Day 3: On Formulas and Models

Peter Hull

Design-Based Regression Inference
Fall 2024

Outline

1. Formula Treatments/Instruments
2. Nonlinear/Structural Models

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
  - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
  - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold
- As it turns out, this gives us a way to robustly estimate causal effects with "formula" treatments/instruments, constructed from some exogenous shocks and other non-random variables

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
  - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold

- As it turns out, this gives us a way to robustly estimate causal effects with "formula" treatments/instruments, constructed from some exogenous shocks and other non-random variables. E.g.:
  - Interacted treatments: $x_i = s_i g_i$, with exogenous shocks g_i

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
  - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold

- As it turns out, this gives us a way to robustly estimate causal effects with "formula" treatments/instruments, constructed from some exogenous shocks and other non-random variables. E.g.:
  - Interacted treatments: $x_i = s_i g_i$, with exogenous shocks g_i
  - Shift-share instruments: $z_i = \sum_k s_{ik} g_k$, with exogenous shocks $\{g_k\}_{k=1}^K$

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
  - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold

- As it turns out, this gives us a way to robustly estimate causal effects with "formula" treatments/instruments, constructed from some exogenous shocks and other non-random variables. E.g.:
  - Interacted treatments: $x_i = s_i g_i$, with exogenous shocks g_i
  - Shift-share instruments: $z_i = \sum_k s_{ik} g_k$, with exogenous shocks $\{g_k\}_{k=1}^K$
  - Spillover treatments: e.g. number of neighbors treated in an RCT

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
  - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold

- As it turns out, this gives us a way to robustly estimate causal effects with "formula" treatments/instruments, constructed from some exogenous shocks and other non-random variables. E.g.:
  - Interacted treatments: $x_i = s_i g_i$, with exogenous shocks g_i
  - Shift-share instruments: $z_i = \sum_k s_{ik} g_k$, with exogenous shocks $\{g_k\}_{k=1}^K$
  - Spillover treatments: e.g. number of neighbors treated in an RCT
  - Instruments for policy eligibility, combining exogenous policy shocks & non-random measures of policy exposure (e.g. income/family structure)

Beyond Simple Treatments/Instruments

- We've seen how a regression of y_i on x_i and w_i identifies a convex average of treatment effects when $E[x_i \mid y_i(\cdot), w_i] = E[x_i \mid w_i] = w_i'\gamma$
    - IV version: $E[z_i \mid y_i(\cdot), w_i] = w_i'\gamma$ and exclusion/monotonicity hold

- As it turns out, this gives us a way to robustly estimate causal effects with "formula" treatments/instruments, constructed from some exogenous shocks and other non-random variables. E.g.:
    - Interacted treatments: $x_i = s_i g_i$, with exogenous shocks g_i
    - Shift-share instruments: $z_i = \sum_k s_{ik} g_k$, with exogenous shocks $\{g_k\}_{k=1}^K$
    - Spillover treatments: e.g. number of neighbors treated in an RCT
    - Instruments for policy eligibility, combining exogenous policy shocks & non-random measures of policy exposure (e.g. income/family structure)

- Let's build up to these slowly...

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

- The design-based approach says we need to adjust for:

$$E[x_i \mid s] =$$

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

- The design-based approach says we need to adjust for:

$$E[x_i \mid s] = E[s_i g_i \mid s] =$$

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

- The design-based approach says we need to adjust for:

$$E[x_i \mid s] = E[s_i g_i \mid s] = s_i E[g_i] =$$

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

- The design-based approach says we need to adjust for:

$$E[x_i \mid s] = E[s_i g_i \mid s] = s_i E[g_i] = s_i \mu$$

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

- The design-based approach says we need to adjust for:

$$E[x_i \mid s] = E[s_i g_i \mid s] = s_i E[g_i] = s_i \mu$$

I.e. need to control for s_i to just leverage the random variation in g_i

Interacted Treatments

- Suppose $g_i \mid y(\cdot), s \overset{iid}{\sim} G$ for some observed s_i
  - E.g. g_i drawn in a simple RCT, with s_i being a baseline characteristics
  - We are interested in the effects of $x_i = s_i g_i$: e.g. heterogeneous effects
  - Perhaps estimated alongside direct effects of g_i (but don't worry about multiple treatments for now)

- The design-based approach says we need to adjust for:

$$E[x_i \mid s] = E[s_i g_i \mid s] = s_i E[g_i] = s_i \mu$$

I.e. need to control for s_i to just leverage the random variation in g_i

  - Can use $s_i g_i$ as an IV controlling for s_i, given exclusion/monotonicity

Interacted Treatments (Cont.)

- Now suppose $g_i \mid y(\cdot), s, q \overset{iid}{\sim} G(q_i)$: e.g., a stratified RCT
  - Again, we want to estimate the effects of $x_i = s_i g_i$ (or use it as an IV)

Interacted Treatments (Cont.)

- Now suppose $g_i \mid y(\cdot), s, q \overset{iid}{\sim} G(q_i)$: e.g., a stratified RCT
  - Again, we want to estimate the effects of $x_i = s_i g_i$ (or use it as an IV)

- Now, for $\mu(q_i) = E[g_i \mid q_i]$:

$$E[x_i \mid s, q] = E[s_i g_i \mid s, q] = s_i E[g_i \mid s, q] = s_i \mu(q_i)$$

Interacted Treatments (Cont.)

- Now suppose $g_i \mid y(\cdot), s, q \overset{iid}{\sim} G(q_i)$: e.g., a stratified RCT
  - Again, we want to estimate the effects of $x_i = s_i g_i$ (or use it as an IV)

- Now, for $\mu(q_i) = E[g_i \mid q_i]$:

$$E[x_i \mid s, q] = E[s_i g_i \mid s, q] = s_i E[g_i \mid s, q] = s_i \mu(q_i)$$

I.e. need to control for s_i interacted with a flexible function of q_i

  - E.g. the interactions of s_i and strata fixed effects

Interacted Treatments (Cont.)

- Now suppose $g_i \mid y(\cdot), s, q \overset{iid}{\sim} G(q_i)$: e.g., a stratified RCT
  - Again, we want to estimate the effects of $x_i = s_i g_i$ (or use it as an IV)
- Now, for $\mu(q_i) = E[g_i \mid q_i]$:

$$E[x_i \mid s, q] = E[s_i g_i \mid s, q] = s_i E[g_i \mid s, q] = s_i \mu(q_i)$$

  I.e. need to control for s_i interacted with a flexible function of q_i

  - E.g. the interactions of s_i and strata fixed effects
- Key point: the design of exogenous shocks g_i + knowledge of the "formula" $s_i g_i$ tells us what controls are needed for identification

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i

- What is the "expected instrument"?

$$E[z_i \mid s] =$$

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i

- What is the "expected instrument"?

$$E[z_i \mid s] = E\left[\sum_k s_{ik} g_k \mid s\right] =$$

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i
- What is the "expected instrument"?

$$E[z_i \mid s] = E\left[\sum_k s_{ik} g_k \mid s\right] = \sum_k s_{ik} E[g_k] =$$

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i

- What is the "expected instrument"?

$$E[z_i \mid s] = E\left[\sum_k s_{ik} g_k \mid s\right] = \sum_k s_{ik} E[g_k] = \left(\sum_k s_{ik}\right) \mu$$

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i

- What is the "expected instrument"?

$$E[z_i \mid s] = E\left[\sum_k s_{ik} g_k \mid s\right] = \sum_k s_{ik} E[g_k] = \left(\sum_k s_{ik}\right)\mu$$

  - I.e. need to control for the "sum of shares" $w_i = \sum_k s_{ik}$ (which may be one, in which case no controls needed!)

Shift-Share Instruments

- Now suppose the shocks $g_k \mid y(\cdot), s \overset{iid}{\sim} G$ vary at a different "level" k, and we want to estimate effects with $z_i = \sum_k s_{ik} g_k$
  - E.g. g_k are shocks to industries k and $s_{ik} \in (0,1)$ are regional measures of shock exposure, perhaps with $\sum_k s_{ik} = 1$ for all i

- What is the "expected instrument"?

$$E[z_i \mid s] = E\left[\sum_k s_{ik} g_k \mid s\right] = \sum_k s_{ik} E[g_k] = \left(\sum_k s_{ik}\right)\mu$$

  - I.e. need to control for the "sum of shares" $w_i = \sum_k s_{ik}$ (which may be one, in which case no controls needed!)

- Cool new twist: we can use design to "translate" shocks from one level (e.g. industries) to estimate effects at another (e.g. regions)!

Shift-Share Instruments (Cont.)

- Now suppose $E[g_k \mid y(\cdot), s, q] = q_k'\mu$, still with $z_i = \sum_k s_{ik} g_k$

Shift-Share Instruments (Cont.)

- Now suppose $E[g_k \mid y(\cdot), s, q] = q_k' \mu$, still with $z_i = \sum_k s_{ik} g_k$
  - E.g. Autor et al. (2014) leverage industry shocks g_k from China over two periods, with q_k being a set of period FE

Shift-Share Instruments (Cont.)

- Now suppose $E[g_k \mid y(\cdot), s, q] = q_k'\mu$, still with $z_i = \sum_k s_{ik} g_k$
  - E.g. Autor et al. (2014) leverage industry shocks g_k from China over two periods, with q_k being a set of period FE
  - Want to only use within-period shock variation: e.g. shocks and unobservables could have different means across time

Shift-Share Instruments (Cont.)

- Now suppose $E[g_k \mid y(\cdot), s, q] = q_k'\mu$, still with $z_i = \sum_k s_{ik} g_k$
  - E.g. Autor et al. (2014) leverage industry shocks g_k from China over two periods, with q_k being a set of period FE
  - Want to only use within-period shock variation: e.g. shocks and unobservables could have different means across time

- What is the "expected instrument"?

$$E[z_i \mid s, q] = E\left[\sum_k s_{ik} g_k \mid s, q\right] = \sum_k s_{ik} E[g_k \mid q_k] = \left(\sum_k s_{ik} q_k\right)' \mu$$

Shift-Share Instruments (Cont.)

- Now suppose $E[g_k \mid y(\cdot), s, q] = q_k' \mu$, still with $z_i = \sum_k s_{ik} g_k$
  - E.g. Autor et al. (2014) leverage industry shocks g_k from China over two periods, with q_k being a set of period FE
  - Want to only use within-period shock variation: e.g. shocks and unobservables could have different means across time
- What is the "expected instrument"?

$$E[z_i \mid s, q] = E\left[\sum_k s_{ik} g_k \mid s, q\right] = \sum_k s_{ik} E[g_k \mid q_k] = \left(\sum_k s_{ik} q_k\right)' \mu$$

  - I.e. need to control for the share-weighted average of shock-level confounders, $w_i = \sum_k s_{ik} q_k$

Shift-Share Instruments (Cont.)

- Now suppose $E[g_k \mid y(\cdot), s, q] = q_k'\mu$, still with $z_i = \sum_k s_{ik} g_k$
  - E.g. Autor et al. (2014) leverage industry shocks g_k from China over two periods, with q_k being a set of period FE
  - Want to only use within-period shock variation: e.g. shocks and unobservables could have different means across time

- What is the "expected instrument"?

$$E[z_i \mid s, q] = E\left[\sum_k s_{ik} g_k \mid s, q\right] = \sum_k s_{ik} E[g_k \mid q_k] = \left(\sum_k s_{ik} q_k\right)' \mu$$

  - I.e. need to control for the share-weighted average of shock-level confounders, $w_i = \sum_k s_{ik} q_k$
  - In Autor et al. (2014), this means controlling for the sum-of-shares interacted with period FE

Illustration: Autor et al. (2014) China Shock

ADH study the effects of rising Chinese import competition on US commuting zones, 1991-2000 and 2000-2007

- Treatment x_{it}: local growth of Chinese imports in \$1,000/worker
- Main outcome y_{it}: local change in manufacturing employment share

Illustration: Autor et al. (2014) China Shock

ADH study the effects of rising Chinese import competition on US commuting zones, 1991-2000 and 2000-2007

- Treatment x_{it}: local growth of Chinese imports in \$1,000/worker
- Main outcome y_{it}: local change in manufacturing employment share

To address endogeneity, ADH use a SSIV $z_{it} = \sum_n s_{int} g_{nt}$

- n: 397 SIC4 manufacturing industries $\times$ two periods
- g_{nt}: growth of Chinese imports in non-US economies per US worker
- s_{int}: lagged share of manufacturing industry n in *total* employment of location i; hence $\sum_n s_{int}$ is i's manufacturing employment share

Illustration: Autor et al. (2014) China Shock

ADH study the effects of rising Chinese import competition on US commuting zones, 1991-2000 and 2000-2007

- Treatment x_{it}: local growth of Chinese imports in \$1,000/worker
- Main outcome y_{it}: local change in manufacturing employment share

To address endogeneity, ADH use a SSIV $z_{it} = \sum_n s_{int} g_{nt}$

- n: 397 SIC4 manufacturing industries $\times$ two periods
- g_{nt}: growth of Chinese imports in non-US economies per US worker
- s_{int}: lagged share of manufacturing industry n in *total* employment of location i; hence $\sum_n s_{int}$ is i's manufacturing employment share

Design-based justification: random industry productivity shocks in China, jointly affecting imports in the U.S. and elsewhere

- Observed g_{nt} proxy for latent as-if-randomly assigned productivity shocks; need other drivers to be unrelated to local U.S. conditions

ADH Balance Tests

Balance variable	Coef.		SE
Panel A: Industry-level balance			
Production workers' share of employment, 1991	−0.011		(0.012)
Ratio of capital to value-added, 1991	−0.007		(0.019)
Log real wage (2007 USD), 1991	−0.005		(0.022)
Computer investment as share of total, 1990	0.750		(0.465)
High-tech equipment as share of total investment, 1990	0.532		(0.296)
No. of industry-periods		794	
Panel B: Regional balance			
Start-of-period % of college-educated population	0.915		(1.196)
Start-of-period % of foreign-born population	2.920		(0.952)
Start-of-period % of employment among women	−0.159		(0.521)
Start-of-period % of employment in routine occupations	−0.302		(0.272)
Start-of-period average offshorability index of occupations	0.087		(0.075)
Manufacturing employment growth, 1970s	0.543		(0.227)
Manufacturing employment growth, 1980s	0.055		(0.187)
No. of region-periods		1,444	

- Panel A regresses industry characteristics on the g_{nt} shocks, controlling for period FE
- Panel B regresses location characteristics on the z_{it} instrument, controlling for manufacturing employment share $\times$ period FE

ADH Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Coefficient	−0.596	−0.489	−0.267	−0.314	−0.310	−0.290	−0.432
	(0.114)	(0.100)	(0.099)	(0.107)	(0.134)	(0.129)	(0.205)
Regional controls							
Autor *et al.* (2013) controls	✓	✓	✓		✓	✓	✓
Start-of-period mfg. share	✓						
Lagged mfg. share		✓	✓	✓	✓	✓	✓
Period-specific lagged mfg. share			✓	✓	✓	✓	✓
Lagged 10-sector shares					✓		✓
Local Acemoglu *et al.* (2016) controls						✓	
Lagged industry shares							✓
SSIV first stage F-stat.	185.6	166.7	123.6	272.4	64.6	63.3	27.6
No. of region-periods	1,444	1,444	1,444	1,444	1,444	1,444	1,444
No. of industry-periods	796	794	794	794	794	794	794

Note: columns 3-7 control for mfg. employment share $\times$ period FE

Contrast: Outome Modeling

- Goldsmith-Pinkham et al. (2020) discuss an alternative interpretation of shift-share IV, which models $E[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}, g] = E[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}]$
  - Condition on the shocks, then assume the shares are exogenous in the sense of $E[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}] = 0$
  - When things are in first-differences, this is effectively parallel trends: e.g. regions with different exposure shares would have had the same outcome trends if not for shocks to treatment

Contrast: Outome Modeling

- Goldsmith-Pinkham et al. (2020) discuss an alternative interpretation of shift-share IV, which models $E[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}, g] = E[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}]$
    - Condition on the shocks, then assume the shares are exogenous in the sense of $E[\varepsilon_i \mid s_{i1}, \ldots, s_{iK}] = 0$
    - When things are in first-differences, this is effectively parallel trends: e.g. regions with different exposure shares would have had the same outcome trends if not for shocks to treatment

- Reasonable enough, but not design-based
    - Need an *ex ante* stand on the right form for y_i (e.g. logs v levels)
    - Any transformation of $(s_{i1}, \ldots, s_{iK})$ is a valid instrument, including the individual shares $\rightarrow$ overidentified, potentially massively so
    - Can't have any unobserved shocks which transmit to y_i through the same or correlated shares (e.g. can't have $\varepsilon_i = \sum_k s_{ik} \nu_k + \eta_i$, even if the ν_k are totally independent of the g_k)

Going Beyond Linear Formula

- All the x_i and z_i we've considered so far are linear in exogenous shocks
  - Linearity means a specification for $E[g \mid w]$ is enough to know $E[z \mid w]$

Going Beyond Linear Formula

- All the x_i and z_i we've considered so far are linear in exogenous shocks
  - Linearity means a specification for $E[g \mid w]$ is enough to know $E[z \mid w]$
- Can we consider nonlinear constructions? E.g. $z_i = \ln(\sum_k s_{ik} g_k)$
  - Yes, but now to know $E[z \mid w]$ we generally need to specify the full *distribution* of $g \mid w$ (let's call it $G(w)$)

Going Beyond Linear Formula

- All the x_i and z_i we've considered so far are linear in exogenous shocks
  - Linearity means a specification for $E[g \mid w]$ is enough to know $E[z \mid w]$
- Can we consider nonlinear constructions? E.g. $z_i = \ln(\sum_k s_{ik} g_k)$
  - Yes, but now to know $E[z \mid w]$ we generally need to specify the full *distribution* of $g \mid w$ (let's call it $G(w)$)
- This need not be as daunting as it seems!

Going Beyond Linear Formula

- All the x_i and z_i we've considered so far are linear in exogenous shocks
  - Linearity means a specification for $E[g \mid w]$ is enough to know $E[z \mid w]$
- Can we consider nonlinear constructions? E.g. $z_i = \ln(\sum_k s_{ik} g_k)$
  - Yes, but now to know $E[z \mid w]$ we generally need to specify the full *distribution* of $g \mid w$ (let's call it $G(w)$)
- This need not be as daunting as it seems!
  - If $g \mid w$ comes from a true experiment, then $G(w)$ is given by the experimental protocol

Going Beyond Linear Formula

- All the x_i and z_i we've considered so far are linear in exogenous shocks
  - Linearity means a specification for $E[g \mid w]$ is enough to know $E[z \mid w]$
- Can we consider nonlinear constructions? E.g. $z_i = \ln(\sum_k s_{ik} g_k)$
  - Yes, but now to know $E[z \mid w]$ we generally need to specify the full *distribution* of $g \mid w$ (let's call it $G(w)$)
- This need not be as daunting as it seems!
  - If $g \mid w$ comes from a true experiment, then $G(w)$ is given by the experimental protocol
  - We might think some of the shocks are *exchangeable*: e.g. swap which industries grew more vs. less in ADH (maybe within sectors)

Going Beyond Linear Formula

- All the x_i and z_i we've considered so far are linear in exogenous shocks
  - Linearity means a specification for $E[g \mid w]$ is enough to know $E[z \mid w]$
- Can we consider nonlinear constructions? E.g. $z_i = \ln(\sum_k s_{ik} g_k)$
  - Yes, but now to know $E[z \mid w]$ we generally need to specify the full *distribution* of $g \mid w$ (let's call it $G(w)$)
- This need not be as daunting as it seems!
  - If $g \mid w$ comes from a true experiment, then $G(w)$ is given by the experimental protocol
  - We might think some of the shocks are *exchangeable*: e.g. swap which industries grew more vs. less in ADH (maybe within sectors)
  - Could exploit other symmetries/discontinuities in the shocks (e.g. RD)

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion
  - Shocks may vary at a different “level” than observations

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion
  - Shocks may vary at a different "level" than observations
- Design-based approach: include controls that span $\mu_i = E[f_i(g, s) \mid w]$

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion
  - Shocks may vary at a different "level" than observations

- Design-based approach: include controls that span $\mu_i = E[f_i(g, s) \mid w]$
  - Or use "recentered" instrument $\tilde{z}_i = z_i - \mu_i$ (same estimand)

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion
  - Shocks may vary at a different "level" than observations

- Design-based approach: include controls that span $\mu_i = E[f_i(g, s) \mid w]$
  - Or use "recentered" instrument $\tilde{z}_i = z_i - \mu_i$ (same estimand)

- For complex designs, μ_i can be computed by simulation:
  1. Redraw counterfactual shocks $g^{(\ell)}$ from $G(w)$, $\ell = 1, \ldots, L$

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion
  - Shocks may vary at a different "level" than observations
- Design-based approach: include controls that span $\mu_i = E[f_i(g, s) \mid w]$
  - Or use "recentered" instrument $\tilde{z}_i = z_i - \mu_i$ (same estimand)
- For complex designs, μ_i can be computed by simulation:
  1. Redraw counterfactual shocks $g^{(\ell)}$ from $G(w)$, $\ell = 1, \ldots, L$
  2. Recompute the IV w/these shocks, holding else fixed: $z_i^{(\ell)} = f_i(g^{(\ell)}, s)$

Formula Instruments: General Case

- We have a $z_i = f_i(g, s)$ for known $f_i(\cdot)$ and observed g and s
  - We assume $g \mid y(\cdot), w \sim G(w)$ for $w = (s, q)$ and known $G(\cdot)$
  - Reduced-form case: $z_i = x_i$; otherwise implicitly assuming exclusion
  - Shocks may vary at a different "level" than observations
- Design-based approach: include controls that span $\mu_i = E[f_i(g, s) \mid w]$
  - Or use "recentered" instrument $\tilde{z}_i = z_i - \mu_i$ (same estimand)
- For complex designs, μ_i can be computed by simulation:
  1. Redraw counterfactual shocks $g^{(\ell)}$ from $G(w)$, $\ell = 1, \ldots, L$
  2. Recompute the IV w/these shocks, holding else fixed: $z_i^{(\ell)} = f_i(g^{(\ell)}, s)$
  3. Compute expected instrument as $\mu_i = \frac{1}{L} \sum_\ell z_i^{(\ell)}$

Example: Borusyak and Hull (2023)

- BH are interested in estimating the effect of market access on employment by leveraging changes in the transportation network
    - Market access specifies (using economic theory) how transportation upgrades affect economic integration across a country (i.e. spillovers)

Example: Borusyak and Hull (2023)

- BH are interested in estimating the effect of market access on employment by leveraging changes in the transportation network
  - Market access specifies (using economic theory) how transportation upgrades affect economic integration across a country (i.e. spillovers)
  - Upgrades (of e.g. rail lines) at a different level than regional outcomes

Example: Borusyak and Hull (2023)

- BH are interested in estimating the effect of market access on employment by leveraging changes in the transportation network
  - Market access specifies (using economic theory) how transportation upgrades affect economic integration across a country (i.e. spillovers)
  - Upgrades (of e.g. rail lines) at a different level than regional outcomes
- They use the differential timing of high-speed rail (HSR) construction in China, conditional on construction plans, as exogenous shocks

Example: Borusyak and Hull (2023)

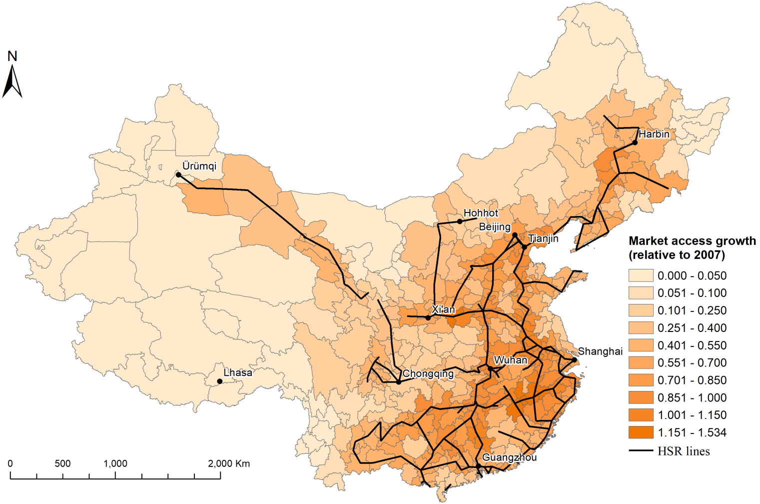
- BH are interested in estimating the effect of market access on employment by leveraging changes in the transportation network
  - Market access specifies (using economic theory) how transportation upgrades affect economic integration across a country (i.e. spillovers)
  - Upgrades (of e.g. rail lines) at a different level than regional outcomes
- They use the differential timing of high-speed rail (HSR) construction in China, conditional on construction plans, as exogenous shocks
  - Basic idea: permute which planned lines opened by some date conditional on line observables to generate counterfactual shocks

Example: Borusyak and Hull (2023)

- BH are interested in estimating the effect of market access on employment by leveraging changes in the transportation network
  - Market access specifies (using economic theory) how transportation upgrades affect economic integration across a country (i.e. spillovers)
  - Upgrades (of e.g. rail lines) at a different level than regional outcomes

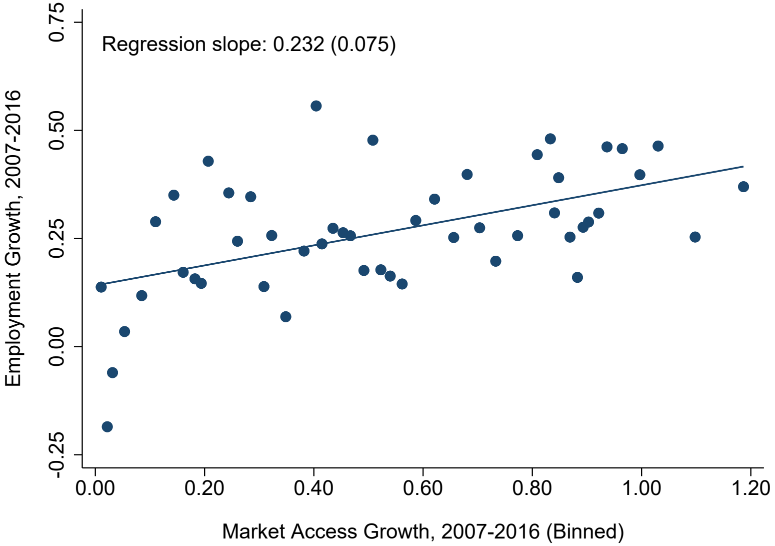
- They use the differential timing of high-speed rail (HSR) construction in China, conditional on construction plans, as exogenous shocks
  - Basic idea: permute which planned lines opened by some date conditional on line observables to generate counterfactual shocks
  - Then either control for or recenter by expected market access growth

HSR Lines and Market Access



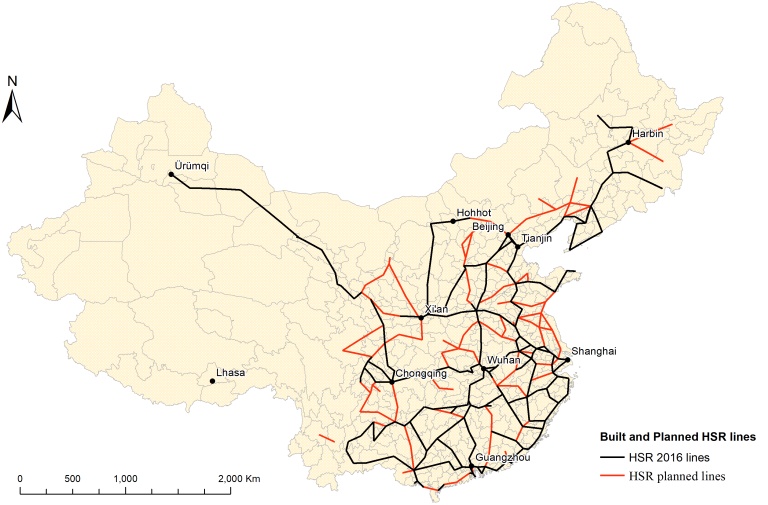
Naive OLS compares dark ("treatment") vs light ("control") regions

Naive OLS Suggests a Big Market Access Effect...



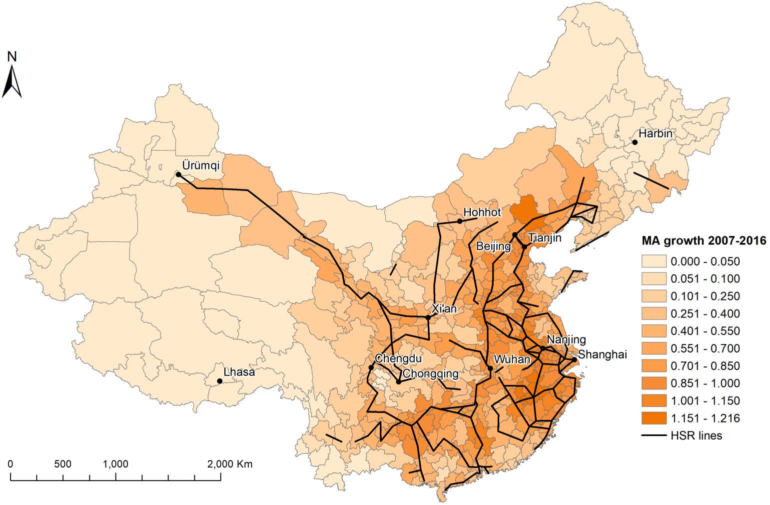
... but we probably shouldn't believe it

HSR Lines and Counterfactuals



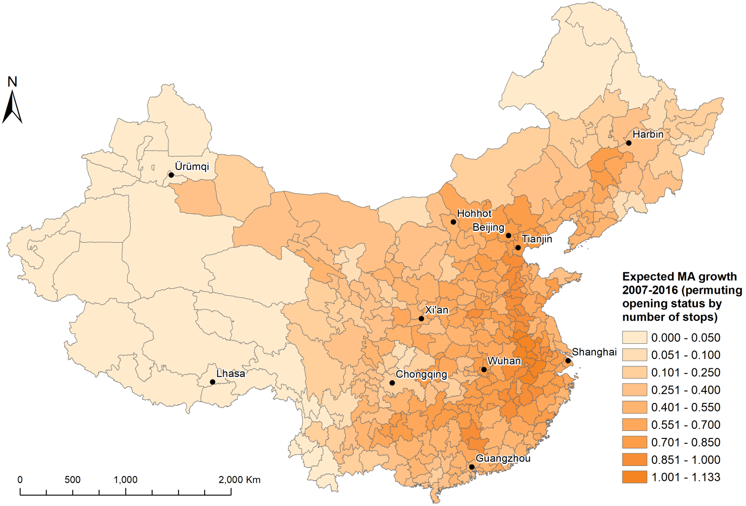
Counterfactuals permute which lines opened by 2016, conditional on length

An Example Counterfactual HSR Network



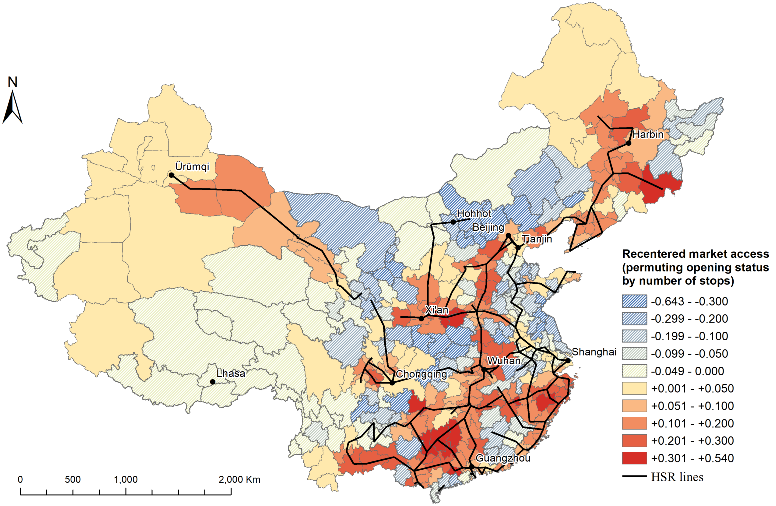
Seems ok...

Expected Market Access Across Counterfactuals



Darker regions see more MA growth regardless of which lines are built first

Recentered Market Access



Recentered IV compares region that saw more MA growth than expected (in red) to those that saw less MA growth than expected (in blue)

Balance Tests

	Unadjusted	Recentered		
	(1)	(2)	(3)	(4)
Distance to Beijing	−0.292	0.069		0.089
	(0.063)	(0.040)		(0.045)
Latitude/100	−3.323	−0.325		−0.156
	(0.648)	(0.277)		(0.320)
Longitude/100	1.329	0.473		0.425
	(0.460)	(0.239)		(0.242)
Expected Market Access Growth			0.027	0.056
			(0.056)	(0.066)
Constant	0.536	0.014	0.014	0.014
	(0.030)	(0.018)	(0.020)	(0.018)
Joint RI p-value		0.489	0.807	0.536
R^2	0.823	0.079	0.007	0.082
Prefectures	274	274	274	274

Recentered MA growth can't be reliably predicted from geography

BH Estimates

	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

Large effect for naive OLS goes away with recentering/controlling

- Once adjusting for μ_i, auxilliary controls don't matter ($\implies$ balance)

Recentering for Power

- Leveraging non-random exposure can dramatically improve precision when leveraging exogenous shocks in an IV...
  - ... as long as you recenter to use this variation "safely"

Recentering for Power

- Leveraging non-random exposure can dramatically improve precision when leveraging exogenous shocks in an IV...
  - ... as long as you recenter to use this variation "safely"
- Ex: Medicaid eligibility is a treatment combining statewide policy shocks and individual exposure (income, family structure, etc)
  - In settings where policy shocks are plausibly exogenous, standard approach is to use them directly as instruments ("simulated IV")
  - BH approach: use eligibility itself, but recenter: e.g. adjust for i's avg. eligibility across permutations of policies (swap MA & RI, say)

Recentering for Power

- Leveraging non-random exposure can dramatically improve precision when leveraging exogenous shocks in an IV...
  - ... as long as you recenter to use this variation "safely"
- Ex: Medicaid eligibility is a treatment combining statewide policy shocks and individual exposure (income, family structure, etc)
  - In settings where policy shocks are plausibly exogenous, standard approach is to use them directly as instruments ("simulated IV")
  - BH approach: use eligibility itself, but recenter: e.g. adjust for i's avg. eligibility across permutations of policies (swap MA & RI, say)
- In an application to ACA Medicaid expansions, we find this helps a lot
  - Design assumption: can swap expansion decisions of states conditional on whether the governor is a Republican or Democrat

Estimates of Private Insurance Crowdout Effects

	Has Medicaid		Has Private Insurance		Has Employer-Sponsored Insurance	
	Simulated IV (1)	Recentered IV (2)	Simulated IV (3)	Recentered IV (4)	Simulated IV (5)	Recentered IV (6)
Panel A. Eligibility Effects						
Eligibility	0.132	0.072	−0.048	−0.023	0.009	−0.009
	(0.028)	(0.010)	(0.023)	(0.007)	(0.014)	(0.005)
	[0.080,0.216]	[0.051,0.093]	[-0.110,0.009]	[-0.040,-0.007]	[-0.034,0.052]	[-0.021,0.004]
Panel B. Enrollment Effects						
Has Medicaid			−0.361	−0.321	0.068	−0.125
			(0.165)	(0.092)	(0.111)	(0.061)
			[-0.813,0.082]	[-0.566,-0.108]	[-0.232,0.421]	[-0.263,0.070]
P-value: SIV=RIV			0.719		0.104	
Exposed Sample	N	Y	N	Y	N	Y
States	43	43	43	43	43	43
Individuals	2,397,313	421,042	2,397,313	421,042	2,397,313	421,042

Inference Can Be Tricky with Formulas

- Common exposure to the exogenous shocks g make $z_i = f_i(g, s)$ correlated across i, potentially in complicated ways
  - E.g. for shift-share $z_i = \sum_k s_{ik} g_k$, if unit i and j are far apart in space but close in terms of $(s_{ik})_{k=1}^{K}$ and $(s_{jk})_{k=1}^{K}$ then $Cov(z_i, z_j) > 0$

Inference Can Be Tricky with Formulas

- Common exposure to the exogenous shocks g make $z_i = f_i(g, s)$ correlated across i, potentially in complicated ways
  - E.g. for shift-share $z_i = \sum_k s_{ik} g_k$, if unit i and j are far apart in space but close in terms of $(s_{ik})_{k=1}^K$ and $(s_{jk})_{k=1}^K$ then $Cov(z_i, z_j) > 0$
  - Following our Day 2 discussion, this suggests we should cluster i and j together.

Inference Can Be Tricky with Formulas

- Common exposure to the exogenous shocks g make $z_i = f_i(g, s)$ correlated across i, potentially in complicated ways
  - E.g. for shift-share $z_i = \sum_k s_{ik} g_k$, if unit i and j are far apart in space but close in terms of $(s_{ik})_{k=1}^K$ and $(s_{jk})_{k=1}^K$ then $Cov(z_i, z_j) > 0$
  - Following our Day 2 discussion, this suggests we should cluster i and j together. But how do we do this? Shares are not "groups"

Inference Can Be Tricky with Formulas

- Common exposure to the exogenous shocks g make $z_i = f_i(g, s)$ correlated across i, potentially in complicated ways
  - E.g. for shift-share $z_i = \sum_k s_{ik} g_k$, if unit i and j are far apart in space but close in terms of $(s_{ik})_{k=1}^K$ and $(s_{jk})_{k=1}^K$ then $Cov(z_i, z_j) > 0$
  - Following our Day 2 discussion, this suggests we should cluster i and j together. But how do we do this? Shares are not "groups"

- For SSIV, Borusyak et al. (2022) show there is an equivalent shock-level IV that yields "exposure-robust" standard errors with ", r"
  - Intuitively: address "clustering" by running the IV at the level of identifying variation. Stata/R package: *ssaggregate*

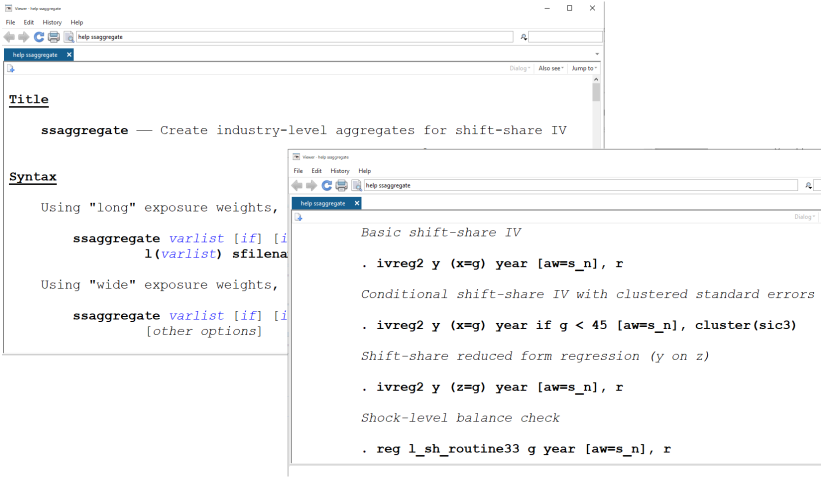
Inference Can Be Tricky with Formulas

- Common exposure to the exogenous shocks g make $z_i = f_i(g, s)$ correlated across i, potentially in complicated ways
  - E.g. for shift-share $z_i = \sum_k s_{ik} g_k$, if unit i and j are far apart in space but close in terms of $(s_{ik})_{k=1}^K$ and $(s_{jk})_{k=1}^K$ then $Cov(z_i, z_j) > 0$
  - Following our Day 2 discussion, this suggests we should cluster i and j together. But how do we do this? Shares are not "groups"

- For SSIV, Borusyak et al. (2022) show there is an equivalent shock-level IV that yields "exposure-robust" standard errors with ", r"
  - Intuitively: address "clustering" by running the IV at the level of identifying variation. Stata/R package: *ssaggregate*

- For other $f_i(\cdot)$, Borusyak and Hull '23 propose randomization inference
  - Use the counterfactual g to simulate the distribution of test statistics under the null and check if the actual test is in the tails

ssaggregate in Stata



This is what I used to get SEs in the previous balance / IV tables

Outline

1. Formula Treatments/Instruments✓

2. Nonlinear/Structural Models

Models as Extrapolators

- I've focused on *linear* regression/IV procedures, showing that they can estimate certain convex averages of heterogeneous effects given design

Models as Extrapolators

- I've focused on *linear* regression/IV procedures, showing that they can estimate certain convex averages of heterogeneous effects given design
  - But such effects needn't coincide w/policy-relevant param's (e.g. ATE)

Models as Extrapolators

- I've focused on *linear* regression/IV procedures, showing that they can estimate certain convex averages of heterogeneous effects given design
  - But such effects needn't coincide w/policy-relevant param's (e.g. ATE)
- Intuitively, we can try to bridge this gap by putting additional structure on (otherwise non-parametric) potential outcomes

Models as Extrapolators

- I've focused on *linear* regression/IV procedures, showing that they can estimate certain convex averages of heterogeneous effects given design
  - But such effects needn't coincide w/policy-relevant param's (e.g. ATE)
- Intuitively, we can try to bridge this gap by putting additional structure on (otherwise non-parametric) potential outcomes
  - In fact, we've already seen this: our Day-1 constant-effect model of $y_i = \beta x_i + \varepsilon_i$ can be understood as extrapolating simply across all i

Models as Extrapolators

- I've focused on *linear* regression/IV procedures, showing that they can estimate certain convex averages of heterogeneous effects given design
  - But such effects needn't coincide w/policy-relevant param's (e.g. ATE)
- Intuitively, we can try to bridge this gap by putting additional structure on (otherwise non-parametric) potential outcomes
  - In fact, we've already seen this: our Day-1 constant-effect model of $y_i = \beta x_i + \varepsilon_i$ can be understood as extrapolating simply across all i
  - Other (nonlinear) procedures can sometimes be seen as imposing different extrapolations to the same underlying (design-based) variation

Primer: Probit in Saturated Models

- Suppose you run an RCT with a binary treatment x_i. A regression of y_i on x_i identifies $E[y_i \mid x_i = 1] - E[y_i \mid x_i = 0] = E[y_i(1) - y_i(0)]$
  - What changes if y_i is also binary?

Primer: Probit in Saturated Models

- Suppose you run an RCT with a binary treatment x_i. A regression of y_i on x_i identifies $E[y_i \mid x_i = 1] - E[y_i \mid x_i = 0] = E[y_i(1) - y_i(0)]$
  - What changes if y_i is also binary? Nothing.

Primer: Probit in Saturated Models

- Suppose you run an RCT with a binary treatment x_i. A regression of y_i on x_i identifies $E[y_i \mid x_i = 1] - E[y_i \mid x_i = 0] = E[y_i(1) - y_i(0)]$
  - What changes if y_i is also binary? Nothing.

- Now suppose the true model for $y_i \mid x_i$ is a Probit:

$$y_i = \mathbf{1}[y_i^* \geq 0]$$
$$y_i^* = \alpha + \beta x_i - \varepsilon_i$$
$$\varepsilon_i \,|x_i \sim \mathrm{N}(0, \sigma^2)$$

Primer: Probit in Saturated Models

- Suppose you run an RCT with a binary treatment x_i. A regression of y_i on x_i identifies $E[y_i \mid x_i = 1] - E[y_i \mid x_i = 0] = E[y_i(1) - y_i(0)]$
  - What changes if y_i is also binary? Nothing.

- Now suppose the true model for $y_i \mid x_i$ is a Probit:

$$y_i = \mathbf{1}[y_i^* \geq 0]$$
$$y_i^* = \alpha + \beta x_i - \varepsilon_i$$
$$\varepsilon_i \,|x_i \sim \mathrm{N}(0, \sigma^2)$$

- The CEF implied by this model is linear, so regression fits it perfectly

$$E[y_i \mid x_i] = Pr(y_i^* \geq 0 \mid x_i) =$$

Primer: Probit in Saturated Models

- Suppose you run an RCT with a binary treatment x_i. A regression of y_i on x_i identifies $E[y_i \mid x_i = 1] - E[y_i \mid x_i = 0] = E[y_i(1) - y_i(0)]$
  - What changes if y_i is also binary? Nothing.

- Now suppose the true model for $y_i \mid x_i$ is a Probit:

$$y_i = \mathbf{1}[y_i^* \geq 0]$$
$$y_i^* = \alpha + \beta x_i - \varepsilon_i$$
$$\varepsilon_i \,|x_i \sim \mathrm{N}(0, \sigma^2)$$

- The CEF implied by this model is linear, so regression fits it perfectly

$$E[y_i \mid x_i] = Pr(y_i^* \geq 0 \mid x_i) = \Phi\left(\frac{\alpha + \beta x_i}{\sigma}\right)$$

$$=$$

Primer: Probit in Saturated Models

- Suppose you run an RCT with a binary treatment x_i. A regression of y_i on x_i identifies $E[y_i \mid x_i = 1] - E[y_i \mid x_i = 0] = E[y_i(1) - y_i(0)]$
  - What changes if y_i is also binary? Nothing.

- Now suppose the true model for $y_i \mid x_i$ is a Probit:

$$
\begin{aligned}
y_i &= \mathbf{1}[y_i^* \geq 0] \\
y_i^* &= \alpha + \beta x_i - \varepsilon_i \\
\varepsilon_i \,|x_i &\sim \mathrm{N}(0, \sigma^2)
\end{aligned}
$$

- The CEF implied by this model is linear, so regression fits it perfectly

$$
\begin{aligned}
E[y_i \mid x_i] &= Pr(y_i^* \geq 0 \mid x_i) = \Phi\left(\frac{\alpha + \beta x_i}{\sigma}\right) \\
&= \underbrace{\Phi\left(\frac{\alpha}{\sigma}\right)}_{\text{constant}} + \underbrace{\left\{\Phi\left(\frac{\alpha+\beta}{\sigma}\right) - \Phi\left(\frac{\alpha}{\sigma}\right)\right\}}_{\text{slope coefficient}} x_i
\end{aligned}
$$

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0, 1\}$. Those with $w_i = 0$ are all untreated

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0, 1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0, 1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?
  - A: The CATE $E[y_i(1) - y_i(0) \mid w_i = 1]$, since $Var(x_i \mid w_i = 0) = 0$

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0, 1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?
  - A: The CATE $E[y_i(1) - y_i(0) \mid w_i = 1]$, since $Var(x_i \mid w_i = 0) = 0$
  - Taking the linear regression/model for $E[y_i \mid x_i, w_i]$ seriously, this is also our estimate of the (otherwise unidentified) $E[y_i(1) - y_i(0) \mid w_i = 0]$

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0, 1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?
  - A: The CATE $E[y_i(1) - y_i(0) \mid w_i = 1]$, since $Var(x_i \mid w_i = 0) = 0$
  - Taking the linear regression/model for $E[y_i \mid x_i, w_i]$ seriously, this is also our estimate of the (otherwise unidentified) $E[y_i(1) - y_i(0) \mid w_i = 0]$

- Suppose y_i is binary and we instead run a Probit on x_i and w_i
  - Taking it seriously, Probit structures potential outcomes:

$$y_i = \mathbf{1}[\alpha + \beta x_i + \gamma w_i \geq \varepsilon_i], \quad \varepsilon_i \mid x_i, w_i \sim N(0,1)$$

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0,1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?
  - A: The CATE $E[y_i(1) - y_i(0) \mid w_i = 1]$, since $Var(x_i \mid w_i = 0) = 0$
  - Taking the linear regression/model for $E[y_i \mid x_i, w_i]$ seriously, this is also our estimate of the (otherwise unidentified) $E[y_i(1) - y_i(0) \mid w_i = 0]$

- Suppose y_i is binary and we instead run a Probit on x_i and w_i
  - Taking it seriously, Probit structures potential outcomes:

$$
\begin{aligned}
y_i =& \mathbf{1}[\alpha + \beta x_i + \gamma w_i \geq \varepsilon_i], \quad \varepsilon_i \mid x_i, w_i \sim N(0,1) \\
\implies& y_i(0) = \mathbf{1}[\alpha + \gamma w_i \geq \varepsilon_i], \;\; y_i(1) = \mathbf{1}[\alpha + \beta + \gamma w_i \geq \varepsilon_i]
\end{aligned}
$$

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0,1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?
  - A: The CATE $E[y_i(1) - y_i(0) \mid w_i = 1]$, since $Var(x_i \mid w_i = 0) = 0$
  - Taking the linear regression/model for $E[y_i \mid x_i, w_i]$ seriously, this is also our estimate of the (otherwise unidentified) $E[y_i(1) - y_i(0) \mid w_i = 0]$

- Suppose y_i is binary and we instead run a Probit on x_i and w_i
  - Taking it seriously, Probit structures potential outcomes:

$$
\begin{aligned}
y_i =& \mathbf{1}[\alpha + \beta x_i + \gamma w_i \geq \varepsilon_i], \quad \varepsilon_i \mid x_i, w_i \sim N(0,1) \\
\implies& y_i(0) = \mathbf{1}[\alpha + \gamma w_i \geq \varepsilon_i], \;\; y_i(1) = \mathbf{1}[\alpha + \beta + \gamma w_i \geq \varepsilon_i]
\end{aligned}
$$

  - In particular, $E[y_i(1) - y_i(0) \mid w_i = 1] = \Phi(\alpha + \beta + \gamma) - \phi(\alpha + \gamma)$ (will match OLS) but $E[y_i(1) - y_i(0) \mid w_i = 0] = \Phi(\alpha + \beta) - \Phi(\alpha)$

Extrapolation by OLS vs. Probit

- Suppose individuals with baseline covariate $w_i = 1$ are randomized into a treatment $x_i \in \{0,1\}$. Those with $w_i = 0$ are all untreated
  - Q: What do we get from regressing y_i on x_i controlling for w_i?
  - A: The CATE $E[y_i(1) - y_i(0) \mid w_i = 1]$, since $Var(x_i \mid w_i = 0) = 0$
  - Taking the linear regression/model for $E[y_i \mid x_i, w_i]$ seriously, this is also our estimate of the (otherwise unidentified) $E[y_i(1) - y_i(0) \mid w_i = 0]$

- Suppose y_i is binary and we instead run a Probit on x_i and w_i
  - Taking it seriously, Probit structures potential outcomes:

$$
\begin{aligned}
y_i =& \mathbf{1}[\alpha + \beta x_i + \gamma w_i \geq \varepsilon_i], \quad \varepsilon_i \mid x_i, w_i \sim N(0,1) \\
\implies& y_i(0) = \mathbf{1}[\alpha + \gamma w_i \geq \varepsilon_i], \;\; y_i(1) = \mathbf{1}[\alpha + \beta + \gamma w_i \geq \varepsilon_i]
\end{aligned}
$$

  - In particular, $E[y_i(1) - y_i(0) \mid w_i = 1] = \Phi(\alpha + \beta + \gamma) - \phi(\alpha + \gamma)$ (will match OLS) but $E[y_i(1) - y_i(0) \mid w_i = 0] = \Phi(\alpha + \beta) - \Phi(\alpha)$
  - Different extrapolation of missing CATE: a feature or a bug?

ExtrapoLATEing

- Consider the simplest design-based IV story: binary x_i, binary z_i, no controls. IV identifies LATE:

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

  where $x_i(z)$ denotes potential treatment when $z_i = z$

  - Avg. treatment effect among compliers (those with $x_i(1) = 1, x_i(0) = 0$)

ExtrapoLATEing

- Consider the simplest design-based IV story: binary x_i, binary z_i, no controls. IV identifies LATE:

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

  where $x_i(z)$ denotes potential treatment when $z_i = z$

  - Avg. treatment effect among compliers (those with $x_i(1) = 1, x_i(0) = 0$)

- Without restrictions on $(y_i(1), y_i(0), x_i(1), x_i(0))$, can't say anything more: IV only reveals effects among i whose x_i is shifted by z_i

ExtrapoLATEing

- Consider the simplest design-based IV story: binary x_i, binary z_i, no controls. IV identifies LATE:

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

where $x_i(z)$ denotes potential treatment when $z_i = z$

  - Avg. treatment effect among compliers (those with $x_i(1) = 1, x_i(0) = 0$)

- Without restrictions on $(y_i(1), y_i(0), x_i(1), x_i(0))$, can't say anything more: IV only reveals effects among i whose x_i is shifted by z_i
  - Actually not quite true: can identify avg. $y_i(1)$ of always-takers (w/ $x_i(1) = x_i(0) = 1$), avg. $y_i(0)$ of never-takers (w/ $x_i(1) = x_i(0) = 0$), 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-if-randomly assigned + excludable
  - Assume a distribution for $(y_i(1), y_i(0), v_i)$ where $x_i = \mathbf{1}[\mu + \pi z_i > v_i]$

Adding Structure to IV

- Suppose we have a z_i which is as-if-randomly assigned + excludable
  - Assume a distribution for $(y_i(1), y_i(0), \nu_i)$ where $x_i = \mathbf{1}[\mu + \pi z_i > \nu_i]$
  - Then we have parametric models for

$$E[y_i \mid x_i = 1, z_i = z] = E[y_i(1) \mid \mu + \pi z > \nu_i] \equiv f_1(z;\theta)$$
$$E[y_i \mid x_i = 0, z_i = z] = E[y_i(0) \mid \mu + \pi z < \nu_i] \equiv f_0(z;\theta)$$

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

Adding Structure to IV

- Suppose we have a z_i which is as-if-randomly assigned + excludable
  - Assume a distribution for $(y_i(1), y_i(0), \nu_i)$ where $x_i = \mathbf{1}[\mu + \pi z_i > \nu_i]$
  - Then we have parametric models for

$$E[y_i \mid x_i = 1, z_i = z] = E[y_i(1) \mid \mu + \pi z > \nu_i] \equiv f_1(z;\theta)$$
$$E[y_i \mid x_i = 0, z_i = z] = E[y_i(0) \mid \mu + \pi z < \nu_i] \equiv f_0(z;\theta)$$

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

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

Adding Structure to IV

- Suppose we have a z_i which is as-if-randomly assigned + excludable
  - Assume a distribution for $(y_i(1), y_i(0), \nu_i)$ where $x_i = \mathbf{1}[\mu + \pi z_i > \nu_i]$
  - Then we have parametric models for

$$E[y_i \mid x_i = 1, z_i = z] = E[y_i(1) \mid \mu + \pi z > \nu_i] \equiv f_1(z;\theta)$$
$$E[y_i \mid x_i = 0, z_i = z] = E[y_i(0) \mid \mu + \pi z < \nu_i] \equiv f_0(z;\theta)$$

    as well as "first stage" models for $Pr(x_i = 1 \mid z_i = z) = g(z;\theta)$
  - 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

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)

Linking Back to LATE

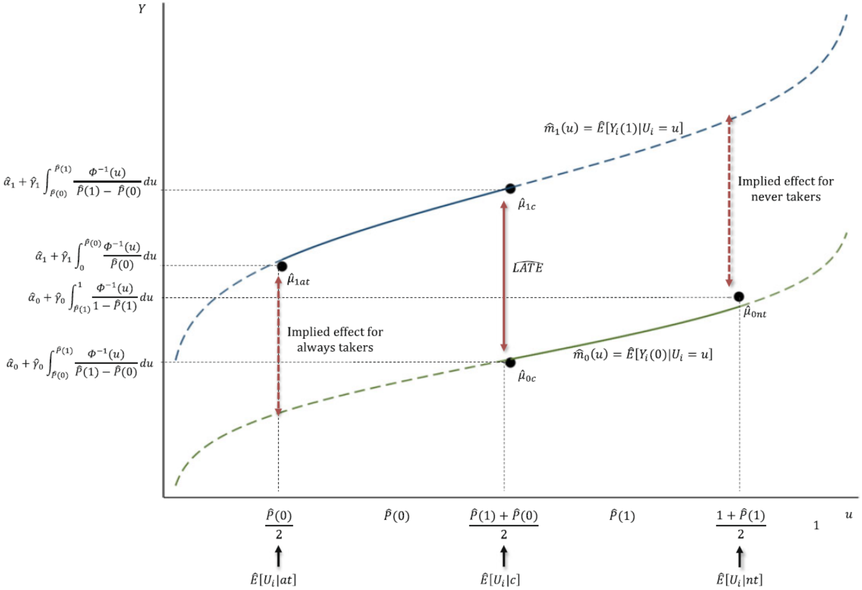
- Kline and Walters (2019) formalize this extrapolation logic in the familiar Imbens and Angrist (1994) setup
  - Key result: in simple binary z_i / no controls setup, control function estimates of LATE are numerically identical to linear IV
  - "Differences between structural and IV estimates therefore 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 (as in our earlier probit example!)

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 (as in our earlier probit example!)

- This can all be extended to conditional as-if-random assignment
  - Design knowledge gives you reduced-form estimands; then plug these into a model to get more!

Heckit Extrapolation of IV Moments



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

Summing Up

- You can do a lot with a solid design-based identification strategy
  - Give a clear *ex ante* rationalization for controls in a linear regression/IV
  - Have confidence in the level of standard error clustering
  - Avoid concerns over “negative weights” / explore alternative weightings
  - Access a large class of formula treatments/IVs
  - Understand what nonlinear/structural models buy you for identification

Summing Up

- You can do a lot with a solid design-based identification strategy
  - Give a clear *ex ante* rationalization for controls in a linear regression/IV
  - Have confidence in the level of standard error clustering
  - Avoid concerns over "negative weights" / explore alternative weightings
  - Access a large class of formula treatments/IVs
  - Understand what nonlinear/structural models buy you for identification
- Of course, design is not the only way to go: outcome modeling (e.g. DiD) may be a good alternative, especially w/o good shock variation
  - But some of the above issues can get murkier – just be clear on what you're assuming!

Summing Up

- You can do a lot with a solid design-based identification strategy
  - Give a clear *ex ante* rationalization for controls in a linear regression/IV
  - Have confidence in the level of standard error clustering
  - Avoid concerns over "negative weights" / explore alternative weightings
  - Access a large class of formula treatments/IVs
  - Understand what nonlinear/structural models buy you for identification
- Of course, design is not the only way to go: outcome modeling (e.g. DiD) may be a good alternative, especially w/o good shock variation
  - But some of the above issues can get murkier – just be clear on what you're assuming!

Thanks for a Great Class!