Difference-in-Differences

Sepehr Ekbatani
(TeIAS)

Spring 2022

What Is Difference-in-Differences?

- A non-random treatment (sometimes called a “natural experiment”) is applied to one or more groups.
- A group of units do not receive the units at the same time (either never, or not yet, for comparison purposes).
- Observations are taken before and after for each group.
- Researcher differences before and after, then differences the difference – hence the name.

Difference-in-Differences

- Very old, conceptually intuitive research design.
- Early attempts at using date back to several health policy debates in the 19th century.
- Brought into labor economics with *Orley Ashenfelter (1978), LaLonde (1985), Card and Krueger (1994)*.
- Has since become the most popular quasi-experimental method, even more than RDD.
  - Over forty thousand hits on google scholar!

John Snow and Cholera

- John Snow, epidemiologist in 19th century, usually credited with first use of DiD.
- Believed cholera was spread through the Thames water supply which contradicted dominant theory about "dirty air" transmission.
- Grand experiment: Lambeth moves its pipe between 1849 and 1854; Southwark and Vauxhall delay.
- How can he use this event to test his hypothesis? Three ways: simple comparisons, interrupted time series of the difference in differences (DiD).

Simple Cross-Sectional Design

Table 1: Lambeth and Southwark and Vauxhall, 1854

Company	Cholera Mortality
Lambeth	$Y = L + D$
Southwark and Vauxhall	$Y = SV$

Interrupted Time Series Design

Table 2: Lambeth, 1849 and 1854

Company	Time	Cholera Mortality
Lambeth	1849	$Y = L$
	1854	$Y = L + (T + D)$

Difference-in-Differences

Table 3: Lambeth and Southwark and Vauxhall, 1849 and 1854

Company	Time	Cholera Mortality	D_1	D_2
Lambeth	Before	$Y = L$		
	After	$Y = L + (T + D)$	$T + D$	
				D
Southwark and Vauxhall	Before	$Y = SV$		
	After	$Y = SV + T$	T	

- We are assuming that there is no time-variant company specific unobservables. i.e., T is the same for all units.

Estimation

- The cholera case is a particular kind of DD design that *Goodman-Bacon (2019)* calls the 2×2 DD design.
- Sample averages:

$$\hat{\delta}_{kU}^{2\times 2} = \left(\bar{y}_k^{post(k)} - \bar{y}_k^{pre(k)}\right) - \left(\bar{y}_U^{post(k)} - \bar{y}_U^{pre(k)}\right)$$

- Population expectations:

$$\hat{\delta}_{kU}^{2\times 2} = (E[Y_k|Post] - E[Y_k|Pre]) - (E[Y_U|Post] - E[Y_U|Pre])$$

Potential Outcomes and The Switching Equation

$$\hat{\delta}^{2\times 2}_{kU} = \underbrace{(E[Y_k|Post] - E[Y_k|Pre]) - (E[Y_U|Post] - E[Y_U|Pre])}_{\text{Switching Equation}}$$
$$+ \underbrace{E[Y^0_k|Post] - E[Y^0_k|Post]}_{\text{Adding Zero}}$$

Parallel Trends Bias

$$\hat{\delta}^{2\times 2}_{kU} = \underbrace{E[Y^1_k|Post] - E[Y^0_k|Post]}_{\text{ATT}}$$
$$+ \underbrace{\Big(E[Y^0_k|Post] - E[Y^0_k|Pre]\Big) - \Big(E[Y^0_U|Post] - E[Y^0_U|Pre]\Big)}_{\text{Non-parallel trends bias in 2x2 case}}$$

- The first term of the bias is counterfactual $\rightarrow$ parallel trends assumption is by definition untestable!
- Unknown bias – could bias it upwards, could bias it downwards.

DiD Assumptions

1. Setting:

	$T = 1$	$T = 2$
Control Group	0% treated	0% treated
Treatment Group	0% treated	100% treated

2. Counterfactual common trends assumption:
    - If treatment units had remained untreated in period 2, then the evolution of their mean outcome between period 1 and 2 would have been the same as that of control units.
3. Exclusion Restriction
    - Nothing else is changing in the treatment group between period 1 and 2.

- Then a two-group/two-period (2x2) DD identifies the average treatment effect on the treated.

ATT

- DD always estimates the ATT because it's only the treatment effect for the treatment group in the post-treatment period.
- It is not the ATE, or the LATE

$$\delta = E[Y_{it}^1 - Y_{it}^0 | D_i = 1]$$

Assumptions

- The common trend assumption will be satisfied if Y_{it}^0 is determined by a model additively separable between a group effect and a time effect.
  - In other words, there are no differential group & time interactions.
- Good practice: Show evidence of parallel trends prior to the policy change.
  - But of course this is not "proof".
- DiD is appropriate when the interventions are as good as random.
  - State A raises the minimum wage and State B doesn't. Is DiD a good idea?
  - Always ask: Why did the intervention happen?

Bias in Our Go-to Estimators

- OLS will identify the ATT with only two groups and no covariates.
- But the more common DD situation is one in which a treatment is adopted by different groups at different times.
- OLS with panel and time fixed effects ("twoway fixed effects" or TWFE) is biased we now know (*Chaisemartin and D'Haultfœuille (2019)*).

Regression DiD

- There are several good reasons to use TWFE.
  - It estimates the ATT under parallel trends,
  - It's easy to calculate the standard errors,
  - It's easy to include multiple periods,
  - We can study treatments with different treatment intensity. (e.g., varying increases in the minimum wage for different states)
- But there are bad reasons, too, which happens when there are differential timing and covariates.

Card and Krueger (1994)

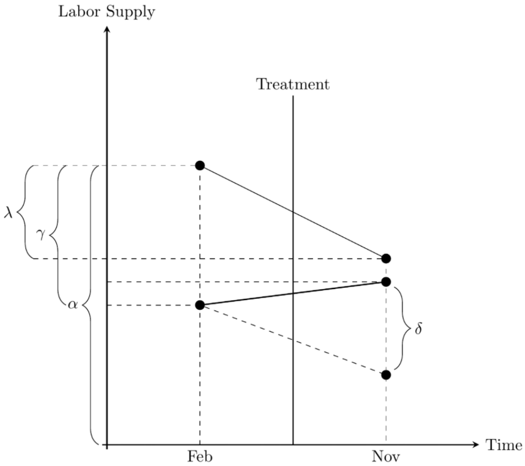
- The equivalent regression includes time and group fixed effects:

$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$

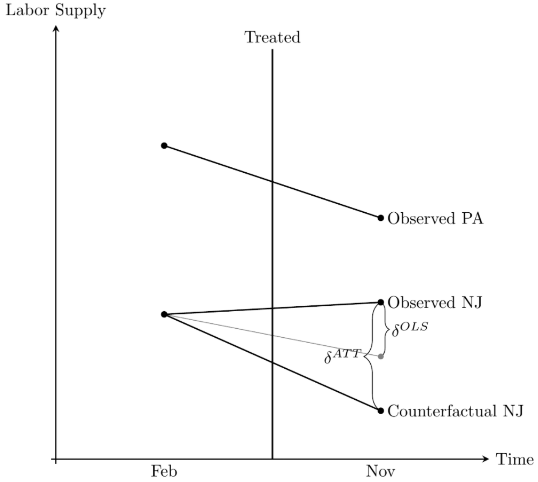
  - NJ is a dummy equal to 1 if the observation is from NJ
  - d is a dummy equal to 1 if the observation is from November (the post period).
- This equation takes the following values
  - PA Pre: α,
  - PA Post: $\alpha + \lambda$,
  - NJ Pre: $\alpha + \gamma$,
  - NJ Post: $\alpha + \gamma + \lambda + \delta$.
- DD estimate: (NJ Post - NJ Pre) - (PA Post - PA Pre) $= \delta$.

DiD Regression Diagram

$$Y_{its} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ \times d)_{st} + \varepsilon_{its}$$



DiD Regression Diagram Without PT



Standard Errors

- Most DiD papers use many years of data and focus on serially correlated outcomes but ignore that the resulting standard errors are inconsistent.
- *Bertrand, Duflo and Mullainathan (2004)* point out, conventional standard errors often severely understate the standard deviation of the estimators: standard errors are biased downward.

Standard Errors: Solutions

1. Block bootstrapping standard errors.
2. Aggregating the data into one pre and one post period.
3. Clustering standard errors at the group level.
    - Where there is only one treatment group, randomization inference is preferred following *Buchmuller et al. (2011)*.

Checks for DiD Design

- Falsification test using data for prior periods.
- Falsification test using data for alternative control group.
- Falsification test using alternative "placebo" outcome that should not be affected by the treatment.

More on Parallel Trends

- One situation where parallel trends would be obviously violated is if the *treatment itself was endogenous*.
- Checking the pre-treatment balance between treatment and control groups.
- Simply show the raw data, which you can do if you have a set of groups who received the treatment at the same point in time.
- What if instead you have differential timing wherein groups of units adopt the treatment at different points?

Cheng and Hoekstra (2013)

- Plot raw data and simply eyeball.

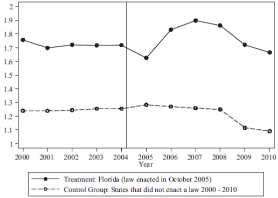


Figure 1a: State Adopting in 2005 (Florida)

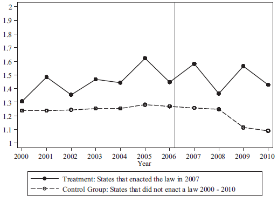


Figure 1d: States Adopting in 2008 (Ohio, West Virginia)

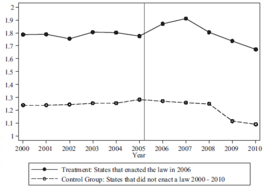


Figure 1b: States Adopting in 2006 (Alabama, Alaska, Arizona, Georgia, Indiana, Kansas, Kentucky, Louisiana, Michigan, Mississippi, Oklahoma, South Carolina, South Dakota)

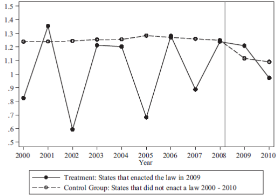


Figure 1e: State Adopting in 2009 (Montana)

Plotting the Raw Data

- Advantage:
  - Transparent display of the raw unadjusted data.
- Disadvantages:
  - Cumbersome when the number of treatment groups is large.
  - May not be beautiful.
  - Necessarily assumes that the only control group is the never-treated group, which in fact is not true. (*Goodman-Bacon (2019)*)
- Any DD is a combination of a comparison between
  - Treated and the never treated,
  - An early treated compared to a late treated, and
  - A late treated compared to an early treated.

Anderson et al. (2013)

- Construct a recentered time path of traffic fatality rates for the control states by assigning random treatment dates to all control counties.

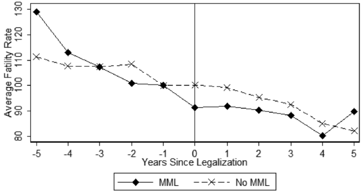


Figure 1. Pre- and postlegalization trends in traffic fatality rates, ages 15–19

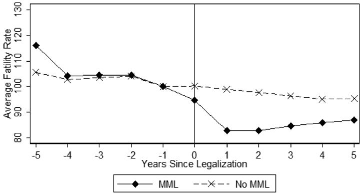
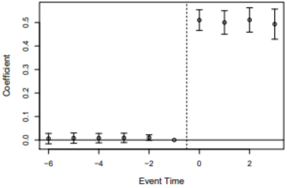


Figure 2. Pre- and postlegalization trends in traffic fatality rates, ages 20–39

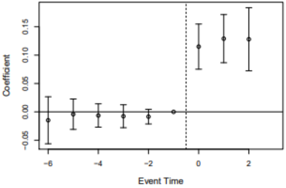
Random Date

- Advantages:
  - It plots the raw data, rather than coefficients from a regression
  - It plots that data against controls.
- Weakness:
  - Technically, the control series is not in fact true.
  - It is chosen so as to give a comparison, but when regressions are eventually run, it will not be based on this series.

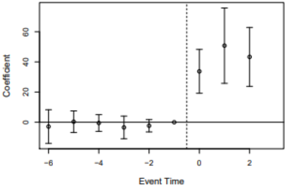
- Estimate a regression model that includes treatment leads and lags.



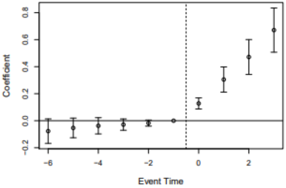
(a) Medicaid Eligibility (ACS)



(b) Any Medicaid Enrollment in Year (CMS)



(c) # Days Medicaid Enrollment in Year (CMS)



(d) Cumulative Years of Medicaid Exposure (CMS)

Treatment Leads and Lags

- As with many contemporary DD designs, *Miller et al. (2019)* evaluate the pre-treatment leads instead of plotting the raw data by treatment and control.

$$Y_{its} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \gamma_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + X_{ist} + \varepsilon_{ist}$$

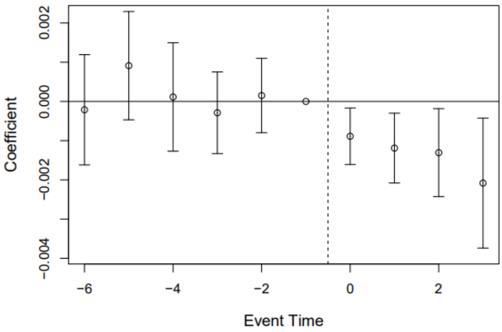
- Treatment occurs in year 0. Include q leads or anticipatory effects and m lags or post-treatment effects.

1. Allows the reader to check both the degree to which the post-treatment treatment effects were dynamic.
2. Whether the two groups were comparable on outcome dynamics pre-treatment.

Miller et al. (2019) Main Result

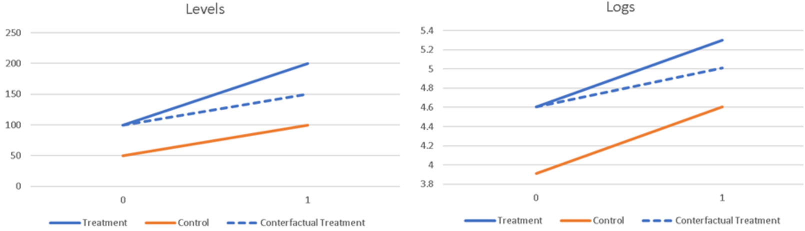
- ACA Medicaid led to a 0.13 percentage point decline in annual mortality, which is a 9.3 percent reduction over the mean.

Figure II: Effect of the ACA Medicaid Expansions on Annual Mortality



One Last Thing

- A last draw-back of DiD is that they are sensitive to the scaling of the outcome.
- This is because the common trend assumption is not invariant to the scaling of the outcome: it cannot hold both for the outcome and for its logarithm!



Importance of Placebos in DiD

- The idea is simple:
  - Say that you are finding some negative effect of the minimum wage on low-wage employment.
  - If you cannot reject the null on an alternative hypothesis, then it provides some credibility to your original analysis.
  - Use data for an alternative type of worker whose wages would not be affected by the binding minimum wage.
  - Minimum wage should not affect the employment of higher wage workers,

Triple Differences

- What if there were state-specific time shocks such as NJ_t or PA_t in *Card and Krueger (1994)*?

Table 4: Triple differences design.

States	Group	Period	Outcomes	D_1	D_2	D_3
NJ	Low-wage	After	$NJ_l + T + NJ_t + l_t + D$	$T + NJ_t +$	$(l_t - h_t) + D$	D
		Before	NJ_l	$l_t + D$		
	High-wage	After	$NJ_h + T + NJ_t + h_t$	$T + NJ_t + h_t$		
		Before	NJ_h			
PA	Low-wage	After	$PA_l + T + PA_t + l_t$	$T + PA_t + l_t$	$l_t - h_t$	
		Before	PA_l			
	High-wage	After	$PA_h + T + PA_t + h_t$	$T + PA_t + h_t$		
		Before	PA_h			

DDD Estimation

- Estimating a DDD model requires estimating the following regression:

$$
\begin{aligned}
Y_{ijs} = \alpha + \psi X_{ijt} + \beta_1 \tau_t + \beta_2 \delta_j + \beta_3 D_i + \beta_4 (\delta \times \tau)_{jt} \\
+ \beta_5 (\tau \times D)_{ti} + \beta_6 (\delta \times D)_{ij} + \beta_7 (\delta \times \tau \times D)_{ijt} + \varepsilon_{ijt}
\end{aligned}
$$

- Parameter of interest is β_7.

Fuzzy DD

- *De Chaisemartin, C., & d'Haultfoeuille, X. (2018)*. Fuzzy differences-in-differences. The Review of Economic Studies, 85(2), 999-1028.

	Older Cohort	Younger Cohort
High Treatment Regions	81.2%	90.0%
Low Treatment Regions	89.8%	94.3 %

Fuzzy DD: Duflo (2005)

- In fuzzy designs, a commonly used estimand is:

$$W_{DID} = \frac{E(Y_{1,1}) - E(Y_{1,0}) - (E(Y_{0,1}) - E(Y_{0,0}))}{E(D_{1,1}) - E(D_{1,0}) - (E(D_{0,1}) - E(D_{0,0}))}$$

- Under common trends what does W_{DID} identifies?
  - Nothing of interest!

Section 1

Salience and Taxation: Theory and Evidence

Raj Chetty, Adam Looney, Kory Kroft (2009)

Definition

- Tax incidence is the study of the effects of tax policies on prices and the distribution of utilities.
- What happens to market prices when a tax is introduced or changed?
  - Increase tax on cigarettes by $1 per pack.
  - Introduction of Earned Income Tax Credit (EITC).
  - Food stamps program.
- Effect on price $\rightarrow$ distributional effects on smokers, profits of producers, shareholders, farmers, ...

Economic vs. Statutory Incidence

- Equivalent when prices are constant but not in general.
- Consider the following argument:
  - Government should tax capital income b/c it is concentrated at the high end of the income distribution.
- Neglects general equilibrium price effects:
  - Tax might be shifted onto workers.
  - If capital taxes $\rightarrow$ less savings and capital fight, then capital stock may decline, driving return to capital up and wages down.
  - Some argue that capital taxes are paid by workers and therefore increase income inequality (*Hassett and Mathur 2009*).

Overview

- Tax incidence is an example of positive analysis
  - Typically the first step in policy evaluation.
  - An input into thinking about policies that maximize social welfare.
- Theory is informative about signs and comparative statics but is inconclusive about magnitudes
  - Incidence of cigarette tax: elasticity of demand w.r.t. price is crucial.
  - Labor vs. capital taxation: mobility of labor, capital are critical.

Overview

- Ideally, we would characterize the effect of a tax change on utility levels of all agents in the economy.
- Useful simplification in practice: aggregate economic agents into a few groups.
- Incidence analyzed at a number of levels:
  1. Producer vs. consumer (tax on cigarettes)
  2. Source of income (labor vs. capital)
  3. Income level (rich vs. poor)
  4. Region or country (local property taxes)
  5. Across generations (social security reform)

Partial Equilibrium Incidence: Key Assumptions

1. Two good economy
    - Only one relative price $\rightarrow$ partial and general equilibrium are same.
    - Can be viewed as an approx. of incidence in a multi-good model if:
        - the market being taxed is "small"
        - there are no close substitutes/complements in the utility function
2. Tax revenue is not spent on the taxed good
    - Tax revenue is used to buy untaxed good or thrown away
3. Perfect competition among producers
    - Relaxed in some studies of monopolistic or oligopolistic markets

Partial Equilibrium Model: Setup

- Two goods: x and y
- Government levies an **excise** tax on good x.
  - **Excise or specific tax**: levied on a quantity (e.g. gallon, pack, ton)
  - **Ad-valorem tax**: fraction of prices (e.g. sales tax)
- Let p denote the pretax price of x and $q = p + t$ denote the tax inclusive price of x.
- Good y, the numeraire, is untaxed.

Partial Equilibrium Model: Demand

- Consumer has wealth Z and has utility $u(x, y)$.
- Let $\varepsilon_D = \frac{\partial D}{\partial q}\frac{q}{D(q)} = \frac{\partial \log D}{\partial \log q}$ denote the price elasticity of demand
  - Elasticity: % change in quantity when price changes by 1%.
  - Widely used concept because elasticities are unit free.

Partial Equilibrium Model: Supply

- Price-taking firms.
- Use $c(S)$ units of the numeraire y to produce S units of x.
- Cost of production is increasing and convex:

$$c'(S) > 0 \text{ and } c''(S) \geq 0$$

- Profit at pretax price p and level of supply S is $pS - c(S)$.
- With perfect optimization, the supply function for good x is implicitly defined by the marginal condition $p = c'(S(p))$
- Let $\varepsilon_S = \frac{\partial S}{\partial p} \frac{p}{S(p)}$ denote the price elasticity of supply.

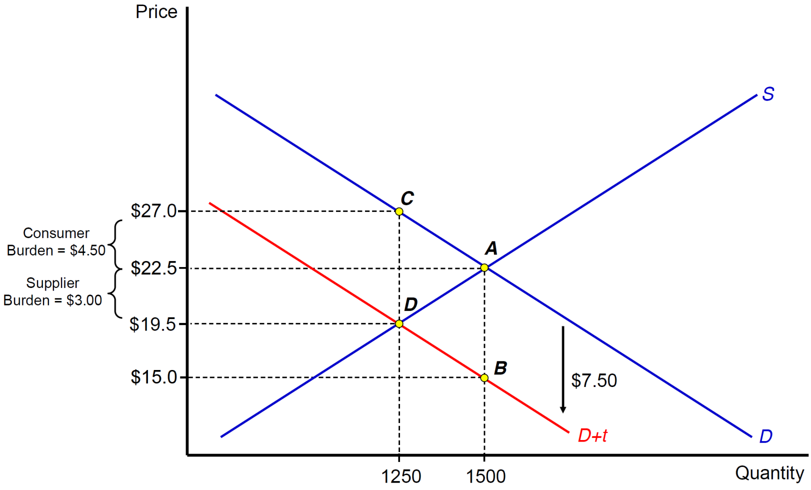
Partial Equilibrium Model: Equilibrium

- Equilibrium condition

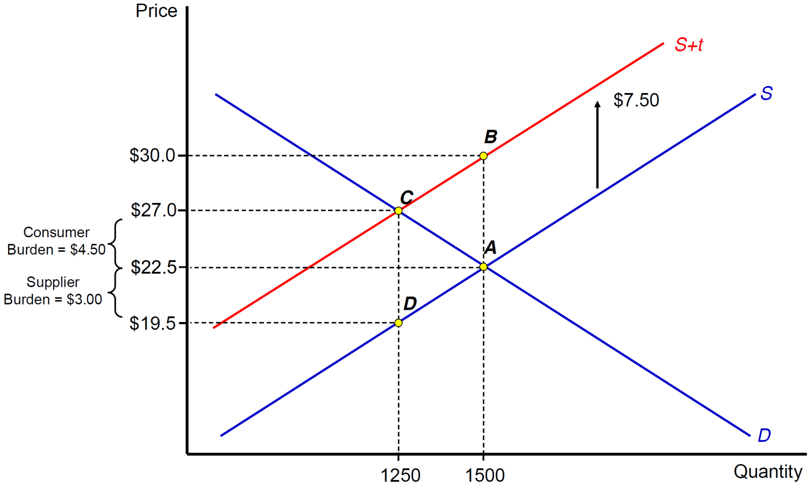
$$Q = S(p) = D(p + t)$$

  defines an equation $p(t)$.
- Goal: characterize $\frac{dp}{dt}$, the effect of a tax increase on price.
- First consider some graphical examples to build intuition, then analytically derive formula.

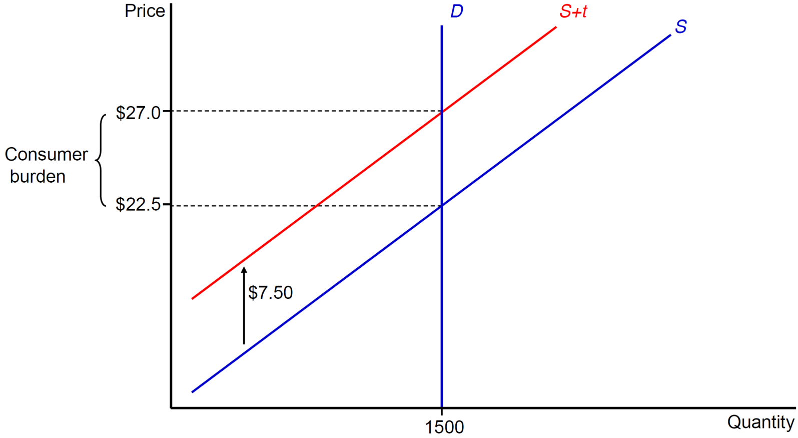
Tax Levied on Consumers



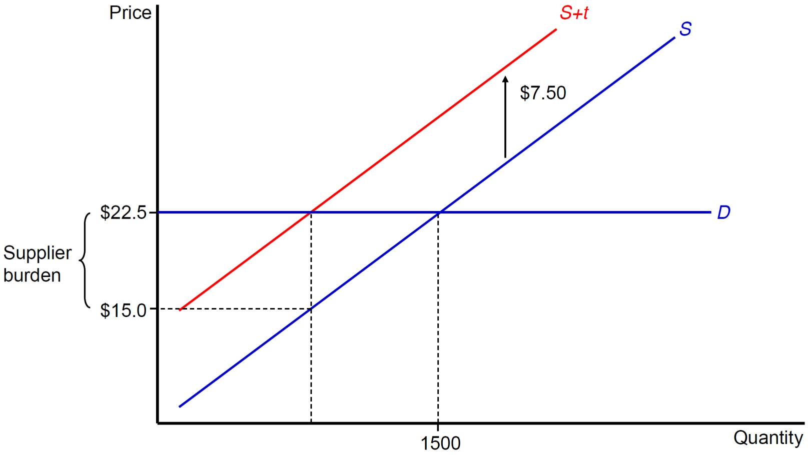
Tax Levied on Producers



Perfectly Inelastic Demand



Perfectly Elastic Demand



Formula for Tax Incidence

- Implicitly differentiate equilibrium condition

$$D(p+t) = S(p)$$

to obtain:

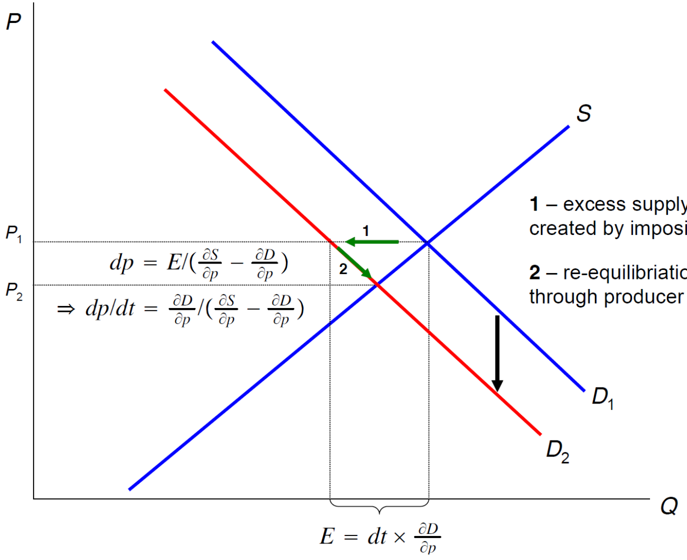
$$\frac{dp}{dt} = \frac{\partial D}{\partial p} \frac{1}{(\frac{\partial S}{\partial p} - \frac{\partial D}{\partial p})}$$

$$\Rightarrow \frac{dp}{dt} = \frac{\varepsilon_D}{\varepsilon_S - \varepsilon_D}$$

- Incidence on consumers:

$$\frac{dq}{dt} = 1 + \frac{dp}{dt} = \frac{\varepsilon_S}{\varepsilon_S - \varepsilon_D}$$

Tax Incidence Depiction



Tax Incidence With Salience Effects

- Central assumption of neoclassical model: taxes are equivalent to prices ($\frac{dx}{dt} = \frac{dx}{dp}$)
- In practice, are people fully aware of marginal tax rates?
- Chetty, Looney, and Kroft (2009) test this assumption and generalize theory to allow for salience effects.
- **Part 1**: Test whether "salience" (visibility of tax-inclusive price) affects behavioral responses to commodity taxation
  - Does effect of a tax on demand depend on whether it is included in **posted** price?
- **Part 2**: Develop formulas for incidence and efficiency costs of taxation that permit salience effects and other optimization errors.

Chetty et al.: Empirical Framework

- Economy with two goods, x and y.
- Prices: Normalize the price of y to 1 and let p denote the (fixed) *pretax* price of x.
- Taxes: y untaxed, x subject to an ad valorem sales tax τ (not included in posted price)
    - Tax-inclusive price of x is $q = (1 + \tau)p$
- Let demand for good x be denoted by $x(p, \tau)$.

Chetty et al.: Empirical Framework

- If agents optimize fully, demand should only depend on the total tax-inclusive price: $x(p,\tau) = x((1+\tau)p, 0)$.
- Full optimization implies price elasticity equals gross-of-tax elasticity:

$$\varepsilon_{x,p} \equiv -\frac{\partial \log x}{\partial \log p} = \varepsilon_{x,1+\tau} \equiv -\frac{\partial \log x}{\partial \log(1+\tau)}$$

- To test this hypothesis, log-linearize demand fn. $x(p,\tau)$ to obtain estimating equation:

$$\log x(p,\tau) = \alpha + \beta \log p + \theta\beta \log(1+\tau)$$

- θ measures degree to which agents under-react to the tax:

$$\theta = \frac{\partial \log x}{\partial \log(1+\tau)} \Big/ \frac{\partial \log x}{\partial \log p} = \frac{\varepsilon_{x,1+\tau}}{\varepsilon_{x,p}}$$

Chetty et al.: Two Empirical Strategies

Two strategies to estimate θ:

1. **Manipulate tax salience**: make sales tax as visible as pre-tax price.
   - Effect of intervention on demand:

$$\nu = \log x((1+\tau)p, 0) - \log x(p, \tau)$$

   - Compare to effect of equivalent price increase to estimate θ:

$$(1-\theta) = -\frac{\nu}{\varepsilon_{x,p}\log(1+\tau)}$$

2. **Manipulate tax rate**: compare $\varepsilon_{x,p}$ and $\varepsilon_{x,1+\tau}$.

$$\theta = \varepsilon_{x,1+\tau}/\varepsilon_{x,p}$$

Chetty et al.: Strategy 1

- Experiment manipulating salience of sales tax implemented at a supermarket that belongs to a major grocery chain.
  - 30% of products sold in store are subject to sales tax
  - Posted tax-inclusive prices on shelf for subset of products subject to sales tax (7.375% in this city)
- Data: Scanner data on price and weekly quantity sold by product.

Tax-Inclusive Price Tags

Source: Chetty, Looney, and Kroft (2009)

Documenting Salience: Mechanism

- Concern with posting tax inclusive prices: may have influenced behavior through various channels besides salience.
- Common problem in field experiments: *Hawthorne effects*.
- Difficult to rule out all mechanisms, but helpful to present evidence that mechanism of interest is very powerful.

Classroom Survey

- Students were asked to choose two items from image.
- Asked to report "Total bill due at the register for these two items."

	Mean	Median	SD
Original Price Tags:			
Correct tax-inclusive price w/in \$0.25	0.18	0.00	0.39
Experimental Price Tags:			
Correct tax-inclusive price w/in \$0.25	0.75	1.00	0.43
T-test for equality of means: $p < 0.001$			
$N = 49$			

Source: Chetty, Looney, and Kroft (2009)

Chetty et al.: Research Design

- Quasi-experimental difference-in-differences.
- Treatment group:
  - Products: Cosmetics, Deodorants, and Hair Care Accessories.
  - Store: One large store in Northern California.
  - Time period: 3 weeks (February 22, 2006 - March 15, 2006).
- Control groups:
  - Products: Other prods. in same aisle (toothpaste, skin care, shave).
  - Stores: Two nearby stores similar in demographic characteristics.
  - Time period: Calendar year 2005 and first 6 weeks of 2006.

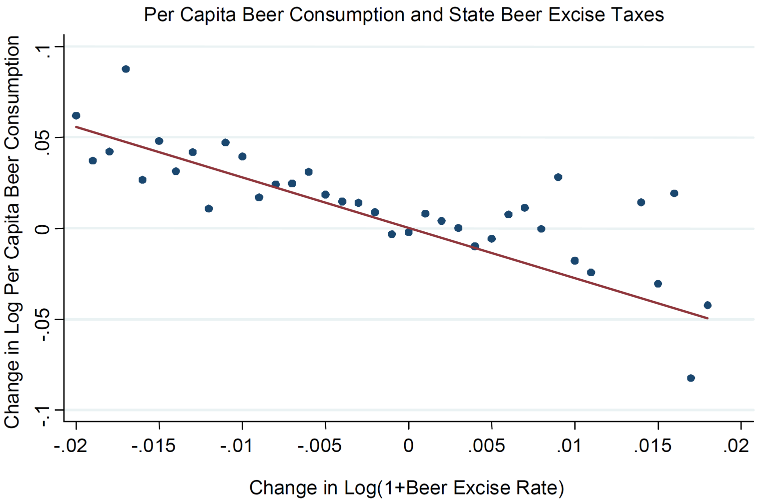
	Treatment Store		
Period	Control Categories	Treated Categories	Difference
Baseline	26.48	25.17	-1.31
	(0.22)	(0.37)	(0.43)
Experiment	27.32	23.87	-3.45
	(0.87)	(1.02)	(0.64)
Difference	0.84	-1.30	$DD_{TS} = -2.14$
over time	(0.75)	(0.92)	(0.64)

	Control Stores		
Period	Control Categories	Treated Categories	Difference
Baseline	30.57	27.94	-2.63
	(0.24)	(0.30)	(0.32)
Experiment	30.76	28.19	-2.57
	(0.72)	(1.06)	(1.09)
Difference	0.19	0.25	$DD_{CS} = 0.06$
over time	(0.64)	(0.92)	(0.90)
		DDD Estimate	-2.20
			(0.58)

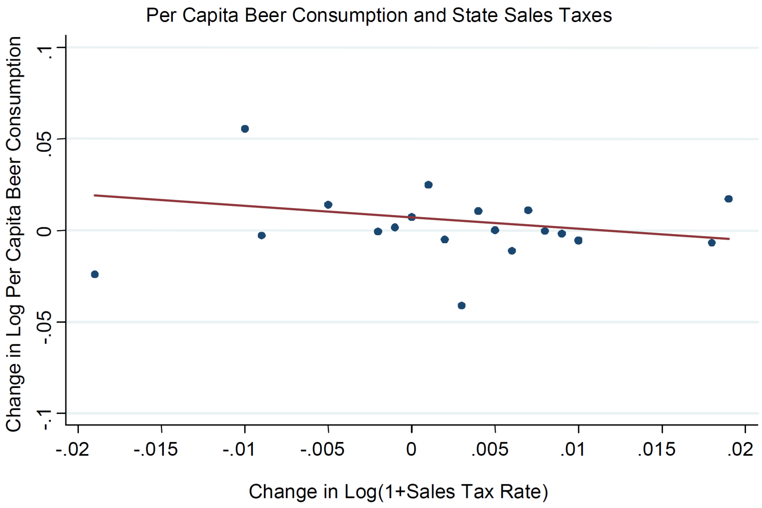
Chetty et al.: Strategy 2

- Compare effects of price changes and tax changes.
- Alcohol subject to two state-level taxes in the U.S.:
  - Excise tax: included in price.
  - Sales tax: added at register, not shown in posted price.
- Exploiting state-level changes in these two taxes, estimate θ.
  - Addresses concern that experiment may have induced a Hawthorne effect.

Excise Taxes



Sales Taxes



Per Capita Beer Consumption and State Sales Taxes
Change in Log Per Capita Beer Consumption
-.1
-.05
0
.05
.1
-.02
-.015
-.01
-.005
0
.005
.01
.015
.02
Change in Log(1+Sales Tax Rate)

Excise And Sales Taxes

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: Δ**Log(per capita beer consumption)**

	Baseline	Bus Cyc, Alc Regs.	3-Year Diffs	Food Exempt
	(1)	(2)	(3)	(4)
ΔLog(1+Excise Tax Rate)	**-0.87**	**-0.89**	**-1.11**	**-0.91**
	(0.17)***	(0.17)***	(0.46)**	(0.22)***
ΔLog(1+Sales Tax Rate)	**-0.20**	**-0.02**	**-0.00**	**-0.14**
	(0.30)	(0.30)	(0.32)	(0.30)
Business Cycle Controls		x	x	x
Alcohol Regulation Controls		x	x	x
Year Fixed Effects	x	x	x	x
F-Test for Equality of Coeffs.	0.05	0.01	0.05	0.04
Sample Size	1,607	1,487	1,389	937

Tax Incidence with Salience Effects

- Let $\{x(p,t,Z), y(p,t,Z)\}$ denote empirically observed demands.
- Place no structure on these demand functions except for feasibility:

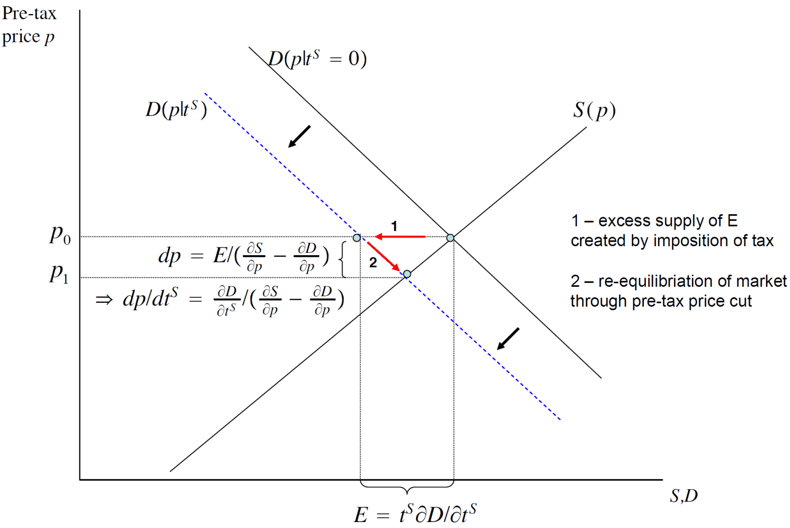
$$(p+t)x(p,t,Z) + y(p,t,Z) = Z$$

- Demand functions taken as empirically estimated objects rather than optimized demand from utility maximization.
- Supply side model same as above.
- Market clearing price p satisfies:

$$D(p,t,Z) = S(p)$$

where $D(p,t,Z) = x(p,t,Z)$ is market demand for x.

Tax Incidence with Salience Effects



Tax Incidence with Salience Effects: Formula

- Incidence on producers of increasing t is

$$\frac{dp}{dt} = \frac{\partial D/\partial t}{\partial S/\partial p - \partial D/\partial p} = -\theta \frac{\varepsilon_D}{\varepsilon_S - \varepsilon_D}$$

  1. Incidence on producers attenuated by θ.
  2. No tax neutrality: taxes on producers have greater incidence on producers than non-salient taxes levied on consumers.

- **Intuition**: Producers need to cut pretax price less when consumers are less responsive to tax.

Section 2

Bride Price and Female Education

Nava Ashraf, Natalie Bau, Nathan Nunn, Alessandra Voena (2020)

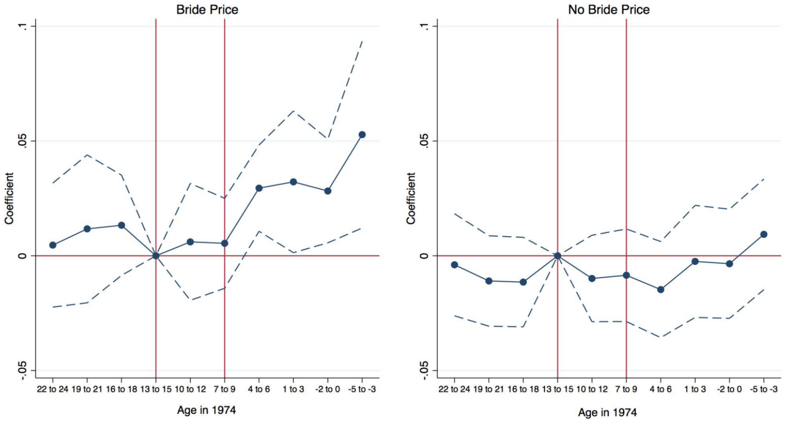
Cultural Traditions and Development Policy

- Cultural traditions may interact with development policies.
  - They shape economic behavior (Fernández et al. 2004, 2009)
  - Recent calls for tailored policies (WDR, 2015)
- This paper: the custom of bride price
  - Payment from the groom to the family of the bride
  - Widespread in parts of Asia and in sub-Saharan Africa
  - Large payments, often above per capita GDP

Bride Price and Female Education

- A possible reward to parents for investing in daughters.
  - Common narrative in our fieldwork.
  - Consistent with anthropologists' interpretation.
  - Correlation between amount paid and a woman's education.
- Bride price may affect the efficacy of education policies.
  - It raises parents' returns to investing in a daughter's schooling.
- Study how the effects of large-scale school construction programs vary with bride price:
  - **Indonesia**: revisit INPRES construction of 68,000 schools studied by *Duflo (2001)*.
    - Previous findings: no detectable effect on female education.
  - **Zambia**: new variation on school construction in '90s and '00s from Ministry of Education.

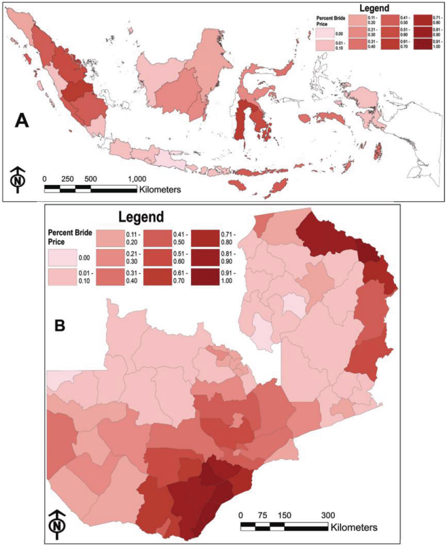
Effect of INPRES by Bride Price Tradition



Becker (1981, 129)

"Bride price then not only compensates parents for the transfer of their property, but also induces them to invest optimally in daughters if girls with appropriate accumulations of human capital command sufficiently high prices."

Geographical Distribution of Bride Price Practice



Some Evidence

- Bride price payments and parent's beliefs about returns to education
  - Qualitative surveys show parents in Zambia believe education is the main driver of bride prices.
- Bride price payments and women's characteristics
  - In both Indonesia and Zambia, higher female education correlates with higher bride prices.

Survey of Parental Beliefs - Zambia

- Think about the factors that affect bride price today: what is the first/ second/ third most important factor?

		(1)	(2)	(3)	(4)
		first	*second*	*third*	*not listed*
Education	**obs**	**543**	**223**	**152**	**538**
		37.29%	**15.32%**	**10.44%**	**36.95%**
Good morals	obs	191	283	216	766
		13.12%	19.44%	14.84%	52.61%
Family values	obs	214	272	206	764
		14.70%	18.68%	14.15%	52.47%
Virginity	obs	137	186	179	954
		9.41%	12.77%	12.29%	65.52%
Age	obs	41	94	141	1180
		2.82%	6.46%	9.68%	81.04%
Tribe	obs	104	117	190	1045
		7.14%	8.04%	13.05%	71.77%
Other	obs	144	118	85	1109
		9.89%	8.10%	5.84%	76.17%

Survey of Parental Beliefs - Zambia Cont.

- If there is an increase in bride price that comes from the bride's education, why do you think that is?

		(1) Unprompted reason	(2) Prompted reason	(3) Not a reason
Bride's parents should to be	obs.	793	461	178
compensated for investments		55.38%	32.19%	12.43%
Improves the bride's skills	obs.	217	485	728
in the house		15.17%	33.92%	50.91%
Improves the woman's	obs.	213	720	481
earning potential		15.06%	50.92%	34.02%
Improves the bride's knowledge	obs.	177	613	631
and skills as a mother		12.46%	43.14%	44.41%
Means her parents	obs.	110	521	789
are rich		7.75%	36.69%	55.56%
Improves the literacy	obs.	89	740	580
of children		6.32%	52.52%	41.16%
Other reasons	obs.	27	-	-
		3.96%	-	-

Education and Bride Price Amounts - Indonesia

- Present discounted value of the return to parents to ensuring their daughter completes primary school is 14% of one year of per capita consumption.

	(1)	(2)	(3)
	Dependent Variable: Log Bride Price Amount		
$I(Completed\ Primary)_i$	0.656***	0.621***	0.544***
	(0.073)	(0.077)	(0.073)
$I(Completed\ Junior\ Secondary)_i$	0.644***	0.653***	0.620***
	(0.073)	(0.077)	(0.072)
$I(College)_i$	0.837***	0.826***	0.887***
	(0.081)	(0.083)	(0.080)
$MarriageAge_i$		0.014	0.006
		(0.014)	(0.013)
$MarriageAge_i^2$		-0.000	-0.000
		(0.000)	(0.000)
Year of Marriage FE	Y	Y	Y
Ethnicity FE	N	N	Y
Survey Round FE	Y	Y	Y
Observations	5,403	5,076	5,076
Adjusted R-Squared	0.426	0.426	0.490

School Construction Programs

- Sekolah Dasar INPRES school construction program of the 1970s in Indonesia.
-

Model Overview

- Draw on seminal models by *Chiappori, Iyigun, and Weiss (2009)* and *Chiappori, Dias, and Meghir (2016).*
- Imperfectly altruistic parents choose whether to educate their children.
- Men and women match in a frictionless marriage market.
- Bride price is modeled as the marital transfer to women that is appropriated by the bride's parents.

Model Predictions

- In equilibrium, the bride price transfer is increasing in the education of the bride.

  $\rightarrow$ The bride price provides an additional monetary incentive for parents to invest in their daughters' education.
- If the bride price custom is present, female education rates are higher.
- As long as education rates are low, female education is more responsive to policies that reduce the costs of schooling.
- Male education rates and their responsiveness to the cost of schooling do not depend on bride price.

Model Setup

- The model has two periods.
- Multiple ethnic groups, e.
- Parents have only one child.
- Unit mass of women (daughters).
- Unit mass of men (sons).

Parents With A Daughter

- Their problem is:

$$\max_{S_i \in \{0,1\}, c_1, c_2 \geq 0} \quad \underbrace{c_1 + \frac{c_2}{1+r}}_{\text{Consumption Utility}} \quad + \gamma \underbrace{\left[a_i S_i + \frac{v_2(S_i, BP_e)}{1+r} \right]}_{\text{Child Well-Being}}$$

$$\max_{S_i \in \{0,1\}, c_1, c_2 \geq 0} c_1 + \frac{c_2}{1+r} + \gamma \left[a_i S_i + \frac{v_2(S_i, BP_e)}{1+r} \right]$$

$r \geq 0$: Discount rate,

$S_i \in \{0, 1\}$: Girl's education,

a_i : Innate ability,

v : Daughter's utility,

$\gamma \in (0, 1)$: Weight parents place on daughter's utility.

$$\text{s.t.} \quad c_1 + k.S_i \leq y \;\text{ and }\; c_2 \leq y + BP_e$$

k : Cost of educating the daughter,

Parents With A Son

- Their problem is:

$$\max_{P_j \in \{0,1\}, c_1, c_2 \geq 0} c_1 + \frac{c_2}{1+r} + \delta \left[a_j P_j + \frac{u_2(P_j, BP_e)}{1+r} \right]$$

$r \geq 0$: Discount rate,

$P_j \in \{0,1\}$: Son's education,

a_j : Innate ability,

u : Son's utility,

$\delta \in (0,1)$: Weight parents place on son's utility.

$$\text{s.t.} \quad c_1 + k.P_j \leq y \ \text{ and } \ c_2 \leq y$$

k : Cost of educating the son,

y : Parent's income in each period.

Bride Price

- For ethnic groups without a bride price tradition ($I_e = 0$), $BP_e = 0$.
- Where $I_e = 1$, bride price is an equilibrium object, $BP(S_i, P_j, k)$, that depends on
  - Education of the bride
  - Education of the groom
  - The cost of schooling.
- Bride price is paid by the groom in the second period, which is reflected by the fact that u_2 depends on the bride price BP_e.

Mariage Market

- For ethnic groups with a bride price tradition ($I_e = 1$), the bride's parents appropriate the marriage market transfer.
- For ethnic groups without a bride price tradition ($I_e = 0$), bride and the groom share the marriage surplus through the intrahousehold allocation of resources.
- Define ζ_i^f and ζ_j^m to be a woman's and man's respective value if they remain single (their labor market earnings.)
- ζ_{ij} be the total value of a marriage between i and j.

Marriage Surplus

- Marriage surplus is defined as $z_{ij} = \zeta_{ij} - \zeta_i^f - \zeta_j^m$.
- Since one's value when single and themarriage surplus depend on only education: $z_{ij} = z_{S_i P_j}$.
- **Assumptions**:
  1. Marriage surplus is increasing in educational attainment:
  $$z_{11} - z_{10} > 0, z_{11} - z_{01} > 0, z_{10} - z_{00} > 0, z_{01} - z_{00} > 0.$$
  2. Supermodularity (Complementarity between the education of the bride and groom):
  $$z_{00} + z_{11} > z_{10} + z_{01}$$
  3. Marriage between two uneducated people is always positive: $z_{00} > 0$. (This ensures that everyone marries in equilibrium.)

Matching

Prediction 1

Matching is assortative by education and is irrespective of the practice of bride price; i.e., educated women marry only educated men, while some educated men marry uneducated women.

Prediction 2

Educated women command a higher bride price payment at marriage than uneducated women.

Bride Price and Investment in Female Education

Prediction 3

The probability that a girl is educated, $Pr(S_i = 1)$, is higher among ethnicities that engage in bride price payments:

$$Pr(S_i = 1 | I_e = 1, k) > Pr(S_i = 1 | I_e = 0, k).$$

Prediction 3 - Model

- Substituting the budget constraint into the objective function:

$$\max_{S_i} y - kS_i + \frac{y + BP_e(S_i).I_e}{1+r} + \gamma \left[a_i S_i + \frac{v_2(S_i)}{1+r} \right]$$

$$S_i = 0 : y + \frac{y + BP_e(S_i = 0).I_e}{1+r} + \gamma \left[\frac{v_2(S_i = 0)}{1+r} \right]$$

$$S_i = 1 : y - k + \frac{y + BP_e(S_i = 1).I_e}{1+r} + \gamma \left[a_i + \frac{v_2(S_i = 1)}{1+r} \right]$$

- Parents choose to educate their daughter $(S_i = 1)$ if her ability exceeds a certain threshold:

$$a_i \geq a^*_{I_e}(k) \equiv \frac{k}{\gamma} - \frac{R^f}{1+r} - \frac{\Delta V_{i,e}}{1+r} \left(1 + I_e \frac{1-\gamma}{\gamma} \right)$$

$R_f \equiv \zeta_1^f - \zeta_0^f$: Labor market returns to schooling,

$\Delta V_{i,e}$: Marriage market returns to education.

Prediction 3 - Model Cont.

- As long as more boys are educated than girls, before and after the school construction program, the ability threshold for girls become:

$$a^*_{I_e}(k) = \frac{k}{\gamma} - \frac{R^f}{1+r} - \frac{z_{11} - z_{01}}{1+r}\left(1 + I_e \frac{1-\gamma}{\gamma}\right)$$

- Since $\gamma \in (0,1)$, we have $a^*_{BP} < a^*_{NoBP}$, hence $Pr(S_i = 1 | I_e = 1, k) > Pr(S_i = 1 | I_e = 0, k)$.

Responsiveness of Female Schooling to Education Policies

Prediction 4

When female education rates are low, a decline in the cost of schooling increases education more for girls from bride price ethnic groups than for girls from non–bride price ethnic groups.

Prediction 4 - Model

- Consistent with the setting, they consider the case in which education levels are low.
- The girl with modal ability does not get educated.
- Then:

$$g(a^*_{BP}) - g(a^*_{NoBP}) > 0.$$

$g()$: Innate ability (a_i) unimodal probability density function.

Prediction 4 - Model Cont.

- The probability that a daughter is educated:

$$\begin{aligned}\Pr(S_i = 1|I_e, k) &= \Pr\left(a_i \geq \frac{k}{\gamma} - \frac{R^f}{1+r} - \frac{\Delta V_{i,e}}{1+r}\left(1 + I_e\frac{1-\gamma}{\gamma}\right)\right)\\ &= 1 - G\left(\frac{k}{\gamma} - \frac{R^f}{1+r} - \frac{\Delta V_{i,e}}{1+r}\left(1 + I_e\frac{1-\gamma}{\gamma}\right)\right)\end{aligned}$$

- Main empirical result then follows directly from the comparison of:

$$\frac{\partial \Pr(S_i = 1|k, I_e = 1)}{\partial k} = -\frac{g(a^*_{BP}(k))}{\gamma} \qquad \text{versus}$$

$$\frac{\partial \Pr(S_i = 1|k, I_e = 0)}{\partial k} = -\frac{g(a^*_{NoBP}(k))}{\gamma}.$$

Bride Price and Male Education

Prediction 5

Neither the probability that a boy is educated nor the responsiveness of male education to a reduction in the cost of schooling differs between bride price and non–bride price ethnic groups.

Prediction 5 - Model

- Probability that a son is educated:

$$\Pr(P_j = 1|k) = 1 - G\left(\frac{k}{\delta} - \frac{R^m}{1+r} - \frac{z_{01} - z_{00}}{1+r}\right)$$

- As long as education rates remain larger among boys than among girls in all groups:

$$\frac{\partial \Pr(P_J = 1|k)}{\partial k} = -\frac{1}{\delta}.g\left(\frac{k}{\delta} - \frac{R^m}{1+r} - \frac{z_{01} - z_{00}}{1+r}\right)$$

which indeed, does not depend on I_e.

Testing the Model - Prediction 1

- **Prediction 1:** *Is Matching Assortative by Education?*
  - There should be a positive relationship between the education of the wife and that of the husband
  - This relationship should be the same for bride price and non–bride price ethnic groups.
- They estimate the following equation:

$$\begin{aligned} I_{iedt}^{Husband\ Primary} =& \alpha_e + \alpha_d + \alpha_t + \beta_1 I_{iedt}^{Wife\ Primary} \\ & + \beta_2 I_{iedt}^{Wife\ Primary} \times I_e^{Bride\ Price} + X_{it}\Gamma + \varepsilon_{iedt}, \end{aligned}$$

where i indexes a marriage, e indexes the ethnicity of the bride, d districts, and t indexes the survey year.

Testing the Model - Prediction 1 Cont.

- Matching on education is positively assortative in all three samples. Strength of this relationship is not statistically different between bride price and non–bride price ethnic groups.

TABLE 3
DEGREE OF ASSORTATIVE MATCHING IN INDONESIA AND ZAMBIA (Dependent Variable: Indicator Variable for Husband Completed Primary)

	INDONESIA INTERCENSUS		IFLS		ZAMBIA POOLED DHS	
	(1)	(2)	(3)	(4)	(5)	(6)
$I_i^{Wife\ Primary}$	.466***	.460***	.445***	.440***	.534***	.510***
	(.005)	(.016)	(.021)	(.021)	(.016)	(.018)
	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
$I_i^{Wife\ Primary} \times I_e^{BridePrice}$	.022	.022	−.041	−.042	−.006	.004
	(.024)	(.020)	(.031)	(.030)	(.023)	(.019)
	[.424]	[.376]	[.204]	[.186]	[.870]	[.850]
Baseline covariates	Yes	Yes	Yes	Yes	Yes	Yes
Wife Muslim controls	No	Yes	No	Yes	No	Yes
Ethnicity interaction controls	No	Yes	No	Yes	No	Yes
Polygynous marriage controls	NA	NA	No	Yes	No	Yes
Mean of dependent variable	.653	.653	.655	.659	.565	.571
Standard deviation of dependent variable	.476	.476	.475	.474	.496	.495
Observations	107,338	107,338	4,847	4,785	22,793	18,574
Clusters	40	40	17	17	29	29
Adjusted R^2	.367	.367	.338	.336	.348	.336

Testing the Model - Prediction 2

- **Prediction 2:** *Do the Bride Price Amounts Increase with the Bride's Education?*
  - Whether the value of the bride price is increasing in the educational attainment of the bride.
- They estimate hedonic regressions where bride price payments are a function of the wife's educational attainment:

$$\begin{aligned}\ln(BP\ Amount)_{iet} =& \alpha_e + \alpha_t + \beta_1 I_{iet}^{Primary} + \beta_2 I_{iet}^{juniorSec} \\ &+ \beta_3 I_{iet}^{SeniorSec} + X_{it}\Gamma + \varepsilon_{iet},\end{aligned}$$

where i indexes a marriage, e indexes the ethnicity of the bride, and t indexes the survey year.

Testing the Model - Prediction 2 Cont.

- The amount of bride price paid at marriage is positively associated with the bride's educational attainment.

TABLE 4
DETERMINANTS OF BRIDE PRICE PAYMENT AMOUNTS (Dependent Variable: Log Bride Price Amount)

	INDONESIA (IFLS)					2SLS	ZAMBIA (ZFPS)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Wife's education:											
$I_i^{Primary}$	.615***	.579***	.366***	.373***	.285**	2.329**	.002	.023	−.015	−.008	−.027
	(.066)	(.071)	(.077)	(.078)	(.111)	(1.173)	(.137)	(.142)	(.141)	(.141)	(.141)
$I_i^{JuniorSecondary}$	.658***	.672***	.471***	.456***	.391***		.134	.258*	.255*	.279**	.289**
	(.066)	(.070)	(.074)	(.074)	(.097)		(.137)	(.137)	(.142)	(.142)	(.143)
$I_i^{SeniorSecondary}$	.865***	.857***	.468***	.457***	.306***		.384***	.391***	.398**	.396**	.358**
	(.077)	(.078)	(.089)	(.091)	(.115)		(.131)	(.147)	(.153)	(.153)	(.154)
Husband's education:											
$I_i^{H:Primary}$			.237***	.212**	.175				.230	.239	.270
			(.084)	(.084)	(.118)				(.197)	(.197)	(.196)
$I_i^{H:JuniorSecondary}$			.414***	.439***	.470***				.173	.176	.158
			(.077)	(.079)	(.103)				(.164)	(.164)	(.163)
$I_i^{H:SeniorSecondary}$			.532***	.540***	.427***				−.080	−.079	−.070
			(.090)	(.090)	(.114)				(.130)	(.130)	(.129)
F-test for first stage						3.04					
Baseline covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wife marriage age controls	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Husband marriage age controls	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Wife's premarital wealth	No	No	No	No	Yes	No	No	No	No	No	Yes
Wife Muslim	No	No	No	No	Yes	No	NA	NA	NA	NA	NA
Polygynous marriage	No	No	No	No	Yes	No	NA	NA	NA	NA	NA

Section 3

The Economics of Density: Evidence form the Berlin Wall

Gabriel M. Ahlfeldt, Stephen J. Redding Daniel M. Sturm, and Nikolas Wolf (2015)

Introduction

- Economic activity is highly unevenly distributed across space in specific locations within cities.
- Strength of agglomeration and dispersion forces is central to a range of economic and policy questions.

Motivation

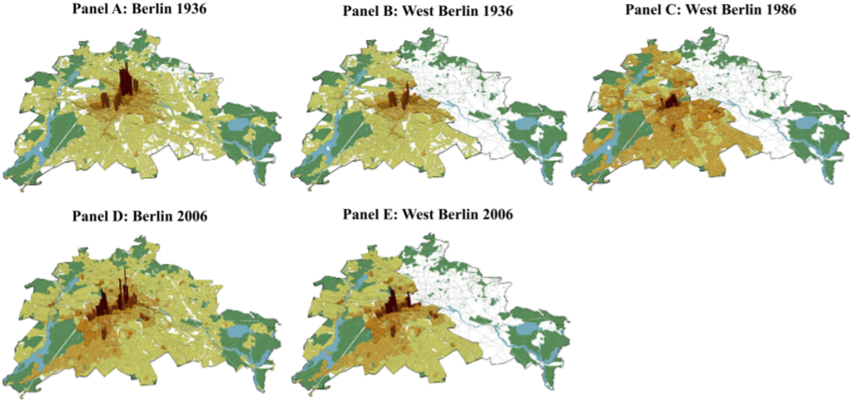


Figure 2: The evolution of land prices in Berlin over time

Contribution

- Distinguishing agglomeration and dispersion forces from variation in locational fundamentals.
  - High land prices and levels of economic activity are consistent with both agglomeration and attractive amenities.
- Exogenous source of variation:
  - Berlin's division in the aftermath of Second World War
  - Its unification following the fall of the Iron Curtain.
- Focus is on the determinants of economic activities within cities.

Findings

- Division leads to a reorientation of the gradient in land prices and employment in West Berlin away from the main pre-war concentration of economic activity in East Berlin.
- Reunification leads to a reemergence of this gradient.
- Elasticity of productivity with respect to the density of workplace employment is 0.07.

Identifying Assumption

- Changes in these structural residuals are uncorrelated with the exogenous change in the surrounding concentration of economic activity induced by Berlin's division and reunification.

Empirical Analysis

1. Theoretical Model
2. Reduced Form

City Structure

- Internal structure of the city is driven by a tension between:
  - Agglomeration forces in form of production and residential externalities.
  - Dispersion forces in form of commuting costs and inelastic supply of land.
- The city consists of discrete locations or blocks, each has an effective supply of floor space L_i, $i = 1, ...S$.
- endogenous fractions of floor space allocated to commercial and residential use are θ_i and $1 - \theta_i$, respectively.
- Blocks differ in terms of their final goods productivity, residential amenities, supply of floor space, and access to the transport network.

Workers

- An endogenous measure of H workers, perfectly mobile within the city and the larger economy, with reservation utility level of $\bar{U}$. Utility for worker o residing in block i and working in block j.

$$U_{ijo} = \frac{B_i z_{ijo}}{d_{ij}} \left(\frac{c_{ijo}}{\beta} \right)^{\beta} \left(\frac{l_{ijo}}{1-\beta} \right)^{1-\beta} \tag{1}$$

B_i: residential amenities

z_{ijo}: individual idiosyncratic shock (Frechet distribution)

d_{ij}: disutility from commuting

c_{ijo}: consumption of the single final good

l_{ijo}: residential floor space

Workers

- Indirect utility of workers

$$u_{ijo} = \frac{z_{ijo} B_i w_j Q_i^{\beta-1}}{d_{ij}} \tag{2}$$

Q_i: residential floor space

w_j: wage paid at workplace j

- Expected worker income conditional on living in block i:

$$E[w_i|i] = \sum_{j=1}^{S} \frac{E_j(w_j/dij)^{\epsilon}}{\sum_{s=1}^{S} E_s(w_s/d_{is})^{\epsilon}} w_j \tag{3}$$

E_j: The average utility derived from working in block j

- First term is the probability that a worker commutes to block j conditional on living in block i.

Production

- Firms produce a single final good, costlessly traded within the city and the larger economy, is chosen as the numeraire
- Perfect competition and constant returns to scale.
- Final good in block j is:

$$y_j = A_j H_{Mj}^{\alpha} L_{Mj}^{1-\alpha} \tag{4}$$

A_j: final good productivity H_{Mj}: measure of workers choosing to commute to block j L_{Mj}: measure of floor space used commercially

Production

- From profit maximization and zero profits, equilibrium commercial floor price:

$$q_j = (1-\alpha)\left(\frac{\alpha}{w_j}\right)^{\alpha/(1-\alpha)} A_j^{1/(1-\alpha)} \tag{5}$$

- Intuitively, firms in blocks with higher productivity (A_j) and/or lower wages (w_j) are able to pay higher commercial floor prices and still make zero profits.

Introducing Agglomeration Forces

- Production externalities impose structure on how the productivity of a given block is affected by the characteristics of other blocks.

$$A_j = a_j Y_j^{\lambda}, \qquad Y_j = \sum_{s=1}^{S} e^{-\delta \tau_{js}} \left(\frac{H_{Ms}}{K_s} \right). \tag{6}$$

  a_j: production fundamentals
  H_{Ms}/K_s: workplace employment density per unit of land area
  δ: determines their rate of spatial decay

- production externalities decline with travel time (τ_{js}) through the iceberg factor $e^{-\delta \tau_{js}}$.

Introducing Agglomeration Forces

- Same for residential externalities:

$$B_i = b_i \Omega_i^{\eta}, \qquad \Omega_i = \sum_{r=1}^{S} e^{-\rho \tau_{ir}} \left(\frac{H_{Rr}}{K_r} \right). \tag{7}$$

Berlin's Division and Reunification

- Focus is on West Berlin since it remained market-based economy.
- Division of Berlin
  - Infinite commuting costs ($\kappa \Rightarrow \infty$)
  - Infinite rate of decay of production externalities ($\delta \Rightarrow \infty$)
  - Infinite rate of decay of residential externalities ($\rho \Rightarrow \infty$)

  1. Loss of employment opportunities in East Berlin,
  2. Loss of commuters from East Berlin,
  3. Loss of production externalities from East Berlin,
  4. Loss of residential externalities from East Berlin.

Berlin's Division and Reunification

- Reduces the expected utility from living in West Berlin, and hence reduces its overall population, as workers out migrate to West Germany.
- these effects is stronger for parts of West Berlin close to employment and residential concentrations in East Berlin
- reducing floor prices, workplace employment, and residence employment in these parts of West Berlin relative to those elsewhere in West Berlin.

Berlin's Division and Reunification

- Reunification of Berlin
  - Need not necessarily have exactly the opposite effects
  - There can be multiple equilibria.
  - reintegration with employment and residential concentrations in East Berlin is predicted to raise relative floor prices, workplace employment, and residence employment in the areas of West Berlin close to those concentrations.

Data Discription

- Workplace employment
- Residence employment
- Price of floor space
- Commuting times between locations

Reduced Form Results

- Difference-in-difference estimation

$$\Delta \ln O_i = \alpha + \sum_{k=1}^{K} 1_{ik}\beta_k + \ln M_i \gamma + u_i \tag{8}$$

$\Delta \ln O_i$: log change in an economic outcome of interest (floor prices, workplace employment, residence employment)

1_{ik}: whether block i lies within a distance grid cell k from the pre-war CBD

β_k: coefficients to be estimated

M_i: time-invariant observable block characteristics (such as proximity to parks and lakes)

Reduced Form Results - Division

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

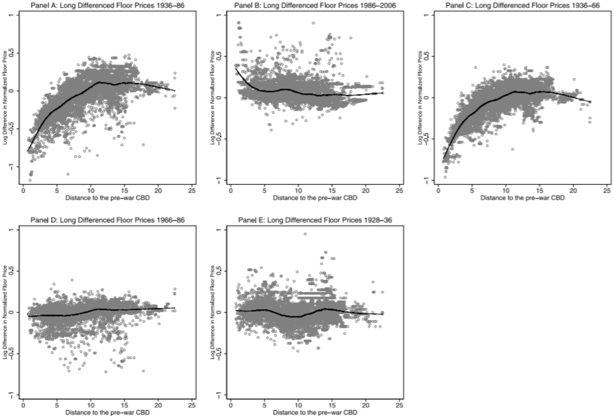
	(1) $\Delta \ln Q$	(2) $\Delta \ln Q$	(3) $\Delta \ln Q$	(4) $\Delta \ln Q$	(5) $\Delta \ln Q$	(6) $\Delta \ln$ EmpR	(7) $\Delta \ln$ EmpR	(8) $\Delta \ln$ EmpW	(9) $\Delta \ln$ EmpW
CBD 1	−0.800***	−0.567***	−0.524***	−0.503***	−0.565***	−1.332***	−0.975***	−0.691*	−0.639*
	(0.071)	(0.071)	(0.071)	(0.071)	(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

Reduced Form Results - Reunification

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

Robustness



Panel A: Long Differenced Floor Prices 1936–86
Panel B: Long Differenced Floor Prices 1986–2006
Panel C: Long Differenced Floor Prices 1936–66
Panel D: Long Differenced Floor Prices 1966–86
Panel E: Long Differenced Floor Prices 1928–36
Log Difference in Normalized Floor Price
Distance to the pre–war CBD

GMM estimates

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)

GMM estimates

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

Conclusion

- Separate out agglomeration and dispersion forces from heterogeneity in locational fundamentals by using Berlin's division and reunification as exogenous source of variation.
- A gravity equation for bilateral commuting flows, successful in accounting for observed commuting patterns.
- Estimate elasticity of productivity with respect to density.
- Importance of taking into account the rapid spatial decay of production/residential externalities.
- Useful quantitative framework for undertaking counterfactuals for changes in organization of economic activity in response to changes in the transport network.

Concerns

- Perfect mobility of workers within the larger economy.
- Both production and residential externalities are highly localized, with exponential rates of decay of $\delta = 0.36$ and $\rho = 0.91$.