ABSTRACT

Title of dissertation: ECONOMIC DEVELOPMENT OF URBAN AREAS

Brian Alexander Quistorff, Doctor of Philosophy, 2016

Dissertation directed by: Professor Sebastian Galiani

Department of Economics

In this dissertation I study the development of urban areas. At the aggregate level I investigate how they may be affected by climate change policies and by being designated the seat of governmental power. At the household level I study with coauthors how microfinance could improve the health of urban residents.

In Chapter 1, I investigate how local employment may be affected by electricity price increases, which is a likely consequence of climate change policies. I outline how previous studies that find large, negative effects may be biased. To overcome these biases I develop a novel estimation strategy that blends border-pair regressions with the synthetic control methodology. I show the conditions for consistent estimation. Using this estimator, I find no effect of contemporaneous price changes on employment. Consistent with the longer time-frame for manufacturing decisions, I do find evidence for negative effects from perceived permanent price shocks. These estimates are much smaller than previous research has found.

National capital cities are often substantially larger than other cities in their countries. In Chapter 2, I investigate whether there is a causal effect from being a

capital by studying the 1960 relocation of the Brazilian capital from Rio de Janeiro to Brasília. Using a synthetic controls strategy I find that losing the capital had no significant effects on Rio de Janeiro in terms of population, employment, or gross domestic product (GDP). I find that Brasília experienced large and significant increases in population, employment, and GDP. I find evidence of large spillovers from the public to the private sector.

Chapter 3 investigates how microfinance could increase the uptake of costly health goods. We study the effect of time payments (micro-loans or micro-savings) on willingness-to-pay (WTP) for a water filter among households in the slums of Dhaka, Bangladesh. We find that time payments significantly increase WTP: compared to a lump-sum up-front purchase, median WTP increases 83% with a six-month loan and 115% with a 12-month loan. We find that households are quite patient with respect to consumption of health inputs. We find evidence for the presence of credit and savings constraints.

ECONOMIC DEVELOPMENT OF URBAN AREAS

by

Brian Quistorff

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Sebastian Galiani, Chair
Professor Maureen Cropper
Professor Raymond Guiteras
Professor John Ham
Professor Kenneth L. Leonard

Foreword

The third chapter of this dissertation was coauthored with Raymond Guiteras at the University of Maryland, David I. Levine at the University of California, Berkeley, and Thomas Polley at Duke University. This chapter was supported financially by the Bill and Melinda Gates Foundation, 3ieimpact, and the Berkeley Initiative for Transparency in Social Sciences.

Dedication

To Callie, for making it all worthwhile.

Acknowledgments

I thank Sebastian Galiani for a being a wonderful adviser and helping me navigate both research and the economics profession. I thank John Ham, Maureen Cropper, Raymond Guiteras, Jessica Goldberg, and all participants at the University of Maryland workshops who have made helpful comments on all my research. I thank Kenneth Leonard for the help reviewing this dissertation and being the Dean's Representative. I thank Giordano Palloni for helping me survive graduate school and for all the research brainstorming. Thank you also to those in my cohort and those on the job market with me, for helping me graduate. I thank all my professors for the great education. Vickie, Terry, and the rest of the staff ensured that I graduated on time.

Chapter 1 benefited from helpful comments by Ben Zou, Paul Carrillo, Matt Notowodigdo, Daniel Bennett, Guido Kuersteiner, Julian Cristia, Diether Beuermann, Yichen (Christy) Zhou, and Magne Mogstad.

Chapter 2 benefited from helpful comments by Karthik Muralidhan, Maria Marta Ferreyra, and Chris Udry.

Chapter 3 benefited from helpful comments by Marcella Alsan, Ariel BenYishay, Glenn Harrison, John Rust, Steve Stern, Ken Train, Chris Udry, Song Yao and seminar participants at the Columbia University Sustainable Development Seminar, Cornell University, Georgetown University gui2de Development Economics Seminar, IFPRI, NEUDC 2015, NCSU, UC Berkeley Development Lunch, University of New South Wales, USC-CESR, WADES and Yale / Citi China India Insights Conference

Contents

List of Tables

List of Figures

Chapter 1:   The Effects of Energy Prices on Manufacturing Employment

1.1   Introduction

Understanding how local employment in electricity-intensive sectors, such as manufacturing, is affected by electricity prices is important for understanding both the potential effects of carbon policies on employment and the geographic concentration of employment. Existing best estimates of this elasticity may be biased by spill-overs between counties in the same labor markets and by selecting low-quality counterfactuals. I develop a new identification strategy to reduce these potential biases and provide new estimates.

Many national policies have differential local impacts due to variations in local context. The Clean Power Plan (CPP), the US's largest climate change policy (released on August 3rd, 2015)[1] is no exception, and will likely affect areas differently because it will affect electricity prices non-uniformly (EPA, 2014b) and because areas may differ in their responses to electricity price changes. Existing studies of similar proposed carbon policies have estimated a potential loss of 460,000 jobs (Deschenes, 2012). The magnitude of these estimate suggests that further understanding of potential effects is important for informed policy- and decision-making.

---

[1]A draft plan was announced on June 2, 2014. It was supplemented in October 28, 2014 to include

In addition to its importance in better understanding the implications of carbon policy on employment, the elasticity of employment with respect to prices is important for understanding the geographic concentration of employment. The geographic concentration of manufacturing responds to competition among many locations (Greenstone et al., 2010). This is especially true for manufacturing as output is not related to local demand. Given electric utilities and prices are highly regulated by the government, it would be useful for policy makers to know if they can use this regulatory authority to attract manufacturing jobs.

Existing best estimates of the local elasticity of manufacturing jobs with respect to electricity prices in the US are from Kahn and Mansur (2013). Their paper is broad in scope and investigates how locations can have comparative advantages for particular manufacturing industries by estimating the effects from electricity prices, environmental regulations, and labor laws. They attempt to remove confounding omitted variables by differencing neighboring counties with each other using a county border-pair design. They find significant negative elasticities between employment and electricity prices, which are quite large for some subsectors. While their design is an improvement upon state-level panel estimates, their estimates may be biased given that neighboring counties may not be independent from each other and may not be similar enough in terms of unobservables to serve as adequate counterfactuals for each other.

To address these limitations, I develop a new identification strategy that differences away omitted variables by selecting weighted averages of non-neighboring

Indian country and multi-jurisdiction partnerships.

counties so that the counterfactuals are observably similar in previous outcomes and covariates. This can improve the match quality of the counterfactual and decreases the chance of interference between counties. Using this new method I find contemporaneous employment elasticities much closer to zero. These small elasticities could be due to either the basic technology of substitution between factors or due to sluggish responses. Slow or delayed adjustment can be caused by costly adjustment of factors or expectations of future prices. Addressing this latter reason, I investigate whether firms are responsive to perceived permanent changes to the electricity price level. I find significant negative results, but much smaller estimates than previously found in the literature.

This paper proceeds as follows. Section 1.2 reviews the literature on this elasticity as well as related identification strategies. Section 1.3 covers the new identification strategy and how it is estimated. Section 1.4 covers the results. Section 1.5 discusses implications of the estimates and other uses of the identification strategy. Section 1.6 concludes.

1.2   Literature Review

In this section I review the literature on both the elasticity of employment with respect to electricity prices and on similar identification strategies.

To define the elasticity of interest, suppose the manufacturing production function is $Y = F(K, L, E)$ where K is capital, L is labor, E is electricity, and F is linear homogeneous. The cross-partial elasticity of labor demand with respect to electricity

3

(p_E) holding output and other input prices constant is $\eta_{LE} = \partial \ln L / \partial \ln p_E$. One can show that

$$\eta_{LE} = s_E \sigma_{LE}$$

where s_E is the cost-share of energy and $\sigma_{LE} = \frac{\partial \ln(E/L)}{\partial \ln(p_L/p_E)}$ is the partial elasticity of substitution holding constant output and other input prices. With more than two factors however, the signs can not be determined from theory.[2] Two factors, such as labor and energy, are (p-)substitutes if $\eta_{LE} > 0$ and compliments if $\eta_{LE} < 0$.

There have been many estimates of this elasticity using manufacturing employment given it is a sector that is likely to be affected by electricity prices due to its high usage and because it is also a crucial intermediary sector (e.g., Linn (2009) finds that manufacturing linkages amplify macro shocks). Earlier estimates of the elasticity were based on large sector aggregates or the full economy where the assumption of long-run equilibrium and constant output are plausible. Hamermesh (1993) reviews these estimates, noting that they find that energy price increases weakly increased employment with the elasticity being positive and small (less than 0.2).

If output is not held constant then there is a scale effect so that

$$\eta'_{LE} = \eta_{LE} - s_E \eta$$

---

[2]Since $\eta_{ii} < 0$, and by homogeneity $\sum_j \eta_{ij} = 0$ (since factor demands in all factor prices are homogeneous of degree zero), then at least one cross-partial elasticity is positive.

where η is the product demand elasticity. This implies that $\eta'_{LE} < \eta_{LE}$ with $\eta'_{LE} < 0$ possible even if $\eta_{LE} > 0$. In a setting dealing with smaller geographic areas, narrower sectors, or more short-run time-horizons, then relaxing the assumption of constant output is more justified. The difference between η_{LE} and η'_{LE}, however, should not be very large given that $s_E \approx .05$ and $\eta \approx 2$ (Aigner and Chu, 1968). More recent studies investigate η'_{LE} directly in the United State and Europe. In the US, Deschenes (2012) uses a state panel from 1976-2008 and finds significant negative employment effects with elasticities of around -0.13. Kahn and Mansur (2013) (henceforth KM) use a border-pair design for Metropolitan Statistical Area (MSA) counties from 1998 to 2009, separating employment into manufacturing sub-sectors and finds that the elasticities range from 0.17 to -1.65 and are significantly correlated with sub-sectoral energy intensity. In the EU there has been mixed evidence for negative employment effects using variation from the EU Emissions Trading Scheme comparing size-eligible vs size-ineligible firms (Chan et al., 2013; Wagner et al., 2014; Abrell et al., 2011) and from tax and electricity rate differences (Flues and Lutz, 2014; Martin et al., 2011; Cox et al., 2013).

Studies of related effects have tended to find negative results. With regards to environmental regulation, Greenstone (2002) finds negative employment effects of the Clean Air Act Amendments (CAAAs). Walker (2013), also investigating the CAAAs, estimates significant earnings losses for workers in regulated sectors that transition to other sectors. The macroeconomics literature has also investigated aggregate-level effects of energy prices. The majority of the work utilizes oil-price shocks: see Davis and Haltiwanger (2001) and the reviews of Hamilton (2008) and Kilian (2008). The

studies emphasize that a large component of the estimated effect is through changes in consumer demand due to gas price changes and therefore their estimates are not from electricity price shocks. One exception is Aldy and Pizer (2012) which estimate, at a national-level, negative output effects due to international competitiveness from changes in energy prices.

This paper advances a novel estimation strategy to limit bias from spatial dependence. Two sources of such bias that I consider are: unobservable unit characteristics causing omitted variable bias and that units may dynamically interact with each other given they are spatially connected (for example, a change to the employment level of one county may affect the level in a neighboring county). For unobservables characteristics, a standard framework is to posit that the error term is decomposed as $\mu_c\lambda_t + e_{ct}$ where there is a vector of factors λ_t each period and μ_c is a vector of unobserved unit (county) characteristics sometimes called factor loadings. Correlation between regressors and $\mu_c\lambda_t$ is allowed and units are correlated cross-sectionally through λ_t. Ahn et al. (2001) provide a GMM estimator that is consistent given a factor model but assume that regressors are i.i.d. across units, which may not be valid in practice. Pesaran (2006) augments an estimated system of equations with generated regressors which are cross-sectional averages of the dependent and independent variables. Bai (2009) uses an iterative process to step repeatedly between identifying the unobserved factors using principal component methods and estimating the main coefficients.

Finally, Abadie and Gardeazabal (2003) and Abadie et al. (2010) provide a matching strategy that removes unobserved factors when estimating the effect of a binary treatment on a single unit in a case study methodology. This is generalized to

multiple treated units and multiple treatment periods by Cavallo et al. (2013). Zou (2015) uses the matching procedure to eliminate unobserved factors from a group of treated units using non-treated units and then estimates a first-differences cross-sectional regression. I generalize the procedure to a panel setting and to the common case where all units can experience changes in their covariates. Additionally, I modify the matching strategy to disallow neighboring units from being compared to each other to reduce the bias from local interactions.[3] Very importantly, I derive the asymptotic properties of the estimator proposed in this paper.

1.3   Empirical Strategy

Estimating the local effect of price changes using a state-level panel would be biased by time-changing state policies affecting both employment and energy prices. The state of the art method, therefore, is to employ a county border-pair design as KM do. This estimation has a pair of observations for every pair of counties that border each other, so that counties with multiple neighbors appear multiple times. The estimation then conditions on pair fixed effects so that the relationship is estimated from differences across the pair.

While this type of design has advantages, there are potentially a few concerns. First, papers that utilize this design may have their comparisons biased by spillovers or other interactions in prices or employment between nearby units. If neighboring

---

[3]If the only identification challenge was spatial spillovers, and a plausible distance metric was available, then on could use methods from the spatial econometrics literature (for example, reviewed in Anselin 2003) that explicitly models either the error structure ($Cov[\varepsilon_i, \varepsilon_j] = \sigma^2 f(d_{ij}, \rho)$) or additional regressors in the form of "spatially-lagged" dependent variables ($\rho \sum_j w(d_{ij}) y_j$).

counties share the same labor market, a shock to electricity prices in County A could cause an employment change that comes at the expense of a neighboring County B given that workers can commute between the two. This would cause the local effect to be biased away from zero. This is a concern given that, in the US the median share of workers from a metro county working in another county in the same metro area is 34.0% (American Community Survey 2009-2013). Similarly, if there are interactions between counties with regards to electricity prices, estimates could also be biased. If in the previous example, County B changes prices in a similar way as County A (either strategically or because they have differing responses to common causes) then the pair difference in prices is now smaller, resulting in increased bias. The bias can increase with the inclusion of the pair fixed-effect as units become compared not to the whole sample but to the unit with which they interact.

Another problem with the border-pairs approach is that it may exacerbate bias if the match is not adequate. Using the factor model, if all units in the same MSA m have the same unobserved characteristics ($\mu_c = \mu_m \forall c \in m$), then differencing any pair will eliminate bias from the factors. If counties differ in their characteristics ($\mu_c = \mu_m + \tilde{\mu}_c$) then this differencing can increase rather than decrease bias. Suppose that prices are determined by the omitted variable and an unobserved exogenous variable, such as $p_{ct} = \alpha_1(\mu_c \lambda_t) + \alpha_2 x_{ct}$, and that the exogenous component also has a regional and individual component ($x_{ct} = x_{mt} + \tilde{x}_{ct}$). Griliches (1979) shows that bias is reduced by differencing only if the regional component accounts for a larger fraction of the variance of $\mu_c \lambda_t$ than of x_{ct}, an assumption that may not hold in this

---

[4]This is a smaller concern with my alternative estimation strategy as I match to a counterfactual

or other cross-county studies.[4] Neumark (1999) also notes that instrumental variable (IV) estimation can exacerbate this bias. Note that in a border-pair design, the match quality does not increase with the addition of more data.[5]

The unobservables that are of concern are labor-market characteristics (e.g., unemployment trends), regulation environment (e.g., how pro-business the government is), and market access. As these are likely to be cross-sectionally correlated they can be well accommodated by a model of unobserved factors. The matching procedure of the synthetic controls methodology (SCM) of Abadie and Gardeazabal (2003) and Abadie et al. (2010) provides a way of reducing the bias from unobserved factors while determining comparison groups in a way that may avoid bias from spillovers. I develop a generalization of the SCM that is consistent in a panel setting.

To explain how this method attempts to control for unobservables, I start with a data generating process (DGP) for county employment, which for now I've assumed has only one sector of manufacturing,

$$l_{ct} = \beta x_{ct} + \lambda_t \mu_c + \varepsilon_{ct} \quad \forall c, t \tag{1.1}$$

where the vector x_{ct} (including prices and other controls) are known, λ_t is a $(1 \times F)$ vector of unobserved common factors, and μ_c is a $(F \times 1)$ vector of factor loadings ($\lambda_t \mu_c$ are known as "interactive fixed-effects"). I assume ε_{ct} are independent $\forall (c, t)$,

---

that should be very close in terms of μ_c.

[5]Alternative models of course exist where border-pair designs eliminate bias but my estimation strategy does not. For instance, there could be a single MSA-level confounder u_{mt} that impacts the counties in an MSA equally with no other interactions between units. My estimation strategy will recover common factors to the extent possible given the cross-sectional correlation of u_{mt}. In this case the bias in synthetic regression reduces to zero as the cross-sectional correlation increases.

the ε_{ct} are mean independent of $\{\mu_c\}$, that $\mathrm{E}[\varepsilon_{ct}] = 0$, and that the x_{ct} are strictly exogenous given the interactive fixed-effects ($x_{ct} \perp \varepsilon_{ct} | \lambda_t \mu_c$). x_{ct} may be correlated with the residual through the interactive fixed-effects.[6]

Interactive fixed effects generalize the controls of a standard panel model. They can reproduce the panel fixed effects by having $\mu_c = (1, \alpha_c)$ and $\lambda_t = (\xi_t, 1)'$. They also can accommodate autoregressive components: $\mu_c = (l_{c,0})$ and $\lambda_t = (\rho^t)$[7]; and unit-specific time trends: $\mu_c = (\alpha_c)$ and $\lambda_t = (t)$. This model allows for violations of the common-trends assumption of difference-in-difference equations.

1.3.1 Synthetic Regression

Equation 1.1 is a similar DGP[8] as that used in the classic SCM. The classic SCM treats x_{ct} as binary, exogenous (conditional on observables and the interactive fixed effects), and that only changes in one period. I extend that methodology to a synthetic regression where x_{ct} may be continuous and change in multiple periods.

1.3.1.1 Estimation

Estimation proceeds in two steps, a match step (similar to SCM) and a regression step. For each observation (county c in year t), I conduct a match against all other counties so that the weighted average is observably similar on data prior to t. The resulting weight vector $\boldsymbol{w}_{ct}$ is constrained to be non-negative and sum to one. This

---

[6]The classic SCM requires that donors not be affected by the treatment variable. This is equivalent to saying that the ε_{it} are independent and that $x_{ct} \perp \varepsilon_{ct}$ conditional on the interactive fixed effects.

[7]Since $l_{ct} = \rho l_{c,t-1} + \varepsilon_{ct}$ can be re-written as $l_{ct} = \rho^t l_{c,0} + \nu_{ct}$ where ν_{ct} is AR(1).

[8]For ease of exposition, the estimate of interest is not time-varying though this is relaxed later.

weighted average serves as an observation's counterfactual. It has a complete history, which is similar on observables to the evaluation county c prior to the evaluation year t but may differ from t onward. In effect, the counterfactual is estimating what one would expect to happen to a county if one only knew its history. Let $\overleftarrow{l}_{ct}$ and $\overleftarrow{\boldsymbol{x}}_{ct}$ be the vectors of the dependent and independent variables for county c prior to year t. Let $\overleftarrow{l}_t$ and $\overleftarrow{\boldsymbol{x}}_t$ be the $N \times (t-1)$ matrices that contain the prior data for all counties. Assume one is able to match such that pre-evaluation characteristics match:

$$\boldsymbol{w}'_{ct}\overleftarrow{l}_t = \overleftarrow{l}_{ct} \quad \& \quad \boldsymbol{w}'_{ct}\overleftarrow{\boldsymbol{x}}_t = \overleftarrow{\boldsymbol{x}}_{ct} \quad \forall c, t. \tag{1.2}$$

In cases where multiple weight vectors can match the histories exactly, I resolve the indeterminacy by choosing the vector that distributes the weight the most (by minimizing the Euclidean distance $\|\boldsymbol{w}_{ct}\|$). This is a change to the existing SCM matching procedure that makes the process smooth in large samples.[9]

Then I construct for every observation (c, t), the contemporary difference between it and its counterfactual, $\Delta l_{ct} = l_{ct} - \boldsymbol{w}'_{ct}l_t$ and $\Delta x_{ct} = x_{ct} - \boldsymbol{w}'_{ct}\boldsymbol{x}_t$. A regression using the observable differences is then

$$\Delta l_{ct} = \beta \Delta x_{ct} + \epsilon_{ct} \quad \forall c, t$$

---

[9]When a complete match is not possible, I follow the SCM and (a) determine a set of variable weights $\boldsymbol{V}$ from regressions of the outcomes on predictors variables (Z), then (b) $\boldsymbol{w}_{ct} = \arg\min_{\boldsymbol{w}} \|Z_{ct} - \boldsymbol{w}'Z_t\|_{\boldsymbol{V}}$.

One can show that $\epsilon_{ct} = \Delta\varepsilon_{ct} - \boldsymbol{\Lambda}_t\Delta\overleftarrow{\bar{\varepsilon}}_{ct}$ with $\boldsymbol{\Lambda}_t = \lambda_t(\overleftarrow{\bar{\boldsymbol{\lambda}}}'_t\overleftarrow{\bar{\boldsymbol{\lambda}}}_t)^{-1}\overleftarrow{\bar{\boldsymbol{\lambda}}}'_t$. I assume that $\overleftarrow{\bar{\boldsymbol{\lambda}}}'_t\overleftarrow{\bar{\boldsymbol{\lambda}}}_t$

is invertible[10]. The composite error has two parts, (a) $\Delta\varepsilon_{ct}$ shows that the synthetic

counterfactual may differ from the evaluation county in the current year even given a

perfect match, and (b) $\boldsymbol{\Lambda}_t\Delta\overleftarrow{\bar{\varepsilon}}_{ct}$ is from incorrectly matching due to historical errors. In

Appendix 1.7, I show that the estimator of β is consistent and asymptotically normal

for fixed T as $N \to \infty$.[11]

In finite sample, not all counties will be able to be matched exactly for pre-

evaluation data. Similar to the classic SC methodology (and analogously to how GLS

optimally downweights observations by their variance) I downweight observations

by their pre-evaluation mean-squared-prediction-error (MSPE). The match quality

increases with the length of historical data so it is standard to set aside an initial T_0

years to be used exclusively for matching.

To see the intuition for the extension to a panel setting, note that if one matched

on the entire history of the independent and dependent variables then there would be

no difference to use in estimation. Additionally, a counterfactual becomes less accurate

the farther into the future it is projected. For these two reasons, in a panel setting,

multiple matches are made for every county, each time with progressively more recent

data used. Changes in the year after matching are what is used in the regression.

In the match step, one can place extra constraints on the weight vector. In

my setting where I am concerned about spill-overs between counties, I constrain the

weight vector to have zero elements for all counties in the same MSA. Without this,

---

[10]This imposes $F \leq t - 1$ by the Cauchy-Binet formula.

[11]It is assumed that β is constant (homogeneous treatment effects). The case of non-constant β (heterogeneous treatment effects) remains to be explored.

they may receive high weights given they are likely similar on observables. Spillovers between MSAs are likely to be much smaller. Compared to the high number of workers that work elsewhere in the same MSA, the median share of workers from a county that work in any other MSA is only 4.0%.

As an example of the process for a county, assume that counties A and B are in an MSA. Examining county A in year t, a weighted average of other counties A' is found (from those not including B) such that A and A' are similar on observables from the initial year to $t-1$. Intuitively, given that the match is on both dependent and independent variables it will end being very similar in terms of the unobservables which will then match μ_A. For year t, a shock may occur for county A and the constituent counties of A'. Given however that A' is likely composed of many counties and I assume that their shocks are independent, then the weighted average of those shocks will be small.

Similar to the classic SCM model, one can estimate delayed effects of current shocks. This would correspond to a DGP with L extra lags, such as,

$$l_{ct} = \sum_{l=0}^{L} \beta_{-l} x_{c,t-l} + \lambda_t \mu_c + \varepsilon_{ct} \quad \forall c, t$$

It can be estimated from a model matched at time $t - L$, and using the estimating equation

$$\Delta l_{ct+L} = \sum_{l=0}^{L} \beta_l \Delta x_{c,t+l} + \epsilon_{ct} \quad \forall c, t$$

so that shocks to covariates in year t can be related to outcomes in year $t + l$.

One advantage of generalizing the exogenous variable x beyond a binary treatment that it was in SCM is that other time-changing regressors can be included. This can improve identification and reduce variation. For example, classic SCM identifies an effect that differentially causes a change in one unit at a particular time. The modeler makes the case that she knows what policy or change happened at this time to attribute the effect to. In reality many variables are changing and it is impossible to distinguish them in SCM. In this synthetic regression one can control for other multiple time-changing variables.

1.3.1.2 Inference

I use the wild bootstrap for inference as it has been shown to be valid under similar conditions. First, it is used with factor-augmented regressions (Gonçalves and Perron, 2014). Second, though asymptotically smooth, the synthetic regression estimator is non-smooth in finite sample (see Appendix 1.7.1). For a similar non-smoothness in k-nearest neighbors matching (which does not become smooth asymptotically) the naive bootstrap is inconsistent (Abadie and Imbens, 2008) whereas the wild bootstrap is consistent (Otsu and Rai, 2015). I also provide Monte-Carlo evidence that this produces valid inference in Section 1.8.

The wild bootstrap takes as given the initial estimations including matches. Specifically, given an initial estimates

$$\Delta l_{ct} = \hat{\beta}\Delta x_{ct} + \hat{\epsilon}_{ct} \quad \forall c, t$$

It then repeatedly modifies the sample by transforming the final residuals and then re-estimating the model. Specifically, it conducts B repetitions to construct the distribution $\{\hat{\beta}^b\}$, where in repetition b, the procedure:

1. Draws $\{v_{ct}^b\}$ for all (c, t) from a resampling distribution. Following Otsu and Rai (2015), I use the Mammen (1993) distribution:

$$v_{ct}^b = \begin{cases} \frac{1-\sqrt{5}}{2} & \text{with probability } \frac{1+\sqrt{5}}{2\sqrt{5}} \\ \\ 1 - \frac{1-\sqrt{5}}{2} & \text{otherwise} \end{cases}$$

2. Constructs $\{\Delta y_{ct}^b\}$ for all (c, t) from $\Delta y_{ct}^b = \hat{\beta}\Delta x_{ct} + v_{ct}^b\hat{\epsilon}_{ct}$.

3. Regresses $\Delta \boldsymbol{y}^b$ on $\Delta \boldsymbol{x}$ yielding estimated $\hat{\beta}^b$.

With the distribution $\{\hat{\beta}^b\}$, I conduct inference.

1.3.2 Estimation

Following KM, I divide manufacturing into 21 subsectors and let the elasticity vary by the electricity intensity of the subsector. I also condition on ozone pollution and environmental regulation measures as they could be correlated with electricity prices

and employment. Ozone is a by product of burning fossil fuels, and if certain indicators of it become too high then the EPA deems the county as ozone "non-attainment" for that year. All ozone emitters in such a county are subject to extra regulatory oversight. This regulation may affect both electricity prices and employment and is therefore included as a control.

Matching variables are picked to expand upon those from KM. For time-changing predictors, I include historical trends since 1990[12] in sectoral employment, energy prices, and EPA county nonattainment status. Fixed characteristics are also included, such as 1970 population, geographic size, distance to central business district, and 1990 house value. To remove the possibility of local spillovers and interference I disallow counties from the same MSA to be in each other's synthetic counterfactual. Matches are also only made within the same subsector.

I then construct the observation-counterfactual differences and estimate an equation similar to KM for 1998-2009:

$$
\begin{aligned}
\Delta l_{ctk} =& \beta_1 \Delta p_{ct} + \beta_2 \Delta p_{ct} \cdot ElecIndex_{kt} + \beta_3 ElecIndex_{kt} + \\
& \beta_4 \Delta Nonattain_{ct} + \beta_5 \Delta Nonattain_{ct} \cdot PollIndex_k + \beta_6 PollIndex_k + \\
& f(\Delta Poll_{ct}) + \varepsilon_{ctk}
\end{aligned}
\tag{1.3}
$$

where l_{ctk} is employment for county c in year t in subsector k, p is log electricity prices, $ElecIndex_{kt}$ is an index of sectoral electricity intensity constructed as electricity costs

---

[12]Following Dube and Zipperer (2015) who determine optimal SCM predictor sets for using cross-

as a proportion of sales, $Nonattain_{ct}$ is whether the EPA designated this county as a Nonattainment county for ozone pollution, $PollIndex_k$ is an index of sectoral ozone pollution intensity constructed as total emissions over value added, $f(Poll_{ct})$ is a cubic polynomial of average ozone concentrations to account for potential regulatory differences away from the non-attainment cutoff.

As a robustness exercise I instrument for the possibility of measurement error or reverse causality, which could be due to demand-side changes or bargaining power of manufacturing firms. Similar to Kahn and Mansur (2013), I instrument p_{ct} with a shift-share IV[13] in the regression step. Each county's electricity fuel type (coal, oil, gas) capacity shares in 1995 are interacted that with the time-series of national prices of those three fuels. This is a valid instrument assuming that initial characteristics are not associated with future unobservables. As it is likely that location characteristics (μ_c) affect both, then it is essential to control for, as I do, the unobserved factors before using the instrument.

1.3.3 Data

My measure of electricity price is the Energy Information Administration data on average yearly utility rates from Form EIA-861 for 1990-2009. This data also includes the counties of operation for each utility. When multiple utilities operate in the same county I compute as a price the average weighted by total utility size.

---

validation, I include as match variables every other value of time-changing variables. This type of cross-validation is not possible in my setting as all counties can experience changes in electricity prices.

[13]This is also known as a Bartik instrument for Bartik (1991)

Manufacturing plants that receive electricity from a source other than their local utility will not be captured in this measure of prices. These establishments will include those that have their own generation or that purchase electricity independently and have it transmitted through their local utility. This latter is more likely for large establishments in states that have a large percentage of independent power producers. Additionally, this price measure is an average and actual rate scheme may be more complicated including different rates for peak-time consumption.

Retail utility electricity prices are usually set by public-utility commissions (PUC), often at the state-level. They approve rates based on average cost pricing.[14] Roughly, this equates to total expenses (e.g., fuel and energy purchases) plus an allowed rate of return on existing assets. PUCs approve major infrastructure changes, large changes to sales areas, mergers and splits, and influence the rates of return. For these reasons it is possible that municipal or state regulatory bodies may take into account, directly or indirectly, local employment when deciding on rates, which is why controlling for unobserved characteristics is important. Negotiations with PUCs can take years so rate plans are often in place for several years. Differences between a county and its synthetic counterfactual in terms of prices can be thought of then as forecast errors that arise from the multi-year plans.

For industrial employment data I use the US Census County Business Patterns (CBP) data from 1990-2009 which has yearly, county employment (total mid-March

---

[14]Rates may also be based on incentivizing other goals, such as energy efficiency, reliability, or transmission capacity. Nuclear plants and earlier contracts made because of the Public Utility Regulatory Policy Act (PURPA) of 1978 caused much of the regional variation in prices earlier in the sample period. Regions that restructured their utilities went through a phase where utilities could recover "stranded" costs and now prices are mostly set by by regional whole-sale costs. See Joskow et al. (1989) and Joskow (2006) for further details.

employees) by detailed industry from firms. Following KM I aggregate the CBP to the three-digit North American Industry Classification System (NAICS) level leaving 21 industry categories. Prior to 1998, the CBP used Standard Industrial Classification (SIC) codes and so I use the National Bureau of Economic Research (NBER) SIC-NAICS concordance to translate older categories. Though there may be some measurement error in this conversion, these data are only used for matching purposes. For privacy considerations many observations with few establishments have their employment counts suppressed in the CBP. Data on the number of establishments in various size classes is always available. Following KM, I therefore impute the suppressed values using size-class midpoints.[15] The imputed observations have mostly lower levels of employment. The median censored observation has 59 imputed employees while the median uncensored observation has 1708 employees.

The NBER-CES Manufacturing Industry Database provides data on factors, including energy and expenditures. It also includes data on value added and sales. The index of electricity intensity is standardized to a zero to one scale.

The EPA provides sectoral emission levels from the 2002 National Emissions Inventory database.[16] These are standardized to a zero to one scale. Data on yearly county nonattainment status is from the EPA's Green Book. The nonattainment standard is measure of the number of days that have ozone concentrations peak above

---

[15]Size class are: 1-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000-1499, 1500-2499, 2500-4999, and 5000 or more. For imputation purposes, those in the 5000 or more category are classified as having 6000 employees.

[16]For ozone, I aggregate tons of nitrogen oxides and volatile organic compounds.

Table 1.1: Summary Statistics

	Mean	Std. dev.
Manufacturing employees	534.0	1996.2
Has manufacturing	0.777	0.416
Electricity price (USD/kWh)	0.0724	0.0202
Ozone monitor in county	0.480	0.500
Ozone (ppm, mean of 8hr)	0.0221	0.0234
Ozone nonattainment (8hr)	0.146	0.353

Notes: Observations are by county (1158 counties in 380 MSAs), year (1998-2009), and 3-digit NAICS sector (21 total). 270444 total observations.

a threshold.[17] Average ozone levels are from the EPA Aerometric Information Retrieval System (AIRS) database.

Summary statistics for the main variables are shown in Table 1.1.

1.4   Results

I begin with re-estimating the main models of Kahn and Mansur (2013) as they are the current state of the art. I then show how my estimation strategy yields different estimates for the same contemporaneous price channel, but slightly stronger effects for more permanent perceived price increases.

---

[17]From 1979-1997, county ozone nonattainment was defined as at least two days per calendar year with maximum hourly average concentration greater than 0.12 ppm. For 1998-2008, county nonattainment was when the annual fourth-highest daily maximum 8-hr concentration, averaged over 3 years was over 0.08 ppm. In 2008 this threshold was changed to 0.075 ppm.

1.4.1 Border-Pair Results

I first present re-estimations of the KM main findings in Table 1.2. The estimation equation is

$$
\begin{aligned}
l_{c,j(c),tk} =& \beta_1 p_{ct} + \beta_2 p_{ct} \cdot ElecIndex_{kt} + \beta_3 ElecIndex_{kt}+ \\
& \beta_4 Nonattain_{ct} + \beta_5 Nonattain_{ct} \cdot PollIndex_k + \beta_7 PollIndex_k+ \\
& \beta_8 Rights_{s(c)} + \beta_9 Rights_{s(c)} \cdot LabCapRatio_{kt} + \beta_{10} LabCapRatio_{kt}+ \\
& \beta_{11} NoMon_c + \beta_{12} NoMon_c \cdot PollIndex_k+ \\
& \delta Z_c + f(Pollution_{ct}) + \phi_{j(c)} + \phi_{s(c),t} + \phi_{kt} + \varepsilon_{c,j(c),tk}
\end{aligned}
\tag{1.4}
$$

where $j(c)$ indexes the county pairs that contain MSA county c, t indexes years (1998-2009), k is the subsector, $Rights_{s(c)}$ denotes whether the state that c is located in is a "Right to Work" state meaning that it prohibits company-union contracts that bar non-union employees, $LabCapRatio_{kt}$ is sectoral labor to capital ratio (in hours worked per value of capital stock), $NoMon_c$ indicates whether the county lacks an ozone monitoring station, ϕ are various fixed effects, and Z_c are county fixed characteristics:population in 1970, miles to central business district, land area, and 1990 housing values.

The main finding of KM is that energy-intensive subsectors experience large employment changes due price change[18]. Their estimated elasticities for sectors with the highest electricity intensity are around -1.6. This is evident in my re-estimation

---

[18]Their results for low energy-intensive sectors are small effect, sometimes negative but often

Table 1.2: Re-estimation of Kahn & Mansur Findings

	(1) N	(2) N	(3) ln N
ln Electricity price	-41.60	-20.62	-0.164
	(88.14)	(90.07)	(0.167)
ln Price * electricity index	-1140.7**	-1611.7**	-1.290***
	(544.2)	(665.3)	(0.455)
Right to work * labor/capital	9245.6***		8.896***
	(2321.3)		(2.789)
Nonattainment county	-5.015		0.0871**
	(24.14)		(0.0428)
Nonattainment * pollution index	-421.0***		-0.0700
	(123.0)		(0.115)
No pollution monitor	2066.0		5.512**
	(2290.3)		(2.562)
No monitor * pollution index	698.6***		0.313***
	(128.7)		(0.0964)
R^2	0.368	0.365	0.467
Observations	969168	969168	783184

Notes: Outcome variable is yearly county employment in manufacturing subsectors (mean=534). Sample is MSA counties from 1998-2009. All regressions include population in 1970, miles to central business district, land area, 1990 housing values, cubic polynomials of ozone concentrations, county-pair, industry-year, and state-year fixed effects. Standard errors clustered at the utility level. * $p < .1$, ** $p < .05$, *** $p < .01$

as well as can be seen in the coefficient on energy prices interacted with the electricity intensity index. The finding is quite robust across the specifications.

There are several differences between my study and that of KM. Equation 1.4 has two main differences from Equation 1.3. First, given its scope is about broader geographic comparative advantage, it includes variables related to union power and presence of pollution monitoring stations. Second, its level of fixed effects is more coarse. Other county border-pair designs, such as Dube et al. (2010), include multiple fixed effects for each pair (e.g., $\phi_c + \phi_{j(c),t}$) which algebraically leads to a difference equation similar to Equation 1.3. Additionally, weighting used by KM is different than some other county border-pairs designs, though the mapping to my design is not exact. KM downweight each observation from a county by the number of pairs that include the county (others weight each border the same). In Table 1.3, I re-estimate the model, changing these characteristics to show that the main result holds. Column 1 is repeated from Table 1.2. The rest of the columns omit variables dealing with union power or presence of pollution monitoring stations. Columns 3 and 4 use the more fine-grained fixed effects. Column 4 weights each border, rather than each county, the same. Across the entire table the coefficient on the interaction of energy prices and sectoral energy intensity is negative, large, and significant. In some of the specifications, the coefficient on log electricity prices is positive indicating that in sectors with the lowest electricity intensity the cross-price elasticity is positive.

positive.

Table 1.3: Border-Pair Sensitivity

	(1)	(2)	(3)	(4)
	N	N	N	N
ln Electricity price	-41.60	-41.58	183.7*	274.5**
	(88.14)	(87.56)	(98.17)	(137.5)
ln Price * electricity index	-1140.7**	-1266.0**	-1041.5**	-1249.9**
	(544.2)	(582.6)	(509.3)	(545.4)
Right to work * labor/capital	9245.6***			
	(2321.3)			
Nonattainment county	-5.015	34.42	20.52	-2.077
	(24.14)	(27.47)	(29.81)	(31.29)
Nonattainment * pollution index	-421.0***	-647.6***	-446.4***	-383.6***
	(123.0)	(163.5)	(123.7)	(115.3)
No pollution monitor	2066.0			
	(2290.3)			
No monitor * pollution index	698.6***			
	(128.7)			
R^2	0.368	0.366	0.360	0.353
FE type	coarse	coarse	fine	fine
Weighting	border	county	county	border
Observations	969168	969168	969168	969168

Notes: Outcome variable is yearly county employment in manufacturing subsectors (mean=534). Sample is MSA counties from 1998-2009. All regressions include cubic polynomials of ozone concentration. Regressions without county FEs also include population in 1970, miles to central business district, land area and 1990 housing values. Standard errors clustered at the utility level. Weight-type county=counties have same total weight. Weight-type border=county comparisons (borders) have same weight. FE-type coarse=county-pair industry#year state#year, fine=county county-pair#year. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 1.4: Base Synthetic Regression Estimation

	(1) L_t	(2) L_{t+1}	(3) L_{t+2}	(4) L_{t+3}
LnEnergy Price	-0.179	15.80***	41.22***	19.81
	[-0.713, 0.330]	[8.521, 24.66]	[14.25, 65.69]	[0.641, 38.35]
ln Price *	34.63	58.75	164.0	123.1
electricity ind.	[-8.645, 95.37]	[-79.04, 244.6]	[-165.5, 543.8]	[-198.5, 442.6]
Nonattainment	1.165	3.274	6.466	5.237
county	[-0.405, 2.829]	[-0.898, 7.690]	[-2.423, 15.33]	[-2.428, 13.20]
Nonattain. *	-2.846	0.524	20.91	13.62
pollution ind.	[-13.06, 5.030]	[-9.915, 10.66]	[-15.97, 61.69]	[-8.181, 42.47]
N	269356	246909	224462	202015

Notes: Outcome variable is yearly county employment in manufacturing subsectors (mean=534). Sample is MSA counties from 1998-2009. Regressions use the difference between an observation and its synthetic counterfactual (weighted average of other counties which in aggregate has observably similar histories of the outcome and regressors). Regressions include cubic polynomial of county Ozone concentration. Columns to the right indicate estimated delayed effects on employment from initial price shocks. 95% CIs in brackets (from 399 bootstraps). * $p < .1$, ** $p < .05$, *** $p < .01$.

1.4.2   Base Synthetic Regression Result

Now I estimate the same equation but pairing a county with its synthetic counterfactual instead of its neighboring county. Column 1 of Table 1.4 contains synthetic regression results from estimating my main specification, Equation 1.3. The later columns show delayed effects from shocks. I find that the effect of energy prices on employment are insignificant and bounded much closer to zero than that of KM. Some delayed effects are significant but quite small.

The sectors may have significant effects that are not correlated by electricity intensity. Table 1.5 shows estimates for when the model is run separately by sector. The results remain insignificant and bounded close to zero except for one of the 21 sectors.

Table 1.5: Sectoral Results

Sector	Electricity index	Point estimate	95% CI
Primary metal manufacturing (mnfct)	1.000	-12.46	[-29.89, 4.97]
Paper mnfct	0.856	-2.32	[-5.78, 1.15]
Textile mills	0.591	-1.07	[-4.58, 2.44]
Nonmetallic mineral product mnfct	0.527	-17.14**	[-31.19, -3.1]
Chemical mnfct	0.459	1.94	[-15.63, 19.51]
Plastics and rubber products mnfct	0.364	6.74	[-2.21, 15.7]
Wood product mnfct	0.265	3.8	[-2.49, 10.09]
Petroleum and coal products mnfct	0.254	28.7	[-6.01, 63.42]
Fabricated metal product mnfct	0.175	3.09	[-5.96, 12.14]
Printing and related support activities	0.154	2.33	[-6.69, 11.35]
Textile product mills	0.149	-16.71	[-40.01, 6.59]
Food mnfct	0.128	-8.07	[-27.75, 11.62]
Elec. equip., appl., and comp. mnfct	0.112	3.81	[-2.7, 10.32]
Furniture and related product mnfct	0.094	0.69	[-2.27, 3.66]
Leather and allied product mnfct	0.077	-0.29	[-1.15, 0.57]
Machinery mnfct	0.068	-2.64	[-21.41, 16.13]
Apparel mnfct	0.067	1.18	[-1.85, 4.22]
Miscellaneous mnfct	0.059	-8.4	[-26.07, 9.27]
Beverage and tobacco product mnfct	0.053	3.13	[-2.68, 8.94]
Transportation equipment mnfct	0.045	3.55	[-18.05, 25.16]
Computer and electronic product mnfct	0.000	-2.28	[-5.41, 0.85]

Notes: Coefficients from separate sector-level regressions of yearly county employment in manufacturing subsectors (mean=534) on county log electricity prices. Sample is MSA counties from 1998-2009. Regressions uses the difference between an observation and its synthetic counterfactual (weighted average of other counties which in aggregate has observably similar histories of the outcome and regressors). Controls include county non-attainments status and a cubic in average county ozone pollution. Inference from 399 bootstraps. * $p < .1$, ** $p < 0.05$, *** $p < 0.01$.

1.4.2.1 Robustness

In Table 1.6 I show that these null results are stable after instrumenting with the shift-share measure. The F-statistic on the excluded instruments is quite strong. Results in column 1 again point to insignificant effects of energy prices.

The national employment trend of manufacturing from 1998-2009 was quite varied, with some years experiencing large overall losses and other years experiencing only small losses. To see if insignificant local results are due to heterogeneity by this national pattern I separately estimate the effect on split samples. The years with small losses (1998-2000 and 2004-2007) experienced an average yearly decline of 134 thousand jobs, while the years with large losses (2001-2003 and 2008-2009) experienced an average yearly decline of 919 thousand jobs. Table 1.7 shows the estimates on contemporaneous employment for these two time periods. The results are similar to those from the whole sample. During years with small national declines there is a small significant positive base effect of price increases but no effect for an industry being more energy-intensive. During years with large national declines there is no significant relationship between local price increases and employment.

1.4.2.2 Source of Discrepancy

To isolate which factor drives the difference between my estimates and those of KM, I estimate an intermediate model that is similar to KM in terms of match

---

[19]Synthetic regression results are by allowing any county to form a counterfactual (including those in the same MSA). This is because, although counties in the same MSA receive a weight that is higher than average, it is small in absolute terms. For this reason, these models still differ from the KM results for the same two basic reasons.

Table 1.6: Base Results (IV)

	(1) L_t	(2) L_{t+1}	(3) L_{t+2}	(4) L_{t+3}
Ln Energy Price	1.581	7.110**	6.839*	-0.673
	[-8.348, 11.66]	[-6.453, 35.28]	[-18.25, 46.23]	[-32.52, 32.20]
ln Price * electricity index	-667.1	-848.3	-937.8	-1043.6
	[-6463.0, 6193.5]	[-17150.5, 9900.1]	[-11228.3, 16341.3]	[-14894.8, 16909.8]
Nonattainment county	1.608	1.504	0.837	2.470
	[-2.630, 6.611]	[-5.974, 11.63]	[-9.675, 15.47]	[-13.56, 13.97]
Nonattainment * pollution index	-10.80	-10.99	6.451	-4.990
	[-70.89, 64.37]	[-186.9, 131.0]	[-139.6, 180.8]	[-180.2, 272.0]
N	202574	182880	163186	143205
Excl. IV F-stat	505.7	508.7	253.1	504.6

Notes: Outcome variable is yearly county employment in manufacturing subsectors (mean=534). Sample is MSA counties from 1998-2009. Regressions use the difference between an observation and its synthetic counterfactual (weighted average of other counties which in aggregate has observably similar histories of the outcome and regressors). Regressions include cubic polynomial of county Ozone concentration. Regressions instrument the electricity price with a shift-share interaction of initial county electricity generation capacity by fuel type (coal, oil, gas) interacted with national trends in those fuel prices. Columns to the right indicate estimated delayed effects on employment from initial price shocks. 95% confidence intervals in brackets (from 399 bootstraps). * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 1.7: Results by National Manufacturing Employment Trend

	(1) Employment	(2) Employment
Ln Energy Price	28.80**	0.582
	[8.593, 50.47]	[-0.663, 1.527]
ln Price * electricity index	125.1	-1.854
	[-131.7, 422.3]	[-24.39, 18.28]
Nonattainment county	-6.920**	5.120*
	[-11.94, -2.523]	[-0.455, 9.365]
Nonattainment * pollution index	42.00	2.988
	[-25.77, 108.1]	[-28.40, 29.75]
National Man Emp Loss	Small	Large
N	157751	112684

Notes: Outcome variable is yearly MSA-county employment in manufacturing subsectors (mean=534). Sample periods split by size of yearly decline in national manufacturing: Small declines (1998-2000 and 2004-2007) and large declines (2001-2003 and 2008-2009). Regressions use the difference between an observation and its synthetic counterfactual (weighted average of other counties which in aggregate has observably similar histories of the outcome and regressors). Regressions include cubic polynomial of county Ozone concentration. Columns to the right indicate estimated delayed effects on employment from initial price shocks. 95% CIs in brackets (from 299 bootstraps). * $p < .1$, ** $p < .05$, *** $p < .01$.

quality but that prevents spill-overs as in my base estimation.[19] To do this, note that for a given evaluation county, the synthetic counterfactual of its neighbor will not be contaminated by spill-overs from the evaluation county but will be at least as bad of a match to the evaluation county as the neighbor. Table 1.8 column 3 shows estimates of this "cross" model, which is estimated similar to the KM border-pair strategy but where counties are paired with the synthetic controls of their neighbors rather than the neighbors themselves. The estimate on the interaction term between price and electricity index of the cross model is similar to that of KM and indeed is contained within the 95% confidence interval of the KM estimate. This suggests that what drives the difference between my main estimates and those of KM is that neighbors serve as poor counterfactuals. Modeled as in Griliches (1979), this implies that the MSA share

Table 1.8: Source of Bias

	KM L_t	Main L_t	Cross L_t
Ln Energy Price	-41.60	-0.179	1018.8**
	[131.2, -214.4]	[-0.713, 0.330]	[939.9, 1075.9]
ln Price * electricity index	-1140.7**	34.63	-2064.1**
	[-74.07, -2207]	[-8.645, 95.37]	[-2235.1, -1860.1]
Nonattainment county	-5.015	1.165	407.4**
	[42.30, -52.33]	[-0.405, 2.829]	[384.8, 429.8]
Nonattainment * pollution index	-421.0**	-2.846	-534.9**
	[-180.0, -662.1]	[-13.06, 5.030]	[-582.2, -494.6]
N	969168	269356	919817

Notes: Dependent variable is county-sector-employment (mean=534). KM regression is a border pair design. SR is a synthetic regression. Cross is a border pair design where a county is matched to the synthetic control of its neighbor. Regressions include cubic polynomial of county Ozone concentration. 95% CIs in brackets (from 299 bootstraps). * $p < .1$, ** $p < .05$, *** $p < .01$.

of the variance in the endogenous component of prices is less than in the exogenous component of prices.

1.4.3 Perceived Permanent Shocks

The fact that transitory, yearly shocks have little effect on employment may not be surprising. Decisions about the opening or closing of establishments and about staffing are likely done over a long time horizon in manufacturing. I therefore check for effects from perceived permanent price changes.

First, I examine the effect of large and persistent shocks to electricity prices. Following Ajzenman et al. (forthcoming), I identify "persistent" shocks that satisfy two criteria: i) for three years the price is consistently higher (lower) than expected, compared to its counterfactual, and ii) the three year difference from what was expected is in the top 10% of all such differences. This includes 9.2% of the county-sector-year

observations. 28.0% of the county-sectors have at least one persistent shock during their history.

I estimate a modification of Equation 1.3 that includes interactions with the persistent shock dummy variable (those on the third line).

$$
\begin{aligned}
\Delta l_{ctk} =& \beta_1 \Delta p_{ct} + \beta_2 \Delta p_{ct} \cdot ElecIndex_{kt} + \beta_3 ElecIndex_{kt} + \\
& \beta_4 \Delta Nonattain_{ct} + \beta_5 \Delta Nonattain_{ct} \cdot PollIndex_k + \beta_6 PollIndex_k + \\
& \beta_7 PersShock_{ct} + \beta_8 PersShock_{ct} \cdot \Delta p_{ct} + \beta_9 PersShock_{ct} \cdot \Delta p_{ct} \cdot ElecInd_{kt} + \\
& f(\Delta Poll_{ct}) + \varepsilon_{ctk} \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (1.5)
\end{aligned}
$$

Results are shown in Table 1.9. One sees that now there is a significant negative effect for persistent shocks to energy intensive sectors. The point estimate stays negative for three years, but becomes insignificant. The estimate for the main effect from energy prices on energy intensity is now positive and significant. This implies that there is a small positive effect from transitory shocks, but, compared to those, the effect of a perceived permanent shock is more negative.

Another signal that prices may change more permanently is the expectation about future electricity prices. As direct county-level data for this does not exist, I construct a measure using existing national-level expectations for future fuel prices interacted with the initial fuel-capacity shares of each county's generation capabilities. The expectations are from the price of fuel futures contracts with delivery dates three

---

[20]I use the yearly average of prices of futures contracts traded on the CME NYMEX exchange. For

Table 1.9: Effects of Persistent Shocks

	(1) L_t	(2) L_{t+1}	(3) L_{t+2}
Ln Energy Price	-0.415	2.732**	9.563**
	[-1.292, 0.361]	[0.664, 5.410]	[0.543, 23.54]
ln Price * electricity index	33.98*	19.60	61.46
	[1.016, 77.75]	[-10.76, 63.94]	[-80.16, 278.8]
Persistent shock	-0.440	-0.437*	-1.225
	[-1.112, 0.0398]	[-0.848, -0.0377]	[-3.165, 0.167]
Persistent * ln Price	-0.252	-3.007**	-10.27
	[-1.844, 0.975]	[-5.705, -0.841]	[-24.85, -0.836]
Persistent * ln Price	-42.41*	-7.724	-30.79
* electricity index	[-90.73, -4.226]	[-57.55, 31.88]	[-262.0, 126.3]
N	202022	202022	202022

Notes: Outcome variable is yearly county employment in manufacturing subsectors (mean=534). Sample is MSA counties from 1998-2009. Regressions use the difference between an observation and its synthetic counterfactual (weighted average of other counties which in aggregate has observably similar histories of the outcome and regressors). Regressions includes controls for Nonattainment county status and a cubic polynomial of county Ozone concentration. Columns to the right indicate estimated delayed effects on employment from initial price shocks. 95% confidence intervals in brackets (from 399 bootstraps). * $p < .1$, ** $p < .05$, *** $p < .01$.

years in the future.[20] I use as expectations the price of the contract three years before expiration, or the closest month to that given the data. This data exists for the three fuels from 2001-2009.

In Table 1.10 I estimate a model with future price measures. I find significant negative results for one-year and two-years after a change in electricity price expectations. The size of these estimated elasticities is still quite small, with the two-year effect corresponding to an elasticity of -0.05.

---

oil I use the CL contract (CME code) for light sweet crude. For natural gas I use the NG contract for natural gas traded at the Henry Hub in Louisiana. For coal I use the QL contract for Central Appalachian Coal.

Table 1.10: Effect of Futures Prices

	(1) L_t	(2) L_{t+1}	(3) L_{t+2}
Ln Energy Price	1.179	5.951**	10.10
	[-1.317, 3.522]	[1.166, 11.12]	[-1.788, 23.49]
ln Price * electricity index	-20.35	45.30	-117.9
	[-63.36, 13.23]	[-64.27, 187.4]	[-263.1, 55.83]
Ln Future Energy Price	-0.145	-0.709	-0.225
	[-0.584, 0.248]	[-1.602, -0.00908]	[-1.753, 1.430]
Ln Future Energy Price *	1.617	-24.28***	-30.29***
electricity index	[-1.153, 5.195]	[-35.80, -13.64]	[-42.81, -20.70]
N	182418	182417	182417

Notes: Outcome variable is yearly county employment in manufacturing subsectors (mean=534). Sample is MSA counties from 1998-2009. Regressions use the difference between an observation and its synthetic counterfactual (weighted average of other counties which in aggregate has observably similar histories of the outcome and regressors). Regressions includes controls for Nonattainment county status and a cubic polynomial of county Ozone concentration. Columns to the right indicate estimated delayed effects on employment from initial price shocks. 95% confidence intervals in brackets (from 299 bootstraps). * $p < .1$, ** $p < .05$, *** $p < .01$.

1.5   Discussion

In this section I discuss the implications of my estimated elasticities for predicting effects of the EPA's Clean Power Plan. I then discuss possible future uses of this estimation strategy.

1.5.1   Effects of CPP

The CPP is the largest US climate change policy. It sets state-by-state limits for CO_2 emissions for the power generation sector to be achieved by 2022 and 2030. Each limit was calculated by devising a federal plan that applies existing solutions for reducing carbon to a state's fleet of generators. States are free, however, to implement

---

[21]The EPA does not attempt to take into account employment effects in non-energy-producing sectors, though it is beginning such a process by convening a panel of experts to examine at this issue.

their own policies to meet their state-level limits. The Environmental Protection Agency's Regulatory Impact Analysis (EPA, 2014b) forecasts that the CPP will affect energy prices non-uniformly by up to 6%.[21]

The EPA's Clean Power Plan provides two compliance options for states, a rate-based cap based on the tons of CO_2 that are emitted per MWh of electricity, or a mass-based cap based on total tons of CO_2 emitted. Table 1.12 shows the Regulatory Impact Assessment predictions for the increase in retail electricity prices based on these two choices for 2020. Employment effects estimated here are based on the rate-based caps, though they are similar for the mass-based caps given the retail price increases are quite similar. Retail price changes are predicted based on EPA Electricity Market Module Regions.[22] These are shown in Figure 1.1. Currently, due to lack of data to establish baselines, the Clean Power Plan only applies to the contiguous US.[23]

The EMM regions are not geographically defined but are defined by energy company membership. I translate counties to EMM regions in several steps. First, some states lie completely within a region. Second, Form EIA-861 data connects utilities to Balancing Authorities which can usually be unambiguously mapped to regions using the Emissions & Generation Resource Integrated Database (eGRID) 2012 data. The remaining counties are assigned to the region that had the most generating capacity in the county's state as noted by the eGRID 2012 data.

Given the estimated price increase at the county level, I compute the expected effect on manufacturing given each county's subsectoral employment and the estimated

---

[22]Equivalent to the EPA Emissions & Generation Resource Integrated Database subregion.

[23]Additionally, as Vermont and Washington D.C. do not have any fossil-fuel electricity generating units, they are not they are not covered directly by the plan.

Figure 1.1: EPA eGRID Subregions
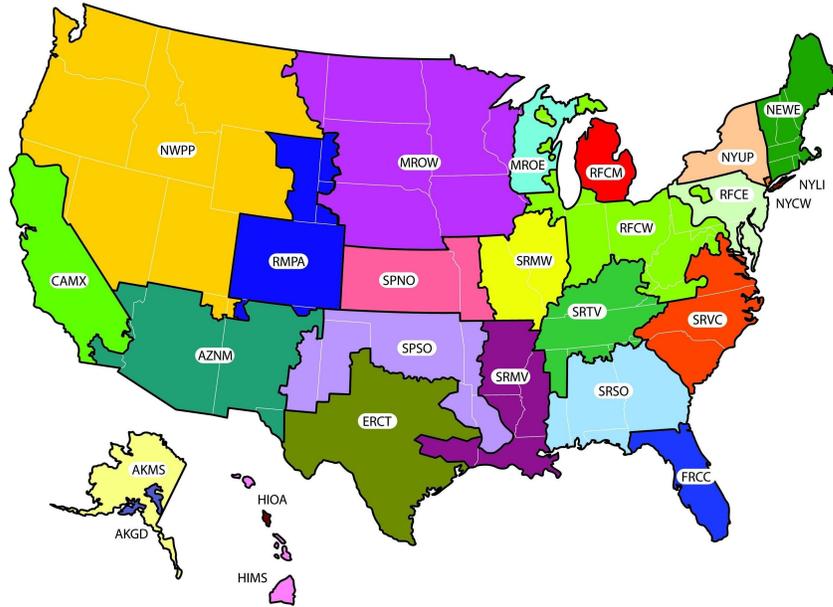
sectoral elasticity. Using the subsectoral elasticities of KM,[24] one would predict a loss of 84,000 jobs (.9% of total manufacturing). The elasticity they estimate is one that is based on contemporaneous prices whereas the policy change is a predicted future change, so this is likely an upper bound. Using the most negative elasticities from my specifications (from the futures prices regressions), I estimate a loss of only 4,100 jobs (.045%). Given my elasticity is estimated from changes to future price expectations, these would be delayed effects happening around two years after the change.

The elasticities estimated are from local-level changes in electricity prices and therefore estimate the extent of local variation in effects and not a national effect. The CPP will affect prices nationally but given my low estimates one would not expect concentrated effects. The predicted local effects are likely to be upper bounds given

---

[24]Elasticities are from KM Table 3 column 1.

that most other locations will also experience an increase in prices and therefore there will be less re-allocation.

1.5.2   Other Border-Pair Studies

While I have applied my identification strategy to electricity prices and manufacturing employment it has broader applications to research that has utilized panel border-pair methodologies. These designs have been used in the study of manufacturing employment since Holmes (1998), in the minimum-wage literature since Dube et al. (2010), finance since Huang (2008), and have been introduced into the economics of education (Naidu, 2012; Dhar and Ross, 2012), armed conflict (Arias et al., 2014) and development (Acemoglu et al., 2012).

In these settings, one may wish to use the synthetic regression for several reason. First, it may reduce bias in situations with spillovers between units or inadequate matches. Second, as this approach attempts to control for a slightly different class of unobservables it can serve as a methodological robustness check or as a substitute if the econometrician has some outside knowledge of possible confounders.

1.6   Conclusion

In this paper I investigate the effect of local electricity prices on local manufacturing employment. Recent best estimates of this elasticity may be biased by dependence between neighboring counties in the same labor market and inadequate match quality. To deal with these potential problems I develop a novel estimation

strategy and provide conditions for its consistency. Applying this methodology I find that the estimated effects of contemporaneous price shocks on contemporaneous employment are in general insignificant.

Given that manufacturing likely faces adjustments costs for their factor inputs, I estimate effects from perceived permanent price changes and for delayed effects. I find effects that are significant and negative for large and persistent shocks and from changes in the expectation of future electricity prices. Economically, though, the estimated elasticities are small, ranging from 0 to -0.05 for the most electricity intensive subsector. These magnitudes are similar to the earlier historical studies of this elasticity. Though firms do not appear to meaningfully change employment levels, they may adjust along other margins, such as hours, wages, or capital investment.

Given these sectoral elasticities, the predicted effects of the EPA's Clean Power Plan on employment are drastically reduced from current best estimates. Though the Clean Power Plan is predicted to raise electricity rates by up to 6%, in a non-uniform way across the US, I find that predicted differential effect across the US is on the order of several thousand jobs, smaller than typical monthly variation in manufacturing employment.

Finally, the identification strategy used in this paper may help others in a panel setting. It provides identification with weaker assumptions than the standard panel model by controlling for unobserved factors. It can be an alternative to border-pair designs, especially when one is concerned about interference between units or non-parallel trends.

1.7 Appendix A: Theoretical Properties of the Estimator

In this section I detail the theoretical properties of my estimator. I first note the need for a modification to the matching procedure. Then, I move to proving the consistency and asymptotic normality.

1.7.1 Smoothness of Matching Procedure

The smoothness by which small changes to the data have predictably small effects on the estimate is a property that is important for both consistency and valid inference in large samples. The matching procedure may not be smooth in finite sample, though, with a modification it will be in large samples.[25]

In finite sample, the predictors of an evaluation units may not be in the convex hull of other units and therefore will not have a synthetic counterfactual that exactly reproduces its characteristics. This is more likely given I match on multiple variables over many years. Perturbing the data of another unit may change the convex hull causing different units to be used for a match. This can cause a large corresponding difference in average outcome.[26] For this reason it is important to use the wild bootstrap as it has been shown to be robust to the same type of non-smoothness in similar matching estimators (Otsu and Rai, 2015).

When a unit is in the convex hull and there are more non-colinear units than the dimension of the matching vector (as in the large-sample case) there is an indeterminacy

---

[25]It should be noted that Abadie et al. (2010) show consistency as $T_0 \to \infty$ where I rely on $N \to \infty$.

[26]Note that the inability to match exactly can be avoided if we allow negative weights and have at least as many (non-colinear) units as predictor variables. This is disallowed in Abadie et al. (2010)

in picking the weight vector as infinitely many will match exactly. Abadie et al. (2010) do not resolve the indeterminacy in their paper and therefore do not guarantee asymptotic smoothness.[27] Asymptotic smoothness is important for determining the asymptotic distribution of the estimator. I therefore extend the matching procedure to resolve the indeterminacy in a smooth manner.

I make a modification the algorithm to find weights $\boldsymbol{w}_i$ that are the most distributed when possible. Let $\boldsymbol{x}_i$ contain the match variables of unit i, and $\boldsymbol{X}$ the $(k \times N)$ matrix for all units. Then, when legal weights can be found such that $\boldsymbol{X}\boldsymbol{w}_i = \boldsymbol{x}_i$, optimize

$$\min_{\boldsymbol{w}_i} \boldsymbol{w}_i'\boldsymbol{w}_i \quad s.t. \tag{1.6}$$

$$\mathbf{1}_N \geq \boldsymbol{w}_i \geq \mathbf{0}_N$$

$$\boldsymbol{w}_i' \cdot \mathbf{1}_N = 1$$

$$\boldsymbol{X}\boldsymbol{w}_i = \boldsymbol{x}_i$$

As is shown in Theorem 1 this causes weights for all units to approach zero. As more units are generated, the disparity between donors with high and low weights is diminished.[28]

---

as it forces the researcher to evaluate whether matches are appropriate given that estimates using data far from the convex hull are likely to be less valid.

[27]Their programmed estimators appears to not guarantee smoothness when picking the weights.

[28]Another possible way to resolve the indeterminacy would be to pick weights so that units receiving large weight are as similar as possible to the evaluation unit. To do so, first construct a distance measure for the similarity of two units $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ such as $d(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{V}) = \sum_k [(x_{ik} - x_{jk})/x_{ik}]^2 v_k$ (where $\boldsymbol{V}$ denotes the vector of predictor weights). Define $\boldsymbol{d}(\boldsymbol{x}_i)$ as the vector with jth element $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Then, if weights can be found such that $\boldsymbol{X}\boldsymbol{w}_i = \boldsymbol{x}_i$, we can change the objective function of Equation 1.6 to $\min_{\boldsymbol{w}_i} \boldsymbol{w}_i'\boldsymbol{d}(\boldsymbol{x}_i)$, keeping the same constraints.

With zero matching variables, the weights become uniform and constant across observations so that the procedure results in OLS (or GLS if down-weighting by the mean square predicted error).

1.7.2 Law of Large Numbers for Weighted Averages

Theorem 1. *Suppose that $\{w_{ni} : n \geq 1, n \geq i \geq 1\}$ is a double sequence of real non-negative numbers from the optimization problem of Equation 1.6. Then for $\{X_i\}$ i.i.d. with $\mathrm{var}(X_1) = \sigma^2$ and $\mathrm{E}[X_1] = \mu$, let $\tilde{X}_n = \sum_{i=1}^{n} w_{ni}X_i$, then*

$$\lim_{n \to \infty} \tilde{X}_n \xrightarrow{p} \mu$$

Proof. First, calculate

$$\mathrm{var}(\tilde{X}_n) = \sigma^2 \sum_{i=1}^{n} w_{ni}^2$$

$$= \sigma^2 \gamma_n$$

where $\gamma_n = \sum_{i=1}^{n} w_{ni}^2$. The standard case is when $w_{ni} = n^{-1}$ and then $\gamma_n = n^{-1}$. By Chebyshev's inequality,

$$\Pr\left(\left|\tilde{X}_n - \mu\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{\varepsilon^2}\gamma_n$$

This would both limit interpolation bias and give more intuition for the weights (concentrating weight to fewer units) so it is more in the spirit of Abadie et al. (2010). However, since evaluation units will be matched with positive weight to other units, this procedure it also not asymptotically smooth. If such an estimator was consistent it would likely rely on $T_0 \to \infty$ (similar to the classic SCM) which is less convincing in my setting given the length of my panel.

For the weak law to hold we need that

$$\lim_{n \to \infty} \gamma_n \to 0$$

Conveniently, this is minimization problem in Equation 1.6. First, note that with the generation of new data points, the optimization problem will only transfer weight from units that had more weight to those that had less (the transfer is linear in weights and it is minimizing the quadratic). This implies that if $\kappa_n = \arg\max_j w_{nj}$ then $w_{n\kappa_n} \geq w_{m\kappa_n}$ for $m > n$. Indeed, $\exists m > n$ such $w_{n\kappa_n} > w_{m\kappa_n}$. In the most degenerate case, eventually the DGP will produce additional points that encompass unit κ_n in their convex hull, at which point w_{κ_n} must diminish. By similar arguments, the maximum weight will eventually fall below any fixed threshold $\omega \in (0, 1)$. This implies that the maximum of all weights approaches zero, $\lim_{n \to \infty} w_{n\kappa_n} \to 0$. Therefore, for all j, $\lim_{n \to \infty} w_{nj} \to 0$. $\square$

1.7.3 Panel Setup

Using the DGP in Equation 1.1 to create the stack of pre-evaluation differences

$$\overleftarrow{\boldsymbol{y}}_{it} - \sum_j w_{jit} \overleftarrow{\boldsymbol{y}}_{jt} = \beta(\overleftarrow{\boldsymbol{x}}_{it} - \sum_j w_{jit} \overleftarrow{\boldsymbol{x}}_{jt}) + \overleftarrow{\boldsymbol{\lambda}}_t(\mu_i - \sum_j w_{jit}\mu_j) +$$

$$(\overleftarrow{\boldsymbol{\varepsilon}}_{it} - \sum_j w_{jit}\overleftarrow{\boldsymbol{\varepsilon}}_{jt}) \; \forall i, t$$

Assuming we can find a match according to Equation 1.2 then this becomes

$$\overleftarrow{\boldsymbol{\lambda}}_t(\mu_i - \sum_j w_{jit}\mu_j) = -(\overleftarrow{\bar{\varepsilon}}_{it} - \sum_j w_{jit}\overleftarrow{\bar{\varepsilon}}_{jt})$$

$$\lambda_t(\mu_i - \sum_j w_{jit}\mu_j) = -\boldsymbol{\Lambda}_t(\overleftarrow{\bar{\varepsilon}}_{it} - \sum_j w_{jit}\overleftarrow{\bar{\varepsilon}}_{jt}) \tag{1.7}$$

where $\boldsymbol{\Lambda}_t = \lambda_t(\overleftarrow{\boldsymbol{\lambda}}_t'\overleftarrow{\boldsymbol{\lambda}}_t)^{-1}\overleftarrow{\boldsymbol{\lambda}}_t'$. Substituting this into the contemporaneous differences becomes then

$$\underbrace{y_{it} - \sum_j w_{jit}y_{jt}}_{\Delta y_{it}} = \beta\underbrace{(x_{it} - \sum_j w_{jit}x_{jt})}_{\Delta x_{it}} + \underbrace{-\boldsymbol{\Lambda}_t(\overleftarrow{\bar{\varepsilon}}_{it} - \sum_j w_{jit}\overleftarrow{\bar{\varepsilon}}_{jt})}_{\Delta r_{1it}} + \underbrace{(\varepsilon_{it} - \sum_j w_{ji}\varepsilon_{jt})}_{\Delta r_{2it}}$$

$$\tag{1.8}$$

Let $\tilde{\boldsymbol{y}}$ be the stacked Δy_{it} observations and similarly for $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{r}}_1, \tilde{\boldsymbol{r}}_2$. Then estimating

$$\tilde{\boldsymbol{y}} = \beta\tilde{\boldsymbol{x}} + \tilde{\epsilon}$$

yields

$$\hat{\beta} = (\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{y}}$$

$$= \beta_0 + (\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{x}}'(\tilde{\boldsymbol{r}}_1 + \tilde{\boldsymbol{r}}_2) \tag{1.9}$$

1.7.4 Consistency

I show consistency for fixed T and $N \to \infty$. In the limit.

$$\plim_{N \to \infty} (\hat{\beta} - \beta_0) = \plim_{N \to \infty} \left(\frac{1}{TN} \tilde{\boldsymbol{x}}' \tilde{\boldsymbol{x}} \right)^{-1} \frac{1}{TN} \tilde{\boldsymbol{x}}' (\tilde{\boldsymbol{r}}_1 + \tilde{\boldsymbol{r}}_2)$$

Expanding using the above yields

$$\plim_{N \to \infty} \left(\frac{1}{TN} \tilde{\boldsymbol{x}}' \tilde{\boldsymbol{x}} \right)^{-1} \frac{1}{T} \sum_t \frac{1}{N} \sum_i (x_{it} - \sum_j w_{jit} x_{jt})$$

$$[-\boldsymbol{\Lambda}_t (\overleftarrow{\varepsilon}_{it} - \sum_j w_{jit} \overleftarrow{\varepsilon}_{jt}) + (\varepsilon_{it} - \sum_j w_{jit} \varepsilon_{jt})] \quad (1.10)$$

Let the cross-sectional means be $\mathrm{E}_i[x_{it}] = \bar{x}_t$. Then by the Weak Law of Large Numbers $\sum_j \frac{1}{N} x_{jt} \xrightarrow{p} \bar{x}_t$. Given the weights are generated from the smooth optimization problem from Section 1.7.1, then by Theorem 1 $\sum_j w_{jit} x_{jt} \xrightarrow{p} \bar{x}_t$. By the LLN and Theorem 1, $\sum_j w_{jit} (\varepsilon_{it} - \varepsilon_{jt}) \xrightarrow{p} \mathrm{E}_j[\varepsilon_{it} - \varepsilon_{jt}] = 0$. Since $\mathrm{E}_j[\overleftarrow{\varepsilon}_{jt}] = \boldsymbol{0}_{t-1}$ (0-vector of length $t-1$), then by the LLN and Theorem 1 $\sum_j w_{jit} \overleftarrow{\varepsilon}_{jt} \xrightarrow{p} \boldsymbol{0}_{t-1}$. Then Equation 1.10 becomes

$$\plim_{N \to \infty} (\hat{\beta} - \beta_0) = \left[\frac{1}{T} \sum_t \plim_{N \to \infty} \frac{1}{N} \sum_i (x_{it} - \bar{x}_t)^2 \right]^{-1} \frac{1}{T} \sum_t \plim_{N \to \infty} \frac{1}{N} \sum_i (x_{it} - \bar{x}_t)[-\boldsymbol{\Lambda}_t \overleftarrow{\varepsilon}_{it}]$$

Assume that $\mathrm{E}_i(x_{it} - \bar{x}_t)^2$ and $\mathrm{E}_i[(x_{it} - \bar{x}_t)\boldsymbol{\Lambda}_t\overleftarrow{\varepsilon}_{it}]$ exist. The summed N terms are now i.i.d., so by the law of large numbers.

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty}(\hat{\beta} - \beta_0) &= \left[\frac{1}{T}\sum_t \mathrm{E}_i(x_{it} - \bar{x}_t)^2\right]^{-1}\frac{1}{T}\sum_t \mathrm{E}_i[(x_{it} - \bar{x}_t)(-\boldsymbol{\Lambda}_t)\overleftarrow{\varepsilon}_{it}] \\
&= \left[\frac{1}{T}\sum_t \mathrm{E}_i(x_{it} - \bar{x}_t)^2\right]^{-1}\frac{1}{T}\sum_t \mathrm{E}_i[(x_{it} - \bar{x}_t)(-\boldsymbol{\Lambda}_t)]\,\mathrm{E}_i[\overleftarrow{\varepsilon}_{it}] \\
&= \left[\frac{1}{T}\sum_t \mathrm{E}_i(x_{it} - \bar{x}_t)^2\right]^{-1}\frac{1}{T}\sum_t \mathrm{E}_i[(x_{it} - \bar{x}_t)(-\boldsymbol{\Lambda}_t)]\cdot\mathbf{0} \\
&= 0
\end{aligned}
$$

Therefore $\hat{\beta}$ is consistent.

For observations that are not matched well, one can exclude them from the regression step. Alternatively, following Abadie et al. (2010) one could downweight them by their match mean-squared prediction error. Note that as $N \to \infty$, the proportion of observations that are poorly matched goes to zero.

1.7.5 Normality

By the above logic

$$
\operatorname*{plim}_{N\to\infty}\sqrt{N}\frac{1}{T}\sum_t\frac{1}{N}\sum_i(x_{it} - \sum_j w_{jit}x_{jt})[-\boldsymbol{\Lambda}_t(\overleftarrow{\varepsilon}_{it} - \sum_j w_{jit}\overleftarrow{\varepsilon}_{jt}) + (\varepsilon_{it} - \sum_j w_{jit}\varepsilon_{jt})] =
$$

$$
\operatorname*{plim}_{N\to\infty}\sqrt{N}\frac{1}{N}\sum_i\frac{1}{T}\sum_t(x_{it} - \bar{x}_t)[-\boldsymbol{\Lambda}_t\overleftarrow{\varepsilon}_{it} + \varepsilon_{it}]
$$

Let $L_i = \frac{1}{T}\sum_t (x_{it} - \bar{x}_t)[-\boldsymbol{\Lambda}_t \overleftarrow{\bar{\varepsilon}}_{it} + \varepsilon_{it}]$. Notice that $E[L_i] = 0$ and that $\{L_i\}$ are i.i.d.. Assume that $\text{var}[L_i] = V < \infty$. Then by the Lindeberg–Lévy Central Limit Theorem

$$\sqrt{N}\frac{1}{N}\sum_i L_i \xrightarrow{d} N(0, V)$$

.

1.8   Appendix B: Monte-Carlo Study of Estimator Properties

To provide evidence that the wild bootstrap gives valid inference and that the estimator is appropriate in finite sample, I conduct the following Monte-Carlo study. The DGP is

$$y_{it} = \beta_0 x_{it} + \mu_i \lambda_t + u_{it}$$

$$u_{it} \sim N(0, 1)$$

$$\beta_0 = 1$$

where μ_i is $1 \times F$, λ_t is $F \times 1$, and both are unobserved by the econometrician. I vary the complexity of the DGP by changing the dimension F. With $F > 0$, x_{it} is made

correlated with the unobservables through the interactive fixed effects as

$$x_{it} = \rho\mu_i\lambda_t + e_{it}$$

$$e_{it} \sim N(0,1)$$

$$\mu_i, \lambda_t \sim N(\mathbf{0}_F, \mathbf{I}_F)$$

$$\rho = 0.5$$

With $F = 0$, there is no interactive fixed-effect and the covariate is exogenous as in the ideal regression setup. I generate data for $T = 8$ and vary the sample size.

I estimate the model using a fixed effect OLS regression on the whole data. I then estimate the model using synthetic regression, setting aside the first $T_0 = 6$ periods just for matching. I use a wild bootstrap with 300 replications. As can be seen in Figures 1.2a and 1.3a, both the synthetic regression and OLS regression are consistent and close to the true β_0 when $F = 0$. Notice that the synthetic regression uses many periods just for matching (and not estimation) and therefore has more uncertainty at small N.

However, when $F = 1$ then the synthetic regression performs much better and improves with sample size. The OLS regression produces estimates much higher and with confidence intervals that always exclude β_0. Nor does the OLS regression estimates improve with sample size.

In Figures 1.2b and 1.3b I plot the proportion of the time that an estimator rejects β_0 at the 95% level. Ideally this would be .05. Synthetic regression approaches

Figure 1.2: Comparison of FE OLS to Synthetic Regression with No Unobserved Factors

(a) Point Estimates ($\beta_0 = 1$)
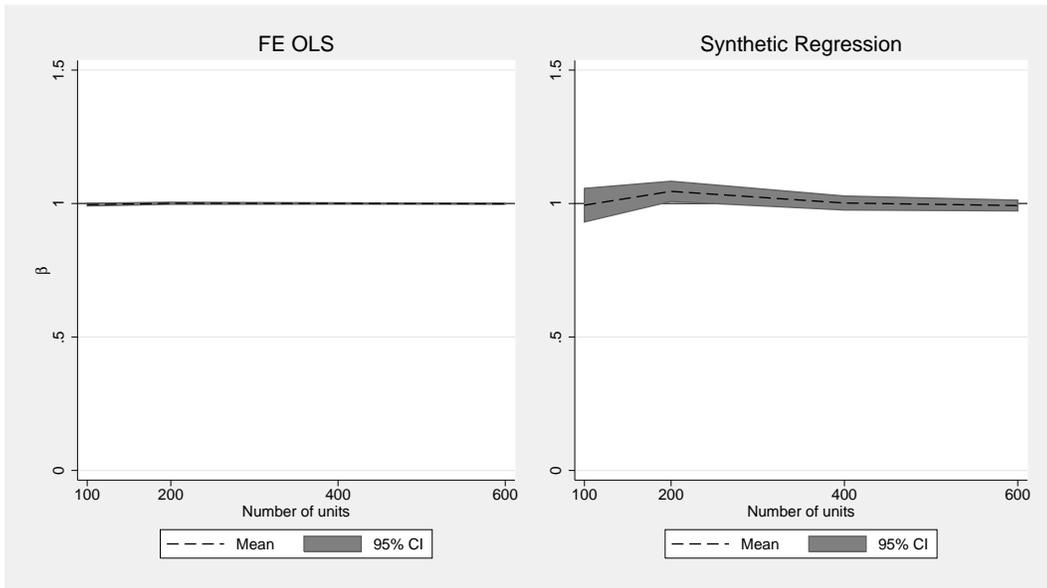
(b) 95% Confidence Interval Rejection Rates
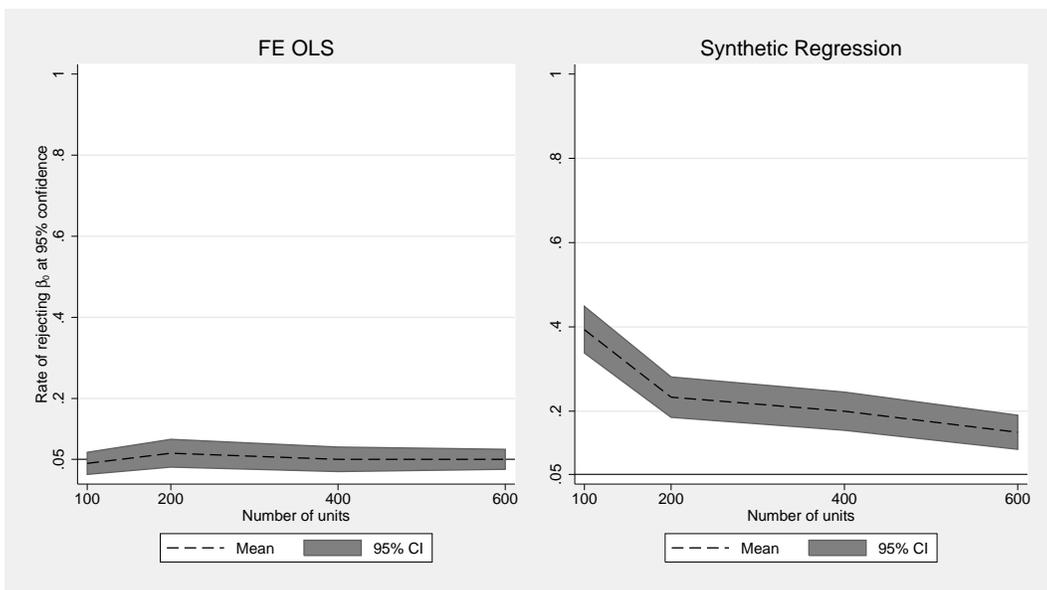
Notes: For 300 Monte Carlo simulations. DGP is $y_{it} = \beta_0 x_{it} + u_{it}$. $\beta_0 = 1$, N varies, T is fixed at 8, $u \sim N(0,1)$. $x \sim N(0,1)$. OLS regression uses unit fixed-effects. Synthetic regression uses all but the last 2 periods for just matching and uses 399 bootstrap replications.

Figure 1.3: Comparison of FE OLS to Synthetic Regression with an Unobserved Factor:

(a) Point Estimates ($\beta_0 = 1$)
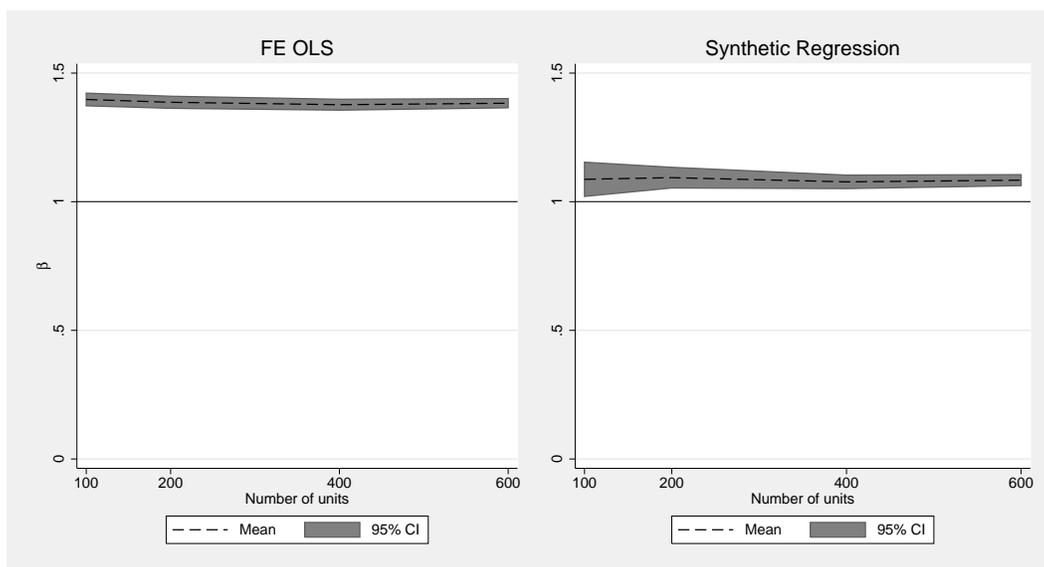
(b) 95% Confidence Interval Rejection Rates
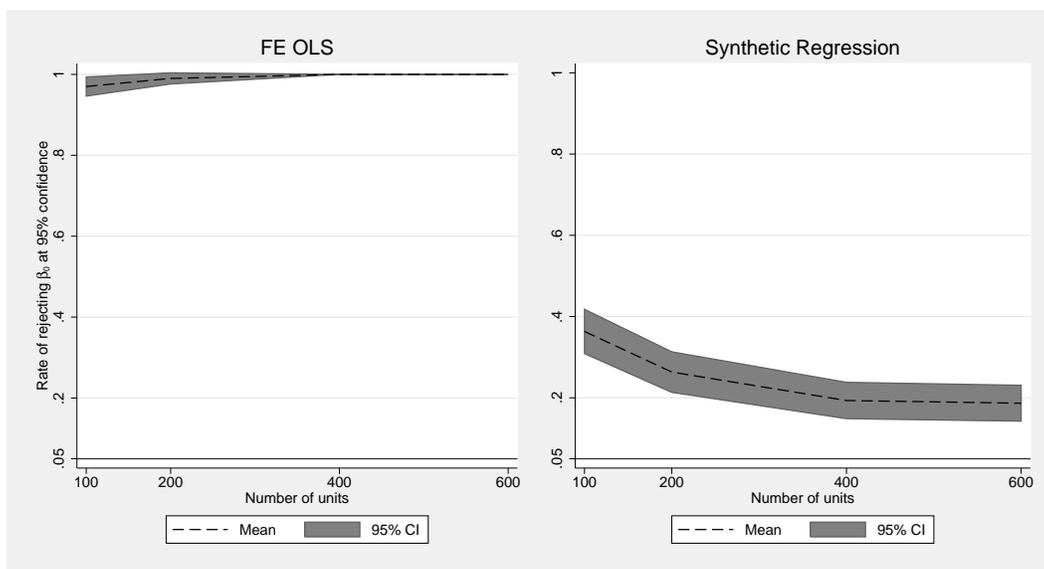
Notes: For 300 Monte Carlo simulations. DGP is $y_{it} = \beta_0 x_{it} + \mu_i \lambda_t + u_{it}$. $\beta_0 = 1$, N varies, T is fixed at 8, $u \sim N(0,1)$. $(\mu_i, \lambda_t) \sim N(0,1)$, and $x = \rho \mu_i + N(0,1)$ so that x is correlated with the unobservables. OLS regression uses unit fixed-effects. Synthetic regression uses all but the last 2 periods for just matching and uses 399 bootstrap replications.

Table 1.11: Border-Pair Sensitivity - All Manufacturing

| | (1) | (2) | (3) | (4) |
	N	N	N	N
ln Electricity price	-7546.1	-7689.3	1583.5	2487.3
	(7123.8)	(7016.9)	(3303.7)	(4373.5)
Nonattainment county	4391.3	4377.4	4821.4*	5348.6*
	(3021.4)	(3030.8)	(2723.7)	(2884.7)
No pollution monitor	66098.2			
	(221300.5)			
R^2	0.925	0.925	0.999	0.999
FE type	coarse	coarse	fine	fine
Weighting	border	county	county	border
Observations	48852	48852	48672	48672

Notes: All regressions include cubic polynomials of ozone concentration and right to work status. Regressions without county FEs also include population in 1970, miles to central business district, land area and 1990 housing values. Standard errors clustered at the utility level. Weight-type county=counties have same total weight. Weight-type border=county comparisons (borders) have same weight. FE-type coarse=county-pair industry#year state#year, fine=county county-pair#year. * $p < .1$, ** $p < .05$, *** $p < .01$.

this for both models as the sample size increases. OLS regression estimates almost never include β_0 in their confidence intervals when there are interactive fixed effects.

1.9 Appendix C: Additional Tables

Table 1.11 shows estimates from a border-pair design that does not separate manufacturing into subsectors. The effect of energy prices on total employment is insignificant and varies in sign depending on the level of detail in the fixed effects.

Table 1.12: 2020 Projected Contiguous U.S. and Regional Retail Electricity Price Changes Due to the Clean Power Plan

EMM region	Rate-based	Mass-based
ERCT	2.5%	2.1%
FRCC	2.0%	1.6%
MROE	4.2%	3.8%
MROW	2.8%	2.3%
NEWE	5.1%	5.5%
NYCW	5.0%	5.3%
NYLI	4.6%	5.1%
NYUP	5.4%	5.3%
RFCE	6.1%	6.1%
RFCM	4.3%	4.3%
RFCW	5.1%	4.8%
SRDA	2.1%	1.7%
SRGW	4.1%	4.8%
SRSE	0.9%	0.5%
SRCE	1.1%	0.8%
SRVC	1.5%	1.2%
SPNO	-0.8%	-0.9%
SPSO	3.2%	2.4%
AZNM	2.1%	2.1%
CAMX	3.3%	3.0%
NWPP	3.2%	2.9%
RMPA	3.1%	2.9%
Contiguous U.S.	3.2%	3.0%

Source: (EPA, 2015)

Notes: Electricity Market Module regions show in Figure 1.1 The mass-based compliance option sets a limit on the total amount of CO_2 emitted while the rate-based cap limits the amount per MWh of generated electricity.

Chapter 2: Capitalitis? Effects of the 1960 Brazilian Capital Relocation

2.1 Introduction

Capitals are often the largest cities in their countries. In a dataset of Latin American cities in 1990, Galiani and Kim (2008) find that being a national capital is associated with a 919% increase in size. Internationally, a national capital is on average in the 7th percentile of the population distribution of cities.[1] Additionally, when ranking all cities in a country by population, capital cities are often bigger than predicted by their position in the ranking. Gabaix (1999) notes that when plotting log-population against log-rank for cities in a country ("Zipf" plots), capital cities commonly lie above the best-fit line.

Capitals could be larger than expected for two potential reasons: governments might elect to have their seats of power in cities with strong growth prospects or city growth could be positively affected by being a government seat. Previous analyses of the relationship between capitals and city size (Ades and Glaeser, 1995; Galiani and Kim, 2008) have used only cross-sectional variation to identify the relation between being a capital and city size. I utilize the 1960 relocation of the Brazilian capital from Rio de Janeiro to Brasília to estimate what, to my knowledge, are the first causal

---

[1]Of countries with at least 20 cities in the UN Statistics Division city size database.

51

estimates of gaining and losing capital status. I use a synthetic control identification strategy (Abadie et al., 2010) to generate counterfactual outcomes for both Rio de Janeiro and Brasília. My main estimates limit potential control cities to other Brazilian cities. As a robustness check, I also estimate the treatment effects using international capital cities as potential comparison cities.

I find that the effect of relocating the capital had insignificant effects on Rio de Janeiro in terms of population and GDP. The effects on Brasília, however, were significant, positive, and persistent for both population and GDP. These results are robust to alternative sets of possible comparisons and to methodologies that omit areas that have also been affected by the relocation.

Identifying the causal impact of capital designation is both theoretically interesting and policy relevant. At a theoretical level, the effect on Brasília suggests that capitals do affect population, but the asymmetry implies that this relationship is more complicated than previous theories account for. Additionally, the effect on Brasília's growth rate (as shown later) disappears over time, implying that this shock did not change the location's long-run growth rate.

The results can also help inform decisions regarding capital relocations and related policies. Developing countries sometimes use capital relocations as a regional development tool (most of the relocations happen in developing countries and reasons given often involved development). For instance, Argentina has twice considered moving its capital: in 1987 to the South and in 2014 to the North. In most countries the relocation of the national capital is a limited tool given there is just one, though some countries divide the national government functions across cities. Subnational

state capitals are also relocated. A final, similar policy is the creation of completely new ("greenfield") cities. For example, India is pursuing plans for four such cities (Sharma, 2010).

Aside from estimating the causal impacts of the capital relocation on population and economic activity, this paper extends the synthetic controls methodology of Abadie et al. (2010) to situations where some units may interact with each other, for example, due to migration. I develop a test to check for violations of the stable unit treatment value assumption (SUTVA) using treated and synthetic control units, a methodology for dealing with such violations, and a robustness test for inference.

The paper proceeds as follows. Section 2.2 reviews the related literature and the historical background for Brazil. Section 2.3 describes the empirical approach and the data used. Section 2.4 details the results. Section 2.6 concludes.

2.2 Background

2.2.1 Literature Review

The first area of related literature is the study of capital cities and population.[2] One of the main political theories put forward for the primacy of capital cities is by Ades and Glaeser (1995). They posit that rulers stay in power by buying off potential revolutionaries (capital residents) by transferring resources from the hinterland to the urban masses. With a simple spatial equilibrium setup this leads then to more in-

---

[2]This is closely related to the study of primate (i.e., largest) cities (many papers treat capitals and primate cities interchangeably as they are so highly correlated). More generally this research focuses on understanding why some countries have a high proportion of their urban population concentrated in a few cities. See for example, Krugman and Elizondo (1996) who develop a trade-

migration into the capital. They theorize, and find evidence to support in cross-section, that primacy is more likely in dictatorships and unstable democracies. Another reason for capital-city bias in public expenditures is put forward by Galiani and Kim (2008). They collect more extensive information on cities in the US and Latin America. In cross-sectional regressions of population on capital city status (both provincial and country) for 1900 and 1990 separately, they show that capital city status is more important for city size in Latin America than in the US. They posit that this could be due to a bias in the provision of public goods to capital cities which could be caused by less political decentralization in Latin America.

A second branch of related literature investigates at how persistent are shocks to cities as compared to their long-run fundamentals. Davis and Weinstein (2002) find that Japanese cities rebounded quickly after WWII bombings. Studies of other wars have found similar results (Brakman et al., 2004; Paskoff, 2008; Miguel and Roland, 2011). On the other side, Bleakley and Lin (2012) examine river portage sites in the US and find that a century after their natural advantage disappeared they are still larger than comparison areas and this effect in levels did not diminish over time. Long-run effects are also found in regards to government programs such as the Tennessee Valley Authority (Kline and Moretti, 2013) and the expansion of the railroad network (Donaldson and Hornbeck, 2013).

Finally, there are papers that examine at the effects of the locations of capital cities. Campante and Do (2014) find, using IV estimates, that isolated US state

---

based economic geography model of urban concentration, Davis and Henderson (2003) who note correlates of urban concentration, and Soo (2005) who models the city sizes as a Pareto distribution ($\ln size_i = const + \beta \cdot \ln rank_i$) and tests Zipf's law by testing if $\beta = -1$.

capital cities have more federal corruption cases per resident and have more money in politics. They believe this is due to the fact that isolation is also associated with less newspaper coverage, voter knowledge, and voter turnout. Campante et al. (2014) focus on the isolation of country capitals and find that their isolation is associated with: misgovernance in autocracies, less power sharing, a larger income premium enjoyed by capital city inhabitants, and lower levels of military spending by ruling elites. Morten and Oliveira (2014) use the same Brazilian capital city change to analyze the effect of migration costs on inter-regional real-wage differences. They use the set of new roads built connecting Brasília to state capitals to identify a structural spatial equilibrium model. They estimate their model on 131 "meso-regions" using data from 1980-2000 and find that migration costs substantially inhibit migration.

2.2.2   Brazilian Context

I use the Brazilian capital relocation for several reasons. First, it transferred all capital functions, which is the most common type of relocation. Second, the distance to the new location was substantial, which helps separate the effects on the original and final locations. Third, the timing of the move did not coincide with any major border change in Brazil, which could have change the importance of the capital city. Finally, Brazil has high quality data on local population and GDP available both before and after the move for all domestic cities. All of the other recent capital relocations (Table 2.1 lists 29 relocations since 1950) lack one or more of these features.

---

[3]Brazil's states are completely divided into municipalities.

Before the 1960 relocation, the capital of Brazil was the municipality[3] of Rio de Janeiro (which today is the capital of the state of Rio de Janeiro). Rio de Janeiro became the capital in 1763 when it was moved from Salvador. The plan to move the capital from Rio de Janeiro to a more interior location was originally conceived in 1827 by José Bonifácio, an adviser to Emperor Pedro I. He presented a plan for a new city called Brasília. The constitutions of 1891, 1934 and 1946 all stated that the capital should be moved to a place closer to the center of the country. The reasons mentioned for this move (Morten and Oliveira, 2014) included regional development, promoting nationalism,[4] and removing the capital from Rio de Janeiro because it was crowded, too "international", received too much national attention, and as a port was vulnerable to military attacks by sea.

The decision to move the capital in 1960 had much to do with the particulars of Juscelino Kubitschek, the president from 1956 to 1961. According to Epstein (1973), Juscelino admitted that when he was campaigning for the presidency he had not thought thoroughly about the issue until a spectator asked whether he would move the capital. He replied that he would. He stuck with that campaign pledge and ordered the construction of Brasília when he became president. Juscelino might also have been inclined to the idea as he previously was the governor of the state of Minas Gerais, which had previously constructed the planned capital city of Belo Horizonte.[5]

Construction on Brasília began in 1956 and the new capital was inaugurated on April 21, 1960 (Epstein, 1973). For my analyses I consider the treatment date as 1956

---

[4]For example, the border state Acre was originally the territory of Bolivia, but was settled by Brazilians who resisted Bolivian control and in 1903 was officially given to Brazil.

[5]The state of Goiás also began constructing the city of Goiânia in 1933 which became the capital

56

(the main results are robust to using the inauguration date of 1960). It became the only municipality of the new Federal District which was created from land removed from the state of Goiás. Brasília is also the name of the central municipal sub-district, but I use the term to refer to the entire municipality. In 1960 the city was far from complete. Many federal agencies stayed in Rio de Janeiro for some time. The Ministry of External Relations, for example, did not move until 1970, with foreign embassies moving even later (Mendes, 1995, p. 27).

The stability and decentralization of Brazil are indicative of what the theories of Ades and Glaeser (1995) and Galiani and Kim (2008) would predict for Brazil's capital. The government of Brazil during this period was quite unstable. Brazil was a constitutional indigenous monarchy from 1822-1889. After a military coup in 1889 it was nominally a constitutional democracy until 1930 but was de facto controlled by oligarchs of the dominant São Paulo and Minas Gerais states. A military junta took power and controlled the country from 1930-1945. A bloodless military coup ushered in a democratic regime from 1945-64. A military dictatorship again took over and ruled between 1964–85. Finally in 1985 Brazil became a democracy. Ades and Glaeser (1995) counted Brazil as an unstable democracy in their earliest data period of 1960-1964 (their earliest data for dictatorship does not begin until 1972) so by their theory the capital status should be contributing to its size.

For the level of decentralization in Brazil, I use the World Bank's Fiscal Decentralization Indicators' percentage of government expenditures that happen below the national level. Brazil first appears in the international data in 1980 with 32% of

---

of Goiás in 1937.

its expenditures being sub-national. Of the 56 countries that have data, Brazil was more decentralized than 43. Using data from Brazil (that is not directly comparable) it appears that this percentage of subnational expenditures was stable from 1956 till 1980. Given the theory suggested in Galiani and Kim (2008), this high level of decentralization could imply that their might be little population advantage to being the capital of Brazil.

Finally, I should note that while I estimate the effects of the relocation, data on the costs of the relocation are hard to determine. Estimates for the cost of building Brasília vary as formal records were not well kept. While the government raised some revenue through the sale of residential lots in Brasília, a typical estimate for the government's spending is 2-3% of GDP over the period 1956-1960 (250-300 billion Cruzeiros or 400-600 million USD in 1960) according to Gordon (2006).

2.3 Empirical Strategy

I estimate the effect that the relocation of the Brazilian capital in 1960 had on the population and GDP of Rio de Janeiro (the old capital) and Brasília (the new capital). In a standard spatial equilibrium model (e.g., Moretti 2011), population is an aggregate measure of the desirability of a locality, encompassing both economic productivity and non-productive amenities. While GDP does not account for non-productive amenities, the GDP data differentiate between private and government sector GDP. This enables me to estimate the effect that the capital relocation had on purely private sector economic activity.

In this setting, identifying causal effects is complicated by the fact that the capital is relocated only once. I therefore use the synthetic controls methodology of Abadie et al. (2010) to generate counterfactual outcomes for the two "treated" localities. The estimated treatment effects are inclusive of any other policies that occurred at the same time as the capital relocation and that differentially affected Rio de Janeiro or Brasília. For instance, they include the effect of new highways that were built to connect the new capital of Brasília to nearby cities. These types of infrastructure investments often accompany capital relocations so they should be thought of as part of the broader program of a capital relocation.

2.3.1   Synthetic Controls

To create counterfactual outcomes for Rio de Janeiro and Brasília I use a synthetic controls strategy of Abadie et al. (2010). Similar to a difference-in-difference design, synthetic controls exploits the difference in treatment and untreated units across the event of interest. However, in contrast to a difference-in-differences design, synthetic controls does not give all untreated units the same weight in the comparison. Instead, it generates a weighted average of the untreated units that closely matches the treated unit over the pre-treatment period. Outcomes for this synthetic control are then projected into the post-treatment period using the weights identified from the pre-treatment comparison. This projection is used as the counterfactual for the treated unit. The identifying assumption is that the weighted average is a valid counterfactual for the post-treatment treated unit.

One data-generating process which satisfies the identifying assumption is the following model of unobserved factors. Let D be an indicator for treatment and the observed outcome variable Y_{jt} (for unit j and time t) is the sum of a time-varying treatment effect $\alpha_{jt}D_{jt}$ and the no-treatment counterfactual Y_{jt}^N, which is specified using a factor model

$$Y_{jt} = \alpha_{jt}D_{jt} + Y_{jt}^N = \alpha_{jt}D_{jt} + (\delta_t + \boldsymbol{\theta}_t\boldsymbol{Z}_j + \boldsymbol{\lambda}_t\boldsymbol{\mu}_j + \varepsilon_{jt}) \tag{2.1}$$

where δ_t is an unknown time factor, $\boldsymbol{Z}_j$ is a $(r \times 1)$ vector of observed covariates unaffected by treatment, $\boldsymbol{\theta}_t$ is a $(1 \times r)$ vector of unknown parameters, $\boldsymbol{\lambda}_t$ is a $(1 \times F)$ vector of unknown factors, $\boldsymbol{\mu}_j$ is a $(F \times 1)$ vector of unknown factor loadings, and the error ε_{jt} is independent across units and time with zero mean. Letting the first unit be the treated unit, the treatment effect is estimated by approximating the unknown Y_{1t}^N with a weighted average of untreated units

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j \geq 2} w_j Y_{jt} \tag{2.2}$$

In my application, the outcome $\boldsymbol{Y}$ is a measure of local population or GDP.

Equation 2.1 simplifies to the traditional fixed effect equation if $\lambda_t\mu_j = \mu_j$. The fixed effect model allows for any unobserved heterogeneity that is time-invariant. The factor model employed by synthetic controls also allows for the existence of non-parallel trends between treated and untreated units after controlling for observables. When dealing with aggregate data as I do, ε_{jt} captures specification error rather than the

standard sampling error. More specifically, ε_{jt} measures the inability of the model to match a "synthetic" control to the treated unit.

2.3.1.1 Estimation

Let T_0 be the number of pre-treatment periods of the T total periods. Index units $\{1, ..., J + 1\}$ such that the first unit is the treated unit and the others are "donors". Let $\boldsymbol{Y}_j$ be $(T \times 1)$ vector of outcomes for unit j and $\boldsymbol{Y}_0$ be the $(T \times J)$ matrix of outcomes for all donors. Let $\boldsymbol{W}$ be a $(J \times 1)$ observation-weight matrix $(w_2, w_3, ..., w_{J+1})'$ where $\sum_{j=2}^{J+1} w_j = 1$ and $w_j \geq 0$ $\forall j \in \{2, ..., J + 1\}$. A weighted average of donors over the outcome is constructed as $\boldsymbol{Y}_0 \boldsymbol{W}$. Partition the outcome into pre-treatment and post-treatment vectors $\boldsymbol{Y}_j = (\overleftarrow{\boldsymbol{Y}}_j \backslash \overrightarrow{\boldsymbol{Y}}_j)$. Let $\boldsymbol{X}$ represent a set of k pre-treatment characteristics ("predictors"). This includes $\boldsymbol{Z}$ (the observed covariates above) and M linear combinations of $\overleftarrow{\boldsymbol{Y}}$ so that $k = r + M$. Analogously, let $\boldsymbol{X}_0$ be the $(k \times J)$ matrix of donor predictors. Let $\boldsymbol{V}$ be a $(k \times k)$ variable-weight matrix indicating the relative significance of the predictor variables.

Given $\boldsymbol{Y}$ and $\boldsymbol{X}$, estimation of synthetic controls consists of finding the optimal weighting matrices $\boldsymbol{W}$ and $\boldsymbol{V}$. Following Abadie et al. (2010), I pick $\boldsymbol{V}$ to minimize the prediction error of the pre-treatment outcome between the treated unit and the synthetic control. Define distance measures $\|\boldsymbol{A}\|_{\boldsymbol{B}} = \sqrt{\boldsymbol{A}'\boldsymbol{B}\boldsymbol{A}}$ and $\|\boldsymbol{A}\| = \sqrt{\boldsymbol{A}'\text{cols}(\boldsymbol{A})^{-1}\boldsymbol{A}}$. $\left\|\overleftarrow{\boldsymbol{Y}}_1 - \overleftarrow{\boldsymbol{Y}}_0 \boldsymbol{W}\right\|$ is then the pre-treatment root mean squared prediction error (RMSPE) with a given weighted average of the control units. $\boldsymbol{W}$ is picked to minimize the RMSPE of the predictor variables, $\|\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}\|_{\boldsymbol{V}}$. In

this way, the treated unit and its synthetic control are similar along dimensions that matter for predicting pre-treatment outcomes. Abadie et al. (2010) suggest nesting the optimization problems where the inner problem is finding $\boldsymbol{W}(\boldsymbol{V})$

$$\boldsymbol{W}(\boldsymbol{V}) = \arg\min_{\boldsymbol{W}} \|\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}\|_{\boldsymbol{V}} \tag{2.3}$$

Then, armed with the optimal $\boldsymbol{W}(\boldsymbol{V})$, I find the optimal $\boldsymbol{V}$ such that

$$\boldsymbol{V} = \arg\min_{\boldsymbol{V}} \left\| \overleftarrow{\boldsymbol{Y}}_1 - \overleftarrow{\boldsymbol{Y}}_0 \boldsymbol{W}(\boldsymbol{V}) \right\| \tag{2.4}$$

The variables in $\boldsymbol{X}$ should be chosen to be good predictors. Including too many variables in $\boldsymbol{X}$ may cause the model to be over fit, predicting well in the pre-treatment periods while predicting poorly in the out-of-sample post-treatment periods. While Abadie et al. (2010) do not address this directly, Dube and Zipperer (2015) use a cross-validation approach to address this model selection problem. To implement this, for a given predictor set $\boldsymbol{X}^l$, identify how well the synthetic control estimation estimates the actual path for the donors. Synthetic controls should ideally predict the actual path of the donors quite well as they are assumed to be unaffected by treatment. For each donor $j \geq 2$, I estimate a donor specific $\boldsymbol{V}_j(\boldsymbol{X}^l)$ and $\boldsymbol{W}_j(\boldsymbol{X}^l)$ that match unit j to the other $(J-1)$ units in the donor pool. I then measure $\left\| \vec{\boldsymbol{Y}}_j - \vec{\boldsymbol{Y}}_{0,-j} \boldsymbol{W}_j(\boldsymbol{X}^l) \right\|$, where $\vec{\boldsymbol{Y}}_{0,-j}$ is the matrix of post-treatment outcomes for all donors except unit j. For possible predictor sets $\{\boldsymbol{X}^1, ..., \boldsymbol{X}^L\}$, I find the $\boldsymbol{X}^l$ that has

the lowest RMSPE in the post-treatment periods when estimating synthetic controls on all the donors,

$$\boldsymbol{X}^* = \arg \min_{\boldsymbol{X}^l \in \{\boldsymbol{X}^1, \dots, \boldsymbol{X}^L\}} \sum_{j=2}^{J+1} \left\| \vec{\boldsymbol{Y}}_j - \vec{\boldsymbol{Y}}_{0,-j} \boldsymbol{W}_j(\boldsymbol{X}^l) \right\| \tag{2.5}$$

The estimated treatment effect is then the difference between the treated unit and the synthetic control during the post-treatment periods. I can look at either a single period difference, $Y_{1t} - \boldsymbol{Y}_{0t}\boldsymbol{W}$ for $t > T_0$, or the RMSPE over all post-treatment years, $\left\| \vec{\boldsymbol{Y}}_1 - \vec{\boldsymbol{Y}}_0 \boldsymbol{W} \right\|$.

If weights can be found such that the synthetic control matches the treated unit in the pre-treatment period:

$$\overleftarrow{\boldsymbol{Y}}_0 \boldsymbol{W} = \overleftarrow{\boldsymbol{Y}}_1 \quad \& \quad \boldsymbol{Z}_0 \boldsymbol{W} = \boldsymbol{Z}_1 \tag{2.6}$$

and $\sum_{t=1}^{T_0} \boldsymbol{\lambda}_t' \boldsymbol{\lambda}_t$ is non-singular, then $\hat{\boldsymbol{\alpha}}_1$ has a bias that goes to zero as the number of pre-intervention periods grows large relative to the scale of the ε_{jt}.

2.3.1.2 Inference

After estimation, significance is determined by running permutation tests. Given an estimated effect $\hat{\alpha}_1$, I can characterize the null-distribution $\{\hat{\alpha}_p\}$ by performing the same synthetic control estimation on the donors. If estimating placebo effects on the donors yields many "effects" as large as $\hat{\alpha}_1$, then it is likely that $\hat{\alpha}_1$ was observed by

chance. This non-parametric test has the advantage of not imposing any distribution on the errors.

The quality of the pre-treatment match in each permutation may vary significantly. If the match is poor, then the post-treatment prediction error is also likely to be large. Therefore, when comparing $\hat{\alpha}_1$ to $\{\hat{\alpha}_p\}$, I take match quality into account. Let s_j be the RMSPE for pre-treatment years when estimating synthetic controls for unit j, so $s_1 = \left\| \overleftarrow{\boldsymbol{Y}}_1 - \overleftarrow{\boldsymbol{Y}}_0 \boldsymbol{W} \right\|$ and $s_{j \geq 2} = \left\| \overleftarrow{\boldsymbol{Y}}_j - \overleftarrow{\boldsymbol{Y}}_{0,-j} \boldsymbol{W}_j \right\|$. There are two ways to take match quality into account.

1. Limit the comparison distribution to only those cases p where $s_p \leq z s_1$ for some cut-off value z. Abadie et al. (2010) list statistics for $z \in \{1, 5, \infty\}$ where $z = \infty$ signifies comparing $\hat{\alpha}_1$ to $\{\hat{\alpha}_p\} = \{\hat{\alpha}_j\}_{j=2}^{J+1}$ without regard to pre-treatment match quality.

2. Construct a t-like statistic, $\tau = \hat{\alpha}/s$ so that effects are weighted by their match quality and then compare τ_1 to $\{\tau_p\}$.

In practice, neither adjustment affects the results.

Confidence intervals for the main effect $\hat{\alpha}_1$ are constructed of all points $\{\alpha_c\}$ such that the test of $H_0 : \alpha_0 = \alpha_c$ against $H_1 : \alpha_0 \neq \alpha_c$ is not rejected at the given significance level γ. For scalar values I employ a two-side test and reject the null hypothesis if $(\hat{\alpha}_1 - \alpha_c)$ is in extreme γ-proportion of $\{\|\hat{\alpha}_p\|\}_p$. Since the p-value changes discretely, this essentially trims the extremes of the null distribution and then fits the width around the main effect. That is $CI_{1-\gamma}^{\alpha} = \{\hat{\alpha}_1 - \hat{\alpha}_p | q(\|\hat{\alpha}_p\|) < 1 - \gamma\}$ where $q(.)$ is the quantile function. If $\dim(\alpha) = K > 1$ then I take into account the variance-

covariance of $\{\hat{\alpha}_p\}$ so that $\|\hat{\alpha}_p\| = \sqrt{(\hat{\alpha}_p - \bar{\alpha})'\hat{\Omega}^{-1}(\hat{\alpha}_p - \bar{\alpha})}$. To calculate the confidence interval for functions of estimated parameters I follow the procedure of Woutersen and Ham (2013). For a function $f(\alpha)$, a function of directly estimated quantities, I apply $f(.)$ to $CI_{1-\gamma}^{\alpha}$. I use functions of estimated parameters in two cases: taking the ratios of estimated effects for two industries and taking the sum of population effects between pairs of cities. In the first one, $\dim(\alpha) = 2$, and $CI_{1-\gamma}^{f(\alpha)} = \{\hat{\alpha}_{p,1}/\hat{\alpha}_{p,2}|\hat{\alpha}_p \in CI_{1-\gamma}^{\alpha}\}$. In the second case $\dim(\alpha) = 1$ and $CI_{1-\gamma} = \{\hat{\alpha}_p + \hat{\alpha}_{p'}|\hat{\alpha}_p, \hat{\alpha}_{p'} \in CI_{1-\gamma}^{\alpha}\}$.

2.3.1.3   Interpolation Bias

If $\boldsymbol{Y}$ is a highly nonlinear function of $\boldsymbol{X}$ and the support of $\boldsymbol{X}$ is large, then there may be interpolation bias so that $\boldsymbol{W}\boldsymbol{Y}_0$ is not be a good approximation for the outcome of a hypothetical unit with characteristics $\boldsymbol{W}\boldsymbol{X}_0$. More formally, interpolation bias implies that $\sum w_j \boldsymbol{Y}(\boldsymbol{X}_j; \cdot) \not\approx \boldsymbol{Y}(\sum w_j \boldsymbol{X}_j; \cdot)$. Abadie et al. (2010) suggest limiting the donor pool to those with similar $\boldsymbol{X}$s. With city growth, a major concern is size-related congestion. To deal with this possibility, I explore limiting the donor pool to localities of similar sizes. I pick the optimal size bandwidth by cross-validation.

2.3.1.4   Identifying Uncontaminated Donors

A key assumption above is that the ε_{jt} are independent across units. This is a version of SUTVA. However, when considering population flows, a treatment shock that increases population in one locality usually does so by increasing migration. This increase in migration must reflect a decrease in population in another unit. These

other, "contaminated" units could detrimentally affect the estimation in two ways. First, if these units are used in the synthetic control for one of the treated units, then it can bias the estimated treatment effect. For example, if the localities adjacent to Brasília were (a) picked for the synthetic control of Brasília, and (b) lost population due to emigration to Brasília, then Brasília's synthetic control will be smaller than it should be, biasing the estimate for Brasília upwards. A second way that contaminated donors could affect estimation is that when estimating synthetic controls on the donors, their prediction errors may be different as a result of the treatment. This distorts the null distribution and leads to mistaken inference.

I use two methods to deal with potentially contaminated donors. First, I assume that all localities spatially close to the treated unit are contaminated and therefore unsuitable as donors. To implement this strategy, I remove all localities that are in the same state as the treated unit.

The second method utilizes placebo tests over time and across units. If the null-distribution is distorted by contaminated donors, this can be used to identify them. With sufficient independence between the contaminated donors and the units that compose the synthetic controls for the contaminated donors, then donors that are contaminated will have larger predicted errors in the post-treatment period. I use the distribution of these errors to determine which units to consider as contaminated. Ideally, I would know the distribution of prediction errors for the population in the absence of treatment. Knowing this, I could identify if there are more units with larger-than-predicted errors, which would result in a distribution with fatter tails. Real

predicted errors with absolute size above some threshold defined by the counterfactual distribution are flagged as having a high likelihood of being contaminated.

As I do not know the counterfactual distribution of prediction errors, I approximate it with a new set of placebo tests that are one period earlier.[6] I do this by generating prediction errors while assuming that the treatment affected units in 1950, ten years before the capital was actually moved.

The procedure is:

1. Estimate synthetic controls on all localities assuming that the first post-treatment period is 1950. Collect the prediction errors for 1950.

2. Estimate synthetic controls on all localities assuming that the first post-treatment period is 1960. Collect the prediction errors for 1960.

3. Consider a unit as contaminated if its 1960 prediction error is in the top/bottom $\alpha\%$ of prediction errors for 1950.

4. Re-run the main estimation with these contaminated donors removed.[7]

2.3.2 Data

The municipal population statistics come from the Brazilian Institute of Geography and Statistics (IBGE) national censuses of 1872, 1890, 1900, 1910, 1920, 1940, 1950, 1960, 1970, 1980, 1991, 2000, 2010 and the National Counts of 1996, 2007. The

---

[6]Using the existing 1960 distribution (e.g., considering as treated all units with real prediction errors in the top and bottom 5%) would just order units by the size of the error and not be informative of how many units were really affected.

[7]This will change the synthetic controls estimated for the other donors and therefore change their

international population data I use for a robustness test comes from the Populstat compilation of historic city populations (Lahmeyer, 2006). Principally, the Populstat data is derived from the *Almanach de Gotha* in early years and from national censuses in later years. I remove some Populstat sources where it appears that they had different definitions of city boundaries.

Estimations of local GDP are available from Brazil's Institute of Applied Economic Research (IPEA) for 1920, 1939, 1949, 1959, 1970, and many more recent years. Local government GDP figures are available for the same years except for 1949 and 1959, where only state figures are available. I estimate the government GDP for 1949 using the municipal government GDP per capita in 1939 and adjust for changes in municipal population and state government GDP from 1939 to 1949. To lengthen the matching period I also extrapolate both the total and government municipal figures to 1900 and 1910. I use the 1920 levels per capita and adjust for changing municipal population and federal government expenditures back to 1910 and 1900.

For information on the size of government employment in the municipalities I transcribe the industry employment count data for selected municipalities from the IBGE census reports for 1940, 1950, 1970, and 1980 (the data from 1960 are incompatible). To lengthen the matching period, I also extrapolate these figures to earlier censuses by using earlier population changes. The construction of Brasília was done by the Urbanization Company of New Capital (Novacap) a state enterprise. The census data, however, records these workers in the larger "Industrial Activities"

prediction errors.

68

category. State-level samples from IPUMS, however, classify the workers more finely, so I am able to add construction workers to the public employees for Brasília.

I use a variety of additional data in the pre-treatment stage that could be predictive of future growth. From the IBGE "Cidades" project I use climatic information (yearly rainfall and mean winter and summer temperatures) and latitude and longitude (I use the center point of the of "bounding box" of the maps of each municipality). For a measure of the distance to the sea, I use NASA's Distance From Coast dataset and use the municipality's center point. As a rough measure of amenities I use a list of municipalities that had soccer teams during the pre-treatment period using the list of early soccer team founding dates from RSSSF Brazil (2014). I use municipal area information from IPEA and some individual-level data from the 1960 and 1970 IPUMS 5% samples.

While the data for Brazil is available at the municipal-level, these boundaries have changed over time. To deal with this, IPEA aggregated municipalities to Minimum Comparable Areas (MCAs) such that the MCAs' boundaries remain constant from 1872 to the present. Rio's municipal boundaries have not changed since 1872. However, the MCA for Brasília consists of 21 small, current and former municipalities. In the sparsely-settled West, where new states and municipalities were recently created, some MCAs are quite large in area while still small in population. All municipal-level estimations are done at the MCA level and I refer to them as "localities".

2.4 Results

To conduct the main estimation I first complete the preliminary tasks of picking the optimization parameters for synthetic controls. Then I analyze local population effects on Brasília and Rio de Janeiro along with separate robustness checks. After that I estimate spillovers to neighboring localities. Finally, I estimate the effects on GDP and its components (public and private) for both treated locations.

2.4.1 Picking Optimization Parameters

I use a data-driven approach to select the optimal predictor sets, size bandwidth, and identifying uncontaminated donors.

2.4.1.1 Optimal Predictor Set

I pick the optimal set of predictors using cross validation. One set of predictor variables that is common in the synthetic control literature is all pre-treatment response variables $\overleftarrow{\boldsymbol{Y}}$. When including other pre-treatment characteristics ($\boldsymbol{Z}$), some of the components of $\overleftarrow{\boldsymbol{Y}}$ must be removed otherwise the optimization will put zero weight on the $\boldsymbol{Z}$ (as noted by Dube and Zipperer 2015). For my predictor sets I then use every other period of $\boldsymbol{Y}$ and then additional $\boldsymbol{Z}$ variables. I break my additional predictors into thematic groups: geographic (area, coordinates, distance to sea), climate (rainfall, summer & winter temperatures), economic (GDP per capita in 1920, 1939, 1949 and GDP's public service administration component for 1920, 1939), facilities (football teams), and other population ranks (1950 rank of population and population density).

With these five groups I estimate all possible combinations. The set of predictors that gives the lowest RMSPE for the donors is the set that includes every other predictor variables along with all the additional predictors.

2.4.1.2 Optimal Size Bandwidth

I also select a bandwidth for a locality's rank in the population distribution via cross-validation. This bandwidth is used to help match each locality to other units. I consider possible quantile values $\{1, 0.5, 0.3, 0.25, 0.1\}$ (where 1 indicates there is no such limit and 0.1 indicates that only localities in the same decile may be considered as matchable donors). The bandwidth with the lowest RMSPE for the donors is 0.3, so I use that in all the following population estimations.

2.4.1.3 Identifying Uncontaminated Donors

I test for contaminated donors by comparing the distribution of prediction errors when treating 1950 as the first post-treatment period to the distribution that (correctly) treats 1960 as the first post-treatment period. The errors are only for the first post-treatment period in each case. I shift the predictor variables back one decade when possible so that the predictors have the same relationship to the treatment date in both cases. The distribution of prediction errors for the 1950 placebo-case is shown in Figure 2.1. One can see that there is a single unit, Brasília, with a very large positive prediction error in 1960. Other than this, the distributions are quite similar, despite not scaling either distribution. A Kolmogorov-Smirnov test can not reject that

Figure 2.1: Population Prediction Errors
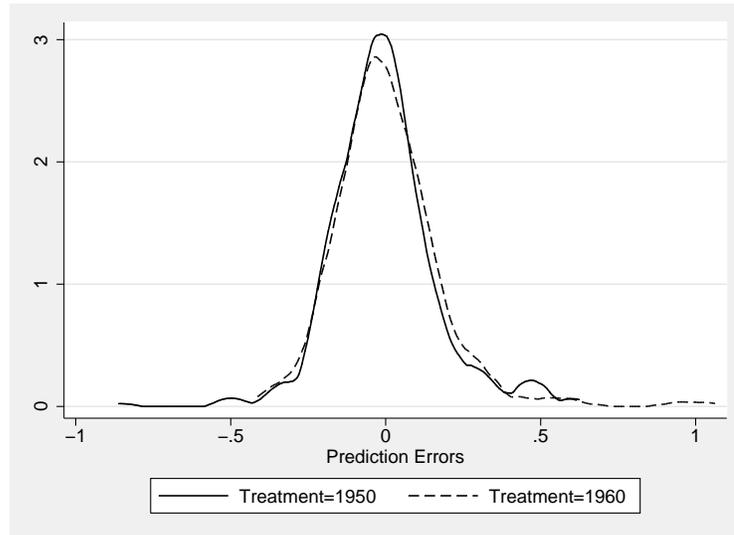
Notes: Kernel density of prediction errors from separate synthetic controls, each using six pre-treatment periods of population.

the two are drawn from the same distribution (p-value=.707). Therefore, it does not appear that there are large general equilibrium effects. The population that migrated to Brasília came likely from Rio de Janeiro and, in insignificant amounts, from the rest of Brazil.

There are three options then for dealing with possibly contaminated donors: (1) drop none, (2) drop the nearby ones (those in the same state), and (3) drop those with 1960 prediction errors above the cut-off levels for the 1950 prediction errors (which in this case are the 25 localities shown in Figure 2.16). As would be suggested by the lack of significant differences between the 1950 and 1960 prediction error distributions, the main results do not vary with the procedure used. I elect to label the results that use option (3) as the Main Results and the others as robustness checks. In particular, the results when using method (2) are shown in Figure 2.15.

2.4.2   Population

There are two details worth noting about how I construct the outcome variable. First, as synthetic controls is essentially a linear estimation method it has smaller prediction errors when matching on log population figures rather than on raw population. Second, synthetic controls match better when the treated unit's pre-treatment data lie in the convex hull of the donor data. This is not always the case for the treated units as Rio de Janeiro is largest locality in Brazil for many years. I therefore demean the log-population figures when matching (I add back in the mean for all figures).[8] When comparing Rio de Janeiro to other national capitals I estimate models both with log population and demeaned log population. In all estimations, the treated unit is never outside the convex hull of the donors for pre-treatment years.[9]

I first use other Brazilian localities as donors to estimate population causal effects. Effects are estimated separately for each "treated" unit (dropping the opposing unit from the donor list). I then perform robustness checks for each treated unit. Finally, I check for spill-overs in neighboring localities.

2.4.2.1   Domestic Comparisons

My preferred specification, is to use the other localities in Brazil as potential donors. It allows me to hold constant many national-level effects. This provides many

---

[8]Assuming $\boldsymbol{Y}_0$ is full rank, demeaning is equivalent to matching on the original data but allowing unrestricted match weights.

[9]I also disregard two other measures common used in the literature. I do not consider population density as Rio has the maximum value in several periods. I also do not consider the urban population in each locality as this distinction was only began in 1940, leaving only two pre-treatment periods.

consistent pre-treatment variables to match on. It also offers many donor units to improve the inference from permutation tests.

The results when matching on demeaned log population are shown in Figure 2.2. For each estimation I show 95% confidence intervals around the estimated counterfactual using 376 permutations. Confidence intervals are shown using the unrestricted set of donor prediction errors. Significance levels, for each year and for different null-distributions are shown in Section 2.8. They do not, however, change the graphical interpretations. Brasília experiences a significant and substantial increase in population for the whole post-treatment period. While the effect on Brasília is large in percentage terms, it was not a significant driver of national population movement. The estimated effect on Brasília's 1960 population was .2% of total Brazilian population and 1.2% of residents who had migrated in the last decade.

Rio de Janeiro experienced a negative but insignificant effect for the whole post-treatment period. In 1950 it had a population of 2.4 million. Its population increased to 3.3 million in 1960 and 6.3 million in 2010. For Rio de Janeiro to have experienced a significant negative impact, its population would have to have remained below 2.9 million in 1960 and 3.5 million in 2010.

The variable weights ($\boldsymbol{V}$), the similarity between $\boldsymbol{X}_1$ and $\boldsymbol{W}\boldsymbol{X}_0$, and the top locality matches ($\boldsymbol{W}$) for the estimation are shown in Appendix 2.10. The pre-treatment population variables are most predictive of future population with minor weight given to summer temperatures and latitude. The treated units match their synthetic controls quite well for variables that have high predictive value. The highest weight any donor locality receives in the synthetic control is .35 which suggests the

74

Figure 2.2: Estimated Population Effects

(a) Rio de Janeiro

Notes: Predictors: response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 376 permutation tests. 68% of the permutation tests had lower pre-treatment RMSPEs.

(b) Brasília

Notes: Predictors: response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 376 permutation tests. 11% of the permutation tests had lower pre-treatment RMSPEs.

results are not driven by a single donor. A map of the estimated effects on the donors is shown in Figure 2.13.

A final check is to see if the procedure does good job of predicting the path for Rio de Janeiro and Brasília for the latter part of the pre-treatment period matching on just earlier data from the pre-treatment period.. I do this by reserving the last pre-treatment period as a validation period. The prediction errors for the last pre-treatment period are insignificant (see tables 2.4 and 2.5) implying that the procedure does well at matching the treated units in general. Matching on growth-rates instead of log population presents similar results and are show in Figure 2.12.

The new residents of Brasília came from all over Brazil, though the largest concentration came from nearby states. This is shown in Table 2.11 using census micro data to identify the state-to-state migration flows for 1960 and 1970 (municipal-level dataa are not in the sample extract until 1980). Similarly, Rio de Janeiro sent most emigrants to its two neighboring states (Brasília is third on the list). Distinguishing how much of Brasília's rise is due to a Rio de Janeiro's loss is difficult. While there was some direct migration it is difficult to say if there were diverted migration flows (Brasília receiving migrants that would have otherwise gone to Rio de Janeiro). In Figure 2.14 I show the sum of the effects for Brasília and Rio de Janeiro. The results are close to zero and consistently insignificant. As I argue later, this implies that we should view the government as merely shifting a fixed quantity of resources between the two locations.

Figure 2.3: Population - Tupaciguara
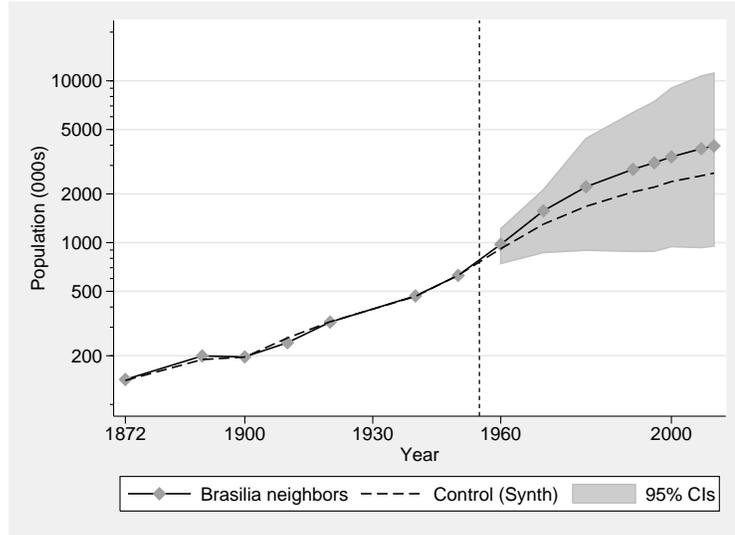
2.4.2.2   Population Robustness

For Brasília, one scenario that would confound interpretation of the effect is if the planning commission picked Brasília because they expected it to grow quickly in the future compared to other similar locations. The planning commission report to the Brazilian senate that suggested the current site of Brasília also noted the alternative site of Tupaciguara. It is similar in terms of geographic factors. It also borders Goiás and Minas Gerais. Brasília's current site was selected by the congress because it was the farther North and there are more Northern states represented in the senate Epstein (1973). If Brasília was positively selected then one might expect the same positive selection to be exist for Tupaciguara. From Figure 2.3, however, one can see

that Tupaciguara did not grow faster than its history would lead us to believe. This suggests that Brasília was unlikely to be positively selected.

For Rio de Janeiro, I perform a robustness test against other country capital cities. In this setup all the donors share the same pre-treatment characteristic that they are national capitals. I identify 37 separate countries that were independent since 1900 and that have consistent historical capital city population data in the Populstat data.[10] The data vary between cities. On average the first data point is in 1850 and they have 16 population statistics before 1956 and 8 after 1956. To make the data comparable, I interpolate using cubic splines and then sample every 10 years from 1880 to 1990. As can be seen in Figure 2.4, the point estimate for the effect on Rio de Janeiro is small and far from significant. While this estimate addresses the concern about some units possibly interacting with each other it does have a few downsides. First, the data come from different sources (and sometimes different sources for the same city over time) so it is noisier. Second, I currently only use population data. Finally, there are fewer units so it is less likely that the synthetic control will be a great match. My preferred estimation for this is matching on demeaned log population as shown in Figure 2.4. The results show a slightly positive but insignificant effect for Rio de Janeiro. The results are similar when matching on log population though the fit is worse.

---

[10]Countries: Argentina, Bolivia, Brazil, Canada, Columbia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Paragua, US, Uruguay, Venezuela, Liberia, Japan, China, Thailand, Sweden, Switzerland, Spain, France, Austria, Belgium, Denmark, Greece, Italy, Luxemborg, Norway, Monaco, Netherlands, Portugal, Romania, Russia

Figure 2.4: Population - Rio de Janeiro (Cross-Country)

(a) Matching on Demeaned Log-Population
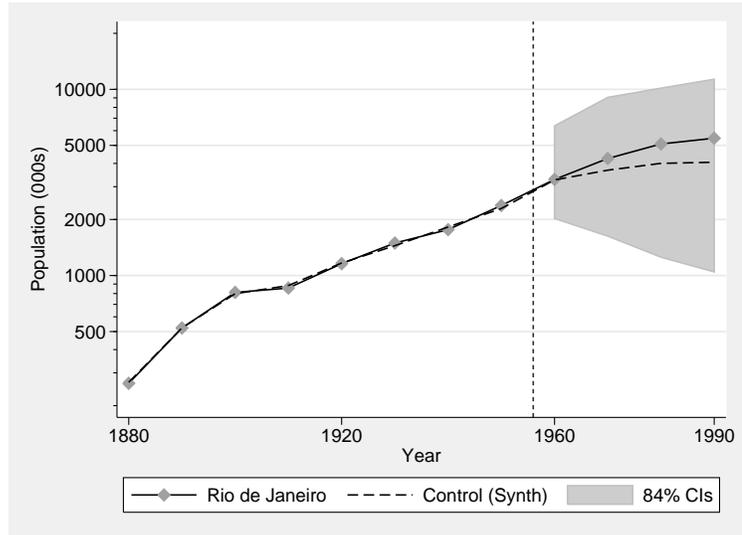
Notes: Predictors: Pre-treatment population. Confidence intervals for control from 37 permutation tests. 29% of the permutation tests had lower pre-treatment RMSPEs.

(b) Matching on Log-Population
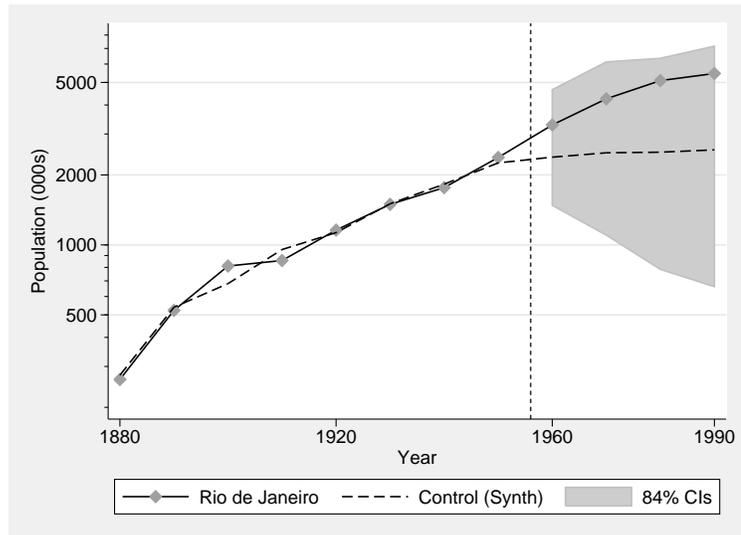
Notes: Predictors: Pre-treatment population. Confidence intervals for control from 37 permutation tests. 42% of the permutation tests had lower pre-treatment RMSPEs.

2.4.2.3 Neighboring Areas

I next estimate spillovers in neighboring localities. I aggregate the localities that border Rio de Janeiro and Brasília into single units and separately estimate their effects. Given the sizes of the localities, the neighbors of Rio de Janeiro can be thought of as part of the larger Rio de Janeiro urban agglomeration. This is not the case for the neighbors of Brasília as they were never part of a larger agglomeration. The results are shown in Figure 2.5. In both cases I find slightly positive, but insignificant effects.

2.4.3 Sectoral Employment

Figure 2.6 shows the effects on public sector employment for Rio de Janeiro and Brasília. Figure 2.7 shows the corresponding figures for private (non-public) employment. In both figures, there is a large and significant effect on Brasília and there is no significant effect on Rio de Janeiro. For Brasília, the ratio of the effect on total employment to the effect on public employees is 2.84 with a 95% confidence interval of (2.59,2.98). This implies that for every additional government job that was created, an addition 1.7 non-public jobs were created. This number, however, should not be viewed as a causal effect of just the government employment as the capital relocation had effects other than shifting government employment. This is on the high end for employment multipliers. The closest would be Moretti (2010) who finds a multiplier of 2.5 for new skilled positions in cities.

Figure 2.5: Effect on Neighbor Populations

(a) Rio de Janeiro's Neighbors
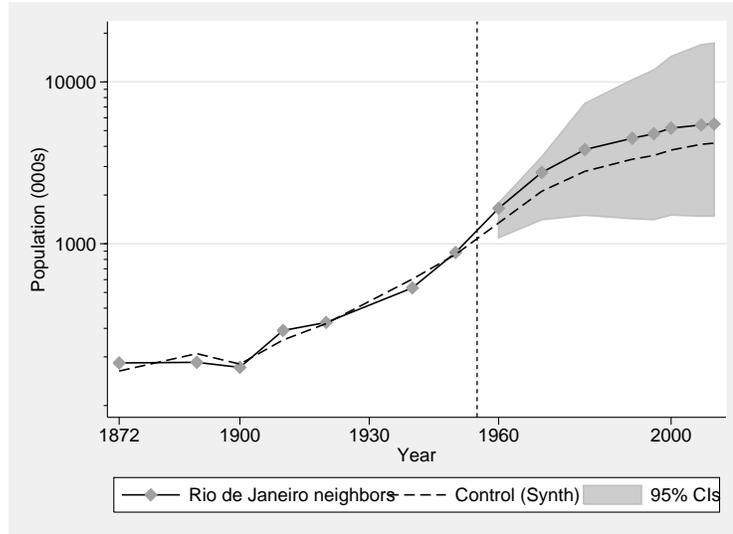
Notes: Predictors: response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 376 permutation tests. 59% of the permutation tests had lower pre-treatment RMSPEs.

(b) Brasília's Neighbors
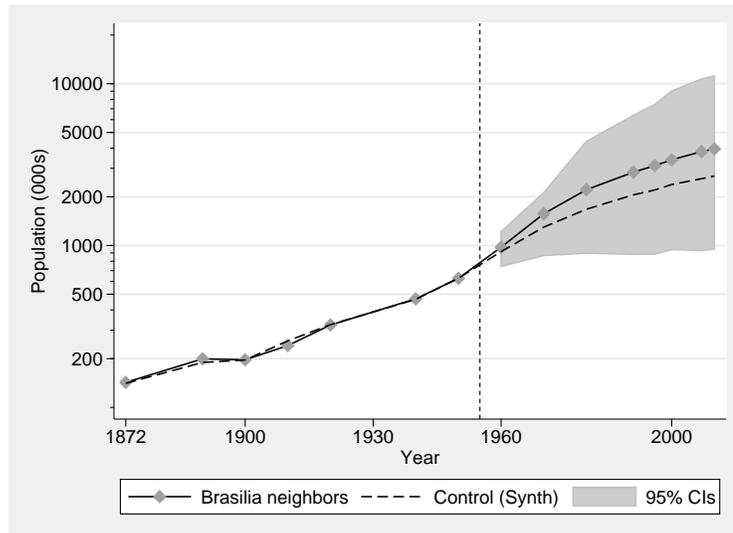
Notes: Predictors: response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 376 permutation tests. 10% of the permutation tests had lower pre-treatment RMSPEs.

One reason that this multiplier is large is that the government intervened in the housing market in Brasília. Novacap built apartment blocks and provided subsides for those living there. This likely controlled one potential channel for congestion by keeping rents down.

2.4.4   Sectoral GDP

Finally, I estimate effects on real municipal GDP. I estimate effects for both total GDP, and its public and private components.

The results in Figures 2.8, 2.9, and 2.10 show the effects for both treated units using total, public, and private GDP. Brasília shows a consistent positive effect in all post-treatment periods and for all GDP measures. Rio de Janeiro shows a negative effect for public GDP that becomes marginally significant after 2004. This is less than convincing as it is 50 years after treatment and no other effects show up in private or total GDP.

For Brasília, the 1970, estimated effects are BRL 1.362 million (all figures are in 2000 BRL) on government GDP, BRL 3.954 million on total GDP, implying an approximate fiscal multiplier of 2.93 with a 95% confidence interval of (2.89,2.95). This should be considered an upper bound on the fiscal multiplier as there were associated actions (new highway construction) that are not counted in the government effect (as they were outside of this territory) but which likely increased total GDP in Brasília. Even so, it is on the high side of existing fiscal multiplier estimates. CBO (2012)

Figure 2.6: Effect on Municipal Public Administration Employment

(a) Rio de Janeiro

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 330 permutation tests.

(b) Brasília

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 330 permutation tests.

83

Figure 2.7: Effect on Municipal Private-Sector Employment

(a) Rio de Janeiro

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 330 permutation tests.

(b) Brasília

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 330 permutation tests.

Figure 2.8: Effect on Municipal GDP

(a) Rio de Janeiro

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950)Confidence intervals for control from 215 permutation tests.

(b) Brasília

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950)Confidence intervals for control from 215 permutation tests.

surveys studies and finds a high estimate of 2.5 for the purchases of goods and services by the federal government in the US.

2.5   Discussion

The 1960 relocation of the Brazilian capital from Rio de Janeiro to Brasília appears to have had a large and significant effect on Brasília. Conversely, the move appears not to have affected Rio de Janeiro to any large extent. However, my estimation strategy estimates differences in proportions, so insignificant effects for Rio de Janeiro could be large in absolute terms when compared to Brasília. While the sum of the effects on total population for Rio de Janeiro and Brasília was very close to zero (Figure 2.14), it was not that the government moved a fixed quantity of resources when it relocated the capital. Figure 2.11[11] shows that the sum of effects on public GDP was positive and significantly different from zero.

It is not surprising that Rio de Janeiro did not suffer a significant decline. There were no significant changes in the public sector employment or GDP for Rio. The shift away from Rio de Janeiro seems to have been gradual enough to not disrupt the local labor markets. Additionally, several recent papers have noted asymmetric responses to shocks. Glaeser and Gyourko (2005) and Henderson and Venables (2009) show that cities lose less population to negative shocks than they gain from positive shocks in settings where there is durable capital in the cities. Rauch (1993) provides

---

[11]I trust the GDP estimates more than the employment estimates are they do not suffer from report bias nor the miscategorization of those indirectly employed by the government.

Figure 2.9: Effect on Municipal Public Administration GDP

(a) Rio de Janeiro

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950)Confidence intervals for control from 215 permutation tests.

(b) Brasília

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950)Confidence intervals for control from 215 permutation tests.

Figure 2.10: Effect on Municipal Private-Sector GDP

(a) Rio de Janeiro

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950)Confidence intervals for control from 215 permutation tests.

(b) Brasília

Notes: Predictors include response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950)Confidence intervals for control from 215 permutation tests.

Figure 2.11: Sum of Public GDP Effects for Rio de Janeiro and Brasília
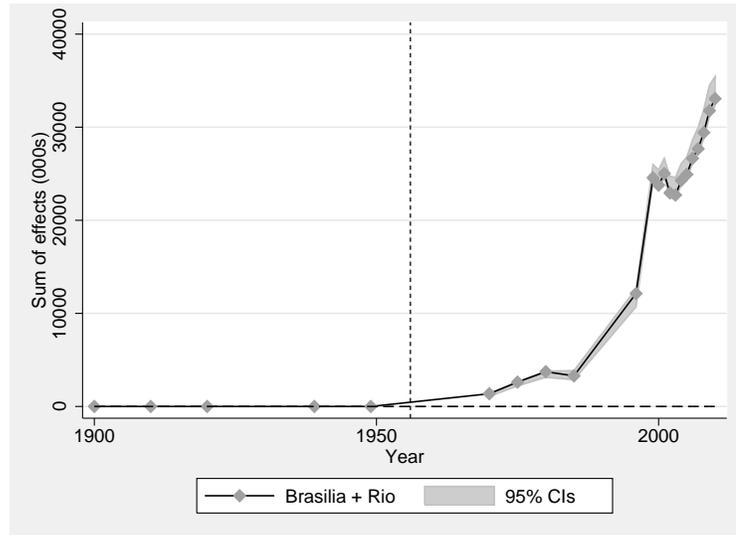
Notes: Confidence intervals for control from 200 permutation tests.

another mechanism for such asymmetry where there are agglomeration economies and there is a first-mover disadvantage for firms leaving a large city.

While I am able to say that the relocation had a large effect on employment, a lack of available data precludes looking at the other main channels of response that one would expect to see in an urban setting. Municipal wage and rent data is not available before the relocation. Even for 1960 and 1970 it is only available at the state-level. Similarly there is no adequate migration data before the relocation. Along with the capital relocation, new radial highways were constructed from Brasília. While these may have lowered transportation-costs between Brasília and state-capital cities, it is unlikely that this was a large driver of Brasília's rise given that effects are seen by 1960 when only small amounts of the new highways were finished (Morten and Oliveira, 2014).

2.6 Conclusion

In this paper I estimate the first causal effects of capital city designation by using the capital relocation-event in Brazil in 1960. I find that receiving the capital designation had large, positive effects on total population; public and private GDP; and public and private employment of Brasília. For Rio de Janeiro, the former capital, I find no significant impact on public sector employment or GDP and consequently no impact on the private sector or total population. The relocation was not a simple transfer of resources, but was accompanied by an increase in public spending pooled over the two municipalities. Likely the speed of withdraw from Rio de Janeiro was slow and its large labor markets facilitated local labor reallocation.

I also propose new tests and procedures for applying synthetic controls to situations where some units may directly influence each other. Hopefully these will increase the use of th newer Synthetic Control Methods in spatial settings.

2.7 Appendix - Capital City Relocations

Table 2.1: Capital City Relocations since 1950

Year	Old	New	Country	ID problems	Notes
1950	Tel Aviv-Jaffo	Jerusalem	Israel	CC	
1954	Trieste	Rome	(Trieste) Italy	CC	
1959	Karachi	Rawalpindi	Pakistan		
1960	(none)	Nouakchott	Mauritania	CC	
1961	*Rio de Janeiro*	*Brasília*	*Brazil*		*PB*
1962	Butare	Kigali	Rwanda	CC	
1962	Ta'izz	Sana'a	North Yemen	CC	
1965	Mafeking	Gaborone	Botswana	CC	
1969	Rawalpindi	Islamabad	Pakistan	MC	PB
1970	Belize City	Belmopan	Belize (colony)		PB, PT
1970	Salalah	Muscat	Oman		PT
1974	Zomba	Lilongwe	Malawi	CC	PT
1974	Madina do Boe	Bissau	Guinea Bissau	CC	
1975	Luang Prabang	Vientiane	Laos		
1976	Baguio		Philippines		PT
1976	Quezon City	Manila	Philippines	MC	
1976	Saigon	Hanoi	(South) Vietnam	CC	
1982	Colombo	Kotte	Sri Lanka	CC, MC	
1983	Abidjan	Yamoussoukro	Cote d'Ivoire		PT
1989	Kolonia	Palikir	F. S. of Micronesia	CC, MC	
1990	Santiago	Valparaíso	Chile		PT
1990	Aden	Sana'a	(South) Yemen	CC	
1991	Lagos	Abuja	Nigeria		PB
1996	Dar es Salaam	Dodoma	Tanzania		PT
1997	Almaty	Astana	Kazakhstan	CC	
1999	Bonn	Berlin	(West) Germany	CC	PT
1999	Kuala Lumpur	Putrajaya	Malaysia	MC	PT
2005	Yangon	Naypyidaw	Myanmar		PB
2006	Koror	Ngerulmud	Palau		PB

Notes: List of national capital city relocations since 1950.
ID - Identification
CC - Country Changed (e.g., major border change) within 10 years. This includes mergers, for example, when Trieste was annexed into Italy. Country changes likely affect the importance of the capital dramatically.
MC - Moved Close (within 25 km)
PT - Partial move of government (only certain functions)
PB - Purpose Built

2.8 Appendix - Permutation Test *P*-values

Table 2.2: *P*-Values for Rio de Janeiro Population

Null dist.	1960	1970	1980	1991	1996	2000	2007	2010	Joint
$\{\hat{\alpha}_p \vert s_p \le s_1\}$	0.55	0.48	0.40	0.34	0.29	0.29	0.28	0.31	0.31
$\{\hat{\alpha}_p \vert s_p \le 5s_1\}$	0.58	0.50	0.43	0.35	0.31	0.31	0.30	0.32	0.34
$\{\hat{\alpha}_p\}$	0.58	0.50	0.43	0.35	0.31	0.31	0.30	0.32	0.34
$\{\tau_p = \hat{\alpha}_p/s_p\}$	0.65	0.60	0.55	0.49	0.46	0.48	0.47	0.49	0.51

Notes: *P*-values from 376 permutation tests. 68% of the permutation tests had lower pre-treatment RMSPEs. Null distribution includes those donors that are matched at least as well as Rio de Janeiro, those matched no more than 5 (or 10) times worse, all donors. The first three rows compare yearly effects while the last row compares effects weighted by pre-treatment match quality. The values correspond to Figure 2.2a.

Table 2.3: *P*-Values for Brasília Population

Null dist.	1960	1970	1980	1991	1996	2000	2007	2010	Joint
$\{\hat{\alpha}_p \vert s_p \le s_1\}$	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
$\{\hat{\alpha}_p \vert s_p \le 5s_1\}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$\{\hat{\alpha}_p\}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$\{\tau_p = \hat{\alpha}_p/s_p\}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Notes: *P*-values from 376 permutation tests. 11% of the permutation tests had lower pre-treatment RMSPEs. Null distribution includes those donors that are matched at least as well as Brasília, those matched no more than 5 (or 10) times worse, all donors. The first three rows compare yearly effects while the last row compares effects weighted by pre-treatment match quality.The values correspond to Figure 2.2b.

Table 2.4: Early Placebo Check: *P*-Values for Rio de Janeiro Population Estimation Using Treatment Date as 1950

Null dist.	Joint
$\{\hat{\alpha}_p | s_p \leq s_1\}$	0.65
$\{\hat{\alpha}_p | s_p \leq 5s_1\}$	0.66
$\{\hat{\alpha}_p\}$	0.66
$\{\tau_p = \hat{\alpha}_p/s_p\}$	0.80

Notes: *P*-values from 376 permutation tests. 82% of the permutation tests had lower pre-treatment RMSPEs. Null distribution includes those donors that are matched at least as well as Rio de Janeiro, those matched no more than 5 (or 10) times worse, all donors. The first three rows compare yearly effects while the last row compares effects weighted by pre-treatment match quality.

Table 2.5: Early Placebo Check: *P*-Values for Brasília Population Estimation Using Treatment Date as 1950

Null dist.	Joint
$\{\hat{\alpha}_p | s_p \leq s_1\}$	0.42
$\{\hat{\alpha}_p | s_p \leq 5s_1\}$	0.44
$\{\hat{\alpha}_p\}$	0.45
$\{\tau_p = \hat{\alpha}_p/s_p\}$	0.38

Notes: *P*-values from 376 permutation tests. 39% of the permutation tests had lower pre-treatment RMSPEs. Null distribution includes those donors that are matched at least as well as Brasília, those matched no more than 5 (or 10) times worse, all donors. The first three rows compare yearly effects while the last row compares effects weighted by pre-treatment match quality.

Figure 2.12: Annualized Growth Rate

(a) Rio de Janeiro

Notes: Predictors: response variables for every pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 382 permutation tests. 90% of the permutation tests had lower pre-treatment RMSPEs.

(b) Brasília

Notes: Predictors: response variables for every pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 382 permutation tests. 61% of the permutation tests had lower pre-treatment RMSPEs.

Table 2.7: *P*-Values for Brasília Population Growth Rates

Null dist.	1960	1970	1980	1991	1996	2000	2007	2010	Joint
$\{\hat{\alpha}_p \mid s_p \leq s_1\}$	0.01	0.01	0.01	0.09	0.08	0.12	0.06	0.40	0.01
$\{\hat{\alpha}_p \mid s_p \leq 5s_1\}$	0.01	0.01	0.01	0.09	0.10	0.12	0.05	0.40	0.01
$\{\hat{\alpha}_p\}$	0.01	0.01	0.01	0.09	0.10	0.12	0.05	0.40	0.01
$\{\tau_p = \hat{\alpha}_p / s_p\}$	0.02	0.02	0.03	0.17	0.17	0.20	0.13	0.48	0.03

Notes: *P*-values from 382 permutation tests. 61% of the permutation tests had lower pre-treatment RMSPEs. Null distribution includes those donors that are matched at least as well as Brasília, those matched no more than 5 (or 10) times worse, all donors. The first three rows compare yearly effects while the last row compares effects weighted by pre-treatment match quality. The values correspond to Figure 2.12.

2.9   Appendix - Growth Rates

Table 2.6: *P*-Values for Rio de Janeiro Population Growth Rates

Null dist.	1960	1970	1980	1991	1996	2000	2007	2010	Joint
$\{\hat{\alpha}_p \mid s_p \leq s_1\}$	0.51	0.24	0.31	0.29	0.42	0.77	0.58	0.88	0.58
$\{\hat{\alpha}_p \mid s_p \leq 5s_1\}$	0.51	0.25	0.31	0.28	0.43	0.77	0.58	0.89	0.58
$\{\hat{\alpha}_p\}$	0.51	0.25	0.31	0.28	0.43	0.77	0.58	0.89	0.58
$\{\tau_p = \hat{\alpha}_p / s_p\}$	0.72	0.51	0.59	0.52	0.68	0.87	0.79	0.93	0.87

Notes: *P*-values from 382 permutation tests. 90% of the permutation tests had lower pre-treatment RMSPEs. Null distribution includes those donors that are matched at least as well as Rio de Janeiro, those matched no more than 5 (or 10) times worse, all donors. The first three rows compare yearly effects while the last row compares effects weighted by pre-treatment match quality. The values correspond to Figure 2.12.

2.10   Appendix - Weights for Matches

The following are the match variables when the response variable is demeaned log-population

Table 2.8: Variable Weights and Balance (Rio de Janeiro)

Variable	Weight	Treated	Control
Log Pop (demeaned)(1950)	0.330	0.980	0.977
Log Pop (demeaned)(1872)	0.257	-1.177	-1.172
Log Pop (demeaned)(1900)	0.143	-0.178	-0.208
Summer temp	0.104	20.542	20.044
Log Pop (demeaned)(1920)	0.068	0.252	0.270
Latitude	0.033	-22.900	-18.964
Per-capita GDP (1920)	0.018	1.206	1.584
Per-capita GDP (1939)	0.015	4.274	6.280
Winter temp	0.008	25.417	24.311
Pc GDP for pub admin (1920)	0.006	0.263	0.121
Per-capita GDP (1949)	0.006	5.388	8.707
Pop density rank (1950)	0.005	0.002	0.173
Longitude	0.003	-43.460	-45.258
Area	0.002	1,167.000	10,615.322
Pc GDP for pub admin (1939)	0.002	0.371	0.457
Distance to the sea	0.002	13.498	206.886
Rainfall	0.000	550.583	548.574
Pop rank (1950)	0.000	0.009	0.126

Notes: List of variables used in the matching procedure that produces the synthetic control for Rio de Janeiro for the main specification where the outcome is demeaned log population. The first column lists the relative weights produced during the matching process. Weight indicates how useful the variable is a predicting the pre-treatment path of the outcome. Columns 2 and 3 show the values for the treated unit and synthetic control. The values should be similar for those variables that have higher weights.

Table 2.9: Variable Weights and Balance (Brasília)

Variable	Weight	Treated	Control
Log Pop (demeaned)(1950)	0.288	0.538	0.542
Log Pop (demeaned)(1872)	0.272	-0.495	-0.510
Log Pop (demeaned)(1900)	0.207	-0.240	-0.237
Summer temp	0.081	21.197	20.891
Log Pop (demeaned)(1920)	0.067	0.182	0.187
Latitude	0.043	-15.386	-14.350
Winter temp	0.020	23.233	23.686
Per-capita GDP (1939)	0.006	2.911	2.776
Longitude	0.004	-47.569	-46.534
Per-capita GDP (1920)	0.004	1.201	0.565
Area	0.003	53,607.000	30,463.245
Pop rank (1950)	0.003	0.364	0.372
Pc GDP for pub admin (1939)	0.002	0.136	0.149
Pop density rank (1950)	0.001	0.955	0.773
Rainfall	0.000	510.047	556.861
Pc GDP for pub admin (1920)	0.000	0.037	0.027
Per-capita GDP (1949)	0.000	1.803	4.264
Distance to the sea	0.000	863.507	309.848

Notes: List of variables used in the matching procedure that produces the synthetic control for Brasília for the main specification where the outcome is demeaned log population. The first column lists the relative weights produced during the matching process. Weight indicates how useful the variable is a predicting the pre-treatment path of the outcome. Columns 2 and 3 show the values for the treated unit and synthetic control. The values should be similar for those variables that have higher weights.

Table 2.10: Top Locality Matches

	Rio de Janeiro			Brasília	
Municipality	Weight	Current pop.	Municipality	Weight	Current pop.
Belo Horizonte	0.35	3,055,786	Patrocínio	0.23	137,788
Jaú	0.26	266,788	Cametá	0.22	239,151
Santos	0.15	918,693	Palmas	0.15	841,270
Barreirinhas	0.14	458,062	São Lourenço	0.14	179,609
São Gonçalo	0.10	1,519,707	Andrelândia	0.12	74,875
			Tremedal	0.08	97,443
			Ipatinga	0.04	567,993
			Catu	0.01	173,339

Notes: List of municipalities with the match weight above .01 used in the creation of the synthetic controls.

Table 2.11: Migration Summary Information

Place	Emigration from RJ % of emigration		Immigration to Brasilia % immigration	
	1960	1970	1960	1970
Rondônia	0.00	0.09	0.00	0.04
Acre	0.00	0.01	0.02	0.06
Amazonas	0.00	0.23	0.00	0.11
Roraima	0.00	0.02	0.00	0.01
Pará	0.00	0.28	0.71	0.41
Amapá	0.00	0.04	0.00	0.08
Maranhão	0.00	0.07	1.56	2.95
Piauí	0.00	0.07	4.02	5.18
Ceará	0.18	0.40	7.48	6.37
Rio Grande do Norte	0.26	0.27	2.03	1.89
Paraíba	0.18	0.39	5.38	5.46
Pernambuco	0.56	1.02	4.00	3.79
Alagoas	0.15	0.11	0.32	0.33
Sergipe	0.11	0.20	0.52	0.33
Bahia	0.67	0.97	6.73	6.60
Minas Gerais	4.70	3.57	15.63	22.06
Espírito Santo	0.00	1.38	1.05	0.98
Rio de Janeiro	80.56	73.17	14.81	13.72
São Paulo	7.23	8.35	7.27	4.75
Paraná	1.76	1.31	0.78	1.06
Santa Catarina	0.00	0.29	0.32	0.47
Rio Grande do Sul	0.65	0.61	0.22	0.58
Mato Grosso	0.23	0.28	0.62	0.85
Goiás	0.25	0.48	26.44	21.51
Distrito Federal	2.51	6.40		
Abroad			0.10	0.38

Source: IPUMS 5% sample.

Notes: Summary information on immigration and emigration for Brazilian states.

2.11   Appendix - Additional Tables and Figures

Figure 2.13: Estimated Population Effects - 1960
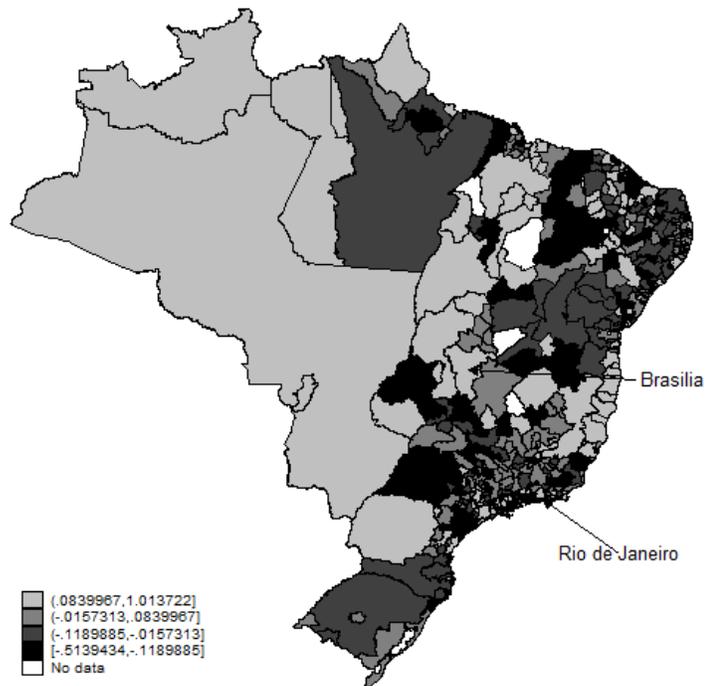
Notes: Estimation differences between each unit and their synthetic control for 1960.Predictors are those from the standard estimations of section 2.4.2.1.

Figure 2.14: Sum of Population Effects for Rio de Janeiro and Brasília
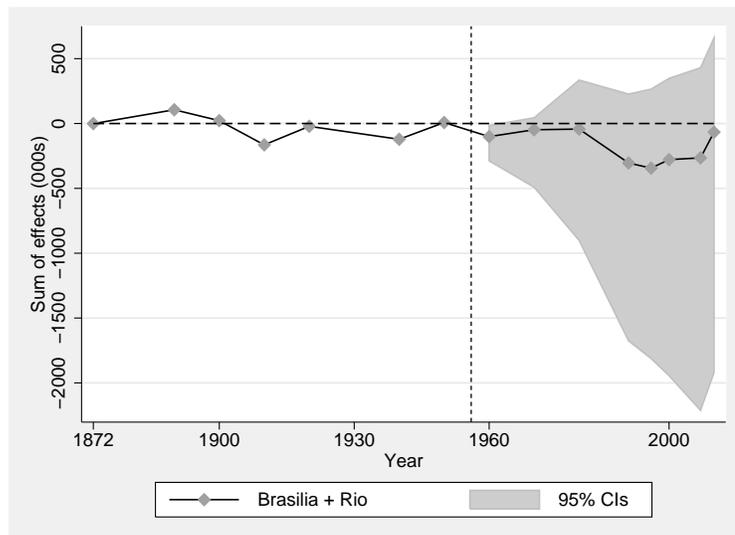
Notes: Confidence intervals for control from 200 permutation tests. The permutation distribution is constructed by drawing random pairs of donor municipalities and summing their effects. Predictors are those from the standard estimations of section 2.4.2.1.

Figure 2.15: Dropping Nearby Localities

(a) Rio de Janeiro

Notes: Predictors: response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 376 permutation tests. 68% of the permutation tests had lower pre-treatment RMSPEs.

(b) Brasília

Notes: Predictors: response variables for every other pre-treatment period, area, distance to sea, latitude, longitude, per capita municipal GDP (1920, 1939, 1949), per capita municipal GDP for public administration (1920, 1939), yearly rainfall, winter & summer temperatures, existence football teams in 1955, population rank (1950), population density rank (1950). Confidence intervals for control from 376 permutation tests. 11% of the permutation tests had lower pre-treatment RMSPEs.

Figure 2.16: Municipalities Dropped
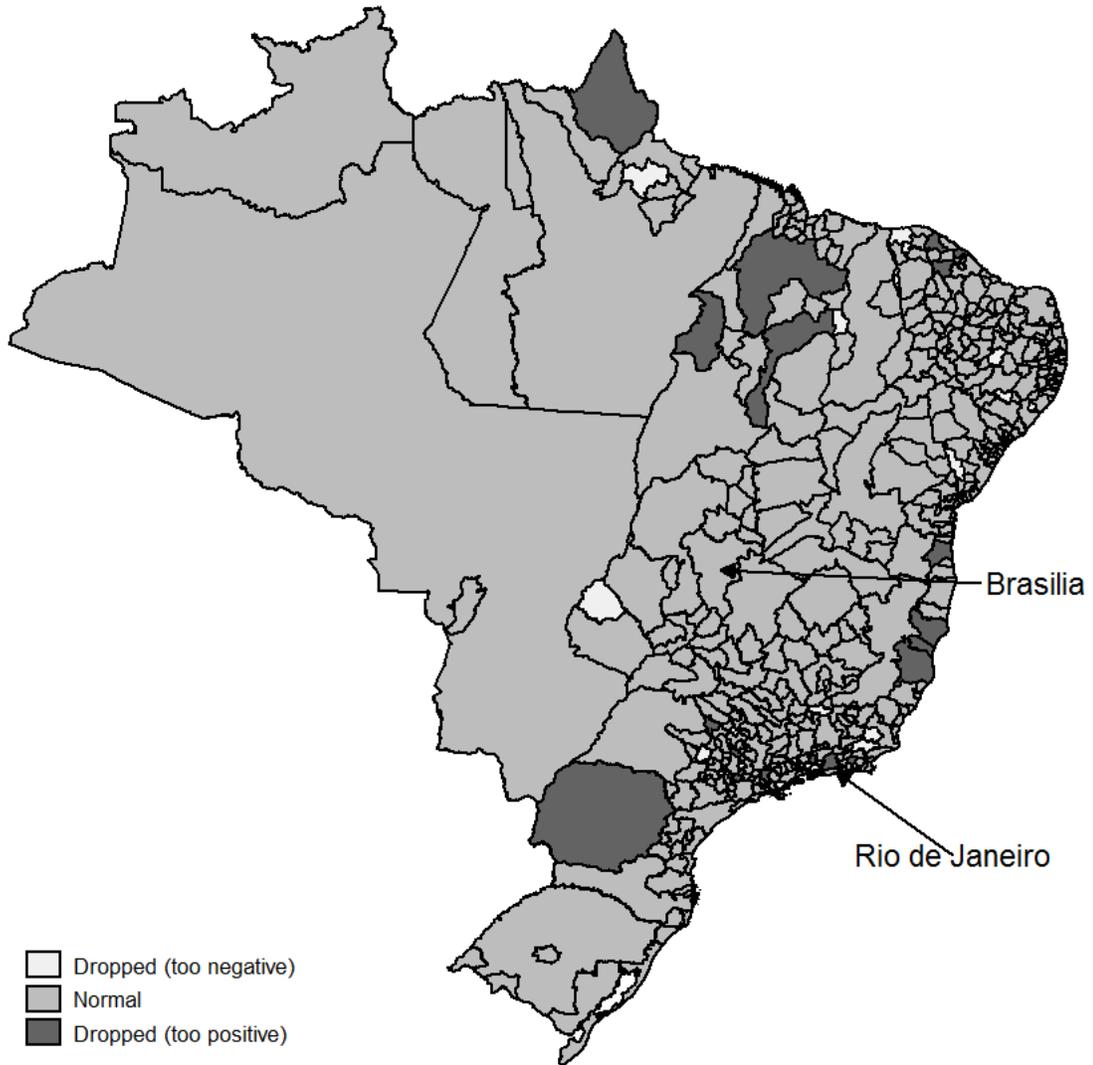
Notes: Municipalities dropped due to possibly contamination. These municipalities have prediction errors when estimating with treatment in 1960 for their first treatment year in the critical region of the distribution of such errors when estimating with treatment in 1950.

Chapter 3:   Credit Constraints, Discounting and Investment in Health:

Evidence from Micropayments for Clean Water in Dhaka

3.1   Introduction

Low rates of adoption of and low willingness to pay (WTP) for preventative health technologies pose an ongoing puzzle in development economics (Dupas, 2011; Abdul Latif Jameel Poverty Action Lab, 2011). In the case of water-borne diseases, the burden is high both in terms of poor health and cost of treatment, and inexpensive preventative technologies are available, but WTP for products such as chlorine treatment or ceramic filters has been observed to be low in a number of contexts (Ahuja et al., 2010; Ashraf et al., 2010; Luoto et al., 2011; Berry et al., 2015; Guiteras et al., 2016).

Many explanations for this puzzle have been proposed, including lack of information, difficulties in learning returns, and inconvenience, effort costs or other non-health, non-financial disutilites associated with use (Foster and Rosenzweig, 2010; Dupas, 2011). We focus on one common characteristic of many health technologies: a relatively large up-front investment is required, while the benefits accrue over time. This is problematic for a number of interdependent reasons. First, households may find it difficult to borrow, especially for non-business purposes. Second, poor households

may have high discount rates or be close to subsistence levels of consumption and therefore be unwilling to sacrifice a large amount of current consumption. Third, households may exhibit time-inconsistency in the form of present bias or hyperbolic discounting (Ashraf et al., 2006). Fourth, households may be unwilling to sink a large sum into a new technology when they are unsure of its benefits. These barriers suggest a number of interventions to increase adoption and improve welfare. Consumers who face liquidity constraints or exhibit present bias may find it difficult to fund purchases even if they are willing to pay substantial amounts over time (Holla and Kremer, 2009). As a result, time payments, either micro-loans or layaways (dedicated savings), may increase adoption and improve welfare (Mahajan and Tarozzi, 2011; Dupas and Robinson, 2013; Tarozzi et al., 2014). When consumers have an uncertain valuation of a new product, a free trial or money-back guarantee can allow learning at low risk (Levine and Cotterman, 2012).

In this paper, we examine how time payment plans (either micro-loans or micro-savings) and interventions to decrease the risk incurred while learning (a free trial and a money-back guarantee) affect WTP and attempt to understand the mechanisms at work. Both of these are empirically challenging. First, individuals with greater access to finance may have a greater taste for health relative to consumption or more resources overall. Second, even if access to finance were randomly assigned, there are many variations possible and we would typically only observe one choice per individual, so it would require an enormous sample size to determine which policies are most attractive. Third, many of the underlying reasons for increased WTP (liquidity constraints, high discount rates, present bias / hyperbolic discounting, value of low-

risk learning) have similar empirical implications, making it difficult to identify the underlying mechanisms.

To address these questions, we measure WTP for a high-quality ceramic water filter in 400 households in slums of Dhaka, Bangladesh, where water quality is poor and the burden of water-borne disease high. We use a modified Becker-Degroot-Marschak (BDM) mechanism to elicit WTP under a variety of time payment plans, including a lump-sum paid immediately, micro-loans and dedicated micro-savings plans of varying duration. Crucially, we obtain valuations from each household across all payment plans, which (a) increases power and (b) helps us investigate the mechanisms behind differences in WTP across plans.

We find that the availability of time payments dramatically increases WTP. While the retail price is BDT 2100 (USD 28[1]), median WTP under a lump-sum, up-front payment is BDT 755 (USD 10.07), but increases to BDT 1260 (USD 16.80) with a simple 6-month loan and BDT 1530 (USD 20.40) for a 12-month loan. To separate time preference from liquidity constraints, we elicited WTP from subjects given layaway (dedicated micro-savings) plans with the same payment schedule as the loans. The intuition for this approach is that, while layaway plans should be less appealing than loans to all consumers, patient consumers who are liquidity constrained will find the layaway relatively more appealing than will impatient consumers. To our surprise, we found that for almost all households, WTP with a loan is virtually identical to WTP with a layaway plan with the same payment schedule. That is, households are willing to pay the exact same amount over 6 months to receive the

---

[1] The exchange rate during the study was roughly BDT 75 = USD 1.

filter in 6 months as they are to receive the filter today. In a standard model where all forms of consumption are discounted at the same rate, this implies that time preferences are unimportant and liquidity constraints alone explain the large increase in WTP from time payments. Alternatively, households could discount utility from future general consumption heavily, but do not discount the utility from owning the filter. To investigate the mechanisms at work, we estimate a simple structural model of liquidity constraints and time preference, and find strong evidence for the existence of credit constraints.

This paper proceeds as follows. In Section 3.2, we provide a brief literature review and conceptual framework. In Section 3.3, we describe the experimental design. In Section 3.4, we discuss the reduced-form evidence provided by our data. In Section 3.5, we propose and estimate a simple structural model of time preferences and credit constraints. Section 3.6 concludes.

3.2 Literature

Under-investment in welfare-enhancing or profitable technologies is thought to be a commonplace problem in developing countries. There are a variety of products, ranging from modern fertilizer to efficient cookstoves, that many poor people do not purchase, in spite of what would appear to be large benefits (Foster and Rosenzweig, 2010). While there are many potential explanations for this seeming underinvestment, in this section we focus on research related to time preference, liquidity constraints and consumers' lack of information on the effectiveness of the new product.

While the study of the relationship between liquidity constraints and consumption has a long history (Deaton, 1991), recent research in developing countries has focused on the impact of credit constraints on business investment and micro-enterprise. A large body of research shows that credit market imperfections are important impediments, (Banerjee and Duflo, 2005; de Mel et al., 2008; Banerjee and Duflo, 2014), but to date microcredit has not proved to be an effective solution (Banerjee et al., 2015b,a). A few recent studies have found that credit-based interventions have increased takeup of health investments (Devoto et al., 2012; Tarozzi et al., 2014; BenYishay et al., 2016)

Mahajan and Tarozzi (2011) (TM) and Dupas and Robinson (2013) (DR) both examine the relationship between non-standard time preferences and health investments, TM studying loans for bednet purchases in Orissa, India, and DR studying commitment savings for subject-chosen health products in Kenya. We highlight two differences between our study and these. First, we directly compare behavior under savings and borrowing. This is useful for policymakers as well as for understanding behavioral mechanisms. Second, we measure effects on WTP rather than share purchasing at a single price (TM) or total health investment or savings accumulated (DR), so our results are informative for pricing policy.

Both consumers and producers are likely to be uncertain about the returns to a new technology, and experimentation can be risky (Foster and Rosenzweig, 1995). Recent empirical research on the relationship between experimentation and adoption has been mixed. Dupas (2014) finds that short-run subsidies increase long-run adoption of insecticide-treated bednets in Kenya. Levine and Cotterman (2012) find that adding a free trial, time payments, and the right to return increased uptake of an efficient

charcoal stove from 5% to 45%. That study showed that either the free trial or time payments increased uptake by about half the total effect, but did not identify what barriers the sales offers overcame. However, experimentation can also lead to decreased adoption if consumers find the product inconvenient or unpleasant to use (Mobarak et al., 2012; Luoto et al., 2012).

There is substantial evidence that many people have present bias, meaning that their subjective discount rate for short-term decisions today is higher than their subjective discount rate for short-term decisions in the future. The most common formulation within economics is a model that assumes there is an exponential discount rate δ for most decisions, but an additional present bias discount rate $\beta < 1$ for all future periods (Laibson, 1997; O'Donoghue and Rabin, 1999). The potential role of present bias in underinvestment in health is discussed in Kessler and Zhang (2015) and, in a development context, in Dupas (2011).

3.3 Experimental Design and Data Collection

3.3.1 Context and Object of Sale

Our sample consists of approximately 400 poor households with young children in slums of Dhaka, Bangladesh. This population is of particular interest because of the low-quality piped water in these neighborhoods and high burden of water-borne disease, both generally and among young children.

The core intervention is the offer for sale of a long-lasting (18-24 months) ceramic water filter with a retail price of approximately BDT 2100. We are interested in the

demand for water filters because in previous research in this population found a strong distaste for chlorine-based treatment: WTP is low, and use is low even when provided free (Guiteras et al., 2016). The ceramic filter was popular in consumer testing in a similar population elsewhere in Dhaka, although few households purchased the filter at the break-even price (Luoto et al., 2011).

We begin with a simple household survey to collect basic data on demographics, socioeconomic status, risk preferences and recent episodes of water-borne disease. We then conduct a marketing meeting in which we explain the dangers of local water and promote the filter as a solution. The promotional message draws on our previous work in Dhaka with similar compounds, and combines both a positive health message as well as a message emphasizing disgust at ingesting fecal matter in unfiltered water. We inform the subject of the possible payment plans that might be offered in the sales visit and instruct her to think how much her household is willing to pay for the filter under each payment plan.[2] We also explain the modified BDM mechanism (Becker et al., 1964), described below, that we use to elicit WTP. To increase understanding, we conduct a real-money practice round of BDM for a token item (a packet of powdered dish detergent with a retail value of approximately BDT 10).

Two weeks later, we return for a sales visit, in which we use BDM to obtain the households' WTP under several different payment plans, listed in Table 3.1 and described at greater length below. Under the lump-sum plan, the filter would be delivered either the same day or, in a few cases when delivery was delayed, the next

---

[2]Our target respondent was the female head of household, since women typically have primary responsibility for water collection and providing water to children. We encouraged all household members to participate in the marketing meeting, discussion of WTP and financing, and sales process.

day, and payment in full required upon delivery. For the loan plans, the delivery of the filter was, again, that day or the next, with the first installment due on delivery. For the layaway plans, the timing of payments was the same as the corresponding loan plan, but the filter was only delivered at the time of the final payment. Payments were collected monthly and the collections officer recorded at each visit if the filter appeared to be in use.

Table 3.1: Offer Types

Offer type	Time of payment(s) (months)	Filter received (month)
Lump sum	0	0
3-month loan	0, 1, 2	0
3-month layaway	0, 1, 2	2
7-month loan	0, 1, 2, 3, 4, 5, 6	0
7-month layaway	0, 1, 2, 3, 4, 5, 6	6
12-month loan	0, 1, 2, …, 11	0
1-month delay	1	0
"75%, X, X"	0, 1, 2	2

Notes: In the "75%, X, X" offer, we fix the household's first payment at 75% of the maximum payment agreed to for a three-month loan, and the household then bids on the amount of the last 2 payments (X). The purpose is to provide variation between current and future payments to help identify present bias. The immediate (month=0) payment was due by the next day.

3.3.2 WTP Data and the BDM Mechanism

To obtain precise data on WTP, for each offer type, we conduct a series of BDM procedures, one for each offer type. In the standard implementation of BDM for a

---

[3]This need not be the case for subjects who are not expected-utility maximizers (Horowitz, 2006). In a study of WTP for water filters in rural Ghana, Berry et al. (2015) find a gap of about USD 1 between WTP as elicited by BDM and as revealed through a take-it-or-leave-it offer at a randomized price. While departures from expected-utility maximization could reduce our confidence in the levels of WTP we obtain from BDM, for our findings on differences between lump-sum and time payments, we need only for these departures not to interact with financing. Furthermore, even if these departures do interact with financing, if this interaction also appears in market behavior then BDM would still be informative about the effects of financing.

single good, the subject states her maximum WTP ("bid"). If there was only one offer type the bid is then compared against a random price ("offer"). If her bid is less than the offer price, she does not purchase the filter. If her bid is greater than or equal to the offer price, she purchases the filter at the offer price. For expected-utility maximizers, the subject's best strategy is to bid her maximum WTP truthfully.[3] To obtain a subject's WTP for a number of different offer types, we adapt BDM into a two-stage procedure. In the first stage, we obtain the subject's bid for each of the 8 offer types shown in Table 3.1, giving us a vector of WTP values for each household $\text{WTP}_h = (\text{WTP}_{h,1}, \ldots, \text{WTP}_{h,8})$. Then, in the second stage, we randomize one offer type for which the BDM draw is actually taken. That is, a random offer type $t \in \{1, \ldots, 8\}$ is chosen, we draw a random offer price $p_{h,t}$, and proceed as in a single-item BDM: if $p_{h,t} \leq \text{WTP}_{h,t}$, the household receives the filter and pays $p_{h,t}$; if $p_{h,t} > \text{WTP}_{h,t}$, the household cannot buy the filter. One disadvantage of our implementation was that, after extensive piloting, we found that it was necessary to provide participants with the minimum and maximum possible lottery prices, and to cap this range at the approximate break-even retail price of BDT 2100. This was necessary to improve participant understanding and to maintain a sense of fairness. However, it does mean that our WTP measure is censored, in that if a household has a very high WTP, we will observe only the top-coded value of BDT 2100. Because of this censoring, we will focus on quantile (median) estimates for demand data.

3.3.2.1 Offer Types

Table 3.1 lists the main offer types. The simplest offer is a lump sum paid on delivery which was scheduled with the family for that day or the next. We offer loans which begin immediately and involve 3, 7 and 12 monthly payments.[4] A parallel set of plans (3 and 7 payments) are for layaway, in which households make regular payments into a dedicated lockbox, according to the payment schedule, until they have accumulated the offer amount. These plans are soft commitments: even though the lockbox key is held by the organization implementing the survey,[5] the lockbox itself remains with the household and the savings will not be confiscated if the household "defaults" by not following through on its commitment. At the time the household is scheduled to make a deposit, field staff visit to confirm that the deposit has been made. Households also have the option to "deposit" their money with the field staff in exchange for a receipt.[6] To control for possible anchoring or ordering effects, the

---

[4]The prompts for BDM bids are framed in terms of the monthly payment rather than the total (e.g., "three monthly payments of BDT 400," rather than "BDT 1,200 over three months." However, we also provide subjects with the total amount implied by their monthly payments if they ask, as most pilot subjects did. The BDM draw, which determines the allocation and total price paid, is in terms of the total amount, which is then converted back into monthly payments for the relevant payment plan. We conduct the BDM draw in terms of the total amount for operational simplicity – otherwise, surveyors would have to carry separate sets of price envelopes for each offer.

[5]The International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDRB).

[6]As discussed in the Introduction, the intent of the layaway plans was to separate liquidity constraints from present bias. An alternative approach to identifying liquidity constraints would have been to give the subject the good in question and perform a reverse BDM in which the subject reveals the minimum amount she is willing to accept (WTA) in exchange for the good. The idea is to remove the liquidity constraint so that any variation in minimum WTA across payment plans could be attributed to time preference. This was not successful in piloting, for two main reasons. First, a large majority of pilot subjects stated that they would not accept any amount in exchange for the filter. We interpret this as some combination of a please-the-implementer effect and an endowment effect, with the former being more likely given that subjects even refused amounts higher than the going retail price. Second, reverse time payment plans were not perceived as credible by the subjects – many were skeptical that we would return multiple times over several months to give them money.

order in which plans were presented to each household was randomized to one of four different sequences: (a) loan plans then layaway plans, where plans in each section were ordered from shortest length to longest, (b) layaway plans then loan plans, all presented from shortest to longest, (c) loan plans then layaway plans, all presented from longest to shortest, (d) layaway plans then loan plans, all presented from longest to shortest.

3.3.2.2 Free Trial and Money-Back Guarantee

We randomized two treatments: a two-week free trial and a money-back guarantee. These were orthogonally randomized, so a quarter received both treatments, a quarter received neither, and a quarter each received only one of the treatments.

Randomized Treatment 1: Free trial. The first treatment is a two-week free trial, giving households an opportunity to learn to use the filter and to confirm whether ease of use, taste, and other characteristics are acceptable. These households received the filter at the marketing meeting and have if for use until the sales meeting. For risk-averse consumers one would expect the free trial to increase WTP (Levine and Cotterman, 2012), although there are counterexamples (Mobarak et al., 2012; Luoto et al., 2012).

Randomized Treatment 2: Money-back guarantee or rent-to-own. One potential barrier to adoption is that households may incur income, health or consumption shocks that ex-post mean that money spent on a filter would have been better spent on something else. To test whether this is an important determinant of WTP, we

randomize whether the loan offer gives the household the option to return the product for a partial or full refund up to a year from the sales date. With no refund, a time payment plan is similar to a "rent-to-own" scheme, in which the subject risks losing only accumulated payments, rather than the full lump sum. With a full or partial refund, the loan comes to resemble the layaway plan, but with the household receiving the flow of benefits from the product while payments are being made.

3.3.3   Data Collection and Summary Statistics

We conducted a baseline household survey at the time of the marketing meeting to collect basic data on demographics, socioeconomic status, risk preferences and recent episodes of water-borne disease. A final end line survey is conducted 6 months later. Main survey collections occurred from September of 2012 to June of 2013.

Table 3.2 shows the means of demographic characteristics in our sample. It also reports the difference across the free trial treatments (mean of those with free trial minus those without) and across the guarantee arms. There do not appear to be large differences across these treatments. Our measure of income is fairly noisy because of several outliers.

3.4   Reduced-Form Evidence

3.4.1   Effects on WTP

The most salient result from the study is that time payments dramatically increase WTP. Figure 3.1 compares the share of households willing to purchase the

Table 3.2: Randomization Check

	(1) Mean	(2) Free trial(1-0)	(3) Guarantee(1-0)
Female	0.877	0.0317	0.0129
Age	31.67	-0.222	1.266
HH Education (years)	5.527	0.713*	0.0632
Married	0.909	0.0137	0.0137
HH size	4.188	0.0309	0.152
Rent > USD 27/month	0.779	-0.0539	0.0274
Has gas-line	0.992	0.000475	0.0157**
Water in compound	0.990	-0.0113	0.0118
HH Income (USD, monthly)	368.8	61.06	10.63

Notes: 471 observations. Differences in means between those with and without the free trial (guarantee) are shown in column 2 (3).
* $p < .1$, ** $p < .05$, *** $p < .01$.

filter given a lump-sum offer with the share using the household's maximum bid across offers. All figures deflate cash flows by 13% per year, the approximate cost of funds for an MFI.[7] The increase in share of households willing to purchase via a time payments plan, relative to the share willing to purchase under a lump-sum offer, is statistically singificant at all prices above BDT 300, and the effect is 30 percentage points or more at all prices above BDT 700. Median WTP increases from BDT 755 to 1530 for a 12-month loan. Figure 3.2 examines differences in individual household WTP. Among households that are not censored (i.e., (i) do not have all bids at the top bid amount, and (ii) express some positive WTP for any offer), WTP increases for most households, with a median increase of BDT 600 (min. 0, IQR 200-1100, max. 1900).

---

[7]The commerical bank prime lending rate at the time of the study was 13% annually (CIA, 2012). Grameen Bank charges a nominal interest rate on loans of about 24% per year and pays approximately 8.5% on savings (Roodman, 2010).

Figure 3.1: Inverse Demand Curves: Time Payments vs. Lump Sum

(a) Levels
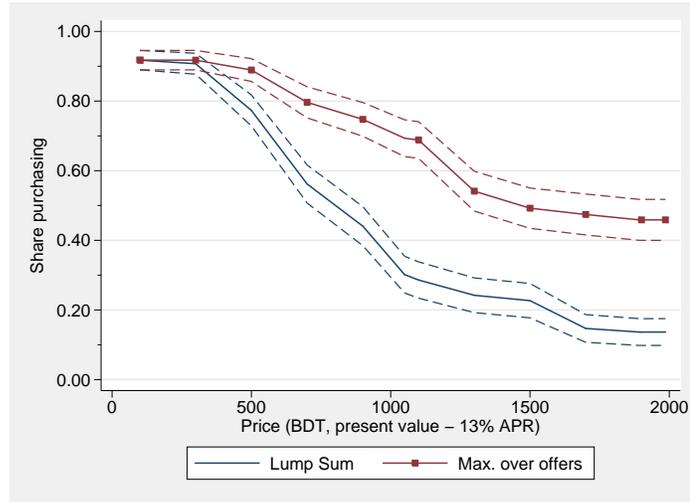
(b) Difference
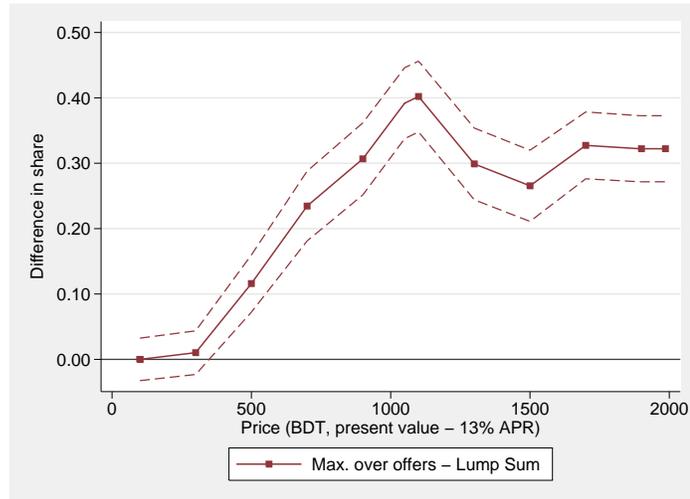
Notes: The top figure plots BDM demand curves, with 90% confidence bands, using households' maximum WTP across all offers (square markers) and households' maximum WTP for an immediate lump sum (no markers). The bottom figure plots the estimated differences (max. across all offers relative to lump sum). Pointwise inference from logit regressions (at prices BDT 100, 300, 500, ..., max). Standard errors clustered at the compound level. 388 observations. 75 BDT = 1 USD.

Figure 3.2: Distribution of Difference in WTP: Time Payments vs. Lump Sum
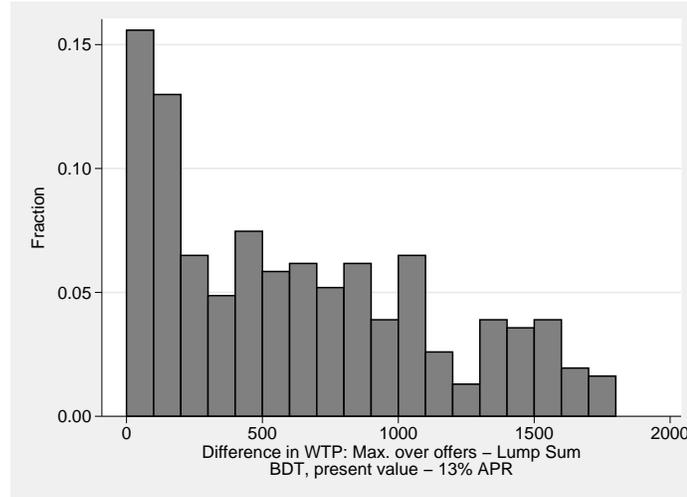
Notes: This figure plots the distribution of difference in household willingness to pay (WTP) under time payments (i.e., the maximum nominal amount across all loan and layaway offers) relative to an up-front lump-sum payment. We exclude 48 households that were top-coded (i.e., both their lump-sum and maximum time payment WTP were at the upper bound price) and the 32 households with zero WTP under all offers (including attriters and refusals), leaving 308 observations. 75 BDT = 1 USD.

Even a short-term (3-month) loan significantly increases demand, which continues as the term of the loan lengthens. This can be seen in Figure 3.3, which plots the share of subjects willing to purchase given each loan offer.

Surprisingly, WTP given time payment layaway plans are almost identical to loans. Figure 3.4 shows that the demand curves lie almost on top of each other, and Figure 3.5 shows that nearly all households have identical WTP for loans and layaway plans of the same duration. This suggests that households do not discount health benefits in the same way as utility from general consumption.[8] It is possible is that households anchor their valuation based on the offer they consider first, or are simply fatigued when considering later offers. However, this does not appear to be the case:

---

[8]It is possible that the layaway plan provides unique benefits, such as access to a lockbox to help save, that offset any utility loss to delaying consumption of the filter. However, it is unlikely that these effects would almost exactly cancel out for almost all households.

Figure 3.3: Demand Across Loan Offers

(a) Levels
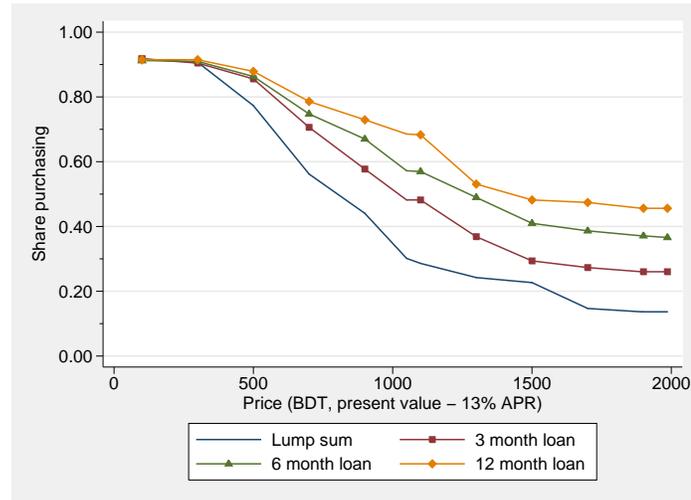
(b) Difference Relative to Lump-Sum
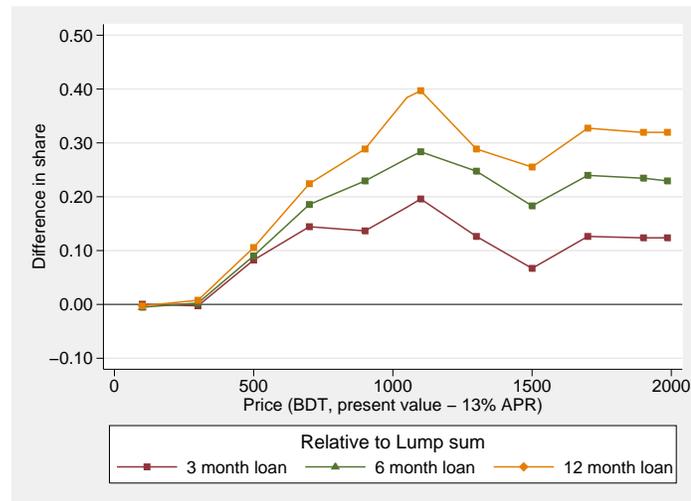
Notes: The top figure compares BDM demand curves across, with 90% confidence bands, loan offers: lump-sum (no markers), 3-month (square markers), 6-month (triangles) and 12-month (diamonds). The bottom figure plots the estimated differences for the three loan plans relative to lump-sum. Pointwise inference from logit regressions (at prices BDT 100, 300, 500, . . . , max). Standard errors clustered at the compound level. 388 observations. 75 BDT = 1 USD.

the randomized order of the offers did not significantly affect the WTP prices (Table 3.3).

Table 3.3: Offer Order Effect

	Loan7	Loan3
Loan before layaway offers	252	114.7
	(175.5)	(118.0)
Constant	735.0	1381.4*
	(1122.1)	(754.7)
Observations	352	352

Notes: Quantile (median) regression. Standard errors in parentheses. Includes controls for age, high rent, water access, gas access, and education. Dependent variable is total nominal amount in BDT (75 BDT = 1 USD). * $p < .1$, ** $p < .05$, *** $p < .01$.

The results from our randomized treatments are somewhat less striking. In neither case (free trial, Figure 3.6; guarantee, Figure 3.7) do we see strong evidence for an increase in demand.

3.4.2 Interpreting Reduced-Form Evidence

In this sub-section, we consider implications of the reduced-form results presented above, in particular for understanding savings and borrowing constraints in the population.

First, we can provide some evidence on whether households have difficulty saving for health investment (Dupas and Robinson, 2013). A simple measure of the ability of households to save is to compare WTP between the up-front lump-sum payment with the one-month delayed plan. If a household's WTP with a one-month delay is no greater than its WTP given a lump-sum today, this suggests that the household

Figure 3.4: Demand: Loans vs. Layaways
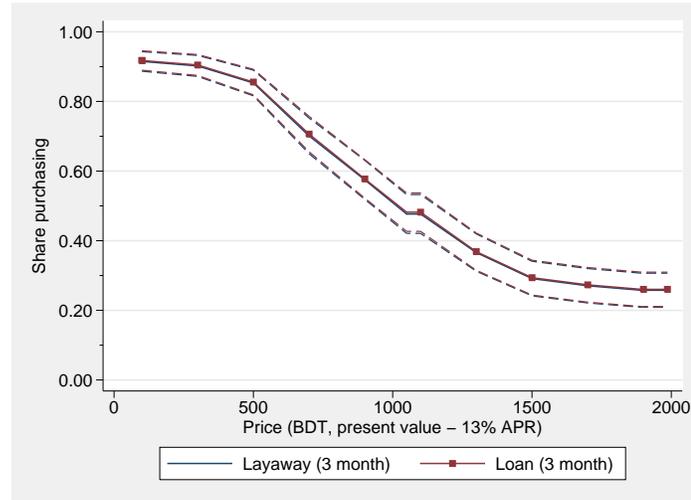
(a) 3 Months

Notes: The figure compares BDM demand curves, with 90% confidence bands, for 3-month loans (square markers) and 3-month layaway plans (no markers). Pointwise inference from logit regressions (at prices BDT 100, 300, 500, ..., max). Standard errors clustered at the compound level. 388 observations. 75 BDT = 1 USD.

(b) 7 Months
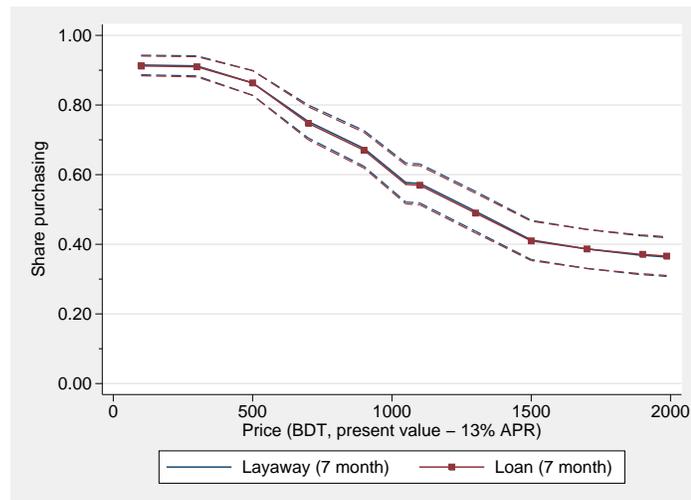
Notes: The figure compares BDM demand curves, with 90% confidence bands, for 7-month loans (square markers) and 7-month layaway plans (no markers). Pointwise inference from logit regressions (at prices BDT 100, 300, 500, ..., max). Standard errors clustered at the compound level. 388 observations. 75 BDT = 1 USD.

Figure 3.5: Difference in Household WTP: Loans vs. Layaways
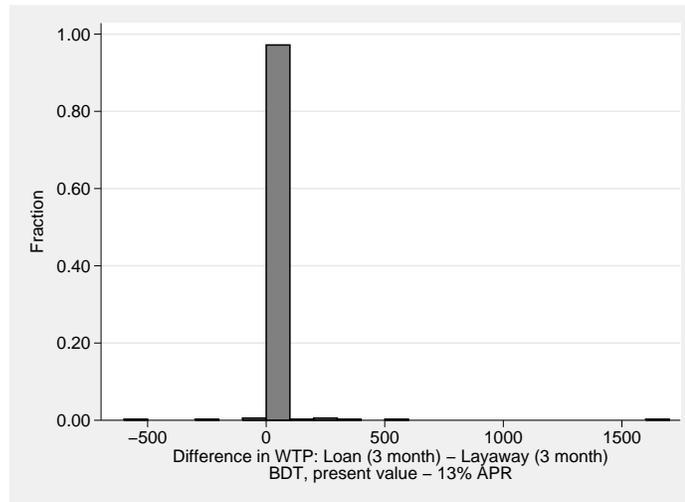
(a) 3 Months

Notes: This figure plots the distribution of difference in household willingness to pay (WTP) for 3-month loans relative to 3-month layaway plans. We exclude 0 households that were top-coded (i.e., both their lump-sum and maximum time payment WTP were at the upper bound price,) and 32 households with zero WTP for both offers (including attriters and refusals), leaving 356 observations. 75 BDT = 1 USD.

(b) 6 Months
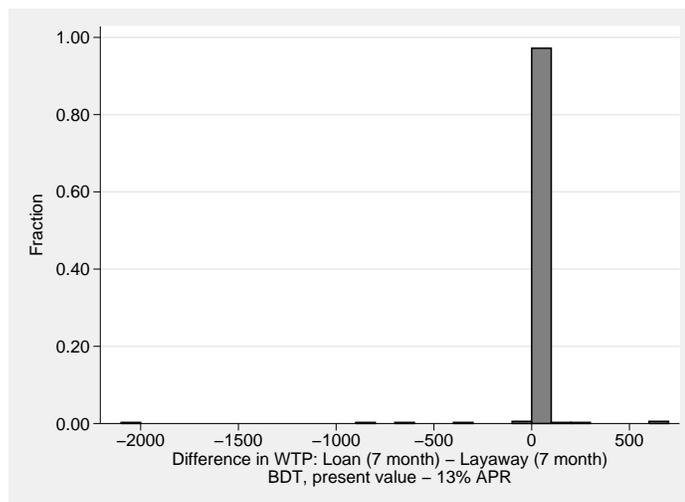
Notes: This figure plots the distribution of difference in household willingness to pay (WTP) for 7-month loans relative to 7-month layaway plans. We exclude 0 households that were top-coded (i.e., both their lump-sum and maximum time payment WTP were at the upper bound price) and 33 households with zero WTP for both offers (including attriters and refusals), leaving 355 observations. 75 BDT = 1 USD.

Figure 3.6: Effect of Free Trial Treatment on Demand

(a) Lump-Sum

Notes: The figure compares BDM demand curves, with 90% confidence bands, between free trial and no free trial households, given a offer. Standard errors clustered at the compound level. 189 observations. 75 BDT = 1 USD.

(b) 6-Month Loan
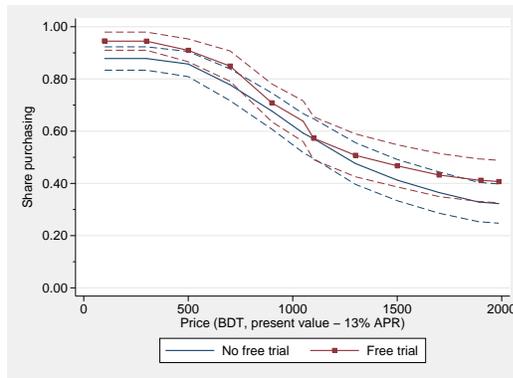
Notes: The figure compares BDM demand curves, with 90% confidence bands, between free trial and no free trial households, given a offer. Standard errors clustered at the compound level. 189 observations. 75 BDT = 1 USD.

(c) Max. WTP across All Offers

Notes: The figure compares BDM demand curves, with 90% confidence bands, between free trial and no free trial households, for the household's maximum (nominal) WTP across all offers. Standard errors clustered at the compound level. 189 observations. 75 BDT = 1 USD.

Figure 3.7: Effect of Money-Back Guarantee on Demand

(a) Lump-Sum

Notes: The figure compares BDM demand curves, with 90% confidence bands, between guarantee and no guarantee households, given a offer. Standard errors clustered at the compound level. 192 observations. 75 BDT = 1 USD.

(b) 6-Month Loan
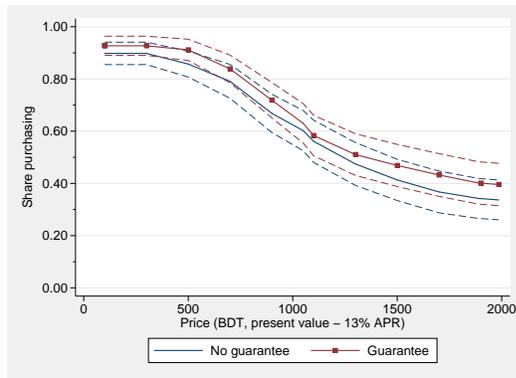
Notes: The figure compares BDM demand curves, with 90% confidence bands, between guarantee and no guarantee households, given a offer. Standard errors clustered at the compound level. 192 observations. 75 BDT = 1 USD.

(c) Max. WTP across All Offers

Notes: The figure compares BDM demand curves, with 90% confidence bands, between guarantee and no guarantee households, for the household's maximum (nominal) WTP across all offers. Standard errors clustered at the compound level. 192 observations. 75 BDT = 1 USD.

finds it difficult to save. In fact, 53.7% of the sample does not have a higher WTP for the delayed plan than lump sum, implying that the majority of households do have difficulty saving. On the other hand, among households that can save, this one-month delay is meaningful, since on average the delayed plan allows households to increase the WTP by 20.6%. This increase by allowing an extra month is roughly consistent in terms of magnitudes with the 45.5% from having two extra months in the 3-month plan.

Second, following Attanasio et al. (2008) and Karlan and Zinman (2008), we can test for the presence of credit constraints by measuring whether household WTP (in net present value terms) increases with the length of the loan. If a household does not face credit constraints in that it can borrow or save at a prevailing market interest rate without restriction, then the household's WTP would be equal in net present value across all payment plans, regardless of the household's time preferences. Therefore, if the household's WTP increases with the length of the loan maturity, this is evidence of credit constraints.[9] In Table 3.4, we show the results of regressions of deflated WTP on the length of the loan. In column 1, the dependent variable is WTP, deflated by a uniform interest rate of 26%, twice the average business business loan rate. We include household fixed effects and each household provides three observations, WTP from the 3-, 7- and 12-month loans. In column 2, we construct individual-specific interest rates by calculating the interest rate necessary for the household's WTP for the 3-month loan to be equal in real terms to the household's WTP under a lump sum. We then apply these individual-specific interest rates to the household's WTP for

---

[9]Friction costs associated with meetings would bias our results in the opposite direction.

the 7- and 12-month loans, and use these deflated WTPs as the dependent variables. Again, we include household fixed effects. In both specifications, the maturity length has a statistically and significant and economically meaningful positive effect on WTP, which suggests that households, on average, face important credit constraints.

Table 3.4: Maturity Effect

	(1) Loan PV (Const)	(2) Loan PV (Ind)
Loan length	22.12***	19.23***
	(1.544)	(1.950)
Constant	755.0***	784.3***
	(9.186)	(9.910)
N	1433	986
HH FEs	yes	yes

Notes: Standard errors in parentheses. The first model calculates the loan present value using a constant annual 26%. The second model calculates loan present values using individual rates derived from the 3-installment plan (compared to lump-sum). Units are BDT (75 BDT = 1 USD). Top-coded values are excluded. * $p < .1$, ** $p < .05$, *** $p < .01$.

Additional reduced-form evidence for credit constraints can be found by looking at the 1-month delay and 3-month plan. With perfect liquidity, we would observe that households' WTP are roughly equal in net present value terms:

$$\frac{p_{1dly}}{R} \approx p_{3m} + \frac{p_{3m}}{R} + \frac{p_{3m}}{R^2},$$

where R is the repayment rate (one plus the interest rate). Then

$$p_{1dly} = p_{3m} \left(R + 1 + R^{-1} \right).$$

With monthly interest rates small (i.e. $R \approx 1$), then the nominal totals of both (p_{1dly} and $3p_{3m}$) should be approximately the same. In reality, the average value of the 3-month loan value is 28.9% higher on average, which is economically significant.

Finally, the magnitude of this maturity effect suggests that some households have very high rates of subjective discounting. Even among households whose bids do not suggest they are savings-constrained, we observe a sharp increase in WTP as loan length increases from 3 months to 7 months and then to 12. It is difficult to attribute these increases to liquidity constraints alone. As a simple illustrative example, consider a household with a WTP of $20 for a 7-month loan. This is approximately equivalent to USD 0.095 per day over the course of the 7- month loan, which is slightly less than 1% of mean household income (approximately USD 10 per day). A 12-month loan reduces the cost to the household's daily consumption to approximately USD 0.055 per day over the course of the 12-month loan. Even extremely risk-averse households are unlikely to find a difference of USD 0.04 per day so substantial that they would strongly prefer the 12-month loan to the 7-month loan, so it seems unlikely that liquidity constraints alone can explain the observed increase in WTP from 7- to 12-month loans.

3.5   Estimating Preferences and Constraints

Our reduced-form empirical analysis provides strong evidence that micro-loans and micro-savings significantly increase WTP. To assess the relative importance of credit constraints and time preferences in explaining this fact, we turn to a simple

structural model. We designed our experiment to provide clean identification of preferences versus constraints by using within-household differences in WTP between micro-loans and micro-savings. However, the surprising fact that subjects' WTP for micro-loans and micro-savings plans were almost identical means that this strategy is no longer viable. Instead, to estimate the structural model, we exploit differences in households' WTP across loan offers with different loan duration and different timing of payments.

3.5.1 Utility

The household receives utility from consumption and from owning the filter. The lifetime benefit to the household of owning the filter is B. Given that households appear indifferent to when they receive the filter, there is no discounting of B. The household's per-period (monthly) utility from general consumption is

$$u\left(c_t\right) = u\left(y - \bar{p}_t\right),$$

where y is monthly income, assumed for simplicity to be constant, and $\bar{p}_t$ is the amount the household pays for the filter in period t.

In the model, we limit the household's planning horizon to 12 months. This is a largely innocuous assumption, because all financial transactions with regards to the filter will be complete within 12 months. We assume the household discounts its utility from general consumption exponentially with discount rate δ. If the household

does not purchase the filter, its total utility is

$$U_0\left(\text{buy}=0\right) = \sum_{t=0}^{11}\left\{\left(\frac{1}{1+\delta}\right)^t u\left(y\right)\right\}.$$

If the household does purchase the filter and makes a sequence of payments $\{\bar{p}_t\}$, its total utility is

$$U_0\left(\text{buy}=1; \{\bar{p}_t\}\right) = \sum_{t=0}^{11}\left\{\left(\frac{1}{1+\delta}\right)^t u\left(y-\bar{p}_t\right)\right\} + B.$$

A household is indifferent between not purchasing and purchasing with a sequence of payments $\{\bar{p}_t\}_{t=0}^{11}$ if $U_0\left(\text{buy}=0\right) = U_0\left(\text{buy}=1; \{\bar{p}_t\}\right)$, i.e. if

$$\sum_{t=0}^{11}\left(\frac{1}{1+\delta}\right)^t \left[u\left(y\right) - u\left(y-\bar{p}_t\right)\right] = B. \tag{3.1}$$

Taking a second-order approximation of the difference (3.1) yields

$$\sum_{t=0}^{11}\frac{1}{(1+\delta)^t}\left[\bar{p}_t + \frac{1}{2}\eta\bar{p}_t^2\right] = w, \tag{3.2}$$

where $\eta = -u''(y)/u'(y)$ measures utility curvature (the coefficient of absolute risk aversion) and $w = B/u'(y)$ is the value of the filter normalized by the marginal utility of income.

3.5.2 Credit Environment

We denote the payment plan (i.e. the payments in our micro-loan) as a sequence $\{p_t\}$, which we distinguish from the household's net payment sequence $\{\bar{p}_t\}$. These two can differ because the household may borrow from other sources or, having previously borrowed, may need to repay these outside loans. We denote these "outside" payments as f_t, so in any period t,

$$p_t = \bar{p}_t + f_t. \tag{3.3}$$

That is, if in a period t the household owes an installment p_t on its micro-loan, it can pay this installment $\bar{p}_t$ by forgoing consumption in period t, by borrowing f_t from an outside source, or some combination of the two. Equivalently, the household's net payment for the filter in a period t is

$$\bar{p}_t = p_t - f_t. \tag{3.4}$$

If a household borrows in a period t, $f_t > 0$, while if a household is repaying outside loans, then $f_t < 0$.

Rather than build credit constraints from microfoundations, we take a reduced-form approach and model credit constraints as a nonlinear cost-of-borrowing function, which for simplicity we approximate as a quadratic. To develop the notation, first suppose that the household does not face any credit market imperfections, and can borrow freely at a gross monthly interest rate R_1. Then, if the household borrows an

amount b, it must make payments $\{p_t\}$ with net present value equal to b:

$$\sum_t \frac{p_t}{R_1^t} = b. \tag{3.5}$$

Now we introduce our reduced-form model of credit constraints: rather than repaying an amount with net present value equal to b, the household must make payments $\{p_t\}$ with net present value

$$\sum_t \frac{p_t}{R_1^t} = \tilde{R}_0 + b + \tilde{R}_2 b^2 = \tilde{q}(b), \tag{3.6}$$

where $\tilde{R}_0 + \tilde{R}_2 b^2$ is the penalty for being constrained. So, for example, if a household borrows an amount b for one period at time τ, then at time $\tau + 1$ it owes an amount $q(b)$ defined by the following quadratic function:

$$q(b) = R_0 + R_1 b + R_2 b^2, \tag{3.7}$$

where $R_0 = \tilde{R}_0 \cdot R_1$, and $R_2 = \tilde{R}_2 \cdot R_1$.

This extends the standard transaction-cost model of loans (e.g., Helms and Reille, 2004) by adding the quadratic term R_2. Adding this term is attractive for several reasons: (a) the observed repayment rate $q(b)/b$ is not necessarily declining in b; (b) it is better able to approximate situations where there are fixed limits on the amounts a household can borrow, i.e. where the borrowing costs function would be vertical. Possible micro-foundations for a quadratic-shaped borrowing cost function include (a) it incorporates the idea that with multiple sources of limited funds a household will

choose the cheaper options first, and (b) if lenders expect that larger loans are less likely to be paid back they will charge higher effective interest rates for larger loans.

We restrict outside borrowing as follows: the household may borrow the same amount b for each month of the micro-loan, and then it must repay its outside borrowing within three months. That is, during the term of the micro-loan, it pays p_t each month, with $\bar{p}_t$ of this amount coming from forgone consumption and b of this amount from outside borrowing, i.e. $f_t = b$ in equations 3.3 and 3.4. Then, in the three months after the formal micro-loan is complete, the household makes payments r with net present value $\tilde{q}(b)$, i.e. $f_t = -r$ in equation 3.4. We make this restriction for simplicity, but it is a reasonable assumption in an environment where most moneylending is fairly short-term. For a plan of N payments, the above leads to the following repayment equation relating the net-present value of the borrowing costs with repayments

$$\sum_{t=0}^{N-1} \frac{1}{R_1^t} \tilde{q}(b) = \sum_{t=N}^{N+2} \frac{1}{R_1^t} r. \tag{3.8}$$

As an example, imagine a household is deciding the maximum monthly payment p they are willing to make in a three-month loan. Rather than forgo the full amount p during the three months of repayment, the household can borrow monthly amounts b from an outside source, so that its consumption only drops by $\bar{p}_{t<3} = p - b$ over the three months of the formal loan. Then, over the three months after the formal loan, the household must repay its source of outside borrowing a monthly amount

132

$\bar{p}_{t \geq 3} = r$ for three months, where r is determined by $q(b)$. Outside credit would be most attractive for shorter plans as there is more opportunity for smoothing.

In specifying the credit environment in this way, we are not claiming that households literally managed their finances as we have described. Rather, our goal is to model the essential features of an environment like that described in Collins et al. (2009), where households have multiple possible sources of funds, whether short-term informal borrowing or deferring other obligations, and that the household incurs a fixed cost for drawing on these funds (thus the transactions cost term R_0) and finds it increasingly difficult or costly to obtain larger amounts (thus the quadratic term R_2).

3.5.3 Household's Optimization Problem

We now consider the household's maximum WTP for the filter with a given micro-loan plan. The micro-loan plan is a sequence of payments $\{p_t\}$, and the household must decide on $\{p_t^*\}$, the most it is willing to pay given its preferences and the constraints that it faces. Recall from Equation 3.2 that a household with preferences $\{B, \delta, \eta\}$ (valuation of the filter, discounting and utility curvature, respectively) is indifferent between not purchasing the filter and purchasing the filter for a sequence of total payments $\{\bar{p}_t\}$ such that

$$\sum_{t=0}^{11} \frac{1}{(1+\delta)^t} \left[\bar{p}_t + \frac{1}{2} \eta \bar{p}_t^2 \right] = w. \tag{3.2}$$

Given the credit environment the $\{R_0, R_1, R_2\}$, the household's best strategy is to make the highest possible bid $\{p_t^*\}$ such that the indifference condition 3.2 holds, while

133

selecting the optimal amount of outside borrowing $\{b_t^*\}$ and implied repayments $\{f_t^*\}$ to maximize its utility. If borrowing costs are sufficiently high, optimal borrowing may be zero.

The solution for the optimal $\{p_t^*\}$ is derived in Appendix 3.7. Intuitively, we can think of the household as optimizing in the following way. Suppose that the household is considering making a bid $\{p_t\}$ for the filter. Given this bid $\{p_t\}$, the household will choose the optimal amount of borrowing to maximize its utility over the 12-month horizon, which will result in net payments $\{\bar{p}_t\}$. If these net payments $\{\bar{p}_t\}$ are such that the left-hand side of 3.2 is less than the right-hand side, then the household should increase its bid $\{p_t\}$. Similarly, if these net payments $\{\bar{p}_t\}$ are such that the left-hand side of 3.2 is greater than the right-hand side, then the household should decrease $\{p_t\}$.

3.5.4 Identification

To build intuition for how the parameters are identified, we show in Figure 3.8 the effects of varying each parameter individually on the profile of BDM bids across payment plans. In Figure 3.8a, we examine the effect of R_0, the fixed cost of borrowing. Initially, as R_0 increases from 0 to 20, the largest effect comes in the household's WTP for the 7-month loan, since the household no longer finds it worthwhile to borrow from outside sources to smooth consumption. WTP for the 12-month loan is unaffected, since the household does not need to resort to outside borrowing for the 12-month loan. WTP for shorter durations are less affected because, in spite of the fixed cost, the

household still finds it worthwhile to borrow from outside sources. As R_0 continues to increase, WTP is depressed for shorter-term plans, since the increased fixed cost makes it no longer worthwhile to borrow from outside sources to supplement the "inside finance" provided by these plans. Figure 3.8b shows that increasing the cost of capital, R_1, pushes down the WTP profile for all durations except the 12-month loan, in which the household is not borrowing. Figure 3.8c shows that increasing the quadratic term the cost of capital, R_2, has the largest effect on short-term plans: the household borrows more from outside sources, so the quadratic term has a bigger impact on borrowing costs for these short-term plans. Figure 3.8d shows that increasing the household's discount rate δ increases the household's WTP in all plans. This is because a household that is more impatient values future consumption less, and so is more willing to sacrifice future consumption to obtain the filter. Increasing δ has a larger effect on longer-duration loans (7-month) since because δ compounds: the household downweights the loss of consumption in period t by $1/\left(1+\delta\right)^t$, and so sacrificing consumption in months $t = 6, 7, 8, \ldots$ carries only a small utility cost for households with high δ. Figures 3.8e and 3.8f show that utility curvature (η) and the monetized filter value (w) are not well identified separately in the current model; they both shift vertically the WTP profile quite evenly. Only one of these is estimated at the individual level.

Figure 3.8: Identifying Structural Parameters from WTP Data

(a) R_0 – Fixed Cost of Borrowing

(b) R_1 – Linear Term in Borrowing Cost

(c) R_2 – Quadratic Term in Borrowing Cost

(d) δ – Discount Rate

(e) η – Curvature of Utility Function

(f) w – Monetized Value of Filter
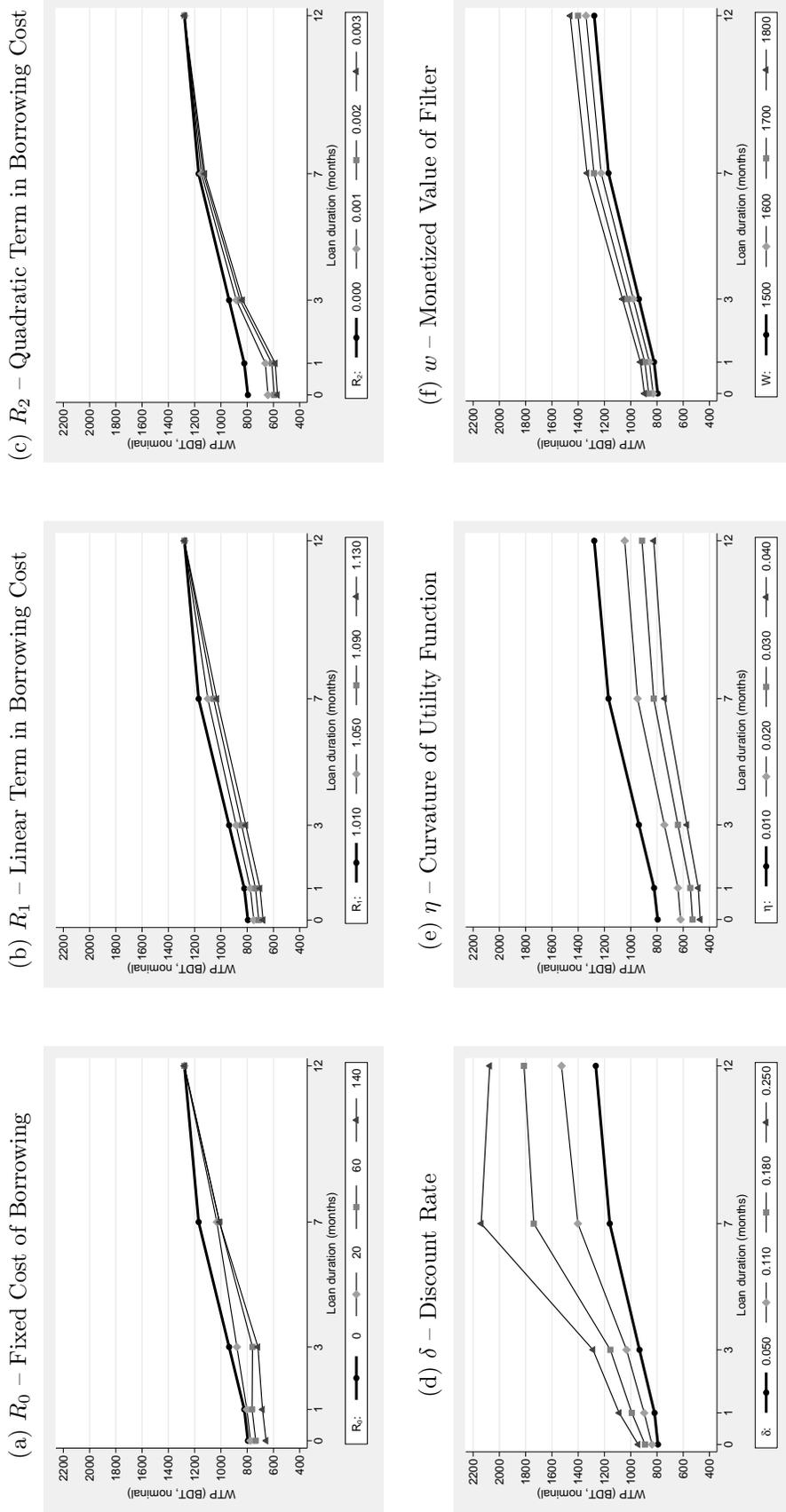
Notes: These figures show how the household's set of bids change as we vary a single parameter, holding the other parameter values constant. Baseline values of parameters: $R_0 = 0$; $R_1 = 1.010$; $R_2 = 0$; $\delta = 0.050$; $\eta = 0.010$; $w = 1500$. This baseline case is plotted with the bold line in all figures. The plan with a single payment delayed one month is plotted at 1 month. BDT 75 = USD 1.

3.5.5 Estimation

For each household i, our data consist of the household's bids on each of the M offers, $p_i = (p_{i,1}, \ldots, p_{i,M})$, i.e., lump-sum, one-month delay, 3-, 7-, and 12-month loan. The parameters of our model are $\{B, \delta, \eta, R_0, R_1, R_2\}$, where $\{B, \delta, \eta\}$ describe a household's preferences (valuation of the filter, discounting and utility curvature, respectively) and $\{R_0, R_1, R_2\}$ describe the household's credit environment. In principle, all of these parameters vary by household. However, since we have only 5 data points per household, we cannot estimate all of the parameters at the household level. We divide the parameters into population-level and individual-level groups. For individual parameters, we choose $\omega_i = \{B_i, R_{2i}, \delta_i\}$, as we believe these are likely to vary the most in the population and allow us to be able to distinguish basic preferences from credit constraints. Population parameters are then $\alpha = \{R_0, R_1, \eta\}$.

We estimate the model in an iterative two-step process. To build intuition, we first describe this process as if p_i were not censored (recall that bids were top-coded at the approximate break-even price of BDT 2100), and then describe our modification to account for censoring.

First, note that the household's optimization provides a mapping from parameters (ω_i, α) to predicted WTP $\hat{p}_i = (\hat{p}_{i,1}, \ldots, \hat{p}_{i,M})$. For each plan m, $\hat{p}_{im}$ is the maximum of the predicted WTP assuming no other borrowing and of the WTP assuming outside borrowing. For the no borrowing condition we solve Equation 3.2 for when monthly forgone consumption is the monthly filter price and there are no effects after the plan

is finished. For the borrowing condition, we use the aggregate repayment equation

plus the first order condition from Equation 3.2 to determine the maximum WTP.

The intuition for our estimation strategy is to choose the value of the parameters

that minimizes the difference between actual bids p_i and predicted bids $\hat{p}_i$. We start

with initial guesses for the population-level parameters $\alpha^{(0)}$. We then iterate the

following procedure:

1. Given current population level estimates $\alpha^{(j)}$, we choose household-specific

    parameters $\omega_i^{(j+1)}$ to maximize:

$$\Gamma_i\left(\omega_i|\alpha^{(j)}\right) = \sum_m \Psi_{im}(\omega_i, \alpha^{(j)}, p_{im}) \; \forall i \qquad (3.9)$$

$$\Psi_{im}(\omega_i, \alpha^{(j)}, p_{im}) = \ln \frac{1}{\sigma} \phi \left(\frac{p_{im} - \hat{p}_m(\omega_i, \alpha)}{\sigma} \right) \qquad (3.10)$$

    where Ψ_{im} is the log-likelihood of the parameters yielding the stated WTP for

    person i and plan m, the summation is over plans m, p_{im} is the observed WTP

    for individual i for plan m, and $\hat{p}_m(\omega_i, \alpha)$ is the predicted WTP for plan m given

    the parameters (i.e. the maximum WTP given parameters $\{\omega_i, \alpha^{(j)}\}$).

2. Given current individual-specific parameters for the sample $\omega^{(j+1)} = \left\{\omega_i^{(j+1)}\right\}_{i=1}^{N}$,

    we choose population level parameters $\alpha^{(j+1)}$ to maximize the sample log-

likelihood:

$$\Gamma\left(\alpha^{(j+1)}|\omega^{(j+1)}\right) = \sum_i \Gamma_i\left(\omega_i|\alpha^{(j)}\right)$$

$$= \sum_i \sum_m \Psi_{im}\left(\omega^{(j+1)}, \alpha^{(j+1)}, p_{im}\right) \qquad (3.11)$$

where the outer summation is over subjects i.

3. We repeat steps 1-2 until convergence.

In each step we estimate the parameters of interest via maximum likelihood. With current candidate parameters and the parameters taken as given in each round, we predict the WTP for each individual. The WTP is the highest price that allows the family through some amount of borrowing to be indifferent between purchasing the filter at the price and having no filter. We solve then for the amount of borrowing from outside sources that maximizes the WTP while keeping the family indifferent. We can then determine the error between predicted and observed WTPs which we assume is normally distributed and independent across individuals and plans. We weight deviations between predicted and observed bids equally for each offer.

Since observed WTPs are censored from above, we adjust the likelihood function in a Tobit-style fashion, replacing $\Psi_{im}(\omega_i, \alpha^{(j)}, p_{im})$ from Equation 3.10 with

$$\Psi_{im}(\omega_i, \alpha^{(j)}, p_{im}) = 1\{p_{im} < p_{top}\} \cdot \ln\left[\frac{1}{\sigma}\phi\left(\frac{p_{im} - \hat{p}_m(\omega_i, \alpha)}{\sigma}\right)\right] \qquad (3.12)$$
$$+ 1\{p_{im} = p_{top}\} \cdot \ln\left[1 - \Phi\left(\frac{p_{im} - \hat{p}_m(\omega_i, \alpha)}{\sigma}\right)\right],$$

where p_{top} is the top-coded amount and $1\{\cdot\}$ is the indicator function.

During estimation, parameters are constrained to be non-negative. For all parameters except R_{2i}, this restriction is natural. For R_{2i}, in principle borrowing costs could be concave $(R_{2i} < 0)$, but we believe this restriction is reasonable given the borrowing environment.

3.5.6 Structural Results

The values of the structural parameters are reported in Table 3.5, and distributions for individual-level parameters are shown in Figure 3.9a. The population-level (i.e., not household-varying) parameters R_0, R_1 and η are significantly different than zero ($p < 0.001$ in all cases). The point estimates for R_0 (BDT 117) and R_1 (a gross monthly interest rate of 1.012, approximately equal to a net annual interest rate of a 2.4%) are reasonable.

Table 3.5: Estimated Structural Parameters

	Estimates
R_0 (borrowing fixed cost)	116.9***
	(11.50)
R_1 (borrowing repayment rate)	1.012***
	(0.0117)
R_{2i} (borrowing quadratic cost): median	0.0000597
	(0.000139)
R_{2i} (borrowing quadratic cost): share positive	0.509***
	(0.0288)
δ_i (monthly discount rate): median	0.180***
	(0.0301)
w_i (monetized filter value, BDT): median	1768.7***
	(174.7)
η (utility curvature)	0.0102***
	(0.000194)
Observations	291

Notes: Estimate of structural parameters from WTP data.
75 BDT = 1 USD. Bootstrap p-values (399 reps).
* $p < .1$, ** $p < .05$, *** $p < .01$.

While the median estimated R_{2i} is not significantly different from zero, over half of the point estimates are positive, as shown in Figure 3.9a. For a large fraction of the population, the effective cost of borrrowing large sums is extremely high. 3.10 plots the amount a household must repay for a one period loan (the $q(b)$ function from Equation 3.7) across the distribution of R_{2i}. At the 75th percentile of R_{2i}, a loan of BDT 250 would require a repayment of nearly BDT 1,000, and for the top quintile of R_{2i}, borrowing just BDT 100 is prohibitively costly.

We estimate quite high discount rates for non-health utility, as shown in Figure 3.9b: the median estimated discount rate δ_i is 18.9% per month, and the 75th percentile is approximately 50% per month. This is extraordinarily high – a 50%

Figure 3.9: CDFs of Estimated Household-Varying Structural Parameters

(a) Borrowing Cost Quadratic Term R_{2i}
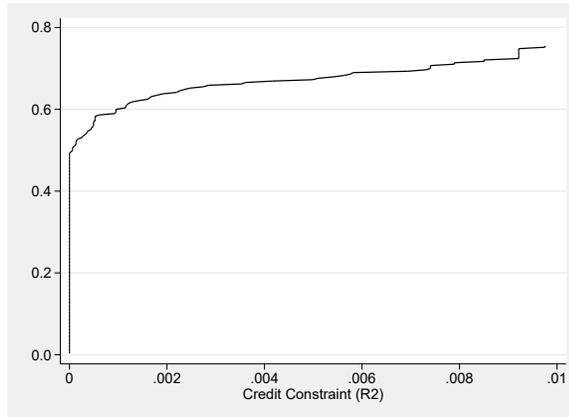
Notes: Values above the 75 percentile omitted from graph

(b) Monthly Discount Rate δ_i
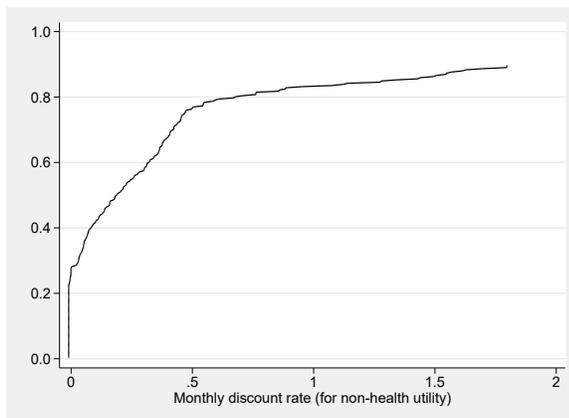
Notes: Values 10 times the median are omitted from graph (10.31% of observations)

(c) Filter Value w_i
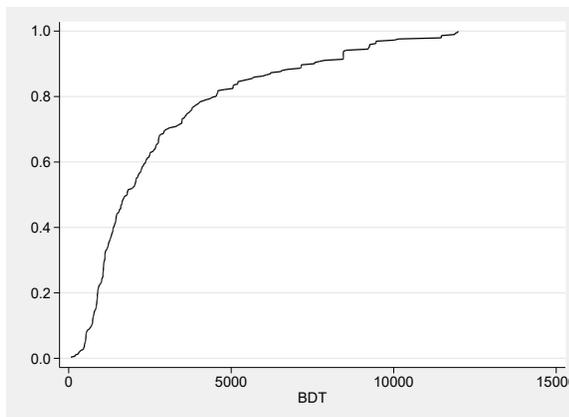
Notes: Distribution of estimated per-month health utilty divided by Marginal Utility of money. 75 BDT = 1 USD.

142

Figure 3.10: Cost of Funds - Variation by Quadratic Cost (R2)
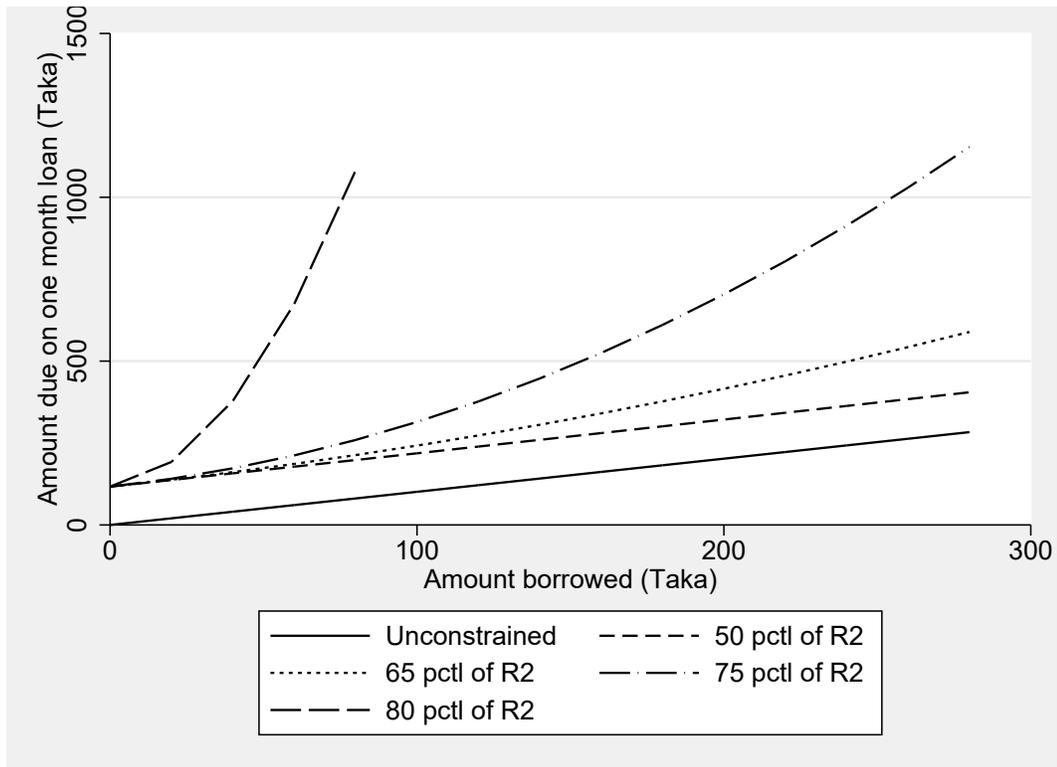
Notes: This figure plots the amount due on a one month loan for different values of the borrowing cost parameters R_0 and R_{2i}. R_1 is held fixed at the estimated value of 1.012. The solid line ("Unconstrained") represents an unconstrained borrower with $R_0 = 0$ and $R_{2i} = 0$. The remaining lines represent borrowing costs for individuals at the indicated quantiles of the estimated distribution of R_{2i}. In each of these cases, the borrower must pay the estimated fixed cost $R_0 = 116.9$.

monthly discount rate implies that USD 100 in one year is worth about USD 0.75 today. While we believe that impatience with respect to general consumption is an important feature of these households' economic lives, we suspect that features not captured by the model – in particular, difficulty saving – are contributing to this result. In our model, utility is discounted exponentially, but generalizing to hyperbolic discounting would be unlikely to explain this particular puzzle, since the estimated discount rates are driven primarily by differences in valuation between the 7-month and 12-month loans, and the difference in immediate payment due between these two plans is small. Finally, Figure 3.9c shows that our estimates of w_i, the monetized utility value of the filter, are plausible, with a median of BDT 1770 (USD 23.5).

Using these estimated parameters, we can conduct the counterfactual experiment of eliminating credit constraints by setting $R_0 = R_2 = 0$. The median WTP for a 3-month loan plan increases by BDT 347, which is equivalent to an increase in the value of the filter by 53%.

3.6   Conclusion

In this paper, we show that micro-loans dramatically increase willingness to pay (WTP) for water filters among poor households in slums of Dhaka, Bangladesh. Surprisingly, micro-savings plans have an equal effect, indicating that households are willing to save to receive a filter in the future. Micro-savings may be a useful option for an NGO that is concerned about its ability to enforce the terms of a micro-loan. Of course, micro-savings requires that the household trust the institution holding

144

its savings, and our context may be somewhat unusual given how highly regarded ICDDRB is in the community. Still, even in settings where it may be more difficult to establish trust, this striking result suggests that more enterprises should experiment with layaway plans and variations, such as dedicated savings accounts held by a trusted third party. Developing technologies to facilitate micro-payments, such as mobile money, may make it feasible to offer micro-payment schemes without incurring large collection costs.

In examining the WTP data from our experiment to understand the mechanisms behind the large increase in WTP, we find evidence of important financial frictions. Many households appear to find it difficult to save, and many households face significant credit constraints. Using the estimates from our structural model, we find a counterfactual setting that removing credit constraints is equivalent to increasing the household's value of the filter by 53%. Furthermore, a large fraction of households appear to discount general consumption heavily, suggesting that plans similar to Save More Tomorrow may be useful for social enterprises marketing health goods (Thaler and Benartzi, 2004). Finally, one important limitation of our model is that it does not capture short-term (within-month) savings constraints, and these may be driving some of the extremely high discount rates we estimate. The existence of such short-term credit constraints points to the value of more research on the value of flexible micro-payments that can be adapted to a household's individual circumstances, such as daily fluctuations in income (Collins et al., 2009).

3.7 Appendix A: Household's Optimal Bid

In this Appendix, we derive the household's optimal BDM bid $\{p_t^*\}$ given its preferences and the credit environment it faces. If the household borrows b during the plan and repays $\hat{p}$ afterward, their repayments paying off their borrowing (updating 3.8)

$$\sum_{t=\gamma}^{N-1+\gamma} \frac{1}{R_1^t} \tilde{q}(b) = b_{total} = \sum_{t=N+\gamma}^{N+\gamma+2} \frac{1}{R_1^t} \hat{p} \tag{3.13}$$

$$(R_0 + R_1 b + R_2 b^2) \left[R_1^2 \frac{(R_1^N - 1)}{(R_1^3 - 1)} \right] = \hat{p}$$

$$(R_0 + R_1 b + R_2 b^2) \tilde{R} = \hat{p} \tag{3.14}$$

and the WTP identifies their indifference point (updating equation 3.1)

$$\left[(p - b) + \frac{1}{2} \eta (p - b)^2 \right] \sum_{t=\gamma}^{N-1+\gamma} \delta^t + \left(\hat{p} + \frac{1}{2} \eta \hat{p}^2 \right) \sum_{t=N+\gamma}^{N+\gamma+2} \delta^t = w \tag{3.15}$$

$$\left[(p - b) + \frac{1}{2} \eta (p - b)^2 \right] d_1 + \left(\hat{p} + \frac{1}{2} \eta \hat{p}^2 \right) d_2 = w$$

$$\frac{1}{2} \eta p^2 + (1 - \eta b)p + \left[\frac{1}{2} \eta b^2 - b + \left(\hat{p} + \frac{1}{2} \eta \hat{p}^2 \right) \frac{d_2}{d_1} - \frac{w}{d_1} \right] = 0 \tag{3.16}$$

This is a quadratic function of p.

$$p(b) = -\lambda_0 + b + \sqrt{\lambda_2 - \lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right)}$$

$$\lambda_0 = \frac{1}{\eta}$$

$$\lambda_1 = 2\frac{1}{\eta}\frac{d_2}{d_1}$$

$$\lambda_2 = \frac{1}{\eta^2} + 2\frac{1}{\eta}\frac{w}{d_1}$$

$$d_1 = \delta^\gamma \frac{1 - \delta^N}{1 - \delta}$$

$$\frac{d_2}{d_1} = \frac{\delta^N(1 - \delta^3)}{(1 - \delta^N)}$$

We then need to pick b to maximize $p(b)$. We use $p'(b) = 0$ to find optimal $\hat{b}$.

$$1 + \frac{1}{2}\left(\lambda_2 - \lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right) \right)^{-\frac{1}{2}} \left[-\lambda_1(1 + \eta\hat{p})\frac{\partial\hat{p}}{\partial b} \right] = 0$$

$$\lambda_1(1 + \eta\hat{p})\frac{\partial\hat{p}}{\partial b} = 2\left(\lambda_2 - \lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right) \right)^{\frac{1}{2}}$$

$$\lambda_1^2(1 + \eta\hat{p})^2((R_1 + 2R_2b)\tilde{R})^2 = 4\left(\lambda_2 - \lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right) \right)$$

$$\lambda_1^2(1 + 2\eta\hat{p} + \eta^2\hat{p}^2)\tilde{R}^2(R_1^2 + 4R_1R_2b + 4R_2^2b^2) = 4\left(\lambda_2 - \lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right) \right)$$

we could substitute in for $\hat{p}$ and solve for b, but we instead substitute out b and solve for $\hat{p}$ using equation .

$$\lambda_1^2(1 + 2\eta\hat{p} + \eta^2\hat{p}^2)\tilde{R}^2 \left(R_1^2 + 4R_2 \left(\frac{\hat{p}}{\tilde{R}} - R_0 \right) \right) - 4\left(\lambda_2 - \lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right) \right) = 0$$

$$\lambda_1^2\tilde{R}(1 + 2\eta\hat{p} + \eta^2\hat{p}^2) \left(\hat{R} + 4R_2\hat{p} \right) - 4\lambda_2 + 4\lambda_1 \left(\hat{p} + \frac{1}{2}\eta\hat{p}^2 \right) = 0$$

where $\hat{R} = \tilde{R}(R_1^2 - 4R_2R_0)$.

$$\lambda_1^2 \tilde{R}\hat{R} + \lambda_1^2 \tilde{R}4R_2\hat{p} - 4\lambda_2 +$$

$$\left(\hat{p} + \frac{1}{2}\eta\hat{p}^2\right)2\lambda_1\left(2 + \eta\lambda_1\tilde{R}\hat{R} + \eta\lambda_1\tilde{R}4R_2\hat{p}\right) = 0$$

$$\left[\lambda_1^2\tilde{R}\hat{R} - 4\lambda_2\right] + \hat{p}2\lambda_1\left[2 + \lambda_1\tilde{R}(2R_2 + \eta\hat{R})\right] +$$

$$\hat{p}^2\lambda_1\eta\left[2 + \lambda_1\tilde{R}(\eta\hat{R} + 8R_2)\right] + \hat{p}^3 4\lambda_1^2\eta^2\tilde{R}R_2 = 0$$

If the household has $b = 0$ (for instance with the 12-month plan) then we just have

$$\left[p + \frac{1}{2}\eta p^2\right]\delta^\gamma\frac{1 - \delta^N}{1 - \delta} = w \tag{3.17}$$

$$\frac{1}{2}\eta p^2 + p - w\frac{1 - \delta}{\delta^\gamma(1 - \delta^N)} = 0$$

The WTP process for a payment plan then identifies $\max_{p,b,\hat{p}} p$ such that either $b, \hat{p} > 0$ and equations 3.15 and 3.13 hold, or $b = \hat{p} = 0$ and equation 3.17 holds.

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, June 2010, *105* (490), 493–505. doi:10.1198/jasa.2009.ap08746.

__ and Guido W. Imbens, "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 2008, *76* (6), 1537–1557. doi:10.3982/ECTA6474.

__ and Javier Gardeazabal, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, March 2003, *93* (1), 113–132. doi:10.1257/000282803321455188.

Abdul Latif Jameel Poverty Action Lab, "The Price is Wrong," Cambridge, MA April 2011. http://www.povertyactionlab.org/publication/the-price-is-wrong.

Abrell, Jan, Anta Ndoye Faye, and Georg Zachmann, "Assessing the Impact of the EU ETS Using Firm Level Data," Working Papers 579, Bruegel July 2011.

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson, "Finding Eldorado: Slavery and Long-Run Development in Colombia," *Journal of Comparative Economics*, November 2012, *40* (4), 534–564. doi:10.1016/j.jce.2012.07.003.

Ades, Alberto F. and Edward L. Glaeser, "Trade and Circuses: Explaining Urban Giants," *The Quarterly Journal of Economics*, February 1995, *110* (1), 195–227. doi:10.2307/2118515.

Ahn, Seung Chan, Young Hoon Lee, and Peter Schmidt, "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, April 2001, *101* (2), 219–255. doi:10.1016/s0304-4076(00)00083-x.

Ahuja, Amrita, Michael Kremer, and Alix Peterson Zwane, "Providing Safe Water: Evidence from Randomized Evaluations," *Annual Review of Resource Economics*, October 2010, *2* (1), 237–256. doi:10.1146/annurev.resource.012809.103919.

Aigner, D. J. and S. F. Chu, "On Estimating the Industry Production Function," *The American Economic Review*, 1968, *58* (4), 826–839.

Ajzenman, Nicolás, Sebastián Galiani, and Enrique Seira, "On the Distributive Costs of Drug-Related Homicides," *Journal of Law and Economics*, forthcoming.

Aldy, Joseph E. and William A. Pizer, "The Competitiveness Impacts of Climate Change Mitigation Policies," Environmental Economics Working Paper 12-01, Duke January 2012.

Anselin, Luc, "Spatial Econometrics," in "A Companion to Theoretical Econometrics," Wiley-Blackwell, January 2003, pp. 310–330. doi:10.1002/9780470996249.ch15.

Arias, María Alejandra, Ana María Ibáñez, and Andrés Zambrano, "Agricultural Production Amid Conflict: The Effects of Shocks, Uncertainty, and Governance of Non-State Armed Actors," DOCUMENTOS CEDE 011005, Universidad de los Andes-CEDE February 2014.

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines," *The Quarterly Journal of Economics*, 2006, *121* (2), 635–672. doi:10.1162/qjec.2006.121.2.635.

_ , James Berry, and Jesse M. Shapiro, "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia," *American Economic Review*, December 2010, *100* (5), 2383–2413. doi:10.1257/aer.100.5.2383.

Attanasio, Orazio P., Pinelopi Koujianou Goldberg, and Ekaterini Kyriazidou, "Credit Constraints in the Market for Consumer Durables: Evidence from Micro Data on Car Loans," *International Economic Review*, May 2008, *49* (2), 401–436. doi:10.1111/j.1468-2354.2008.00485.x.

Bai, Jushan, "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 2009, *77* (4), 1229–1279. doi:10.3982/ECTA6135.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman, "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, January 2015, *7* (1), 1–21. doi:10.1257/app.20140287.

_ , Esther Duflo, Rachel Glennerster, and Cynthia Kinnan, "The Miracle of Microfinance? Evidence from a Randomized Evaluation," *American Economic Journal: Applied Economics*, January 2015, *7* (1), 22–53. doi:10.1257/app.20130533.

Banerjee, Abhijit V. and Esther Duflo, "Chapter 7 Growth Theory through the Lens of Development Economics," *Handbook of Economic Growth*, 2005, pp. 473–552. doi:10.1016/s1574-0684(05)01007-5.

_ and _ , "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program," *The Review of Economic Studies*, February 2014, *81* (2), 572–607. doi:10.1093/restud/rdt046.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?*, W.E. Upjohn Institute for Employment Research, September 1991. doi:10.17848/9780585223940.

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak, "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 1964, *9* (3), 226–232. doi:10.1002/bs.3830090304.

BenYishay, Ariel, Andrew Fraker, Raymond P. Guiteras, Giordano Palloni, Neil Shah, Stuart Shirrell, and Paul Wang, "Microcredit and Willingness to Pay for Environmental Quality: Evidence from a Randomized-Controlled Trial of Finance for Sanitation in Rural Cambodia," Working Paper, University of Maryland 2016.

Berry, James, Gregory M. Fischer, and Raymond P. Guiteras, "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana," MPRC Population Working Paper PWP-MPRC-2015-017 December 2015.

Bleakley, Hoyt and Jeffrey Lin, "Portage and Path Dependence," *The Quarterly Journal of Economics*, May 2012, *127* (2), 587–644. doi:10.1093/qje/qjs011.

Brakman, Steven, Harry Garretsen, and Marc Schramm, "The Strategic Bombing of German Cities during World War II and its Impact on city growth," *Journal of Economic Geography*, April 2004, *4* (2), 201–218. doi:10.1093/jeg/4.2.201.

Campante, Filipe R. and Quoc-Anh Do, "Isolated Capital Cities, Accountability and Corruption: Evidence from US States," *American Economic Review*, 2014, *104* (8), 2456–2481. doi:10.1257/aer.104.8.2456.

_ , _ , and Bernardo V. Guimaraes, "Capital Cities, Conflict, and Misgovernance: Theory and Evidence," Technical Report 2014-13, Sciences Po September 2014.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, "Catastrophic Natural Disasters and Economic Growth," *Review of Economics and Statistics*, December 2013, *95* (5), 1549–1561. doi:10.1162/rest_a_00413.

CBO, "Estimated Impact of the American Recovery and Reinvestment Act on Employment and Economic Output from October 2011 through December 2011," Technical Report, CBO, Washington, DC. 2012.

Chan, Hei Sing (Ron), Shanjun Li, and Fan Zhang, "Firm Competitiveness and the European Union Emissions Trading Scheme," *Energy Policy*, December 2013, *63*, 1056–1064. doi:10.1016/j.enpol.2013.09.032.

CIA, "World Factbook," 2012. Retrieved August 28, 2012. https://www.cia.gov/library/publications/the-world-factbook/geos/bg.html.

Collins, Daryl, Jonathan Morduch, Stuart Rutherford, and Orlanda Ruthven, *Portfolios of the Poor: How the World's Poor Live on \$2 a Day*, Princeton University Press, 2009.

Cox, Michael, Andreas Peichl, Nico Pestel, and Sebastian Siegloch, "Labor Demand Effects of Rising Electricity Prices: Evidence for Germany," Policy Paper 74, IZA December 2013.

Davis, Donald R and David E Weinstein, "Bones, Bombs, and Break Points: The Geography of Economic Activity," *American Economic Review*, December 2002, *92* (5), 1269–1289. doi:10.1257/000282802762024502.

Davis, James C. and J.Vernon Henderson, "Evidence on the Political Economy of the Urbanization Process," *Journal of Urban Economics*, January 2003, *53* (1), 98–125. doi:10.1016/S0094-1190(02)00504-1.

Davis, Steven J. and John Haltiwanger, "Sectoral Job Creation and Destruction Responses to Oil Price Changes," *Journal of Monetary Economics*, December 2001, *48* (3), 465–512. doi:10.1016/s0304-3932(01)00086-1.

de Mel, Suresh, David McKenzie, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, November 2008, *123* (4), 1329–1372. doi:10.1162/qjec.2008.123.4.1329.

Deaton, Angus, "Saving and Liquidity Constraints," *Econometrica*, September 1991, *59* (5), 1221–1248. doi:10.2307/2938366.

Deschenes, Olivier, "Climate Policy and Labor Markets," in Don Fullerton and Catherine Wolfram, eds., *The Design and Implementation of U.S. Climate Policy*, University of Chicago Press, 2012.

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons, "Happiness on Tap: Piped Water Adoption in Urban Morocco," *American Economic Journal: Economic Policy*, November 2012, *4* (4), 68–99. doi:10.1257/pol.4.4.68.

Dhar, Paramita and Stephen L Ross, "School District Quality and Property Values: Examining Differences along School District Boundaries," *Journal of Urban Economics*, January 2012, *71* (1), 18–25. doi:10.1016/j.jue.2011.08.003.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," Working Paper 19213, National Bureau of Economic Research July 2013. doi:10.3386/w19213.

Dube, Arindrajit and Ben Zipperer, "Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies," Technical Report 8944, IZA March 2015.

_ , T. William Lester, and Michael Reich, "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties," *Review of Economics and Statistics*, November 2010, *92* (4), 945–964. doi:10.1162/rest_a_00039.

Dupas, Pascaline, "Health Behavior in Developing Countries," *Annual Review of Economics*, 2011, *3* (1), 425–449. doi:10.1146/annurev-economics-111809-125029.

_ , "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment," *Econometrica*, January 2014, *82* (1), 197–228. doi:10.3982/ECTA9508.

_ and Jonathan Robinson, "Why Don't the Poor Save More? Evidence from Health Savings Experiments," *American Economic Review*, June 2013, *103* (4), 1138–1171. doi:10.1257/aer.103.4.1138.

Environmental Protection Agency, "eGRID 9th edition Version 1.0 - Summary Tables for year 2010 data," 2014. https://www.epa.gov/energy/egrid-9th-edition-version-10-summary-tables-year-2010-data.

_ , "Regulatory Impact Analysis for the Proposed Carbon Pollution Guidelines for Existing Power Plants and Emission Standards for Modified and Reconstructed Power Plants," June 2014. http://www2.epa.gov/carbon-pollution-standards/clean-power-plan-proposed-rule-regulatory-impact-analysis.

_ , "Regulatory Impact Analysis for the Clean Power Plan Final Rule," 2015. http://www2.epa.gov/cleanpowerplan/clean-power-plan-final-rule-regulatory-impact-analysis.

Epstein, David G., *Brasília, Plan and Reality: A Study of Planned and Spontaneous Urban Development*, Berkeley: University of California Press, 1973.

Flues, Florens and Benjamin Johannes Lutz, "The Effect of Electricity Taxation on the German Manufacturing Sector: A Regression Discontinuity Approach.," December 2014. Mimeo. https://ideas.repec.org/p/zbw/zewdip/15013.html.

Foster, Andrew D. and Mark R. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209. doi:10.1086/601447.

_ and _ , "Microeconomics of Technology Adoption," *Annual Review of Economics*, 2010, *2* (1), 395–424. doi:10.1146/annurev.economics.102308.124433.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, August 1999, *114* (3), 739–767. doi:10.1162/003355399556133.

Galiani, Sebastian and Sukkoo Kim, "Political Centralization and Urban Primacy: Evidence from National and Provincial Capitals in the Americas," in Dora L. Costa and Naomi R. Lamoreaux, eds., *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*, University of Chicago Press, November 2008, chapter 4, pp. 121–153.

Glaeser, Edward L. and Joseph Gyourko, "Urban Decline and Durable Housing," *Journal of Political Economy*, April 2005, *113* (2), 345–375. doi:10.1086/427465.

Gonçalves, Sílvia and Benoit Perron, "Bootstrapping Factor-Augmented Regression Models," *Journal of Econometrics*, September 2014, *182* (1), 156–173. doi:10.1016/j.jeconom.2014.04.015.

Gordon, David L. A., *Planning Twentieth Century Capital Cities*, London; New York: Routledge, 2006.

Greenstone, Michael, "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977: Clean Air Act Amendments and the Census of Manufactures," *Journal of Political Economy*, December 2002, *110* (6), 1175–1219. doi:10.1086/342808.

_ , Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, June 2010, *118* (3), 536–598. doi:10.1086/653714.

Griliches, Zvi, "Sibling Models and Data in Economics: Beginnings of a Survey," *Journal of Political Economy*, January 1979, *87* (S5), S37–S64. doi:10.1086/260822.

Guiteras, Raymond P., David I. Levine, Stephen P. Luby, Thomas H. Polley, Kaniz Khatun e Jannat, and Leanne Unicomb, "Disgust, Shame and Soapy Water: Tests of Novel Interventions to Promote Safe Water and Hygiene," *Journal of the Association of Environmental and Resource Economists*, 2016. Forthcoming.

Hamermesh, Daniel, *Labor Demand*, Princeton, NJ: Princeton University Press, 1993.

Hamilton, James D., "Oil and the Macroeconomy," *The New Palgrave Dictionary of Economics*, February 2008, pp. 172–177. doi:10.1057/9780230226203.1215.

Helms, Brigit and Xavier Reille, "Interest Rate Ceilings and Microfinance: The Story So Far," Technical Report, CGAP 2004.

Henderson, J. Vernon and Anthony J. Venables, "The Dynamics of City Formation," *Review of Economic Dynamics*, April 2009, *12* (2), 233–254. doi:10.1016/j.red.2008.06.003.

Holla, Alaka and Michael Kremer, "Pricing and Access: Lessons from Randomized Evaluations in Education and Health," Working Paper 158, Center for Global Development 2009.

Holmes, Thomas J., "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy*, August 1998, *106* (4), 667–705. doi:10.1086/250026.

Horowitz, John K., "The Becker-DeGroot-Marschak Mechanism Is Not Necessarily Incentive Compatible, Even for Non-Random Goods," *Economics Letters*, October 2006, *93* (1), 6–11. doi:10.1016/j.econlet.2006.03.033.

Huang, Rocco R., "Evaluating the Real Effect of Bank Branching Deregulation: Comparing Contiguous Counties across US State Borders," *Journal of Financial Economics*, March 2008, *87* (3), 678–705. doi:10.1016/j.jfineco.2007.01.004.

Joskow, Paul L., "Markets for Power in the United States: An Interim Assessment," *The Energy Journal*, January 2006, *27* (1). doi:10.5547/issn0195-6574-ej-vol27-no1-2.

_ , Douglas R. Bohi, and Frank M. Gollop, "Regulatory Failure, Regulatory Reform,

and Structural Change in the Electrical Power Industry," *Brookings Papers on Economic Activity: Microeconomics*, 1989, *1989*, 125–208. doi:10.2307/2534721.

Kahn, Matthew E. and Erin T. Mansur, "Do Local Energy Prices and Regulation Affect the Geographic Concentration of Employment?," *Journal of Public Economics*, May 2013, *101*, 105–114. doi:10.1016/j.jpubeco.2013.03.002.

Karlan, Dean S and Jonathan Zinman, "Credit Elasticities in Less-Developed Economies: Implications for Microfinance," *American Economic Review*, May 2008, *98* (3), 1040–1068. doi:10.1257/aer.98.3.1040.

Kessler, Judd B. and C. Yiwei Zhang, "Behavioural Economics and Health," in Roger Detels, Martin Guillford, Quarraisha Abdool Karim, and Chorh Chuan Tan, eds., *Oxford Textbook of Public Health*, 6 ed., Vol. 2, Oxford University Press, 2015, pp. 775–789.

Kilian, Lutz, "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature*, November 2008, *46* (4), 871–909. doi:10.1257/jel.46.4.871.

Kline, Patrick and Enrico Moretti, "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *The Quarterly Journal of Economics*, November 2013, *129* (1), 275–331. doi:10.1093/qje/qjt034.

Krugman, Paul and Raul Livas Elizondo, "Trade Policy and the Third World Metropolis," *Journal of Development Economics*, April 1996, *49* (1), 137–150. doi:10.1016/0304-3878(95)00055-0.

Lahmeyer, Jan, "Population Statistics: Historical Demography of All Countries, Their Divisions and Towns," 2006. www.populstat.info. Retrieved January 1, 2015.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 1997, *112* (2), 443–477. doi:10.1162/003355397555253.

Levine, David I. and Carolyn Cotterman, "What Impedes Efficient Adoption of Products? Evidence from Randomized Variation in Sales Offers for Improved Cookstoves in Uganda," Working Paper Series, UC Berkeley: Institute for Research on Labor and Employment March 2012.

Linn, Joshua, "Why Do Energy Prices Matter? The Role of Interindustry Linkages in U.S. Manufacturing," *Economic Inquiry*, July 2009, *47* (3), 549–567. doi:10.1111/j.1465-7295.2008.00168.x.

Luoto, Jill, Minhaj Mahmud, Jeff Albert, Stephen Luby, Nusrat Najnin, Leanne Unicomb, and David I. Levine, "Learning to Dislike Safe Water Products: Results from a Randomized Controlled Trial of the Effects of Direct and Peer Experience on Willingness to Pay," *Environmental Science & Technology*, 2012, *46* (11), 6244–6251. doi:10.1021/es2027967.

_, Nusrat Najnin, Minhaj Mahmud, Jeff Albert, M. Sirajul Islam, Stephen Luby, Leanne Unicomb, and David I. Levine, "What Point-of-Use Water Treatment Products Do Consumers Use? Evidence from a Randomized Controlled Trial among the Urban Poor in Bangladesh," *PLoS ONE*, October 2011, *6* (10). doi:10.1371/journal.pone.0026132.

Mahajan, Aprajit and Alessandro Tarozzi, "Time Inconsistency, Expectations and Technology Adoption: The Case of Insecticide Treated Nets," Economic Research Initiatives at Duke (ERID) Working Paper, Duke 2011.

Mammen, Enno, "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *Annals of Statistics*, March 1993, *21* (1), 255–285. doi:10.1214/aos/1176349025.

Martin, Ralf, Laure de Preux, and Ulrich Wagner, "The Impacts of the Climate Change Levy on Manufacturing: Evidence from Microdata," Technical Report, National Bureau of Economic Research September 2011. doi:10.3386/w17446.

Mendes, Manuel P., *O Cerrado de Casaca*, Thesaurus Editora, 1995.

Miguel, Edward and Gérard Roland, "The Long-Run Impact of Bombing Vietnam," *Journal of Development Economics*, September 2011, *96* (1), 1–15. doi:10.1016/j.jdeveco.2010.07.004.

Mobarak, Ahmed Mushfiq, Puneet Dwivedi, Robert Bailis, Lynn Hildemann, and Grant Miller, "Low Demand for Nontraditional Cookstove Technologies," *Proceedings of the National Academy of Sciences*, 2012, *109* (27), 10815–10820. doi:10.1073/pnas.1115571109.

Moretti, Enrico, "Local Multipliers," *American Economic Review*, May 2010, *100* (2), 373–377. doi:10.1257/aer.100.2.373.

_, "Local Labor Markets," *Handbook of Labor Economics*, 2011, pp. 1237–1313. doi:10.1016/s0169-7218(11)02412-9.

Morten, Melanie and Jaqueline Oliveira, "Migration, Roads and Labor Market Integration: Evidence from a Planned Capital City," July 2014. http://www.stanford.edu/~memorten/Melanie_Morten_files/Brasilia_March2014.pdf.

Naidu, Suresh, "Suffrage, Schooling, and Sorting in the Post-Bellum U.S. South," Working Paper 18129, National Bureau of Economic Research June 2012. doi:10.3386/w18129.

Neumark, David, "Biases in Twin Estimates of the Return to Schooling," *Economics of Education Review*, April 1999, *18* (2), 143–148. doi:10.1016/s0272-7757(97)00022-8.

O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, March 1999, *89* (1), 103–124. doi:10.1257/aer.89.1.103.

Otsu, Taisuke and Yoshiyasu Rai, "Bootstrap Inference of Matching Estimators for

Average Treatment Effects," STICERD - Econometrics Paper Series /2015/580, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE January 2015.

Paskoff, Paul F., "Measures of War: A Quantitative Examination of the Civil War's Destructiveness in the Confederacy," *Civil War History*, 2008, *54* (1), 35–62. doi:10.1353/cwh.2008.0007.

Pesaran, M. Hashem, "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, July 2006, *74* (4), 967–1012. doi:10.1111/j.1468-0262.2006.00692.x.

Rauch, James E., "Does History Matter Only When It Matters Little? The Case of City-Industry Location," *The Quarterly Journal of Economics*, August 1993, *108* (3), 843–867. doi:10.2307/2118410.

Roodman, David, "Quick: What's the Grameen Bank's Interest Rate?," Microfinance Open Book Blog September 2010. http://www.cgdev.org/blog/quick-whats-grameen-banks-interest-rate.

RSSSF Brazil, "Rec.Sport.Soccer Statistics Foundation - Brazil," 2014. www.rsssfbrasil.com/ (accessed 2015-01-01).

Sharma, Shantanu Nandan, "Urban Dreams: New Cities to Sprout across India," The Economic Times March 2010.

Soo, Kwok Tong, "Zipf's Law for Cities: A Cross-Country Investigation," *Regional Science and Urban Economics*, May 2005, *35* (3), 239–263. doi:10.1016/j.regsciurbeco.2004.04.004.

Tarozzi, Alessandro, Aprajit Mahajan, Brian Blackburn, Dan Kopf, Lakshmi Krishnan, and Joanne Yoong, "Micro-loans, Insecticide-Treated Bednets, and Malaria: Evidence from a Randomized Controlled Trial in Orissa, India," *American Economic Review*, 2014, *104* (7), 1909–1941. doi:10.1257/aer.104.7.1909.

Thaler, Richard H. and Shlomo Benartzi, "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 2004, *112* (S1), S164–S187. doi:10.1086/380085.

Wagner, Ulrich J., Mirabelle Muuls, Ralf Martin, and Jonathan Colmer, "The Causal Effect of the European Union Emissions Trading Scheme: Evidence from French Manufacturing Plants," December 2014. Mimeo.

Walker, W. Reed, "The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce," *The Quarterly Journal of Economics*, August 2013, *128* (4), 1787–1835. doi:10.1093/qje/qjt022.

Woutersen, Tiemen and John C. Ham, "Confidence Sets for Continuous and Discon-

tinuous Functions of Parameters," Cemmap CWP23/13, University College London October 2013.

Zou, Ben, "The Local Economic Impacts of Military Personnel Contractions," February 2015. mimeo. http://zouben.weebly.com/research.html.