The Role of Models in “Reduced Form” Empirics

Peter Hull

Brown University
Spring 2024

Model-Free Empirics?

- The so-called "Credibility Revolution" (Angrist & Pischke 2010) has forever changed empirical practice in microeconomics
- In the '80s, economists would carefully/parametrically model an individuals' decision to select into a treatment (e.g. job training) and use this structure to fight selection bias (Heckman 1974, 1976, 1979)
  - Sometimes the resulting estimates were "preposterously large or outlandishly negative" (Lewis 1986), and not always taken seriously
- Angrist, Ashenfelter, Card, Krueger, Pischke, and others changed the game by thinking hard about *variation*, instead of model structure
  - E.g. find a series of as-good-as-random shocks (e.g. draft lottery #s) or a "clean" policy change (e.g. NJ minimum wage increase)
- Sometimes this approach is termed "reduced form" and viewed as "model-free".... but is it really?

Outline for Today

1. OLS & IV Review

2. Models for Specification

3. Models for Extrapolation

4. Models for Identification

Causal Models

- Let X_i be a "treatment" of interest and Y_i be an "outcome." Suppose:

$$Y_i = \beta X_i + \varepsilon_i \tag{1}$$

specifies how Y_i is generated from X_i across units (e.g. individuals) i. Of interest: the *model parameter* β

- **Ex. 1** *(Potential outcomes)* Let $D_i \in \{0,1\}$ indicate treatment receipt and let $Y_i(0)$ and $Y_i(1)$ be i's outcome when $D_i = 0$ or $D_i = 1$. Then:

$$Y_i = Y_i(0)(1 - D_i) + Y_i(1)D_i = (Y_i(1) - Y_i(0))D_i + Y_i(0)$$

so (1) follows when effects $Y_i(1) - Y_i(0) = \beta_i$ are constant across i

  - **Q:** What is a way the general potential outcomes model is restrictive?

Causal Models (Cont.)

- **Ex. 2** *(Production function)* Let Y_i measure firm i's output, L_i measure labor, K_i measure capital, and suppose:

$$Y_i = Q_i K_i^{\alpha} L_i^{\beta}$$

  for unobserved TFP shocks Q_i. Suppose $\alpha = 1 - \beta$ to get to (1):

$$\ln(Y_i/K_i) = \beta \ln(L_i/K_i) + \ln Q_i$$

- **Q:** Suppose we have data on (Y_i, K_i, L_i) across firms. What would we need to be true in order to estimate β?

Natural Experiments

- Given our model, $Y_i = \beta X_i + \varepsilon_i$, suppose we know (or assume) that X_i is drawn at random conditional on ε_i
  - E.g. we flip a fair coin for each i, regardless of ε_i, to determine X_i

- Regression slope coefficient:

$$\frac{Cov(X_i, Y_i)}{Var(X_i)} = \frac{Cov(X_i, \beta X_i + \varepsilon_i)}{Var(X_i)} = \beta + \frac{Cov(X_i, \varepsilon_i)}{Var(X_i)} = \beta$$

  I.e. the regression of Y_i on X_i *identifies* the model parameter β

- More generally, if $E[\varepsilon_i \mid X_i, W_i] = E[\varepsilon_i \mid W_i]$ for some vector W_i, a regression of Y_i on X_i that flexibly controls for W_i identifies β
  - How much does this result rely on our constant-fx model $Y_i = \beta X_i + \varepsilon_i$?

The Robustness of Regression

- Angrist (1998) first showed that a regression on a binary treatment X_i and strata dummies W_i identifies a convex average of treatment effects when X_i is as-good-as-randomly assigned within strata

- More generally, suppose $Y_i = \beta_i X_i + \varepsilon_i$ with $X_i \perp\!\!\!\perp (\beta_i, \varepsilon_i) \mid W_i$. The regression of Y_i on X_i w/controls spanning $\mu_i = E[X_i \mid W_i]$ identifies:

$$\beta^{OLS} = E[\sigma_i^2 \beta_i]/E[\sigma_i^2]$$

  where $\sigma_i^2 = Var(X_i \mid W_i)$ (see, e.g., Borusyak and Hull 2024)

- Thus, regression identifies a sensible convex average of causal effects even when the constant-effect model motivating it is misspecified

What About IV?

- Back to the model $Y_i = \beta X_i + \varepsilon_i$. Suppose we don't think X_i is as-good-as-randomly assigned but we do think another Z_i is
  - We also think Z_i only affects Y_i through X_i, making $Cov(Z_i, \varepsilon_i) = 0$
- IV slope coefficient:

$$\frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)} = \frac{Cov(Z_i, \beta X_i + \varepsilon_i)}{Cov(Z_i, X_i)} = \beta + \frac{Cov(Z_i, \varepsilon_i)}{Cov(Z_i, X_i)} = \beta$$

  I.e. the IV regression of Y_i on X_i *identifies* the model parameter β

- More generally, if $E[\varepsilon_i \mid Z_i, W_i] = E[\varepsilon_i \mid W_i]$, an IV regression of Y_i on X_i that flexibly controls for W_i and instruments with Z_i identifies β
  - How much does this result rely on our constant-fx model $Y_i = \beta X_i + \varepsilon_i$?

The Robustness of IV

- Imbens and Angrist (1994) first showed that an IV regression with binary X_i and binary Z_i identifies a *local average treatment effect* (LATE): the avg effect for those who comply with the instrument
  - Requires a monotonicity condition: nobody defies the instrument

- More generally, suppose $Y_i = \beta_i X_i + \varepsilon_i$ with $Z_i \perp\!\!\!\perp (\beta_i, \varepsilon_i) \mid W_i$. The regression of Y_i on X_i w/controls spanning $\mu_i = E[Z_i \mid W_i]$ identifies:

$$\beta^{IV} = E[\sigma_i^2 \pi_i \beta_i]/E[\sigma_i^2 \pi_i]$$

where $\sigma_i^2 = Var(Z_i \mid W_i)$ and $\pi_i = \frac{Cov(Z_i, X_i | W_i, \beta_i)}{Var(Z_i | W_i, \beta_i)}$ (see again BH '24)

- Thus, as long as $\pi_i \geq 0$ (monotonicity), IV regression is robust to misspecification of the constant-effects model motivating it

Models for Specification

- Even so far, we've seen a role for models in "reduced-form" analyses: specifying the target parameter
  - OLS is remarkably robust to unspecified heterogeneity so long as the treatment is as-good-as-randomly assigned and the controls are flexible
  - IV too, as long as exclusion + monotonicity hold
- But as-good-as-random with respect to what? The model is also needed for saying what ε_i is / what you should be concerned about
  - E.g. unobserved TFP shocks in example 2
- Models might also point to other ways basic OLS/IV specs fall short
  - E.g. spillovers, violating even the general potential outcomes model

Modeling Spillovers

- Suppose a set of new federal subsidies/tariffs g_k are as-good-as-randomly assigned across industries k
  - They shift labor demand; we want to use them to estimate labor supply
  - Why wouldn't we want to just regress industry wages on industry employment, instrumenting with g_k?

- General equilibrium models tell us worker mobility causes spillovers: g_k affects wages/employment of *all* industries within local markets i

- Suppose these models specify regional supply curves $Y_i = \beta_i X_i + \varepsilon_i$ for wages Y_i, employment X_i, and unobserved supply shocks ε_i
  - Can we still use the national g_k shocks to say something about β_i?

Shift-Share Instruments

- Consider $Z_i = \sum_k s_{ik} g_k$ for some "exposure shares" s_{ik} with $\sum_k s_{ik} = 1$
  - E.g. baseline industry employment shares, prior to subsidies g_k

- Borusyak et al. (2022) show how Z_i can be a valid instrument for X_i in the location-level IV regression of Y_i, even with endogenous s_{ik}
  - Intuition: imagine full industry specialization, i.e. $s_{ik} \in \{0,1\}$. Then it's like we have a "clustered" RCT with industry-level shocks
  - More generally, need g_k to be as-good-as-random with respect to a share-weighted average of the ε_i

- The shift-share Z_i acts as a "translation device," bringing the exogenous industry-level shocks to the specified location-level model
  - See Borusyak and Hull (2023) for a generalization of this approach, for general "formula instruments" $Z_i = f_i(s, g)$

Example: Market Access

- Models of economic geography tell us transportation upgrades affect local outcomes (e.g. land value) by increasing market access (MA):

$$\Delta \log V_i = \beta \Delta \log MA_i + \varepsilon_i, \quad (2)$$

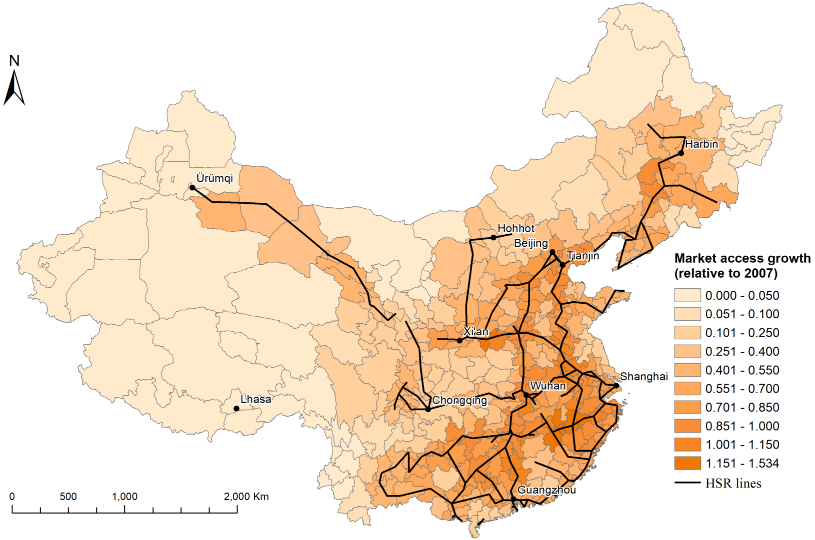
$$\text{where } MA_{it} = \sum_j \tau(g_t, loc_i, loc_j)^{-1} pop_j, \quad (3)$$

  for railroad network g_t in periods $t = 1, 2$, region locations loc_j (co-determining travel cost τ), and regional population pop_j

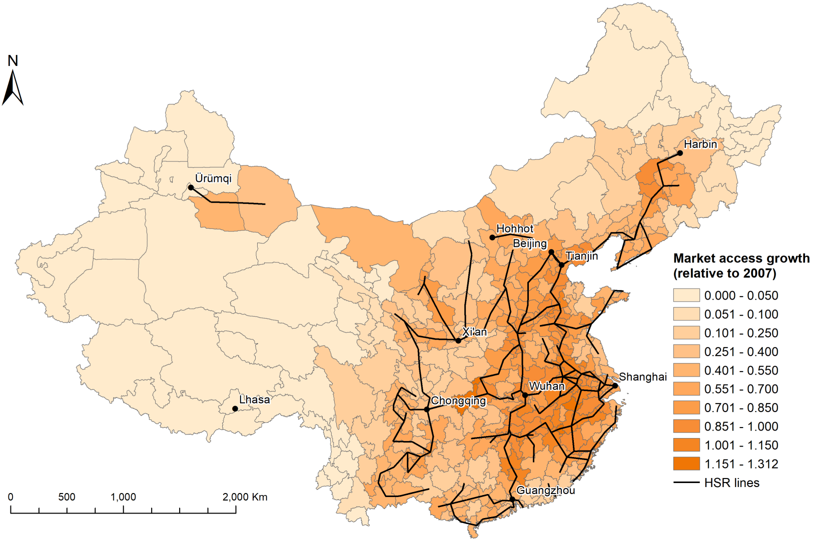
- Suppose we have a natural experiment in g_t: e.g. timing of new line construction is as-good-as-random conditional on plans
  - Further assume these shocks don't directly affect local "amenities" ε_i

- Could use some shift-share $z_i = \sum_k w_{ik} \mathbf{1}[\text{line } k \text{ opens}]$ for some w_{ik}
  - Or could use the model more directly: z_i as predicted MA growth from predetermined variables + opening shocks
  - BH '23 show how you can "recenter" these predictions to just use the exogenous shocks for identification....

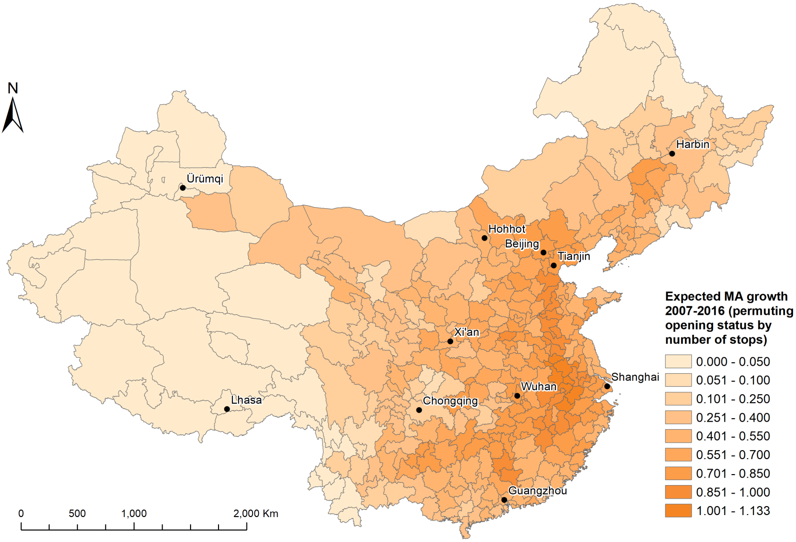
BH '23: Market Access Growth from HSR Construction



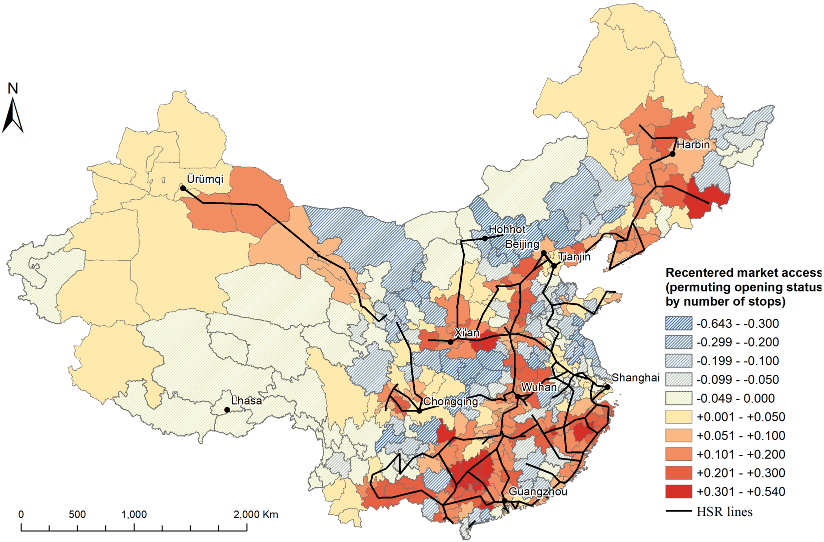
BH '23: Counterfactual HSR Line Openings + MA Growth



BH '23: Expected MA Growth Over Counterfactuals



BH '23: "Recentered" (Realized - Expected) MA Growth



BH '23: Estimated Market Access Effects

	Unadjusted OLS (1)	Recentered IV (2)	Controlled OLS (3)
Panel A. No Controls			
Market Access Growth	0.232	0.081	0.069
	(0.075)	(0.098)	(0.094)
		[-0.315, 0.328]	[-0.209, 0.331]
Expected Market Access Growth			0.318
			(0.095)
Panel B. With Geography Controls			
Market Access Growth	0.132	0.055	0.045
	(0.064)	(0.089)	(0.092)
		[-0.144, 0.278]	[-0.154, 0.281]
Expected Market Access Growth			0.213
			(0.073)
Recentered	No	Yes	Yes
Prefectures	274	274	274

Regressions of log employment growth on log market access growth in 2007–2016.
Spatial-clustered standard errors in parentheses; permutation-based 95% CI in brackets

Models for Extrapolation

- Consider the simplest version of Imbens and Angrist (1994): binary D_i, binary Z_i, no controls. IV identifies:

$$\beta^{IV} = E[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)]$$

  where $D_i(z)$ denotes potential treatment when $Z_i = z$

  - LATE: average treatment effect $(Y_i(1) - Y_i(0))$ among compliers (those with $D_i(1) = 1, D_i(0) = 0$)

- Without restrictions on $(Y_i(1), Y_i(0), D_i(1), D_i(0))$, can't say anything more: IV only reveals effects among i whose D_i is shifted by Z_i
  - Actually not quite true: can identify avg. $Y_i(1)$ of always-takers (w/ $D_i(1) = D_i(0) = 1$), avg. $Y_i(0)$ of never-takers (w/ $D_i(1) = D_i(0) = 0$), as well as avg. $Y_i(1)$ & $Y_i(0)$ separately for compliers
  - By adding a (semi-)parametric model of selection, we can extrapolate these objects to identify other parameters, e.g., ATE $E[Y_i(1) - Y_i(0)]$

Adding Structure to IV

Suppose we have a Z_i which is as-good-as-randomly assigned + excludable

- Assume a distribution for $(Y_i(1), Y_i(0), v_i)$ where $D_i = \mathbf{1}[\mu + \pi Z_i > v_i]$
- Then we have parametric models for

$$
\begin{aligned}
E[Y_i \mid D_i = 1, Z_i = z] &= E[Y_i(1) \mid \mu + \pi z > v_i] \equiv f_1(z;\theta) \\
E[Y_i \mid D_i = 0, Z_i = z] &= E[Y_i(0) \mid \mu + \pi z < v_i] \equiv f_0(z;\theta)
\end{aligned}
$$

  as well as "first stage" models for $Pr(D_i = 1 \mid Z_i = z) = g(z;\theta)$

- With enough variation in Z_i, the parameter vector θ (and thus ATE) can be identified from these moment restrictions

Key point: the model allows us to extrapolate "local" IV variation to estimate more "policy relevant" parameters

- When Z_i has limited support, the model is doing more "work"
- With full support, we have "identification at infinity" (w/o a model)

Linking Back to LATE

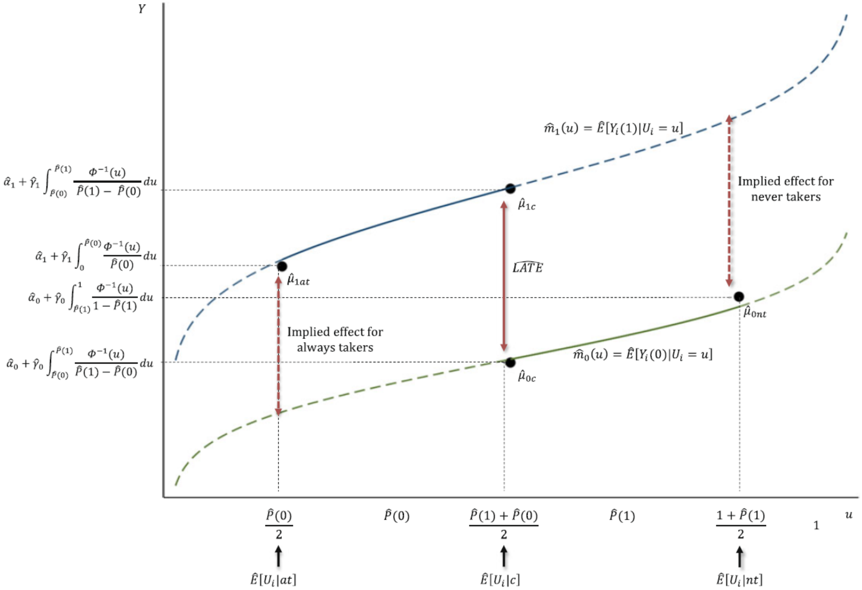
Kline and Walters (2019) formalize this extrapolation logic in the familiar Imbens and Angrist (1994) setup

- Key result: in simple binary Z_i / no controls setup, control function estimates of LATE are numerically identical to linear IV
- "Differences between structural and IV estimates therefore stem in canonical cases entirely from disagreements about the target parameter rather than from functional form assumptions" (p. 678)
- Functional form instead shapes the extrapolation to other parameters

They conclude with a nice point about validating "structural" models:

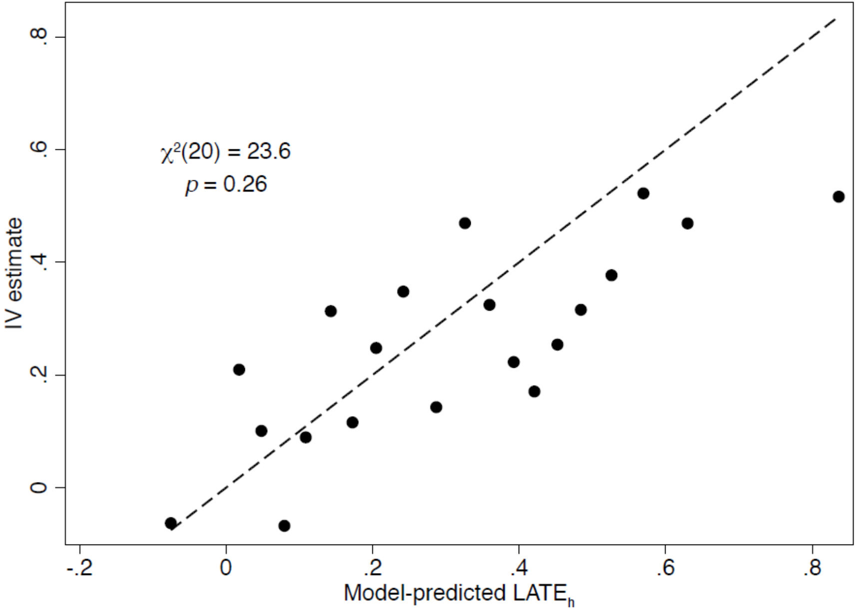
- "Comparing the model-based LATEs implied by structural estimators with unrestricted IV estimates provides a transparent assessment of how conclusions regarding a common set of behavioral parameters are influenced by the choice of estimator" (p. 678)

Heckit Extrapolation of IV Moments



"Heckit" model: $E[Y_i(d)|U_i] = \alpha_d + \gamma_d \Phi^{-1}(U_i)$

Validating Structural Models: Kline and Walters (2016)



Models for Identification

- So far, we've assumed we have as-good-as-randomly assigned and excludable "shocks" and shown how models can help us use them
  - No need to restrict how model unobservables (e.g. potential outcomes) relate to each other or observables (e.g. demographics), except if useful for extrapolating "local" variation to more policy-relevant parameters

- When we do not have such shocks, certain restrictions on unobservables can yield (relatively) transparent identification
  - Popular example: parallel trends. Sse in a panel of $t \in \{0,1\}$ we assume $E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0]$
  - Then in a canonical diff-in-diffs setup:

$$
\begin{aligned}
&E[Y_{i1} - Y_{i0} \mid D_i = 1] - E[Y_{i1} - Y_{i0} \mid D_i = 0] \\
&= E[Y_{i1}(1) - Y_{i0}(0) \mid D_i = 1] - E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0] \\
&= E[Y_{i1}(1) - Y_{i0}(0) \mid D_i = 1] - E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] \\
&= E[Y_{i1}(1) - Y_{i1}(0) \mid D_i = 1] = ATT
\end{aligned}
$$

"Model-Based" Identification

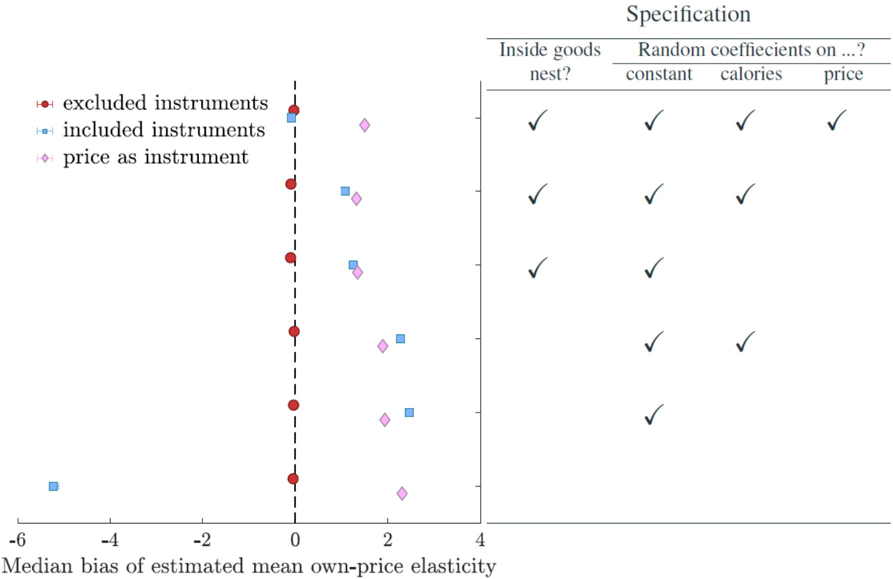
- Notice how $E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid D_i = 0]$ is a different flavor of identifying assumption: it models unobservables
  - As a result, DiD is highly sensitive to the chosen functional form of Y_i (Roth and Sant'Anna 2023)
  - This difference is also the source of the recent "negative weights" freakout for TWFE regression (Goldsmith-Pinkham et al. 2024)

- General "model-based" ID strategy: for causal model $Y_i = \beta X_i + \varepsilon_i$, sse $E[\varepsilon_i \mid X_i, W_i] = W_i'\gamma$. Then $Y_i = \beta X_i + W_i'\gamma + v_i$ is a regression
  - E.g. if $i = (j,t)$ indexes a panel with W_i including j and t FE, $Y_i = \beta X_i + W_i'\gamma + v_i$ is TWFE and $E[\varepsilon_i \mid W_i] = W_i'\gamma$ is "parallel trends"
  - This works just as well for IV: suppose $E[\varepsilon_i \mid Z_i, W_i] = W_i'\gamma$. Then regressing Y_i on X_i instrumenting w/Z_i and controlling for W_i ID's β
  - Works for fancier models too: e.g. BLP models for demand sometimes assume product unobservables are mean-zero conditional on observed characteristics $\rightarrow$ "BLP instruments"

The Fragility of "Internal" Instruments

- Andrews et al. (2023) show instruments constructed from restrictions like $E[\varepsilon_i \mid Z_i, W_i] = W_i'\gamma$ are inherently sensitive to model structure
  - In contrast to "strongly excluded" instruments that are as-good-as-randomly assigned + excludable from potential outcomes
- Specifically, they show strong exclusion is sufficient and (essentially) necessary for structural IV to satisfy *sharp-zero consistency*:
  - If true causal effects are zero, the IV estimand reflects this
- Borusyak and Hull (in progress) show how "recentering" can ensure strong exclusion in structural models, given exogenous shocks
  - Stay tuned...

Andrews et al. (2023) BLP Simulations



Note: simulations based on Miller and Weinberg (2017)

Takeaways

- "Reduced-form" estimation is not "model-free"
  - At minimum, we need models to guide our choice of Y_i and X_i, and to assess when "exogeneity" of X_i or Z_i holds via specification of ε_i
  - Given a well-specified model and natural experiment, we can identify key model parameters (e.g. by recentering)
- Models can help us extrapolate "local" estimands to more generalizable or policy-relevant parameters
  - Great to visualize/test the extrapolation, if possible
- Restrictions on model unobservables can subsitute for clean "design" of exogneous observed shocks
  - But beware: estimates are more model-dependent and may be less robust to seemingly harmless deviations (e.g. heterogeneous effects)