Title
Essays in Urban Economics and International Trade

Author
Heilmann, Kilian Tobias

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in Urban Economics and International Trade

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Kilian Tobias Heilmann

Committee in charge:

Professor Gordon Hanson, Chair
Professor Eli Berman
Professor Lawrence Broz
Professor Gordon McCord
Professor Marc-Andreas Muendler

2017

The dissertation of Kilian Tobias Heilmann is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2017

DEDICATION

To my family and friends.

EPIGRAPH

I may not have gone where I intended to go,

but I think I have ended up where I needed to be.

— D.A.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor Gordon Hanson for invaluable guidance and support. I am also grateful to my dissertation committee members Marc-Andreas Muendler, Eli Berman, Gordon McCord, and Lawrence Broz for academic and personal advice.

I further benefited from the rest of the Economics faculty with their . Special thanks to Ran Goldblatt for opening up a new door in spatial research.

I would also like to thank my friends and colleagues at UCSD for their support during six challenging but great years. Without you, I would not be here. You made me a better researcher and a better person.

Chapter 1, in full, is currently being prepared for submission for publication of the material. The dissertation author, Kilian Tobias Heilmann, was the sole author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in: Kilian Heilmann, "Does political conflict hurt trade? Evidence from consumer boycotts", *Journal of International Economics*, 99, 2016. The dissertation author was the sole author of this paper.

Chapter 3 is coauthored with Ran Goldblatt, Gordon Hanson, Amit Khandelwal, and Yonatan Vaizman, and in full, is currently being prepared for submission for publication of the material. I thank my coauthors for the permission to use this chapter in my dissertation. The dissertation author, Kilian Tobias Heilmann, was the primary author of this paper.

VITA

2009	Magister Artium, Johannes Gutenberg University Mainz
2010	MSc in Economics and Econometrics, University of Nottingham
2011-2017	Graduate Teaching Assistant, University of California, San Diego
2017	Ph. D. in Economics, University of California, San Diego

PUBLICATIONS

Kilian Heilmann, "Does political conflict hurt trade? Evidence from consumer boycotts", *Journal of International Economics*, 99, 2016.

ABSTRACT OF THE DISSERTATION

Essays in Urban Economics and International Trade

by

Kilian Tobias Heilmann

Doctor of Philosophy in Economics

University of California, San Diego, 2017

Professor Gordon Hanson, Chair

This dissertation explores three topics in Urban Economics and International Trade. Chapter 1 measures the effect of transit access on neighborhood incomes by exploiting a quasi-experimental setting in Dallas. I show that income in neighborhoods that received rail access increases compared to neighborhoods that were promised to receive access, but did not receive it. The treatment effect is positively correlated with initial neighborhood income and highlights the role of transit as an incubator for income segregation. Chapter 2 estimates the impact of international conflict on bilateral trade relations using several incidents of politically motivated boycotts. I find large reductions in exports from the boycotted to the boycotting countries. Product-level results are in line with intuition and most effective for consumer goods while having at most a temporary effect on intermediates and capital goods. Chapter 3 explores the usage of Landsat satellite imagery for the measurement of economic outcomes at small geographies.

Chapter 1

Transit Access and Neighborhood Segregation. A Study of the Dallas Light Rail System

Abstract: I study the effect of transit access on neighborhood incomes by exploiting a quasi-experimental setting of an extensively planned, but only partially built urban rail system in Dallas. I show that neighborhood income in census tracts that received rail access increases compared to neighborhoods that were promised to receive access, but did not due to funding cuts. The treatment effect is positively correlated with initial neighborhood income and negative for the poorest tracts. This reconciles gentrification and "poverty magnet" effects of rail infrastructure found in the earlier literature and highlights the role of transit as a potential incubator for income segregation.

JEL classification: R3, R4, L92

Keywords: Transit provision; Income segregation; Gentrification; Spatial sorting

1.1 Introduction

Increasing housing prices not only in US metropolitan areas, but in cities worldwide, have highlighted the issue of the spatial distribution of income. Income segregation, that is the uneven geographic distribution of income groups within a city, has been on the rise. Gentrification of formerly lower-income neighborhoods and the resulting displacement have left many poor residents unable to afford housing in but the most remote and dilapidated neighborhoods of urban areas. In economics, the issue of spatial sorting by income has traditionally been thought of as an optimal outcome induced by different preferences for public goods (Tiebout, 1956). The literature on the spatial clustering of the poor however has documented sizable negative spillovers of low-income neighborhoods (Oreopoulos, 2003) and recently gained interest in the spatial determinants of intergenerational mobility (Chetty et al., 2014).

While the surge in income segregation can be partially explained by increased income inequality in the US over the last four decades (Reardon and Bischoff, 2011; Fry and Taylor, 2012), and is deeply entangled with historic and current racial segregation (Cutler et al., 1999), it is a phenomenon in itself and its causes are not well understood. This paper sets out to identify one potential mechanism through which income segregation can be exacerbated, namely the transportation system of a city. I focus on public transportation as a specific part of the general transportation infrastructure that has seen a recent rebound in the US.

Over the past thirty years, many American cities have invested heavily into the construction of rail-based transit systems. These include the large western cities of San Jose (1987), Sacramento (1987), Los Angeles (1990), Seattle (2003), and the fast-growing sunbelt urban areas Dallas (1996), Houston (2004), and Phoenix (2008) that have been centered around the private automobile and experienced severe urban sprawl. Many of these cities have opted to build light rail systems in order to relieve car dependence, fight traffic congestion, and revitalize impoverished neighborhoods. More cities are planning to build new transit systems or are contemplating extending their existing lines.[1]

Despite the boom of urban rail, new transit infrastructure receives fierce opposition in the form of the "not in my backyard" (NIMBY) type. Residents of neighborhoods affected by transit plans often oppose new infrastructure and block transit projects (Altshuler and Luberoff, 2003). The resistance to rail transit stems from fears of transit-induced crime, noise, traffic, physical separation of neighborhoods by rail tracks, and undesired change in the composition of residents. Local home owners often worry about the attraction of poor transit-dependent households and the decline of neighborhoods in response to rail construction. To give a historical example, Osofsky (1966) argues that the construction of subways in Harlem, New York, turned the previously thriving area into a low-income neighborhood. Proponents of rail-based transit in contrast argue that

[1]For example, voters in Phoenix, AZ passed proposition 104 in August 2015 that increases sales taxes for 35 years to fund further extension of the local light rail system. San Diego plans to add an additional line to its trolley system until 2021.

rail infrastructure has positive local effects on the communities along transit corridors. These stated benefits include increased business investment, job creation, and residential development through better transit access.[2]

This discourse exemplifies the complex mechanisms through which transit infrastructure can influence neighborhoods and their economic and demographic composition. Arguments between proponents and opponents have been exchanged in debates and at the ballot box. For example, before the successful passing of the streetcar proposal in 2012, Kansas City residents rejected earlier proposals in every year between 1997 and 2003, and again in 2008. Two years after approval, a ballot measure to add new lines to the system failed again. Most recently, measures to introduce urban mass transit proposals failed in Austin, Texas (2014) and St Petersburg, Florida (2014), but were successful in Phoenix (2016).

Empirical studies in urban economics that could settle the opposing opinions on the effect of rail transit on neighborhood composition have shown contradictory results. On the one hand, Glaeser et al. (2008) observe slight increases in poverty rates around newly-built rail stations. This is consistent with the notion that poor, transit-dependent residents move towards rail access, thus confirming fears of local residents that transit stations act as "poverty magnets". Kahn (2007) however, in a study of 14 cities that expanded rail from 1970-2000, documents heterogeneous effects within and between cities. He shows that rail transit can cause gentrification, especially around "walk-and-ride" stations, as measured by a higher share of college graduates, but that suburban park-and-ride stations often experience increases in poverty.

This paper aims to advance the economic evidence by estimating the causal effect of public transportation infrastructure on neighborhood composition in a quasi-experimental setting. I analyze demographic and economic characteristics of neighbor-

[2]For example, the American Public Transportation Association summarizes the growth arguments for public transport on http://www.publictransportation.org/benefits/Pages/default.aspx.

hoods in Dallas, Texas, before and after the expansion of its urban rail network DART (Dallas Area Rapid Transit). The local impacts of public transport improvements are difficult to study due to a lack of exogenous variation in transit provision. The planning process of large transit networks necessarily leads to selection into the treatment. This paper deals with the endogeneity issue by exploiting a natural experiment of a planned but only partially implemented urban rail network that was pared down due to funding cuts in the wake of the Savings and Loan crisis of the 1980s. The institutional history of the Dallas transit system provides an unusual opportunity to reduce selection bias by observing the initial selection decision of transit planners that was eventually overturned by the general economic climate and was thereafter on pure cost considerations. Building on Billings (2011) and Mayer and Trevien (2013), I use historical transit plans to implement a difference-in-differences design comparing neighborhoods that received transit access to those that were promised to receive access, but did not receive it. The Dallas-Fort Worth metropolitan lends itself well to the study of rail transit as, like many other cities that have implemented or are considering implementing rail systems, it superimposed a rail system on a largely car-dependent, sprawling city with high employment decentralization. Furthermore, mostly unhindered by geographical limitations that could confound the effect of transportation, the area comes close to von Thünen (1826)'s postulated "flat featureless plain".

I find an economically large effect on income in neighborhoods that received rail access. The effect is positive on average, and represents an increase of about $5,300 or one fifth of a standard deviation. The treatment effect, however, correlates positively with initial neighborhood income. In fact, poor tracts become poorer while rich areas become richer, increasing the spatial dispersion of income. This reconciles the conflicting findings of the previous literature by documenting that transit can have both a gentrification as well as a "poverty magnet" effect. While it has been recognized in the literature that improvements in road infrastructure can lead to a more extreme spatial population distributions through its effect on suburbanization (Baum-Snow, 2007) and political polarization (Nall, 2015), the spatial sorting effect of transit has received only modest attention. Employing an equilibrium sorting model, Bayer and McMillan (2012) predict that a drop in commuting costs leads to increased income segregation. My empirical results show that public transportation indeed leads to a more extreme spatial distribution of incomes. This finding is in contrast to cross-sectional studies that document a negative correlation between income inequality and transit provision (Blumenberg and Ong, 2001; Sanchez, 2002). Employing a new dataset of precisely geocoded industry locations, I find evidence of a small increase in business activity around rail stations, but no shift in the industry composition towards urban amenities such as restaurants or retail businesses that are in general associated with gentrification. Instead, the increase of new businesses is largely captured by chain stores.

This paper builds up on a small but growing economic literature on urban rail infrastructure and its effect on the shape of cities. The focus of most studies has been to document increasing property values around stations, consistent with the economic argument that improved travel options are capitalized in rents (Bowes and Ihlanfeldt, 2001; Gibbons and Machin, 2005; Billings, 2011; Hewitt and Hewitt, 2012). Cervero and Duncan (2002) and Ko and Cao (2013) find that this also holds for commercial and industrial properties, while Ryan (2005) contests the latter findings. Beyond the focus on property values, other facets of rail provision on city structure have been studied. Bollinger and Ihlanfeldt (1997) show that Atlanta's MARTA rail system had little impact on population and employment levels in transit neighborhoods. Studying the same system, Ihlanfeldt (2003) analyzes rail transit's effect on crime and finds mixed results with increasing crime rates in the inner city and decreasing crime in suburbs. Brooks and Lutz (2013) provide evidence for long lasting effects of rail transit by showing that abandoned interurban rail stations in Los Angeles still predict higher density today. Anderson (2014) documents transit's high marginal impact on easing roadway delays, thus rationalizing public support for public transportation by both transit users and non-users, who would otherwise face increased road congestion. On a global scale, Gonzalez-Navarro and Turner (2014) show that while subway systems have little effect on the size of cities, they cause them to be more decentralized. Concerning methodology, my paper is in the spirit of previous studies that have exploited exogenous variation in the provision of transport infrastructure, most notably for interstate highways (Baum-Snow, 2007), road lane length (Duranton and Turner, 2011), and railroads (Donaldson, forthcoming).

The paper is organized as follows: Section 2 portrays the historical background of the Dallas rail system while Section 3 describes in detail the data sources used. Then, Section 4 outlines the identification strategy and empirical implementation. Section 5 presents the results and Section 6 concludes the findings.

1.2 Background: The Dallas Area Rapid Transit System

The history of the Dallas Area Rapid Transit (DART) system provides an unusual opportunity to measure the causal effect of transit infrastructure due to the observed selection process. The system was planned during the early 1980s, when the Dallas economy was profiting from the Texas oil boom that led to a building boom.[3] In this euphoric setting, plans for a modern transportation system were emerging that aimed at relieving the perennial traffic congestion in the city[4] and in 1983, the electorate in Dallas

[3] To illustrate the importance of that period for the shape of the city of Dallas: 14 out of the tallest 25 buildings in Dallas as of today were completed between 1979 and 1985.

[4] "Dallas Rail Plan Throws Texas-Sized DART at Traffic Congestion", Washington Post, Aug. 16, 1983.

and 13 suburban municipalities approved a one-cent sales tax to fund a regional transit authority to implement an urban rail system.[5]

The 1983 proposal planned to provide $8.9 billion to improve the existing bus network and included 160 miles of guided rail.[6] Figure 1.1 shows the approved transit service plan which featured 10 radial lines into the suburbs. With exception of the downtown part that required new tracks, the transportation planners made use of pre-existing freight rail right-of-way in in the metropolitan area that would have been converted into light rail lines. This approach lowered construction costs substantially and it did not require expensive tunnel boring or rezoning of land for transit use.

The financial situation changed dramatically in the mid-1980s, when Dallas experienced a severe recession during the Savings and Loan crisis that eventually led to the failure of many banking institutions based in the city.[7] The crumbling economy also had its effect on the metropolitan transit plans which depended heavily on sales taxes for financing. The original financial plan assumed a yearly tax revenue growth of 1.5% (DART and Peat, 1983) and projected a continuously high price of oil (DART, 1986a) that was not realized. In 1985, a financial review of DART's books revealed that funding was not sufficient to implement the entire 1983 rail plan.

Faced with the new budget situation,[8] the transit planners had to pare down the system. With the explicitly stated goal to build as much rail as possible, they made made use of the historical coincidence that several freight companies in the Dallas-Fort Worth area had folded or merged. DART was able to acquire their available freight track at low cost and convert it to light rail lines. Subsequently, the new transit plan that was

[5]"Transit plan wins in Dallas Region", New York Times, Aug. 15, 1983.

[6]"Dallas Voters Approve Tax to Finance $8.75 Billion Rail System", Washington Post, Aug 15, 1983.

[7]Out of the 2,231 banks that failed or required financial assistance by the government between 1983 and 1990, more than one third (748) were based in Texas. 123 of these banks were headquartered in the Dallas-Fort Worth metropolitan area. Source: Federal Deposit Insurance Corporation.

[8]The construction cost of the 1983 plan was estimated to cost 4.5 billion dollars, but only 3 billion were available (DART, 1986b).

approved in 1989 featured lines that were running on unused freight track in the west and northeast and otherwise axed the branches in the southern and western suburbs.[9] This new rail system plan is depicted in Figure 1.2.

Construction of the transit system commenced in 1990 and a starter system of two light rail lines opened in 1996.[10] At the same time the Trinity Rail Express (TRE), a commuter rail line to the neighboring city of Irving, began its service. Major extensions of the system followed. The expansion of the commuter rail to downtown Fort Worth was completed in 2001 and extensions to the light rail starter system to the northern suburbs followed in 2002. Further extensions in late 2009, 2010, 2012, and 2014 added links to the DFW airport and more suburbs in the northwest and southeast, making DART currently the largest light rail system in the US.[11]

Usage of the system remains low when put in contrast to other modes of transportation. DART reports a rail ridership of 29.9 million passenger trips for the 2015 fiscal year, while during the same time 36.5 million bus trips were made.[12] Earlier commuting data from the ACS reports that in 2009, less than 2% of all work trips were made by any kind of public transportation in the Dallas-Fort Worth-Arlington MSA (McKenzie and Rapino, 2011). Regarding ridership demographics, the rail system is mainly used by the poor. According to a 2014 passenger survey (NCTCG, 2014), almost 80% of the

[9] In specific, the Rock Island Railroad filed bankruptcy in 1875 and DART subsequently constructed the line to Fort Worth on its tracks. Similarly, the Denver and Rio Grande Western Railroad merged with the Southern Pacific Railroad in 1988 and DART converted the latter's right of way into two light rail lines to Plano and Garland.

[10] "Dallas Opening Southwest's First Rail Transit", The New York Times, June 14, 1996.

[11] Except for the single commuter rail line to Fort Worth with which it shares the ticketing system, the DART system operates as a light rail. While lacking a formal definition, light rail systems are usually characterized by electric-propelled vehicles that have an intermediate capacity between classic streetcars (trams) and heavy commuter rail. As in Dallas, they often run on mixed guideway that includes both exclusive right-of-way (such as running on embankments, elevated tracks, or in subway tunnels) as well as at-grade sections on urban streets. Unlike classic heavy rail metro systems such as the New York subway, light rail systems like DART have a higher visual and audial impact on their environment as stations and tracks are primarily above ground. Light rail has been the dominant mode of choice of city planners in the US due to their cost-advantage. While cheaper to construct than subterranean metro lines, they nevertheless provide the perceived comfort of fixed rail and are not constrained by federal rail regulation.

[12] https://www.dart.org/about/dartfacts.asp.

respondents reported an annual household income of less than \$50,000, thus the majority of DART riders earn below the Dallas-Fort Worth median household income of \$59,175. At the same time, 81.2% of all respondents indicated that they work at least part-time, while 10.4% of riders were actively seeking work.

1.3 Data Sources

The empirical analysis is based on a balanced census tract-level panel dataset covering the neighborhoods affected by the Dallas rail system. In this section, I describe the data collection process for the economic and geographic data.

1.3.1 Census Data

Data on neighborhood demographics and economic characteristics are from the long form of the censuses (1970-2000) and the 2006-2010 American Community Survey (ACS) 5-Year estimates that replaced the census long forms in the early 2000s. The census long form provides extensive data on racial characteristics, incomes, and employment and is based on a sample of about one in six households. The ACS samples about 2% of the population each year and combines the estimates of the different years. The unit of observation is the census tract, an area that is designed to encompass around 4,000 people and to approximate a city neighborhood of similar socioeconomic characteristics. In practice, tract populations vary from about 2,500 to 8,000 residents. To alleviate the issue of changing boundaries over time, I use census tract data provided by the Neighborhood Change Database (NCDB). The NCDB contains historical census data that has been re-weighted to 2010 census tract boundaries and thus allows to compare economic and population characteristics for consistent geographic units over time. Additionally, I use reweighted census block group data as a robustness check. Census block groups are finer than census tracts, but change their boundaries more frequently and thus are more difficult to compare over time. I obtain census boundaries from the US Census TIGER database.

1.3.2 Business Data

To measure business activity, I make use of official sales tax license data collected and provided by the office of the Texas Comptroller of Public Accounts. The Texas business sales register provides data for every business license issued in the state of Texas. It entails names and contact details of the license holder, establishment name, and business address along with the date of the first sale. In addition, it supplies the 6-digit NAICS industry classification. I retrieved the data for the Dallas metropolitan area for the years from 1995 to 2014 inclusive. The establishment addresses were geocoded using ESRI ArcMap 10.2 and merged to the census tracts described above. I collapse the daily observations to a yearly balanced panel data set of business openings at the census tract level for the years 1995-2014.

1.3.3 Transit Data

The main source for constructing my treatment and control neighborhoods is the 1983 Dallas Transit Service Plan which I obtained as a digitized scan (Figure 1.1). This document summarizes the then current planning stage of the Dallas transportation projects that were approved by voters in the same year. It shows the corridors of the proposed rail system as well as their funding status and the type of track (elevated or at-grade). The plan further identifies approximate areas that were supposed to serve as stations of the rail system. I used ArcMap to overlay this plan with an actual street map and then geo-referenced the lines and stations to be saved as a GIS shape file. As described above, this plan never got fully realized. I obtained data on the actually built rail system by geo-referencing Google Maps imagery. I follow the tracks of the urban rail system as of June 1999 and create a shape file of the then current lines. I also use the exact geographic location of station platforms to create a geo-referenced dataset of transit access points.

1.3.4 Geographic Data

To support the demographic and business data with city-location specific data, I draw further geographic data from various sources. To quantify the location of each census tract within the metropolitan area, I define a central point in both the Dallas and Fort Worth central business districts (CBD). These locations are based on the 1982 Census of Retail Trade CBD definitions. For each census tract I then calculate the minimum distance of each tract centroid to either CBD. For road infrastructure, I downloaded the street network shape file from the OpenStreetMap project including all roadways in the state of Texas and calculated the linear distance to the nearest expressway for each census tract centroid.

1.4 Methodology

1.4.1 Identification

As with other large infrastructure projects like highways and airports, rail lines and stations are not randomly assigned to neighborhoods, but are designed in response to current and future expected demand. If the planning process is correlated with other characteristics that also drive neighborhood incomes, then it is not possible to recover the causal effect with correlational studies that simply compare places with rail against places without rail. The selection process for transit is not always clear: If rail lines are prioritized towards economically declining areas in order to develop them, the estimate might pick up part of the decline that the rail system is designed to alleviate. On the other hand, if the goal is to connect economically prospering neighborhoods with public transportation, then the bias might be positive as the estimate falsely absorbs the positive trend that would have been taking place even in absence of any rail construction.

The planning process of the Dallas light rail system allows me to address the fundamental selection issue in evaluating large infrastructure projects. Firstly, similar to Donaldson (forthcoming), Billings (2011), and Mayer and Trevien (2013) I use the not built portion of the network as the control group to transit neighborhoods that were excluded due to a state-wide funding shock unrelated to local conditions. Comparing neighborhoods that were both chosen to be part of the initial system are likely similar on characteristics that might be only observable to the transit planners but unobservable to the econometrician. This helps to overcome the problem of selection, as it differences out the transit planners' initial preferences. For example, if the initial plan favored neighborhoods that were predicted to grow faster than areas that were never considered to receive rail access, then comparing neighborhoods within the initial plan would cancel out this positive trend.

Secondly, I use the cost consideration of the transit planners after the funding shock as an exogenous source of transit variation between neighborhoods. To estimate the true causal effect of rail access, the planned rail lines of the initial system that were excluded after the funding cut must have been excluded for reasons that are not correlated with unobserved neighborhood characteristics that also directly affect neighborhood income. This rules out that neighborhoods along rail lines that were defunded experienced different trends than those along the actually built lines. As mentioned earlier, the decision in the case of DART was based on availability of previously unused freight track that became available due to consolidation in the freight industry. The exogenous variation in transit access then comes from the fact that some track could be bought cheaply by DART because the owner company ceased to exist.[13]

Large infrastructure projects often go through lengthy planning and construction processes and economic agents might react to plans and internalize future benefits and costs before the system is in place. Anticipatory effects in response to future transit stations have been documented by McDonald and Osuji (1995) and McMillen and McDonald (2004). However, if anticipatory effects are important for the outcome variables in my study, this will work against finding an impact as all adjustment to the opening transit system would already have been completed by time of opening. Hence, this would cause the difference-in-differences estimates to be biased towards zero. The fact that I still find significant effects shows that I indeed capture at least part the effect of the actual transit access that is unrelated to eventual property appreciation around rail stations. Thus, my estimated treatment can be interpreted as a lower bound estimate of the true local average treatment effect (LATE) of converting one more freight line to a light rail line.[14]

1.4.2 Regression Setup

I implement a difference-in-differences design to estimate the causal effect of rail transit on neighborhood incomes by comparing areas that received rail access to neighborhoods that were planned to receive it, but eventually did not. I restrict my sample

[13]Both the Rock Island and Denver & Rio Grande Railroads were large companies operating in several states and their folding was not driven by the local situation within Dallas. Similarly, since these freight companies used their tracks merely to haul goods along the line and operated freight yards in other part of the city, there is little concern that the stop of freight service affected Dallas neighborhoods differently.

[14]The results are further biased towards zero by the fact that some of the rail lines in the control group were merely postponed and opened at later times in 2009, 2010, 2012, and 2014. Therefore, anticipatory effects could have taken place in the control group and thus changed the counterfactual in the same direction as the treatment group, making it even more difficult to detect effects.

to suburban lines that opened in the early 2000s and exclude the Downtown starter system of 1996.[15] I use median family income by census tract as the main dependent variable in my study and estimate the differential change of tract income between treatment and control group before and after the extensions opened in 2001/2002. As the pre-treatment period, I use data in the year 2000 and compare that to data collected in 2010. I estimate the difference-in-differences specification as in equation (1.1):

$$y_{it} = \alpha + \beta_1 \times treat_i + \beta_2 \times post_t + \beta_3 \times post_t \times treat_i + \gamma \times controls_{it} + \varepsilon_{it} \qquad (1.1)$$

where $i \in (2000, 2010)$, β_1 and β_2 control for initial differences between the two groups and time periods respectively, and γ measures the effect of several control variables specified below. The estimate for β_3 is the parameter of interest and can be interpreted as the causal impact of transit infrastructure on tract income. Figure 1.3 summarizes the timeline of the natural experiment and the econometric setup.

1.4.3 Construction of the Treatment and Control Groups

I construct the treatment and control group in multiple ways. In the simplest way, I draw a one mile buffer around the planned rail lines to define the "catchment area" of rail. I then assign to the treatment group all census tracts whose centroids are within one mile of the planned and actually built rail lines. Similarly, I assign to the control group all census tracts within a mile of the planned lines of the 1983 Dallas Transit Service Plan that were eventually not built. Figure 1.4 depicts the census tract boundaries in dark

[15]Excluding the downtown part of the rail system has several advantages: At first, I avoid selection issues as most of the downtown starter system consisted of newly built light rail track with at-grade crossing. Secondly, like most other American cities, downtown Dallas already featured a dense bus transit system whereas the suburbs were treated with a fast transit system. Comparing only suburban neighborhoods to other suburban areas allows for a clearer interpretation of the treatment effect.

gray against the buffers of the built (red) and non-built (green) rail lines. As noted above, I exclude the downtown starter system by subtracting all census tracts within a one-mile buffer of the 1996 from both groups.

The choice of the catchment area around the rail lines is a critical one. Transit planners often use distances of half a mile to define transit catchment around rail access points, as this represents the maximum distance people are willing to walk to stations.[16] Studies on the economic impact of transit infrastructure however have shown that there exist effects at larger distances. Kahn (2007) finds effects of rail infrastructure within a one-mile and 2km (= 1.24 miles) radius. Bowes and Ihlanfeldt (2001) report significant effects at even further distances and determine that property values right next to rail stations decline, but increase in a one to three mile radius. I opt for the one mile radius because this yields a sensible compromise between increasing sample size and restricting the spatial effect of transit. Using this buffer results in adding 55 census tracts to the treatment and 108 tracts to the control group respectively.[17]

In robustness checks in the appendix, I show that my results are not sensitive to changes in the exact specification of the treatment and control neighborhoods and also report regression results using alternative definitions of the two groups.

1.4.4 Pre-Trends

To see whether there are obvious violations of the random assignment to the treatment group, I perform several checks. Firstly, I test for differences in the means

[16]For example,Guerra et al. (2011) define this distance as the industry standard and the surveyed studies in Vessali (1996) use similar distances below the one-mile mark.

[17]For my preferred specification, I do not use the distance to stations alone. The location of stations is obviously crucial as these are the only access points to enter the transit network. However, the impact of the rail tracks is also part of the treatment effect and can play a sizable role especially for non-transit users. In fact, opposition to urban rail projects is often based on arguments that rail tracks create noise and physically separate neighborhoods. Ignoring these effects by focusing on station areas only would focus on only one, presumably positive, aspect of the transit infrastructure.

between treatment and control group in the year 2000, i.e. immediately before the construction of the extensions. Table 1.1 compares the means of the two groups together with a test statistic for differences in means. Using several income measures (average and median family income), I show that, before the opening of the new lines, the census tracts in the treatment and control group did not differ significantly. On other dimensions however they are different. The treated tracts experience lower unemployment and poverty rates and show a higher labor force participation. The control group is also more ethnically diverse. While the higher share of Black population in the control group is only marginally significant, the share of Hispanics is almost 10% larger than in the treatment group.

In a second step, I investigate the parallel trend assumption of the difference-in-differences estimator. I use data from earlier censuses in 1970, 1980, and 1990 and plot the means of several neighborhood characteristics in Figure 1.5. The graphs show very similar trends between the treatment and the control group prior to the opening of the rail extension in 2001/2002. For income, the two groups track each other very closely and only diverge in the decade after the treatment. The exception is unemployment, where the treatment and control group tend to have slightly diverging trends in the pre-treatment period. Unemployment seems to increase more in the control group between 1980 and 1990 relative to the treatment group. In the two periods immediately prior to the opening of the extensions however the differences remain constant.

1.5 Empirical Results

In this section, I present the empirical results of the difference-in-differences estimation. At first, I show the estimates of the effect of rail transit on neighborhood socio-economic characteristics and also describe how the treatment effect varies by initial conditions. I then present results on business patterns and discuss potential mechanisms driving these results.

1.5.1 Neighborhood Income, Poverty, and Employment

Table 1.2 shows the results of the basic difference-in-differences regression for median family income. In the simplest specification I only include locational controls (a quadratic polynomial in distance to downtown, linear distance to the nearest freeway, linear distance to the initial rail system of 1996). In column (1) the coefficients on the interaction term $Post \times Treat$ signifies a considerable increase in median income of \$5,229 for census tracts that received rail access compared to those that did not between 2000 and 2010. This represents an increase of about one fifth of the standard deviation (\$26,698) of the median family income distribution in 2000. In column (2), I show that the results are not primarily driven by race characteristics. When controlling for the initial shares of Black and Hispanic population in the census tracts, these predictors are highly significant. The point estimate of the interaction, however, changes only modestly to \$5,365. When adding census tract fixed effects (column 3) to control for time-invariant omitted variables, the coefficient remains is robust and only increases slightly to \$5,309.9.

In conclusion, the estimated coefficients are positive and stable over different specifications. The effect of being close to transit infrastructure on median family income of around \$5,300 is economically large. To put it into comparison, the coefficient on *Post* is equal to around \$4,000 in all specifications and represents the average increase in median family income between 2000 and 2010. Treated neighborhoods therefore increased their income by more than twice as much as non-treated ones, pointing towards a strong gentrification effect.

Income Heterogeneity

I now explore neighborhood heterogeneity in my sample. Figure 1.6 shows the distribution of median incomes for census tracts before (dashed lines) and after (solid lines) the rail expansion for both the treatment and the control group. The graph shows that the estimated positive impact of improved transit infrastructure is not a mere shift of the distribution of the treatment group, but that there are more complex changes present. While there is little change in the between-distribution for the control group (right panel) with only slight movements at the very top end of the income distribution, the treatment group (left panel) shows a stark spreading of the median income distribution with many more very rich but only slightly more very poor census tracts. While this imbalance explains the increase in the mean, it also points to important non-uniform treatment effects.

The heterogeneous impact of transit infrastructure depending on neighborhood characteristics has been documented in the literature. Kahn (2007) and Baum-Snow and Kahn (2005) highlight the non-uniform impact of rail transit on neighborhoods according to their urban/suburban location. I now investigate how other characteristics such as income and race interact with the treatment. To quantify the differential effects of rail access on census tracts based on their pre-existing characteristics, I allow the treatment effect to vary by initial characteristics in 2000. In a first specification, I interact the treatment effect on neighborhood income with its lagged variable from the year 2000 by estimating a set of heterogeneous treatment effects in the following augmented difference-in-differences specification:

$$inc_{it} = \alpha + \beta_1 \times treat_i + \beta_2 \times post_t + \beta_3 \times post_t \times treat_i + \beta_4 \times post_t \times treat_i \times inc_{2000,i}$$
$$+ \beta_5 \times inc_{2000,i} + \beta_6 \times post_t \times inc_{2000,i} + \beta_7 \times treat_i \times inc_{2000,i}$$
$$+ \gamma \times controls_{it} + \varepsilon_{it} \tag{1.2}$$

where inc_{it} is the outcome variable median family income per census tract. The term $post_t \times inc_{2000,i}$ controls for potentially different trends of rich versus poor census tracts that affect all neighborhoods regardless of treatment status. Controlling for $treat_t \times inc_{2000,i}$ captures initial differences in the income composition of the treatment and control group that might persist over time. The coefficient β_5 on the lag of income controls for persistence in the neighborhood median incomes. As it turns out, this persistence is prevalent in all specifications.

The results for this setup are reported in Table 1.2, column (4). The treatment effect is strongly positively related to initial median family income in 2000, indicating that richer neighborhoods see larger increases in income from rail expansion. A census tract with a $1,000 higher income in 2000 will experience an increase in the treatment effect of $334. For better interpretation of the coefficients, I calculate the marginal effects at the 25th and the 75th percentile of the initial income distribution. In the baseline specification, the treatment effect is negative (estimated at -$327.7) for the poorest quarter of the tracts while for the richest ones it is positive and large ($8,707).[18] The treatment effect of around $5,300 in the simple difference-in-differences regression is therefore an average of a much larger range of the treatment, including negative and highly positive values. This indicates that the rail transit expansion causes richer tracts to become richer

[18] The estimated threshold for a zero effect is at a median family income of $33,859, which represents the 30th percentile.

while it causes the poorest to become poorer, thus making the spatial distribution of incomes between tracts more extreme within the treatment area and leading to increased income segregation.

The range of the effect also demonstrates that there can be both a gentrification and a “poverty magnet” effect acting on neighborhoods at the same time.[19] This effect is consistent with the patterns in Figure 1.6 where the between-tract income distribution for the treatment group is spreading out leaving more mass at both tails. The same heterogeneous effect is confirmed when using the smaller (yet less consistent) census block groups (Table 1.3). Similarly, I use a quantile regression in Table 1.4 to estimate the treatment effect at different percentiles of the income distribution. The coefficient is the stronger the higher the percentile, and even though they are only statistically significant at higher quantiles, this shows the high dependence of the treatment effect on initial income.

Other Neighborhood Outcomes

In a next step, I look at the census tract poverty rate as another measure of neighborhood income. Table 1.5, column (1) summarizes the results for the simple difference-in-differences setup using the share of households whose incomes were below the household-size adjusted poverty threshold as the outcome variable. The estimated treatment effect is positive at around 0.8 percentage points. Again, the coefficient does not change much when including racial controls or fixed effects (columns 2 and 3), but is not statistically significant in either specification. The treatment effect however varies strongly and positively with initial poverty (column 4), suggesting that the share of poor households increases in neighborhoods that were initially poor while it decreases in

[19]In regressions not reported, I find that the treatment effect on median family income varies positively with the initial share of college graduates living in the neighborhood and negatively with the presence of non-white residents. The effects are however estimated with large standard errors.

richer neighborhoods. This points towards the same segregation as found with median family income before. In addition, the treatment effect is positively associated with the initial unemployment rate (column 5). These results highlight the heterogeneity of the treatment effect of rail infrastructure and are consistent with the earlier findings that less prosperous neighborhoods decline in response to gaining transit access.

The changes in income measures are not accompanied by similar effects on neighborhood population. Implementing the same difference-in-differences regression with the absolute tract population as the outcome variable reveals that there was no excess population growth in the treated areas. The results in Table 1.6 show that the transit tracts actually slightly lost population when compared to the untreated control areas, although the negative coefficient is not statistically significant. If rail brings along higher incomes, one would expect people to move towards these areas. The absence of such an effect points towards a migration channel that I discuss in section 1.5.4. Looking at housing values, I use self-reported housing values from the census. Regressing median housing values of owner-occupied units in Table 1.7, I find an imprecisely measured increase in housing values in the transit neighborhoods. Again the treatment effect is highly heterogeneous and positively correlated with initial housing values in 2000.

Finally, I examine unemployment. Unemployment might be expected to go up in transit areas because unemployed transit-dependent households move towards rail stations in order to improve job accessibility. This is consistent with the fact that 10.4% of DART riders are job seekers, a share that is above the metropolitan area's unemployment rate. Indeed, in Table 1.8 I find a slight uptick in unemployment rates around transit lines of about 1.2 percentage points. The effect is however imprecisely measured and not statistically significant. When interacting the treatment effect with initial unemployment rates in 2000, I find the surprising result that it is highly negatively correlated. Unemployment rates thus decrease in areas with initially high unemployment which could just indicate regression to the mean.

Putting the different results together, the empirical evidence points towards a positive, but heterogeneous treatment effect of rail infrastructure on neighborhood economic characteristics. Transit access leads to a more extreme spatial distribution of neighborhood income measures in areas that received rail infrastructure as compared to the untreated control neighborhoods. Median family income stands out within this polarization as incomes increase the most in already wealthy tracts and decrease in the bottom 30% of the income distribution. The census tract poverty rate follows this result. The estimated treatment effect shows that poverty goes up in neighborhoods that were initially poor and had high unemployment rates.

1.5.2 Business Patterns

One factor that could explain the increase in neighborhood income in the treatment group is through an effect on businesses in transit areas. An increase in economic activity around rail stations could result into higher incomes in transit neighborhoods. In contrast, improving transit could bring a change in the business structure of neighborhoods through transit-oriented development.[20] Transit could induce a business structure of specialty shops, restaurants, and related stores that are easily accessible by walking or convenient transit trips. These kinds of amenities are naturally valued more by households with higher income and thus may attract these households to relocate towards transit access stations (Dutzik et al., 2014).

[20] Transit think tank Reconnecting America defines transit-oriented development as a *"mixture of housing, office, retail and/or other amenities integrated into a walkable neighborhood and located within a half-mile of quality public transportation."*

Regression Setup

To analyze whether overall business activity increases and whether the business structure in transit tracts has changed after the introduction of the system, I use geocoded administrative data on the issuance of business licenses in the Dallas-Fort Worth MSA from 1995 to 2014. As before, I implement a difference-in-differences regression comparing business openings in areas that received rail transit against areas that were supposed to gain transit access, but eventually did not.[21] I add up all new businesses openings at the census tract level and merge them with the socioeconomic data from the census. This allows me to calculate the yearly number of new business openings at the census tract level for both the treatment and control areas. Table 1.9 compares the distribution of the yearly openings on the census tract level by treatment and control group. The most apparent feature of the data is that business openings are rather rare. More than 25% of all tract-year combinations see at most one new establishment and zero values are quite prominent. The treatment group also has significantly less openings per year and a much smaller standard deviation.

The crucial question is whether this flow variable is a good proxy for economic activity, which would be best measured by the count and size of businesses in terms of the number of employees and sales revenue. Unfortunately, the data does not provide information about the number of business closings, and neither does it indicate revenue or employment for the individual businesses. However, if rail transit infrastructure indeed induces economic rejuvenation and changes in the industry composition, this should

[21]In contrast to the earlier analysis, I now use one-mile buffers around the rail stations rather than the lines laid out in the 1983 transit plan to create treatment and control areas. The rationale behind this is that businesses choose their location primarily to maximize access to customers and suppliers and are less sensitive to the potential negative influence of rail tracks such as noise.

be detectable through an increased frequency of new businesses opening around rail infrastructure or a relative change in openings of different industry classifications.[22]

Effect on Business Openings

Figure 1.7 depicts the log openings aggregated at the treatment and control group with a vertical line denoting the opening of the new rail lines in 2001/2002. Visual inspection of the time series data on total openings does not depict a large effect of the rail opening on log business openings. Before the onset of the treatment, the two series follow a very similar trend. Only after the opening of the new lines is there a slight narrowing of the gap between the control and the treatment group. This visual result is confirmed by a difference-in-differences regression at the census tract level reported in Table 1.10, column (1) where the treatment effect on $Post \times Treat$ is estimated to .215. The coefficient is significant at the 5% level. To account for serial correlation of local unobservables, I cluster the standard errors at the municipality level and include municipality and year fixed effects. Rail tracts therefore tend to see a slight uptick in

[22] To investigate the correlation between flow and stock variables at higher level geographies, I make use of the County Business Patterns (CBP) data that reports the number of establishments, employment and payroll data at the ZIP code level from 1998-2013. I first aggregate the business openings to the 86 ZIP codes in Dallas for which data is available and then correlate these numbers with statistics from the CBP. The average correlation between business openings and the number of establishments at the ZIP code level is 0.73 for the years 1998 to 2013 and never drops below 0.63 for a single year. The correlation with the number of jobs (0.49) and payroll (0.31) is lower, but always strongly positive. This shows that at this higher aggregation, the average yearly inflow of new businesses correlates well with indicators of economic activity. The correlation of the number of new openings with the change in establishments however is less pronounced. While the coefficient estimate for β of the regression

$$\Delta establishments_{i,t} = \alpha + \beta \, openings_{i,t} + \varepsilon_{i,t}$$

is positive and highly significant ($\widehat{\beta} = .111$, standard error $= .043$), the overall correlation between the two variables is rather small at only 0.052. This small correlation is the result of very little change in the number of establishments in the ZIP-code level county business patterns data. It appears that while the the number of business openings fluctuates, the number of establishments at larger geographies is relatively constant, possibly due to local zoning restrictions that limit the expansion of commercially zoned areas. The number of new businesses should therefore be interpreted as a proxy for industry composition change rather than as an approximation for overall economic activity.

new businesses openings of about 21%. This effect is however small and translate to one additional business opening at the census tract per year.

Effect on Industry Composition

The construction of the rail system might not only change the total amount of business investment, but also alter a neighborhood's industry composition. For example, a new light rail station might cause previously existing businesses to close and be replaced by new shops or restaurants. To investigate a change in the composition, I break down the business data by its industry classification. In a first step, I look only at business openings that can be classified as urban consumer amenities, especially retail and hospitality services, and are potentially higher valued by high-income households.[23] I also look at restaurants and bars in general and fast-food restaurants in specific. In a further exercise, I examine the effect on chain stores.

The results in columns (2)-(5) do not show a clear pattern of how the business structure of neighborhoods changes in response to receiving rail transit access. The coefficient for amenities is small and imprecisely estimated (column 2). This does not suggest that transit infrastructure creates a significant increase in shopping, entertainment, and dining establishments around rail stations. The treatment effect for hospitality businesses (column 3) is, although positive, similarly imprecise and does not allow for a definite conclusion about its response to transit access. The treatment effect for fast-food restaurants which are often seen as predictors of community decline is even negative, although the standard error of the estimate is large (column 4). When considering chain stores in column (5), I see an uptick of about 19.3%. This indicates that part of the increase in total business openings is absorbed by new chain stores. Estimating augmented difference-in-differences models as in equation (1.2) does not

[23]In specific, these urban amenities include the NAICS codes 44, 45 ("Retail Trade"), 71 ("Arts, Entertainment, and Recreation"), and 72 ("Accommodation and Food Services").

yield any dependence of the treatment effect on initial characteristics such as income or geographical location within the city.

In conclusion, gaining access to transit does on average increase the number of new business openings at the census tract level. The absolute impact however is small and seems to be mostly realized as a surge in chain stores. Classic transit-oriented establishments that are usually linked to gentrification like restaurants and retail stores did not exhibit an increase and there is no evidence for a massive change in the industry composition of the treated neighborhoods. This casts doubt on transit as an incubator for economic redevelopment of declining neighborhoods.

1.5.3 Placebo Tests

To demonstrate that the estimated effects are indeed caused by the treatment of transit access and do not pick up merely a slow capitalization effect, I run a placebo regression for the years 1990 (pre-treatment) and 2000 (post-treatment). If the rail access is indeed causing the effects, then I should not find any significant impacts in the placebo regression. Table 1.11 shows the estimation results of the same regressions as in the initial analysis for the earlier time period. The results do not show any effect of the placebo treatment as neither of the coefficients is large or significant. The treatment effect estimate of -$182 is small and negative, with a large standard error of more than $4,600. This confirms the parallel trend assumption of the difference-in-differences approach. The triple interaction term is also small and statistically insignificant with an estimate of about -0.05 and a large standard error of 0.11. There is therefore no evidence for increased sorting due to the announcement of the rail infrastructure and the gap between the 25th and 75th percentile estimates [239,-626] is rather small.

1.5.4 Mechanisms

The estimated treatment effect above must be thought of as a combination of two effects: Firstly, rail infrastructure affects the incumbent residents' income through directly. Secondly, neighborhood characteristics change as a result of changes in the composition of the neighborhood caused by people moving into and out of the affected areas. Given the lack of individual-level data, these two effects are difficult to disentangle. In this section I try to evaluate these two effects.

Direct Effects of Transit on Incomes

In a first step, I look at channels through which light rail can directly change the income of residents in transit neighborhoods. Better rail access could increase neighborhood incomes through an increase in economic activity and provide jobs to residents in transit neighborhoods that previously had no or worse jobs. As the analysis in the previous section has shown, the light rail indeed induced an increase in business activity around transit stations. However, the rather small estimated treatment effect is unlikely to cause the large changes of more than $5,000 in median family income, especially since the increase in business activity is largely accounted for by chain stores that include small businesses such as convenience stores.

Similarly, light rail could change incomes directly by improving job access of transit-dependent households and help them to find (and keep) jobs outside their immediate neighborhood. The physical disconnect of low-income households from job centers is known as the spatial mismatch hypothesis (Kain, 1992) and often used as an argument in favor of rail transit. However, the observed heterogeneous treatment effect of rail depending on initial neighborhood income is difficult to rationalize with a pure access story. If the light rail overcomes spatial mismatch, then the treatment effect should be larger for initially poor neighborhoods where households are more likely to use transit.

However, I observe the opposite effect with the largest impacts in rich neighborhoods where residents are likely to drive regardless of transit access. In addition, the negative effect on the poorest neighborhoods is inconsistent with a transit as being a pure choice.

In the next sub-section, I instead propose a sorting mechanisms that could rationalize the more extreme income distribution between census tracts in transit neighborhoods by a simple resorting of the poor in response to better transit access. This mechanism is consistent with the data and can explain heterogeneous effects on neighborhood incomes even when only few people take transit.

Migration Effects

One way to disentangle the treatment effect is to look at household characteristics that do not change with transit access, but are correlated with income. Race is a natural candidate and if the racial composition of a neighborhood changes in response to light rail access, it must be due to a migration effect. In the Dallas-Fort Worth MSA, race is persistently correlated with income as Black and Hispanic residents report less per-capita income than white residents across all years (see Table 1.12). To quantify changes in the racial makeup, I use the same difference-in-differences design with the shares of population for different ethnic groups as the outcome variable (Table 1.13). While there is no statistically significant effect for whites and Hispanics, the estimated treatment effect for Blacks is statistically significant at the 10% level. I estimate a 2.17 percentage point increase in the share of Black population. This represents more than a 10% increase compared to the mean of the study area of 19.39% in the year 2000. More interestingly, as indicated in Table 1.14, the treatment effect is dependent on initial neighborhood characteristics as the interaction term with the initial share of Black residents in the tract (column 4) is positive and significant. Again, the treatment is negative for the tracts that had an initially low share of Black residents and rises by 1.5 percentage points for every 10% increase in the share of Black residents. This indicates that migration plays a role.

Like the distribution of income, the spatial distribution of black residents becomes more extreme in transit areas. This result points towards the possibility of transit causing more sorting along income and race lines. The simplest mechanism to reconcile this increase in segregation is through a resorting of poor households from (initially) richer to poorer neighborhoods within the newly connected transit areas only. If the poor of the initially rich neighborhoods move to the initially poor neighborhoods, median family income in the richer areas increases (due to the poorest leaving) and income in the poorer areas decreases further as new poor arrive.[24]

Consistent with that, the interaction term with the initial median family income by tract (column 5) is negative and borderline significant (p-value = 0.137). This indicates that the share of Black population increased more in neighborhoods that were initially poorer. If the Black residents that moved were comparatively poor, this could provide evidence that low-income households move towards low-income neighborhoods. This simple resorting of the poor can then explain the spreading of the distribution in the treatment group in Figure 1.6.[25]

Discussion: A Sorting Mechanism of the Poor

A potential reason for the poor moving from richer to poorer households is through a change in transportation costs due to the new light rail option in transit neighborhoods.

[24]For this to be true, the poor in the richest neighborhoods that move must be poorer than the median in the poorest neighborhoods to avoid Simpsons's Paradox.

[25]Such a sorting mechanism is consistent with the the high residential turnover in Dallas between the decennial censuses. Data from the ACS in Table 1.15 shows that about 10% of all households living in the study area (treatment and control group combined) in 2010 had moved within the previous year. 62.7% were not living in the same residence as they did five years ago and more than 70% had moved within the previous decade. For renter-occupied housing units, this number even approaches 90%. These statistics do not differ significantly from the non-study area in the Dallas-Fort Worth metroplex and confirm that households do change their residence frequently, indicating that frictions in the housing market and moving costs are not preemptively large to prevent residential readjustment.

The new rail system increased travel speeds primarily for the transit-dependent poor. While it shortened travel times for a set of origin-destination pairs compared to the previous bus network, driving is still much faster for most point-to-point trips within the Dallas-Fort Worth metropolitan area.[26] Given the high job decentralization in Dallas, it is therefore not surprising that the actual take-up of the rail system is small and that rail trips make up at most 2% all work commutes in the MSA and 4.5% within the rail areas (American Community Survey, 2010). It is therefore plausible that people who can afford a car continued to drive to work and that most DART riders were former transit-dependent bus users, a fact that has been observed for many US cities (Baum-Snow and Kahn, 2005).

One mechanism that could explain the more extreme income distribution between transit neighborhoods is that with increased transit opportunities, poorer transit-dependent households are now able to move out of high-income, high-rent neighborhoods to poorer areas with lower housing costs while still having access to their jobs. In a city with poor public transportation and high job decentralization like Dallas, transit-dependent workers have to live close to their jobs. At the same time, proximity to jobs is highly valued by high-income households due to their higher opportunity cost of time and this drives housing prices up. Better public transportation therefore allows the poor to segregate themselves away from housing competition with high-income households in high-rent areas. Using their comparative advantage of lower opportunity cost of time, they can trade off cheaper housing costs for longer commutes.

The link between transportation infrastructure and spatial segregation by income has been recognized in the literature (LeRoy and Sonstelie, 1983). Baum-Snow (2007) and Nall (2015) have shown that lower within-city transportation costs, through the

[26]The North Central Texas Council of Governments maintains travel time matrices for 5,386 traffic analysis zones (TAZ) within the Dallas-Forth Worth metropolitan area. Among the 2,432,344 pairs for which rail travel is possible, light rail was faster than peak-hour car travel for only 52. For these 52 origin-destination pairs, the average time saving was 1.5 minutes.

introduction of the interstate highway system, can lead to spatial polarization in form of suburbanization. Glaeser and Kahn (2004) and Glaeser et al. (2008) have suggested similar mechanisms for transit and predict that providing different transport modes allows for residential segregation where the poor use the slow, inexpensive mode and the rich use the fast but expensive mode. In this way, the low-income households can sort themselves away from competition for housing with the rich by using the transportation mode that fits their lower marginal cost of time.

While other mechanisms cannot be ruled out, the resorting of the poor is consistent with the data and the previous literature. Such a mechanism would confirm the strong relationship between transportation costs and spatial sorting. The results indicate that even if changes in transportation costs only affect a small share of the population, namely transit riders, they can have a potentially large effect on the spatial dispersion of incomes within a city. This sorting effect adds another dimension to the overall impact of transit infrastructure on welfare. While the addition of a new transportation mode in form of a rail system helps to facilitate mutually beneficial income sorting in the spirit of Tiebout (1956), the resulting clustering of the poor could have potential negative externalities in form of negative peer effects on education (Kling et al., 2007), crime (Damm and Dustmann, 2014), and labor market outcomes (Chyn, 2016). Interestingly, the empirical evidence runs counter to the view of transit as a tool to reduce urban income inequality (Blumenberg and Ong, 2001; Sanchez, 2002).

1.6 Conclusion

In this paper, I exploit exogenous variation in rail transit access to analyze its effect on neighborhood composition using a natural experiment setting in the city of Dallas, Texas. The results show that urban rail access on average drives census tracts' median family incomes up compared to the control group. The estimated impact of around $5,300 represents one fifth of a standard deviation of the income distribution among census tracts and is larger than the average increase in the study period. This is a surprisingly strong effect in economic terms given the overall low ridership of the transit system in the Dallas-Fort Worth metropolitan area.

The treatment effect, however, is strongly positively correlated with initial neighborhood income, implying that initially richer neighborhoods benefit disproportionately more from transit access than poor ones. In fact, the estimated treatment effect is negative for the poorest census tracts. The results document that improved transit access can lead to increased spatial dispersion of family incomes and other economic characteristics within a city, identifying transit as a mechanism for increased neighborhood segregation. The heterogeneity of effects by initial neighborhood income reconciles two opposing findings in the literature, namely that transit infrastructure extension appears to cause gentrification in some studies while it acts as a poverty magnet in others. The positive correlation of the treatment effect with initial neighborhood income implies that these mechanisms are at work at the same time and that the average impact depends on the initial configuration. The results also confirm the important heterogeneity of the treatment effect of transit access. Consistent with the findings in previous studies, they highlight that treatment effects are not uniform, but that interaction terms matter for evaluating the impact of improved transit infrastructure. In the case of Dallas, these effects vary strongly with income, race, and other economic characteristics.

Analysis of business openings shows that the transit-induced neighborhood segregation can only be partially explained by the light rail's effect on the local economy. While, on average, rail presence does increase the inflow of new businesses at the neighborhood level relative to the control group, this impact is small in absolute terms and largely captured by an increase in chain stores. Sectors that are associated with gentrification such as the entertainment and hospitality sector do not see an increase in the number of establishments. Given the high frequency of housing turnover, the observed income segregation is best interpreted as an equilibrium outcome in response to a shock in transportation costs. One potential mechanism for polarization is that, with increased transit opportunities, poorer households are now able to relocate from job-rich, high-income neighborhoods to areas with lower housing costs while still having access to their jobs. Increased income sorting in response to transit extension might simply be a result of a mutually beneficial housing rearrangement where rail access allows poorer households to make use of their lower opportunity cost of time. While spatial sorting by income might then be alleviate housing pressure, potential negative spillovers created by the clustering of poor households in space, such as negative impacts on public health, education, and crime, need to be taken into account

Table 1.1: Comparison of Groups

	Treatment Group	Control Group	p-value
Median Household Income	49494.4	48666.2	.852
Average Family Income	61898.5	60635.5	.841
Unemployment	5.276	7.33	.023
Poverty Rate	13.11	17.10	.038
Labor Force Participation	69.64	65.99	.041
Share of Black Population	16.08	21.08	.188
Share of Hispanic Population	25.18	34.70	.018
Observations	55	108	

p-value of a t-test of equality of means between groups

Table 1.2: Dependent Variable: Median Family Income

	(1)	(2)	(3)	(4)
Post	4170.9***	4035.3***	4090.9***	-751.1
	(1522.8)	(1508.6)	(1495.4)	(3620.6)
Treat	-672.7	-11526.4***		1533.8
	(3779.6)	(2952.2)		(1164.1)
Post $\times$ Treat	5229.9*	5365.5*	5309.9*	-11324.0
	(3149.3)	(3157.9)	(3125.1)	(8133.6)
Share of Black Population in 2000		-1015.5***		
		(140.4)		
Share of Hispanic Population in 2000		-982.2***		
		(117.3)		
Median Family Income in 2000				1.030***
				(0.0143)
Post $\times$ Median Family Income in 2000				0.0995
				(0.0901)
Treat $\times$ Median Family Income in 2000				-0.0229
				(0.0212)
Post $\times$ Treat $\times$ Median Family Income in 2000				0.334*
				(0.177)
Census Tract Fixed Effects	No	No	Yes	No
Marginal Effect at 25th percentile				-594.2
Marginal Effect at 75th percentile				8707.5
Observations	325	325	325	325
R squared	0.0900	0.603	0.122	0.868

Standard errors clustered at the census tract level: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Treatment group: census tracts within 1 mile of built rail lines

Control group: census tracts within 1 mile of planned, but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

Table 1.3: Dependent Variable: Median Family Income at Block Group Level

	(1)	(2)	(3)	(4)
Post	4847.1***	4847.1***	4847.1***	11838.5***
	(1164.3)	(1305.6)	(1299.9)	(3789.0)
Treatment	2835.3	-8413.3*		
	(4370.2)	(4441.0)		
Post × Treat	7175.7**	7175.7**	7175.7**	-13767.5**
	(2831.1)	(3590.5)	(3574.8)	(6762.5)
Share of Black		-761.2***		
		(55.91)		
Share of Hispanic		-724.4***		
		(56.31)		
Post × Income in 2000				-0.168
				(0.102)
Post × Treat × Income in 2000				0.436***
				(0.132)
Constant	31311.5***	104488.1***	44584.9***	44584.9***
	(4830.5)	(6858.6)	(659.6)	(619.5)
Block Group Fixed Effects	No	No	Yes	Yes
Marginal Effect at 25th percentile				-1242.8
Marginal Effect at 75th percentile				8870.2
Observations	804	804	804	804
R squared	0.104	0.516	0.105	0.153

Standard errors clustered at the census tract level: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Treatment group: census block groups within 1 mile of built rail lines

Control group: census block groups within 1 mile of planned, but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2008-2013

Regressions include location controls

Table 1.4: Quantile Regression. Dependent Variable: Median Family Income

	(0.1)	(0.25)	(0.5)	(0.75)	(0.9)
Post	-1127.2	924.0	3654.2**	6338.5***	6763.5*
	(1424.7)	(1476.2)	(1787.5)	(2422.5)	(3745.5)
Treatment	-2527.4	-3210.6	-5567.2**	-9660.7**	-13656.6**
	(2218.1)	(2298.2)	(2782.8)	(3771.4)	(5831.1)
Post × Treat	-3424.8	-2826.4	2918.2	10354.4**	15118.0**
	(2699.2)	(2796.8)	(3386.4)	(4589.5)	(7096.0)
Share of Black	-344.3***	-437.4***	-590.9***	-821.8***	-1031.0***
	(29.07)	(30.12)	(36.47)	(49.43)	(76.42)
Share of Hispanic	-280.6***	-384.2***	-529.9***	-790.7***	-1079.0***
	(25.16)	(26.07)	(31.56)	(42.78)	(66.14)
Observations	804	804	804	804	804
Pseudo R-squared	0.196	0.231	0.296	0.392	0.476

Standard errors clustered at the census tract level: $^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Treatment group: census block groups within 1 mile of built rail lines

Control group: census block groups within 1 mile of planned, but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2008-2013

Regressions include location controls

Table 1.5: Dependent Variable: Poverty Rate

	(1)	(2)	(3)	(4)	(5)
Post	5.251***	5.263***	5.179***	6.061***	6.355***
	(0.763)	(0.765)	(0.755)	(1.383)	(1.402)
Treat	-2.353	0.838		-0.343	1.221
	(1.506)	(1.041)		(0.436)	(1.495)
Post × Treat	0.821	0.808	0.892	-4.607***	-3.322*
	(1.244)	(1.248)	(1.234)	(1.767)	(1.783)
Share of Black Population in 2000		0.373***			
		(0.0433)			
Share of Hispanic Population in 2000		0.275***			
		(0.0258)			
Poverty Rate in 2000				1.018***	
				(0.0296)	
Post × Poverty Rate in 2000				-0.0517	
				(0.0806)	
Treat × Poverty Rate in 2000				0.0337	
				(0.0343)	
Post × Treat × Poverty Rate in 2000				0.404***	
				(0.0959)	
Unemployment in 2000					1.722***
					(0.173)
Post × Unemployment in 2000					-0.160
					(0.174)
Treat × Unemployment in 2000					-0.201
					(0.182)
Post × Treat × Unemployment in 2000					0.737***
					(0.199)
Census Tract Fixed Effects	No	No	Yes	No	No
Marginal Effect at 25th percentile				-1.744	-0.926
Marginal Effect at 75th percentile				3.536	2.730
Observations	325	325	325	325	325
R squared	0.345	0.641	0.345	0.851	0.630

Standard errors clustered at the census tract level: $^{*} p < 0.10$, $^{**} p < 0.51$, $^{***} p < 0.01$
Treatment group: census tracts within 1 mile of built rail lines
Control group: census tracts within 1 mile of planned, but actually not built rail lines
Pre-Treatment Period: 2000. Post-Treatment Period: 2010
Regressions include location controls

Table 1.6: Dependent Variable: Tract Population

	(1)	(2)	(3)	(4)
Post	195.8	203.7*	184.1	-129.5
	(120.8)	(121.7)	(119.9)	(226.6)
Treat	-130.1	66.82		-1331.4**
	(260.6)	(244.8)		(659.8)
Post × Treat	-253.8	-261.7	-242.1	-668.6
	(229.2)	(230.2)	(227.5)	(585.4)
Share of Black Population in 2000		3.647		
		(6.017)		
Share of Hispanic Population in 2000		20.53***		
		(5.980)		
Median Family Income in 2000				-0.0188***
				(0.00473)
Post × Median Family Income in 2000				0.00669*
				(0.00403)
Treat × Median Family Income in 2000				0.0239**
				(0.0109)
Post × Treat × Median Family Income in 2000				0.00826
				(0.0108)
Census Tract Fixed Effects	No	No	Yes	No
Marginal Effect at 25th percentile				-403.6
Marginal Effect at 75th percentile				-173.9
Observations	325	325	325	325
R squared	0.0234	0.0952	0.0137	0.0785

Standard errors clustered at the census tract level: $^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$
Treatment group: census tracts within 1 mile of built rail lines
Control group: census tracts within 1 mile of planned, but actually not built rail lines
Pre-Treatment Period: 2000. Post-Treatment Period: 2010
Regressions include location controls

Table 1.7: Dependent Variable: Median Housing Values

	(1)	(2)	(3)	(4)
Post	35775.1***	35057.5***	34507.2***	33482.3***
	(4462.5)	(4605.7)	(4363.8)	(6406.3)
Treat	-6324.4	-39642.5***		-5522.5
	(11604.5)	(14926.7)		(3719.4)
Post × Treat	4567.8	3562.6	4607.6	-44251.4***
	(7977.7)	(7736.3)	(7535.0)	(11657.2)
Share of Black Population in 2000		-3103.4***		
		(761.5)		
Share of Hispanic Population in 2000		-2984.5***		
		(658.2)		
Median Housing Value in 2000				1.002***
				(0.00319)
Post × Median Housing Value in 2000				0.00909
				(0.0606)
Treat × Median Housing Value in 2000				0.0482*
				(0.0287)
Post × Treat × Med. Housing Value 2000				0.473***
				(0.116)
Census Tract Fixed Effects	No	No	Yes	No
Marginal Effect at 25th percentile				-16841.1
Marginal Effect at 75th percentile				11704.5
Observations	314	314	314	314
R squared	0.0625	0.475	0.405	0.923

Standard errors clustered at the census tract level: $^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$

Treatment group: census tracts within 1 mile of built rail lines

Control group: census tracts within 1 mile of planned, but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

Table 1.8: Dependent Variable: Unemployment

	(1)	(2)	(3)	(4)
Post	1.933***	1.913***	1.892***	3.823***
	(0.474)	(0.473)	(0.469)	(0.674)
Treat	-1.711**	-0.866		0.229
	(0.691)	(0.558)		(0.219)
Post × Treat	1.212	1.232	1.253	3.061***
	(0.847)	(0.849)	(0.841)	(0.941)
Share of Black Population in 2000		0.163***		
		(0.0185)		
Share of Hispanic Population in 2000		0.0611***		
		(0.00870)		
Unemployment in 2000				1.036***
				(0.0313)
Post × Unemployment in 2000				-0.263**
				(0.104)
Treat × Unemployment in 2000				-0.0707**
				(0.0331)
Post × Treat × Unemployment in 2000				-0.446***
				(0.112)
Census Tract Fixed Effects	No	No	Yes	No
Marginal Effect at 25th percentile				1.609
Marginal Effect at 75th percentile				-0.605
Observations	325	325	325	325
R squared	0.297	0.552	0.190	0.711

Standard errors clustered at the census tract level: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$
Treatment group: census tracts within 1 mile of built rail lines
Control group: census tracts within 1 mile of planned, but actually not built rail lines
Pre-Treatment Period: 2000. Post-Treatment Period: 2010
Regressions include location controls

Table 1.9: Annual Business Openings by Neighborhood. Summary Statistics by Group

Group	**mean**	**sd**	**min**	**p25**	**p75**	**max**	**N**
Treatment	5.15	6.96	0	1	7	63	600
Control	6.70	15.3	0	1	6	258	1860

Source: Dallas Tax Register
Years from 1995-2014

Table 1.10: Dependent Variable: Log Business Openings

	(1) All	(2) Amenities	(3) Hospitality	(4) Fast Food	(5) Chain Stores
Distance to Freeway	-0.0310	-0.0681	-0.114*	-0.0409	-0.0751
	(0.0748)	(0.0662)	(0.0642)	(0.0488)	(0.0780)
Distance to Downtown	0.137*	0.0933	-0.00712	-0.0302	0.118
	(0.0783)	(0.0612)	(0.0508)	(0.0365)	(0.0772)
Treat	-0.562***	-0.405***	-0.319***	0.00315	-0.574***
	(0.149)	(0.105)	(0.108)	(0.0571)	(0.144)
Post	1.692***	1.562***	0.878***	0.464***	1.117***
	(0.0876)	(0.0928)	(0.157)	(0.0896)	(0.105)
Post $\times$ Treat	0.215**	0.0413	0.112	-0.124	0.193*
	(0.0861)	(0.0855)	(0.105)	(0.0791)	(0.0979)
Observations	2048	1674	801	401	1700
R^2	0.348	0.338	0.227	0.111	0.246

Standard errors in parentheses (clustered at the municipality level)
Treatment group: census tracts within 1 mile of built rail stations
Control group: census tracts within 1 mile of planned, but actually not built rail stations
Amenities includes NAICS 44-45 (Retail Trade), 71 (Arts, Entertainment, and Recreation), and 72 (Accommodation and Food Services)
Hospitality includes NAICS 722 (Food Services and Drinking Places)
Fast Food includes NAICS 722513 (Limited Service Restaurants)
Chain Stores include business with more than one outlet
Regression includes municipality and year fixed effects
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.11: Dependent Variable: Median Family Income Placebo

	(1)	(2)	(3)	(4)
Post	12193.7***	12193.7***	12193.7***	4210.6***
	(1328.5)	(1332.7)	(1318.2)	(1568.1)
Treat	-534.1	-11486.9***		-2167.2
	(2702.6)	(3076.4)		(1409.0)
Post × Treat	-182.0	-182.0	-182.0	1405.7
	(1890.5)	(1896.5)	(1875.8)	(4450.5)
Share Black in 1990		-715.0***		
		(131.4)		
Share Hispanic in 1990		-941.2***		
		(155.0)		
Median Family Income in 1990				1.001***
				(0.0121)
Post × Median Family Income 1990				0.219***
				(0.0591)
Treat × Median Family Income 1990				0.0545*
				(0.0330)
Post × Treat × Median Family Income 1990				-0.0483
				(0.118)
Census Tract Fixed Effects	No	No	Yes	No
Marginal Effect at 25th percentile				239.3
Marginal Effect at 75th percentile				-626.6
Observations	326	326	326	326
R squared	0.199	0.563	0.486	0.886

Standard errors clustered at the census tract level: $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$
Treatment group: census tracts within 1 mile of built rail lines
Control group: census tracts within 1 mile of planned, but actually not built rail lines
Placebo pre-treatment period: 1990. Placebo post-treatment period: 2000
Regression includes location controls

Table 1.12: Per Capita Income by Race

	Study Area			Dallas-Fort Worth MSA		
Year	**Black**	**Hispanic**	**White**	**Black**	**Hispanic**	**White**
1990	10,743	10,743	18,057	11,894	10,861	17,966
2000	13,196	13,196	23,806	18,372	15,829	26,150
2010	16,818	16,818	28,584	23,712	19,713	31,905

Source: Census, American Community Survey
Study area: Treatment and control group

Table 1.13: Dependent Variable: Racial composition

	(1) Black	(2) Hispanic	(3) White
Post	-0.147	10.38***	0.102***
	(0.769)	(1.020)	(0.0113)
Post × Treat	2.170*	-1.288	0.00883
	(1.239)	(1.963)	(0.0214)
Census Tract Fixed Effects	Yes	Yes	Yes
Observations	325	325	325
R squared	0.0234	0.444	0.426

Standard errors clustered at the census tract level

$^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

Treatment group: census tracts within 1 mile of built rail lines

Control group: census tracts within 1 mile of planned,
but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

Table 1.14: Dependent Variable: Share of Black Population

	(1)	(2)	(3)	(4)	(5)
Post	0.122	-0.147	-0.162	2.875***	-4.082***
	(0.811)	(0.769)	(0.776)	(0.775)	(1.537)
Treat	-1.861		-0.524	0.249	-0.808
	(2.802)		(0.340)	(0.301)	(0.835)
Post × Treat	1.901	2.170*	2.185*	-1.037	6.876**
	(1.271)	(1.239)	(1.252)	(1.272)	(3.481)
Share of Hispanic Population in 2000			-0.0290**		-0.0617***
			(0.0112)		(0.0190)
Share of Black Population in 2000			0.942***	1.018***	0.908***
			(0.0172)	(0.00585)	(0.0249)
Post × Share of Black Population in 2000				-0.143***	
				(0.0343)	
Treat × Share of Black Population in 2000				-0.0116	
				(0.00936)	
Post × Treat × Share Black Population 2000				0.154*	
				(0.0816)	
Median Family Income in 2000					-0.0671***
					(0.0202)
Post × Median Family Income in 2000					0.0804***
					(0.0263)
Treat × Median Family Income in 2000					-0.00121
					(0.0165)
Post × Treat × Median Family Income 2000					-0.0959
					(0.0642)
Census Tract Fixed Effects	No	Yes	No	No	No
Marginal Effect at 25th percentile				-0.257	3.798
Marginal Effect at 75th percentile				2.705	1.130
Observations	325	325	325	325	325
R squared	0.215	0.0234	0.946	0.951	0.949

Standard errors clustered at the census tract level: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$
Treatment group: census tracts within 1 mile of built rail lines
Control group: census tracts within 1 mile of planned, but actually not built rail lines
Pre-Treatment Period: 2000. Post-Treatment Period: 2010
Regressions include location controls

Table 1.15: Migration Statistics

	Study Area	Non-Study Area
Moved within the last year	9.76%	10.53%
Not living in the same house as five years ago	62.7%	64.4%
Moved in within the last decade	70.6%	70.2%

Study Area: Treatment and control tracts

Non-Study Area: Census tracts of the Dallas-Fort Worth MSA not in the treatment or control area

Source: American Community Survey 2010

Figure 1.1: 1983 Dallas Transit Service Plan depicting planned rail infrastructure and High Occupancy Vehicle (HOV) lanes

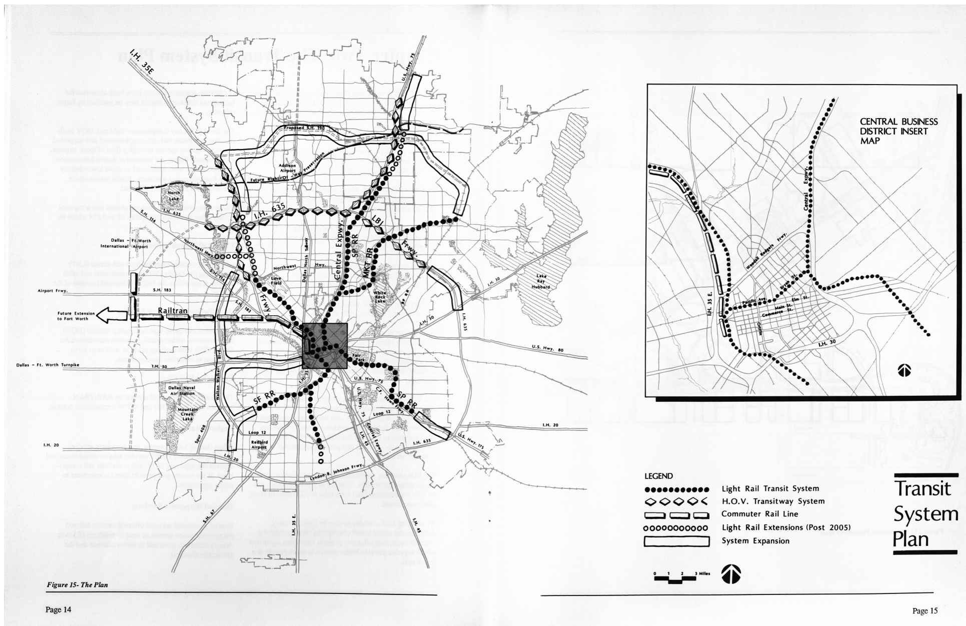

Figure 1.2: Dallas 1989 Transit Plan. Rail infrastructure as planned by the 1989 Dallas Transit Service Plan. Source: New Directions for Dallas Area Rapid Transit. Transit System Plan 1989.

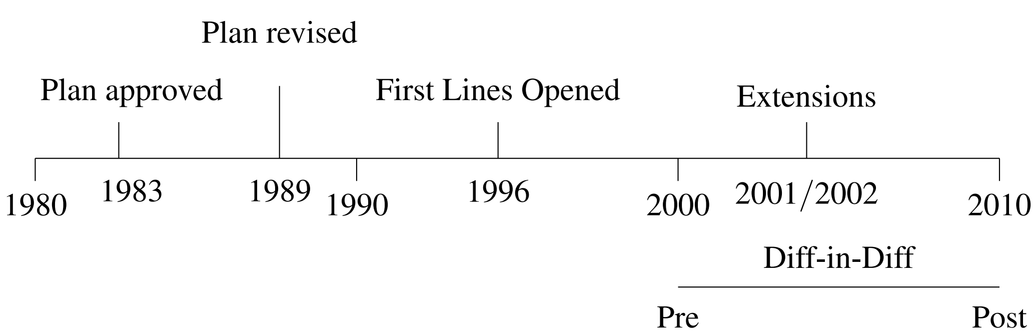

Figure 1.3: Detailed Timeline of Natural Experiment

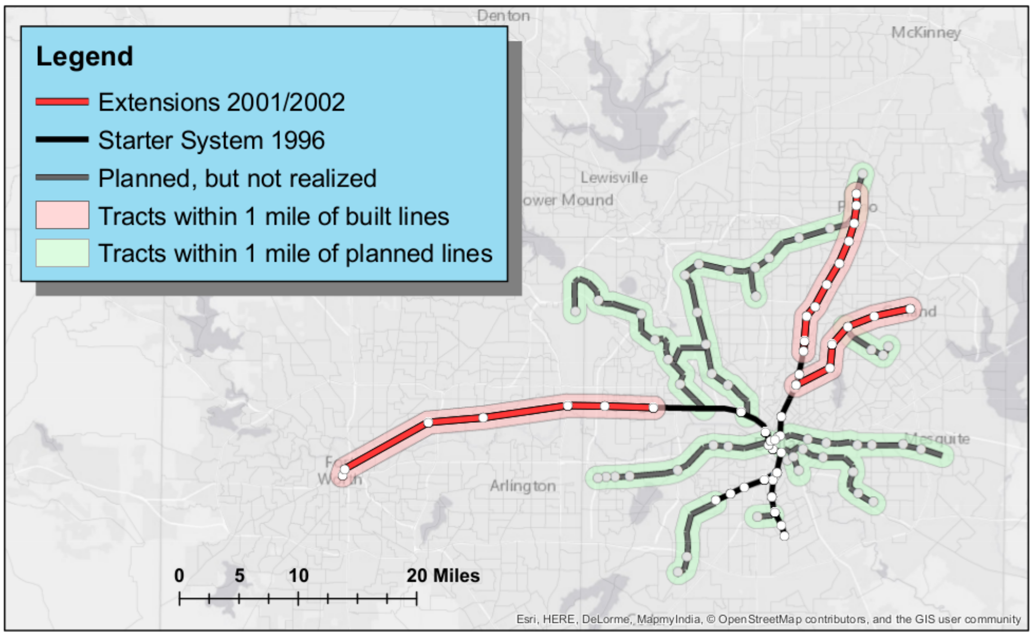

Figure 1.4: Rail infrastructure as planned by the 1983 Dallas Transit Service Plan. Dark bold lines represent the light rail starter system as it was in operation in 1996. Red lines represent extensions built between 2001 and 2002. Gray lines represent lines that were included in the 1983 plan, but were not built until 2010. White lines represent major express highways. Treatment neighborhoods are within the light red buffer. Similarly, control neighborhoods are in the light green buffer.

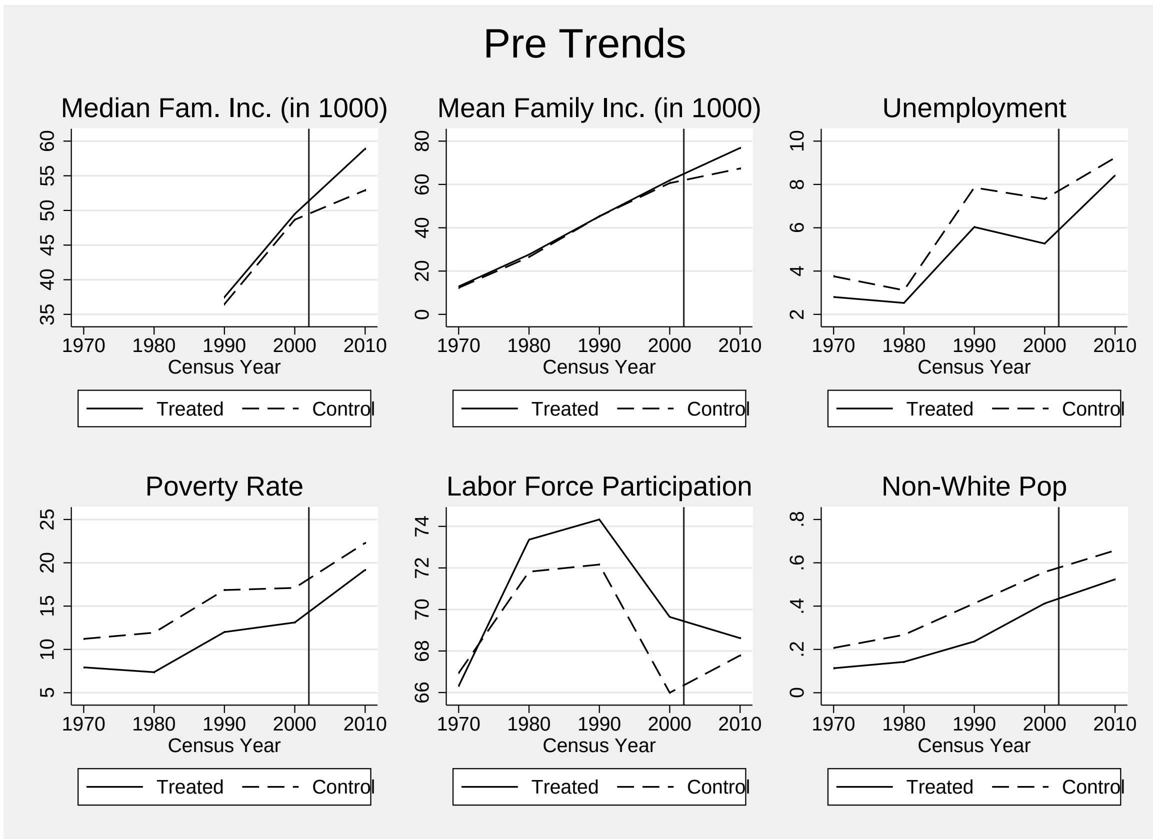

Figure 1.5: Trends of major census tract outcomes for the pre-treatment (1970-2000) and treatment period (2000)

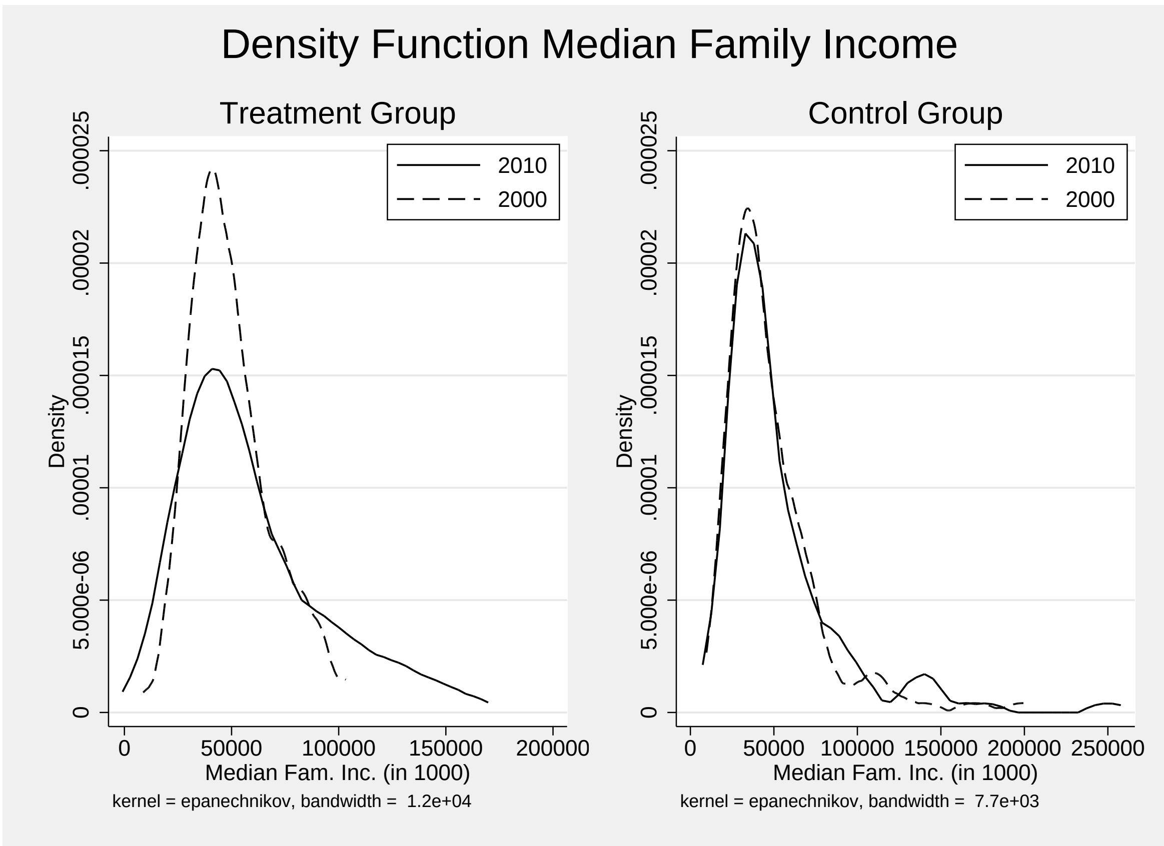

Figure 1.6: Empirical density function of median family income for the treatment group (left panel) and control group (right panel). Solid lines represent the distribution in 2010. Dotted lines represent the distribution in 2000.

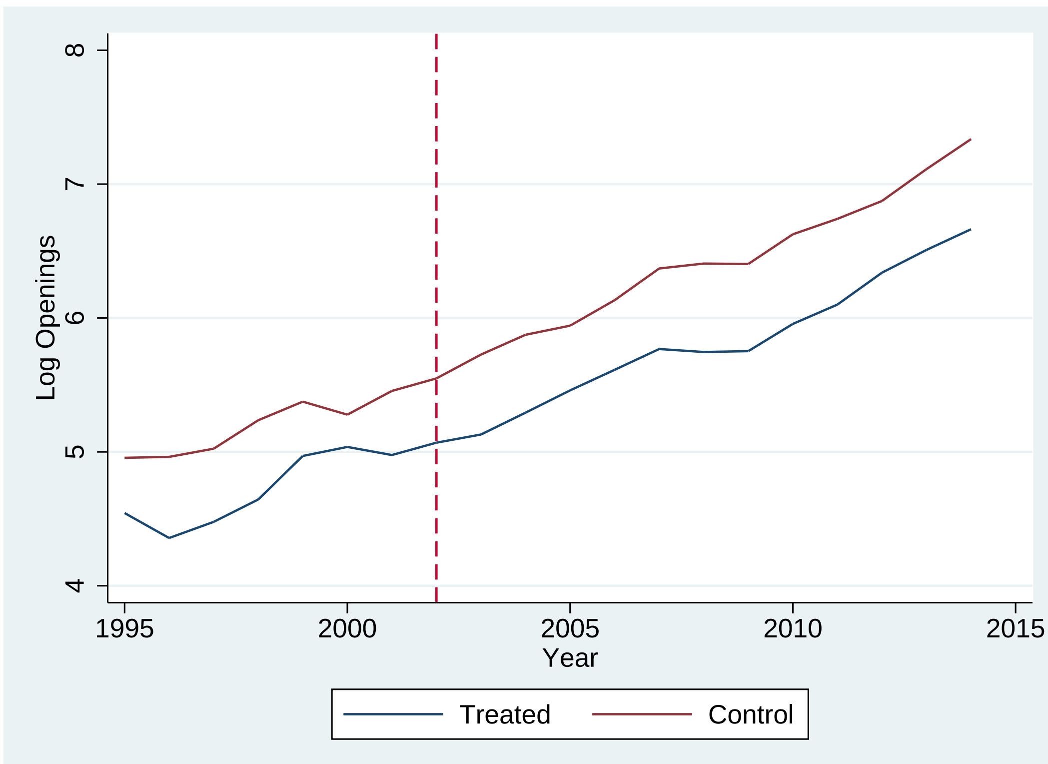

Figure 1.7: Trends in Business Openings within treatment tracts (blue line) and control tracts (red line) from 1995-2014. Vertical dashed line represents opening of the first rail extension in 2001.

1.7 Appendix

1.7.1 Robustness Checks

Alternative Distance Measures

In a first robustness check, vary the distance measure used. The initial choice of a one-mile radius around rail lines was arbitrary. In this section I perform robustness checks to see whether the results are strongly dependent on the initial one-mile choice. I first use a two-miles radius around light rail lines instead of the one-mile distance. This specification will likely include tracts that did not benefit from rail access due to the large distance to the treatment group. The results are consistent with this notion as the estimated treatment effect in Table 1.16, column (3) drops to $3,805.

I then use a half-mile circle around rail stations in column (5). This reduces the sample size significantly, but the treatment effect is still strongly significant and increases to $7,332. This larger estimate is in line with the notion that the effect of rail transit decays with distance to the rail lines. Estimating the augmented difference-in-differences specification confirms the strongly heterogeneous effect of rail transit by initial incomes and observes the same attenuating effect with distance.

Alternative Priority Schedules and Rail Standard

Next, I look at different priority schedules and rail standards. As described above, the original transit plan of Dallas was laid out in distinct phases. The higher the phase number, the later the lines of that phase were to be built. The eventually built rail lines I use in this study as the treatment group came primarily from phases one and two, while only a small stub of phase three was constructed. If planners had reason to believe that some neighborhoods were on different growth paths and assigned different phases according to these predictions, then the control group would consist of neighborhoods that are fundamentally different from each other. This might cause the control group to be a proper counterfactual for the treatment group. I therefore change the definition of the control group to tracts along rail lines that were planned only in phases 1 and 2. This definition excludes some smaller sections at the end of the light rail lines in the north and east and a longer stretch leading towards the DFW airport.

In another robustness exercise, I exclude the portion that runs as a commuter rail and focus on lines that were constructed according to light rail standards. This excludes the western stretch of built extensions that operate as the Trinity Rail Express (TRE) to the cities of Irving and Fort Worth. Commuter rail might potentially have different effects on the surrounding neighborhoods than light rail. Due to the heavy rail setup, distances between stations are longer and operation times are usually limited to peak hours with only sporadic service on evenings and weekends. The TRE is therefore geared almost completely towards commuters and does not facilitate recreational or other non-work travel.

The results are reported in Table 1.17. The findings are remarkably robust to the changes in the control group. The estimated increase in median family income without lines in phase 3 (column 3) changes little to \$5,717.2, while the increase for light rail only is larger at \$8,812.9, indicating that the commuter rail effect is lower. The estimated coefficients on the interaction term with initial neighborhood income for the regression with only phases 1 and 2 (column 4) and for the light rail only (column 6) are almost exactly the same as in the baseline setup. The standard errors however increase slightly with the lower sample size. There are small changes in the absolute magnitude of the treatment as the marginal effect at the lower quartile is now positive when excluding the TRE, indicating that the average effect of light rail on neighborhood income is slightly higher than that for heavy rail.

These robustness exercises show that the income sorting effect is not an artifact in the data driven by different trends in the outer neighborhoods. It also provides evidence that the measured effect of transit systems on local incomes is independent of the actual transit technology and holds for both light and heavy rail. Given the similarity of the results it also makes me more confident that transit access was randomly assigned and not part of a strategic decision to connect neighborhoods on faster growth paths first.

Inconsequential Unit Approach

An issue in identification of transport infrastructure is the potential positive selection into the treatment of fast growing areas. Naturally, transportation planners want to connect places that have high demand and that create enough ridership for the transit system to operate at economically reasonable capacity. In Dallas, the eventually built rail lines reached Plano and Garland, both suburban job centers with high employment density. If these areas received transit access because they were expected to grow even in absence of rail access, the selection into the treament is positive and the treatment effect consequently upward biased.

In a further robustness check, I therefore apply the inconsequential units approach pioneered by Chandra and Thompson (2000) to deal with this endogeneity issue. The inconsequential units approach assumes that transportation infrastructure is assigned to connect major places and that the areas en-route are inconsequential to the routing, and thus received transit access quasi-randomly (Redding and Turner, 2015). In Dallas, this assumption is plausible as most rail lines were planned to be built on pre-existing freight track, thus allowing transportation planners little discretion to re-route the radial lines extending from the city center.

To implement the inconsequential unit approach, I remove census tracts directly located at the terminal station of the built lines. In particular, I drop all neighborhoods from the sample that are within one mile of any planned or realized terminal station. This restriction drops 12 census tracts from the sample. Table 1.18 reports the results from this robustness check. The coefficient in column (3) drops only slightly from \$5309.9 (in the full sample in column 1) to \$5013.3. Although this estimate is not statistically significant at conventional significance levels under the smaller sample size anymore, it shows that the average increase in family incomes in transit neighborhoods is not driven by the areas surrounding the rail termini. Instead, the average increase is also present for the intermediate neighborhoods between downtown and the suburban job centers for which the identification is more plausible.

Comparing the results of the augmented difference-in-differences specification, the point estimate of the treatment effect interacted with initial neighborhood income in the inconsequential unit approach (column 4) is almost exactly the same as in the main specification (column 2), dropping only minimally from 0.334 to 0.329. The marginal effect for the lowest quartile of the neighborhood income distribution is still negative at -\$860. The similarity of estimates under the inconsequential units approach suggests that the sorting effect is not an artifact of the developments at the destination neighborhoods that might suffer from the strongest selection bias.

Propensity Score Matching

In this robustness check, I follow Ahlfeldt et al. (2016) and implement a propensity score weighting approach. The main concern with the identification strategy is that there might have been some implicit ordering of the rail lines and the reduced tax revenues were used to build the lines with the highest expected impact. This approach assumes that there is a latent variable guiding the selection and that this variable can be approximated reasonably well with the expected value of being in the treatment group from an auxiliary regression of the treatment status on neighborhood characteristics. Next, the threshold value of this variable that best predicts the designation of the treatment status is found and the identifying variation is restricted to observations near the threshold. The idea behind this approach is that there are neighborhoods for which it is easy to predict whether they will end up in the treatment or control group and are thus more heavily selected. Yet there are also neighborhoods for which, given the observables, the treatment status is ambiguous and the method assumes that, among them, treatment status is as good as randomly assigned. A weighting scheme then puts more weight on these latter observations that are more plausibly exogenous.

To implement this approach, I regress the eventual treatment status on a number of neighborhood characteristics from the pre-treatment, pre-planning year 1980. The explanatory variables include average family income, unemployment rate, poverty rate, the share of ethnic minorities, and a set of locational controls such as distance to the nearest freeway, distance to downtown, and residential density. The regression predicts the treatment status reasonably well with an R-squared of about 0.275. I next calculate the propensity score and determine the cut-off value that best predicts whether a neighborhood received transit access or not. I next restrict the identifying variation to observations around the cut-off by weighting the observations with a Gaussian kernel[27] as depicted in Figure 1.9. This weighting scheme oversamples observations close to the cutoff where random assignment is more plausible and puts low weight on observations where selection is potentially severe.

Table 1.19 reports the results from the propensity score approach and shows that the basic results of an on average positive, but highly heterogeneous effect still holds when applying the propensity score weighting. The point estimate on the treatment effect in the weighted specification in column (3) is estimated at \$5791 and differs only

[27] Following Ahlfeldt et al. (2016), the weights are calculated as $w_s = \frac{1}{\lambda\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{S_s-\bar{S}}{\lambda}\right)^2\right)$ where S_s is the propensity score of observation s, $\bar{S}$ is the cut-off value, and $\lambda = 1.06 \times \sigma N^{\frac{-1}{5}}$ is the "golden rule" bandwidth size.

slightly from the baseline estimate of $5300 without the weighting scheme in column (1). Even though the standard errors in this setup are much higher, this indicates that the positive effect of transit access on incomes is also present for neighborhoods that received it quasi-randomly. I find a similar result in the augmented difference-in-differences regression in column (4) with a point estimate on the interaction term of 0.310, that is statistically significant and similar to the baseline regression in column (2).

1.8 Acknowledgments

Chapter 1, in full, is currently being prepared for submission for publication of the material. The dissertation author, Kilian Tobias Heilmann, was the sole author of this paper.

I would like to thank Gordon Hanson, Marc Muendler, Eli Berman, Gordon McCord, Lawrence Broz, Prashant Bharadwai, Michelle White, Roger Gordon, Karthik Muralidharan, Matthew Kahn, Gabriel Ahlfeldt, Craig McIntosh, and participants of the IEB/UEA Summer School in Urban Economics, the European Meeting of the UEA, the UC Irvine Transportation, Urban and Regional Workshop, the North American Regional Science Conference, and the UCSD seminar series for helpful comments. Data on business openings were kindly provided by the Texas Comptroller of Public Accounts. Travel times were supplied by the North Central Texas Council of Governments.

Table 1.16: Dependent Variable: Median Family Income (different distance measures)

	(1) One Mile	(2) One Mile	(3) Two Mile	(4) Two Mile	(5) Half Mile	(6) Half Mile
Post	4090.9***	-751.1	394.1	6459.9	3819.5***	698.4
	(1495.4)	(3620.6)	(2029.7)	(4260.3)	(1061.0)	(2496.4)
Treat		1533.8		-1081.1		190.2
		(1164.1)		(1466.3)		(1188.4)
Post × Treat	5309.9*	-11324.0	7332.0**	-10586.7*	3805.5**	-7008.1*
	(3125.1)	(8133.6)	(3244.4)	(6259.1)	(1734.1)	(3760.9)
Median Family Income in 2000		1.030***		1.019***		1.029***
		(0.0143)		(0.0221)		(0.0135)
Post × Median Family Income 2000		0.0995		-0.127		0.0591
		(0.0901)		(0.117)		(0.0564)
Treat × Median Family Income 2000		-0.0229		0.0202		-0.00281
		(0.0212)		(0.0246)		(0.0169)
Post × Treat × Median Family Income 2000		0.334*		0.374**		0.210**
		(0.177)		(0.153)		(0.0814)
Census Tract Fixed Effects	Yes	No	Yes	No	Yes	No
Marginal Effect at 25th percentile		-594.2		313.7		143.7
Marginal Effect at 75th percentile		8707.5		11908.7		6696.1
Observations	325	325	128	128	558	558
R squared	0.122	0.868	0.105	0.924	0.136	0.890

Standard errors clustered at the census tract level: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

Table 1.17: Dependent Variable: Median Family Income (different phases)

	(1) Phase 1-2-3	(2) Phase 1-2-3	(3) Phase 1-2 only	(4) Phase 1-2 only	(5) Light rail only	(6) Light Rail only
Post	4090.9***	-751.1	3683.6**	-1332.4	3669.9***	-1191.8
	(1495.4)	(3620.6)	(1572.6)	(4324.3)	(1395.5)	(3649.7)
Treat		1533.8		1821.3		2058.9*
		(1164.1)		(1326.6)		(1115.8)
Post × Treat	5309.9*	-11324.0	5717.2*	-10742.6	8812.9**	-8668.9
	(3125.1)	(8133.6)	(3164.1)	(8482.2)	(4062.8)	(9960.6)
Median Family Income in 2000		1.030***		1.031***		1.029***
		(0.0143)		(0.0183)		(0.0137)
Post × Median Family Income 2000		0.0995		0.110		0.100
		(0.0901)		(0.115)		(0.0901)
Treat × Median Family Income 2000		-0.0229		-0.0239		-0.0309
		(0.0212)		(0.0242)		(0.0195)
Post × Treat × Med. Fam. Inc. 2000		0.334*		0.324*		0.331
		(0.177)		(0.192)		(0.201)
Census Tract Fixed Effects	Yes	No	Yes	No	Yes	No
Marginal Effect at 25th percentile		-594.2		-382.5		1917.1
Marginal Effect at 75th percentile		8707.5		7257.5		11136.7
Observations	325	325	299	299	307	307
R squared	0.122	0.868	0.122	0.859	0.139	0.869

Standard errors clustered at the census tract level: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Treatment group: census tracts within 1 mile of built rail lines

Control group: census tracts within 1 mile of planned, but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

Table 1.18: Inconsequential Unit Approach, Dep. Variable: Median Family Income

	(1) Baseline	(2) Baseline	(3) IUA	(4) IUA
Post	4090.9***	-751.1	4200.8***	-739.5
	(1495.4)	(3620.6)	(1594.8)	(3631.5)
Treat		1533.8		1586.5
		(1164.1)		(1159.3)
Post × Treat	5309.9*	-11324.0	5013.3	-11385.1
	(3125.1)	(8133.6)	(3331.5)	(8316.8)
Median Family Income in 2000		1.030***		1.031***
		(0.0143)		(0.0162)
Post × Median Family Income 2000		0.0995		0.101
		(0.0901)		(0.0907)
Treat × Median Family Income 2000		-0.0229		-0.0216
		(0.0212)		(0.0223)
Post × Treat × Median Family Income 2000		0.334*		0.329*
		(0.177)		(0.181)
Census Tract Fixed Effects	Yes	No	Yes	No
Marginal Effect at 25th percentile		-594.2		-860.3
Marginal Effect at 75th percentile		8707.5		8464.7
Observations	325	325	303	303
R squared	0.122	0.868	0.115	0.868

Standard errors clustered at the census tract level: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Treatment group: census tracts within 1 mile of built rail lines (excluding termini)

Control group: census tracts within 1 mile of planned, but not built rail lines (excluding termini)

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

Table 1.19: Propensity Score Weighting. Dependent Variable: Median Family Income

	(1) Baseline	(2) Baseline	(3) Weighted	(4) Weighted
Post	4090.9***	-751.1	4595.7***	-1422.0
	(1495.4)	(3620.6)	(1491.2)	(3377.7)
Treat		1533.8		2549.9
		(1164.1)		(1789.8)
Post × Treat	5309.9*	-11324.0	5791.8	-12469.0
	(3125.1)	(8133.6)	(3770.8)	(8132.5)
Median Family Income in 2000		1.030***		1.048***
		(0.0143)		(0.0253)
Post × Median Family Income in 2000		0.0995		0.125*
		(0.0901)		(0.0751)
Treat × Median Family Income in 2000		-0.0229		-0.0473
		(0.0212)		(0.0321)
Post × Treat × Median Family Income in 2000		0.334*		0.310*
		(0.177)		(0.180)
Census Tract Fixed Effects	Yes	No	Yes	No
Marginal Effect at 25th percentile		-594.2		-2533.6
Marginal Effect at 75th percentile		8707.5		5977.6
Observations	325	325	323	323
R squared	0.122	0.868	0.158	0.856

Standard errors clustered at the census tract level: $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Treatment group: census tracts within 1 mile of built rail lines

Control group: census tracts within 1 mile of planned, but actually not built rail lines

Pre-Treatment Period: 2000. Post-Treatment Period: 2010

Regressions include location controls

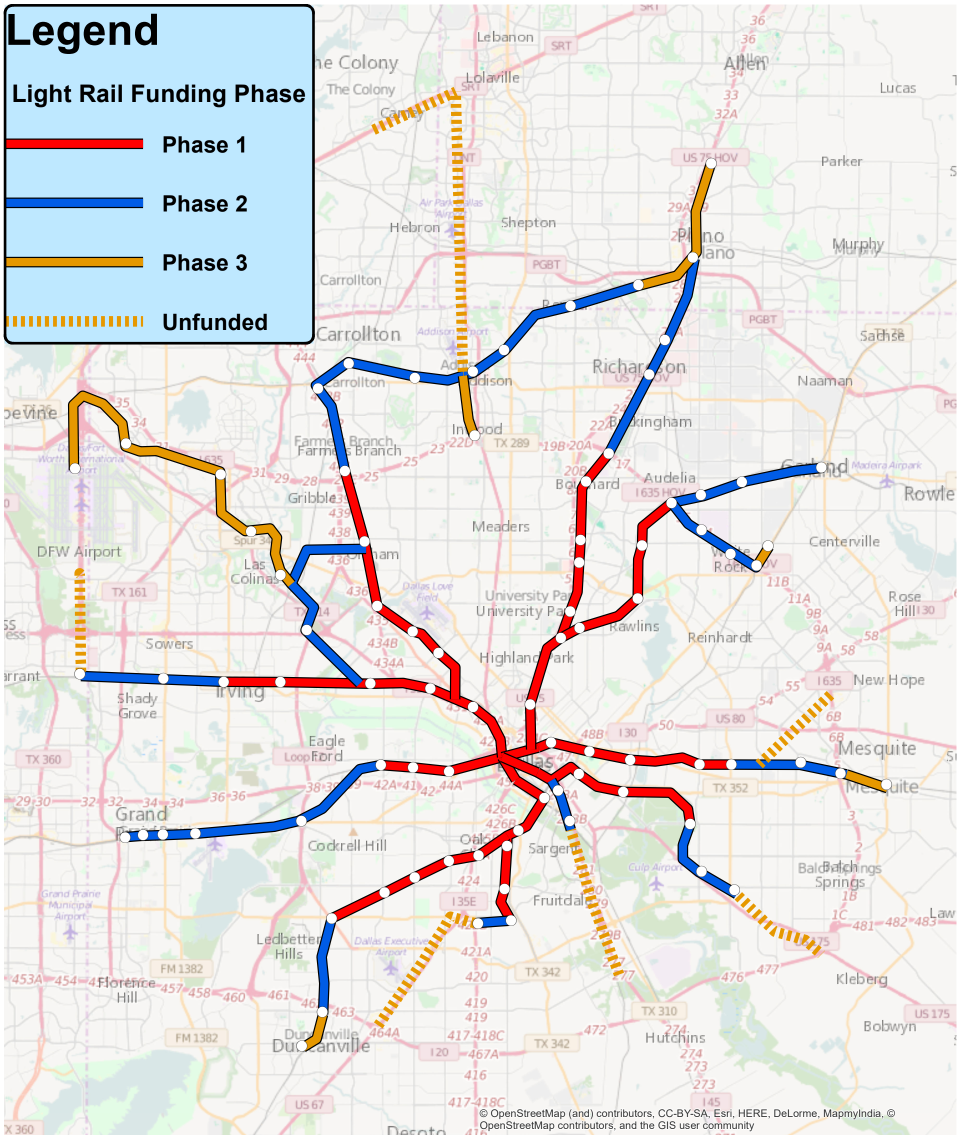

Figure 1.8: Reproduced 1983 Dallas Transit Service Plan depicting different planning phases.

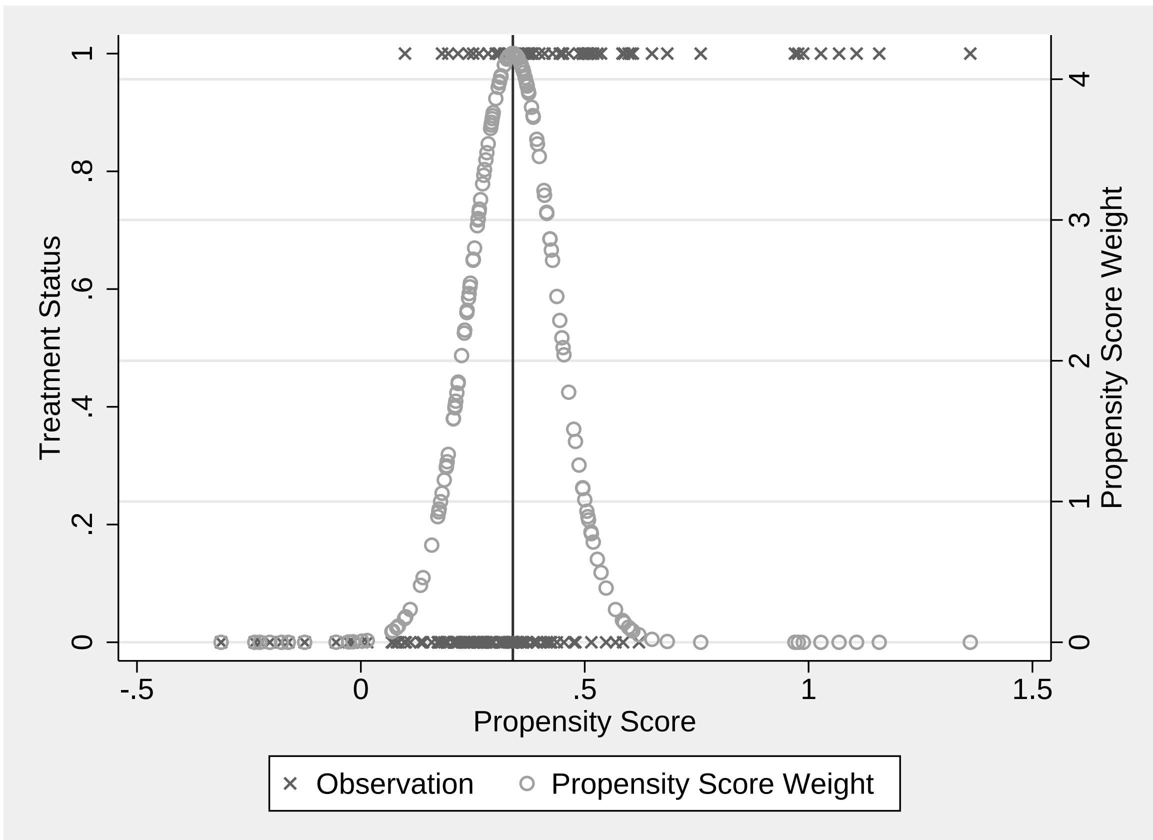

Figure 1.9: Propensity Score Weighting Scheme. Propensity scores are obtained as predicted values from a regression of the binary treatment status on a variety of neighborhood characteristics from the 1980 census and locational controls. The propensity score weight is calculated with a Gaussian kernel around the threshold value of .3395.

Chapter 2

Does political conflict hurt trade? Evidence from consumer boycotts

Abstract: I estimate the impact of international conflict on bilateral trade relations using several incidents of politically motivated boycotts: The boycott of Danish goods by Muslim countries following the Muhammad Comic Crisis in 2005/2006, the Chinese boycott of Japanese goods in response to the Senkaku/Diaoyu Island conflict in 2012, the boycott of French products in the US over the Iraq War in 2003, and Turkey's boycott of Israel over the Gaza conflict in 2014. The results from difference-in-differences regressions and the synthetic control group method show that boycotts can have strong negative effects on bilateral trade in both goods and services. I estimate an average one-year trade disruption of 18.8% in the case of Denmark, 2.7% for Japan, and 1.7% of French imports, where in the latter two cases this effect is only short term. For all boycott instances, this is only a minor share of overall exports of the boycotted country over the same period. For the Iraq and Gaza conflicts, there is a reciprocal negative effect on the boycotted countries' imports from the boycotter. Product-level results are in line with intuition: Boycotts are most effective for consumer goods, especially highly-branded signature export goods such as Japanese cars, while having at most a temporary effect on intermediates and capital goods. An event study on Japanese stock market returns suggests that the Chinese boycott depressed stock values of explicitly boycotted Japanese firms only temporarily.

Consumer boycotts ; International trade ; International political economy ; Economic diplomacy

JEL classification: F14 ; F51 ; F52

2.1 Introduction

Trade policy has long been a popular tool in relations between states. Trade agreements can strengthen inter-state relations and a large literature in political science has worked on international trade's role in promoting peace and interstate cooperation (e.g. Gartzke et al. (2001), Barbieri (2002), Li and Reuveny (2011), Massoud and Magee (2012)). At the same time, international trade can be used as a policy means in the case of conflict through sanctions, embargoes, and boycotts. Trade boycotts between countries are a special form of these policy tools. They have been used throughout history to punish or coerce specific behavior among trading partners. Examples of international conflicts where boycotts were used include the repeated boycotts of Japan by China throughout the 1930s in response to the Japanese invasion (Lauterpacht, 1933), the boycott of Israel by the Arab League after formation of the Jewish state in 1948, the worldwide boycott movement in protest of South Africa's apartheid system in the late 1950s, and the consumer boycott against French products over nuclear testing in the 1990s. Most recently, the importance of international trade boycotts has been highlighted by Russia's state-led import ban of agricultural products from Europe in response to sanctions over Russian interference in neighboring Ukraine.

These events share the common characteristic that they are not motivated by economic rationale, such as inferior product quality, but rather by political events and thus allow us to learn about how shocks to international relations affect trade. In contrast to the more frequent boycotts against specific firms, such as the boycott against Shell in 1995, they are directed against entire countries. They seem to become an option when other means of coercion, such as war or the severing of diplomatic relationships appear to be infeasible. The latter boycotts of the 21st century seem to be a simple continuation of earlier practices, but several developments portend an increase in the importance of boycotts as policy tools and warrant further research.

In a world characterized by less violence and decreasing tolerance for militarized conflict between states Pinker (2011), trade policy is the prevailing tool to carry out international disputes.[1] In addition, international trade has surged over the past decades, making boycotts potentially more harmful to trading partners. This is especially true since the nature of trade has changed from a simple exchange in final goods to a system of international production sharing. The advent of the internet has also changed international relations and the importance of governments. Being able to communicate and coordinate their actions online, consumer boycotts enable the public to become a political agent in international relations. In the case of the Chinese consumer boycott against Japan in 2012 that I study in this paper, the internet may have played a crucial role in organizing the boycott, with the Chinese government having limited control over the reaction on the streets. This raises questions on how governments and the populace interact when it comes to foreign relations (Weiss, 2013, 2014) and how different regime types favor the emergence of consumer boycotts.

An important question is whether these new types of boycotts are effective. Aside from a reduction in import demand, international conflicts might hurt trade by putting business partners at personal risk when traveling, through latent government intervention or even through the boycotted country's refusal to export in response to the aggression. Similarly, boycotts can fail in many dimensions. At first, if the boycotted country's exporters can easily redirect their sales to domestic or other foreign markets, the potential economic loss may be small. Secondly, even if disrupted exports hurt the exporting country significantly, boycotts are a costly tool, since the boycotting country is also giving up on its gains from trade. This is even more true in a world characterized by increasing international integration of production, often within firms (Zeile, 1997). Today, trade is not primarily in final goods anymore, but the share of processing trade is rising. If production of the boycotting country depends heavily on imports from the boycotted

[1] Besides boycotts, this also includes trade sanctions. While not the focus of this paper, the prevalence of trade policy in solving international conflicts is reflected in the recent economic sanctions against Iran and North Korea.

country, this will raise the costs of the boycott, and it might render it an incredible threat. Furthermore, consumer-led trade boycotts rely on collective action that can be difficult to organize. Friedman (1999) and John and Klein (2003) study consumer boycotts and their inherent small-agent problem, i.e. the success of the boycott depends on a mass of participants, but every individual's impact and motivation to join in is low. To explain that consumer boycotts do happen, they propose a variety psychological motivations, such as guilt and self-esteem or simply an exaggerated sense of one's own effectiveness. These theoretical studies suggest that consumer-organized boycotts are short-lived.

The empirical literature on the impact of boycotts on international trade has found contradicting results, mainly from boycotts in the aftermath of the Iraq War of 2003. Michaels and Zhi (2010) estimate that US-French trade deteriorated by about 9% in 2003 when France's favorability rating in the US fell sharply over its refusal to intervene in Iraq. (Pandya and Venkatesan, 2016), using supermarket scanner data, find that brands that are perceived as being French lose market shares in weeks with high media attention of the boycott. They estimate the implied costs of this boycott to be similar to the costs of an average product recall. Similarly, Chavis and Leslie (2009) find a 26% reduction in weekly sales of French wine in the US, but Ashenfelter et al. (2007) attribute this decline to boycott-unrelated influences. Clerides et al. (2015) find a significant but short-lived drop in sales of US soft drinks in the Middle East, but cannot find a similar effect on other goods. These studies are based on local sales and do not investigate the effect on trade. Davis and Meunier (2011) study quarterly trade relationships between the US and France as well as between China and Japan for the years 1990-2006, thus including the boycott of French goods. They do not find any significant link between negative events involving these countries and the level of goods exchanged, but find that trade, as well as foreign direct investment, continued to grow sharply in the period studied.

Besides the focus on explicitly announced boycotts, there is a new literature studying the relationship between other political conflicts and international economic relations. Fuchs and Klann (2013) study countries' trade with China if they officially receive the Dalai Lama. China perceives any formal relations with the Tibetan spiritual leader as an interference into internal political affairs and threatens countries that do so with a reduction of trade. The authors find a significant negative short-term effect of state visits on trade volumes and confirm that, even though the effect dies out after one year, countries are willing to use trade as a tool to enforce their political will. Fisman et al. (2014) and Govella and Newland (2010) study the effects of Sino-Japanese conflicts in the 21st century on the stock market value of Japanese firms using an event study approach. They find that stocks of Japanese companies with a high share of sales to China lose value compared to companies with a low exposure to China.

The aim of this paper is to evaluate the effectiveness of international trade boycotts, to quantify their impact, and thus to learn about the consequences of international conflict on trade relationships. The contributions that distinguish it from previous studies are manifold: At first, with the Mohammad Comics boycott I study an international conflict that was unexpected and plausibly exogenous to unobserved trade-related confounding effects, thus providing superior identification to study the impact of political conflict on bilateral trade. The previous literature has largely focused on the US boycott of French products, an incidence that might be confounded with other trade-related effects of the looming Iraq War. Secondly, monthly product-level data allows me to study the boycotts' complex short-term impacts which cannot be uncovered using only yearly or quarterly data. The availability of only low-frequency data might be the major reason why the previous literature came to contradicting results regarding the effect of consumer boycotts. Furthermore, the high frequency of the data enables me to extend the analysis to recent incidents that would be impossible to study with yearly data, such as the Chinese boycott of Japanese goods in the aftermath of the Senkaku/Diaoyu Island conflict in 2012, and Turkey's boycott of Israel over the Gaza conflict in 2014, thus expanding the set of conflicts to learn from considerably. In addition, the fine product disaggregation of the data allows me to estimate different impacts for consumer, intermediate, and capital goods based on the full range of traded products, rather than having to a priori choose boycott-prone products like French wine or US soft drinks. This dimension of the data offers insight into the main drivers behind the boycotts. Finally, I apply the synthetic control group methodology to construct data-driven counterfactuals showing that the results are robust to omitted variable bias.

The results show strong heterogeneity in the response among the boycotting countries, with an average one-year reduction in imports of about 18.8%, 2.7%, 1.7% of total trade in the Muslim boycott case, Senkaku conflict, and the US consumer boycott against France respectively. I do not find a negative effect for Turkish imports from Israel following the Gaza war in 2014, but instead observe that Israel reduces its imports from Turkey by 12.3%. Product-level analysis shows that the impact is concentrated in consumer goods and especially in highly branded goods such as Japanese cars. I find only minor effects for intermediates and capital goods, being consistent with the notion that international trade boycotts are mainly carried out by consumers and not by firms or governments. This is confirmed by results from the multi-country Muhammad Comic boycott, where countries with a higher press freedom boycott more, indicating that consumers find it easier to organize and participate in boycotts in open regimes. While the estimated disruption in imports from the boycotted country can be large, the reduction in total exports of the boycotted country is low in all boycott cases (0.4% for Denmark, 0.5% for Japan, and 0.4% for the US). This suggests that even though an individual firm of the boycotted country might be hit hard, the overall effect on the export sector is small.

An event-study analysis based on time series variation does not hint towards substitution of imports or exports towards non-boycotting countries.

The paper is organized as follows: Section 2 provides background information on the events studied, while section 3 outlines the empirical implementation. Section 4 presents the findings on both aggregate and product-level data. Section 5 concludes.

2.2 Background

In this section, I provide background information on the international conflicts used in the study and describe the events leading up to the boycotts as well as their political consequences.

2.2.1 Muhammad Cartoon Crisis

On September 30, 2005 the Danish newspaper Jyllands-Posten published a series of cartoons depicting Islamic prophet Muhammad in an unfavorable manner, the most striking one showing him with a bomb in his turban.[2]. Not only is the depiction of the prophet forbidden in several branches of Islam, but Muslims felt that the comics equated them to terrorists, thus the comics had a religious as well as political dimension. Even though Danish Muslims protested the publication from the very beginning, it was not until early 2006 that the controversy became international after the comics had been reprinted in Arabic newspapers. Violent protests sparked in many Middle Eastern countries, leading the ambassadors of several Muslim countries to unsuccessfully demand an official apology by the Danish government and prosecution of the cartoon artists.

The months of January and February 2006 saw further escalation of the conflict with Western embassies being attacked in Damascus, Beirut, and Tehran, leaving several

[2]For a detailed narrative of the events, see Rask Jensen (2008)

dozen people dead. With the Danish government refusing an official apology, religious leaders in Saudi Arabia called for a boycott of Danish goods on January 26, 2006, publishing a boycott list of Danish firms.[3] Soon other Muslim countries joined the boycott. The French supermarket chain Carrefour preemptively removed Danish goods from its shelves in the Middle East and several Danish food producers, such as Arla Foods, reported large losses.[4] At the same time, a counter-boycott campaign called "Buy Danish" was called for, but it remains unclear whether this campaign gained enough media attention to have any large scale effects.[5]

The scandal about the Muhammad cartoons eventually lost public attention and the protests calmed down, though several incidents in later years were linked to the cartoons, e.g. the 2008 and 2010 attempts to assassinate the creator of the most controversial of the cartoons which could be prevented by police.

2.2.2 Senkaku/Diaoyu Islands Conflict

The Senkaku (in Japanese) or Diaoyu (in Chinese) Islands are a small group of islets unsuited for settlements in the East China Sea approximately 170 km North-East of Taiwan. In the aftermath of the First Sino-Japanese War (1884-85) and the subsequent invasion of Taiwan, Japan began to survey the islands and claimed them as its territory. After the Treaty of San Francisco formally established peace after World War II, Japan ceded all its claims to Taiwan and the nearby Okinawa islands came under US control. When the Okinawa islands were returned to Japan in 1972, it tacitly took control of the Senkaku islands as well and retains a military presence on the islands until today.

In 1968, possible oil reserves were found in the area surrounding the Senkaku/-Diaoyu islands leading to claims of both Mainland China and the Republic of China

[3]Examples of these lists can be found on http://shariahway.com/boycott/index.htm.
[4]http://www.nytimes.com/2006/01/31/international/middleeast/31danish.html?_r=0.
[5]http://www.foxnews.com/story/2006/02/16/muslim-boycotts-hurt-danish-firms/.

(Taiwan) to the islets that were rejected by Japan, leaving the territorial conflict remained unsolved. It was not until the 2000s when several incidents brought the Sen”-kaku”-/Diao”-yu conflict back to public attention. Between 2006 and 2011 several activist groups from Mainland China, Taiwan and Hong Kong arrived at the islands to proclaim Chinese sovereignty and were expelled by the Japanese navy immediately.

While these events worsened Japanese-Chinese relationships, the conflict only escalated after Japan announced to purchase the islands from their private owner in August 2012 and de facto established sovereignty over the archipelago. This led to anti-Japanese protests in several Chinese cities that later turned violent. Japanese businesses in China were attacked and protesters called for a boycott of Japanese goods. Japanese-Chinese relations deteriorated drastically when further naval standoffs near the disputed islands occurred, leading to worldwide fears over a military conflict. While the dispute has calmed down and lost media attention, the major issue is still unresolved and remains a major problem in Japanese-Chinese relations.

2.2.3 US Boycott of France

The months preceding the invasion of Iraq by US-led forces in March 2003 caused widespread conflicts in international relations. While some European countries supported action against Saddam Hussein’s regime, others, notably France and Germany, vocally opposed any intervention that was not backed by the UN. France’s favorability ratings in the US began to plummet starting in February 2003 and conservative media outlets called for boycotts of French goods to punish the perceived betrayal of a supposedly close ally. Relations between the two states deteriorated so much that even Congress’s food menu was officially relabeled French fries as “freedom fries”.[6]

[6] While US consumers were boycotting French products, the US itself became the victim of a boycott movement. The eventual invasion of Iraq triggered a boycott movement against US-American products in the Middle East. Clerides et al (2015) report the existence of boycott lists of American brands and find

2.2.4 Turkey's Boycott of Israel

On July 8th 2014, the long-lasting conflict between Israel and the Palestinians escalated again when Israeli military launched airstrikes on Gaza after heavy shelling of Israeli territory by Hamas. Two weeks later, the Israeli Defense Force led a ground invasion into the Gaza strip to destroy smuggling tunnels which resulted in the death of more than 2,000 Palestinians, around 1,500 of them being civilians[7]. Public outcry over the humanitarian toll of the conflict sparked anti-Israel protests in Turkey with Turkish prime minister equating Israel's actions to genocide. The Turkish trade union TESK launched a boycott call against Israel in late July. At the same time, polls in Israel showed that Israelis were boycotting Turkey and especially its holiday destinations.

2.3 Methodology

To evaluate the impact of the boycotts on trade, I estimate difference-in-differences models of logged exports from the boycotted country $Y_{j,t}$ to all its trading partners j at time t at monthly frequency. I determine treatment status by participation in the boycott and thus use non-boycotting countries as the control group. I include the typical gravity regressors GDP and distance provided by CEPII and control for a time trend and monthly fixed effects. The regression equation is given by

$$Y_{j,t} = \alpha + \beta_1\, Treat_j + \beta_2\, Post_t + \beta_3\, Treat_j \times Post_t + \beta_4 \log GDP_t + \beta_5 \log dist_j + \beta_6 t + \varepsilon_{j,t}. \tag{2.1}$$

The difference-in-differences approach might suffer from omitted variable bias if important determinants of trade are not controlled for. Despite the empirical success of

a negative effect of US softdrink sales in the Middle East, but are unable to detect a similar effect for detergents.

[7]Source: OCHAOPT (http://reliefweb.int/sites/reliefweb.int/files/resources/annual_humanitarian_overview_2014_english_final.pdf)

parsimonious gravity equations in the cross-section, this relationship describes long-time averages and in the short-term, there may be many more unobserved confounding factors, such as a country's industry composition. To avoid this problem and to consistently construct a suitable control pool, I follow the synthetic control group method first used in Abadie and Gardeazabal (2003) and later further developed in (Abadie et al., 2010, 2015). The synthetic control group method follows a pragmatic data-driven approach to choose the right control group by creating a weighted average of all the available control units. The weights are chosen such that the synthetic control group resembles the actual treatment unit in both the outcome variable as well as in any known explanatory characteristics in the pre-treatment period. The idea behind the method is to indirectly control for any unobserved factor by matching on previous outcomes. An estimate for the treatment effect can then be calculated by the difference between treatment unit and the synthetic control unit in the post-treatment period.

One problem of the synthetic control group methodology is the inability to calculate standard errors. In practice, the fit between treatment group and the synthetic control group in the pre-boycott period will not be perfect, but subject to idiosyncratic shocks captured in the error term $\varepsilon_{j,t}$. This will bring randomness into the estimate of the treatment effect β_t. The exact distribution of the estimate depends on the unobserved parameter vector λ_t and therefore cannot be computed. A pragmatic ad-hoc approach to evaluate the significance of the parameter estimates is to compare them to the prediction error in the pre-boycott period. The intuition is that if the synthetic control group fits the actual treatment unit poorly before the boycott happened, this would undermine the confidence in the estimate of the treatment effect. If the fit between the actual treatment country and its synthetic control, however, is close in the pre-boycott period, we can be more confident in assuming that any divergence after the treatment is actually caused by the boycott and not due to unrelated shocks.

I formalize this idea by testing for a structural break in the time series of the error term $\varepsilon_{j,t}$ and test the model $\varepsilon_{j,t} = \sum_{k=1}^{6} \rho_k \varepsilon_{j,t-k} + \beta_j, t + u_j, t$ against the simple alternative $\varepsilon_{j,t} = \sum_{k=1}^{6} \rho_k \varepsilon_{j,t-k} + u_t$. The six-month autoregressive specification and inclusion of clustered standard errors allows for the possibility of a correlation over time and between countries. In specific, I report p-values of an F-test with the null hypothesis $H_0 : \sum_{t=T_0+1}^{T_0+d} \beta_t = 0$ where d denotes the horizon of the effect.

To complement the analysis, I also perform placebo tests that traditionally have been used in the context of synthetic control groups. There are two dimensions where a placebo test can detect wrongful inference: Within a single time series, a random assignment of a treatment time should not break the close fit between actual and synthetic control group and should not produce large estimates of the treatment effect. If both series deviate even though there is no boycott, then this should warn us that the synthetic control group is merely picking up unrelated idiosyncratic effects. Furthermore, we can estimate the same treatment effect for the control countries. If these countries are indeed unaffected by the boycott, the synthetic control group method should not find large treatment effects. If however the control countries seem to be negatively affected by the boycott, this would hint to mis-specification in the model and would greatly undermine our confidence in the method.

2.4 Results

This section presents the data sources, descriptive statistics, and the estimation results of both the difference-in-differences and synthetic control group methods for each boycott case.

2.4.1 Mohammad Cartoon Crisis

Data and Descriptive Statistics

I use data from the online portal of Statistics Denmark. This dataset covers Danish export values in local Danish krona (DKK) to virtually all trade partners at monthly frequency at the two-digit and five-digit SITC classification from the late 1980s onward. Unsurprisingly, being a small country, imports from Denmark make up only a small share of the Muslim world's total trade. On average, only 0.29% of all imported goods of the 34 countries with at least 75% Muslim population[8] stem from Denmark. Similarly, Danish exports to the Muslim world as a share of its total exports are relatively small accounting for 2.66% of Danish exports to all trading partners in 2004. Even the biggest Muslim trading partner, Saudi Arabia, accounted for less than half a percent of Danish exports in 2004 (see Table 2.16 in the appendix).

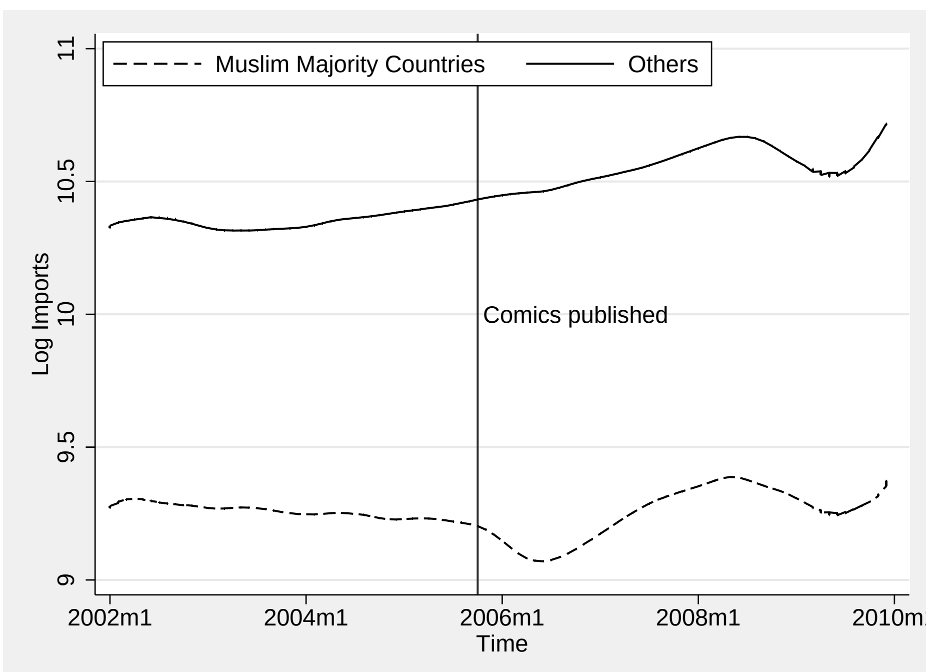

Figure 2.1: Lowess Plot (Imports from Denmark)

[8]For the exact list of these countries, see Table 2.15 in the appendix.

Examining export values from Denmark to the boycotting countries shows that monthly trade data is characterized by high volatility, seasonal patterns, and possibly changing time trends. It is not uncommon that Danish exports to these countries increase by a multitude over one month or that trade completely collapses even in the pre-boycott period. The strong month-to-month swings are more prominent for the smaller export partners, so I exclude countries with zero values from the analysis.[9] A lowess plot of imports from Denmark by Muslim majority and minority countries in Figure 2.1 reveals a pronounced dip for Muslim countries at the end of 2005 which is not present for non-Muslim countries. Table 2.1 summarizes the descriptive statistics of the time series for the three treatment countries with the largest imports from Denmark: Saudi Arabia, Turkey, and the United Arab Emirates.

Table 2.1: Descriptive Statistics (Log Danish Exports)

Country	Saudi Arabia	Turkey	UAE	Aggregate
Mean (in DKK)	176,710	205,965	129,765	1,036,268
Standard Deviation	31,738	87,048	37,468	188,399
Std Dev as Mean	18.0%	42.3%	28.9%	18.2%
Minimum	100,143	77,810	82,677	710,530
Maximum	272,301	462,941	422,072	1,596,562
Min % Change	-33.0%	-60.2%	-54.4%	-29.1%
Max % Change	67.8%	92.6%	194.7%	46.9%
Seasonality p-value	0.37	0.32	0.68	N/A

Statistics over the pre-boycott period October 2000 to September 2005.
Seasonality p-value is the p-value of a F-test testing for joint significance of monthly indicator variables in a linear time series regression.

Difference-in-Differences Results

Since the comics were published on the last day of September 2005 and a same-day effect is unlikely, I define October 2005 to be the first treatment period in the sample.

[9] For example, a complete disruption of trade with Kyrgyzstan would reduce Danish exports by only 0.004%.

This is considerably earlier than the official announcement of the consumer boycott in January 2006, but allows for undeclared boycotts as an immediate reaction to the insult. Instead of a binary treatment status, I use the share of Muslim population as a continuous treatment. Data on the Muslim population for each country is provided by the Pew Research Center. The results in Table 2.2 indicate that the treatment effect is negative and robust to including fixed effects. The coefficient on $Post \times Muslim$ suggests that a ten percent higher share of Muslim population reduces imports from Denmark by 3.7%. The elasticities with respect to GDP and distance have the expected positive and negative signs respectively. The negative coefficient on the share of Muslim population indicates that the treatment countries in general import less from Denmark than similar non-Muslim countries. Controlling for potentially endogenous exchange rate fluctuations, the treatment effect is still significant, but reduced to 2.2%.

Heterogeneity among the boycotting countries allows me to investigate the importance of different regime types for the effectiveness of the boycott. Consumer-organized boycotts are only possible if the populace is able to interact and draw masses to its cause. I estimate a triple difference model by interacting the treatment effect with a variable measuring the freedom of press as reported by Reporters Sans Frontières in the Quality of Government database. The coefficient on the triple interaction term in column (4) suggests that Muslim countries with a one-unit higher press freedom score reduce their imports from Denmark by an additional 1.18%. This suggests that more open countries allow for more organized action of their people and this strengthens the theory of the conflict being a consumer boycott.[10] In column (5), I interact the treatment effect with elasticities of substitution at the five-digit SITC level as measured by Broda and Weinstein (2006). This addition shows no significant effect on the boycott. If at all, highly substitutable goods are boycotted less, but the coefficient is imprecisely estimated.

[10]In regressions not reported here, I show that this result is robust against using alternative governance indicators such as the Polity IV score that ranks countries according to constitutional and practical criteria.

Table 2.2: Muhammad Comic Crisis: Results

Dependent Variable: Danish Exports	(1)	(2)	(3)	(4)	(5)
log GDP	0.950***	0.455***	0.511***	0.491***	0.258*
	(0.0366)	(0.115)	(0.111)	(0.112)	(0.147)
log Distance	-0.880***				
	(0.0787)				
Post	0.043	0.065*	0.049	-0.166	0.691***
	(0.0389)	(0.0376)	(0.037)	(0.129)	(0.072)
Muslim	-0.106				
	(0.179)				
Post × Muslim	-0.370***	-0.302***	-0.217**	0.204	-0.457***
	(0.098)	(0.091)	(0.090)	(0.215)	(0.175)
log Exchange Rate			-0.015		
			(0.085)		
Press Freedom				-0.0024	
				(0.0047)	
Post × Press Freedom				0.0039**	
				(0.0019)	
Press Freedom × Muslim				-0.0006	
				(0.0116)	
Post × Press Freedom × Muslim				-0.0118**	
				(0.0049)	
Elasticity (in 100)					-0.00801
					(0.00509)
Post × Elasticity					0.00298
					(0.00487)
Elasticity × Muslim					0.00595
					(0.0138)
Post × Elasticity × Muslim					0.0159
					(0.0155)
Country Fixed Effects	No	Yes	Yes	Yes	Yes
Trend and Month FE	Yes	Yes	Yes	Yes	Yes
N	13,518	13,518	10,267	12,954	5,037,984
adj. R^2	0.857	0.944	0.959	0.943	0.113

Standard errors in parentheses (clustered at country level)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To analyze the potentially heterogeneous effects on different product groups, I break up the analysis into three main product types: Consumer goods, intermediate goods, and capital goods.[11] Unlike consumer goods which merely reduce consumption, a boycott of intermediate and capital goods may have direct effects on the economy of the boycotting country if it depends heavily on foreign inputs. This drives up the cost of the boycott and we expect a weaker effect for these goods if countries choose their boycott strategy rationally. In addition, knowing which goods are boycotted allows us to gain some insight about who is the main driver behind the boycott. If the boycott is mainly consumer-driven, we should expect a higher trade disruption in consumer goods as compared to non-consumer goods. A large effect for non-consumer goods would suggest that local producers engage in the boycott as well or that governments restrict imports indiscriminately.

The results show the heterogeneity of the treatment effect for the different product types. While there is no statistically significant treatment effect for intermediate goods, we observe that a ten percent higher Muslim population is associated with a 6.5% and 2.8% drop in consumer and capital goods imports from Denmark respectively. This confirms that the boycott was most effective for products that individual consumers purchase and suggests that while capital goods are also affected, the nature of the boycott is primarily a consumer boycott. The difference between capital and intermediate goods might be explained by the fact that capital goods tend to be branded and are thus easier to recognize as of Danish origin than intermediate goods. I also use yearly data from the International Trade Centre on trade in services and analyze its response to the boycott. The negative effect is very strong at 4.8% suggesting that Muslim countries readily reduced travel, communication, financial and other services from Denmark. This is

[11] Where available, I use the Broad Economic Categories (BEC) classification developed by the UN Statistics Department to categorize SITC5 codes. Trade codes that are not available in the BEC were coded by my own judgment in close concordance with the logic of the BEC classification. The complete conversion table can be found in the online appendix.

not surprising as compared to trade in goods, trade in services requires more personal interaction between people in the conflict parties.

To answer the question whether the boycott announcement caused a two-way trade disruption, that is whether Danish consumers retaliated against the Muslim states, I apply the above methodology to Danish import data. The results in Table 2.19 indicate that the Danish did not boycott. While the estimates indicate a reduction in imports from Muslim countries after the comics were published, the standard errors are too high to conclude that there was a significant effect. Anecdotal evidence from newspaper articles also suggests that the boycott was a one-way trade disruption.

Table 2.3: Muhammad Comic Crisis: Results by Product Type

Dependent Variable: Log Danish Exports	**Consumer**	**Intermediate**	**Capital**	**Services**
log GDP	0.043	0.437***	0.637***	0.528***
	(-0.183)	(-0.116)	(-0.136)	(0.151)
Post	0.039	0.041	-0.055	0.464
	(-0.067)	(-0.053)	(-0.064)	(0.06)
Post × Muslim	-0.558***	-0.073	-0.221**	-0.481***
	(-0.173)	(-0.094)	(-0.11)	(0.166)
Constant	11.57***	9.393***	7.767***	-22.61
	(-1.083)	(-0.736)	(-0.859)	(24.73)
Country Fixed Effects	Yes	Yes	Yes	Yes
Trend and Month FE	Yes	Yes	Yes	Yes
Frequency	monthly	monthly	monthly	yearly
N	16,149	16,942	15,243	1,527
adj. R^2	0.862	0.909	0.841	0.951

Standard errors in parentheses (clustered at country level)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Synthetic Control Group Results

The synthetic control group method requires a binary treatment status, so I assign all countries that have a share of Muslim population of the total population of more than 75% into the treatment group. Conversely, I assign countries for which this share is less than 10% into the control group and drop all other countries to avoid contamination of the control group.[12] This leaves me with 34 countries in the treatment group and 100 countries in the control group (see Table 2.15 in the appendix).

Since the number of potential control units is large, I restrict the pool of controls to countries that are close in both distance and GDP in the month prior to the boycott. In specific, I allow the GDP to differ by 100% in both directions and distance to deviate by 4,000km. This avoids that the relatively small and close economies of the Middle East are replicated by large and distant countries like Japan and the US. While these restrictions seem arbitrary, they shrink the pool of control countries to an average of no more than ten units, a reasonable number to avoid overfitting 60 pre-boycott time periods. Experimenting with different specifications, the results tend to be fairly robust to these restrictions.

To calculate the value of the foregone trade, I simply add up the treatment effects of all treatment countries for each month and calculate the percentage loss as a share of total trade levels. Table 2.4 shows the estimated aggregate percentage reduction for a period of three, twelve, and 24 months. The results indicate that there was a statistically significant fall in imports from Denmark in the treatment countries which is robust to changes in the specification of the control group and sampling frequency. My preferred estimate in column (1) with all three predictors for the 19 treatment countries that take up at least 0.02% of all Danish exports shows that the short-term reduction in imports

[12]The distribution of Muslim percentage between countries is bimodal and most countries exhibit either a very high or very low Muslim population. Only few countries fall between the thresholds and the results are robust to changes in the cutoffs.

reaches 12.4% after three months. The boycott then intensifies to a 18.8% trade loss within one year; after which the impact is reduced to 14.7% after 24 months.

Table 2.4: Estimated Treatment Effect (Synthetic Control)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
3 Months	**-12.4%**	**-11.8%**	**-9.1%**	**-16.6%**	**-15.1%**	**-13.7%**	**-8.7%**
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
12 Months	**-18.8%**	**-17.6%**	**-16.0%**	**-19.3%**	**-20.9%**	**-19.3%**	**-18.5%**
	(.0000)	(.000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
24 Months	**-14.7%**	**-13.7%**	**-11.1%**	**-14.6%**	**-16.1%**	**-15.2%**	**-13.2%**
	(.0019)	(.0000)	(.0000)	(.0001)	(.0000)	(.0000)	(.0000)
GDP	100%	100%	none	100%	100%	100%	100%
Distance	4000km	4000km	none	4000km	4000km	4000km	4000km
Frequency	monthly	monthly	monthly	quarterly	monthly	monthly	monthly
Excluded	none	none	none	none	GDP	Distance	Lags
Controls	9.5	8.1	14.7	8.0	9.4	9.4	23.2
Correlation	40.2%	32.7%	62.7%	62.8%	38.3%	36.7%	5.9%
CV	40.8%	67.8%	36.2%	26.6%	41.2%	41.5%	50.5%
Countries	19	29	19	19	19	19	19

Excluded: Variable excluded from the matching procedure.
Controls: Average number of control countries per treatment country.
Correlation: Average value of correlation coefficient in the pre-treatment period between treatment and synthetic control.
CV: Average value of coefficient of variation.
Countries: Number of treatment countries.
p-values in parentheses.

Including the ten smaller Muslim countries introduces more noise to the analysis without changing the results much. Releasing the restrictions on the control pool significantly increases the average number of control countries from 9.5 to 14.7 and consequently the average pre-period correlation, leading to slightly lower estimates of the treatment effect. Using quarterly instead of monthly data, the reduction in noise leads to a similarly high pre-treatment fit. While the short-term estimates are slightly higher, the treatment effect after 24 months remains basically the same. To further strengthen the robustness of the results, I shut down one predictive variable (GDP, distance, previous trade levels) at a time in columns (5) - (7). While the short-term results differ slightly, the long-term effects after 24 months are close to the baseline result of -14.7%.

The results by country depicted in Table 2.17 in the appendix show strong heterogeneity between the different Muslim countries. Some larger export partners like Algeria, Egypt, Kuwait, and Saudi Arabia see a strong and persistent negative effect, while some countries even show a positive reaction. Most notably, the second and third largest Danish trading partners Turkey and UAE show no reaction to the boycott at any time horizon.

Adding up the estimates for all countries, I calculate the total disruption of trade due to the boycott to be about 0.51 billion DKK after three months, 2.86 billion DKK after twelve months, and 4.28 billion DKK after two years. The US-Dollar equivalents after taking into account fluctuations of the exchange rate are 198 million USD after three months, 444 million USD after one year, and 758 million USD after two years.

While the percentage loss for all the Muslim countries combined is sizable, this loss is marginal when compared to the total exports of Denmark. Over the period from October 2005 to September 2007, Danish exports to all its trading partners summed to 1.08 trillion DKK (185 billion USD). The implied overall disruption of trade caused by the boycott is then only 0.4% of all Danish exports during this period. While the boycott might have hit individual Danish companies hard, the effect on the total Danish export sector is negligible.

Product-level Results To assess the treatment effect by product type, I first add up the Danish exports to all the treatment countries and then separate them by product type. Table 2.2 shows the realized and counterfactual log Danish imports of each classification. Consistent with the boycott being consumer-driven, I see the largest relative decline in consumer goods with long-term reductions in this category of 27.5% and 24.8% after one and two years respectively. This suggests that the publication of the comics itself did not cause a major consumer reaction, but only after the official boycott announcement did imports from Denmark decline.

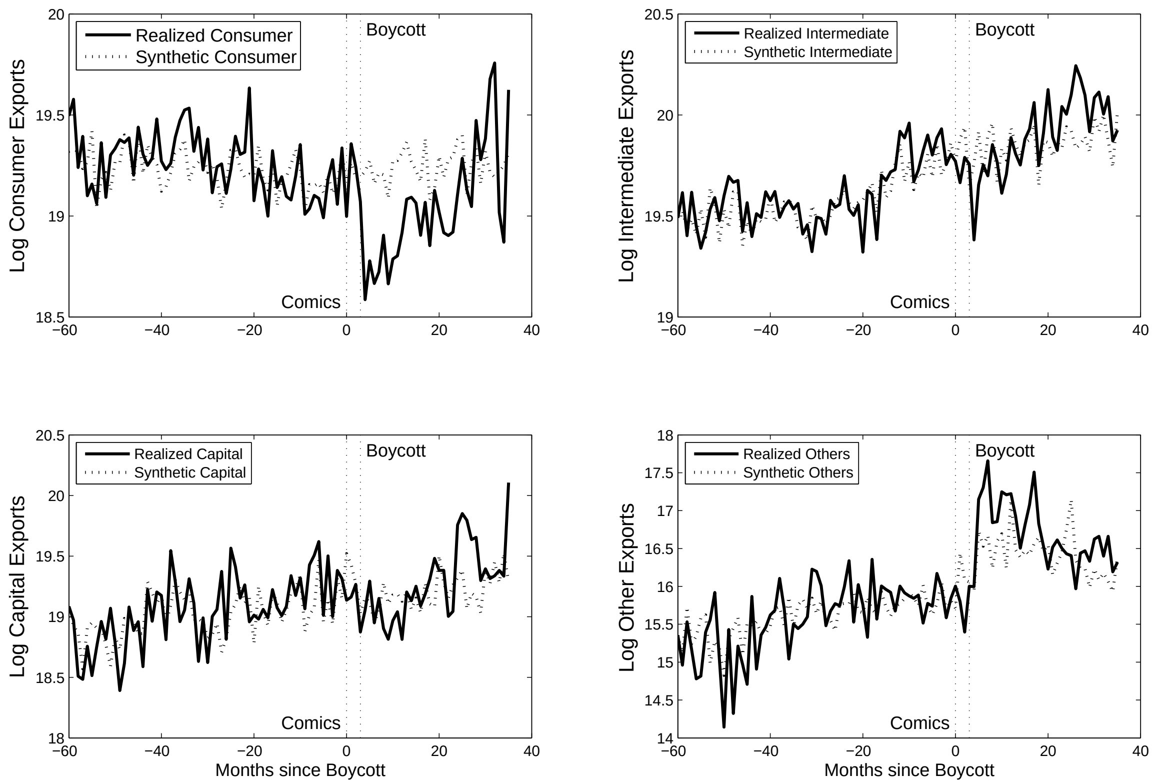

Figure 2.2: Realized and Counterfactual Trade Levels by Class

For non-consumer goods, the reaction is less strong and in many cases not statistically significant. Danish capital goods exports to the Muslim world seem to decline marginally in the short and medium run, but the large prediction errors render this result statistically insignificant. Over two years, this decline is reduced to less than 2%. For intermediate goods, we do see a significant reduction in imports from Denmark of about 9.0% and 10.9% after 3 and 12 months respectively. The reduction for these goods is reduced to 1.7% after two years. This is inconsistent with the idea of a pure consumer

Table 2.5: Treatment Effect by Product Type (Synthetic Control)

Period	**Consumer**	**Intermediate**	**Capital**
3 Month	1.5%	-9.0%***	-13.4%
	(.9894)	(.0001)	(.6646)
12 Month	-27.5%***	-10.9%***	-12.0%
	(.0006)	(.0002)	(.8539)
24 Month	-24.8%**	-1.7%	-1.7%
	(.0337)	(.1502)	(.8518)

p-values in parentheses.

boycott and could be explained by nationalistic sentiment of business owners or official trade restrictions such as complicating the processing of imports at custom offices.

Placebo Tests To check whether the results depend strongly on the parametrization of the synthetic control group, I assign placebo treatment times and estimate the trade disruption for these false boycott instances. I restrict the robustness checks to the three largest export partners Saudi Arabia, Turkey, and the United Arab Emirates. I estimate the cumulative treatment effect over six months for the 30 months preceding the publication of the comics. 25 of these placebo treatment times are not related to the boycott, but the five 6-month estimates prior to the actual treatment month will contain at least one of the actual treatment months respectively. Figure 2.8 in the appendix shows the distribution of the estimated treatment effects. Some of the placebo treatments do create negative treatment effects, but in general are of smaller magnitude and not as persistent as the estimated trade disruption of the actual treatment. For Saudi Arabia and UAE, all six-month estimates including the actual treatment month are negative and large. For Turkey, the estimate of the actual treatment is still negative, but at a much smaller scale especially compared to previous large negative and positive effects. These random fluctuations are in line with Turkey's estimated, non-significant effect of about 0%.

2.4.2 Senkaku Island Conflict

Data and Descriptive Statistics

The data for the Senkaku/Diaoyu conflict comes from the Monthly Comtrade dataset that, in addition to the standard Comtrade data, reports trade flows at monthly frequency for all Harmonized System (HS) product codes. The very fine disaggregation of the data is however offset by very limited availability of trade from January 2010 to January 2014 only. Data prior to 2010 is at the moment only available at annual frequency.

Table 2.6: Descriptive Statistics (Japanese Exports)

Country	**PR China**	**Taiwan**	**Hong Kong**
Mean (in USD)	12,800,000	4,175,000	3,502,000
Standard Deviation	1,378,000	363,400	362,700
Std Dev as Mean	10.8%	8.7%	10.4%
Minimum	9,626,000	3,090,000	2,674,000
Maximum	15,420,000	4,770,000	4,149,000
Min % Change	-30.4%	-25.2%	-29.4%
Max % Change	32.6%	22.9%	39.2%
Seasonality p-value	0.01	0.00	0.00
Share of Japanese Exports	19.2	6.2	5.2
Japanese hare of Total Imports	11.8	19.6	8.8

Statistics over the pre-boycott period January 2010 to August 2012.
Seasonality p-value is the p-value of a F-test testing for joint significance of monthly indicator variables in a linear time series regression.

Unlike the Danish-Muslim boycott where all the boycotting countries take up only a small share of total exports, the People's Republic of China is the largest export partner for Japan in the pre-boycott period from January 2000 to August 2012. China alone accounts for 19.23% of all Japanese exports. The Special Administrative Region of

Hong Kong and Taiwan[13] report separate trade statistics. Including the trade with these entities, the total percentage of exports to the Chinese-speaking world amounts to 30.8%

For the Japanese-Chinese trade data, the month-to-month fluctuations are lower but can still reach percentage changes of more than 30% in either direction. The time series is marked by a stark drop in March 2011, the effect of the devastating Tohoku earthquake and tsunami that resulted in more than 50,000 deaths. Seasonality might be an issue especially in the winter months in which trade appears to slow down and the F-test testing for the joint significance of the monthly indicator variables suggests seasonal patterns.

Difference-in-Differences Results

For the Senkaku Island Crisis case, I identify three political entities that are potentially affected by the boycott announcement: The People's Republic of China, its Special Administrative Region (SAR) Hong Kong and the Republic of China (Taiwan). All these entities claim sovereignty of the Diaoyu Islands and sent activists to them. I esimtate the model in (2.1) where $Chinese_j$ is an indicator variable that takes the value of one if the country is either China, Taiwan, or Hong Kong.

The results in Table 2.7 show that the treatment effect is negative in all specifications and is estimated to be -12.3% when including country fixed effects. As before, the coefficients on GDP and distance have the expected signs and positive results for *Chinese* indicate that Japan exports more to the treatment countries than to similar non-Chinese countries to begin with. To analyze heterogeneity in the response to the boycott, I re-estimate the model above for each Chinese country separately. The estimates for the individual countries indicate that the results are mainly driven by the PR China with a strong negative estimate of 29% whereas the Taiwan and Hong Kong show smaller

[13]For political reasons, monthly trade data for Taiwan is not officially available in the Comtrade Monthly dataset, but can be inferred from the country code 490, "Other Asia, nes".

estimates of -6.4% and -5.7%. In the opposite direction, I do not find any effect for Japan boycotting imports from China as seen in Table 2.19.

Table 2.7: Senkaku Crisis: Results

Dependent Variable: Log Imports from Japan					
Countries	All	All	PR China	Hong Kong	Taiwan
log GDP	1.068***	1.419***	1.427***	1.441***	1.443***
	(0.0516)	(0.255)	(0.256)	(0.257)	(0.256)
log Distance	-1.091***				
	(0.217)				
Post	-0.131***	-0.117***	-0.120***	-0.120***	-0.120***
	(0.0316)	(0.0313)	(0.0317)	(0.0316)	(0.0315)
Chinese	1.016				
	(0.719)				
Post × Chinese	-0.123**	-0.138**	-0.292***	-0.0647**	-0.0573
	(0.0596)	(0.0689)	(0.0347)	(0.0284)	(0.0348)
Constant	1.621	-15.95***	-16.22***	-16.54***	-16.56***
	(2.461)	(5.263)	(5.277)	(5.289)	(5.285)
Country Fixed Effects	no	yes	yes	yes	yes
Trend and Month FE	yes	yes	yes	yes	yes
N	5760	5760	5664	5664	5664
adj. R^2		0.965	0.963	0.962	0.962

Standard errors in parentheses (clustered at country level)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Synthetic Control Group Results

The nature of Japanese trade with Mainland China creates challenges with the synthetic control group method. As discussed above, Mainland China is not only Japan's largest export partner over the pre-treatment period but it is also geographically close. It is thus at the end of the distribution of both outcome as well as explaining variables and it is impossible to replicate its imports from Japan with a weighted average with the strong restrictions on the weights given in equations (2.4) and (2.5). The other treatment units, Taiwan and Hong Kong, have smaller shares of 6.2% and 5.2% respectively, but there are still only two control countries that import more from Japan (USA and Korea). I therefore relax the conditions of the weights to be in the unit interval and instead allow for arbitrary weights.

To avoid overfitting, I restrict the number of control units to countries that have a similar GDP.[14] In general, a small number of countries is able to replicate the Chinese trade patterns rather well according to the correlation coefficients in Table 2.8.

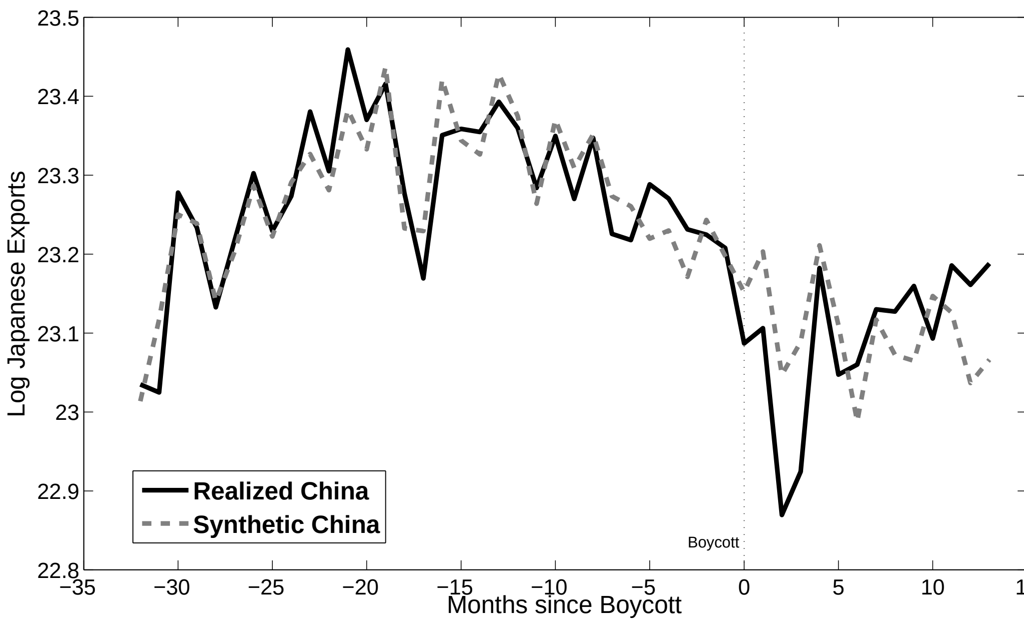

Figure 2.3: Realized and Counterfactual Japanese Exports to China

Figure 2.3 shows the realized and counterfactual exports from Japan to China on a log scale. The strong decline in realized exports for about six months after the boycott is easily visible and trade levels even fell below those that followed the devastating earthquake in 2011. Yet Chinese imports from Japan were on a downward trend and only a portion of the decline can be attributed to the boycott, as the counterfactual trade figures

[14]In specific, the replicating country's GDP should have at least 20% of GDP of the treatment country and it should not exceed it by the factor 1.8. While arguably arbitrary, this creates control pools of around 10 control countries.

implied by the synthetic control group decline as well. The effect seems to die out after half a year and then trade values catch up with the control unit. Table 2.8 shows this short term effect with a highly significant three and six month effect that is not statistically significant at 12 months anymore. The total reduction in Japanese exports within one year of the boycott amounts to 2.69% and is equivalent to 3.48 billion USD. This estimated trade disruption amounts to a share of 0.5% of total Japanese exports over the same time period. As in the case of the Muhammad Comic boycott, this is a rather small percentage of the total Japanese export economy.

For the other Chinese entities there is no significant negative effect, but Hong Kong and Taiwan experience a positive reaction to the boycott. This hints towards substitution of exports from Mainland China towards these entities, significantly reducing the overall negative impact of the Mainland boycott. One can conclude that the boycott was effective only in Mainland China and that the movement was unable to encourage Chinese people in Taiwan and Hong Kong to participate in the boycott.

The short pre-boycott period does not allow for a sensible placebo assignment of the treatment time. I instead estimate the treatment effect for the control countries that should not be affected by the boycott. I calculate the percentage losses of Japanese imports to the countries of France, Germany, Russia, India, Thailand, the UK, and the US which are all major trading partners of Japan. The results in Table 2.8 show that for the majority of the controls, the boycott did not have a significant effect on imports from Japan. Russia is the exception as it shows a significant negative impact over a 6-month period. This effect however disappears at the one-year window. The US and Thailand show a positive reaction to the Chinese boycott, suggesting that the Japanese exporters substituted their goods towards these countries.

Table 2.8: Estimated Trade Disruption

Country	Correlation	3 Months	6 Months	9 Months	12 Months
PR China	0.913	-10.39%***	-9.09%**	-3.59%*	-2.69%
		(.0021)	(.0147)	(.0854)	(.1376)
Taiwan	0.840	0.87%	6.75%***	13.41%**	10.20%*
		(.2363)	(.003)	(.0388)	(.0648)
Hong Kong SAR	0.917	5.15%***	3.00%***	9.28%***	7.03%***
		(.0030)	(.0057)	(.0007)	(.0022)
Placebo					
France	0.860	-2.48%	-5.40%	1.43%	1.09%
		(.3527)	(.1029)	(.958)	(.7423)
Germany	0.860	3.57%	1.47%	0.20%	0.15%
		(.0757)	(.2998)	(.7073)	(.6525)
Russia	0.861	-10.59%***	-7.88%**	-0.19%	-0.15%
		(.0001)	(.0199)	(.2628)	(.2901)
India	0.845	-1.02%	4.76%**	-3.42%	-2.56%
		(.6073)	(.032)	(.5129)	(.2121)
Thailand	0.883	8.78%***	11.30%***	6.94%*	5.21%
		(.0003)	(.006)	(.051)	(.1534)
UK	0.730	3.30%	7.12%	-4.25%	-3.16%
		(.7451)	(.9986)	(.3841)	(.2811)
USA	0.963	3.93%**	7.18%***	11.60%**	8.62%*
		(.039)	(.0012)	(.0163)	(.0886)

Correlation is the pre-treatment correlation coefficient between treatment and synthetic control unit.
p-values in parentheses.

Identifying Consumer Industries

Beyond dividing trade into consumer, intermediate, and capital goods the data allows me to look at a more detailed product level to trace out the effect of the boycott for six-digit HS categories.[15] I make use of publications of the Chinese boycott movement itself to identify consumer goods that are most prone to the boycott, i.e. goods that can be clearly identified by Chinese consumers as being Japanese. These publications are

[15]Product-class series for China will suffer from the same problem as the total trade values as they will be the largest and cannot be reproduced without non-negative weights. This problem is less severe for product-level HS6 codes, as China is not be the biggest export market for all of them.

two flyers that were circulated on the internet at the height of the conflict and contain pictures of Japanese brands that Chinese consumers should avoid (see Figure 2.4). I report the brand names and their industry in Table 2.22 in the appendix. Most of these firms are concentrated in a few industries, namely automotive, consumer electronics, foods, clothing, and cosmetics, while the remaining companies engage in industries as diverse as toys, cigarettes, and airline services.

Figure 2.4: Internet Flyers Calling for Boycott

I searched through the companies' internet representations and identify the their major export products. I then classify these products into the corresponding HS codes using the official description and the commercial website http://hs.e-to-china.com/ that allows searching for keywords and outputs the relevant HS code. These signature products can be subsumed into seven product codes which show a significant amount of trade between Japan and China. These codes contain highly branded goods such as passenger cars, make-up and beauty articles, foods, and a variety of consumer electronics such as cameras and video recording devices.

I estimate the impact of the boycott on these consumer goods and Table 2.9 summarizes the results. The category that sees the most drastic decline in trade is unsurprisingly 8703 which includes passenger cars. Figures 2.5 shows the realized and counterfactual log trade levels for Mainland China. Clearly visible is the massive drop in car imports and although they catch up to the control group after about nine months, Japanese car exports to China drop by a 32.3% within a single year. While the effect of the boycott is very clear for vehicles, evidence for other product codes is not obvious. The estimated percentage disruption in trade in highly-branded goods like beverages, beauty products, and cameras is large, but in absolute values the estimates of a mere 12, 14, and 4 million USD for these categories are dwarfed by the huge trade disruption of almost two billion USD for passenger cars. The other product categories seem not to be negatively affected by the boycott.

Table 2.9: Estimated One Year Trade Disruption by HS Code (Synthetic Control)

HS Code	Description	**in USD**	**in %**	**p-value**
1902	Pasta	550,166	124.6%	0.1736
22	Beverages	-12,884,956	-26.5%	0.2679
3304	Make-up	-14,596,832	-7.7%**	0.0286
8508	Electromechanical tools	1,066,317	40.4%*	0.0855
8521	Videorecording apparatus	5,677,785	32.1%	0.9600
8703	Motor cars	-1,888,019,883	-31.8%**	0.0295
9006	Still cameras	-4,630,183	-26.1%**	0.0286

Notable is the lack of a negative reaction in the area of consumer electronics, even though 34 Japanese companies in this sector were mentioned on the flyers. One explanation could be the outsourcing of Japanese firms' production to China. As it is well known, China is the "world's workshop" and produces consumer electronics under foreign brands (Feenstra and Wei, 2009). If this is true, even though sales of Japanese-branded products in China might decline, this will not show up in the trade data as these products are manufactured in China.

Event Study

Even though a boycott does not appear in the trade data, it should affect companies' profits and thus be reflected in their stock prices. Following Govella and Newland

(2010), I implement an event study for different incidents with daily market prices of 23 boycotted firms and 12 domestic control firms using the market model with the Nikkei225 index as the market proxy (MacKinley, 1997).[16]

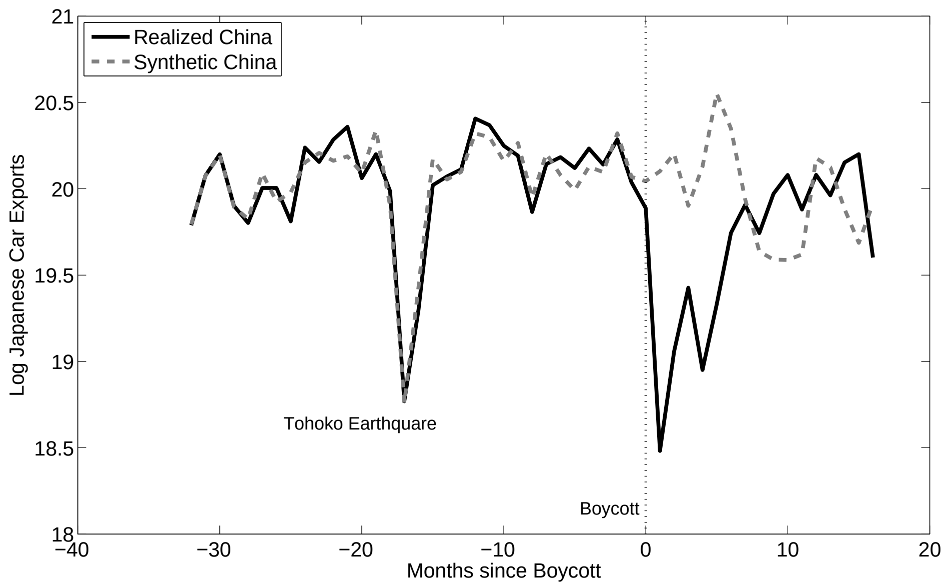

Figure 2.5: Realized and Counterfactual Japanese Exports (HS8703 Motor Cars)

The event study results indicate that none of the events mentioned above had a statistically significant effect on the Japanese firms who were mentioned in the boycott calls. Neither the start of the protests on August 20 nor the nationalization of the Senkaku Islands, which is believed to be the main cause of the boycott, do affect stock market prices significantly. This surprising results is most likely due to the fact that the treatment countries make up a big share of the market index. Plotting the cumulative abnormal returns in Figure 2.9 in the appendix for both the treatment and control country and the

[16]The events include the announcement of the nationalization of the islands by the Japanese government on 11 September 2012 that most likely caused the boycott. I also use the onset of the actual protests on September 17 as the beginning of the event period. In addition, I analyze the effect of the protests in front of the Japanese embassy in Beijing on 15 August and a protest call by Chinese internet users on 20 August. The control countries are taken from "Japan's Best Domestic Brands 2012", published by Interbrand, and have less than 10% foreign sales each. Most of these firms are engaged in non-traded services like broadband telecommunication, construction, and real estate.

market return shows that abnormal returns do not deviate much from zero until about seven trading days after the boycott announcement, when both groups start to diverge. This coincides with negative returns of the market index which tracks the treatment group pretty well. The cumulative abnormal returns however start to converge again after 20 trading days and eventually are close to zero. This suggests that the boycott did have a lagged effect not only on the stock prices of Japanese firms mentioned in the boycott flyers but also on the whole market index. This effect however is temporary and dies out after roughly one month.

2.4.3 US-French Boycott

Data for the trade boycott that resulted from the Iraq War are taken from the US Census Bureau's Foreign Trade database that provides SITC-denoted import and export data at monthly frequency. US-French trade relations are unbalanced: While France's share of US imports and exports amounts to only 2.4% and 2.8%, US shares of France's trade is much higher with 8% of imports and 8.1% of exports as measured in 2002.

For the French case, I estimate equation (2.1) for both US imports and exports, thus I now use data sources from the boycotting country. I set the treatment timing to February 2003, a period when public opinion polls detected a sharp fall in positive sentiment of France in America. I exclude Middle Eastern states from the control group to avoid contamination as there is mild evidence that some of these countries might have been involved in a boycott of US product over the Iraq War (Clerides et al, 2013). The results show that imports from France fell by almost 20% in the post-treatment period, a result that is robust to the inclusion of country fixed effects. The effect is slightly smaller for exports to France, which fell by about 16% indicating that the Iraq conflict did not merely lead to an unilateral boycott, but that it seriously hurt trade relations in both directions.

Table 2.10: Freedom Fries: Results

Dependent Variable	US Imports from France		US Exports to France	
	(1)	(2)	(3)	(4)
log GDP	0.847***	0.748***	0.866***	0.794***
	(0.043)	(0.192)	(0.040)	(0.075)
log Distance	-0.545**		-1.284***	
	(0.218)		(0.203)	
Post	-0.106***	-0.078*	-0.073	-0.060
	(0.037)	(0.040)	(0.045)	(0.045)
France	-0.314		-0.366**	
	(0.198)		(0.172)	
Post × France	-0.197***	-0.225***	-0.164***	-0.177***
	(0.053)	(0.048)	(0.027)	(0.023)
Constant	0.754	-3.142	6.802***	-3.892***
	(1.941)	(1.996)	(1.688)	(0.814)
Country Fixed Effects	No	Yes	No	Yes
Trend and Month FE	Yes	Yes	Yes	Yes
N	10219	10219	8924	8924
adj. R^2	0.708	0.952	0.744	0.964

Standard errors in parentheses (clustered at country level)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results however are not robust in the synthetic control group specification, where we see much higher drop in US exports to France than in imports from it. Figure 2.6 shows the gap in exports to France compared to the synthetic control group[17] in the year following the boycott announcement. This sums up to a 15% trade disruption equivalent to 3.020 billion USD, a share of 0.4% of all US exports during that time. The reduction in imports instead is short-term and amounts to only 4.3% within the first three months. The one-year estimate of -2.8% (or 493 million USD) is not statistically

[17]Notable is the high weight of Germany in the synthetic control group (Table2.21.) The inclusion of Germany is somewhat questionable since the country itself was a major opponent of the Iraq War. However, (Pandya and Venkatesan, 2016) point out that the boycott was primarily directed against France and not against all the European countries that did not support the US in invading Iraq. In calculations not shown I find a similar effect if Germany is excluded from the controls.

Table 2.11: Estimated Trade Disruption (Synthetic Control)

US	Imports	Exports	**Turkey**	Imports	Exports
3 Months	**-4.3%**	**-9.1%**	3 Months	14.5%	-10.5%
	(0.0399)	(0.0052)		(0.0550)	(0.0774)
12 Months	-1.7%	**-15.0%**	6 Months	11.1%	**-12.3%**
	(0.3242)	(0.0001)		(0.6285)	(0.0000)
24 Months	-2.8%	**-8.9%**			
	(0.2190)	(0.0021)			

p-values in parentheses.

significant. This suggests that the consumer boycott in the US was only a temporary shocks to imports, but that it severely harmed exports through other channels. The possibilities include a consumer boycott in France or governmental intervention. That the US administration is willing to use commerce to punish can be seen by the policy to exclude French and other nations' firms from reconstruction contracts in Iraq.[18]

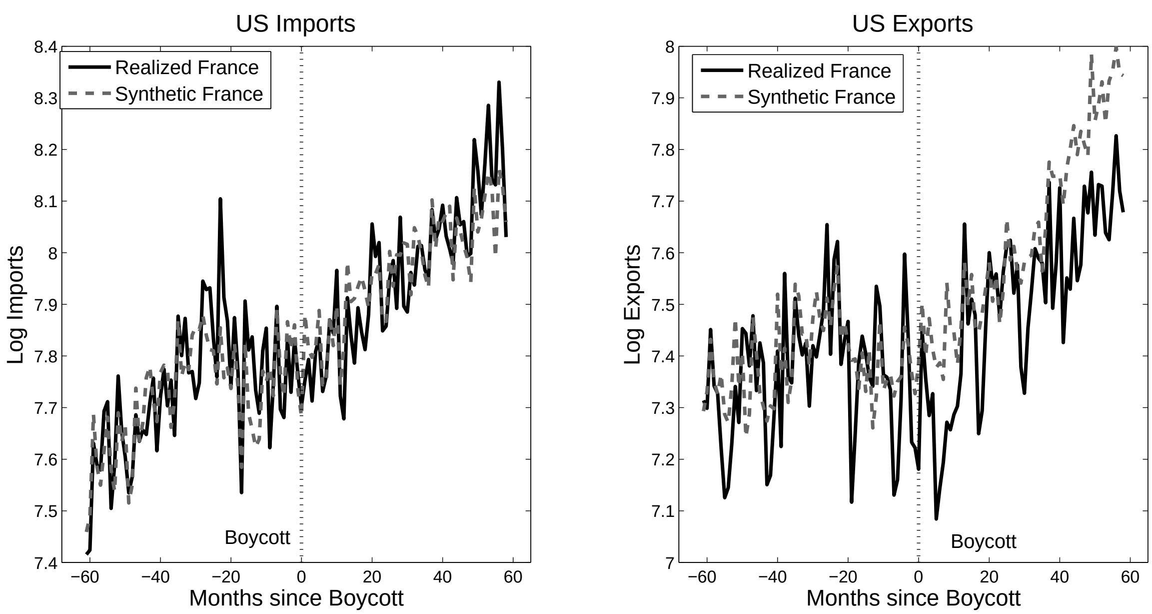

Figure 2.6: Realized and Counterfactual Trade Levels (US)

[18]http://www.nytimes.com/2003/12/10/world/a-region-inflamed-the-reconstruction-pentagon-bars-three-nations-from-iraq-bids.html.

2.4.4 Gaza

I use data for Turkish imports and exports from Comtrade Monthly with availability until January 2015. In 2013, Israel only made up a small share of Turkish trade with 1% of imports and 1.7% of exports. Turkey had a higher share with 3.3% and 3.8% of Israeli imports and exports respectively.

The difference-in-differences results for the Gaza boycott in table 2.12 show a surprising uptick in Turkish imports from Israel that is robust to the inclusion of fixed effects. The synthetic control group approach confirms the positive effect, although statistical significance holds only for the short period of the first three months after the Gaza war. Figure 2.7 shows imports well above the synthetic control group until the very last period of available data when imports drop sharply. In summary, there is no evidence of any reduction of Turkish imports from Israel due to the crisis for a horizon of at least half a year and only with future availability of data will one be able to draw a conclusion. Instead, we observe a pronounced drop in exports to Israel of around 7.5% in the post-treatment period and the synthetic control group confirms the strongly significant effect. Israeli consumers appear to be more efficient at carrying out boycotts which again is confirmed by the synthetic control group approach.

2.4.5 Substitution Effects

Applying the difference-in-differences methodology implicitly assumes that the control group is not affected by the boycott, i.e. that there is no substitution of exports from boycott to non-boycott countries. If these substitutions happen, the control group will be affected positively and the treatment effect will be upward biased in absolute values. However, the estimated treatment effect can still be interpreted as an upper bound of the true causal effect of the boycott on bilateral trade. Alternatively, we can interpret

Table 2.12: Gaza: Results

	Turkish Imports		Turkish Exports	
	(1)	(2)	(3)	(4)
log Distance	-0.198		-0.436***	
	(0.152)		(0.119)	
Post	-0.106***	-0.105***	-0.114***	-0.114***
	(0.0346)	(0.0351)	(0.0274)	(0.0276)
Israel	0.599***		1.387***	
	(0.189)		(0.129)	
Post × Israel	0.221***	0.218***	-0.0751***	-0.0751***
	(0.0394)	(0.0415)	(0.0248)	(0.0250)
Constant	17.02***	15.46***	15.76***	12.33***
	(1.487)	(0.793)	(1.165)	(0.573)
N	3717	3717	3904	3904
adj. R^2	0.025	0.879	0.178	0.926

Standard errors in parentheses (clustered at country level)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the estimate as the combination of the import-reducing direct effect of the boycott and the substitution effect, where the latter can at most be one half of the former.

Substitution can happen on both ends: The boycotted country might shift its export to other trading partners and thus alleviate the negative impact of the boycott, while the boycotting country might import the same products from other sources. Either way, this substitution is likely to be costly as importers source from the cheapest supplier and exporters supply to the country with the highest willingness to pay. The issue is loosely related to the literature on trade creation and trade diversion (Krueger 1999, Magee 2008) which distinguishes between substitution of domestic production with trade as well as substitution of trade between trading partners, and suffers from similar identification problems.

To explore the issue of substitution, I use an event study approach and estimate a post-boycott period dummy for all countries separately according to equation (2.2),

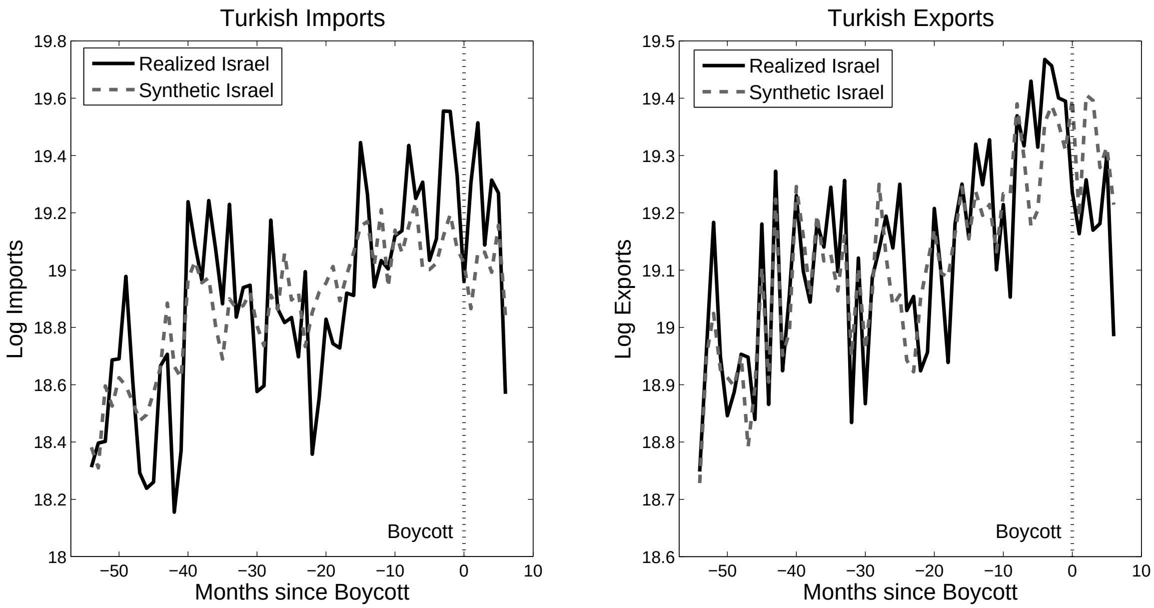

Figure 2.7: Realized and Counterfactual Trade Levels (Turkey)

controlling for GDP and time trends. The advantage is that I do not have to specify a control group, but relying on only the time series variation results in imprecise estimates. I estimate equation (2.2) for Danish exports and then correlate the post-boycott dummies with the share of Muslim population. I find that on average, a 10% higher Muslim population is associated with a 3% drop of Danish exports to this country, a result which is very close to the initial difference-in-differences findings. Furthermore, the results suggest that a hypothetical country with zero Muslim population imports a statistically insignificant 6.5% more from Denmark. While this does not completely rule out substitution effects, it shows that there is no obvious export substitution to non-boycotting countries.

$$Y_{j,t} = \alpha_j + \beta_1\, Post_{j,t} + \beta_2 \log GDP_t + \beta_3 t + month_t + \varepsilon_{j,t}. \tag{2.2}$$

To check for substitution on the import side, I run the same regression for US imports in the context of the Iraq War boycott. I find a stark (though statistically

Table 2.13: Substitution Effects

	(1) Danish Exports	(1) US Imports
Muslim Share	-0.308***	
	(0.086)	
France		-0.233
		(0.484)
Constant	0.065	-0.039
	(0.041)	(0.049)
N	122	96

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

insignificant) drop in US imports from France which again is close to the difference-in-differences estimate. The average drop in imports for all non-boycotted countries is 3.9% which is not statistically significant from zero. Thus, again there is no clear evidence for substitution away from France to other trade partners.

2.5 Conclusion

The analysis of the case studies has shown that boycotts can have a significant effect on trade relations, and that political conflict has sizable spillovers to international trade. There is no such thing as a typical reaction to a boycott, but the impact on trade is very heterogeneous. Estimates of trade reductions in the range of 30% and 40% show that some countries are very willing to carry out boycotts. However, there are also many countries that do not seem to boycott, but even increase their trade with the boycotted country. Comparing the results of the Danish-Muslim and the Japanese-Chinese boycotts, the similarities are that they both cause one-time reductions in trade, reverting to previous levels after several months. The boycotts are also impacting consumer goods much more than non-consumer goods and countries with more open political regimes tend to boycott more. Both findings are in line with the theory of boycotts being mainly consumer-driven with no or very moderate official government assistance. This suggests that agents do take the economic consequences of disrupting trade into account when choosing their conflict strategy, as boycotting consumer goods is plausibly less costly to the boycotter than stopping the import of intermediate and capital goods.

While the boycotts of Denmark and Japan appear to be one-directional with Danish and Japanese imports from the boycotting countries being mostly stable, the US-French dispute in 2003 and the Turkish-Israeli conflict in 2014 seem to have bilateral effects on both imports and exports. This is indicative of consumers in the boycotted country retaliating against the perceived aggression and the results suggests that the impact of this reverse boycott can be even larger than the initial boycott, especially in the case of Israel. The US-French boycott nevertheless confirms the transient nature of the boycotts, while the poor data availability for the very recent boycott of Turkey prevents a clear conclusion on the nature of this event.

The boycotted countries in this analysis - Denmark, Japan, the US, and Israel - happen to be well-diversified export economies. In the case of Denmark, the Muslim countries that are prone to boycotting made up only 2.6% of total exports. This a priori limits the total impact of boycotts on the boycotted country's export sector. Even though the trade disruption was large for some boycotting countries, overall exports declined by only 0.4% in the case of Denmark. Very similar numbers hold true for the US where France's export share is 2.6% and the overall trade disruption is also 0.4%. Contrary to this is the boycott of Japan, where the boycotting country, China, makes up almost one fifth of total exports. The rather low estimate for the boycott effect however causes the total trade disruption to be only 0.5%. One explanation for this small estimate is that trade from Japan to China is still dominated by intermediate and capital goods. This suggests that for exporting countries that have a diverse range of export goods and destinations, the intended punishment effect of a boycott is not likely to have a major impact; however, a boycott could be potentially more harmful to more specialized countries. The inability of the empirical approach to account for substitution effects towards non-boycotting countries is likely further strengthening this argument.

This conclusion is not necessarily true for firms within the boycotted country. While the overall disruption of exports for both Denmark and Japan is low, some firms might have suffered heavily. The event study approach on individual firms' stock market prices, however, shows that explicitly boycotted Japanese firms did experience large negative abnormal returns compared to domestic firms during the boycott, but that these relative losses are reverted within a few weeks' time. It is likely that these large companies serve widely diversified markets with different products, so that a boycott by a single country is not overly damaging.

Appendix

The Synthetic Control Group Methodology

Suppose that there are $J+1$ units in a balanced dataset with T observations which consists of one treatment unit and J potential control units. Denote the number of pre-treatment periods as T_0 and the first period with the treatment in place as T_0+1. Without loss of generality, we can define the first unit to be the treated unit and specify the units $2\ldots J+1$ to be the control units. If there are more treated units, one can simply remove them from the data and repeat the same procedure for these units.

Assume that log export values $Y_{j,t}$ of the boycotted country to its trading partners j are given by the following factor model

$$Y_{j,t} = \delta_t + \theta_t X_j + \lambda_t \mu_j + \beta_t \, boycott_{j,t} + \varepsilon_{j,t} \tag{2.3}$$

where δ_t is a time trend common to all export partners, θ_t and λ_t are $(1 \times r)$ and $(1 \times F)$ vectors of common factors, X_j and μ_j are and $(r \times 1)$ and $(F \times 1)$ vectors of factor loadings, and $\varepsilon_{j,t}$ is an iid error term with mean zero that captures idiosyncratic shocks. The parameters of interest are $\{\beta_t\}_{t=T_1}^{T}$ that measure the dynamic impact of a boycott captured by the dummy variable $boycott_{j,t}$. The difference between X_j and μ_j is that the former one is known to the econometrician and includes observable trade determinants like GDP and bilateral distance while the latter factors are unobserved and might include variables like industry composition and consumer preferences.

The ideal experiment to estimate the impact of the boycott would be to compare the outcome of the boycotting country to the outcome of a non-boycotting country that has the same factor loadings X_j and μ_j. However, this is infeasible for two reasons: Firstly, most likely no such unit exists and secondly, μ_j is unobserved. While the former problem could be resolved by a regression-based approach, the unknown factor loadings cause the regression to be necessarily misspecified. Instead, the synthetic control group method constructs the counter-factual as a weighted average of the available control countries.

Denote as $Y_{j,t}^{I}$ the counterfactual log value of the exports in case the boycott had not happened. For all the countries in the control group, I assume that $Y_{j,t}^{I} = Y_{j,t}$ for $j = 2 \dots J$ and all periods $t = 1 \dots T$. The goal is to construct the counterfactual export levels for the boycott country so that we can obtain the estimator $\widehat{\beta}_t = Y_{1,t} - Y_{1,t}^{I}$ for the treatment effect at time t.

A synthetic control group is defined by a set of J weights w_j for $j = 2 \dots J+1$ that determine a weighted average of the control units. The ideal synthetic control group would match both the factors X and μ of the treatment unit, yet this is impossible since μ is unobserved. However, Abadie et al. (2010) show that under mild regularity conditions, the synthetic control group can only match μ if it also matches a long period of pre-treatment outcome variables $Y_{j,t}$. This motivates to choose the weights to minimize both the deviation in the known characteristics X as well as the pre-treatment outcomes $Y_{j,t}$.

Assume that there exist weights w_j^* with $\sum_{j=2}^{J+1} w_j^* = 1$ and $0 < w_j^* < 1 \ \forall \ j = 2 \dots J+1$ such that

$$\sum_{j=2}^{J+1} w_j^* X_j = X_1 \tag{2.4}$$

$$\sum_{j=2}^{J+1} w_j^* Y_{j,t} = Y_{1,t} \ \forall \ t = 1 \dots T_0, \tag{2.5}$$

that is the synthetic control group resembles both the pre-treatment outcomes as well as the known explanatory variables of the treatment unit perfectly. The restrictions on w_j^* ensure that no extrapolation outside the support of the data takes place.[19]

The model in (2.3) implies that for the synthetic control group it holds that

$$\sum_{j=2}^{J+1} w_j^* Y_{j,t} = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j^* X_j + \lambda_t \sum_{j=2}^{J+1} w_j^* \mu_j + \sum_{j=2}^{J+1} w_j^* \varepsilon_{j,t}$$

and that the difference between the actual treatment unit and the synthetic control group in the pre-treatment period $t = 1 \ldots T_0$ is

$$\underbrace{Y_{1,t} - \sum_{j=2}^{J+1} w_j^* Y_{j,t}}_{=0} = \theta_t \underbrace{\left(X_1 - \sum_{j=2}^{J+1} w_j^* X_j \right)}_{=0} + \lambda_t \left(\mu_1 - \sum_{j=2}^{J+1} w_j^* \mu_j \right) + \sum_{j=2}^{J+1} w_j^* \left(\varepsilon_{1,t} - \varepsilon_{j,t} \right)$$

Rearranging, summing over all pre-treatment periods, and dividing by T_0 yields

$$\left(\mu_1 - \sum_{j=2}^{J+1} w_j^* \mu_j \right) \frac{1}{T_0} \sum_{t=1}^{T_0} \lambda_t = - \sum_{j=2}^{J+1} w_j^* \frac{1}{T_0} \sum_{t=1}^{T_0} \left(\varepsilon_{1,t} - \varepsilon_{j,t} \right)$$

Note that as the number of pre-treatment periods T_0 becomes large, the right hand side of the equation goes to zero. As long as $\frac{1}{T} \sum_{t=1}^{T} \lambda_t \neq 0$ (that is the average effect of the unobserved factors is not equal to zero over time), this implies that the difference in the unobserved characteristics goes to zero as well.

This suggests to use $\sum_{j=2}^{J+1} w_j^* Y_{j,t}$ as the counter-factual and subsequently calculate the treatment effect as

[19]This is the crucial difference to a regression-based construction of the counterfactual. Abadie et al. (2014) show that a regression-based counterfactual can be interpreted as a weighted average of the controls with weights that also sum up to one, but allow for negative values or values larger than one.

$$\widehat{\beta}_t = Y_{1,t} - \sum_{j=2}^{J+1} w_j^* Y_{j,t} \ \forall\, t > T_0$$

In practice however, one will not be able to find weights such that equations (2.4) and (2.5) hold exactly. This is the case if the characteristics of the treatment country are not in the convex hull of the characteristics of the control countries and thus cannot be replicated with the restrictions on w_j. In this case the weights are chosen such that the equations hold approximately. Formally, define $Z_j = (Y_{j,1}, Y_{j,2} \ldots Y_{j,T_0}, X_j')'$ as the column vector that stacks all export values for the pre-treatment period $1 \ldots T_0$ and the known characteristics of country j. Similarly, define the matrix $Z_C = [Z_2\ Z_3 \ldots Z_{J+1}]$ that collects these column vectors for all the potential control countries.

The $(J \times 1)$ vector W^* is then the solution to the following minimization problem

$$W^* = \underset{W}{\operatorname{argmin}} \|Z_1 - Z_C W\| = \underset{W}{\operatorname{argmin}} \sqrt{(Z_1 - Z_C W)'\ V\ (Z_1 - Z_C W)} \qquad (2.6)$$

for a given weighting matrix V. The choice of V allows to assign different importance to the explanatory variables or specific pre-treatment outcomes. In order to reduce the deviation between treatment and synthetic control group, the factors with the largest predictive power should be given the highest relative weights.

If the number of pre-treatment periods T_0 is small and the number of potential controls J is large, the characteristics of the treatment unit will be mechanically replicated by a combination of the control units. To circumvent this problem of overfitting, Abadie et al. (2014) suggest restricting the pool of potential controls to countries that have similar characteristics as the treatment country to break the spurious fit. Restricting the controls to units that are "close" in terms of the characteristics also helps to reduce potential interpolation bias.

Table 2.14: Composition of Synthetic Control Group I

Saudi Arabia		UAE		Turkey	
Antigua and Barbuda	0.024	Antigua & Barbuda	0.012	Cape Verde	0.018
Armenia	0.019	Finland	0.124	Croatia	0.09
Congo, Republic	0.011	Iceland	0.029	Luxembourg	0.003
Dominican Republic	0.039	Kenya	0.087	Moldova	0.044
Greece	0.011	Croatia	0.163	Mongolia	0.033
Haiti	0.001	Macao	0.007	Switzerland	0.01
Ireland	0.104	Malta	0.045	Slovakia	0.011
Jamaica	0.029	Mongolia	0.004	Slovenia	0.204
Cape Verde	0.032	Nepal	0.001	Sweden	0.289
Croatia	0.051	Thailand	0.142	Ukraine	0.251
Moldova	0.029	Trinidad & Tobago	0.002	Hungary	0.045
Nepal	0.025	Ukraine	0.084	Austria	0.003
Norway	0.392	Zimbabwe	0.007		
Slovakia	0.058	Austria	0.292		
Sri Lanka	0.018				
Sweden	0.061				
Trinidad & Tobago	0.01				
Ukraine	0.043				
Venezuela	0.032				
Austria	0.012				
Correlation	0.619		0.429		0.637

Correlation is the correlation coefficient between treatment and synthetic control group in the pre-treatment period.

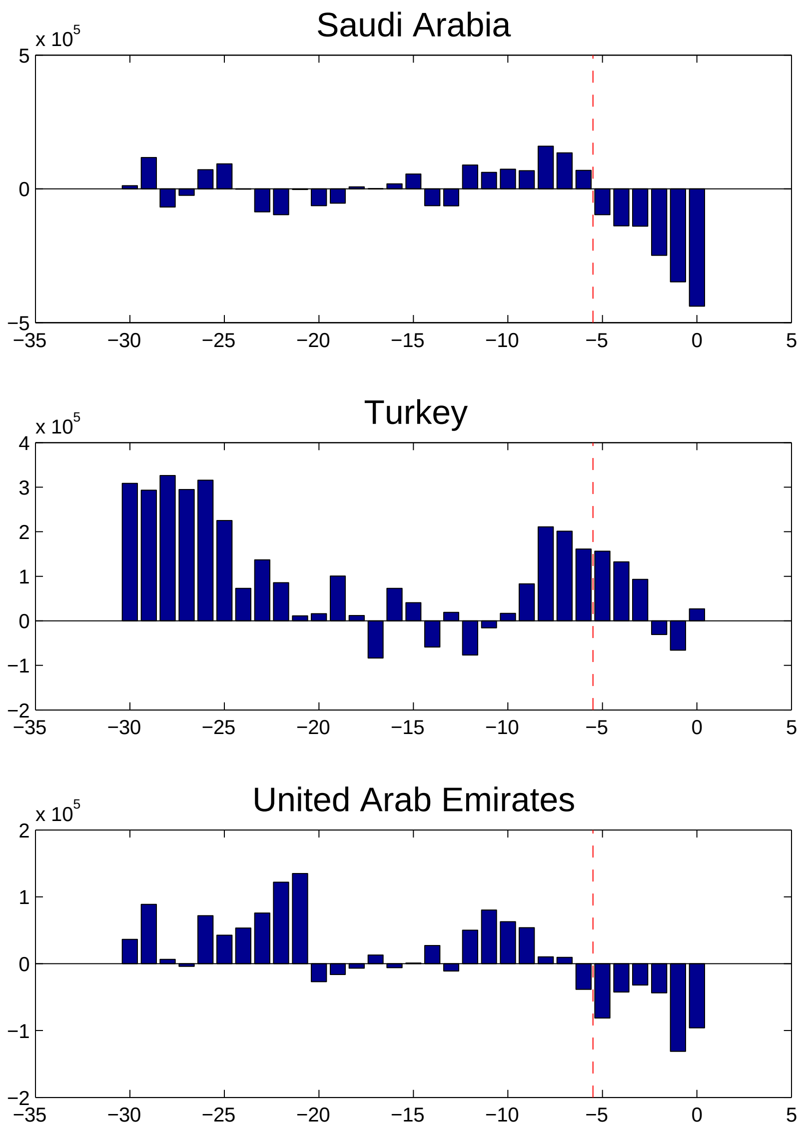

Figure 2.8: Treatment Effects of Placebo Treatment Times. Cumulative estimated trade disruption in thousand DKK over 6 months for placebo treatment assignments for the actual treatment (normalized to zero) and the 29 months prior to the actual treatment. The dashed line marks estimates for periods that include at least one actual treatment month.

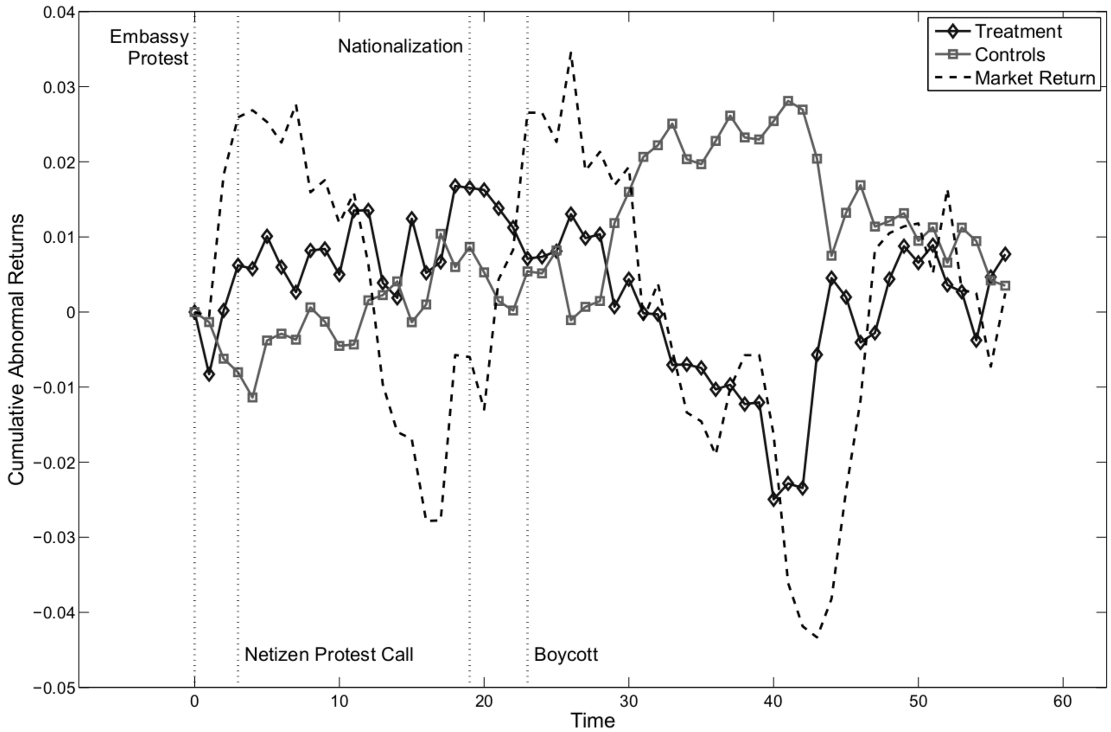

Figure 2.9: Average Cumulative Returns

2.6 Acknowledgments

I would like to thank Gordon Hanson, Marc Muendler, Eli Berman, Lawrence Broz, Gordon McCord, Xinyi Zhang, Sieuwerd Gaastra, Koji Takahashi, Matthew Goldman, and participants of the UCSD seminar series for helpful comments.

Chapter 2, in full, is a reprint of the material as it appears in: Kilian Heilmann, “Does political conflict hurt trade? Evidence from consumer boycotts”, *Journal of International Economics*, 99, 2016. The dissertation author, Kilian Tobias Heilmann, was the sole author of this paper.

Table 2.15: Treatment and Control Countries (Muhammad Boycott)

Country and Share of Muslim Population					
Treatment Countries					**N = 34**
Afghanistan	99.8%	Jordan	98.8%	Saudi Arabia	97.1%
Albania	82.1%	Kuwait	86.4%	Senegal	95.9%
Algeria	98.2%	Kyrgyzstan	88.8%	Syria	92.8%
Azerbaijan	98.4%	Libya	96.6%	Tajikistan	99.0%
Bahrain	81.2%	Maldives	98.4%	Tunesia	99.8%
Bangladesh	90.4%	Mali	92.4%	Turkey	98.6%
Djibouti	97.0%	Mauritania	99.2%	Turkmenistan	93.3%
Egypt	94.7%	Morocco	99.9%	United Arab Emirates	76.0%
Gambia	95.3%	Niger	98.3%	Uzbekistan	96.5%
Guinea	84.2%	Oman	87.7%	Yemen	99.0%
Indonesia	88.1%	Pakistan	96.4%		
Iran	99.6%	Qatar	77.5%		
Control Countries					**N = 100**
Angola	1.0%	France	7.5%	Norway	3.0%
Antigua & Barbuda	0.6%	Gabon	9.7%	Palau	0.0%
Argentina	2.5%	Germany	5.0%	Panama	0.7%
Armenia	0.0%	Greece	4.7%	P. New Guinea	0.0%
Australia	1.9%	Grenada	0.3%	Paraguay	0.0%
Austria	5.7%	Guatemala	0.0%	Peru	0.0%
Barbados	0.9%	Guyana	7.2%	Philippines	5.1%
Belarus	0.2%	Haiti	0.0%	Poland	0.1%
Belgium	6.0%	Honduras	0.1%	Portugal	0.6%
Belize	0.1%	Hong Kong	1.3%	Rwanda	1.8%
Bhutan	1.0%	Hungary	0.3%	Samoa	0.0%
Bolivia	0.0%	Iceland	0.1%	Slovakia	0.1%
Botswana	0.4%	Ireland	0.9%	Slovenia	2.4%
Brazil	0.1%	Italy	2.6%	South Africa	1.5%
Burundi	2.2%	Jamaica	0.0%	South Korea	0.2%
Cambodia	1.6%	Japan	0.1%	Spain	2.3%
Canada	2.8%	Kenya	7.0%	Sri Lanka	8.5%
Cape Verde	0.1%	Laos	0.0%	Swaziland	0.2%
Central African Rep.	8.9%	Latvia	0.1%	Sweden	4.9%
Chile	0.0%	Lesotho	0.0%	Switzerland	5.7%
China	1.8%	Lithuania	0.1%	Thailand	5.8%
Colombia	0.0%	Luxembourg	2.3%	Tonga	0.0%
Congo, Republic	1.4%	Macao	0.0%	Trinidad & Tobago	5.8%
Costa Rica	0.0%	Madagascar	1.1%	USA	0.8%
Croatia	1.3%	Malta	0.3%	Ukraine	0.9%
Czech Republic	0.0%	Marshall Islands	0.0%	United Kingdom	4.6%
Dominica	0.2%	Mexico	0.1%	Uruguay	0.0%
Dominican Republic	0.0%	Moldova	0.4%	Vanuatu	0.0%
Ecuador	0.0%	Mongolia	4.4%	Venezuela	0.3%
El Salvador	0.0%	Namibia	0.4%	Vietnam	0.2%
Equatorial Guinea	4.1%	Nepal	4.2%	Zambia	0.4%
Estonia	0.1%	Netherlands	5.5%	Zimbabwe	0.9%
Fiji	6.3%	New Zealand	0.9%		
Finland	0.8%	Nicaragua	0.0%		

Muslim Population as Percent of Total Population. Source: PEW Center

Table 2.16: Top Danish Export Partners and Import Shares

Rank	Country	Share of Danish Exports	Share of Total Imports	Freedom of Press	Polity IV Score
1	Germany	18.197%	0.81%	84.3	10
2	Sweden	13.207%	4.60%	90.3	10
3	United Kingdom	8.996%	0.57%	81.5	10
4	USA	5.998%	0.13%	83.6	10
5	Norway	5.745%	3.67%	90.5	10
26	Saudi Arabia	0.487%	0.40%	18	-10
31	Turkey	0.426%	0.18%	47.8	7
33	UAE	0.325%	0.12%	28.8	-8
35	Iran	0.290%	0.26%	18.2	-6
45	Egypt	0.164%	0.22%	33.1	-3
48	Kuwait	0.121%	0.28%	45.5	-7
51	Indonesia	0.106%	0.08%	61.9	8
54	Algeria	0.101%	0.22%	36.5	2
57	Pakistan	0.074%	0.12%	39.4	-5
58	Morocco	0.071%	0.15%	38.4	-6
61	Libya	0.059%	n/a	6.9	-7
62	Jordan	0.059%	0.18%	37.7	-2
63	Oman	0.056%	0.27%	28.8	-8
66	Yemen	0.052%	0.50%	24.8	-2
68	Bangladesh	0.042%	0.19%	36.5	6
69	Syria	0.038%	0.09%	18.9	-7
74	Qatar	0.029%	0.11%	37	-10
75	Bahrain	0.029%	0.14%	28.5	-7
76	Afghanistan	0.028%	n/a	25.7	n/a
77	Tunesia	0.027%	0.12%	19.6	-4
86	Senegal	0.012%	0.11%	55.6	8
87	Azerbaijan	0.011%	0.08%	24.5	-7
93	Albania	0.008%	0.16%	49.6	9
96	Uzbekistan	0.007%	n/a	11.3	-9
98	Maldives	0.007%	0.17%	38	n/a
104	Gambia	0.005%	3.30%	26.8	-5
114	Guinea	0.004%	0.19%	31	-1
115	Kyrgyzstan	0.004%	0.05%	74.3	3
116	Mali	0.004%	0.16%	75.5	7
118	Djibouti	0.003%	n/a	31.2	2
120	Turkmenistan	0.003%	n/a	5.6	-9
123	Niger	0.003%	0.25%	41.7	6
131	Mauritania	0.002%	0.05%	40.6	-5
133	Tajikistan	0.002%	n/a	23.2	-3
	Average Treatment	0.078%	0.291%	34.1	-2.3
	Subtotal Treatment	2.660%			

Share of Danish Exports: Percentage of total Danish exports absorbed by country in 2004.

Share of Total Imports: Percentage of Danish imports as share of total imports.

Polity IV: Score assessing regime type from -10 (full autocracy) to +10 (full democracy).

Table 2.17: Estimated Percentage Reduction in Trade by Country

Country	3m	12m	24m	Country	3m	12m	24m
Algeria	-43.5%*	-25.3%**	-22.2%*	Morocco	-44.9%	-44.5%	-41.7%
	(.0595)	(.0420)	(.0897)		(.2147)	(.1066)	(.0867)*
Bahrain	-12.7%***	-2.9%**	9.2%*	Oman	11.7%**	-29.0%***	-12.2%
	(.0009)	(.0384)	(.0927)		(.0500)	(.0010)	(.1175)
Bangladesh	10.9%	21.8%**	12.9%	Pakistan	79.4%	48.5%	31.5%
	(.2224)	(.0327)	(.4986)		(.2076)	(.1227)	(.1416)
Egypt	-41.6%***	-42.2%*	-28.9%	Qatar	41.8%	26.4%	67.9%
	(.0001)	(.0854)	(.2304)		(.5418)	(.4999)	(.8173)
UAE	-6.9%***	-2.9%*	4.1%	Saudi Arabia	-30.9%***	-37.2%***	-31.6%***
	(.0034)	(.0601)	(.8396)		(.0000)	(.0005)	(.0019)
Indonesia	31.3%***	6.6%	-9.1%	Syria	-18.9%	-20.5%***	-9.9%**
	(.0000)	(.6763)	(.1487)		(.6720)	(.0002)	(.0117)
Iran	-18.1%***	-20.5%***	-22.3%**	Tunisia	22.1%	21.8%	43.5%**
	(.0011)	(.0020)	(.0166)		(.3998)	(.7135)	(.0181)
Jordan	-5.4%*	-14.5%*	-8.5%	Turkey	5.3%	-2.5%**	0.9%
	(.0753)	(.0910)	(.1202)		(.3939)	(.0226)	(.4098)
Kuwait	-26.0%	-47.6%***	-51.3%***	Yemen	-41.4%***	-46.6%**	-37.8%**
	(.1087)	(.0003)	(.0025)		(.0000)	(.0223)	(.0385)
Libya	-2.7%**	-37.4%***	-3.6%				
	(.0133)	(.0041)	(.2102)				

p-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.18: Muhammad Comic Crisis: Trade in Services

	Service Exports		Service Imports	
	(1)	(2)	(3)	(4)
log GDP	0.864***	0.528***	0.789***	0.179
	(0.039)	(0.151)	(0.048)	(0.126)
log Distance	-0.368***		-0.437***	
	(0.090)		(0.092)	
Post	0.367***	0.464***	0.267***	0.343***
	(0.073)	(0.060)	(0.087)	(0.066)
Muslim	-0.283		-0.422*	
	(0.229)		(0.239)	
Post × Muslim	-0.506**	-0.481***	-0.542***	-0.170
	(0.220)	(0.166)	(0.197)	(0.152)
Constant	63.60***	-22.61	86.42***	-57.08**
	(21.13)	(24.73)	(22.30)	(27.70)
Country Fixed Effects	No	Yes	No	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
N	1527	1527	1512	1512
adj. R^2	0.769	0.951	0.701	0.960

Standard errors in parentheses (clustered at country level)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.19: Muhammad Comic Crisis: Danish and Japanese Imports

Dependent Variable	Imports Denmark		Imports Japan	
	(1)	(2)	(3)	(4)
log GDP	1.310***	0.438*	0.998***	0.616
	(0.064)	(0.239)	(0.088)	(0.541)
log Distance	-1.063***		0.483***	
	(0.148)		(0.132)	
Post	0.246***	0.288***	-0.308**	-0.214*
	(0.088)	(0.082)	(0.125)	(0.121)
Muslim	-2.397***			
	(0.436)			
Post × Muslim	-0.365	-0.183		
	(0.222)	(0.199)		
China			0.123	
			(0.579)	
Post × China			0.037	-0.098
			(0.121)	(0.125)
Constant	19.93***	14.76***	-15.50***	-6.455
	(1.637)	(1.581)	(4.466)	(14.61)
Country Fixed Effects	No	Yes	No	Yes
Trend and Month FE	Yes	Yes	Yes	Yes
N	13602	13602	2717	2717
adj. R^2	0.704	0.866	0.476	0.836

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.20: Composition of Synthetic Control Group II

China		Taiwan		Hong Kong	
Australia	0	Bangladesh	0.08	Azerbaijan	0
Canada	-0.01	Indonesia	-0.40	Bangladesh	0
France	-0.08	Kazakhstan	-0.04	Sri Lanka	0.06
Germany	0.17	Malaysia	0.85	Finland	0.02
Indonesia	-0.01	Pakistan	0.05	Kazakhstan	-0.12
Italy	0.04	Philippines	0.48	Malaysia	0.64
Korea	0.37	Singapore	0.14	Oman	0.12
Mexico	0.05	Vietnam	-0.18	Pakistan	0.02
Netherl.	0.21	Thailand	0.07	Philippines	0.17
Russia	0.08			Singapore	-0.08
India	0.12			Vietnam	0.19
Spain	0.14			Thailand	0.03
Turkey	0.00				
UK	0.01				
Corr.	.913	Corr.	.840	Corr.	.917

Unrestricted country weights for synthetic control group.
Corr. is the correlation coefficient between treatment and synthetic control group in the pre-treatment period.

Table 2.21: Composition of Synthetic Control Group III

US Imports		US Exports		Turkey Imports		Turkey Exports	
Belgium	0.114	Australia	0.099	Bosnia Herzegov.	0.136	Bangladesh	0.043
Germany	0.376	Belgium	0.058	Canada	0.113	Croatia	0.046
Italy	0.24	Brazil	0.067	Czech Rep	0.03	Germany	0.446
Korea, South	0.111	Germany	0.596	Greece	0.009	Hong Kong	0.016
Netherlands	0.006	Italy	0.139	India	0.419	Kyrgyzstan	0.016
Philippines	0.152	Malaysia	0.041	Rep of Korea	0.167	Norway	0.049
				Philippines	0.004	Marshall Isl	0.033
				Slovakia	0.008	Portugal	0.029
				Vietnam	0.025	Vietnam	0.06
				Turkmenistan	0.088	Spain	0.073
						Thailand	0.027
						Turkmenistan	0.045
						USA	0.102
Correlation	0.803		0.622		0.841		0.907

Table 2.22: Japanese Brands mentioned on Boycott Flyers

Automotive					
Acura	Bridgestone	Daihatsu	Denso	Hino	Honda
Infiniti	Isuzu	Kawasaki	Kubota	Lexus	Mazda
Mitsubishi	Nissan	Subaru	Suzuki	Toyota	Yokohama
Electronics					
audio technica	Brother	Canon	Capcom	Casio	Epson
Fujifilm	Fujitsu	Hitachi	JVC	Kenwood	Konami
Konica	Korg	Kyocera	Mitsumi	NEC	Nikon
Nintendo	OKI	Olympus	Panasonic	Pentax	Pioneer
Ricoh	Sansui	Sanyo	Sega	Sharp	Sony
TDK	Toshiba	Vaio	Yamaha		
Food					
Asahi	Biore	BOSS	Glico	Kewpie	Kikkoman
Kirin	Meiji	Nissin	Pocari Sweat	Suntory	UCC Coffee
UHA	Yakult	Yoshinoya			
Cosmetics					
DHC	Dongyangzhihua	Kanebo	Kao	Kose	Shiseido
Shu Uemura	SK-II	Sofina			
Clothing					
Asics	Kenzo	Uniqlo			
Other					
ANA	Bandai	Butterfly	Citizen	Daikin	INAX
Kato	Komatsu	Mild Seven	Mitsukoshi	Mizuno	Muji
Nippon	Noritz	Omron	Orient	Rinnai	Roland
Seiko	Sogo	Sumitomo	Tadano	Toray	Toto
Yasaka	Yonex	Zojirushi			

Chapter 3

Properties of Satellite Imagery for Measuring Economic Activity at Small Geographies

Abstract: We explore the potential and the limits of remotely sensed data as a proxy for economic activity at small geographic units using a commune-level dataset from Vietnam. We compare the performance of commonly used nightlight data and higher resolution Landsat imagery. Contrary to common belief, we find that nightlights perform reasonably well at predicting economic activity and expenditure once controlling for the size of geographic units. Landsat imagery has similar predictive power in the cross-section and a simple combination of the first two moments of the Landsat spectral bands can explain a large share of differences in enterprise and employment density. We however find poor prediction power of either satellite product in the time-series which severely limits the usage of remote sensing for predicting economic changes over time at small geographies.

JEL classification: E01, R1, O11

Keywords: Remote sensing; Night time lights; Landsat

3.1 Introduction

The analysis of satellite imagery is now a key methodology in economics and other applied scientific research. Coming straight from impartial satellites, remotely sensed data has the advantage of not being filtered through national data agencies that are potentially inefficient or biased. As Donaldson and Storeygard (2016) lay out its main benefits, remote sensing allows researchers to access information that would otherwise be difficult to obtain due to low state capacity, provides high spatial resolution, and a wide (if not global) geographic coverage. Since the marginal cost of collecting more data is low, repeated consistent samples are often available to researchers to learn about the world. Lastly, satellite imagery ignores administrative boundaries and can therefore be flexibly combined with other data at any geographical unit. Satellites have enabled development economists to study geographic entities that were previously inaccessible because of insufficient data coverage and in future, new commercial satellite projects with continually improving spatial and temporal coverage will only reinforce the importance of remote sensing for academic studies.

Especially the use of nighttime lights as a proxy for economic activity is important to scholars in economics and other social sciences. Night light intensity has been used to approximate economic activity as light is believed to be a normal good that is consumed more at higher incomes. Henderson et al. (2012) have found strong correlation with GDP at the national level and researchers have gone on to use nightlights at smaller geographic units. For example, Bleakley and Lin (2012) use night time light intensity to measure economic activity in their study on the economic persistence of defunct portage sites. Similarly, Storeygard (2016) studies the intercity transport costs and their impact on income of sub-Saharan African cities while Harari (2016) employs night lights to measure shapes of urban areas in India.

However, the predictive power of nighttime light at smaller geographies than the national level has recently disputed. For example, while Mellander et al. (2015) conclude that nightlight data and economic activity have a reasonably high correlation in the developed world context of Sweden, Bickenbach et al. (2016) find severe parameter instability between regions in India, Brazil, and the US, and cast doubt that the correlation between nightlights and GDP at the country level carries over to the subnational level. The reasons for this are threefold: The common night light data produced by the National Oceanic and Atmosphere Administration (NOAA) is only available at a geospatial resolution of 30 arc seconds (about 1km) that might be too coarse for small geographical units. In addition, nighttime lights have a tendency to extend into neighboring regions, the so called blooming effect, (Small et al., 2005; Abrahams et al., 2016), which complicates the identification of the actual source of lights. Lastly, nighttime light data is saturated at a certain threshold of light intensity which is often exceeded in very bright urban cores, thus making the analysis of night time light data at small geographies difficult.

As an alternative to nightlights, other remotely sensed data has been proposed. In this paper, we look at the Landsat program that has been used to describe earth over the last thirty years and its correlation with economic activity. In contrast to the night lights, Landsat satellites measure light reflectance during daytime at different wavelengths of the spectral band. Having much finer spatial (30m) and temporal resolution (16 days) than nightlights, Landsat imagery might be able to detect very different correlates of economic activity and income on the ground.

We explore the potential and limits of both nightlight and Landsat imagery at very small geographic entities in Vietnam. Using counts of enterprises and employees as well as expenditure data at the commune level, we compare the performance of nightlight and Landsat imagery as predictors of economic activity. Contrary to common belief, we find that a linear model of nightlights performs reasonably well at predicting differences in economic outcome variables in the cross-section once controlling for the size of geographic units. We find similar prediction power when using only simple Landsat indices that are designed to capture certain geographic features. We find a much stronger fit between our simple linear model and the data when using several characteristics of the distribution of all Landsat band values. Using LASSO regularization and machine-learning type cross-validation, we conclude that these relationships are stable and not just fitting statistical noise. In contrast to the cross-sectional results, we find virtually no prediction power of our remotely sensed data for changes in the time series. While we do find some rather stable coefficients for the Landsat indices, neither the nightlights nor the spectral Landsat bands can explain any useful share of the variation over time. This suggests severe limits to the use of remote sensing for forecasting economic growth at small geographies.

This paper proceeds as follows: Section 2 describes the economic and satellite data used in the study, Section 3 compares the predictive power of nightlights and Landsat in the cross-section while Section 4 does the same in the time series. Section 5 concludes and discusses the results.

3.2 Data and Descriptive Statistics

This paper uses satellite imagery to preedict commune-level economic data from Vietnam. This section introduces the data sources.

3.2.1 Economic Data

We use economic data at the production commune level in Vietnam for the years between 2008 and 2012. In Vietnam, communes are the third-level administrative subdivision after provinces and districts, and cover the whole country. Using a shapefile provided by the World Bank, we were able to match of 8,650 of the 8,916 communes with non-missing economic data to the boundaries in that shapefile. These communes have a large variety of size and cover areas between 0.05 and 1,567 square kilometers, with a median area of 14.7 square kilometers. Communes therefore represent very different entities. While in cities, communes tend to be small, there are several large communes which are sparsely populated and often sit along the western border with Laos. Even though the exact decision process of outlining commune borders is unknown to us, we suspect the formation of these units to be highly endogenous. Unfortunately, our data does not come with population data from which we could deter a mechanism such as a target population size within a commune. Figure 3.1 provides a graphical introduction into the shape of our commune units in relation to province boundaries. Because the communes are too small to sensibly plot them for the whole of the country, we provide two map inlets that display the heterogeneity of their boundaries. While in rural areas, boundaries can be large with diameters of more than ten miles, they can be very small in urban areas.

We received economic data for communes from the World Bank on enterprises and employees collected through an enterprise survey. This survey provides the total count of enterprises and counts on establishments within certain employee number bins. It also provides the total count of employees per commune as well as a disaggregation by industry. We can therefore approximate the economic activity of each commune by the number of workers and the number of enterprises employing them. Additionally, the dataset contains net revenue and profits in the aggregate and by industry. The data is limited by the fact that only formal employment is recorded and that the sampling scheme led to fully surveying enterprises with more than 10 employees only, while a random sample is used for smaller firms. In a country with a potentially very high share of the informal economy (Cling et al., 2011), these numbers can severely underestimate the actual employment numbers.

We further have access to real per-capita expenditure data (measured in 2010 US dollars) from a household survey. Again, there are issues with the survey design. While the household survey design makes sure that the sample is representative at the province level, it might not be so at the commune level. The household survey has a smaller sample size than the enterprise data and we were able to match 3,039 communes to the shapefile provided.

Below in 3.2, we plot the distribution of three variables that will serve as our main measures of economic activity together with commune size on a log scale. While the commune size, the number of employees, and per-capita expenditure approximately follow a log-normal distribution, the number of enterprises follows is highly skewed and has the majority of its mass in the left tail.

3.2.2 Satellite Imagery

We use remote sensing imagery from the DMSP-OLS night light dataset and the Landsat program. This section introduces the Landsat program and describes our processing of the data.

Landsat data The Landsat program is organized by the United States Geological Survey (USGS) and consists of several satellites that capture global imagery at frequent intervals. Unlike satellites that measure nightlights, the Landsat satellites records daytime reflection of light from the sun. Depending on their structure, objects on earth reflect different portions of the electromagnetic spectrum. Landsat data is organized into images and covers all of Earth except for the polar regions.

The Landsat program dates back to the early 1970s and has launched several satellites since then. The newest satellites (Landsat 7 and Landsat 8) provide a spatial resolution of 30x30 meters, and their output can therefore be classified as “medium-resolution” imagery. While Landsat imagery lacks the detail internet users are used to from commercial mapping data such as Google Maps, it is fine enough to recognize structures and other land use. Besides visible light that can be recognized by the human eye, the satellites also capture non-visible light such as infrared and thermal heat.

For this paper, we use Landsat 7 which launched in 1999 and is still operating, whereas Landsat 8 was only launched in 2013. Landsat 7 records eight different spectral bands and has a temporal resolution (the time until the satellite revisits a certain position on earth) of 16 days. Table 3.1 provides an overview of the recorded bands and their resolution. We work with annual simple composite images of Vietnam. The simple composite algorithm corrects for the disturbance by cloud-coverage that is perennial in tropical regions and stitches together cloud-free images collected at different times. This comes at the loss of temporal accuracy as the data is not sourced from a single snapshot in time.[1] We extract several characteristics of the band distribution within each commune. In specific, we calculate the mean, median, and standard deviation, as well as the 10th and 90th percentile of all bands listed in Table 3.1.

While widely used in environmental sciences, the vast Landsat dataset has only recently been introduced into the social sciences, mostly in urban studies. Several papers have examined the predictive power of Landsat imagery for population counts and urban boundaries, e.g. Goldblatt et al. (2016). There is very little usage of Landsat in Economics besides the fact that Landsat imagery is available easily for free.

Nightlights For nighttime lights, we use the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) product. In contrast to daytime Landsat imagery, these satellites measure light emitted from the globe at night. The dataset has a spatial resolution of about 1km. Each pixel is coded with an integer value between zero (no light) and 63 (maximum light). This top coding due to saturation is an issue in bright city areas that easily hit the maximum sensitivity of the satellite sensor. While nightlights are in principle measured every 24 hours, the raw data requires significant ex-post processing and datasets are typically released every year.

In this paper, we use the stable lights product that removes unstable light sources such as moonlight, clouds, fires, and gas flares that create large outliers in the data (Baugh et al., 2010). We extract annual images and calculate the mean light night intensity as well as other characteristics of the light distribution within a commune. In Figure 3.4, we correlate this measure with the individual Landsat band means during daylight. The nightlights show the highest correlation with visible red, green, and blue light with

[1] In specific, we use a standard Top-of-Atmosphere (TOA) calibration on all USGS Landsat 7 Raw Scenes in one year with less than 10% cloud coverage and then use the median of each pixel that satisfies this restriction.

correlation coefficients of around 0.5, while it is less correlated with the non-visible spectrum.

Data Processing For both satellite databases, we use Google Earth Engine (GEE) to access and manage the data. GEE is a cloud-based computational platform that allows to integrate data storage and data manipulation within a single framework. Coming with a JavaScript-based library of geospatial tools similar to the environment of ArcGIS but not being restricted to a single computer system, it allows to easily scale the geospatial analysis across space and time.

3.3 Predictive Power of Satellite Imagery in the Cross-Section

3.3.1 Econometric Setup

We start by exploring the predictive power of satellite imagery in the cross-section. At first, we focus on one single year. Since commune boundaries change, our shapefile does not perfectly match the economic data provided for all years. We were able to have the highest matching rate in the year 2012 and therefore proceed with estimating cross-sectional regressions.

In a first step, we use simple linear ordinary least squares (OLS) to predict economic variables using the remotely sensed satellite data. We run a log-log specification in the form of

$$\log y_i = \alpha + \beta \log satellitedata_i + \varepsilon_i \tag{3.1}$$

where log is the natural logarithm, y_i denotes the economic outcome of commune i and $satellitedata$ are statistics from the remotely sensed satellite products. The coefficient β can therefore be interpreted as the elasticity of the economic variable with respect to the satellite measure. Besides the elasticity, we are also interested in the R-squared measure as an indicator of the the predictive power of remotely sensed data for economic outcomes.

In later specifications, we control for the area of the communes which turns out to be a important predictor for economic outcomes. Commune area is not random and accounting for the endogenous choice of commune boundaries is important. Figure 3.5 depicts the strong negative relationship between the number of enterprises and employees, and commune size. On average, small communes have higher economic activity with estimated elasticities of -.52 (enterprises) and -.65 (employment) respectively. The same negative relationship hold for per capita expenditure with an albeit smaller elasticity of -.22.

3.3.2 Night Lights

We first correlate our measures of economic activity (employment, number of enterprises) and expenditure with the night light data from DMSP-OLS. We proceed by using the sum of lights (SOL) approach that is commonly applied for indicating light intensity. The sum of lights approach counts up all nightlight sensor values (coded from 0-63) of pixels that fall within the area of a commune. We then divide this value by the area of the commune to arrive at the light intensity of a commune. Similarly, we divide the number of enterprises and employment by the area and take the natural logarithm.

Table 3.3 reports the results from the simple linear regression model in (1) for the above mentioned outcome variables and real per capita expenditure. In column (1), the estimated elasticity of enterprise density with respect to nightlight density is 1.18, indicating a strong relationship between economic activity and nightlight even at small geographies. The R-squared measure of 0.541 indicates high predictive power of the nightlights in the cross-section. Similarly, the elasticity with respect to employment is 1.36 and again explains a considerable share of the between variation in employment ($R^2 = 0.517$). In contrast, the elasticity of per capita expenditure with nightlight density is much smaller at 0.181, indicating that nightlights are very responsive to changes in production but less so with consumption. At least in the development country setting of Vietnam, nightlights could be much more related to production and the justification of light as normal good is questionable.

When controlling for potentially endogenous commune sizes, we confirm the strong positive relationship between nightlights and economic activity. In column (2), we include the log area (in square kilometers) as a regressor. The coefficient is strongly negative and indicates a elasticity of enterprise density of close to -1. The coefficient on log nightlight density decreases to 0.598 but is still strongly significant. Similarly, in column (3) the employment density elasticity drops to 0.806 while again signaling an approximate unit elasticity with respect to commune size. The considerable increase in the R-squared measure in both regressions indicates the predictive power of communize size independently of night light density and confirms our notion that commune boundaries are endogenous. For expenditure, the elasticity is reduced even further from 0.181 to 0.126.

3.3.3 Landsat Satellite Imagery

We now focus on the question whether Landsat band values have similar predictive power for ground truth data of economic activity.

Landsat Spectral Indices Since Landsat bands measure the reflectance of light of a certain wavelength and are difficult to interpret numerically, we first proceed by using spectral indices derived from these bands that measure certain outcomes and can be more easily interpreted. We use the normalized difference vegetation index (NDVI), the normalized difference water index (NDWI) and the normalized difference built-up index (NDBI) that are non-linear combinations of two Landsat bands each. These indices are defined as

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad NDWI = \frac{Green-NIR}{Green+NIR} \quad NDBI = \frac{NIR-SWIR}{NIR+SWIR}$$

where *Red* and *Green* correspond to the Landsat 7 bands 3 (red light) and 2 (green light) respectively, NIR is the Near Infrared measurement of band 4, and SWIR is the value of the Shortwave Infrared band 5. These indices are designed to measure vegetation, water coverage, and urban areas by capturing typical spectral signatures of these features. Their values ranges between -1 and 1, and a higher index value corresponds to more vegetation, water, and built-up area presence respectively. For example, the NDVI measure is designed to capture live green vegetation on the ground. For photosynthesis, live plants absorb visible light (low wavelengths) but reflect infrared light (higher wavelengths), thus a higher value of $NIR - Red$ indicates presence of vegetation. The NDVI has been used in the environmental sciences to distinguish vegetation from other land uses. Similarly, NDWI captures the unique reflection signature of water bodies while NDBI is designed to detect urban areas (Zha et al., 2003). Figure 3.3 plots the distribution of these three indices together with the log sum of light measure from the nighttime light data. All measures follow a distribution which can be somewhat approximated with a normal distribution, although there are thick tails in the right (NDBI, NDWI) and the left (nightlights, NDVI) end of the distributions.

We proceed by estimating equation (1) with the three Landsat indices above. Since these indices are already normalized and well bounded, we do not apply the natural logarithm. Table 3.4 summarizes the results. In column (1), all indices are statistically significant in explaining differences in the log density of enterprises between communes. While a higher measure of built-up areas is positively correlated, the coefficients on NDVI and NDWI are negative and indicate that the presence of vegetation and water bodies predicts lower enterprise density. This is not surprising as high NDBI but low NDVI and NDWI values indicate the presence of cities and as self-employed farmers are not covered by the enterprise survey.

The coefficients in the regressions on employment density and expenditure are largely similar and have the same sign, indicating that employment and consumption is higher in more urban places. The Landsat indices predict 49.3% and 41.8% of the squared variation in enterprise and employment density respectively and thus slightly less than the pure nightlight approach (54.1% and 51.7%). The difference in predictive power of both remote sensing approaches is less pronounced for expenditure ($R^2 = 0.291$ in the nighlights regression versus $R^2 = 0.284$ in the Landsat regression).

When controlling for commune size, the coefficient on NDBI remains largely stable while the coefficient on NDVI drops drastically. Similarly, the NDWI coefficient drops and is now statistically insignificant. This indicates high correlation between commune area and vegetation and water presence, highlighting the endogenous nature of commune boundaries.

Landsat Spectral Bands We then proceed with exploring the predictive power of the full set of spectral Landsat bands. We regress a “kitchen sink” specification that includes all Landsat bands as explanatory variables and then compare the fit with the results from the regressions with the simple indices only. We use band averages as well as other characteristics (such as the standard deviation) of the band distributions as predictors. While the Landsat indices such as NDBI provide an easy to interpret measure of ground characteristics that might be correlated with economic outcomes, the researcher who is interested in remote sensing economic activity needs to form a prior on which spectral signatures to use as predictors. In contrast, using all spectral Landsat bands as predictive variables allows for an agnostic and flexible (yet difficult to interpret) way of recovering the statistical relationship between measures of economic activity and remotely sensed data.

However, the large number of potential variables poses the risk of overfitting the data and of mistaking noise as a valuable signal. To alleviate this issue, we perform two kinds of analyses to guide our econometric approach. We first use LASSO techniques (Belloni et al., 2014) to restrict our variable space and secondly, we apply cross-validation methods from the machine learning literature to divide our sample into a training and testing dataset to judge out-of-sample validity of our estimated parameters. In specific, we randomly attribute 70% of our sample to a training dataset, estimate the coefficients and then examine the out-of-sample fit for the remaining 30% testing dataset. We repeat this exercise 500 times and calculate the cross-validated R-squared as the average out-of-sample R-squared of these 500 draws.

Table 3.5 compares the regression results for the Landsat indices with the ones of more flexible models. We first only use the mean band values of each commune and then augment them with the standard deviation. We then account for skewness of the distribution by including medians and the interquartile range (25th and 75th percentile). Comparing specification (1) and (2) for enterprise density, using only the band means as predictors yields already a better fit than NDBI, NDVI, and NDWI only, indicating that other band values than those to calculate the indices provide valuable information. The LASSO regression picks up all of the band means and the cross-validated R-squared is reduced only marginally compared to the full in-sample R-squared. Adding the standard deviations as predictors further increases both in-sample and out-of-sample

R-squared measures, while adding further information about the distribution yields only very marginal improvements in the fit.

For employment density and per capita expenditure, the results are similar. Using band means and standard deviations has higher prediction power than the Landsat indices, while adding further distribution characteristics beyond the first two moments yields only small improvements in R-squared. Throughout the analysis, the LASSO technique tends to select all potential regressors as informative. This is the case despite the very high correlation between certain Landsat bands. Surprisingly, the cross-validation exercise yields very little evidence of overfitting and only in the expenditure regression are there small differences between in-sample and out-of-sample fit. This suggests that in the cross-section, the parameters of the Landsat bands are highly stable throughout the whole country for predicting economic activity.

3.3.4 Combining Nightlights and Landsat

We now explore combinations of nighttime light data and Landsat imagery as a predictor for our economic outcome variables. The two satellites might pick up very different correlates of economic activity and combining them could increase the predictive power. Nighttime lights are positively correlated with NDBI ($\rho = 0.30$) and NDWI ($\rho = 0.49$), but negatively with NDVI ($\rho = -0.52$). Table 3.6 summarizes the regression results for several combinations of Landsat indices and nighttime light density. Independent of the Landsat indices used, the nighttime light density remains a strong predictor of economic activity. Even after controlling for area, NDBI, NDVI, and NDWI, the elasticity of enterprise density, employment density, and per capita expenditure is still 0.48, 0.70, and 0.1 respectively. R-squared measures increase slightly when incorporating both satellite datasets. This indicates that nightlights and Landsat satellites are picking up features that are independently correlated with economic activity and can thus used in combination to improve prediction accuracy.

3.4 Evaluating Changes in the Time Series

We next examine the predictive power of satellite imagery in the time series and look at changes in economic activity over time.

3.4.1 Econometric setup

We continue to use a simple linear prediction model and estimate differences in the economic outcome variables on changes in the remotely sensed satellite data. For that we difference our estimating equation in (1) and arrive at

$$\Delta \log y_{i,t} = \alpha + \beta \, \Delta \log satellitedata_{i,t} + \nu_{i,t} \tag{3.2}$$

where the operator Δ denotes the change between t and $t-1$ and $\nu_i, t = \varepsilon_t - \varepsilon_{t-1}$ is the potentially serially correlated error term. Our key objects of interest are whether the parameter β is stable between the cross-sectional regressions in Section 3 and the analysis in the time series and whether the reduced variation in the satellite measure has enough predictive power for economic activity at the commune level.

We first deal with the potential serial correlation issue by estimating equation (2) in long differences. We observe the number of enterprises and employment for the years 2004 to 2012, and expenditure for the years 2004, 2008, 2010, 2012, and 2014. Although our economic activity data goes back to 2004, inconsistencies in the commune boundaries make us believe that our data quality is highest from 2009 onward. We therefore begin with regressing the medium-run changes of enterprise and employment density from

2009 to 2012, and the change in expenditure between 2008 and 2012 on the respective changes in the nightlight and Landsat data.

3.4.2 Empirical Results

Nightlights and Landsat Indices Table 3.7 summarizes the results for the long difference regression in equation (2) for both the nighttime lights and the simple Landsat indices as well as a combination of both measures. In contrast to the cross-section regressions, we find very low predictive power and parameter instability for predicting changes in the economic outcome variables of employment and enterprise density. For the percentage change in employment density (column 1), the estimated elasticity is -0.017 which is statistically significant at the 5% level. This finding differs vastly from the cross-sectional regression where the elasticity was positive and close to 1. We find a similarly small but imprecisely estimated coefficient for employment in column (4). The R-squared measure in both regressions is very close to zero and we conclude that in our context of communes in Vietnam, nighttime lights have very little predictive power for changes in employment or enterprise density.

We now turn to the simple Landsat indices and assess the changes in the mean NDBI, NDVI, and NDWI as predictors for changes in the economic outcome variables in columns (2) and (5). Again, we find very little predictive power with R-squared measures of virtually zero. We do find the same signs on the coefficients as in the cross-section as increases in NDBI and decreases in NDVI and NDWI are correlated with increases in enterprise and employment density, although some of the coefficients are imprecisely measured. While NDBI changes are highly significant for employment in the column (5), this relationship is less strong and statistically insignificant for enterprises in (2).

When combining both measures in the same regression, we again find little predictive power in the time series. The estimated parameters change only slightly, reflecting the fact that changes in nightlights and changes in the Landsat indices are only minimally correlated (maximum correlation between ΔNTL and $\Delta NDVI$ is 0.097). We conclude that neither changes in nightlights or Landsat indices individually, nor a combination of both measures help in predicting changes of economic outcome variables on the ground between 2009 and 2012 in a useful way. Especially the use of nightlights at very small geographies as an indicator of economic activity in the time series appears questionable. We do however confirm our cross-sectional results that the NDBI and NDVI are correlates of economic activity, even though they only explain a marginal share of its variation over time.

We find very similar results for expenditure in the long differences regression between 2008 and 2012. Neither of the coefficients of either satellite measure is significant in explaining variation in per capita expenditure over time and predictive power is virtually zero. This again lets us conclude that in the developing country context of Vietnam, satellite imagery is a poorer correlate of consumption than it is of production.

Spectral Landsat Bands We next turn to the spectral Landsat bands and examine their predictive power. Table 3.9 summarizes our findings. In this exercise, we restrict our attention to the first two band moments that showed the highest predictive power in the cross-section. In this table, we only report the estimates on coefficients that were significant at the 5% level in any of the specifications. Column (2) reports the results of a linear regression of enterprise density on changes in the raw band moments only. While the R-squared increases by a factor of 5, it is still very close to zero. Adding the changes in the standard deviations does yield only a tiny increase in predictive power and none of the coefficients are significant. Adding these predictors does not change the estimated parameters on the means and it appears that the first three Landsat bands (red, blue, green) do carry some information.

Looking at employment density, the results are very similar and again the first three bands are significant and have the same sign as in the enterprise regression. Again adding changes in the means and standard deviations of the spectral Landsat bands does not increase the predictive power in any meaningful way. Consistently, the LASSO technique chooses less satellite bands as useful predictors. We therefore conclude that even when using the full power of spectral Landsat bands, linear prediction models are unable to capture changes in the economic outcome variables. Figure 3.6 summarizes the differences between cross-section and time-series by plotting the actual versus the predicted values for all three economic outcome variables in levels and in long differences.

3.5 Discussion and Conclusion

This paper has introduced the Landsat image data and evaluated its usefulness for prediction of economic activity that is difficult to measure through surveys. Unlike the commonly used nightlight data, Landsat imagery comes at a much higher spatial resolution, is measured more frequently, and provides data on a multitude of spectral bands during daytime. While nighttime lights from the DMSP-OLS measures human activity in form of light consumption, Landsat imagery captures a bigger picture of earth shaped by both nature and man. Thus, there is potential for detecting a variety of economically relevant features on the ground that correlate with socioeconomic data.

Using small-scale geographic data, our analysis shows that Landsat imagery can act as a strong predictor for enterprise and employment density in the cross-section in the context of Vietnamese communes. Simple combinations of Landsat bands that were developed to detect urban areas and vegetation already have reasonable predictive power, while flexible combinations of band means and standard deviations can explain a large share of the differences in firm counts and employment between communes.

Cross-validation exercises and LASSO regressions indicate strong parameter stability and do not suggest overfitting due to the large number of potential parameters. The relatively poor prediction power of satellite imagery for expenditure data suggest that, at least in the context of developing countries, remote sensing is mainly taking up ground features that correlate with production rather than consumption.

Comparing the Landsat data to the often-used nighttime light approach, we find that night lights do reasonably well at predicting economic outcomes even at small geographies and that both satellite sources have similar predictive power in the cross-section. Surprisingly, given their very different nature of data collection, both satellite measures are highly correlated and combining both does not vastly improve the precision of predicting economic activity.

When looking at changes over time, we find that neither nightlights nor Landsat bands have much prediction power. Although some coefficients on certain Landsat bands are significant in our regression, they can only explain a tiny share of the variation over time and the R-squared of our linear models is virtually zero. This casts doubt of the usefulness of remote sensing for predicting economic growth at small entities and suggests that the strength of remote sensing lies in capturing stable features on the ground that correlate with economic activity.

A general problem that remote sensing of economic activity faces is the endogenous formation of geographic units and in our setting, commune size in itself is a strong negative predictor of economic activity. This makes remote sensing applications very context-specific, and parameters estimated on one dataset might not act as a good model for prediction in another context.

3.6 Acknowledgments

Chapter 3 is coauthored with Ran Goldblatt, Gordon Hanson, Amit Khandelwal, and Yonatan Vaizman, and in part, is currently being prepared for submission for publication of the material. I thank my coauthors for the permission to use this chapter in my dissertation. Richard Ferrera, Travis Holtby, and Johannes Verkamp provided excellent research assistance The dissertation author, Kilian Tobias Heilmann, was the primary author of this paper.

Table 3.1: Bands of Landsat 7

Bands	**Wavelength**	**Resolution**
Band 1 - Blue	0.45-0.52μm	30m
Band 2 - Green	0.52-0.60μm	30m
Band 3 - Red	0.63-0.69μm	30m
Band 4 - Near Infrared (NIR)	0.77-0.90μm	30m
Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75μm	30m
Band 6 - Thermal	10.40-12.50μm	60m
Band 7 - Shortwave Infrared (SWIR) 2	2.09-2.35μm	30m
Band 8 - Panchromatic	0.52-0.90μm	15m

Note: Landsat records data for Band 6 at two different sensitivities. Since they are highly correlated as shown in Figure 3.4, we proceed with only the less sensitive setting.
Source: USGS Landsat

Table 3.2: Descriptive Statistics

	Observations	Mean	Std Dev	Min	Max
Total number of enterprises	8493	39.8	134.7	1	4722
Total number of employees	8493	1242.0	4956.4	1	134574
Per capita expenditure (2010 USD)	2990	989.6	735.9	106.8	9024.8
Sum of Lights (SOL)	8757	13473	22551	0	573895
Average NDVI	8757	.405	.147	-.172	.710
Average NDWI	8757	-.315	.134	-.585	.293
Average NDBI	8757	-.285	.089	-.653	.086
Area in (sqkm)	8757	28.6	56.2	.05	1567.11

Table 3.3: Nighttime Lights. Year 2012

	(1) Enterprises	(2) Enterprises	(3) Employment	(4) Employment	(5) Expenditure	(6) Expenditure
Log(Nightlight/area)	1.180***	0.589***	1.361***	0.806***	0.181***	0.126***
	(0.0239)	(0.0194)	(0.0266)	(0.0254)	(0.00761)	(0.00844)
Log Area		-1.042***		-0.978***		-0.0994***
		(0.0173)		(0.0224)		(0.00877)
Observations	7643	7643	7643	7643	2629	2629
Adjusted R^2	0.541	0.726	0.517	0.635	0.292	0.328

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.4: Landsat Indices. Year 2012

	(1) Enterprises	(2) Enterprises	(3) Employment	(4) Employment	(5) Expenditure	(6) Expenditure
Average NDBI	5.720***	5.609***	6.455***	6.335***	0.423*	0.410**
	(0.434)	(0.289)	(0.490)	(0.357)	(0.166)	(0.155)
Average NDVI	-23.19***	-4.268***	-25.82***	-5.343***	-5.376***	-3.010***
	(1.902)	(1.101)	(2.191)	(1.459)	(0.665)	(0.630)
Average NDWI	-14.70***	-1.437	-16.73***	-2.368	-3.936***	-2.232***
	(2.045)	(1.143)	(2.353)	(1.530)	(0.701)	(0.660)
Log Area		-1.216***		-1.316***		-0.143***
		(0.0137)		(0.0194)		(0.00830)
Observations	8493	8493	8493	8493	2990	2990
Adjusted R^2	0.493	0.744	0.418	0.626	0.284	0.352

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.5: Landsat Kitchen Sink Regressions

Dependent Variable:	**Log(Enterprises/Area)**				
	(1)	(2)	(3)	(4)	(5)
Variables included	Indices	Means	+Std. Dev.	+Median	+Interquartile Range
Average NDBI	5.720***				
	(-0.298)				
Average NDVI	-23.19***				
	(-1.258)				
Average NDWI	-14.70***				
	(-1.355)				
Observations	8493	8493	8493	8493	8493
Adjusted R-squared	0.493	0.637	0.694	0.706	0.711
Crossvalidated R-squared	0.492	0.635	0.691	0.702	0.706
Variables selected by LASSO	3/3	9/9	18/18	25/27	45/45
Dependent Variable:	**Log(Employment/Area)**				
	(1)	(2)	(3)	(4)	(5)
Variables included	Indices	Means	+Std. Dev.	+Median	+Interquartile Range
Average NDBI	6.455***				
	(-0.38)				
Average NDVI	-25.82***				
	(-1.602)				
Average NDWI	-16.73***				
	(-1.726)				
Observations	8493	8493	8493	8493	8493
Adjusted R-squared	0.418	0.543	0.608	0.618	0.625
Crossvalidated R-squared	0.417	0.541	0.605	0.614	0.619
Variables selected by LASSO	3/3	9/9	18/18	24/27	45/45
Dependent Variable:	**Log(Per Capita Expenditure)**				
	(1)	(2)	(3)	(4)	(5)
Variables included	Indices	Means	+Std. Dev.	+Median	+Interquartile Range
Average NDBI	0.423**				
	(-0.137)				
Average NDVI	-5.376***				
	(-0.561)				
Average NDWI	-3.936***				
	(-0.602)				
Observations	2990	2990	2990	2990	2990
Adjusted R-squared	0.284	0.356	0.397	0.404	0.413
Crossvalidated R-squared	0.284	0.320	0.378	0.388	0.396
Variables selected by LASSO	3/3	9/9	16/18	26/27	37/45

Crossvalidated R-squared is the average R-squared from 500 replications of splitting the data into a 70% training to estimate the model parameters and a 30% testing dataset to calculate the out-of-sample R-squared.

Table 3.6: Landsat and Nightlight Combination. Year 2012

	(1) Enterprises	(2) Enterprises	(3) Employment	(4) Employment	(5) Expenditure	(6) Expenditure
Log Area	-1.216***	-0.900***	-1.316***	-0.849***	-0.143***	-0.0606***
	(0.0137)	(0.0163)	(0.0194)	(0.0224)	(0.00830)	(0.00914)
Average NDBI	5.609***	5.876***	6.335***	6.590***	0.410**	0.781***
	(0.289)	(0.323)	(0.357)	(0.384)	(0.155)	(0.179)
Average NDVI	-4.268***	1.999	-5.343***	4.086**	-3.010***	-0.406
	(1.101)	(1.214)	(1.459)	(1.532)	(0.630)	(0.684)
Average NDWI	-1.437	4.117***	-2.368	6.054***	-2.232***	0.205
	(1.143)	(1.248)	(1.530)	(1.587)	(0.660)	(0.704)
Log(Nightlight/area)		0.484***		0.702***		0.104***
		(0.0168)		(0.0232)		(0.00799)
Observations	8493	7643	8493	7643	2990	2629
Adjusted R^2	0.744	0.773	0.626	0.672	0.352	0.362

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.7: Long Differences Regression: Enterprise and Employment Density

	(1) Enterprises	(2) Enterprises	(3) Enterprises	(4) Employment	(5) Employment	(6) Employment
Change in NTL	-0.0170*		-0.0176*	0.00333		-0.00258
	(0.00856)		(0.00861)	(0.0160)		(0.0161)
Change in NDBI		0.136	0.195		0.630***	0.618***
		(0.108)	(0.111)		(0.177)	(0.180)
Change in NDVI		-0.914*	-0.871		-0.0106	-0.0472
		(0.456)	(0.478)		(0.794)	(0.838)
Change in NDWI		-0.929	-0.940		-0.229	-0.331
		(0.500)	(0.525)		(0.879)	(0.926)
Constant	0.255***	0.245***	0.265***	0.166***	0.173***	0.184***
	(0.00733)	(0.00649)	(0.00786)	(0.0130)	(0.0115)	(0.0137)
Observations	6940	7908	6940	6940	7908	6940
Adjusted R^2	0.000	0.001	0.001	-0.000	0.001	0.002

Standard errors in parentheses
Long Differences Regression 2009-2012
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.8: Long Differences Regression: Real Per Capita Expenditure

	(1) Expenditure	(2) Expenditure	(3) Expenditure
Change in NTL	-0.00483		-0.00421
	(0.0277)		(0.0276)
Change in NDBI		-0.382	-0.286
		(0.240)	(0.241)
Change in NDVI		-1.635	-1.618
		(1.150)	(1.212)
Change in NDWI		-1.657	-1.592
		(1.309)	(1.368)
Constant	0.386***	0.385***	0.401***
	(0.0189)	(0.0198)	(0.0243)
Observations	958	1056	958
Adjusted R^2	-0.001	0.001	-0.001

Standard errors in parentheses

Long Differences Regression 2008-2012

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.9: Long Differences Regression. Raw Landsat Bands

	(1) Enterprises	(2) Enterprises	(3) Enterprises	(4) Employment	(5) Employment	(6) Employment
Average NDBI	0.136			0.630***		
	(0.108)			(0.177)		
Average NDVI	-0.914*			-0.0106		
	(0.456)			(0.794)		
Average NDWI	-0.929			-0.229		
	(0.500)			(0.879)		
Average B1		-0.0302**	-0.0319**		-0.0699***	-0.0700***
		(0.01000)	(0.0103)		(0.0173)	(0.0178)
Average B2		0.0695**	0.0724**		0.167***	0.172***
		(0.0233)	(0.0245)		(0.0404)	(0.0420)
Average B3		-0.0323**	-0.0371**		-0.0740***	-0.0772**
		(0.0125)	(0.0136)		(0.0222)	(0.0238)
Average B4		-0.0135*	-0.0112		-0.00737	-0.00961
		(0.00659)	(0.00682)		(0.0115)	(0.0119)
Average B5		0.0175*	0.0169*		0.0193	0.0252
		(0.00809)	(0.00840)		(0.0134)	(0.0140)
Std Dev B4			-0.00452			0.0163*
			(0.00461)			(0.00806)
Std Dev B7			0.0162			0.0514**
			(0.00964)			(0.0178)
Constant	0.245***	0.239***	0.248***	0.173***	0.176***	0.196***
	(0.00649)	(0.00882)	(0.00932)	(0.0115)	(0.0155)	(0.0164)
Observations	7908	7908	7908	7908	7908	7908
Adjusted R^2	0.001	0.005	0.006	0.001	0.004	0.006
Variables selected by LASSO	3/3	9/9	12/18	2/3	8/9	16/18

Standard errors in parentheses

Long Differences Regression. All variables in changes between 2009-2012

Results on coefficients that did not reach significance at the 5% level in any regression are excluded in the table.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

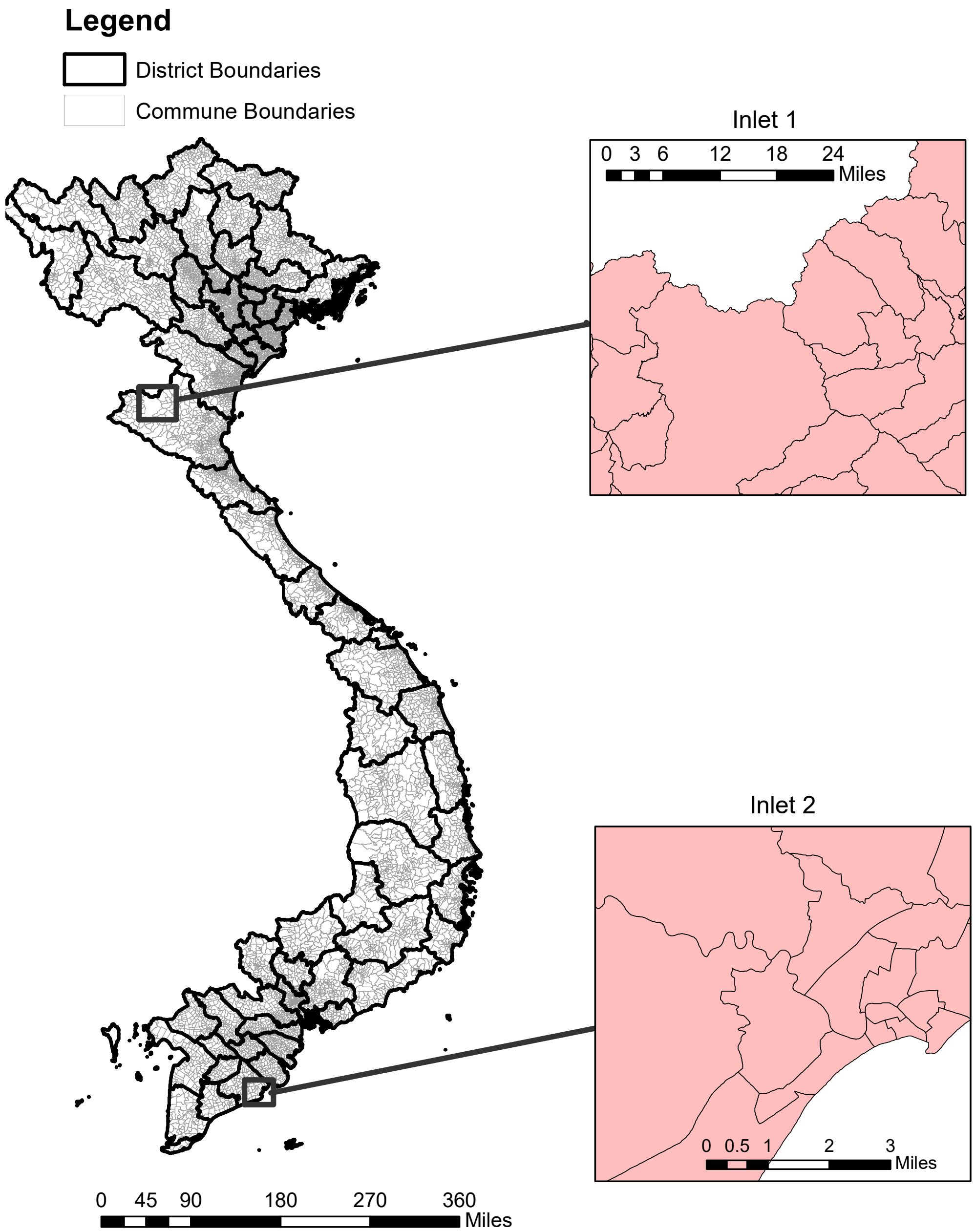

Figure 3.1: Administrative Boundaries of Vietnam

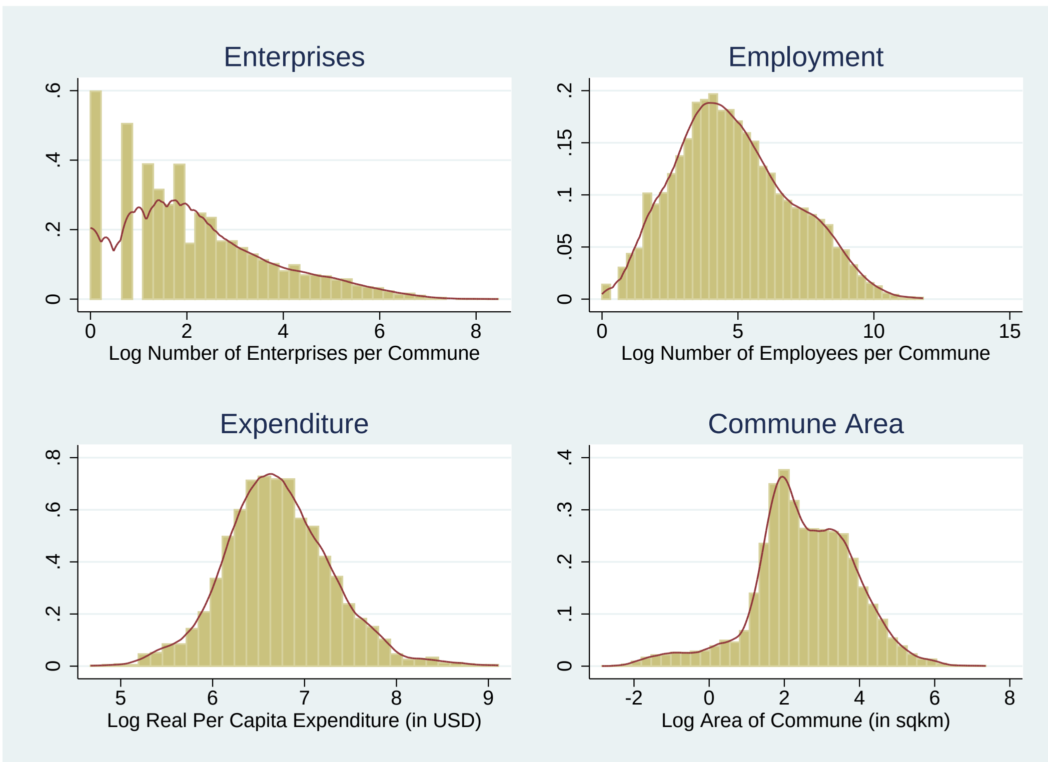

Figure 3.2: Histograms of Economic Variables

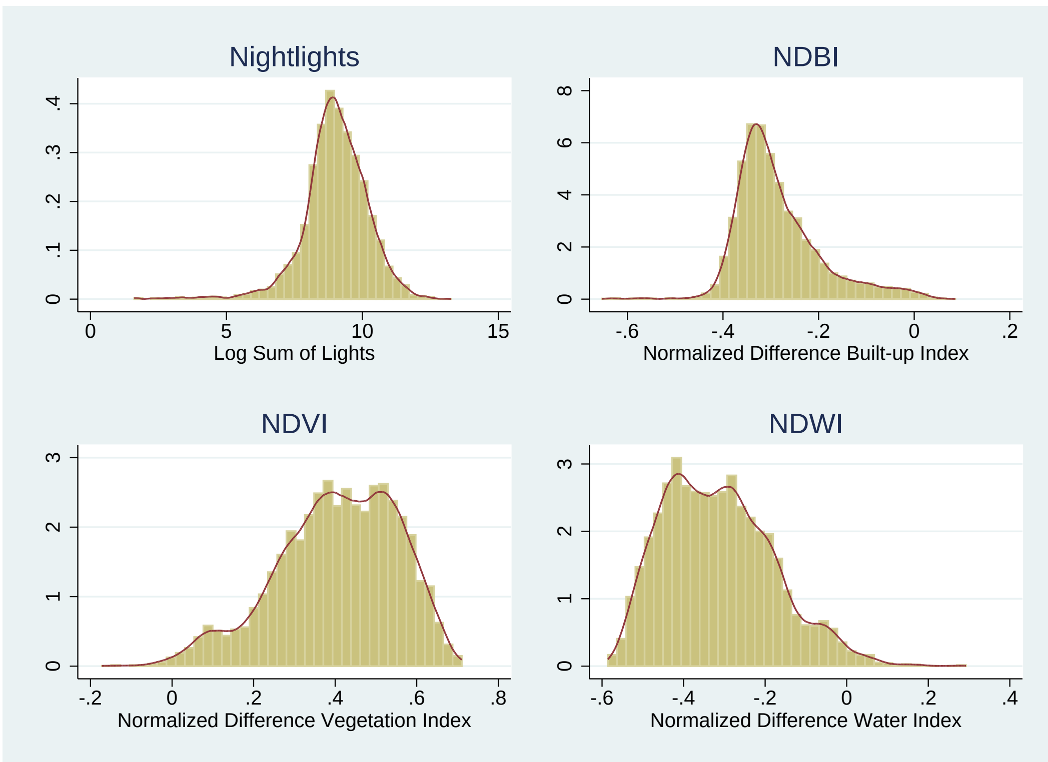

Figure 3.3: Histograms of Satellite Measures

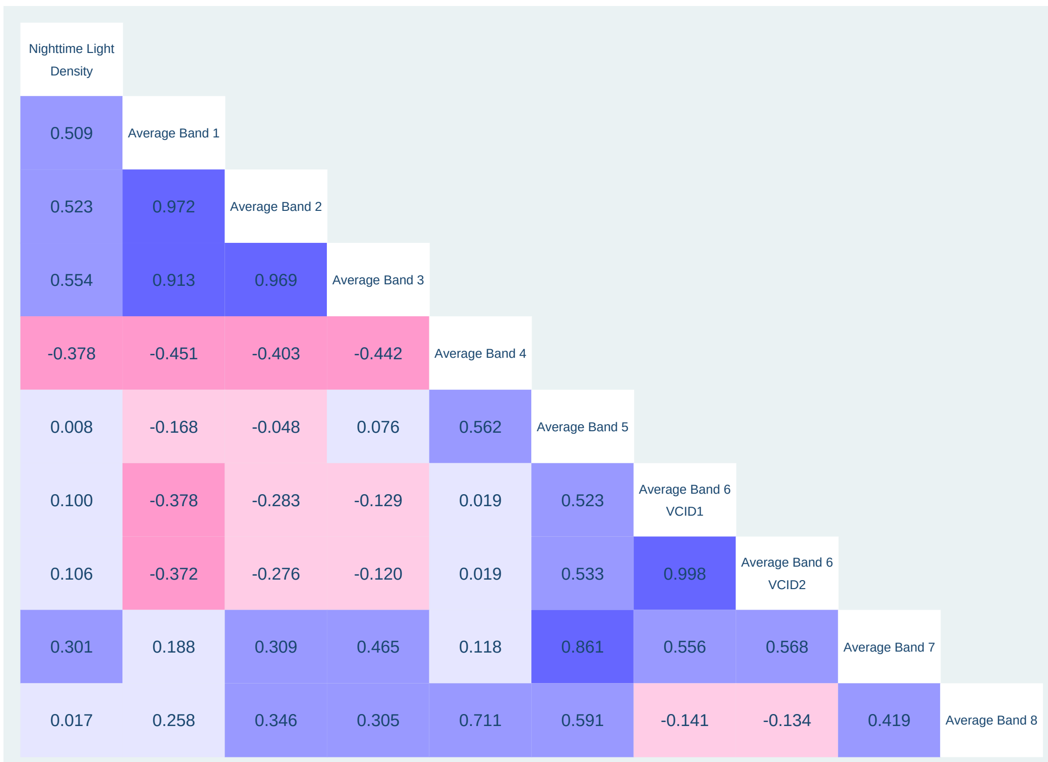

Figure 3.4: Correlation Heatmap between Nighttime Light Density and Landsat Bands.

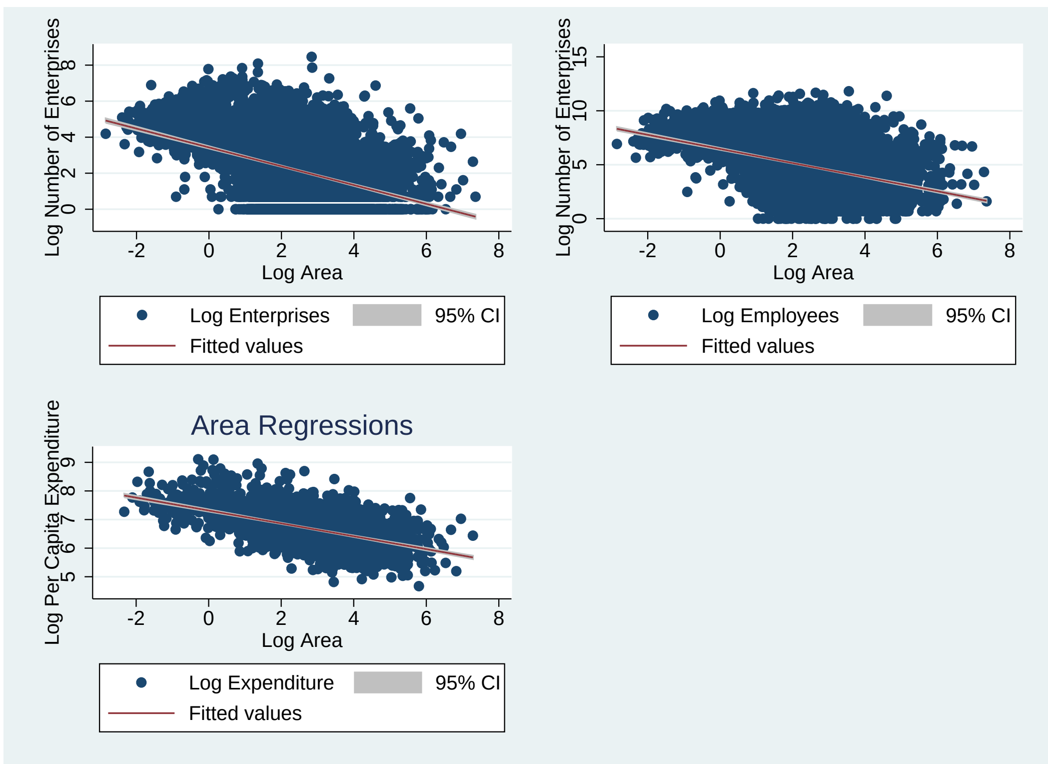

Figure 3.5: Relationship between Economic Activity and Commune Size

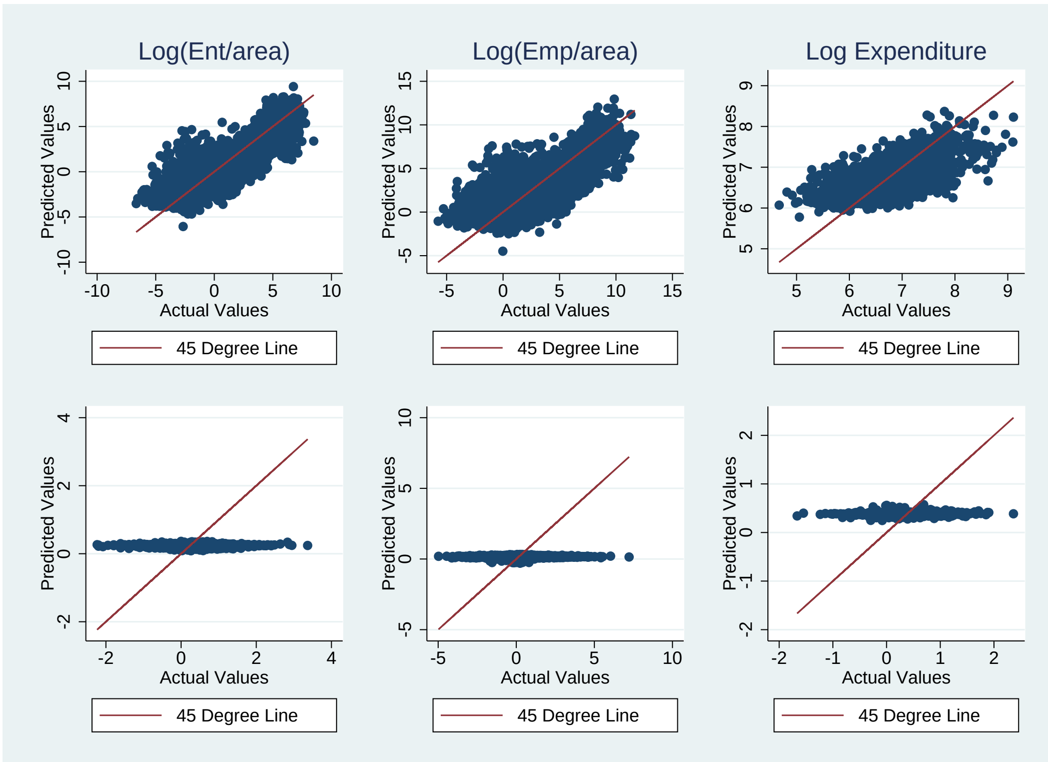

Figure 3.6: Scatterplots: Predicted vs Actual Values. Note: These scatterplots show the predicted versus the actual values from the models with the highest R-squared for log enterprise density (column 1), log employment density (column 2), and log per capita expenditure (column 3) for both the cross-section (top) and the time series (bottom). While in the cross-section, our models create some overestimation of low-activity communes and tend to underestimate high-activity neighborhoods, the time-series models are unable to replicate the large variation in the changes of economic outcomes between 2009 and 2012 (2008 and 2012 for expenditure).

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of Californias tobacco control program," *Journal of the American statistical Association*, 2010, *105* (490), 493–505.

_ , _ , and _ , "Comparative politics and the synthetic control method," *American Journal of Political Science*, 2015, *59* (2), 495–510.

_ and Javier Gardeazabal, "The economic costs of conflict: A case study of the Basque Country," *The American Economic Review*, 2003, *93* (1), 113–132.

Abrahams, Alexei, Nancy Lozano-Gracia, and Christopher Oram, "Deblurring DMSP Nighttime Lights," *Working Paper*, 2016.

Ahlfeldt, Gabriel M., Wolfgang Maennig, and Felix J. Richter, "Urban renewal after the Berlin Wall: a place-based policy evaluation," *Journal of Economic Geography*, 2016.

Altshuler, Alan and David Luberoff, *Mega Projects. The Changing Politics of Urban Public Investment*, Brookings Institution Press, 2003.

Anderson, Michael L., "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Highway Congestion," *American Economic Review*, 2014, *104* (9), 2763–2796.

Ashenfelter, Orley, Stephen Ciccarella, and Howard J Shatz, "French wine and the US boycott of 2003: Does politics really affect commerce?," Technical Report, National Bureau of Economic Research 2007.

Barbieri, Katherine, *The liberal illusion: Does trade promote peace?*, University of Michigan Press, 2002.

Baugh, Kimberly, Christopher D. Elvidge, Tilottama Ghosh, and Daniel Ziskin, "Development of a 2009 Stable Lights Product using DMSPOLS data," *Proceedings of the Asia-Pacific Advanced Network*, 2010, *30*, 114–130.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, 2007, *122* (2), 775–805.

— **and Matthew E. Kahn**, "Effects of Urban Rail Transit Expansions: Evidence from Sixteen Cities, 1970-2000," *Brookings-Wharton Papers on Urban Affairs*, 2005, pp. 147–206.

Bayer, Patrick and Robert McMillan, "Tiebout Sorting and Neighborhood Stratification," *Journal of Public Economics*, 2012, *96* (1112), 1129 – 1143. Fiscal Federalism.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, May 2014, *28* (2), 29–50.

Bickenbach, Frank, Eckhardt Bode, Peter Nunnenkamp, and Mareike Söder, "Night lights and regional GDP," *Review of World Economics*, 2016, *152* (2), 425–447.

Billings, Stephen B., "Estimating the Value of a New Transit Option," *Regional Science and Urban Economics*, 2011, *41* (6), 525–536.

Bleakley, Hoyt and Jeffrey Lin, "Portage and Path Dependence," *The Quarterly Journal of Economics*, 2012, *127* (2), 587–644.

Blumenberg, Evelyn A. and Paul M. Ong, "Cars, Buses, and Jobs: Welfare Participants and Employment Access in Los Angeles," *Transportation Research Board*, 2001, *1756*, 22–31.

Bollinger, Christopher R. and Keith R. Ihlanfeldt, "The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's MARTA," *Journal of Urban Economics*, 1997, *42* (2), 179 – 204.

Bowes, David R. and Keith R. Ihlanfeldt, "Identifying the Impacts of Rail Transit Stations on Residential Property Values," *Journal of Urban Economics*, 2001, *50* (1), 1 – 25.

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Brooks, Leah and Byron Lutz, "Vestiges of Transit: Urban Persistence at a Micro Scale," *Working Paper, Trachtenberg School of Public Policy and Public Administration, George Washington University*, 2013.

Cervero, Robert and Michael Duncan, "Transit's Value-Added Effects: Light and Commuter Rail Services and Commercial Land Values," *Transportation Research Record: Journal of the Transportation Research Board*, 2002, *1805*, 8–15.

Chandra, Amitabh and Eric Thompson, "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, 2000, *30* (4), 457 – 490.

Chavis, Larry and Phillip Leslie, "Consumer boycotts: the impact of the Iraq war on French wine sales in the US," *QME*, 2009, *7* (1), 37–67.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, 2014.

Chyn, Eric, "Moved to Opportunity," *Unpublished Manuscript*, 2016.

Clerides, Sofronis, Peter Davis, and Antonis Michis, "National Sentiment and Consumer Choice: The Iraq War and Sales of US Products in Arab Countries," *The Scandinavian Journal of Economics*, 2015, *117* (3), 829–851.

Cling, Jean-Pierre, Mireille Razafindrakoto, and François Roubaud, "The informal economy in Viet Nam," *Hanoi: International Labour Organisation*, 2011.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor, "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 1999, *107* (3), 455–506.

Damm, Anna Piil and Christian Dustmann, "Does growing up in a high crime neighborhood affect youth criminal behavior?," *The American Economic Review*, 2014, *104* (6), 1806–1832.

DART, Dallas Area Rapid Transit, "Independent Financial Feasibility Study," 1986.

_ , "Integrated Work Program. Report No 3. Rail System," 1986.

_ **and Marwick Peat**, "Regional Transit Service Plan Financial Analysis: Final Report," 1983.

Davis, Christina L and Sophie Meunier, "Business as usual? Economic responses to political tensions," *American Journal of Political Science*, 2011, *55* (3), 628–646.

Donaldson, Dave, "Railroads of the Raj. Estimating the Impact of Transportation Infrastructure," *American Economic Review*, forthcoming.

_ **and Adam Storeygard**, "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, November 2016, *30* (4), 171–98.

Duranton, Gilles and Matthew A. Turner, "The Fundamental Law of Road Congestion: Evidence from US Cities," *The American Economic Review*, 2011, *101* (6), 2616–2652.

Dutzik, Tony, Jeff Inglis, and Phineas Baxandall, "Millennials in Motion. Changing Travel Habits of Young Americans and the Implications for Public Policy," 2014.

Fisman, Raymond, Yasushi Hamao, and Yongxiang Wang, "Nationalism and economic exchange: Evidence from shocks to sino-japanese relations," *Review of Financial Studies*, 2014, *27* (9), 2626–2660.

Friedman, Monroe, *Consumer boycotts: Effecting change through the marketplace and the media*, Psychology Press, 1999.

Fry, Richard and Paul Taylor, "The Rise of Residential Segregation by Income," *Pew Research Center*, 2012.

Fuchs, Andreas and Nils-Hendrik Klann, "Paying a visit: The Dalai Lama effect on international trade," *Journal of International Economics*, 2013, *91* (1), 164–177.

Gartzke, Erik, Quan Li, and Charles Boehmer, "Investing in the peace: Economic interdependence and international conflict," *International organization*, 2001, *55* (02), 391–438.

Gibbons, Stephen and Stephen Machin, "Valuing Rail Access using Transport Innovations," *Journal of Urban Economics*, 2005, *57* (1), 148–169.

Glaeser, Edward L and Matthew E. Kahn, "Sprawl and urban growth," *Handbook of Regional and Urban Economics*, 2004, *4*, 2481–2527.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport, "Why do the poor live in cities? The role of public transportation," *Journal of Urban Economics*, 2008, *63* (1), 1–24.

Goldblatt, Ran, Wei You, Gordon Hanson, and Amit K. Khandelwal, "Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine," *Remote Sensing*, 2016, *8*.

Gonzalez-Navarro, Marco and Matthew A. Turner, "Subways and Urban Growth: Evidence from Earth," *Working Paper*, 2014.

Govella, Kristi and Sara Newland, "Hot Economics, Cold Politics? Reexamining Economic Linkages and Political Tensions in Sino-Japanese Relations," 2010.

Guerra, Erick, Robert Cervero, and Daniel Tischler, "The Half-Mile Circle: Does It Best Represent Transit Station Catchments?," *University of California Transportation Center Working Papers*, 2011.

Harari, Mariaflavia, "Cities in Bad Shape. Urban Geometry in India," *Working Paper*, 2016.

Henderson, J Vernon, Adam Storeygard, and David N Weil, "Measuring economic growth from outer space," *The American Economic Review*, 2012, *102* (2), 994–1028.

Hewitt, Christopher M. and W.E. Hewitt, "The Effect of Proximity to Urban Rail on Housing Prices in Ottawa," *Journal of Public Transportation*, 2012, *15* (4), 43–65.

Ihlanfeldt, Keith R, "Rail Transit and Neighborhood Crime: The Case of Atlanta, Georgia," *Southern Economic Journal*, 2003, *70* (2), 273–294.

Jensen, Hans Rask, "The Mohammed cartoons controversy and the boycott of Danish products in the Middle East," *European Business Review*, 2008, *20* (3), 275–289.

John, Andrew and Jill Klein, "The boycott puzzle: consumer motivations for purchase sacrifice," *Management Science*, 2003, *49* (9), 1196–1209.

Kahn, Matthew E., "Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems," *Real Estate Economics*, 2007, *35* (2), 155–182.

Kain, John F, "The spatial mismatch hypothesis: three decades later," *Housing policy debate*, 1992, *3* (2), 371–460.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Ko, Kate and Xinyu Jason Cao, "The Impact of Hiawatha Light Rail on Commercial and Industrial Property Values in Minneapolis," *Journal of Public Transportation*, 2013, *16* (1), 3.

Lauterpacht, Hersch, "Boycott in International Relations," *Brit. YB Int'l L.*, 1933, *14*, 125.

LeRoy, Stephen F. and Jon Sonstelie, "Paradise lost and regained: Transportation innovation, income, and residential location," *Journal of Urban Economics*, 1983, *13* (1), 67–89.

Li, Quan and Rafael Reuveny, "Does trade prevent or promote interstate conflict initiation?," *Journal of Peace Research*, 2011, *48* (4), 437–453.

Massoud, Tansa G and Christopher S Magee, "Trade and Political, Military, and Economic Relations," *Peace Economics, Peace Science and Public Policy*, 2012, *18* (1), 1–37.

Mayer, Thierry and Corentin Trevien, "Urban Public Transportation and Firm Location Choice Evidence from the Regional Express Rail of Paris Metropolitan Area," Technical Report November 2013.

McDonald, John F. and Clifford I Osuji, "The Effect of Anticipated Transportation Improvement on Residential Land Values," *Regional science and urban economics*, 1995, *25* (3), 261–278.

McKenzie, Brian and Melanie Rapino, "Commuting in the United States: 2009," *American Community Survey Reports*, 2011.

McMillen, Daniel P. and John F. McDonald, "Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 19831999," *Real Estate Economics*, 2004, *32* (3), 463–486.

Mellander, Charlotta, José Lobo, Kevin Stolarick, and Zara Matheson, "Night-Time Light Data: A Good Proxy Measure for Economic Activity?," *PloS one*, 2015, *10* (10), e0139779.

Michaels, Guy and Xiaojia Zhi, "Freedom fries," *American Economic Journal: Applied Economics*, 2010, *2* (3), 256–281.

Nall, Clayton, "The Political Consequences of Spatial Policies: How Interstate Highways Facilitated Geographic Polarization," *The Journal of Politics*, 2015, *77* (2), 394–406.

NCTCG, North Central Texas Council of Governments, "On-board Transit Survey," 2014.

Oreopoulos, Philip, "The Long-Run Consequences of Living in a Poor Neighborhood," *The Quarterly Journal of Economics*, 2003, *118* (4), 1533–1575.

Osofsky, Gilbert, *Harlem. The Making of a Ghetto. Negro New York, 1890-1930*, Harper and Row, 1966.

Pandya, Sonal S and Rajkumar Venkatesan, "French Roast: Consumer Response to International ConflictEvidence from Supermarket Scanner Data," *Review of Economics and Statistics*, 2016, *98* (1), 42–56.

Pinker, Steven, *The better angels of our nature: Why violence has declined*, Vol. 75, Viking New York, 2011.

Reardon, Sean F. and Kendra Bischoff, "Income Inequality and Income Segregation," *American Journal of Sociology*, 2011, *116* (4), 1092–1153.

Redding, Stephen J. and Matthew A. Turner, "Transportation Costs and the Spatial Organization of Economic Activity," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Elsevier, 2015, pp. 1339 – 1398.

Ryan, Sherry, "The Value of Access to Highways and Light Rail Transit: Evidence for Industrial and Office Firms," *Urban Studies*, 2005, *42* (4), 751–764.

Sanchez, Thomas W., "The Impact of Public Transport on US Metropolitan Wage Inequality," *Urban Studies*, 2002, *39* (3), 423–436.

Small, Christopher, Francesca Pozzi, and Christopher D Elvidge, "Spatial analysis of global urban extent from DMSP-OLS night lights," *Remote Sensing of Environment*, 2005, *96* (3), 277–291.

Storeygard, Adam, "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

Tiebout, Charles M., "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 1956, *64* (5), 416–424.

Vessali, Kaveh, "Land Use Impacts of Rapid Transit: A Review of Empirical Literature," *Berkeley Planning Journal*, 1996, *11* (1), 71–105.

von Thünen, Johann Heinrich, *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalkonomie*, Warthe, 1826.

Weiss, Jessica Chen, "Authoritarian signaling, mass audiences, and nationalist protest in China," *International Organization*, 2013, *67* (01), 1–35.

_ , *Powerful patriots: nationalist protest in China's foreign relations*, Oxford University Press, 2014.

Zeile, William J, "US intrafirm trade in goods," *Survey of Current Business*, 1997, *77* (2), 23–38.

Zha, Yong, Jay Gao, and Shaoxiang Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *International Journal of Remote Sensing*, 2003, *24* (3), 583–594.