MIT 14.582: PhD International Economics II Sp 2026, Lectures 18-19: Economic Geography and Urban Economics (Urban Models)

Dave Donaldson

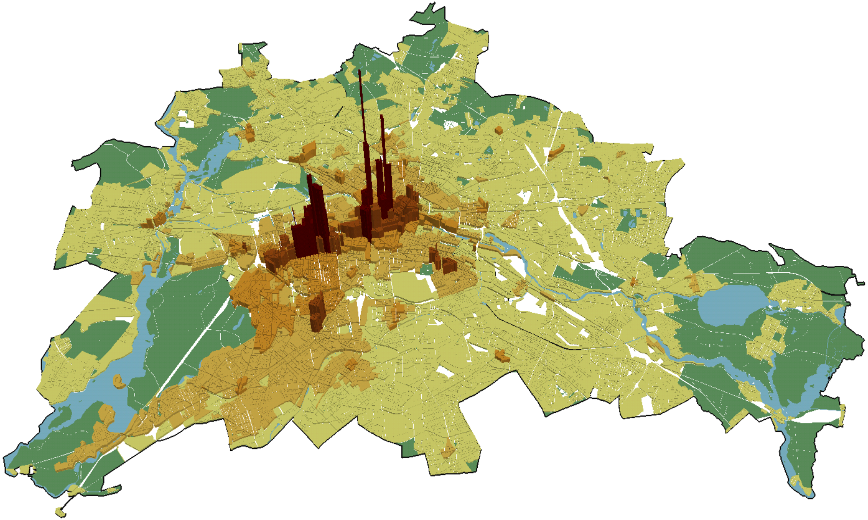
Plan for This Lecture

- Explore a model of within-city economic geography: urban economics
- Detailed look at Ahlfeldt, Redding, Sturm and Wolf (ECMA, 2015)
- Appendix material on estimating and testing quantitative spatial models (from a lecture I gave at the 2024 UEA Summer School on "how much should we trust QSM estimates?")

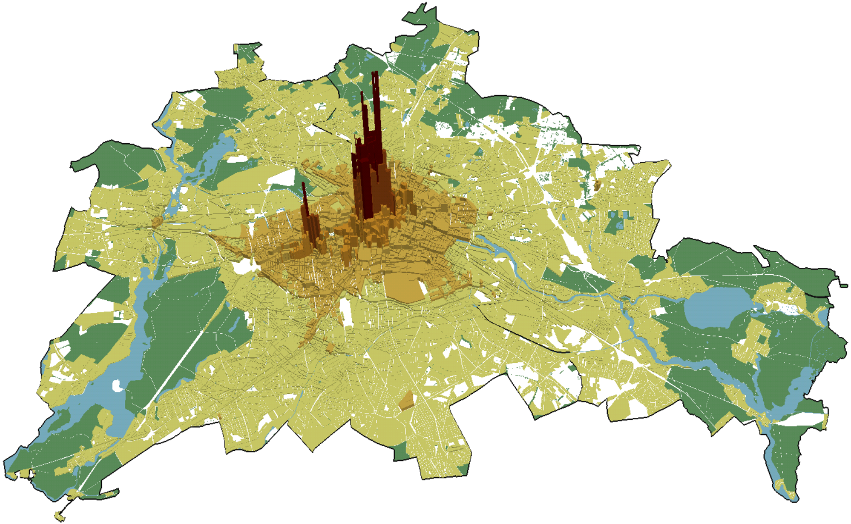
What Makes Urban Different from Economic Geography?

- No simple answers of course
- But some common model features will be different:
  - Free migration in/out of the city
  - Commuting costs within the city (so choosing both where to live and where to work)
  - Housing is important (either in fixed supply or requiring land that is in fixed supply)
  - Some goods freely tradable, some non-tradable (like housing)
  - Small spatial units so cross-location externality spillovers allowed for
- For more on Urban, great resources include:
  - Glaeser (2011), *Triumph of the City*
  - Glaeser and Gottlieb (JEL 2008)—covered in recitation
  - Textbooks such as Glaeser (2008), Brueckner (2011) and Fujita and Thisse (2012)
  - Redding chapter in 2025 *Handbook of Regional and Urban Economics*

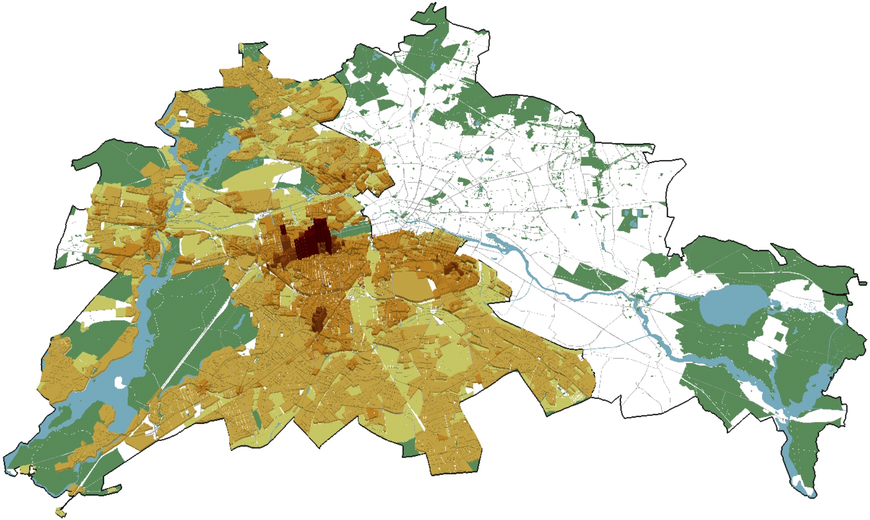
The type of thing we are trying to model: land rents in Berlin in 2006



The type of thing we are trying to model: land rents in Berlin in 1936



The type of thing we are trying to model: land rents in West Berlin in 1986



Monocentricity of Cities

- All three of these city-years show strong evidence for near-monocentricity of land rents

- This is a common pattern for:
  - Cities around the world and throughout history
  - And other outcomes, such as density of residency and employment

- These patterns are predicted by canonical urban models such as:
  - Linear city: Alonso (1964), Muth (1969), Mills (1967)
  - Disk city: Lucas and Rossi-Hansberg (2002)
  - Arises from equilibrium interaction between scarce land, commuting costs, and local (within-city) agglomeration economies

- Is there evidence for the implied causal effect of proximity to the "center" on outcomes like rents and density?

Ahlfeldt, Redding, Sturm and Wolf (2015)

- Exploits the fact that Berlin was divided into East Berlin and West Berlin during the cold war
  - Boundaries drawn post-1945
  - But in 1961 West Berlin segregated into its own city by Berlin Wall: very limited interactions with rest of larger Berlin
  - Reunification of East and West Berlin (with full economic integration) in 1989
- Paper studies locations ("blocks") in West Berlin before and after the division shock (36-66 or -86) and the reunification shock (86-06)
- Estimates diff-in-diff regressions relating changes in outcomes (like rents or density) to distance from block to "Mitte" (the original central business district of unified Berlin, CBD)

Diff-in-diff on Division of Berlin

(Q = price of housing ("floor price"), $EmpR$ = number of residents, $EmpW$ = number of workers)

TABLE I

BASELINE DIVISION DIFFERENCE-IN-DIFFERENCE RESULTS (1936–1986)[a]

	(1) $\Delta \ln Q$	(2) $\Delta \ln Q$	(3) $\Delta \ln Q$	(4) $\Delta \ln Q$	(5) $\Delta \ln Q$	(6) Δ ln EmpR	(7) Δ ln EmpR	(8) Δ ln EmpW	(9) Δ ln EmpW
CBD 1	−0.800***	−0.567***	−0.524***	−0.503***	−0.565***	−1.332***	−0.975***	−0.691*	−0.639*
	(0.071)	(0.071)	(0.071)	(0.071)	(0.077)	(0.383)	(0.311)	(0.408)	(0.338)
CBD 2	−0.655***	−0.422***	−0.392***	−0.360***	−0.400***	−0.715**	−0.361	−1.253***	−1.367***
	(0.042)	(0.047)	(0.046)	(0.043)	(0.050)	(0.299)	(0.280)	(0.293)	(0.243)
CBD 3	−0.543***	−0.306***	−0.294***	−0.258***	−0.247***	−0.911***	−0.460**	−0.341	−0.471**
	(0.034)	(0.039)	(0.037)	(0.032)	(0.034)	(0.239)	(0.206)	(0.241)	(0.190)
CBD 4	−0.436***	−0.207***	−0.193***	−0.166***	−0.176***	−0.356**	−0.259	−0.512***	−0.521***
	(0.022)	(0.033)	(0.033)	(0.030)	(0.026)	(0.145)	(0.159)	(0.199)	(0.169)
CBD 5	−0.353***	−0.139***	−0.123***	−0.098***	−0.100***	−0.301***	−0.143	−0.436***	−0.340***
	(0.016)	(0.024)	(0.024)	(0.023)	(0.020)	(0.110)	(0.113)	(0.151)	(0.124)
CBD 6	−0.291***	−0.125***	−0.094***	−0.077***	−0.090***	−0.360***	−0.135	−0.280**	−0.142
	(0.018)	(0.019)	(0.017)	(0.016)	(0.016)	(0.100)	(0.089)	(0.130)	(0.116)
Inner Boundary 1–6			Yes	Yes	Yes		Yes		Yes
Outer Boundary 1–6			Yes	Yes	Yes		Yes		Yes
Kudamm 1–6				Yes	Yes		Yes		Yes
Block Characteristics					Yes		Yes		Yes
District Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,260	6,260	6,260	6,260	6,260	5,978	5,978	2,844	2,844
R^2	0.26	0.51	0.63	0.65	0.71	0.19	0.43	0.12	0.33

[a] Q denotes the price of floor space. EmpR denotes employment by residence. EmpW denotes employment by workplace. CBD1–CBD6 are six 500 m distance grid cells for distance from the pre-war CBD. Inner Boundary 1–6 are six 500 m grid cells for distance to the Inner Boundary between East and West Berlin. Outer Boundary 1–6 are six 500 m grid cells for distance to the outer boundary between West Berlin and East Germany. Kudamm 1–6 are six 500 m grid cells for distance to Breitscheid Platz on the Kurfürstendamm. The coefficients on the other distance grid cells are reported in Table A.2 of the Technical Data Appendix. Block characteristics include the log distance to schools, parks and water, the land area of the block, the share of the block's built-up area destroyed during the Second World War, indicators for residential, commercial and industrial land use, and indicators for whether a block includes a government building and urban regeneration policies post-reunification. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors in parentheses (Conley (1999)). * significant at 10%; ** significant at 5%; *** significant at 1%.

Diff-in-diff on Reunification of Berlin

TABLE II

BASELINE REUNIFICATION DIFFERENCE-IN-DIFFERENCE RESULTS (1986–2006)[a]

	(1) $\Delta \ln Q$	(2) $\Delta \ln Q$	(3) $\Delta \ln Q$	(4) $\Delta \ln Q$	(5) $\Delta \ln Q$	(6) $\Delta \ln \text{EmpR}$	(7) $\Delta \ln \text{EmpR}$	(8) $\Delta \ln \text{EmpW}$	(9) $\Delta \ln \text{EmpW}$
CBD 1	0.398***	0.408***	0.368***	0.369***	0.281***	1.079***	1.025***	1.574***	1.249**
	(0.105)	(0.090)	(0.083)	(0.081)	(0.088)	(0.307)	(0.297)	(0.479)	(0.517)
CBD 2	0.290***	0.289***	0.257***	0.258***	0.191**	0.589*	0.538*	0.684**	0.457
	(0.111)	(0.096)	(0.090)	(0.088)	(0.087)	(0.315)	(0.299)	(0.326)	(0.334)
CBD 3	0.122***	0.120***	0.110***	0.115***	0.063**	0.340*	0.305*	0.326	0.158
	(0.037)	(0.033)	(0.032)	(0.032)	(0.028)	(0.180)	(0.158)	(0.216)	(0.239)
CBD 4	0.033***	0.031	0.030	0.034	0.017	0.110	0.034	0.336**	0.261
	(0.013)	(0.023)	(0.022)	(0.021)	(0.020)	(0.068)	(0.066)	(0.161)	(0.185)
CBD 5	0.025***	0.018	0.020	0.020	0.015	−0.012	−0.056	0.114	0.066
	(0.010)	(0.015)	(0.014)	(0.014)	(0.013)	(0.056)	(0.057)	(0.118)	(0.131)
CBD 6	0.019**	−0.000	−0.000	−0.003	0.005	0.060	0.053	0.049	0.110
	(0.009)	(0.012)	(0.012)	(0.012)	(0.011)	(0.039)	(0.041)	(0.095)	(0.098)
Inner Boundary 1–6			Yes	Yes	Yes		Yes		Yes
Outer Boundary 1–6			Yes	Yes	Yes		Yes		Yes
Kudamm 1–6				Yes	Yes		Yes		Yes
Block Characteristics					Yes		Yes		Yes
District Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,050	7,050	7,050	7,050	7,050	6,718	6,718	5,602	5,602
R^2	0.08	0.32	0.34	0.35	0.43	0.04	0.07	0.03	0.06

[a] Q denotes the price of floor space. EmpR denotes employment by residence. EmpW denotes employment by workplace. CBD1–CBD6 are six 500 m distance grid cells for distance from the pre-war CBD. Inner Boundary 1–6 are six 500 m grid cells for distance to the Inner Boundary between East and West Berlin. Outer Boundary 1–6 are six 500 m grid cells for distance to the outer boundary between West Berlin and East Germany. Kudamm 1–6 are six 500 m grid cells for distance to Breitscheid Platz on the Kurfürstendamm. The coefficients on the other distance grid cells are reported in Table A.4 of the Technical Data Appendix. Block characteristics include the log distance to schools, parks and water, the land area of the block, the share of the block's built-up area destroyed during the Second World War, indicators for residential, commercial and industrial land use, and indicators for whether a block includes a government building and urban regeneration policies post-reunification. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors in parentheses (Conley (1999)). * significant at 10%; ** significant at 5%; *** significant at 1%.

Diff-in-diff on Division and Reunification of Berlin

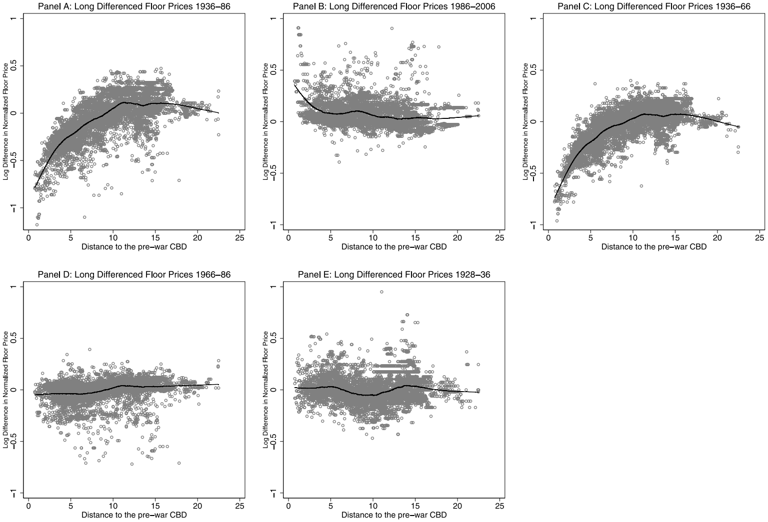


FIGURE 3.—Division and reunification treatments and placebos. Note: Log floor prices are normalized to have a mean of zero in each year before taking the long difference. Solid lines are fitted values from locally-weighted linear least squares regressions.

Ahlfeldt, Redding, Sturm and Wolf (2015)

- ARSW propose a model that:
  - Has all of the core features of urban economics mentioned above
  - And nests the AMM framework (monocentricity) but with a finite number of locations on a realistic geography

- Goals?
  - Workhorse quantitative model for urban settings (same "quantitative model" characteristics as we covered in last lecture: lots of scope for exogenous heterogeneity and realistic geographic features)
  - A tool for mapping data into "structural" parameter estimates
  - Counterfactual simulations using estimated model—useful to the extent that the counterfactual questions are not already answered by the previous diff-in-diff regressions

Consumption

- Utility for worker o residing in block i and working in block j:

$$U_{ijo} = \frac{B_i z_{ijo}}{d_{ij}} \left(\frac{c_{ij}}{\beta}\right)^{\beta} \left(\frac{\ell_{ij}}{1-\beta}\right)^{1-\beta}, \qquad 0 < \beta < 1,$$

  - Consumption of the final good (c_{ij}), chosen as numeraire ($p_i = 1$)
  - Residential floor space (ℓ_{ij})
  - Residential amenity B_i
  - Commuting costs $d_{ij} = e^{\kappa \tau_{ij}}$, where τ_{ij} is travel time
  - Idiosyncratic shock z_{ijo} that captures idiosyncratic reasons for worker o to live in block i and work in block j
- Indirect utility

$$U_{ijo} = \frac{z_{ijo} B_i w_j Q_i^{\beta-1}}{d_{ij}},$$

- The idiosyncratic shock to worker utility is drawn from a Fréchet distribution:

$$F(z_{ijo}) = e^{-T_i E_j z_{ijo}^{-\epsilon}}, \qquad T_i, E_j > 0, \; \epsilon > 0,$$

Commuting Decisions

- Probability worker chooses to live in block i and work in block j is:

$$\pi_{ij} = \frac{T_i E_j \left(d_{ij} Q_i^{1-\beta} \right)^{-\epsilon} (B_i w_j)^{\epsilon}}{\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s \left(d_{rs} Q_r^{1-\beta} \right)^{-\epsilon} (B_r w_s)^{\epsilon}} \equiv \frac{\Phi_{ij}}{\Phi}. \tag{1}$$

- Residential and workplace choice probabilities

$$\pi_{Ri} \equiv \sum_{j=1}^{S} \pi_{ij} = \frac{\sum_{j=1}^{S} \Phi_{ij}}{\Phi}, \qquad \pi_{Mj} \equiv \sum_{i=1}^{S} \pi_{ij} = \frac{\sum_{i=1}^{S} \Phi_{ij}}{\Phi}.$$

- Conditional on living in block i, the probability that a worker commutes to block j follows a gravity equation:

$$\pi_{ij|i} = \frac{E_j \left(w_j / d_{ij} \right)^{\epsilon}}{\sum_{s=1}^{S} E_s \left(w_s / d_{is} \right)^{\epsilon}}, \tag{2}$$

Commuting Decisions

- Workplace employment in block j (H_{Mj}) equals the sum across all blocks i of residence "employment" (H_{Ri}) times the probability of commuting from i to j, conditional on living in i (equation 2):

$$H_{Mj} = \sum_{i=1}^{S} \frac{E_j \left(w_j/d_{ij}\right)^{\epsilon}}{\sum_{s=1}^{S} E_s \left(w_s/d_{is}\right)^{\epsilon}} H_{Ri}$$

- Expected utility ($\gamma =$ Euler's constant)

$$\mathbb{E}\left[U\right] = \gamma \left[\sum_{r=1}^{S} \sum_{s=1}^{S} T_r E_s \left(d_{rs} Q_r^{1-\beta} \right)^{-\epsilon} \left(B_r w_s \right)^{\epsilon} \right]^{1/\epsilon}$$

- City residents freely mobile across cities so get $\mathbb{E}\left[U\right] = \bar{U}$ for some exogenous $\bar{U}$

Production

- A single final good (taken to be the numeraire) is produced under conditions of perfect competition and zero trade costs (both within the city and with respect to other cities)
- Production function for rep. firm in location j is:

$$y_j = A_j \left(H_{Mj}\right)^{\alpha} \left(L_{Mj}\right)^{1-\alpha}, \qquad 0 < \alpha < 1,$$

- Where:
  - H_{Mj} is workplace employment (as above)
  - L_{Mj} is floor space used commercially

Land Market Clearing

- Floor space in location i (L_i) can be allocated to either residential (at price Q_i) or commercial (at price q_i) use.
  - Let θ_i be share put to commercial use.
  - Assume floor space will be put to its most profitable use (so actual price is $\max\{Q_i, q_i\}$)
  - Paper also allows for a tax (or quantity restriction) wedge between commercial and residential uses

- Floor space produced competitively by rep. firm in each location with production function:

$$L_i = M_i^{\mu} K_i^{1-\mu}$$

- Where:
  - M_i is capital; assumed to be perfectly elastically supplied to the city
  - K_i is land; in fixed supply in each location i

Externalities

- Like we saw in earlier lectures, allow for two types of agglomeration externalities
- Amenities (B_i) depend on fundamentals (b_i) and spillovers:

$$B_i = b_i \Omega_i^{\eta}, \qquad \Omega_i \equiv \left[\sum_{s=1}^{S} e^{-\rho \tau_{is}} \left(\frac{H_{Rs}}{K_s} \right) \right]$$

- Productivities (A_j) depend on fundamentals (a_j) and spillovers:

$$A_j = a_j \Upsilon_j^{\lambda}, \qquad \Upsilon_j \equiv \left[\sum_{s=1}^{S} e^{-\delta \tau_{js}} \left(\frac{H_{Ms}}{K_s} \right) \right]$$

- Note that in both cases, the spillovers spill beyond the location itself (when $\delta < \infty$ or $\rho < \infty$)

Equilibrium Existence and Uniqueness

- ARSW include a proof of this for the no-spillovers ($\eta = \lambda = 0$) case
- Allen, Arkolakis and Li (2025) show how their method (that we saw last lecture) for studying existence/uniqueness can be applied to the ARSW model
- As we'd expect from previous lectures, can still have uniqueness despite positive spillovers if they are not strong enough to overcome the "congestion forces":
  - Exogenous heterogeneity in taste for locations (really, work-home pairs): low ϵ means more heterogeneity, so more likely to get uniqueness
  - Land in fixed supply: low α and low β make floorspace important, low μ makes fixed land more important for producing floorspace; combination of these being low means more likely to get uniqueness.

Structural Estimation

- Unknown parameters to be estimated are:
  - Elasticities: ϵ, κ, η, ρ, λ, δ (estimation described below)
  - Cobb-Douglas shares: α, β, μ (calibrated to shares in data)
  - Primitives: B_i, T_i, E_i, A_i (residuals after the others are known)

- ARSW propose a recursive procedure for estimating the elasticities above

- ARSW's estimation method was pioneering (c. 2015) because part of the estimation procedure relied on exogeneity assumptions that are similar to those evoked in the earlier diff-in-diff "reduced-form" estimation
  - Such a practice has become more common since 2015
  - But ARSW still remains a great example of how to think this through

Step #1: Commuting Gravity

- First step: estimate composite elasticity $\nu \equiv \epsilon\kappa$
- Note that commuting flows from residence i to workplace j take the gravity form (taking log of equation 1):

$$\ln \pi_{ij} = -\nu\tau_{ij} + \vartheta_i + \varsigma_j + e_{ij}, \tag{3}$$

- Where:
  - ϑ_i are residence fixed effects
  - ς_j are workplace fixed effects
- ARSW estimate this equation using survey of commuting flows:
  - From 2008—so assuming that ν is stable back to 1936
  - Survey reports origin, destination, travel times
  - At the district ($N = 12$) level. So paper discusses aggregation bias due to estimating block-level gravity model from district-level data; reports that bias is small in their model-based simulations.

Commuting Gravity Equation: Results

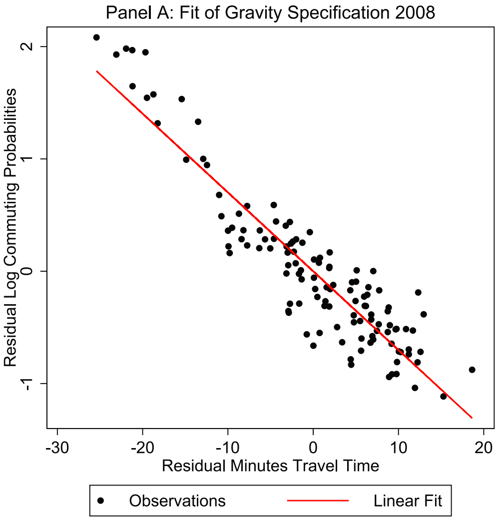
TABLE III

COMMUTING GRAVITY EQUATION[a]

	(1) ln Bilateral Commuting Probability 2008	(2) ln Bilateral Commuting Probability 2008	(3) ln Bilateral Commuting Probability 2008	(4) ln Bilateral Commuting Probability 2008
Travel Time ($-\kappa\varepsilon$)	-0.0697^{***}	-0.0702^{***}	-0.0771^{***}	-0.0706^{***}
	(0.0056)	(0.0034)	(0.0025)	(0.0026)
Estimation	OLS	OLS	Poisson PML	Gamma PML
More than 10 Commuters		Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Observations	144	122	122	122
R^2	0.8261	0.9059	–	–

[a] Gravity equation estimates based on representative micro survey data on commuting for Greater Berlin for 2008. Observations are bilateral pairs of 12 workplace and residence districts (post 2001 Bezirke boundaries). Travel time is measured in minutes. Fixed effects are workplace district fixed effects and residence district fixed effects. The specifications labelled more than 10 commuters restrict attention to bilateral pairs with 10 or more commuters. Poisson PML is Poisson Pseudo Maximum Likelihood estimator. Gamma PML is Gamma Pseudo Maximum Likelihood Estimator. Standard errors in parentheses are heteroscedasticity robust. * significant at 10%; ** significant at 5%; *** significant at 1%.

Commuting Gravity Equation: Fit



Note: Residuals from conditioning on workplace and residence fixed effects.

Step #2: Uncovering (adjusted) wages

- Recall that "commuting market clearing equation" was (in any time period t):

$$H_{Mjt} = \sum_{i=1}^{S} \frac{E_{jt}\left(w_{jt}/d_{ijt}\right)^{\epsilon}}{\sum_{s=1}^{S} E_{st}\left(w_{st}/d_{ist}\right)^{\epsilon}} H_{Rit}$$

- With data on H_{Mjt} and H_{Rjt} (number of residents and number of employees, by block), and with a measure of $d_{ijt}^{-\epsilon} = e^{-\nu\tau_{ijt}}$, can solve this system of equations for "adjusted wages" $\omega_{jt} \equiv E_{jt} w_{jt}^{\epsilon}$. ARSW show that this solution exists and is unique.

- ARSW measure τ_{ijt} using:
  - Travel time computed from map of roads (with speeds), train/subway networks (with schedules), etc. in each time period t
  - With travel time computed using the Dijkstra (shortest-path) algorithm in each time period.

Step #3: Estimating ϵ

- ARSW estimate this parameter using micro-data on wages from 1986. Let $\sigma^2_{\ln w_i}$ denote the variance of log wages across workplace locations in 1986.
- Then use the moment:

$$E[(1/\epsilon^2)\ln(\omega_i)^2 - \sigma^2_{\ln w_i}] = 0$$

- Since $\omega_{jt} \equiv E_{jt} w_{jt}^{\epsilon}$ this is assuming that $Var(\ln E_{jt}) = 0$.
- They stack this moment with a longer list of moments and use (over-identified) GMM; results below.

Step #4: Uncovering productivity and amenity terms

- Firms' FOCs can be written as:

$$\ln \widetilde{A}_{it} = \chi_t + (1-\alpha)\ln \widehat{Q}_{it} + \frac{\alpha}{\epsilon}\ln \omega_{it}$$

  - Where $\widetilde{A}_{it} \equiv A_i E_i^{\alpha/\epsilon}$, $\widehat{Q}_{it} \equiv \max\{q_{it}, Q_{it}\}$, and χ_t is a year fixed-effect.
- And the labor mobility and commuting expressions can be written as:

$$\ln \widetilde{B}_{it} = \eta_t + \frac{1}{\epsilon}\ln H_{Rit} + (1-\beta)\ln \widehat{Q}_{it} - \ln W_{it}$$

  - Where $\widetilde{B}_{it} \equiv B_i T_i^{1/\epsilon}$, $W_{it} \equiv \sum_s \omega_{st} e^{\widehat{\nu}\tau_{ist}}$, and η_t is a year fixed-effect
- So with data on $\widehat{Q}_{it}$ and estimates of parameters ϵ, α and β, can solve for the productivity and amenity terms $\widetilde{B}_{it}$ and $\widetilde{A}_{it}$ (up to the year-specific fixed effects, which won't matter for anything)

Do productivity ($\widetilde{A}_{it}$) and amenity ($\widetilde{B}_{it}$) terms correlate with diff-in-diff Berlin Wall treatment (CBD distance)?

TABLE IV

PRODUCTIVITY, AMENITIES, AND COUNTERFACTUAL FLOOR PRICES[a]

	(1) $\Delta \ln A$ 1936–1986	(2) $\Delta \ln B$ 1936–1986	(3) $\Delta \ln A$ 1986–2006	(4) $\Delta \ln B$ 1986–2006	(5) $\Delta \ln QC$ 1936–1986	(6) $\Delta \ln QC$ 1986–2006
CBD 1	−0.207***	−0.347***	0.261***	0.203***	−0.408***	−0.010
	(0.049)	(0.070)	(0.073)	(0.054)	(0.038)	(0.020)
CBD 2	−0.260***	−0.242***	0.144**	0.109*	−0.348***	0.079**
	(0.032)	(0.053)	(0.056)	(0.058)	(0.017)	(0.036)
CBD 3	−0.138***	−0.262***	0.077***	0.059**	−0.353***	0.036
	(0.021)	(0.037)	(0.024)	(0.026)	(0.022)	(0.031)
CBD 4	−0.131***	−0.154***	0.057***	0.010	−0.378***	0.093***
	(0.016)	(0.023)	(0.015)	(0.008)	(0.021)	(0.026)
CBD 5	−0.095***	−0.126***	0.028**	−0.014*	−0.380***	0.115***
	(0.014)	(0.013)	(0.013)	(0.007)	(0.022)	(0.033)
CBD 6	−0.061***	−0.117***	0.023**	0.001	−0.354***	0.066***
	(0.015)	(0.015)	(0.010)	(0.005)	(0.018)	(0.023)
Counterfactuals					Yes	Yes
Agglomeration Effects					No	No
Observations	2,844	5,978	5,602	6,718	6,260	7,050
R^2	0.09	0.06	0.02	0.03	0.07	0.03

[a]Columns (1)–(4) based on calibrating the model for $\nu = \varepsilon\kappa = 0.07$ and $\varepsilon = 6.83$ from the gravity equation estimation. Columns (5)–(6) report counterfactuals for these parameter values. A denotes adjusted overall productivity. B denotes adjusted overall amenities. QC denotes counterfactual floor prices (simulating the effect of division on West Berlin). Column (5) simulates division holding A and B constant at their 1936 values. Column (6) simulates reunification holding A and B for West Berlin constant at their 1986 values and using 2006 values of A and B for East Berlin. CBD1–CBD6 are six 500 m distance grid cells for distance from the pre-war CBD. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors in parentheses (Conley (1999)). * significant at 10%; ** significant at 5%; *** significant at 1%.

Step #4: Constructing moments

- Now let $\widetilde{a}_{it} \equiv a_{it} E_{it}^{\alpha/\epsilon}$ and assume that exogenous component ($\widetilde{a}_{it}$) of the productivity terms A_i does not change before/after the Berlin Wall is built/removed in a way that is correlated with distance to the CBD.

- That is, for each distance band (from the CDB) I_k assume:

$$E[I_k \times \Delta \ln \widetilde{a}_{it}] = 0 \tag{4}$$

- And $\ln \widetilde{a}_{it} = \ln \widetilde{A}_{it} - \lambda \ln \Upsilon_{it}$. So given $\widetilde{A}_{it}$ identified in Step #3, can therefore use moments (4) to identify λ and δ.

- The exogeneity assumptions in (4) are analogous to those that identify the diff-in-diff responses we studied earlier.

- Similar approach applies to amenity side, $\widetilde{B}_{it}$, to identify η and ρ.

- **Question:** How does this work if there is multiplicity?

Estimated Parameters (GMM)

TABLE V

GENERALIZED METHOD OF MOMENTS (GMM) ESTIMATION RESULTS[a]

	(1) Division Efficient GMM	(2) Reunification Efficient GMM	(3) Division and Reunification Efficient GMM
Commuting Travel Time Elasticity ($\kappa\varepsilon$)	0.0951***	0.1011***	0.0987***
	(0.0016)	(0.0016)	(0.0016)
Commuting Heterogeneity (ε)	6.6190***	6.7620***	6.6941***
	(0.0939)	(0.1005)	(0.0934)
Productivity Elasticity (λ)	0.0793***	0.0496***	0.0710***
	(0.0064)	(0.0079)	(0.0054)
Productivity Decay (δ)	0.3585***	0.9246***	0.3617***
	(0.1030)	(0.3525)	(0.0782)
Residential Elasticity (η)	0.1548***	0.0757**	0.1553***
	(0.0092)	(0.0313)	(0.0083)
Residential Decay (ρ)	0.9094***	0.5531	0.7595***
	(0.2968)	(0.3979)	(0.1741)

[a]Generalized Method of Moments (GMM) estimates. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors in parentheses (Conley (1999)). * significant at 10%; ** significant at 5%; *** significant at 1%.

Estimated Parameters (GMM)—Implications

TABLE VI

EXTERNALITIES AND COMMUTING COSTS[a]

	(1) Production Externalities $(1 \times e^{-\delta\tau})$	(2) Residential Externalities $(1 \times e^{-\rho\tau})$	(3) Utility After Commuting $(1 \times e^{-\kappa\tau})$
0 minutes	1.000	1.000	1.000
1 minute	0.696	0.468	0.985
2 minutes	0.485	0.219	0.971
3 minutes	0.338	0.102	0.957
5 minutes	0.164	0.022	0.929
7 minutes	0.079	0.005	0.902
10 minutes	0.027	0.001	0.863
15 minutes	0.004	0.000	0.802
20 minutes	0.001	0.000	0.745
30 minutes	0.000	0.000	0.642

[a] Proportional reduction in production and residential externalities with travel time and proportional reduction in utility from commuting with travel time. Travel time is measured in minutes. Results are based on the pooled efficient GMM parameter estimates: $\delta = 0.3617$, $\rho = 0.7595$, $\kappa = 0.0148$.

What about testing the model?

TABLE VII

COUNTERFACTUALS[a]

	(1) $\Delta \ln QC$ 1936–1986	(2) $\Delta \ln QC$ 1936–1986	(3) $\Delta \ln QC$ 1936–1986	(4) $\Delta \ln QC$ 1936–1986	(5) $\Delta \ln QC$ 1986–2006	(6) $\Delta \ln QC$ 1986–2006	(7) $\Delta \ln QC$ 1986–2006
CBD 1	−0.836***	−0.613***	−0.467***	−0.821***	0.363***	1.160***	0.392***
	(0.052)	(0.032)	(0.060)	(0.051)	(0.041)	(0.052)	(0.043)
CBD 2	−0.560***	−0.397***	−0.364***	−0.624***	0.239***	0.779***	0.244***
	(0.034)	(0.025)	(0.019)	(0.029)	(0.028)	(0.044)	(0.027)
CBD 3	−0.455***	−0.312***	−0.336***	−0.530***	0.163***	0.594***	0.179***
	(0.036)	(0.030)	(0.030)	(0.036)	(0.031)	(0.045)	(0.031)
CBD 4	−0.423***	−0.284***	−0.340***	−0.517***	0.140***	0.445***	0.143***
	(0.026)	(0.019)	(0.022)	(0.031)	(0.021)	(0.042)	(0.021)
CBD 5	−0.418***	−0.265***	−0.351***	−0.512***	0.177***	0.403***	0.180***
	(0.032)	(0.022)	(0.027)	(0.039)	(0.032)	(0.038)	(0.032)
CBD 6	−0.349***	−0.222***	−0.304***	−0.430***	0.100***	0.334***	0.103***
	(0.025)	(0.016)	(0.022)	(0.029)	(0.024)	(0.034)	(0.023)
Counterfactuals	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agglomeration Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,260	6,260	6,260	6,260	7,050	6,260	7,050
R^2	0.11	0.13	0.07	0.13	0.12	0.24	0.13

[a] Columns (1)–(6) are based on the parameter estimates pooling division and reunification from Table V. Column (7) is based on the parameter estimates for division from Table V. QC denotes counterfactual floor prices. Column (1) simulates division using our estimates of production and residential externalities and 1936 fundamentals. Column (2) simulates division using our estimates of production externalities and 1936 fundamentals but setting residential externalities to zero. Column (3) simulates division using our estimates of residential externalities and 1936 fundamentals but setting production externalities to zero. Column (4) simulates division using our estimates of production and residential externalities and 1936 fundamentals but halving their rates of spatial decay with travel time. Column (5) simulates reunification using our estimates of production and residential externalities, 1986 fundamentals for West Berlin, and 2006 fundamentals for East Berlin. Column (6) simulates reunification using our estimates of production and residential externalities, 1986 fundamentals for West Berlin and 1936 fundamentals for East Berlin. Column (7) simulates reunification using division rather than pooled parameter estimates, 1986 fundamentals for West Berlin, and 2006 fundamentals for East Berlin. CBD 1–CBD6 are six 500 m distance grid cells for distance from the pre-war CBD. Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors in parentheses (Conley (1999)). * significant at 10%; ** significant at 5%; *** significant at 1%.

Counterfactuals and Subsequent Work

- Not main focus of paper, but can now put the entire estimated model to work for some extra purpose—answering counterfactual questions
  - E.g. What was the causal impact of division/reunification on total population of West Berlin? (This sounds not counterfactual, since this event did happen. But of course what happened in the real world was not "division, holding all else constant" but "division, with lots of other things changing too". So the question posed is indeed counterfactual.)
  - E.g. What would be the impact on the 2006 economy of eliminating commuting via automobile (one of the modes of transport in the model)?
  - Formally, all of these can be thought of as just some other element of θ in my $\theta = g(y, x)$

- General idea can be found in lots of other recent papers, e.g.:
  - Allen, Arkolakis and Li (2015) on Chicago
  - Tsivanidis (2017) on Bogota
  - Heblich, Redding and Sturm (QJE 2020) on historical London

Appendix Slides: How Much Should We Trust QSM Estimates?

Some Questions

Some Questions

1. What was the impact of the 1961 division of Berlin on the welfare of West Berlin landowners?
   - (Ahlfeldt, Redding, Sturm and Wolf, 2015)

Some Questions

1. What was the impact of the 1961 division of Berlin on the welfare of West Berlin landowners?
   - (Ahlfeldt, Redding, Sturm and Wolf, 2015)

2. What was the impact of the 1950-2010 rise in demand for tourism in Mexico on the welfare of Mexican residents?
   - (Faber and Gaubert, 2018)

Some Questions

1. What was the impact of the 1961 division of Berlin on the welfare of West Berlin landowners?
    - (Ahlfeldt, Redding, Sturm and Wolf, 2015)
2. What was the impact of the 1950-2010 rise in demand for tourism in Mexico on the welfare of Mexican residents?
    - (Faber and Gaubert, 2018)
3. What was the impact of the 2018-19 US-China trade war on the welfare of US residents?
    - (Fajgelbaum, Goldberg, Khandelwal and Kennedy, 2020)

Some Questions

1. What was the impact of the 1961 division of Berlin on the welfare of West Berlin landowners?
    - (Ahlfeldt, Redding, Sturm and Wolf, 2015)
2. What was the impact of the 1950-2010 rise in demand for tourism in Mexico on the welfare of Mexican residents?
    - (Faber and Gaubert, 2018)
3. What was the impact of the 2018-19 US-China trade war on the welfare of US residents?
    - (Fajgelbaum, Goldberg, Khandelwal and Kennedy, 2020)

- For each of these studies, imagine:
    - Normalizing the researchers' answer to "1"
    - What is the probability distribution of your belief about the true answer?

Some Questions

1. What was the impact of the 1961 division of Berlin on the welfare of West Berlin landowners?
    - (Ahlfeldt, Redding, Sturm and Wolf, 2015)

2. What was the impact of the 1950-2010 rise in demand for tourism in Mexico on the welfare of Mexican residents?
    - (Faber and Gaubert, 2018)

3. What was the impact of the 2018-19 US-China trade war on the welfare of US residents?
    - (Fajgelbaum, Goldberg, Khandelwal and Kennedy, 2020)

- For each of these studies, imagine:
    - Normalizing the researchers' answer to "1"
    - What is the probability distribution of your belief about the true answer?
    - Y/N: does the true answer lie within $(0.5, 1.5)$ with probability $> \frac{2}{3}$?

How do researchers answer questions like these?

How do researchers answer questions like these?

- Common scenario: seems that no purely data-driven answer exists

How do researchers answer questions like these?

- Common scenario: seems that no purely data-driven answer exists
- Recent rise in Spatial Economics in the use of *structural econometric models*
  - (“quantitative models” is perhaps a more accurate name for cases where relatively little focus on estimation)

How do researchers answer questions like these?

- Common scenario: seems that no purely data-driven answer exists
- Recent rise in Spatial Economics in the use of *structural econometric models*
    - ("quantitative models" is perhaps a more accurate name for cases where relatively little focus on estimation)
- In this talk I hope to stimulate a discussion about:
    - How useful such methods are for achieving their stated goals
    - And what might be done to improve the credibility of the answers provided

How do researchers answer questions like these?

- Common scenario: seems that no purely data-driven answer exists
- Recent rise in Spatial Economics in the use of *structural econometric models*
    - ("quantitative models" is perhaps a more accurate name for cases where relatively little focus on estimation)
- In this talk I hope to stimulate a discussion about:
    - How useful such methods are for achieving their stated goals
    - And what might be done to improve the credibility of the answers provided
- Background reading:
    - Donaldson (JEP 2022) is a gentle introduction to many of the themes discussed here
    - FN #1 in that paper contains a long list of broader methodological papers about structural models in Economics that I have found useful
    - Adao, Costinot and Donaldson (QJE 2025) "Putting Quantitative Models to the Test" is newer material that I will emphasize today

Why might we not trust QSMs?

- The world is a complicated place:
  - Even firms have no idea what their production functions are. (And consumers' utility functions?!)
  - What is a good? What is a market? Who is the Walrasian auctioneer?
- Models are always abstractions. That's the whole point of them! ("All models are wrong. But some are [hopefully] useful." - G. Box)
- But why would we think that the abstractions chosen are useful ones?
  - Substantive observations:
    - Even when we do allow for market failures, basically infinite set of different types of strategic interactions, asymmetric info, etc. we could choose.
    - Even symmetric uncertainty seems very hard for modeler to know extent of
  - Sociological observations:
    - Consider how much basic modeling choices differ across fields of Econ.
    - Why do all Econ models seem to have about 2-10 non-trivial parameters, no matter the scope of abstraction (e.g. super micro or super macro)?
  - Of course there is Friedman's (1966) argument about billiard players, but that rested on an *empirical* claim.

Why would we trust QSMs?

- Seems to me like the answer would come from one or both of:

1. Introspection:
   - But see comments on previous slide
   - My personal suspicion is that we choose the functional forms of structural models largely for their tractability, because we follow norms/precedent, and/or our own lack of imagination. E.g. CES is simple (elasticity both common and constant), has one parameter, is symmetric, satisfies IIA, etc.
   - Tractability is great for getting quickly to *some* estimate, but why would it help with delivering a trustworthy one?

2. Evidence
   - But what evidence really tells us that the models are trustworthy?
   - And surely such evidence would be based on the end use of the model: evaluating Box's "is the model useful?"
   - Who does this? [For a laugh, see Section 6 of Ed Leamer's chapter in the 2007 *Handbook of Econometrics* on "Linking the Theory with the Data"]

A running example: FGKK (2020)

- From abstract: *"After decades of supporting free trade, in 2018 the U.S. raised import tariffs and major trade partners retaliated. [...] the aggregate real income loss was $7.2 billion, or 0.04% of GDP"*
- Let's use this example to think about how researchers might arrive at an answer like that

A General Economic Model

- Setup/notation follows Adao, Costinot, Donaldson and Sturm (2023)

A General Economic Model

- Setup/notation follows Adao, Costinot, Donaldson and Sturm (2023)
- **Domestic technology:** firm f produces net output $\tilde{y}(f) \in \Upsilon(f)$
- **Domestic preferences:** individual n has utility $u(n) = u(c(n); n)$
- **Domestic ownership:** individual n owns share $\phi(f, n)$ of firm f (and endowments of factors are just simple "firms")
- **Domestic competition:** high-level conditions s.t. any change in the envt. results in (even at the firm's max) $d\pi(f) = \tilde{y}(f) \cdot dp + p \cdot d\tilde{y}(f)$
- **Domestic taxes and transfers:**
  - Specific trade taxes on different goods g: $p_g = p_g^w + t_g$
  - Uniform lump-sum transfer: τ
  - Production/sales taxes t^y drive wedge between dom. cons. prices (p) and producer prices (q)
  - Income tax schedule $T(\pi \cdot \phi(n); n)$; marginal rate $t(n)$
- **Domestic externalities:** let $\Upsilon(f)$, $u(n)$, $\Omega(\cdot)$ all depend on arbitrary externalities z (which can itself depend arbitrarily on allocation and prices)

Counterfactuals

- Consider any change caused by a small change in trade taxes (or foreign shocks)
  - but similar expressions easy to derive if domestic tech/prefs/other taxes change
- If consumers and firms are optimizing, endog. changes must satisfy:

$$\begin{aligned}\sum_n \nu(n) du(n) &= \underbrace{\beta \cdot d(\omega - \bar{\omega})}_{\text{Dom. redistribution}} + \underbrace{(t \cdot dm)}_{\text{Fiscal ext.: trade taxes}} - \underbrace{(m \cdot dp^w)}_{\text{Redistribn. from abroad}} \\ &+ \underbrace{(t^y \cdot d\tilde{y}^{\text{total}})}_{\text{Fiscal ext.: other taxes}} + \underbrace{\sum_n \sum_f \beta(n)\phi(f,n)(p \cdot dy(f))}_{\text{``Markup/down''} \times \Delta\text{allocation}} \\ &+ \underbrace{(\sum_n \sum_f \beta(n)(\phi(f,n)\pi_z(f) - e_z(n)) \cdot dz}_{\text{Effect on un-internalized externalities}}\end{aligned}$$

- Where $\nu(n)$ is arbitrary set of (marginal) "SWF" weights, $\mu(n)$ is MU of income, and $\beta(n) \equiv \frac{\mu(n)\nu(n)}{\sum_n' \mu(n')\nu(n')}$; and as usual the ET implies that $d\omega(n) \equiv (1 - t(n))(\sum_f \phi(f,n) d\pi(f)) - c(n) \cdot dp$
- If not optimizing, add additional wedges (FOC gaps) for "internalities"

Counterfactuals

- Previous expression contains a ton of insight about the forces that must be driving any changes in essentially any economy.
- In other words: a key part of understanding the counterfactuals that come from structural models relies on mapping the answers they spit out to the terms above
- And if we are to trust structural models we need assurances that they get these terms right

Establishing Trust

- How could a researcher establish trust that their counterfactual prediction about $\sum_n \nu(n) du(n)$ is accurate?
- Each term of $\sum_n \nu(n) du(n)$ takes form ("j" could index goods, firms, or people):

$$\text{"Effect"} = \sum_j \omega_j dy_j \equiv N \times E_\omega[dy]$$

- Where:
  - ω_j is a "weight" (but NB: $\omega_j < 0$ and $\sum_j \omega_j \neq 1$ are possible) that is a "level" characteristic of the pre-shock economy (i.e. not something to do with the counterfactual, and not a causal "response" of any sort)
  - dy_j is the causal response in the endogenous outcome y_j induced by our shock of interest (probably $\neq$ the observed change dy in the data, since other shocks will likely have happened too)
- We should trust structural models iff they get $N \times E_\omega[dy]$ right. Which means that they should correctly measure/model each of:
  1. The weights ω_j
  2. The average causal response $E_\omega[dy]$ (easier than every single causal response)

Establishing Trust: In the Weights ω_j (Part I)

- Sometimes ω_j is just the **initial level of some outcome** (e.g. the m in the term $m \cdot dp^w$).
  - So these should be "easy" to measure in the raw data.
  - Intuition here is related to why DEK/"exact-hat algebra" works (and indeed works for essentially any model not just CES/gravity/etc—see Adao, Costinot and Donaldson (2017).)
  - Indeed, sometimes extra easy due to convenient aggregation:
    - May have a prior that the shock vector dy will only move in certain aggregate ways, so that only the corresponding *aggregate* of weights needs to be known.
    - E.g. in competitive models, for many shocks dp^w only moves because countries' factor prices move. So m that matters is the (very low-dim.) factor content of trade, not (very high-dim.) trade. See Adao, Costinot and Donaldson (2017).
  - But when data not available, a role that some structural models can play is to effectively make up the raw data on the initial levels.
    - So premium on basic measurement should be high.
    - That said, Dingel and Tintelnot (2023) provide warnings about "over-fitting" like problems when the granularity of data required for ω_j is high

Establishing Trust: In the Weights ω_j (Part II)

- Sometimes, ω_j is simply the **value of a tax in the pre-period**.
  - Again, in principle "easy" to measure (tax rates = legislation).
  - In practice, can worry about tax avoidance and evasion. But see Feldstein (1999) for when that does and does not matter

- Sometimes ω_j is a **non-tax wedge** (e.g. a markup or an externality wedge) in the pre-period.
  - Now the "basic measurement" is not at all easy
  - But in principle it could be done in a relatively "model-free" way (i.e. far less model-dependent than the prediction of *dy* responses will be).
  - E.g. think of the Hall/de Loecker-Warzynski approach to markup estimation. (Or common approaches to agglomeration externality estimation.)

Establishing Trust: In the Weights ω_j (Part III)

- Sometimes, ω_j involves **SWF weights** $\nu(n)$.
  - Surely very hard to know what to do with these
  - But also reminds us that the choices we make here are unavoidably important.
  - All too common in applied work to see $\nu(n) = 1$ (and hence $\beta(n) \propto \mu(n)$) used without commentary.
  - Indeed, some choices of SWF can lead to paradoxical-seeming results (e.g. a strict technology improvement in a closed and undistorted economy actually lowers researcher's notion of aggregate welfare) until one realizes what is going on.
- Bottom line: know your weights!
  - Know your distortions (why/where departing from First Welfare Theorem?)
  - Know your SWF (when have multiple agents...which usually is the case in any spatial model with imperfectly mobile factors).

Establishing Trust: In the Causal Responses $E_\omega[dy]$ Part I

- Sometimes (and for some "Effect"s above) we do have special settings in which $E_\omega[dy]$ *is* identified from some quasi-experimental variation (and data on the weights ω).
- This is the heart of the Chetty (2009) exposition of the "sufficient statistics" approach.
  - Also see Kleven (2022) for a very nice critical review. But to me Kleven's critiques seem minor relative to my distrust of the *much* stronger assumptions made in QSMs.
- Aside: if measuring $E_\omega[dy]$ in situations where wedges are changing in complicated but known ways (e.g tax contexts, or even not changing at all), then trick in Lee et al (2021) can be useful/simplifying.

Establishing Trust: In the Causal Responses $E_\omega[dy]$ Part II

- But at a more general level, the whole point of using a structural model is that we are interested in things for which purely data-driven approaches, based on quasi-experimental variation, don't apply.
- That is, literally, the counterfactual embodied in dy (the change in y that occurs only because of our policy change of interest) is truly counterfactual, and that counterfactual can't be observed (even in expectation) nonparametrically in the data.
- Examples of how this can occur:
  - Ex-ante: the policy of interest has just never happened
  - Out of support: a similar policy happened, but we want to consider enlarging it
  - No quasi-experimental variation in the policy (but it did happen)
  - Not "enough" quasi-experimental variation in the policy: e.g. SUTVA violations (spillovers across Treatment and Control units)

Establishing Trust: In the Causal Responses $E_\omega[dy]$ Part III

- Clearly, when faced with a given question, we can only do the best we can with the plausibly valid quasi-experimental variation we can find.
- But thinking about what is inside each $E_\omega[dy]$ is, to me, by far the most useful way of thinking about: "How can I best use the quasi-experimental variation available to me so as to identify 'as much' of $E_\omega[dy]$ as possible before having to rely on a structural model for the rest?"
- How did FGKK do this?

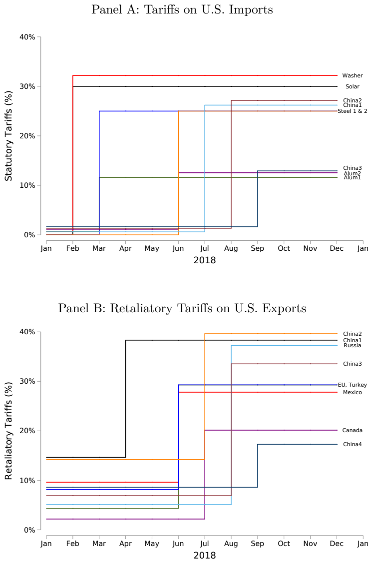
Fajgelbaum et al (QJE 2020)

- Fajgelbaum et al. look at impacts of the US-China "trade war" tariff changes of 2017-18.
- In particular, they start with event studies (for various outcomes "y"):

$$\begin{aligned} \Delta \ln y_{igt} &= \eta_{ig} + \eta_{gt} + \eta_{it}^{m} + \sum_{j=-6,6} \beta_{0j} \mathbf{I}(event_{igt} = j) \\ &+ \sum_{j=-6,6} \beta_{1j} \mathbf{I}(event_{igt} = j) \times target_{ig} + \varepsilon_{igt} \end{aligned}$$

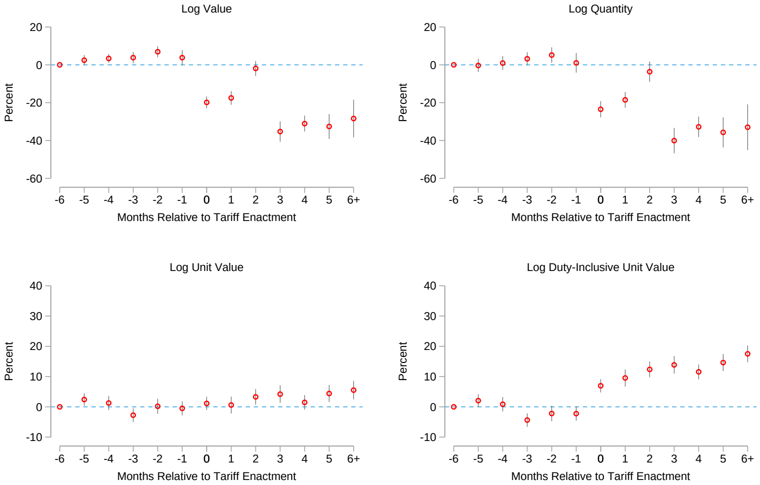
- Where i is the foreign country, g is the product, t is time (month), and $target_{ig}$ denotes products that were targeted (for tariff changes during the trade war) by the country i.

Fajgelbaum et al (QJE 2020)



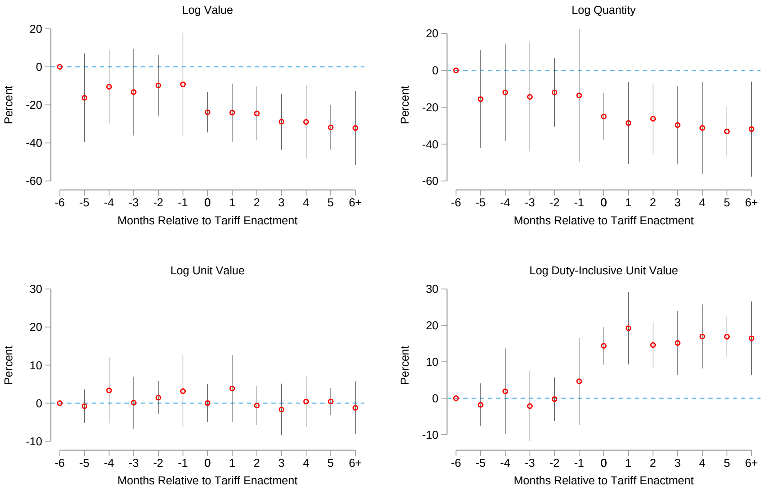
Fajgelbaum et al (QJE 2020)

Figure II: Variety Event Study: Imports



Fajgelbaum et al (QJE 2020)

Figure III: Variety Event Study: Exports



What do these responses imply for US agents' welfare?

- Recall that in general the answer is:

$$\sum_n \nu(n) du(n) = \underbrace{\beta \cdot d(\omega - \bar{\omega})}_{\text{Dom. redistribution}} + \underbrace{(t \cdot dm)}_{\text{Fiscal ext.: trade taxes}} - \underbrace{(m \cdot dp^w)}_{\text{Redistribn. from abroad}}$$

$$+ \underbrace{(t^y \cdot d\tilde{y}^{\text{total}})}_{\text{Fiscal ext.: other taxes}} + \underbrace{\sum_n \sum_f \beta(n)\phi(f,n)(p \cdot dy(f))}_{\text{"Markup/down"} \times \Delta\text{allocation}}$$

$$+ \underbrace{(\sum_n \sum_f \beta(n)(\phi(f,n)\pi_z(f) - e_z(n)) \cdot dz}_{\text{Effect on un-internalized externalities}}$$

- How could we possibly go from the previous diff-in-diff regressions to the "Effects" in this equation?

- Only game in town: write down a full GE model that makes assumptions needed to *extrapolate* from the diff-in-diffs to the aggregate effect $\sum_n \nu(n) du(n)$

FGKK (2020): Domestic Households

- Household in region r and sector s has endowment of $L_{rs,t}$ units of labor
- All households have common nested CES preferences:

$$U_t = (C_{NT,t})^{\beta_{NT,t}}(C_{T,t})^{\beta_{T,t}}$$

$$C_{T,t} = \prod_{s \in \mathcal{S}} (C_{Ts,t})^{\beta_{s,t}}, \quad C_{Ts,t} = \left[(A_{Ds,t})^{\frac{1}{\kappa}} (D_{s,t})^{\frac{\kappa-1}{\kappa}} + (A_{Ms,t})^{\frac{1}{\kappa}} (M_{s,t})^{\frac{\kappa-1}{\kappa}} \right]^{\frac{\kappa}{\kappa-1}}$$

$$D_{s,t} = \left[\sum_{g \in \mathcal{G}_s} (a_{Dg,t})^{\frac{1}{\eta}} (d_{g,t})^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}, \qquad M_{s,t} = \left[\sum_{g \in \mathcal{G}_s} (a_{Mg,t})^{\frac{1}{\eta}} (m_{g,t})^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}$$

$$m_{g,t} = \left[\sum_{i \in F} (a_{ig,t})^{\frac{1}{\sigma}} (m_{ig,t})^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

FGKK (2020): Domestic Firms

- Competitive firms in each region r and sector s take good and factor prices as given

- Nested CES technologies:

$$Q_{NTr,t} = Z_{NTr,t} L_{NTr,t}$$

$$Q_{sr,t} = Z_{sr,t} \left(I_{sr,t}\right)^{\alpha_{Is,t}} \left(L_{sr,t}\right)^{\alpha_{Ls,t}}, \qquad \alpha_{Is,t} + \alpha_{Ls,t} < 1$$

$$I_{sr,t} = \prod_{k \in \mathcal{S}} \left(I_{ksr,t}\right)^{\alpha_{ks,t}}, \qquad \sum_{k \in \mathcal{S}} \alpha_{ks,t} = 1$$

$$\sum_{g \in \mathcal{G}_s} \frac{q_{gs,t}}{z_{gs,t}} = \sum_{r} Q_{sr,t}$$

FGKK (2020): Foreign Import Demand and Export Supply

- Given export price $p^{x}_{ik,t}$, exports (given by foreign import demand):

$$x_{ig,t} = a^F_{ig,t}\left((1+\tau^F_{ig,t})p^X_{ig,t}\right)^{-\sigma_F}$$

- Given (pre-tariff) import price $p^F_{ig,t}$, foreign export supply is

$$m_{ig,t} = (p^*_{ig,t})^{\frac{1}{\omega_F}}(z^F_{ig,t})^{\frac{1}{\omega_F}}$$

- Government at Home imposes import tariffs so that import price is

$$p_{ig,t} = (1+\tau_{ig,t})p^*_{ig,t}$$

- Government uses a lump-sum transfer T_t to rebate tariff revenue and foreign transfer D_t

How does the model help answer the question?

- Recall that in general the answer is:

$$\sum_n \nu(n) du(n) = \underbrace{\beta \cdot d(\omega - \bar{\omega})}_{\text{Dom. redistribution}} + \underbrace{(t \cdot dm)}_{\text{Fiscal ext.: trade taxes}} - \underbrace{(m \cdot dp^w)}_{\text{Redistribn. from abroad}}$$

$$+ \underbrace{(t^y \cdot d\tilde{y}^{\text{total}})}_{\text{Fiscal ext.: other taxes}} + \underbrace{\sum_n \sum_f \beta(n)\phi(f,n)(p \cdot dy(f))}_{\text{“Markup/down”} \times \Delta\text{allocation}}$$

$$+ \underbrace{(\sum_n \sum_f \beta(n)(\phi(f,n)\pi_z(f) - e_z(n)) \cdot dz}_{\text{Effect on un-internalized externalities}}$$

- How did the model assumptions simplify this stuff?

How does the model help answer the question?

- Recall that in general the answer is:

$$\sum_n \nu(n) du(n) = \underbrace{\beta \cdot d(\omega - \bar{\omega})}_{\text{Dom. redistribution}} + \underbrace{(t \cdot dm)}_{\text{Fiscal ext.: trade taxes}} - \underbrace{(m \cdot dp^w)}_{\text{Redistribn. from abroad}}$$

$$+ \underbrace{(t^y \cdot d\tilde{y}^{\text{total}})}_{\text{Fiscal ext.: other taxes}} + \underbrace{\sum_n \sum_f \beta(n) \phi(f,n)(p \cdot dy(f))}_{\text{“Markup/down”} \times \Delta\text{allocation}}$$

$$+ \underbrace{(\sum_n \sum_f \beta(n)(\phi(f,n)\pi_z(f) - e_z(n)) \cdot dz}_{\text{Effect on un-internalized externalities}}$$

- How did the model assumptions simplify this stuff?
- We are left with

$$\frac{1}{\mu} du = (t \cdot dm) - (m \cdot dp^w)$$

- Simpler. But how do we estimate objects like $t \cdot dm$ and $m \cdot dp^w$?

Back to the diff-in-diff regressions

- In this model, the US import-demand and export-supply functions can be expressed as:

$$\begin{aligned}\Delta \ln m_{igt} &= \eta^m_{gt} + \eta^m_{it} + \eta^m_{is} - \sigma \Delta \ln p_{igt} + \varepsilon^m_{igt} \\ \Delta \ln p^F_{igt} &= \eta^{p^F}_{gt} + \eta^{p^F}_{it} + \eta^{p^F}_{is} + \omega_F \Delta \ln m_{igt} + \varepsilon^{p^F}_{igt}\end{aligned}$$

- This is suggestive of the earlier diff-in-diff regressions we saw

- In particular, if we think of (US and foreign retaliatory) tariff changes as potential IVs in these structural equations, the previous regressions are formally the corresponding reduced-form equations

- Identification: Tariffs create wedge between what importer pays and exporter receives. Shifts down the D curve for any given price received by the exporter, tracing S curve. Shifts up supply curve for any given price paid by the consumer, tracing D curve. (If uncorrelated with other S/D shifters.)

Estimates of σ and ω_F from US tariff changes

(NB: I'm calling ω_F what they call ω^*. Similarly for p^*.)

Table IV: Variety Import Demand (σ) and Foreign Export Supply (ω^*)

	(1) $\Delta \ln p^*_{igt} m_{igt}$	(2) $\Delta \ln m_{igt}$	(3) $\Delta \ln p^*_{igt}$	(4) $\Delta \ln p_{igt}$	(5) $\Delta \ln p^*_{igt}$	(6) $\Delta \ln m_{igt}$
$\Delta \ln(1+\tau_{igt})$	-1.52***	-1.47***	0.00	0.58***		
	(0.18)	(0.24)	(0.08)	(0.13)		
$\Delta \ln m_{igt}$					-0.00	
					(0.05)	
$\Delta \ln p_{igt}$						-2.53***
						(0.26)
Product × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Country × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Country × Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
1st-Stage F					36.5	21.2
Bootstrap CI					[-0.14,0.10]	[1.75,3.02]
R2	0.13	0.13	0.11	0.11	0.00	.
N	2,993,288	2,454,023	2,454,023	2,454,023	2,454,023	2,454,023

So $\hat{\sigma} =??$ and $\hat{\omega}_F =??$

Back to the diff-in-diff regressions (again)

- Symmetrically (due to the way the model was set up!) the foreign import-demand function can be expressed as:

$$\Delta \ln x_{igt} = \eta^x_{gt} + \eta^x_{it} + \eta^x_{is} - \sigma_F \Delta \ln((1 + \tau^F_{igt})p^F_{igt}) + \varepsilon^x_{igt}$$

- FGKK also estimate an analogous US (product-level) export-supply function, but this is not a structural equation in the model (so ω is not a structural parameter):

$$\Delta \ln p^X_{igt} = \eta^p_{gt} + \eta^p_{it} + \eta^p_{is} + \omega \Delta \ln x_{igt} + \varepsilon^p_{igt}$$

- So again these are related to the earlier diff-in-diff regressions we saw (and the elasticities are identified via a similar intuition as before)

Estimates of σ_F and ω from retaliatory tariff changes

(NB: I'm calling τ^F what they call τ^*.)

	(1) $\Delta \ln p^X_{igt} x_{igt}$	(2) $\Delta \ln x_{igt}$	(3) $\Delta \ln p^X_{igt}$	(4) $\Delta \ln p^X_{igt}(1+\tau^*_{igt})$	(5) $\Delta \ln p^X_{igt}$	(6) $\Delta \ln x_{igt}$
$\Delta \ln(1+\tau^*_{igt})$	-0.99***	-1.00***	-0.04	0.96***		
	(0.28)	(0.36)	(0.16)	(0.16)		
$\Delta \ln x_{igt}$					0.04	
					(0.16)	
$\Delta \ln p^X_{igt}(1+\tau^*_{igt})$						-1.04***
						(0.32)
Product × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Country × Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Country × Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
1st-Stage F					7.8	38.2
Bootstrap CI					[-0.30,0.26]	[0.73,1.39]
R2	0.07	0.07	0.06	0.06	.	0.51
N	3,306,766	2,564,731	2,564,731	2,564,731	2,564,731	2,564,731

So $\hat{\sigma}_F =$?? and $\hat{\omega} =$??

What about the remaining elasticities κ and η?

- Think about η. This is the elasticity across products g for US demand (whereas σ was the within-product, cross-origin elasticity for US demand).
- Standard exercise in nested-CES system algebra to show that:

$$\Delta \ln(\text{Expenditure})_{Mgt} = FE_{st} + (1-\eta)\Delta \ln(p_{Mgt}) + \varepsilon_{Mgt}$$

where $(\text{Expenditure})_{Mgt} \equiv p_{Mgt} m_{gt} \equiv \sum_{i \in F} p_{igt} m_{igt}$, and p_{Mgt} is the exact (Feenstra, 1994) price index for the product g "bundle"

$$\Delta \ln p_{Mgt} \equiv \frac{1}{1-\sigma} \ln \left(\sum_{i \in C_{gt}} s_{igt} \exp\left((1-\sigma)\Delta \ln(p^F_{igt}(1+\tau_{igt})) + a_{igt} \right) \right) - \frac{1}{1-\sigma} \ln \left(\frac{S_{g,t+1}(C_{gt})}{S_{g,t}(C_{gt})} \right)$$

and where s_{igt} is the share of continuing variety i in all continuing varieties (between t and $t+1$) and $S_{g,t}(C)$ is the share of the varieties in the set C among the total imports of product g at time t.

- And can estimate all a_{igt} from the residuals of the σ regressions

What about the remaining elasticities κ and η?

- But how to identify η? Simultaneity bias just as severe, potentially, at any level of aggregation.
- Natural to use IVs based on weighted aggregations of the same IVs as we used in the earlier within-product nest (i.e. tariffs). (I first "saw" this in Costinot, Donaldson and Smith, 2016 but it is surely an old idea.)
- If we believed in exogeneity of tariffs before, we probably still believe in (aggregated versions of) that same exogenity now.
- In this spirit, FGKK use:

$$\Delta \ln Z_{Mgt} = \ln \left(\frac{1}{N_{gt}^C} \sum_{i \in C_{gt}} \exp \left(\Delta \ln (1 + \tau_{igt}) \right) \right)$$

  where N_{gt}^C is the number of continuing varieties in g between t and $t+1$
- So now this is just another diff-in-diff IV regression but on product-level aggregates
- And then estimating κ proceeds analogously but with *sector-level* aggregates, with price indices and IV built from estimates from η regression.

Estimates of η from aggregated tariff changes

(NB: What FGKK call s_{Mgt} is—for the purposes of this regression, given the included FE_{st}—the same as what I'm calling $(\text{Expenditure})_{Mgt}$.)

TABLE V

PRODUCT ELASTICITY η

	$\Delta \ln s_{Mgt}$ (1)	$\Delta \ln p_{Mgt}$ (2)	$\Delta \ln s_{Mgt}$ (3)
$\Delta \ln Z_{Mgt}$	-0.81^{**}	1.52^{***}	
	(0.39)	(0.40)	
$\Delta \ln p_{Mgt}$			-0.53^{*}
			(0.27)
Sector-time FE	Yes	Yes	Yes
1st-stage F			14.6
$\hat{\eta}$ (se[$\hat{\eta}$])			1.53 (0.27)
Bootstrap CI			[1.15, 1.89]
R^2	0.01	0.10	—
N	371,916	371,916	371,916

So $\hat{\eta} =??$

Estimates of κ from doubly-aggregated tariff changes

(NB: What FGKK call P_{Mst} is the sector-level Feenstra price index—analog of p_{Mgt} we saw above.)

TABLE VI

SECTOR ELASTICITY κ

	$\Delta \ln(\frac{P_{Mst} M_{st}}{P_{Dst} D_{st}})$ (1)	$\Delta \ln\left(\frac{P_{Mst}}{p_{st}}\right)$ (2)	$\Delta \ln(\frac{P_{Mst} M_{st}}{P_{Dst} D_{st}})$ (3)
$\Delta \ln Z_{Mst}$	0.30	−1.59	
	(0.36)	(3.49)	
$\Delta \ln\left(\frac{P_{Mst}}{p_{st}}\right)$			−0.19
			(0.49)
Sector FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
1st-stage F			0.2
$\hat{\kappa}$ (se[$\hat{\kappa}$])			1.19 (0.49)
Bootstrap CI			[0.89, 1.71]
R^2	0.24	0.67	—
N	2,041	2,041	2,041

So $\hat{\kappa} = ??$

FGKK (2020): Putting it together

- How do we go from estimates of $\theta \equiv (\kappa, \eta, \sigma_F, \sigma, \omega_F)$ to estimates of

$$\sum_n \nu(n) du(n) = t \cdot dm - m \cdot dp^w?$$

- FGKK have isolated all elasticities (beyond those in the Cobb-Douglas and perfect-transformation assumptions in the GE model) down to just these 5.

- But the responses that matter (the vectors dm and dp^w) are still enormously high-dimensional (hundreds of thousands of elements each)

- And even in FGKK's assumed model structure it is not the case that dm or dp^w take a simple or near-homogeneous structure—the structural elasticities θ are low-dimensional, but the reduced-form responses dm and dp^w are not.

- But by populating the remaining unknown parameters (i.e. Cobb-Douglas shares) in the GE model with observed data shares (eg from IO tables, labor use table, trade data) in the 2018 baseline, completes the model. And then solve the nonlinear system of equations to solve for dm or dp^w and hence dW.

FGKK (2020): Estimates of $EV \equiv \frac{1}{\mu} du$

TABLE VIII

AGGREGATE IMPACTS

	EV^M (1)	EV^X (2)	ΔR (3)	EV (4)
2018 trade war				
Change ($ b)	−51.0	9.4	34.3	−7.2
	[−54.8,−47.2]	[4.1,15.6]	[32.3,36.1]	[−14.4,0.8]
Change (% GDP)	−0.27	0.05	0.18	−0.04
	[−0.29,−0.25]	[0.02,0.08]	[0.17,0.19]	[−0.08,0.00]
2018 U.S. tariffs and no retaliation				
Change ($ b)	−50.9	16.6	34.8	0.5
	[−52.9,−49.0]	[13.2,20.3]	[32.8,36.5]	[−4.0,5.7]
Change (% GDP)	−0.27	0.09	0.19	0.00
	[−0.28,−0.26]	[0.07,0.11]	[0.18,0.20]	[−0.02,0.03]

NB: FGKK use slightly different conventions from what I've used above. They use:

- $EV^X \equiv x \cdot \Delta p^X = \sum_s \sum_{g \in G_s} \sum_{i \in F} x_{ig} \Delta p^X_{ig}$
- $\Delta R \equiv \Delta(p^F(1+\tau) \cdot m) = \sum_s \sum_{g \in G_s} \sum_{i \in F} \Delta(p^F_{ig}(1+\tau_{ig}) m_{ig})$
- $EV^M \equiv m \cdot \Delta p = \sum_s \sum_{g \in G_s} \sum_{i \in F} m_{ig} \Delta p_{ig}$

Establishing Trust: Beyond Estimation?

- Is there anything further that the modeler can do?

Establishing Trust: Beyond Estimation?

- Is there anything further that the modeler can do?
- Testing!

Establishing Trust: Beyond Estimation?

- Is there anything further that the modeler can do?
- Testing!
- Will talk now about recent work I've been doing with Rodrigo Adao and Arnaud Costinot
- **A new test procedure...**
  - Compare model and data responses of outcomes to exogenous variation
  - But weight outcomes by relevance to the question

FAQs

FAQs

1. **Why test models? (All models are wrong!)**
   - Yes but some are hopefully useful. Design test statistic around the use one has in mind

FAQs

1. **Why test models? (All models are wrong!)**
   - Yes but some are hopefully useful. Design test statistic around the use one has in mind
2. **What's the use of a p-value? (A fancy measure of N?)**
   - Design test statistic (and hence CI) with interpretability under some misspecification

FAQs

1. **Why test models? (All models are wrong!)**
   - Yes but some are hopefully useful. Design test statistic around the use one has in mind
2. **What's the use of a p-value? (A fancy measure of N?)**
   - Design test statistic (and hence CI) with interpretability under some misspecification
3. **What's wrong with an "untargeted moment" test?**
   - Typically unclear how such moments relate to the causal question at hand

FAQs

1. **Why test models? (All models are wrong!)**
   - Yes but some are hopefully useful. Design test statistic around the use one has in mind
2. **What's the use of a p-value? (A fancy measure of N?)**
   - Design test statistic (and hence CI) with interpretability under some misspecification
3. **What's wrong with an "untargeted moment" test?**
   - Typically unclear how such moments relate to the causal question at hand
4. **Isn't it an old idea to compare causal responses in model and data?**
   - Yes. E.g. "Lucas program" of Christiano et al (1999). New here: test with end use in mind

FAQs

1. **Why test models? (All models are wrong!)**
   - Yes but some are hopefully useful. Design test statistic around the use one has in mind
2. **What's the use of a p-value? (A fancy measure of N?)**
   - Design test statistic (and hence CI) with interpretability under some misspecification
3. **What's wrong with an "untargeted moment" test?**
   - Typically unclear how such moments relate to the causal question at hand
4. **Isn't it an old idea to compare causal responses in model and data?**
   - Yes. E.g. "Lucas program" of Christiano et al (1999). New here: test with end use in mind
5. **Where do I get exogenous variation from?**
   - If doing estimation, you already have it! (And can "re-use" it for testing.)

FAQs

1. **Why test models? (All models are wrong!)**
   - Yes but some are hopefully useful. Design test statistic around the use one has in mind
2. **What's the use of a p-value? (A fancy measure of N?)**
   - Design test statistic (and hence CI) with interpretability under some misspecification
3. **What's wrong with an "untargeted moment" test?**
   - Typically unclear how such moments relate to the causal question at hand
4. **Isn't it an old idea to compare causal responses in model and data?**
   - Yes. E.g. "Lucas program" of Christiano et al (1999). New here: test with end use in mind
5. **Where do I get exogenous variation from?**
   - If doing estimation, you already have it! (And can "re-use" it for testing.)
6. **How can I test my model when it (by design) fits the data exactly?**
   - Residuals may perfectly fit data. But can still test whether orthogonal to exog. variation.

Setup

Setup

- **Consider reduced-form of *researcher's model*:**

$$y_{n,t} = f_n(\tau_t, \epsilon_t; \theta)$$

  - $y_{n,t}$: endogenous outcome of interest $n \in \mathcal{N}$
  - τ_t: vector of all "policy" (etc.) variables of interest
  - ϵ_t: vector of all time-varying parameters—"other shocks"
  - $f_n(\cdot)$: mapping implied by market structure, preferences, technology, etc.; suppress θ for now

- **Goal is to answer counterfactual question about causal impact of policy change. WLOG write as:**

$$W(\Delta x) \equiv \sum_n \omega_n \Delta x_n, \quad \text{with } \{\omega_n\}_n \text{ observed}$$

$$\text{where} \quad \Delta x_n \equiv f_n(\tau_{t+1}, \epsilon_{t+1}) - f_n(\tau_t, \epsilon_{t+1})$$

FGKK (2020): arrive at $y_t = f(\tau_t, \epsilon_t; \theta)$

- Time-varying shocks to preferences, technology, and endowments:

$$\epsilon_t \equiv \{\beta_{NT,t}, \beta_{s,t}, A_{Ms,t}, a_{Dg,t}, a_{Mg,t}, a_{ig,t}, Z_{NTr,t}, Z_{sr,t}, \alpha_{Is,t}, \alpha_{Ls,t}, \alpha_{ksr,t}, a^F_{ig,t}, z^F_{ig,t}, D_t, L_{sr,t}\}$$

- Governments' policy vector:

$$\tau_t \equiv \{\tau^H_{ig,t}, \tau^F_{ig,t}\}$$

- UMP + PMP + GMC + LMC + GBC $\Longrightarrow$ **reduced-form** $y_t = f(\tau_t, \epsilon_t; \theta)$

FGKK (2020): Causal Effect of Interest

- Recall, researcher's goal (where $\Delta x_n \equiv f_n(\tau_{t+1}, \epsilon_{t+1}) - f_n(\tau_t, \epsilon_{t+1})$):

$$W(\Delta x) \equiv \sum_n \omega_n \Delta x_n, \quad \text{with } \{\omega_n\}_n \text{ observed}$$

- FGKK: What was impact of "Trump's trade war" on US welfare?

- Proportional change in welfare of US rep. agent due to tariff changes (up to first-order):

$$W(\Delta x) = \sum_{i,g} [\omega_{ig}^X (\Delta x_{ig}^X) - \omega_{ig}^M (\Delta x_{ig}^M) + \omega_{ig}^T (\Delta x_{ig}^T)],$$

  where:
  - $\Delta x_{ig}^X \equiv$ change in log US *export price* of good g in country i (pre-foreign tariff)
  - $\Delta x_{ig}^M \equiv$ change in log US *import price* of good g from country i (post-US tariff)
  - $\Delta x_{ig}^T \equiv$ change in US *tariff revenue* on good g from country i (as share of import spending)
  - $\omega_{ig}^X \equiv$ share of export revenues in 2016 US GDP accounted by country i and good g
  - $\omega_{ig}^M = \omega_{ig}^T \equiv$ share of import spending in 2016 US GDP accounted by country i and good g

Potential Misspecification

- **Data generated by <u>true</u> model:**

$$y_{n,t} = f_n^*(\tau_t, \epsilon_t^*), \qquad \Delta x_n^* \equiv f_n^*(\tau_{t+1}, \epsilon_{t+1}^*) - f_n^*(\tau_t, \epsilon_{t+1}^*)$$

Potential Misspecification

- **Data generated by <u>true</u> model:**

$$y_{n,t} = f_n^*(\tau_t, \epsilon_t^*), \qquad \Delta x_n^* \equiv f_n^*(\tau_{t+1}, \epsilon_{t+1}^*) - f_n^*(\tau_t, \epsilon_{t+1}^*)$$

- **True and researcher's model agree on weights $\{\omega_n\}_n$:**

$$W(\Delta x) \equiv \sum_n \omega_n \Delta x_n \qquad \text{vs.} \qquad W(\Delta x^*) \equiv \sum_n \omega_n \Delta x_n^*$$

Potential Misspecification

- **Data generated by <u>true</u> model:**

$$y_{n,t}=f_n^*(\tau_t,\epsilon_t^*), \qquad \Delta x_n^* \equiv f_n^*(\tau_{t+1},\epsilon_{t+1}^*) - f_n^*(\tau_t,\epsilon_{t+1}^*)$$

- **True and researcher's model agree on weights $\{\omega_n\}_n$:**

$$W(\Delta x) \equiv \sum_n \omega_n \Delta x_n \qquad \text{vs.} \qquad W(\Delta x^*) \equiv \sum_n \omega_n \Delta x_n^*$$

- **That is, agree on goal and how endogenous outcomes map into it:**
  - Could follow from a "sufficient statistics"-like argument
  - If agree on size of economy's distortions, then agree on the Taylor expansion to $W(\cdot)$ in price and quantity changes (and can then define set of outcomes $\mathcal{N}$ to include such changes)

FGKK (2020): Agreeing on weights $\{\omega_n\}_n$

- Recall, FGKK (2020) use

$$W(\Delta x) = \sum_{i,g} [\omega_{ig}^X(\Delta x_{ig}^X) - \omega_{ig}^M(\Delta x_{ig}^M) + \omega_{ig}^T(\Delta x_{ig}^T)]$$

- So true and researcher's models agree that:
  - objective is US rep. agent welfare
  - welfare derives from standard goods (not, e.g., pollution or national security)
  - only distortion is tariff revenue
  - first-order approximation to $W(\Delta x)$ or $W(\Delta x^*)$ is sufficiently accurate

An IV-Based Test Statistic

An IV-Based Test Statistic

- Want to test $W(\Delta x) = W(\Delta x^*)$—not $f = f^*$!

An IV-Based Test Statistic

- Want to test $W(\Delta x) = W(\Delta x^*)$—not $f = f^*$!
- Empirical challenge: don't observe Δx_n^* (obviously)
- But suppose we observe change in outcomes before and after the policy change

$$\Delta y_n = f_n^*(\tau_{t+1}, \epsilon_{t+1}^*) - f_n^*(\tau_t, \epsilon_t^*) = \Delta x_n^* + \Delta\eta_n^*$$

where $\Delta\eta_n^* \equiv f_n^*(\tau_t, \epsilon_{t+1}^*) - f_n^*(\tau_t, \epsilon_t^*)$ denotes the causal impact of the other shocks

An IV-Based Test Statistic

- Want to test $W(\Delta x) = W(\Delta x^*)$—not $f = f^*$!
- Empirical challenge: don't observe Δx_n^* (obviously)
- But suppose we observe change in outcomes before and after the policy change

$$\Delta y_n = f_n^*(\tau_{t+1}, \epsilon_{t+1}^*) - f_n^*(\tau_t, \epsilon_t^*) = \Delta x_n^* + \Delta \eta_n^*$$

where $\Delta \eta_n^* \equiv f_n^*(\tau_t, \epsilon_{t+1}^*) - f_n^*(\tau_t, \epsilon_t^*)$ denotes the causal impact of the other shocks

Definition: IV-based test statistic

Suppose we have some "instrument" z. Then IV-based test statistic *is*

$$\hat{\beta}_z \equiv \frac{1}{N_W} \sum_{n \in \mathcal{N}_W} z_n (\Delta y_n - \Delta x_n)$$

where N_W denotes the number of observations in $\mathcal{N}_W \equiv \{n : \omega_n \neq 0\}$.

An IV-Based Test: Moment Restriction

An IV-Based Test: Moment Restriction

A3 [The causal impact of tariffs in the researcher's model is correct]

For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \Delta x_n$.

An IV-Based Test: Moment Restriction

A3 [The causal impact of tariffs in the researcher's model is correct]

For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \Delta x_n$.

Proposition 1 [Expected value of the goodness of fit measure]

Take any IV z *that satisfies* $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$. *If A3 holds, then* $E_t[\hat{\beta}_z] = 0$.

An IV-Based Test: Moment Restriction

A3 [The causal impact of tariffs in the researcher's model is correct]

For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \Delta x_n$.

Proposition 1 [Expected value of the goodness of fit measure]

Take any IV z that satisfies $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$. *If A3 holds, then* $E_t[\hat{\beta}_z] = 0$.

- **Proof:** Substitute $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$ into definition of $\hat{\beta}_z$ and use identity $\Delta y_n = \Delta x_n^* + \Delta \eta_n^*$. Then A3 implies

$$E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_{n \in \mathcal{N}_W} z_n (\Delta x_n^* - \Delta x_n)] = 0$$

An IV-Based Test: Moment Restriction

A3 [The causal impact of tariffs in the researcher's model is correct]

For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \Delta x_n$.

Proposition 1 [Expected value of the goodness of fit measure]

Take any IV z that satisfies $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$. *If A3 holds, then* $E_t[\hat{\beta}_z] = 0$.

- **Proof:** Substitute $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$ into definition of $\hat{\beta}_z$ and use identity $\Delta y_n = \Delta x_n^* + \Delta \eta_n^*$. Then A3 implies

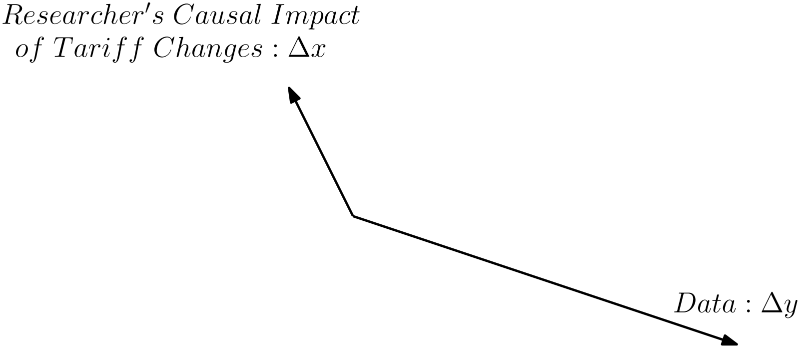
$$E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_{n \in \mathcal{N}_W} z_n (\Delta x_n^* - \Delta x_n)] = 0$$

- **NB:** Given z that satisfies $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$, $E_t[\hat{\beta}_z]$ is a weighted sum of misspecifications, $\Delta x_n^* - \Delta x_n$, along all welfare-relevant variables

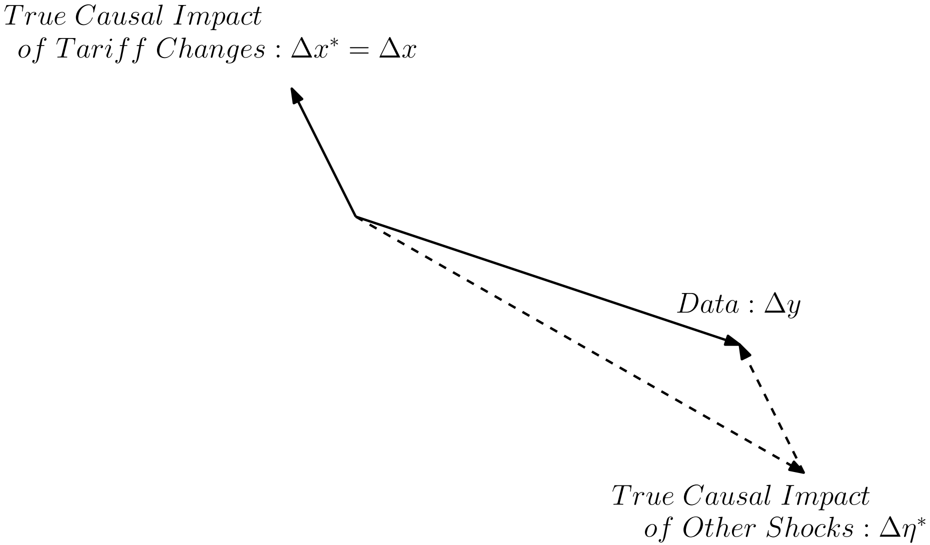
Intuition behind IV-Based Test

$Researcher's\ Causal\ Impact$
$of\ Tariff\ Changes : \Delta x$

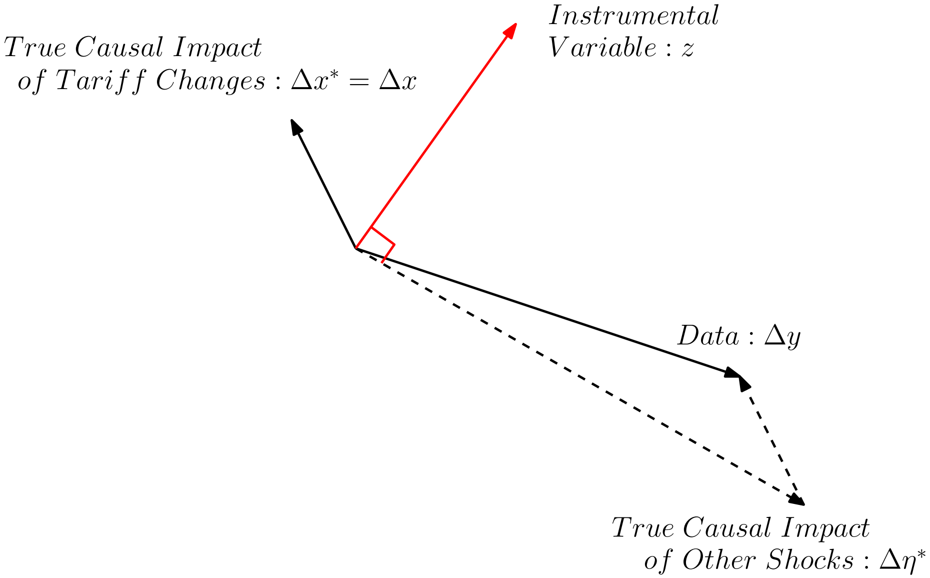
Intuition behind IV-Based Test



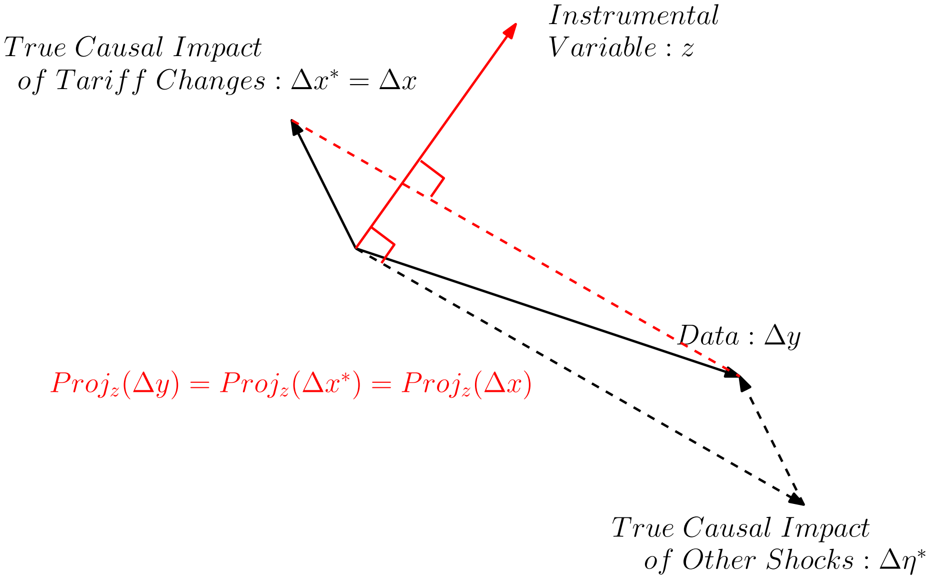
Intuition behind IV-Based Test



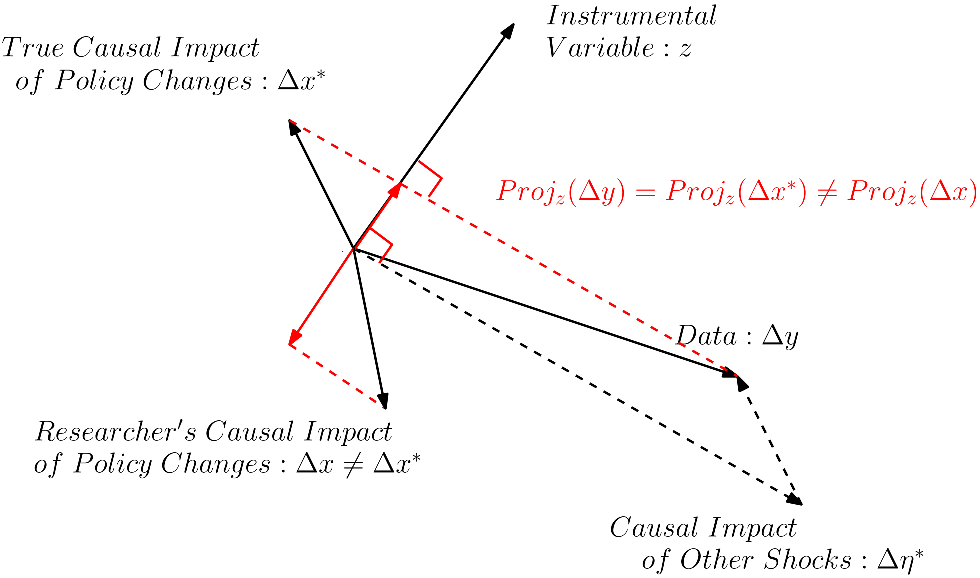
Intuition behind IV-Based Test



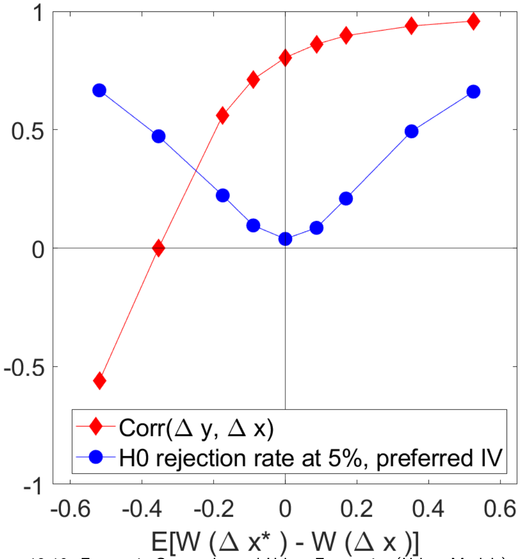
IV-Based Test (when does not reject)



IV-Based Test (when does not reject)



FGKK (2020): Monte Carlo simulation



From Exogenous Policy Shifters to a Candidate IV

- So far, taken as given a z that satisfies $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$
- Empirical literature offers vector of exogenous policy shifters $\Delta \tau_{IV} \equiv \{\Delta \tau_{IV,k}\}_k$:
  - Could just be observed policy change (as in Fajgelbaum, Goldberg, Kennedy, Khandelwal 20)
- We confine attention to instruments z that satisfy:

From Exogenous Policy Shifters to a Candidate IV

- So far, taken as given a z that satisfies $E_t[\sum_{n\in\mathcal{N}_W} z_n \Delta\eta_n^*] = 0$
- Empirical literature offers vector of exogenous policy shifters $\Delta\tau_{IV} \equiv \{\Delta\tau_{IV,k}\}_k$:
  - Could just be observed policy change (as in Fajgelbaum, Goldberg, Kennedy, Khandelwal 20)
- We confine attention to instruments z that satisfy:

A1 [Shift-share structure]

For any $n \in \mathcal{N}_W$, the instrumental variable takes the form $z_n = \sum_k s_{nk}\Delta\tau_{IV,k}$, where the vector of "shares" $\{s_{nk}\}$ may be a function of, and only of, the realization of period t's tariffs and other shocks, (ϵ_t^*, τ_t).

From Exogenous Policy Shifters to a Candidate IV

- So far, taken as given a z that satisfies $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$
- Empirical literature offers vector of exogenous policy shifters $\Delta \tau_{IV} \equiv \{\Delta \tau_{IV,k}\}_k$:
  - Could just be observed policy change (as in Fajgelbaum, Goldberg, Kennedy, Khandelwal 20)
- We confine attention to instruments z that satisfy:

A1 [Shift-share structure]

For any $n \in \mathcal{N}_W$, the instrumental variable takes the form $z_n = \sum_k s_{nk} \Delta \tau_{IV,k}$, where the vector of "shares" $\{s_{nk}\}$ may be a function of, and only of, the realization of period t's tariffs and other shocks, (ϵ_t^*, τ_t).

A2 [Independence of the shifters]

Conditional on the realization of period t's tariffs and other shocks, policy shifters are mean zero and independent of other shocks in period $t+1$: $\Delta \tau_{IV} \perp\!\!\!\perp \epsilon_{t+1}^* | (\epsilon_t^*, \tau_t)$.

- Trivial to show that if z satisfies A1 and A2 then $E_t[\sum_{n \in \mathcal{N}_W} z_n \Delta \eta_n^*] = 0$

Asymptotic Null Distribution of Test Statistic

- How to do inference? Haven't yet taken any stand on distribution of shocks ϵ_{t+1} (and hence of the data Δy under the null

- Can apply "design-based" results on consistency (Borusyak et al., 2022) and inference (Adao et al., 2019) of shift-share IV

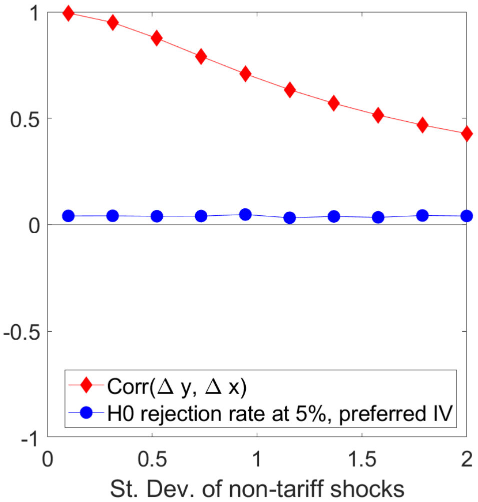
Asymptotic Null Distribution of Test Statistic

- How to do inference? Haven't yet taken any stand on distribution of shocks ϵ_{t+1} (and hence of the data Δy under the null
- Can apply "design-based" results on consistency (Borusyak et al., 2022) and inference (Adao et al., 2019) of shift-share IV

Proposition 2 [Asymptotic behavior of test statistic]

Take IV z that satisfies A1 and A2. If A3 holds and (i) $\Delta\tau_{IV,k}$ *are i.i.d.,* (ii) $\frac{1}{N_W^2}\sum_k (S_k)^2 \to 0$ *with* $S_k \equiv \sum_n |s_{nk}|$, *and* (iii) $Var_t[\Delta\tau_{IV,k}]$ *and* $\Delta\eta_n^*$ *are uniformly bounded, then* $\hat{\beta}_z \to_p 0$.
If, in addition, (iv) $\frac{\max_k(S_{k,t})}{\sum_k S_{k,t}^2} \to 0$; (v) $E_t[(\Delta\tau_{IV,k})^4]$ *is uniformly bounded; and* (vi) $\frac{1}{\sum_k S_k^2} Var_t[\sum_{n\in\mathcal{N}_W} z_n \Delta\eta_n^* | \epsilon_{t+1}^*] \to_p V_\beta > 0$, *then* $r_\beta \hat{\beta}_z \to_d \mathcal{N}(0, V_\beta)$ *with* $r_\beta \equiv N_W / \sqrt{\sum_k S_k^2}$.

FGKK (2020): In Monte Carlo, coverage of test statistic when no misspecification



Extensions

- **Estimation uncertainty:**
  - If f is known up to estimation of structural parameter θ, then can compute asymptotic distribution of $\hat{\beta}_z(\hat{\theta})$ whenever
    - $\hat{\theta}$ is independent of $\hat{\beta}_z(\theta)$ (e.g. when estimation has been conducted on a different sample)
    - $\hat{\theta}$ is an IV estimator, potentially based on the same policy shifters (as in our application)

- **Clustering:**
  - Weaken such that $\Delta\tau_{IV,k}$ is only i.i.d across *groups* of observations

- **Controls:**
  - Weaken A2 such that indep. of $\Delta\tau_{IV}$ holds only after controlling for linear determinants of $\Delta\eta^*$
  - Need to then residualize shares $\{s_{nk}\}$ w.r.t. those controls

Economic Interpretation of Test Statistic $\hat{\beta}_z$

- **Question:** How should we interpret goodness of fit measure? Ideally, we would like it to measure, at least on average, misspecification in the counterfactual of interest, i.e.,

$$E_t[W(\Delta x^*) - W(\Delta x)] = E_t[\sum_n \omega_n (\Delta x_n^* - \Delta x_n)]$$

Economic Interpretation of Test Statistic $\hat{\beta}_z$

- **Question:** How should we interpret goodness of fit measure? Ideally, we would like it to measure, at least on average, misspecification in the counterfactual of interest, i.e.,

$$E_t[W(\Delta x^*) - W(\Delta x)] = E_t[\sum_n \omega_n(\Delta x_n^* - \Delta x_n)]$$

- **Challenge:** $E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_n z_n(\Delta x_n^* - \Delta x_n)] \neq E_t[\sum_n \omega_n(\Delta x_n^* - \Delta x_n)]$ for arbitrary z

Economic Interpretation of Test Statistic $\hat{\beta}_z$

- **Question:** How should we interpret goodness of fit measure? Ideally, we would like it to measure, at least on average, misspecification in the counterfactual of interest, i.e.,

$$E_t[W(\Delta x^*) - W(\Delta x)] = E_t[\sum_n \omega_n(\Delta x_n^* - \Delta x_n)]$$

- **Challenge:** $E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_n z_n(\Delta x_n^* - \Delta x_n)] \neq E_t[\sum_n \omega_n(\Delta x_n^* - \Delta x_n)]$ for arbitrary z

A3' [Misspecification of causal impacts]

For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \alpha_n \Delta x_n$, with the misspecification parameter α_n a function of the realization of period t's tariffs and other shocks, $(\epsilon_t^*, \epsilon_t, \tau_t)$, but not period $t+1$'s tariffs, τ_{t+1}.

Economic Interpretation of Test Statistic $\hat{\beta}_z$

- **Question:** How should we interpret goodness of fit measure? Ideally, we would like it to measure, at least on average, misspecification in the counterfactual of interest, i.e.,

$$E_t[W(\Delta x^*) - W(\Delta x)] = E_t[\sum_n \omega_n(\Delta x_n^* - \Delta x_n)]$$

- **Challenge:** $E_t[\hat{\beta}_z] = \frac{1}{N_W} E_t[\sum_n z_n(\Delta x_n^* - \Delta x_n)] \neq E_t[\sum_n \omega_n(\Delta x_n^* - \Delta x_n)]$ for arbitrary z

A3' [Misspecification of causal impacts]

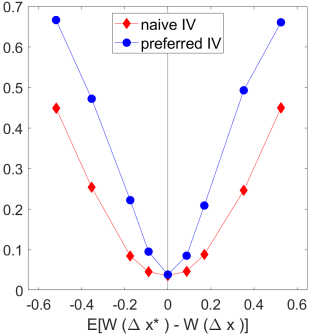
For any $n \in \mathcal{N}_W$, $\Delta x_n^* = \alpha_n \Delta x_n$, with the misspecification parameter α_n a function of the realization of period t's tariffs and other shocks, $(\epsilon_t^*, \epsilon_t, \tau_t)$, but not period $t+1$'s tariffs, τ_{t+1}.

Proposition 3 [IV-based test stat and average welfare misspecification]

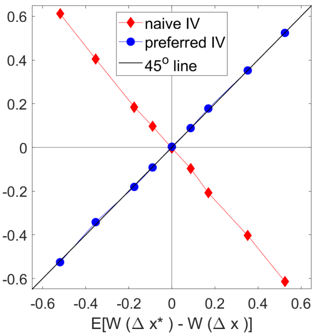
Take IV z that satisfies A1 and A2 and define z', with $z_n' \equiv z_n \omega_n E_t[\Delta x_n]/E_t[z_n \Delta x_n]$ for all $n \in \mathcal{N}_W$. If A3' holds, then $E_t[\hat{\beta}_{z'}] = E_t[W(\Delta x^*) - W(\Delta x)]$.

FGKK (2020): Monte Carlo comparing IV-Based Tests

"Preferred IV" follows method in Proposition 3. "Naive IV" only uses tariff shifters on product of interest.



(a) Rejection rate



(b) Mean of $\hat{\beta}_z$

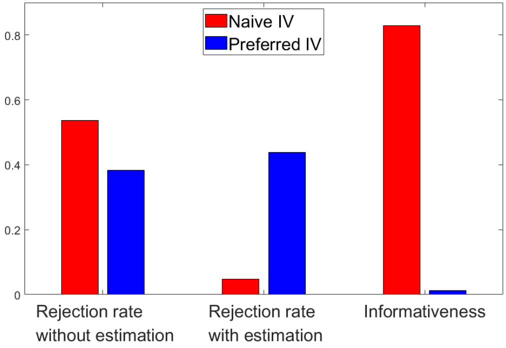
Choosing IVs to Improve Statistical Power

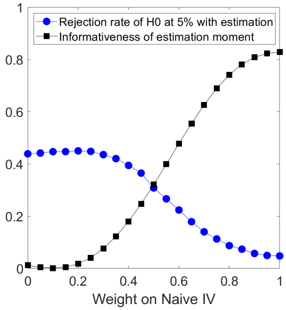
- Three potential reasons for low-power of arbitrary IV-based test:
  1. **Lack of first stage:** $E_t[z_n \Delta x_n] = E_t[z_n \Delta y_n] = 0$ because z is noise
  2. **Mechanical fit:** Estimation moments "mechanically" related to testing moments
  3. **Precision:** Too much variance in $\Delta y_n - \Delta x_n \Rightarrow$ too much variance in $\hat{\beta}_z$

- Three potential solutions:
  1. To address lack of first stage, **use** causal impact of shifters predicted by **researcher's model,** i.e. $s_{nk} = \partial f_n / \partial \tau_k \Rightarrow z_n = \sum_k (\partial f_n / \partial \tau_k) \Delta \tau_{IV,k}$
  2. To address mechanical fit, use IV z such that **estimation moments** are **less informative** about $\hat{\beta}_z$ in the sense of Andrews et al. (2020)
  3. To improve precision, project z on a vector of controls and **use residuals**

FGKK (2020): Monte Carlo for Estimation, Informativeness and Mechanical Fit

"Preferred IV" as before. "Naive IV" further residualized with respect to product-specific fixed effects. σ estimated as in FGKK using product-specific fixed effects. Import quantities are misspecified.





Related Literature

- **Testing via model "forecasts"/backcasts—e.g. correlation(data,model)=1**
  - Lai and Trefler (2002), Costinot and Donaldson (2012), Kehoe et al. (2017), Desmet et al. (2018)
  - $\hat{\beta}_{z'} = \frac{1}{N_W} \sum_n z_n^{'} (\Delta x_n^* - \Delta x_n)$ is very different from $\text{corr}(\Delta y_n, \Delta x_n) \propto \frac{\text{var}(\Delta x_n^*)}{\text{var}(\Delta \eta_n^*)}$

Related Literature

- **Testing via model "forecasts"/backcasts—e.g. correlation(data,model)=1**
  - Lai and Trefler (2002), Costinot and Donaldson (2012), Kehoe et al. (2017), Desmet et al. (2018)
  - $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is very different from $\text{corr}(\Delta y_n, \Delta x_n) \propto \frac{\text{var}(\Delta x_n^*)}{\text{var}(\Delta \eta_n^*)}$
- **Testing via "untargeted moments"**
  - Edmond et al. (2011); Costinot et al. (2016); Antras et al. (2017)
  - Testing $W(\Delta x^*) = W(\Delta x)$ not $f = f^*$

Related Literature

- **Testing via model "forecasts"/backcasts—e.g. correlation(data,model)=1**
  - Lai and Trefler (2002), Costinot and Donaldson (2012), Kehoe et al. (2017), Desmet et al. (2018)
  - $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is very different from $\text{corr}(\Delta y_n, \Delta x_n) \propto \frac{\text{var}(\Delta x_n^*)}{\text{var}(\Delta \eta_n^*)}$

- **Testing via "untargeted moments"**
  - Edmond et al. (2011); Costinot et al. (2016); Antras et al. (2017)
  - Testing $W(\Delta x^*) = W(\Delta x)$ not $f = f^*$

- **Testing via "untargeted causal responses"**
  - "Lucas (1980) Program"—Christiano et al. (1999, 2005), Todd and Wolpin (2006), Nakamura and Steinsson (2014, 2018), Ahlfeldt et al. (2015), Adao et al. (2022)
  - Tests of conduct in IO/Labor: Bresnahan (1982), Berry-Haile (2014), Rousille-Scuderi (2022)
  - $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is weighted avg. of responses that matter for counterfactual
  - How to do inference (dependence, prior estimation)? How to avoid mechanical success?

Testing vs. Estimation

- **Even if agree that moment $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is "useful", why use it for testing rather than estimation?**
  - E.g. could impose $\hat{\beta}_{z'} = 0$ as an additional moment in GMM for estimating θ
  - Would be efficient: minimize asymptotic $\text{Var}(\widehat{\theta})$ under null of correctly specified model
  - Could then also do J-test for purposes of testing A3

Testing vs. Estimation

- **Even if agree that moment $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is "useful", why use it for testing rather than estimation?**
  - E.g. could impose $\hat{\beta}_{z'} = 0$ as an additional moment in GMM for estimating θ
  - Would be efficient: minimize asymptotic Var$(\widehat{\theta})$ under null of correctly specified model
  - Could then also do J-test for purposes of testing A3

- **Two advantages to the testing-based approach developed here:**

Testing vs. Estimation

- **Even if agree that moment $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is "useful", why use it for testing rather than estimation?**
  - E.g. could impose $\hat{\beta}_{z'} = 0$ as an additional moment in GMM for estimating θ
  - Would be efficient: minimize asymptotic Var$(\widehat{\theta})$ under null of correctly specified model
  - Could then also do J-test for purposes of testing A3

- **Two advantages to the testing-based approach developed here:**

  1. **Economic interpretation:**
     - J-test statistic is weighted sum of moment gaps
     - How then to assess errors in the model's counterfactual prediction?

Testing vs. Estimation

- **Even if agree that moment $\hat{\beta}_{z'} = \frac{1}{N_W}\sum_n z_n^{'}(\Delta x_n^* - \Delta x_n)$ is "useful", why use it for testing rather than estimation?**
  - E.g. could impose $\hat{\beta}_{z'} = 0$ as an additional moment in GMM for estimating θ
  - Would be efficient: minimize asymptotic Var($\widehat{\theta}$) under null of correctly specified model
  - Could then also do J-test for purposes of testing A3

- **Two advantages to the testing-based approach developed here:**

  1. **Economic interpretation:**
     - J-test statistic is weighted sum of moment gaps
     - How then to assess errors in the model's counterfactual prediction?
  2. **Power:**
     - Moments used for θ are often relatively "partial equilibrium", but counterfactual is more "GE"
     - GMM: low-variance moments get more weight (for estimation and testing)
     - If GE moments are inherently noisier, this tilts power away from testing the counterfactual

FGKK (2020): Now on actual data from Trump's Trade War, 2016-2019

- Everything exactly as in previous simulations, except...

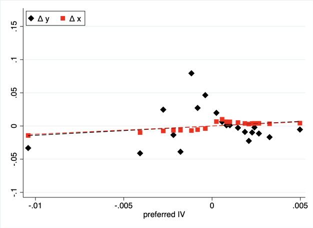
1. Use actual US and foreign tariff changes:
   - $\tau_t \equiv \{\tau^H_{ig,t}, \tau^F_{ig,t}\}$: avg. Jan-Dec, 2016
   - $\tau_{t+1} \equiv \{\tau^H_{ig,t+1}, \tau^F_{ig,t+1}\}$: avg. Jan-April, 2019
2. Use actual data on post-shock outcomes y_{t+1}

FGKK (2020): An IV-based test

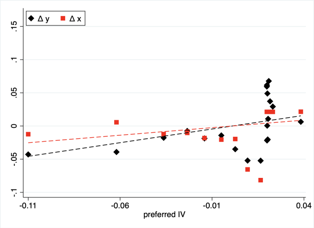
Goodness of fit measure:	Correlation	IV-Based Test	
		Naive IV	Preferred IV
	$Corr\left(\Delta y_n, \Delta x_n(\hat{\theta})\right)$	$\hat{\beta}_{z^{\text{naive}}}(\hat{\theta})$	$\hat{\beta}_{z^{\text{pref}}}(\hat{\theta})$
	(1)	(2)	(3)
Point estimate	0.08	-0.01	-0.09
Inference ignoring estimation uncertainty			
Std. error		0.18	0.15
p-value of H0: $\hat{\beta} = 0$		0.96	0.56
Inference accounting for estimation uncertainty			
Std. error		0.24	0.18
p-value of H0: $\hat{\beta} = 0$		0.97	0.63

Under A3' column (3) $\Rightarrow$ $E_t[W(\Delta x^*) - W(\Delta x)] =$ –\$16 B [CI: +/- \$64 B].
(Recall that $W(\Delta x) =$ –\$7 B.)

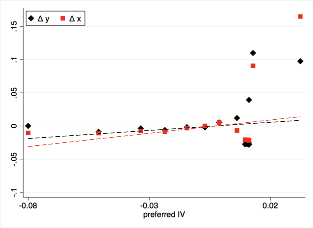
FGKK (2020): A Final Diagnosis



(a) Export Prices (0.16 (0.73))



(b) Import Prices (0.18 (0.07))



(c) Tariff Rev. (-0.16 (0.04))

- Kehoe and Prescott (1995): "... shortcomings in [counterfactual] predictions of a model would then provide motivation for further theoretical development and further testing."

Summary: Questions about your own research questions

- What is the main question I am trying to answer?
- What are the sufficient statistics—weighted averages of causal responses—that answer my question?
- Can I measure those statistics using quasi-experimental variation?
- If not, how is the model I've written down (perhaps coupled with the estimation of parameters in that model) filling in the "missing" suff stats?
- How can I design a test that would tell me if these "missing" suff stats are correctly estimated under reasonable forms of misspecification?
- Does that test reject (and/or does test stat have a worryingly wide CI)?