Lecture 13:
Using Micro Data for Macro

Kunal Sangani

ECON 416-1

Roadmap

- In first part of course, we covered how micro shocks affect aggregate outcomes.
- Often, the sufficient statistics that determine these effects need to be estimated:
  - E.g., households or firms' elasticities of substitution in consumption / production.
  - E.g., the elasticity of labor supply or the elasticity of intertemporal substitution.
- We may also want cleanly identified moments to understand where to improve models:
  - E.g., Shleifer (1985) event study on demand curves for stocks.
  - E.g., Mehra and Prescott (1985) compare equity premium against identified risk aversion.

Roadmap

- Next few lectures on using micro data for macro.
- Today: Linear regression, instruments, and shift-share (Bartik) designs.
  - Objective: introduce/review commonly used methods in empirical macroeconomics.
  - Flag intuitions and rules of thumb that empiricists use.
- Next lecture: Micro-to-macro aggregation.
  - This will take us into the terrain of some “uniquely macroeconomic” issues, such as spillovers and the missing intercept problem.

Table of Contents

Internal and external validity

- Internal validity: When does a regression have a causal interpretation?
  - Given study context, can we draw the inference that differences in an outcome variable were *caused* by differences in the independent variable?
  - I.e., does β from regression $y_i = \beta x_i + \varepsilon_i$ identify causal effect of varying x_i on y_i?
- External validity: Can we apply the causal effects we identify to counterfactuals beyond the narrow study context?
  - Is the statistic the relevant measure for questions of interest?
  - Is it portable to other contexts / settings?

Internal and external validity

- In general, there is a trade-off between internal validity and economic relevance.
- Economic relevance encompasses both external validity and whether the causal effects speak to some big / deep / interesting question.
- Most big, interesting economics questions are pretty confounded.
  - What caused the 1970s and 2020s inflations? Why did they end?
  - What is the effect of new technologies (e.g., LLMs) on growth?
- To get internal validity, we often need to restrict ourselves to special settings.
- Particularly for the job market, internal validity / identification is increasingly important.

Mechanism experiments

- Aside: Macro often special in that we can answer big / deep economic questions through the interplay between theory and well-identified evidence.
- Has the flavor of "mechanism experiments" (Ludwig, Kling, and Mullainathan, 2011):
  - "Suppose that the U.S. Department of Justice wanted to help local police chiefs decide whether to implement "broken windows" policing [....] Suppose that there is no credibly exogenous source of variation in the implementation or intensity of broken windows policing across areas. [...] To an experimentally minded economist, the most obvious next step goes something like: the Department of Justice should choose a representative sample of cities, randomly select half of their high-crime areas to receive broken windows policing (or perhaps randomly assign half the cities to get citywide broken windows policing), and carry out a traditional policy evaluation."
  - "Now consider an alternative experiment: Buy a small fleet of used cars. Break the windows of half of them. Park the cars in a randomly selected subset of neighborhoods, and then measure whether more serious crimes increase in response. Despite the potential Institutional Review Board concerns, a study with this basic design was carried out in the late 1960s, which in turn led eventually to the implementation of broken windows policing at scale in New York City during the 1990s."

Mechanism (quasi-)experiments in macro

- Emi Nakamura and Jon Steinsson on monetary nonneutrality:
  - “The effect of monetary shocks on output can be broken into two separate questions: First, how much do monetary shocks affect various relative prices? Second, how much do these various relative prices affect output? [...] All models—neoclassical and New Keynesian—imply that relative prices affect output. However, only in some models does monetary policy affect relative prices.”
- Tom Sargent on rational expectations:
  - “An implication of [the rational expectations] view is that inflation can be stopped much more quickly than advocates of the ”momentum” view have indicated and that their estimates of the length of time and the costs of stopping inflation in terms of foregone output ($220 billion of GNP for one percentage point in the inflation rate) are erroneous.”
- Andrei Shleifer on financial markets:
  - “Several important propositions in finance rely on the ability of investors to buy and sell any amount of the firm’s equity without significantly affecting the price. [...] If the demand curve is horizontal, inclusion of a stock into the S&P 500 should not be accompanied by a share price increase. In contrast, if the demand curve slopes down, we should observe a share price increase at the announcement of the inclusion.”

Common threats to internal validity

1. Omitted variables: x_i correlated with omitted variable included in error term.
   - E.g., estimating the effect of firm size on wages without controlling for worker ability.
2. Simultaneity: x_i endogenous because it is jointly determined with y_i.
   - E.g., prices and quantities, monetary policy and output; reverse causation.
3. Selection: Assignment to treatment correlated with outcome even w/out causal effect.
   - E.g., assignment correlated with lagged, auto-correlated variable (e.g., firms that receive subsidies are already expanding).
4. Improper statistical inference: Understate variance of $\hat{\beta}$ and overstate significance.
   - E.g., neglect time series or cross-sectional correlation between residuals.
5. Measurement error: If in x_i, biases estimate toward zero (attenuation).
   If in y_i, inflates variance of $\hat{\beta}$ and reduces statistical power.

Common threats to external validity

1. Unrepresentative sample: Treatment effects in study population differ from target population (by demographics, location, or time period).
   - E.g., firm responses in one industry or recession may not generalize to others.
2. Wrong statistic: Estimated parameter does not correspond to relevant policy statistic.
   - E.g., estimating intent-to-treat when policy always affects treated (nudge vs. mandate).
3. Nonlinearities / GE effects: Pilots vs. scaled-up interventions.
   - E.g., Moving to Opportunity vs. large-scale housing mobility.
4. Lucas critique: Structural relationships shift with expectations / policy regime.

Roadmap

1. Linear regression.
   - Internal validity problems: omitted variables, selection, statistical inference (clustering).
2. Instrumental variables.
   - Internal validity problems: Selection, simultaneity.
3. Shift-share instruments.
   - Internal validity problems: Selection, simultaneity, statistical inference.

Table of Contents

Regressions and causality: Potential outcomes

- Suppose we are interested in effect of technology adoption on firm prices, sales.
- Technology adoption treatment $D_i \in \{0, 1\}$.
- Two potential outcomes: $Y_i(0)$ and $Y_i(1)$.
- Rubin Causal Model: We observe

$$Y_i = Y_i(0) + D_i \underbrace{(Y_i(1) - Y_i(0))}_{\text{Causal effect of treatment on } i}.$$

- Question 1: Which statistic is of interest for macroeconomic counterfactuals?
  - Average treatment effect (ATE): $\mathbb{E}[Y_i(1) - Y_i(0)]$.
  - Average treatment effect for the treated (ATT): $\mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1]$.
  - Something else?

Regressions and causality: Potential outcomes

- Question 2: What do we measure if we compare unconditional means?

$$\underbrace{\mathbb{E}[Y_i|D_i=1]-\mathbb{E}[Y_i|D_i=0]}_{\text{Coefficient from regression of } Y_i \text{ on } D_i} =$$

$$\underbrace{\mathbb{E}[Y_i(1)-Y_i(0)|D_i=1]}_{\text{Avg. Treatment Effect for Treated (ATT)}} + \underbrace{(\mathbb{E}[Y_i(0)|D_i=1]-\mathbb{E}[Y_i(0)|D_i=0])}_{\text{Selection bias}}.$$

- Selection bias can go same or opposite direction as treatment effect.
  - E.g., more successful firms more likely to adopt technology.
  - E.g., laggard firms adopt technology because existing strategies are not working.
- Another way of writing this,

$$Y_i = \underbrace{(Y_i(1)-Y_i(0))}_{\substack{\text{Causal effect of}\\ \text{treatment on } i}} \times D_i + \underbrace{Y_i(0)}_{\substack{\text{Error}\\ \text{term}}} = \beta D_i + \varepsilon_i.$$

If $Cov[D_i,\varepsilon_i]\neq 0$, selection bias in regression of Y_i on D_i to estimate β.

Regressions and causality: Conditional independence

- Of course, if $Cov[D_i, \varepsilon_i] = 0$, regression result $\hat{\beta}$ estimates treatment effect.
- E.g., if treatment status is randomly assigned, then $Y_i(D_i) \perp D_i$.
- In empirical work, we often argue for conditional independence assumption:

$$Y_i(D_i) \perp D_i | X_i,$$

  i.e., given covariates X_i, treatment status is as-good-as-randomly assigned.
- E.g., for firms in same industry with similar characteristics, tech adoption is random.
- Of course, one can never prove that conditional independence assumption holds.
  - Challenge to convince a skeptical and clever audience that it does.

Regressions and causality: Conditional independence

- Suppose true model is
  $$Y_i(D_i) = \alpha + \beta D_i + \gamma X_i + \eta_i.$$
  E.g., D_i is tech adoption, X_i is manager ability.

- Conditional independence assumption is that $Y_i(D_i) \perp D_i | X_i$.
  $$\mathbb{E}[Y_i(D_i)|D_i, X_i] = \mathbb{E}[Y_i(D_i)|X_i] = \alpha + \beta D_i + \gamma X_i.$$

- Thus, if we estimate
  $$Y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon_i,$$
  conditional independence implies $\mathbb{E}[\varepsilon_i|D_i, X_i] = 0$, and β has causal interpretation.

Omitted variable bias

- What would happen if we left out X_i when estimating regression?

$$\hat{\beta} = \frac{Cov(Y_i, D_i)}{Var(D_i)} = \beta + \gamma \frac{Cov(X_i, D_i)}{Var(D_i)}.$$

- Bias depends on
  - $\frac{Cov(X_i, D_i)}{Var(D_i)}$: coefficient in univariate regression of X_i on D_i.
  - γ: coefficient on X_i from regression of Y_i on D_i and X_i.
- One option is to argue that omitted variable bias would go opposite way of results.
  - For example, suppose you find large positive effects of tech adoption on sales.
  - Could you argue that higher manager ability firms were less likely to adopt?
- If it is crucial to argue that results not driven by some control, control nonparametrically.

Omitted variable bias: Oster (2016)

- Often common to argue that if coefficient is stable as additional controls are added, then bias from remaining omitted variables is small.

- Oster (2016) formalizes this intuition. Suppose data generating process is:

  $$Y_i = \beta X_i + W_{1i} + W_{2i} + \varepsilon_i, \qquad \text{where} \qquad W_{1i} = \gamma' w_i.$$

  We observe X_i and w_i but not W_{2i}.

- Assume $Cov(W_{1i}, W_{2i}) = 0$. (Can always redefine W_{2i} as part orthogonal to w_i).

- Let δ be the strength of selection on unobservables vs. selection on observables,

  $$\delta \frac{Cov(X_i, W_{1i})}{Var(W_{1i})} = \frac{Cov(X_i, W_{2i})}{Var(W_{2i})}.$$

Omitted variable bias: Oster (2016)

- Define:
  - For $Y_i \sim X_i$, denote $\hat{\beta}_{\text{short}}$ and R^2_{short}.
  - For $Y_i \sim X_i + w_i$, denote $\hat{\beta}_{\text{long}}$ and R^2_{long}.
  - For $Y_i \sim X_i + w_i + W_{2i}$, denote R^2_{max}. Note: Why might $R^2_{\text{max}} < 1$?
- Key result: If researcher knows δ and R^2_{max}, then there is an omitted variable bias-adjusted estimated $\hat{\beta}^*(\delta, R^2_{\text{max}}) \to_p \beta$.

$$\hat{\beta}^*(\delta, R^2_{\text{max}}) \approx \hat{\beta}_{\text{long}} + \delta \left(\hat{\beta}_{\text{long}} - \hat{\beta}_{\text{short}} \right) \frac{R^2_{\text{max}} - R^2_{\text{long}}}{R^2_{\text{long}} - R^2_{\text{short}}}.$$

Adjust $\hat{\beta}_{\text{long}}$ in direction of $\hat{\beta}_{\text{long}} - \hat{\beta}_{\text{short}}$, more so if selection on unobservables (δ) is big, $\hat{\beta}_{\text{long}}$ very far from $\hat{\beta}_{\text{short}}$, or $R^2_{\text{short}} \approx R^2_{\text{long}} \ll R^2_{\text{max}}$.

Omitted variable bias: Oster (2016)

- How to use Oster (2016) result?

1. Fix R^2_{max} and find δ^* such that $\hat{\beta}^*(\delta, R^2_{max}) = 0$.
   - If δ^* large, you would need extreme selection on unobservables to overturn result.
2. Use formula to compute an identified set for β.
   - Vary δ from 0 to $\bar{\delta}$ (Oster 2016 argues $\bar{\delta} = 1$ works well in practice).
   - Vary R^2_{max} between R^2_{long} and 1.
   - Calculate $\hat{\beta}^*$ for each set of values.

Bad controls

- Using a bad control can be as bad as omitting a needed control.

- Suppose
$$Y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon_i.$$
But instead of observing X_i we observe proxy $W_i = \phi_D D_i + \phi_X X_i$.

- Regression of Y_i on D_i and W_i yields
$$Y_i = \alpha + \left(\beta - \frac{\phi_D}{\phi_X}\gamma\right) D_i + \frac{\gamma}{\phi_X} W_i + \varepsilon_i.$$

- Using bad control W_i contaminates our estimate for β.

Standard errors

- Getting your standard errors right is table stakes.
- Always worth thinking about clustering before you interpret regression output.
  - $t = 2$ implies $p = 4.6\%$.
  - $t = 3$ implies $p = 0.27\%$.
  - $t = 4$ implies $p = 0.006\%$.
  - $t = 5$ implies $p = 0.00006\%$ is standard in particle physics.
- If you find t-statistics > 5 in standard macro datasets for a non-mechanical relationship, you should worry you have a clustering problem.

Standard errors

- In standard regression model,

$$Y_i = \beta' X_i + \varepsilon_i,$$

asymptotic normality with independent sampling of observations implies,

$$\sqrt{N}\left(\hat{\beta} - \beta\right) \to_d \mathcal{N}\left(0, \left(\mathbb{E}\left[XX'\right]\right)^{-1} \mathbb{E}\left[\varepsilon^2 XX'\right] \left(\mathbb{E}\left[XX'\right]\right)^{-1}\right).$$

- Under assumption of homoskedasticity, $\mathbb{E}[\varepsilon^2 | X] = \sigma^2$, and we get

$$\sqrt{N}\left(\hat{\beta} - \beta\right) \to_d \mathcal{N}\left(0, \sigma^2 \left(\mathbb{E}\left[XX'\right]\right)^{-1}\right).$$

- But more generally,

$$Var\left(\sum_{i=1}^{N} X_i \varepsilon_i\right) = \sum_{i=1}^{N} Var(X_i \varepsilon_i) + \sum_{i \neq j} Cov(X_i \varepsilon_i, X_j \varepsilon_j).$$

  - Heteroskedasticity: $Var(X_i \varepsilon_i)$ depends on X_i.
  - Covariance terms are not zero if there is correlation across units (clustering problems) or serial correlation (panel problems).

Standard errors: Some heuristics

- Suppose observations i are grouped into $j = 1, \ldots, J$ clusters, and error terms obey

$$\mathbb{E}[\varepsilon_{ij}\varepsilon_{i'j'}] = \begin{cases} \sigma^2 & \text{if } ij = i'j' \\ \rho\sigma^2 & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \Rightarrow AVar(\hat{\beta}) &= \left[X'X\right]^{-1}\left[X'\mathbb{E}[\varepsilon\varepsilon']X\right]\left[X'X\right]^{-1} \\ &= \sigma^2\left[X'X\right]^{-1}\left[I + \rho\left(X'ZZ'X\left[X'X\right]^{-1} - I\right)\right] \end{aligned}$$

  where Z is is an $N \times J$ matrix that identifies cluster membership.

- Bias in variance increases with ρ. "Replicating the data" is like $\rho = 1$.
- In cross section, clustering by industry adjusts for correlated ε_i across firms.
- In panel data, clustering by firm adjusts for correlation within X_i and ε_i over time.
- In panel data, Newey-West and Driscoll-Kraay allow for serial correlation up to K lags and downweight higher order lags.

Table of Contents

Instrumental variables: Motivations

- Suppose sales depend on tech adoption and manager ability:

$$Y_i = \beta D_i + \underbrace{\gamma X_i + \varepsilon_i}_{\eta_i}.$$

- Suppose we do not observe X_i, but we have an instrumental variable Z_i that satisfies
  - **Relevance**: $Cov(Z_i, D_i) \neq 0$.
  - **Exclusion**: $Cov(Z_i, \eta_i) = Cov(Z_i, \gamma X_i + \varepsilon_i) = 0$.

Then we can use instrument to overcome **omitted variable bias**:

$$b_{IV} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_i)} = \frac{Cov(Z_i, Y_i)/Var(Z_i)}{Cov(Z_i, D_i)/Var(Z_i)} = \frac{\pi_{\text{rd}}}{\pi_{\text{fs}}} = \beta,$$

where π_{rd} is coefficient in “reduced-form” regression of Y_i on Z_i and π_{fs} is coefficient in “first-stage” regression of D_i on Z_i.

Instrumental variables: Motivations

- **Simultaneity**: Suppose

$$q_t^D = \beta p_t + \varepsilon_t^D, \qquad \text{(Demand: } \beta < 0\text{)}$$

$$q_t^S = \gamma p_t + \varepsilon_t^S. \qquad \text{(Supply: } \gamma > 0\text{)}$$

  Then:

$$b_{\text{OLS}} = \frac{Cov(p_t, q_t)}{Var(p_t)} = \frac{\sigma_D^2}{\sigma_D^2 + \sigma_S^2}\gamma + \frac{\sigma_S^2}{\sigma_D^2 + \sigma_S^2}\beta.$$

- Identify supply curve using demand-shifters z_t^D that satisfy $Cov(z_t^D, \varepsilon_t^D) \neq 0$ and $Cov(z_t^D, \varepsilon_t^S) = 0$:

$$b_{\text{IV}} = \frac{Cov(z_t^D, q_t)}{Cov(z_t^D, p_t)} = \gamma.$$

- Identify demand curve using supply-shifters z_t^S that satisfy $Cov(z_t^S, \varepsilon_t^S) \neq 0$ and $Cov(z_t^S, \varepsilon_t^D) = 0$:

$$b_{\text{IV}} = \frac{Cov(z_t^S, q_t)}{Cov(z_t^S, p_t)} = \beta.$$

Instrumental variables: Motivations

- **Measurement error**: Suppose $Y_i = \beta X_i^* + \varepsilon_i$, with $Cov(X_i^*, \varepsilon_i) = 0$.
- However, we only observe $X_i = X_i^* + \eta_i$, where $Cov(X_i^*, \eta_i) = 0$.
- Then: $Y_i = \beta(X_i - \eta_i) + \varepsilon_i = \beta X_i + \xi_i$. Note that $Cov(X_i, \xi_i) = -\beta \sigma_\eta^2$, so

$$b_{\text{OLS}} = \frac{Cov(Y_i, X_i)}{Var(X_i)} = \frac{\sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_\eta^2}\beta.$$

- We can use an instrument Z_i for X_i to identify β that satisfies:
    - Relevance: $Cov(Z_i, X_i^*) \neq 0$.
    - Exclusion: $Cov(Z_i, \varepsilon_i) = 0$.
    - Uncorrelated with measurement error: $Cov(Z_i, \eta_i) = 0$.

Instrumental variables: Motivations

- Whenever you use an instrument, need to be clear on:

1. Why OLS is biased (omitted variable, simultaneity, selection, measurement error).
2. What direction OLS is biased.
3. Evidence that relevance condition holds (i.e., strong first stage).
4. Argument for why exclusion restriction holds.

Example IVs: Boehm and Pandalai-Nayar (2022)

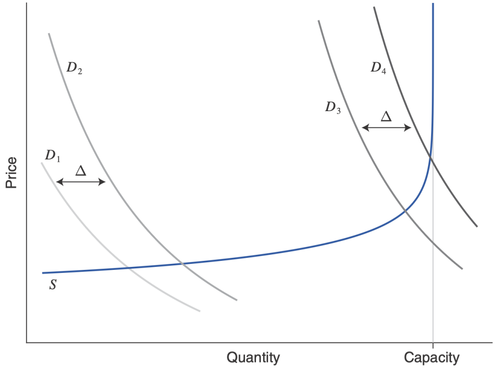


Figure 3. The Utilization Rate as a Sufficient Statistic for the Supply Elasticity

- Boehm and Pandalai-Nayar (2022) argue that industry supply curves are convex.
- Develop model in which supply curve becomes vertical as demand approaches capacity.
- In the model, utilization rate is sufficient statistic for relative slope of the supply curve.

Example IVs: Boehm and Pandalai-Nayar (2022)

- Model yields estimating equation:

$$\Delta \log P_{it} = \alpha + \beta_{Yu}(\Delta \log Y_{it}) \times (u_{it-1} - \bar{u}_i) + \beta_{Qu}(\Delta \log Q_{it}) \times (u_{it-1} - \bar{u}_i) \\ + \beta_Y \Delta \log Y_{it} + \beta_Q \Delta \log Q_{it} + \beta_u(u_{it-1} - \bar{u}_i) + \beta_{mc} \Delta \log mc_{it} + \varepsilon_{it}.$$

- Coefficient β_{Yu} measures how slope of supply curve changes as utilization increases.

- *"In this framework, the initial utilization rate u_{it-1} is a sufficient statistic for the inverse supply elasticity. This feature allows us to estimate the curvature of the supply curve with an interaction term $\Delta \log Y_{it} \cdot (u_{it-1} - \bar{u}_i)$ rather than a squared term in output $(\Delta \log Y_{it})^2$."*

- *"Unlike conventional specifications of supply curves, equation (16) includes the change in capacity as an endogenous supply shifter. All else equal, greater production capacity implies that fewer firms in the industry are constrained and reduces the industry's price index. [...] Since such changes in capacity shift the supply curve, they trace out the demand curve and therefore generate downward-biased slope and curvature estimates when subsumed into the error term."*

Example IVs: Boehm and Pandalai-Nayar (2022)

1. World import demand instrument (Hummels et al. 2014):

$$\mathrm{WID}_{it} = \sum_{j} s_{jit-1} \Delta \log \mathrm{GDP}_{jt},$$

   Time FEs interacted with industries' foreign sales shares absorb global business cycle.

2. Shea (1993) downstream demand instrument:

$$\mathrm{Shea}_{it} = \sum_{j \in \mathcal{J}_i^{\mathrm{Shea}}} s_{jit-1} \Delta \log M_{jt},$$

   where $\mathcal{J}_i^{\mathrm{Shea}}$ is a set of industries such that:

   - j demands a large share of i's output, and materials from i are a small share of j's costs.

3. Effective exchange rate:

$$\Delta e_{it} = \sum_{j} s_{jit-1} \Delta \log E_{jt}.$$

Example IVs: Boehm and Pandalai-Nayar (2022)

TABLE 1—ESTIMATES OF THE LINEAR MODEL

Dependent variable: $\Delta \ln P^Y_{i,t}$					
Estimator:	OLS	OLS	OLS	OLS	2SLS
Instruments:	–	–	–	–	WID, Shea, $\Delta e_{i,t}$
	(1)	(2)	(3)	(4)	(5)
$\Delta \ln Y_{i,t}$	−0.06	0.09	0.13	0.17	0.24
	(0.09)	(0.02)	(0.02)	(0.02)	(0.09)
$\Delta \ln Q_{i,t}$			−0.16	−0.11	−0.16
			(0.03)	(0.03)	(0.08)
$\Delta \ln \mathrm{UVC}_{i,t}$		0.90	0.90	0.89	0.89
		(0.02)	(0.02)	(0.03)	(0.03)
R^2	0.004	0.869	0.876	0.910	0.908
Fixed effects	No	No	No	Yes	Yes
First stage and instrument diagnostics					
F main effect					17.37
Hansen J (p-value)					0.538

Notes: The estimates are based on equation (16). Driscoll-Kraay standard errors are reported in parentheses. Fixed effects include industry fixed effects, time fixed effects, and time fixed effects interacted with industries' lagged foreign sales share $(\sum_{j \in \mathcal{J}^F} s_{j,i,t-1})$. First-stage estimates for column 5 are reported in online Appendix Table D1. 2SLS abbreviates to two-stage least squares.

- "As expected, the slope estimate is insignificantly different from zero since unobserved supply shocks confound the estimation. For instance, an adverse supply shock raises prices while lowering quantities, thereby biasing the slope estimate downward."

Example IVs: Boehm and Pandalai-Nayar (2022)

Table 1—Estimates of the Linear Model

	Dependent variable: $\Delta \ln P^Y_{i,t}$				
Estimator:	OLS	OLS	OLS	OLS	2SLS
Instruments:	–	–	–	–	WID, Shea, $\Delta e_{i,t}$
	(1)	(2)	(3)	(4)	(5)
$\Delta \ln Y_{i,t}$	−0.06	0.09	0.13	0.17	0.24
	(0.09)	(0.02)	(0.02)	(0.02)	(0.09)
$\Delta \ln Q_{i,t}$			−0.16	−0.11	−0.16
			(0.03)	(0.03)	(0.08)
$\Delta \ln \mathrm{UVC}_{i,t}$		0.90	0.90	0.89	0.89
		(0.02)	(0.02)	(0.03)	(0.03)
R^2	0.004	0.869	0.876	0.910	0.908
Fixed effects	No	No	No	Yes	Yes
First stage and instrument diagnostics					
F main effect					17.37
Hansen J (p-value)					0.538

Notes: The estimates are based on equation (16). Driscoll-Kraay standard errors are reported in parentheses. Fixed effects include industry fixed effects, time fixed effects, and time fixed effects interacted with industries' lagged foreign sales share $(\sum_{j \in \mathcal{J}^F} s_{j,i,t-1})$. First-stage estimates for column 5 are reported in online Appendix Table D1. 2SLS abbreviates to two-stage least squares.

- "Controlling for unit variable costs partially addresses the simultaneity problem by removing a large fraction of the confounding variation from the error term. Column 3 further adds the change in capacity to the equation. Its negative coefficient estimate is consistent with the model prediction that, all else equal, industries with greater capacity charge lower prices."

Example IVs: Boehm and Pandalai-Nayar (2022)

Table 1—Estimates of the Linear Model

Dependent variable: $\Delta \ln P^{Y}_{i,t}$					
Estimator:	OLS	OLS	OLS	OLS	2SLS
Instruments:	–	–	–	–	WID, Shea, $\Delta e_{i,t}$
	(1)	(2)	(3)	(4)	(5)
$\Delta \ln Y_{i,t}$	−0.06	0.09	0.13	0.17	0.24
	(0.09)	(0.02)	(0.02)	(0.02)	(0.09)
$\Delta \ln Q_{i,t}$			−0.16	−0.11	−0.16
			(0.03)	(0.03)	(0.08)
$\Delta \ln \mathrm{UVC}_{i,t}$		0.90	0.90	0.89	0.89
		(0.02)	(0.02)	(0.03)	(0.03)
R^2	0.004	0.869	0.876	0.910	0.908
Fixed effects	No	No	No	Yes	Yes
First stage and instrument diagnostics					
F main effect					17.37
Hansen J (p-value)					0.538

Notes: The estimates are based on equation (16). Driscoll-Kraay standard errors are reported in parentheses. Fixed effects include industry fixed effects, time fixed effects, and time fixed effects interacted with industries' lagged foreign sales share $(\sum_{j \in \mathcal{J}^F} s_{j,i,t-1})$. First-stage estimates for column 5 are reported in online Appendix Table D1. 2SLS abbreviates to two-stage least squares.

- "In column 4 we also add industry fixed effects, time fixed effects, and time fixed effects interacted with the industry's export share. The estimate of the inverse supply elasticity rises to 0.17. When we [use the instruments], we obtain a slope estimate of 0.24. It is greater than the OLS estimate, suggesting that despite the controls, the error term in column 4 might still contain supply disturbances."

Example IVs: Boehm and Pandalai-Nayar (2022)

TABLE 2—ESTIMATES OF THE NONLINEAR MODEL

	Dependent variable: $\Delta \ln P^Y_{i,t}$					
Estimator:	OLS	2SLS	2SLS	2SLS	2SLS	2SLS
Instruments						
Main effect:		WID, Shea, $\Delta e_{i,t}$				WID, Shea
Interaction $(\cdot\,(u_{i,t-1} - \bar{u}_i))$:		WID	Shea	$\Delta e_{i,t}$	All	WID, Shea
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \ln Y_{i,t} \cdot (u_{i,t-1} - \bar{u}_i)$	−0.28	1.12	1.20	1.37	1.18	1.17
	(0.24)	(0.30)	(0.60)	(0.67)	(0.29)	(0.29)
$\Delta \ln Y_{i,t}$	0.17	0.28	0.28	0.29	0.27	0.27
	(0.02)	(0.08)	(0.07)	(0.09)	(0.07)	(0.07)
$u_{i,t-1} - \bar{u}_i$	−0.01	0.02	0.02	0.03	0.02	0.02
	(0.02)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
$\Delta \ln Q_{i,t}$	−0.10	−0.23	−0.23	−0.24	−0.22	−0.22
	(0.04)	(0.09)	(0.09)	(0.11)	(0.09)	(0.09)
$\Delta \ln Q_{i,t} \cdot (u_{i,t-1} - \bar{u}_i)$	−0.26	−1.08	−1.12	−1.21	−1.11	−1.10
	(0.41)	(0.44)	(0.44)	(0.51)	(0.43)	(0.42)
$\Delta \ln \mathrm{UVC}_{i,t}$	0.89	0.89	0.89	0.89	0.89	0.89
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
R^2	0.910	0.900	0.899	0.897	0.900	0.900
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
First stage and instrument diagnostics						
F main effect		13.93	37.30	12.46	30.38	26.73
F interaction		26.13	4.18	7.45	69.35	43.41
Cragg-Donald Wald F		8.63	8.17	6.87	6.47	9.49
SW F main effect		37.71	36.39	26.39	28.33	46.41
SW F interaction		27.80	14.85	4.21	32.45	51.58
Hansen J (*p*-value)		0.464	0.440	0.452	0.756	0.748

Table of Contents

Shift-share instruments

- A popular set of instruments in macro take the form of "shift-share IVs":

$$z_i = \sum_{k=1}^{K} \underbrace{s_{ik}}_{\text{Share}} \underbrace{g_k}_{\text{Shift}},$$

  where

  - $(g_1, \ldots, g_K)$ are set of shocks (i.e. shifts) common to all observations,
  - $(s_{i1}, \ldots, s_{iK})$ are exposure weights (i.e. shares) that vary across observations $i = 1, \ldots, N$.

- Shift-share z_i used as instrument for endogenous treatment x_i in structural model

$$y_i = \beta x_i + \gamma' w_i + \varepsilon_i.$$

- Econometric analysis by Goldsmith-Pinkham, Sorkin, and Swift (2020) and Borusyak, Hull, and Jaravel (2022).
  - My discussion will draw from survey by Borusyak, Hull, and Jaravel (2024).

Shift-share instruments: Sample applications

- Bartik (1991), Blanchard and Katz (1992) use SSIV to estimate elasticity of labor supply.
- Card (2009) uses SSIV to estimate effect of migrants on migrant vs. native wages.
- Autor et al. (2013) use SSIV to estimate effect of import penetration from China on local manufacturing employment.
- Chodorow-Reich (2014) uses SSIV to estimate effect of lender shock on employment.
- Nakamura and Steinsson (2014) use SSIV to estimate regional fiscal multiplier.
- Nunn and Qian (2014) use SSIV to estimate effect of food aid from U.S. on conflict.
- Jaravel (2019) uses SSIV to estimate effect of changes in demand on inflation and innovation.
- Chodorow-Reich, Simsek, and Nenov (2021) use SSIV to estimate effect of stock market wealth on consumption.
- Aghion et al. (2022) use SSIV to estimate effect of automation technologies on employment.
- Mohenen (2024) uses SSIV to estimate effect of retirement on job market outcomes for young.

Shift-share instruments: Relevance

- Often, endogenous variable x_i can be decomposed into initial shares and changes.

- Consider Bartik (1991) and Blanchard and Katz (1992), who want to identify regional labor supply elasticity:

$$\underbrace{y_i}_{\substack{\text{Wage}\\\text{growth}}} = \beta \underbrace{x_i}_{\substack{\text{Employment}\\\text{growth}}} + \gamma' w_i + \varepsilon_i.$$

- Employment growth x_i depends on changes in employment ΔX_{ik} across industries:

$$x_i = \frac{\sum_k X_{ik1} - \sum_k X_{ik0}}{\sum_k X_{ik0}} = \sum_k \underbrace{\frac{X_{ik0}}{\sum_{k'} X_{ik'0}}}_{\text{Share}} \underbrace{\frac{X_{ik1} - X_{ik0}}{X_{ik0}}}_{\text{Shift}}.$$

- Therefore, SSIV z_i will be correlated with x_i if exposure shares or shifts/shocks used in z_i are correlated with shares/shifts.

Shift-share instruments: Exclusion

- If exposure shares s_{ik} and shocks g_k used to construct z_i are correlated with shares/shifts in x_i, will guarantee relevance condition.

$$z_i = \sum_{k=1}^{K} \underbrace{s_{ik}}_{\text{Share}} \underbrace{g_k}_{\text{Shift}} .$$

- What about exclusion restriction? In Blanchard and Katz (1992) example:
  - Employment growth x_i can reflect both shifts in labor supply and labor demand.
  - To isolate demand variation, need z_i orthogonal to shocks to local labor supply.

1. **Exogenous shares**: Shares s_{ik} are as-good-as-randomly assigned, and thus affect outcome only through treatment.
2. **Exogenous shocks**: Shares s_{ik} endogenous but shocks g_k exogenous.

Shift-share instruments: Exogenous **shares**

- Shares s_{ik} as-good-as-randomly assigned, i.e., affect outcome only through treatment.
- Analogous to parallel trends assumption: If not for shock, no difference in outcomes.

Shift-share instruments: Exogenous **shares**

- Shares s_{ik} as-good-as-randomly assigned, i.e., affect outcome only through treatment.
- Analogous to parallel trends assumption: If not for shock, no difference in outcomes.
- E.g., Card (2009) studies effect of Mariel Boatlift on relative wages of migrants vs. natives across regions.
  - Exogenous shares $=$ assumption that regions more/less exposed to inflow (as captured by existing Cuban migrant share) would have seen similar trends in relative wages.
- E.g., suppose Kellogg MBA students randomly assigned to sections.
  - Some sections have more students that came from entrepreneurship.
  - Use shares to estimate peer effects of selecting into entrepreneurship after graduation.

Shift-share instruments: Exogenous **shares**

- Shares s_{ik} as-good-as-randomly assigned, i.e., affect outcome only through treatment.
- Analogous to parallel trends assumption: If not for shock, no difference in outcomes.
- E.g., Card (2009) studies effect of Mariel Boatlift on relative wages of migrants vs. natives across regions.
    - Exogenous shares $=$ assumption that regions more/less exposed to inflow (as captured by existing Cuban migrant share) would have seen similar trends in relative wages.
- E.g., suppose Kellogg MBA students randomly assigned to sections.
    - Some sections have more students that came from entrepreneurship.
    - Use shares to estimate peer effects of selecting into entrepreneurship after graduation.
- BHJ recommend "tailoring": i.e., choose shares closely related to research question.
    - Concern is that general shares like industry composition could measure exposure to all other types of unobserved industry shocks.

Shift-share instruments: Exogenous **shares**

1. Are shares plausibly exogenous?
    - One may be tempted to use further lagged shares as more "plausibly exogenous."
    - However, further lagging shares also weakens first stage.
    - Quasi-experiment that leads to variation in shares? (E.g., exogenous variation in why Cuban migrants ended up in some regions vs. others.)
2. Use appropriate additional controls for "conditional" parallel trends.
    - E.g., if exploring effect of Cuban migrant share, control for total migrant share.
3. GPSS Rotemberg weights identify which shares matter most for estimate.
4. Since each of K shares is valid instrument on its own, can explore how pooling shares in different ways affects results.

Shift-share instruments: Exogenous **shocks**

- In many macro applications, approach is to instead argue that shocks are exogenous, even if shares (e.g., industry composition) are endogenous.

- Blanchard and Katz (1992): suppose lottery assigned subsidies g_k to each industry.

- Even though regional industry composition endogenous, the share-weighted average of industry subsidies assigned through the lottery is as-good-as-random.
    - Must have enough industries to ensure law of large numbers.
    - BHJ show that effective number of shocks scales with inverse HHI, $1 / \sum_k \bar{s}_k^2$.
    - If there are a few industries with large average shares, they can dominate the SSIV.
    - Shares must also sum to one (if they don't, regions with higher sum of shares will systematically be more exposed to shocks).

Shift-share instruments: Exogenous **shocks**

- In many macro applications, approach is to instead argue that shocks are exogenous, even if shares (e.g., industry composition) are endogenous.
- Blanchard and Katz (1992): suppose lottery assigned subsidies g_k to each industry.
- Even though regional industry composition endogenous, the share-weighted average of industry subsidies assigned through the lottery is as-good-as-random.
  - Must have enough industries to ensure law of large numbers.
  - BHJ show that effective number of shocks scales with inverse HHI, $1/\sum_k s_k^2$.
  - If there are a few industries with large average shares, they can dominate the SSIV.
  - Shares must also sum to one (if they don't, regions with higher sum of shares will systematically be more exposed to shocks).

Shift-share instruments: Exogenous **shocks**

1. Typically desirable to use shares before shock.
   - Using shares after shock can introduce bias (though does not always).
   - Lagging shares before initial period reduces instrument relevance.
2. Include appropriate shock-level controls.
   - E.g., suppose some sectors more likely to receive subsidy than other sectors.
   - Then, want to control for $\sum_k s_{ik} q_k$, where q_k is control for k's characteristic.
3. Include appropriate unit-level controls.
   - E.g., we might think Rust Belt regions may have differential employment trends.
   - E.g., if shares do not sum to one, include $\sum_k s_{ik}$ control.
4. Balance tests for both units and shocks.
   - Unit-level balance tests (i.e., regions with large shocks similar to those with small shocks).
   - Shock-level balance tests: e.g., since identification relies on equal subsidies possible for all industries, test if industries that have large subsidies systematically different.

Shift-share instruments: Standard errors

- Adao, Kolesar, and Morales (2019) run simulations of Autor-Dorn-Hanson (2013) SSIV.
- OLS specification (reduced form for IV):

$$y_i = \beta z_i + w_i + \varepsilon_i,$$

  where

  - y_i are changes in employment rates and wages from 2000–2007.
  - $z_i = \sum_k s_{ik} g_k$.
  - s_{ik} are employment shares in 1990.
  - g_k are placebo shocks drawn i.i.d. from normal distribution.

- Since shifters are independent of shares and outcomes, true parameter $\beta = 0$.

Shift-share instruments: Standard errors

Table 1: Standard errors and rejection rate of the hypothesis $H_0: \beta = 0$ at 5% significance level.

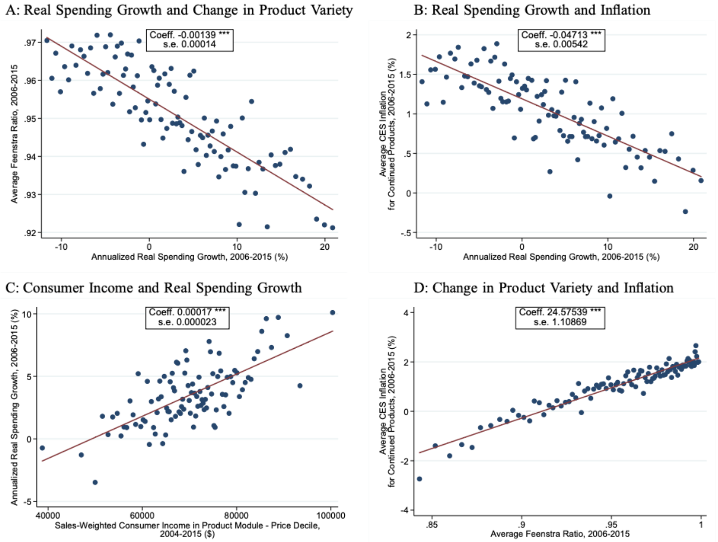
	Estimate		Median std. error		Rejection rate	
	Mean (1)	Std. dev. (2)	Robust (3)	Cluster (4)	Robust (5)	Cluster (6)
Panel A: Change in the share of working-age population						
Employed	−0.01	2.00	0.73	0.92	48.5%	38.1%
Employed in manufacturing	−0.01	1.88	0.60	0.76	55.7%	44.8%
Employed in non-manufacturing	0.00	0.94	0.58	0.67	23.2%	17.6%
Panel B: Change in average log weekly wage						
Employed	−0.03	2.66	1.01	1.33	47.3%	34.2%
Employed in manufacturing	−0.03	2.92	1.68	2.11	26.7%	16.8%
Employed in non-manufacturing	−0.02	2.64	1.05	1.33	45.4%	33.7%

Notes: For the outcome variable indicated in the leftmost column, this table indicates the mean and standard deviation of the OLS estimates of β in eq. (1) across the placebo samples (columns (1) and (2)), the median standard error estimates (columns (3) and (4)), and the percentage of placebo samples for which we reject the null hypothesis $H_0: \beta = 0$ using a 5% significance level test (columns (5) and (6)). *Robust* is the Eicker-Huber-White standard error, and *Cluster* is the standard error that clusters CZs in the same state. Results are based on 30,000 placebo samples.

Shift-share instruments: Standard errors

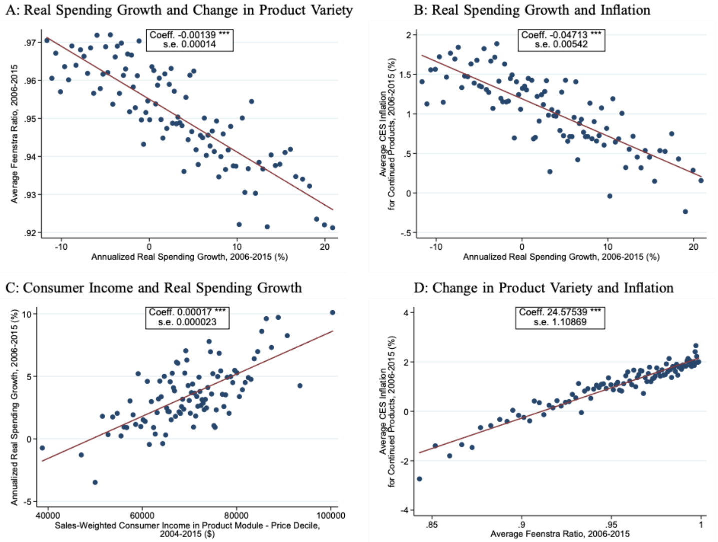
- Heteroskedasticity-robust, geographically clustered SEs show severe overrejection.
- Intuition: Two regions with same shares will mechanically have same z_i and also may have correlated ε_i due to common exposure to unobserved shocks.
- Typical clustering (e.g., by geography) will not capture correlation in error term across geographically separate regions with similar industry composition.
- Adao, Kolesar, and Morales (2019) offer a variance estimator for any correlation structure of errors across regions, as long as shocks are clustered by industry.
  - Corrected standard errors for Autor, Dorn, and Hanson (2013) as 23–66 percent larger.
- BHJ show that one can also run a particular shock-level IV regression that gives an identical coefficient $\hat{\beta}$ as the shift-share regression but gives valid standard errors.

Shift-share instruments: Jaravel (2019)



- Jaravel (2019) documents lower inflation / more varieties in growing product categories.

Shift-share instruments: Jaravel (2019)



- Jaravel (2019) documents lower inflation / more varieties in growing product categories.
- Uses shift-share instrument for changes in demand to isolate causal effect.

Shift-share instruments: Jaravel (2019)

- *"The shift-share design uses variation in Q_l that comes from the variation in the size of household groups only, not from changes in prices or preferences."*

- Define household groups using age, education, race, presence of children, and state.

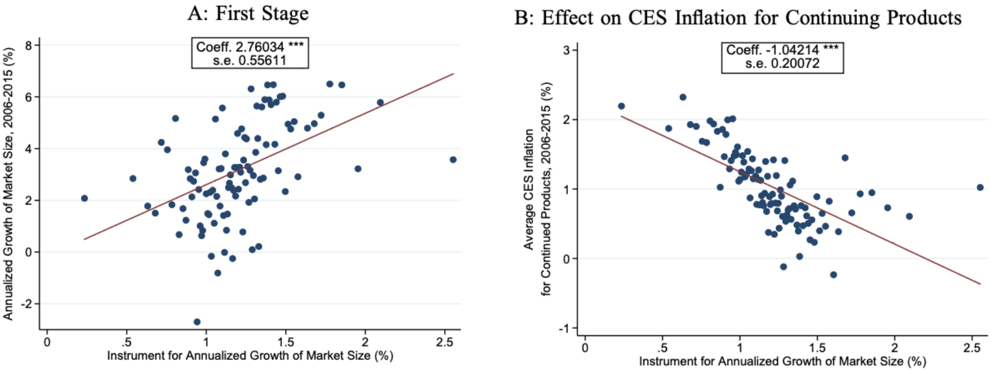
- Let g^h indicate change in size of household group h from 2000–2004 to 2012–2016.

- Use first year data (2004) to construct baseline spending share from each group.

- *"During the sample period older household groups tend to grow faster (i.e., they have a high g^h). It may be intrinsically more difficult to innovate and reduce prices in product categories that sell more to older household groups, for instance because these households are likely to have defined their tastes earlier in life and to be less likely to adopt new products."*

- Approach: Construct residualized shocks after controlling for age FEs, etc.

Shift-share instruments: Jaravel (2019)



- Baseline results suggest 1pp increase in demand lead to 0.418pp fall in inflation.
- Compare to OLS estimate of 0.047!

Summary

- Empirical macroeconomics uses applied micro tools to inform macro models.
- Intuitions about omitted variable bias, measurement error, correlation in residuals, relevance / exclusion all very useful to guide research.
- Next class, we will talk about aggregating micro estimates.