly, the „real world“ is not a controlled
•
can’t be fully controlled for experimentally.
practice levels of treatment vary randomly in a way that
of effort by participants and the training agency, so in
outcomes and one can’t experimentally control for the level
•
in practice, the treatment leads to a distribution of
answer only a limited subset of the interesting questions.’’

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

methods with random assignment of subjects to
various policy changes, then classical experimental
government funding is used to predict the impacts of
Since that time the Congress has mandated that it
won out in the U.S. Congress.
the Lalonde "anti-striucturalist" an "experimentalist" view
Despite the congency and clarity of Heckman's arguments,

can't be taken "seriously" for use in policy forecasting.

This problem implies that it is only feasible to solve highly
necessarily to the solve the model increase exponentially
entering the model (the computational requirements
increases as quantified by the number of "state variables"
and "realism" of a life-cycle, dynamic programming model
Richard Bellman in 1957, is that the as the level of detail

The Curse of Dimensionality: This problem, noted by

Other Problems with the dynamic, structural approach to

Policy Forecasting.
computational standpoint their effective dimensionality is \( I \). Problems are said to be strongly tractable and from a
difficulty as solving one dimensional problem. These high dimensional problems with essentially the same degree
methods and that there exist methods for solving extremely
to break the curse of dimensionality using deterministic
provided that with additional special structure, it is possible
Rust, Traub and Woźniakowski (Econométrica forthcoming) (Econométrica forthcoming) •
dimensionality, randomization can be used to break the curse of
dynamic programming problems (discrete choice DP
Rust (Econométrica 1997) proved that in certain classes of
different private insurance contracts and financial assets. 

health care, private pensions, and purchase of an array of 

housing decisions, career decisions, detailed models of 

models that include education, marriage/divorce, fertility,

will be possible to formulate and solve fairly detailed life cycle 

of detail to be of considerable use in policy forecasting. It 

be possible to develop life cycle models at a sufficient level 

With more significant investment in this area, it will soon

the life cycle models that we are able to solve numerically.

steadily increase the level of realism of „virtual reality“ in 

to parallel processing (and better algorithms we have been able

Via a combination of better hardware (including massive

supercomputers I used to use in the mid 1980s

laptop processes at a faster rate than the first generation

the speed of computer hardware at an exponential rate. My

Technologically, Moore’s Law has lead to improvements in
the historical data equally well. Various models in the equivalence class succeed in explaining determining which prediction is more likely to occur since the policy interventions, then it will be no objective basis for preferences and beliefs produce different predicted responses to that succeed in rationalizing the data. If different pairs of preferences and beliefs class with uncountably many pairs of preferences and beliefs corollary to this result is that there is generally an equivalence succeeded in “rationalizing” any observed behavior pattern. A succeeded in “rationalizing” any observed behavior patterns that possible to find some pair of preferences and beliefs that models are non-parametrically unidentified. That is, it is always Decision Processes” I proved that dynamic programming and Econometrica chapter “Structural Estimation of Markov

The identification problem: In my (1994) Handbook of
behavior and predicting behavioral responses (or other ad hoc methods of prediction) in summarizing
better job than any other comparable parsimonious model

The real test of a parametric model is whether it does a
policy interventions.

summarizing behavior and predicting behavioral responses

view a parametric model as a parsimonious way of
I realize that models are just approximations to reality, and
identifiable.

Generically, parametric classes of preferences and beliefs are
parameterized families. (we work with simple, parsimoniously
families of preferences and beliefs, but instead (due to the
In practice, econometricians don’t search over infinite

•

•

•

•

•
Impacts of other policy changes:

In the ex ante predictions from dynamic structural models, it provides policy makers with greater confidence that the model can be relied on to predict the control and small uncontrolled policy experiments. In a classic controlled experiments, the behavior responses are observed in closely mimicking the actual behavior of responses. If the environment that was not controlled for:

Treatment and not due to some other change in the treatment and the control groups is due to the more certain that changes in behavioral responses between models, in a classic controlled experiment we are much more structural models. Although uncontrolled policy test the predicted behavioral responses from dynamic views. Classical controlled experiments as an ideal way to
experiments since the Negative Income Tax
plains, it will constitute the largest controlled policy
WIA project to be implemented according to current
TWCWIA) If subsequent appropriations enable the
Ticket to Work Act and the Work Incentives Improvement
Social Security is undertaking in response to the 1999
I discuss a recent new controlled experiment that the U.S.

as arch enemies.
structuredists and experimentalists have viewed each other
ought to go hand in hand. However in practice,
this sense structural estimation and experimental methods
forecasts behavioral responses in controlled experiments. In

Thus the real test of a structural model is how well it
By the Social Security Administration, electronic for the use of more advanced life cycle policy models. These experiments could be the rapidly and cost-effectively. These experiments could be the much broader range of policy interventions much more activities and enable policy makers to analyze the impacts. Modeling and experimental methods can be complementary for policy analyses, and how a combination of structural limitations of an exclusive reliance on experimental methods limitations. I will use the TWWIIA Project to demonstrate the inherent models of retirement and social insurance that I have been opportunity to rigorously test the dynamic structural I view this "demonstration project" as a great scientific. I am one of the academic advisors to the TWWIIA project.
Outline for Rest of Talk

1. Review of the U.S. Disability Insurance and Old Age Insurance Program
2. Summary of the LWVIA Demonstration Project and Its Intellectual Origins
3. Review of Micro Panel Data on Retirement and Disability from the Health and Retirement Survey (HRS)
7. Conclusions

Dutch social insurance system.
the „annuity puzzle“ and Heyma (a structural model of the
6. Discussion of related models: Benitez-Silva (a solution to
test
c) predicting the impact of the elimination of the „earnings
project,
b) predicting the impact of „induced entry“ in the TVWIIA
a) assessing the winners and losers in privatization,
5. Illustration the value of the life cycle model for policy
4. Introduce a life cycle model of labor supply, consumption,
work period ends. Eventually have their benefits terminated after the trial.
percent of DI beneficiaries voluntarily return to work an to work without losing any benefits, fewer than 0.18 of 1 month trial work period where a DI beneficiary can return currently despite reasonably strong incentives including a 9 return to work.

to help create better incentives for disabled individuals to Two economists on the panel suggested some new policies headed by Jerry Mashaw of Yale law school: 1996 National Academy of Social Insurance (NASI) panel.

IMPROVEMENT ACT (TWIIWA)
origins of the 1999 TICKET to WORK and WORK INCENTIVES
significant disabilities who return to work. 

Credit (ETC) that would be payable to individuals with 
a tax credit similar to the existing Earned Income Tax 
proposed the use of a disabled worker tax credit. This is 
Richard Burkhauser, a labor economist at Cornell, 

Berkwitz, an emeritus economist from Rutgers,

maintain preretirement living standards in retirement.

saving averages only about one-third of that needed to

baby boomer retirement index is that boomers' retirement

simulation results. His primary finding, summarized in a

He then compares households' actual saving with the

earnings, age, social security, pensions and other factors. Consumption choices as a function of family size, education,

Skepticism of the Validity of the Life Cycle Model

• Bernheim (1992) "Models households' optimal saving and
...
nonetheless saving optimally for retirement.\textquoteleft\textquoteleft

some households have very low wealth earnings ratios are observationally equivalent. This distribution implies that wealth-earnings ratios among households that are earnings, the model generates a distribution of optimal accumulation benchmarks. Because of uncertainty of stochastic life-cycle model to generate optimal wealth. Our study differs from previous work in that it uses a Household Saving\textquoteleft\textquoteleft

Engen, Gale and Uccello, 1999\textquoteleft\textquoteleft The Adequacy of Model

Recent Work Supportive of the Validity of the Life Cycle
results are in fact largely consistent with ours:” (p. 142). Careful interpretation of previous work indicates that earlier interpretations provided in previous research. However, a glance, to be significant, is more optimistic than 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 significant more optimistic than the 50th percentiles of the wealth-earnings distribution exceed the simulation. In addition, households at the 75th and households exceed the median wealth-earnings ratios from this base specification, that more than half of applying the model to data from the HRS and SDF.
the case for this approach." Revision, if anything, represents a step backward in making
elsewhere to understand retirement decisions. And this
demonstrably add value to simpler efforts being undertaken
that thinking. The complicated modeling here does not
can, and that is not far enough. But I remain unconvinced
Reduced form models in this area have gone as far as they
ambitiously about the dynamic modeling of retirement.

Critique I: "There is clearly some merit in thinking

and Disability, "Benitez-Silva, "Dynamic Structural Models of Retirement
Buchinsky and
NIH Panel Evaluation of Proposal by Rust, Buchinsky and
been thus far somewhat limited."

that the general benefits to the research community have
not a reason to deny funding to the research. But it means
decisions. This may be because it is so difficult, and that is
investigators has used this framework to model retirement
mistaken, if remaining true that no one beyond this set of
simply asserting that it is not true. But, unless I am

The revised proposal does nothing to refute this beyond
adherence in either the research or policy communities.
approach was technically impressive but had found little
found troubling. The first is that this very complicated
previous round. There were three main criticisms that I
are remarkably unresponsive to the criticisms in the

Criticue 2: "This is a revised proposal, and the researchers
Critique 3: "The research agenda is founded upon a sound

It does not get nearly the attention it deserves."
models highly as a whole.‘"

Still, many economists do not view these
programming models is a state-of-the-art application of dynamic
proposals depend on whether one believes the dynamic

Critique 4: ‘Fundamentally, the assessment of this
Research questions:

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

Panel Recommendation: REJECT
Robert Moffitt, Johns Hopkins University
Jonathan Gruber, MIT
David Cutler, Harvard University
David Card, University of California at Berkeley

Economists on the NIH Review Panel: