

Model-Based or Design-Based?
Competing Approaches in
“Empirical Micro”

David Card, UC Berkeley

I. Three important background Issues

- models and causality
- working models in social sciences
- alternative meanings of "identification"

II. A brief history

III. Insights from the design-based approach

IV. Sources of conflict between approaches

V. What next?

I. Background Issues

a. Models and Causality

- a primary goal in economic research is to draw causal inferences. What role do models play in this process?
- one view (e.g., Heckman, 2008): *causality is model-based*: causality only exists within the framework of a theory that says 'x causes y'
- an opposing view (e.g., Holland): *causality is design-based*: a claim of causality requires that you can design a manipulation in which 'x causes y'

Between these extremes?

Haavelmo (1944) argued that a theoretical model

...will have an economic meaning only when associated with a design of actual experiments that describes -- and indicates how to measure -- a system of true variables X_1, X_2, \dots, X_n that are to be identified with the corresponding variables in the theory. (1944, p.8)

b. Working Models in Economics

- economic studies always have a “working model”
- framework that specifies the agent(s), the decision variables, the institutional setting
- the working model specifies (at least implicitly)
 - whose behavior are we modeling?
 - what are the constraints?
 - what are endogenous choice vars?
 - what variables are known/observed ?
 - what time horizon, level of aggregation, etc
- a successful working model causes a “suspension of disbelief”: readers/listeners narrow their focus

Nevertheless....

- exogenous variables in one working model are often the endogenous outcomes of another model (e.g., partial equilibrium vs. GE; political economy). Likewise, unobserved tastes/abilities etc. can be endogenous outcomes (peer effects, identify models, etc)
- wide debate about what is a legitimate working model
- trend in many fields is toward more complex working models that incorporate dynamic behavior, imperfect information, and GE (DSGE)

Three classes of working models in emp. micro:

a. explicit/approximating

Archetype: consumer demand/labor supply models of the 1950's through 1970s.

-theory implies $y_i = f_i(X_i)$, with restrictions on f_i

-researcher estimates approximating model

$$y_i = X_i\beta + \varepsilon_i \quad (\text{White, 1980})$$

-error “tacked on”: meas. error, specification error, ...

-focus on testing ‘general predictions’ of theory (e.g. Slutsky symmetry) and falsification (Popper)

b. explicit/exact

Archetype: McFadden's (1974) multinomial logit

- model gives complete d.g.p. ("all causes" model)
e.g. utility function + unobserved heterogeneity in tastes
yields complete demand system
- focus on estimating model parameters **and** distribution of unobservables
- requires specific functional forms, but once estimated can be used for prediction and welfare analysis

c. implicit working model

Archetypes:

1) program evaluations (training, student loans, etc)

2) studies of non-market behavior

-e.g., Levitt (1997) police and crime

-no precisely specified behavioral model. Focus instead on a “structural” outcome equation with a signed coefficient

-can measure causal effect of existing policy, but cannot extrapolate very far (e.g. Mariel Boatlift; ALMP evals.)

Background Issues, continued

c. The two meanings of identification

a) “econometric” identification. Exclusion/ functional form assumptions that invert the data to the model parameters

- examples: linear S+D models (Working, 1927)
CES production function (Fisher, 1969)
- recent work (e.g. Imbens-Angrist LATE) extends the inversion to models with heterogeneous responses
- partial identification: invert to a set or region

b) substantive identification

- ruling out alternative explanations for the data (i.e., get past the 'identification police')
- gold standard = randomized design
 - even with RCT: did randomization work? partial compliance? sample attrition?
- for observational data: tests and checks
 - over-id testing (Sargan, Newey, GMM)
 - “pre-treatment” tests (event studies/diff-in-diffs)
 - falsification exercises
- more art than science (Popper again)

II. A Brief History

pre-1970

- most econometric modeling focused on macro data, using systems of linear equations
- problems of identification stressed by T.C. Liu (1955, 1960), Sargan (1964); others.

In labor economics:

- H. Gregg Lewis and union wage effects
- late 1960s: first large scale micro data samples
- NIT demonstrations (debate over design)

1970s:

- micro data and computer power
- McFadden's logit model:
 - fully specified behavioral model (stochastic and deterministic components). Used to predict effect of BART system *before it was built*

In labor economics:

- labor supply studies (mainly explicit/approximating)
NIT's analyzed using 'structural' methods
- microeconomic studies of union wage premium
- Heckman's sample selection model

Early 1980s

The “crisis” in econometric practice

- Lucas (1976) critique
- Hendry (1980) “Alchemy or Science”
- Leamer (1983) “specification searches”
- failure of replication efforts (JMCB study)

In labor economics

- Lalonde (1985) critique of program eval. Methods
- Lewis’ (1986) review **dismisses** IV and Heckit methods for estimating union wage gaps

1980s/1990s

Emergence of the “design-based” approach:

- random instruments (draft lottery)
- quarter-of-birth; *arguably exogenous* instruments
- quasi-experiments - Mariel boatlift, NJ min. wage
- RD (Angrist-Lavy, Lee's vote share idea)

Markers of Success/Influence

- adoption of design-based approaches in public finance, development econ, econ. history
- explosion in lab-based and field experiments

III. Insights from Design-Based Approach

a) counterfactuals and causality

- identification equated with research design (Haavelmo)
- research design defines the counterfactual
- randomization is 'gold standard' for asserting causality
- “credible” instruments

b) substantive specification testing

- randomization tests in RCT's
- pre-treatment comparisons, falsification exercises, multi-way differencing for observational designs

c) replicability

- growing emphasis on “replicability” as a condition for credibility
- most top journals (QJE excepted) now require authors to deposit programs and data
- replication is still a (very) risky enterprise

Two other implicit messages

1. channels in standard economic models often explain only a small part of observed behavior

- “small responders” (true elasticities are small)
- “long run” vs “short run” changes
- sluggish institutional adaptation
- inattention, procrastination, other “failures”

2. straightforward predictions don’t always hold up

- minimum wage studies
- Dinardo-Lee union vote study
- Medicare eligibility and disparities in health care use

IV. Conflict between model-based and design-based approaches

Background factors

1. DSGE model becomes the working model in macro. (Lucas, Prescott, ...). Models that ignore dynamics, imperfect information, and GE are seen as incomplete and potentially misleading

2. re-emergence of empirical research in IO, heavily influenced by McFadden's logit framework

- limited data and sources of exogenous variation (e.g., only market shares directly observable)
- relatively complete model needed for anti-trust policy analysis, and analysis of dynamic games (entry)
- structural model needed to infer much from observed bids in auctions

3. Developments in econometric methodology

- mixed logit and extensions (BLP)
- Rust, Hotz-Miller, ... adaptation of nested logit to intertemporal context
- approximating methods for solving DP's
- advances in simulation/integration techniques
- partial identification

Key Areas of Disagreement

a. Completeness/complexity of the working model

- design-based papers use simplified working models (e.g., single equation causal model) with no formal derivation of the DGP for the observed data. Emphasis is on “sharp” predictions and falsification of theories.
- structural papers follow McFadden program: derive DGP from assumed model (though estimation may only use moment conditions). Emphasis on sophisticated agents, equilibrium behavior.

b. What are the “parameters of interest”?

- design-based approach estimates $E[y | X] = X\beta$,
with focus on “clean identification” of β
- model-based approach poses model $g(y, x, e; \theta) = 0$
with focus on estimating θ , sometimes using the implied
linear projection $L(y | X) = X\beta(\theta)$
- which is more useful: β or θ ?
- Is θ itself endogenous? (e.g., welfare takeup model)

c. Identification

- design-based practitioners emphasize "credible" identification, with testing/evaluation of assumptions

- model-based studies are often identified by combination of stochastic assumptions, functional form and exclusion restrictions, and “outside information” (e.g. interest or discount rates are calibrated).

d. Inherent confidence in the working model

(Is model falsification a valid pursuit?)

- some design-based studies attempt to test basic predictions. E.g., min wage studies, lab experiments
- in program evaluations: null hypothesis of a zero effect => "proof of concept"

-model-based studies usually stipulate a relatively complete model. Goal of research is to estimate parameters and conduct welfare analysis.

(Some attempts at out-of sample prediction)

e. Choice of Topic

- design based approach has crossed over into many areas on the fringe of economics (including non-neoclassical topics/questions) - e.g. peer effects, happiness
- model-based approach tends to emphasize neoclassical topics, in part because of need for model development
- nevertheless: some model-based behavioral economics (DLM charitable giving study)

V. What Next?

Three key trends:

1. the rise of experiments; the growing backlash
2. behavioral economics and the limits of "uber rationality"
3. normative vs. positive economics

1. The rise of experiments and the growing backlash

- lab and field-based studies with purposeful randomized design represent a growing fraction of papers published in top journals

2010: lab experiments = 10% of papers in top 5
field experiments = 2.5%

- many other papers are "sold" as experiments
 - e.g.: - Bandiera, Barankay, and Rasul (2007, 2009)
 - Chetty, Looney and Kroft (2009)
 - both pre-post designs with no control group

Backlash:

- e.g., Heckman and Smith (1995); Deaton (2010)
- limitations of randomized field experiments:
 - flaws in designs (e.g., incomplete compliance)
 - generalizeability (bike messengers?)
 - lack of theoretical foundations
 - uninterpretable reduced form parameters
 - distributions of treatment effects
- what questions are economists supposed to study?
 - should we be doing experiments on (e.g.) pedagogical methods? If so, should we change the Ph.D. “core”?

2. Behavioral economics and limits to "uber-rationality"

- 1940s: Lester-Machlup debate
- 1950's: Simon and others: do firms really maximize profits?
- 1970s: Solow and others - concerns over 'rational expectations' as basis for macro policy
- recent decade:
 - defaults and reference points
 - inattention and other cognitive failures
 - decision theory: sharp departures from v-N&M

Behavioral models often posit simple “rules of thumb”

- state-of-art model based studies often assume simple preferences over time and risks (e.g., linear preferences with .95 discount rate) and sophisticated agents
- is it useful to test these maintained assumptions?
 - e.g.: - testing for credit constraints
in education choices
- does rejection of the maintained assumptions threaten the validity of the analysis?
- are some agents naïve, others sophisticated?
- do institutions etc evolve to support sophisticated decision-making?

3. Normative vs. Positive Economics

- economists are increasingly interested in normative questions; less interested in “basic science”
- model based approach is inherently suited to normative analysis: policy counterfactuals, etc.
- design-based research building on opportunistic designs often cannot identify the full set of parameters needed for policy analysis
- carefully designed lab/field experiments can go further

Final Thoughts

- Unlike physics, there is no ‘standard model’ of microeconomic behavior
- Many basic premises in conventional working models – such as strongly separable preferences (over time and states) – are probably wrong for many/most agents
- There is a need for many kinds of applied work:
 - “test the model”
 - “fit the best available model”
 - “descriptive work” that can point to new models