2020

Three papers in regional dynamics and panel econometrics

Kevin Davey Duncan
Iowa State University

Three papers in regional dynamics and panel econometrics

by

Kevin Davey Duncan

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Helle Bunzel, Co-major Professor
Otavio Bartalotti, Co-major Professor
Peter Orazem
Brent Kreider
Arnold Cowan

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2020

DEDICATION

I would like to dedicate this thesis to my wife Jennifer, whose support and sacrifices that enabled me to complete this work. This is written with my son Silas, and his brother to be, heavily in my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I have neither been a great student, nor an easy one, and for that I am thankful for the handful of professors that have participated most in my training while at Iowa State University. In the order they entered my life, includes Peter Orazem, Georgeanne Artz, Gray Calhoun, Helle Bunzel, and Otavio Bartalotti. Their patience and mentorship has been a foundation that enables me to confidently take my next steps.

I also owe a limitless debt to my wife, Jennifer, and my parents, Patricia Davey and Scott Duncan.

ABSTRACT

This dissertation includes three chapters that cover broad topics in economics. The first chapter explores how the US Government's Capital Purchase Program, a large capital injection to local and regional banks through a stock purchase agreement, impacted local establishment dynamics such as entry, exit, employment expansion, and employment contraction following the 2008 Financial Crisis. The Capital Purchase Program dispersed over $200 billion dollars to banks hoping to prevent failure and ease tightened lending conditions. I estimate the direct effects of a county having a bank receive Capital Purchase Program funds on local business dynamics in the seven years following treatment, as well as spillover effects as entrepreneurs and business in neighboring regions travel to gain access to credit. Estimates show the CPP had no effect on establishment entry and exit, nor employment expansion and contraction. This paper establishes that the business-lending aims of the CPP were not realized in the communities and regions that received funds, and casts further doubt on meaningful pass through of CPP funds to desirable local economic activity.

The second chapter develops a joint hypothesis centered Wald test over fixed effects in large N small T panel data models with symmetric serial correlation within cross sectional observations. The enables joint hypothesis tests over inconsistently estimated fixed effects, such as the traditional varying intercept model as well as models with individual specific slope coefficients. I establish two different set of assumptions where feasible tests exist. The first assumption requires that individual errors follow a stationary $AR(p)$ process. Under this assumption all second and fourth cross product moments can be consistently estimated while allowing for individual specific hypothesis and covariates to vary across individuals and time with individual specific slopes. The second feasible test requires individuals to have coefficient slopes that are shared among all individuals in a known grouping structure under the null. This set of assumptions enables estimation of a completely unconstrained variance-covariance matrix and higher cross product moments for

individuals. Examples of these tests arise in wanting to establish latent panel structure, such as unobserved grouping of individuals, wanting to compare different models of teacher or firm value added against each other, or testing whether or not fixed effects can be approximated by Mundlak-Chamberlain devices.

Finally, the third paper estimates how messages displayed on Dynamic Message Boards, large signs either adjacent to or above roads, impact near to sign accidents. In this research, I look at the traffic-related messages such as "drive sober," "x deaths on roads this year," and "click it or ticket," displayed on major highways, on reported near-to-sign traffic accidents. This provides estimates of the impact of different types of nudges on road safety behavior. To estimate the causal effect of these nudges, we build a new high-frequency panel data set using the information on the time and location of messages, crashes, overall traffic levels, and weather conditions using the data of the state of Vermont over a three year time period. I estimate models that control for endogeneity of displayed messages, or allow for spillover effects from neighboring messages.

CHAPTER 1. INTRODUCTION

This dissertation is comprised of three papers, each of which are loosely connected by a shared interest in impacts of policies across regions. They are ultimately concerned with understanding and estimating how government policies impact local dynamics. The first paper estimates the impacts of the Capital Purchase Program, the bailout of banks through buying preferred stock, on local establishment dynamics following the 2008 financial crisis. The second provides a new method to conduct joint hypothesis tests over fixed effects when panels are both short and feature serial correlation. The final chapter estimates the impacts of behavioral and informational nudges provided through Dynamic Message Boards, large electronic signs that often sit next to or above major roads, on near to sign accidents.

The overarching question for these papers are how can we improve evaluation of government programs? How do these policies impact regional communities, from across the country, to even variation within a single state. The first and third paper tackle this question empirically, using a variety of extensions of existing causal inference procedures. Somewhat surprisingly, both papers show that government programs do not accomplish the goals they set out for. The second chapter helps develop new a new method of exploring regional effects by allowing researchers to impose increasingly granular joint hypothesis on data as the sample size grows, such as testing whether or not effects are the same across different regions as more are added.

In Chapter 1 I show that the government's bailout of local and regional banks under the Troubled Asset Relief program did not improve business dynamics in counties that had a bank receive CPP funds, nor did any of the surrounding counties. This effect is important since the government spent over $250 billion dollars, and almost $125 billion towards local and regional banks, to help prevent bank failure and promote lending. Chapter 2 provides a new joint test over inconsistently estimated parameters with applications to regional economics, but also has value to many other problems such as employer-employee fixed effects, teacher value added, and Chamberlain-Mundlak devices. The final paper shows that signs displayed over highways do not impact near-to-sign accidents after controlling for sequential exogeneity of traffic accidents to previous message states, and endogeneity between displayed message content and contemporaneous road hazard around a Dynamic Message Board.

CHAPTER 2. THE CAPITAL PURCHASE PROGRAM'S EFFECTS ON ESTABLISHMENT DYNAMICS OVER THE BUSINESS CYCLE

Kevin D. Duncan

Iowa State University

Modified from a manuscript to be submitted to *American Economics Journal: Eocnomic Policy*

Abstract

Using census data on county level business dynamics this paper estimates the impacts of the Treasury Department's Capital Purchase Program on establishment entry, establishment exit, employment expansion, and employment contraction following the 2008 Financial Crisis. I estimate the direct effects of a county having a bank receive Capital Purchase Program funds on local business dynamics in the seven years following treatment, as well as spillover effects as entrepreneurs and business in neighboring regions travel to gain access to credit. Estimates show the CPP had no effect on establishment entry and exit, employment expansion, or contraction. This paper establishes that the business-lending aims of the CPP were not realized in the communities and regions that received funds, and casts further doubt on meaningful pass through of CPP funds to desirable local economic activity.

"The breakdown of key markets for new securities has constrained the ability of even credit worthy small businesses and families to get the loans they need.... It is essential that we get these markets working again so that families and businesses can have access to credit on reasonable terms."
- Tim Geithner, Treasury Secretary 4/21/09

2.1 Introduction

This paper estimates the impact of the Treasury Department's Capital Purchase Program (CPP) on establishment entry, establishment exit, employment expansion, and employment contraction in the 7 years after the 2008 financial crisis. The CPP provided $205 billion dollars to more than

700 banks in over 400 US counties in order to prevent bank failure and stimulate loan supply as part of the broader Troubled Asset Relief Program (TARP). The CPP was one of the largest fiscal responses of the US government to the great financial crisis. The Treasury Department explicitly stated that benefits of the CPP ideally would be passed along to individual households and non-financial firms to bolster beliefs about the government's willingness to loosen credit markets. The aim would be that this would lead to otherwise improved economic activity throughout the worst part of the crisis, where households could gain loans for mortgages, and entrepreneurs or existing businesses could keep loans to start or stay in business at existing levels of employment.

We answer the question on whether or not the Capital Purchase Program impacted local establishment dynamics. If the CPP cased eased lending standards and actual pass through to local households and businesses, prospective entrepreneurs and business owners could have either opened new stablishments or expanded employment more than otherwise in communities with consumer demand for new goods and services. Alternatively, entrepreneuers and business owners might have been granted bridge loans to avoid excess or layoffs if they expected consumer demand to return soon, helping mitigate establishment exit or employment contraction. Previous work by Sheng (2015) shows that large firms that borrowed from banks that received CPP funds did not increase investment or R&D spending, and instead altered firms liquidity and financial decisions. Improved local establishment dynamics in comparison provides clear measures of positive economic value in comparison.

Positive impacts of the CPP would have lead to (1) increased establishment entry, (2) decreased establishment exit, (3) increased the number of establishment expansions, and (4) restricted the number of establishment contractions. Positive firm dynamics are a main contributor to TFP growth (Lee and Mukoyama (2015), Clementi and Palazzo (2016)) and lead to lower unemployment and stronger economic growth out of economic depressions. In practice the CPP can be viewed as a loan guarantee scheme, programs where the government takes up a guarantor of loans that financial institutes pass along to enterprises, where now the government precommits to back loan creation. Previous work on loan guarantee schemes has found they can provide an efficient means of job creation, but guaranteed projects are marginally more likely to fail, that they do induce funds from banks that otherwise would not be lent, and widening to larger firms and loans may hurt program benefits (Riding and Haines, 2001; Parker, 2005).

This paper estimates the direct impacts of a county receiving CPP funds utilizing census data on aggregate county level establishment dynamics. I further estimate the spillover effect of the CPP on neighboring counties that did not receive funds directly but were within 50 miles of a county that received treatment. These results extend previous work by Berger and Roman (2017) showing commercial real estate lending and off-balance-sheet real estate guarantees increased net job creation and net hiring establishments while decreasing business and personal bankruptcies. This paper further provides evidence on how young firm activity is tied to location financial health and credit supply (Davis and Haltiwanger, 2019). A major concern is that previous work estimating whether or not the CPP induced increased commercial and industrial lending from banks. Many studies have come to inconclusive and often contradictory results (Contessi and Francis, 2011; Cole, 2012; Black and Hazelwood, 2013; Blau et al., 2013; Li, 2013; Bassett et al., 2017; Berger et al., 2019a). Importantly, Jang (2017) shows that TARP money provided to distressed areas had spillover effects into neighboring, better performing, counties.

Complementary research has further explored other bank level responses to the Capital Purchase Program. Carow and Salotti (2014) show the Treasury Department gave CPP funds to weaker banks only if they had better performing loan portfolios. Operating efficiency of TARP banks generally decreased relative to non-TARP banks (Harris et al. (2013)). TARP receiving banks gained a competitive advantage by increasing market shares and power due to perceived safety of consumers (Berger and Roman (2016)), and were able to buy up other failed banks for substantial positive abnormal stock returns (Cowan and Salotti (2015)). Banks that received TARP money contributed less to economy wide systemic risk (Berger et al. (2019b)). That CPP funds provided only short term relief to participating commercial banks (Calabrese et al. (2017)). Broad overviews of research in this area have also been generated in Calomiris and Khan (2015) and Berger (2018).

Analysis of the CPP benefits from several stylized facts; the CPP had statutory requirements where the Treasury could only purchase stock valued between 1%-3% of a banks troubled assets, up to $25 billion and that among counties that received money only a few banks received CPP funds. Combined, these facts allows me to view a county as treated as long as at least one bank received CPP funds. We provide estimates of models with just direct and indirect effects, and then differentiate by timing differences on when counties had banks receive CPP funds to define potential outcomes of both own-treatment in either 2008 or 2009, and whether or not a county was adjacent- defined as being within 50 miles of a neighbor counties center- to a treated counties. Treatment effects might be differentiated across time due to both differences in when banks where mandates to apply by, and the type of banks and communities that might have received treatment in each period.

Treatment effects are estimated using a panel data method similar to Hsiao et al. (2012). Since the number of treated and untreated counties is much larger than the number of pre-treatment time periods, we augment the procedure with a Least Absolute Shrinkage and Selection (LASSO) technique such as in Doudchenko and Imbens (2016). With far more untreated counties than time periods traditional methods will uniquely fit on the pre-trend time periods. LASSO fixes this by selecting only the counties that most closely match a given treated county. This is different than the synthetic-control style estimators as it removes the convex hull assumptions such as in Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015); Ferman and Pinto (2016). We show that sample splitting techniques across different treated groups allow for easy estimation of the Average Treatment on the Treated even with spillover effects, and many treated counties that neighbor each other. This relaxes the shared spillover effects as in Cao and Dowd (2018), and allowing for two treated units to be adjacent to each other as ommitted from Di Stefano and Mellace (2020) as comparisons within the synthetic control literature. The downside is for counties with both individual specific direct and indirect treatment effects our estimates can only recover the mean effect for the group instead of individual specific effects.

The results indicate that both direct and spillover effect of a county having a bank receive CPP funds on establishment entry, establishment exit, employment expansion, and employment contraction where non-existant. Establishment entry among treated counties decreased around 10 fewer entrants a year, exits increased 40 additional exits a year, but showed long run improvement. The number of establishments increasing employment decreased by 50 directly following receiving CPP funds, and about 45 additional esetablishments contracted employment. However, five to six years after receiving treatment, firm entry returned to its previous levels, about 40 fewer firms exited treated counties starting in 2011, and there were about 50 more employment expanding- and 50 fewer employment contracting- firms.

Even as average causal effects showed generally no to undesirable outcomes among treated counties, county level heterogeneity shows many counties saw marked improvement. All treatment effects are highly correlated with each other, with a major driver being the large number of firms that enter and exit in a single year. This is not surprisingly since trying to provide funds directly to banks is similar to the pass through of monetary policy changes to credit markets which have previously been shown to have considerable heterogeneity (Blau et al., 2013). One of the most striking results is that immediately following treatment employment expansions (contractions) are almost strictly negative (positive), indicating that few small and medium firms got access to bridge loans to stop them from having to lay off works in the face of contracting consumer demand. The lack of pass through to small and medium establishments is important as most small businesses do not have access to equity markets, and rely on local or regional banks for credit. Relationship lending has been recently established as a major way in which banks recover underlying firm specific behavior (Berger and Udell, 2002).

Motivation for synthetic control methods are provided in robustness checks, where tests for pretrends are rejected across a variety of multiple difference-in-differences estimators and instrumental difference-in-differences specifications. Instead direct estimations of interactive fixed effects difference-in-difference models are carried out using a number of specifications that confirm with earlier synthetic control methods. Overall, this paper provides clear evidence that the CPP did not generate pass through to improved local establishment dynamics. Banks might have preferred providing pass through to households seeking home mortgages, or alternatively might have parked the money as a risk free loan from the banks to pay off other existing balance sheet effects.

The paper proceeds as follows. Section 2.2 describes the Capital Purchase Program in greater detail. Section 2.3 describes the data, providing preliminary data analysis and provides summary statistics. Section 2.4 formalizes the empirical design and estimation processes. Section 2.5 provides our preferred LASSO-synthetic control estimation results. Section 2.6 provides robustness checks. Section 2.7 concludes.

2.2 The Capital Purchase Program

The Capital Purchase Program provided extra capital to banks by buying non-voting senior preferred shares on standardized terms to offset now-high risk assets remaining on bank's balance sheets. The CPP provided $205 billion to more than 700 banks. The first 10 banks received just over $125 billion. These banks include Bank of America, Bank of New York Mellon, Citigroup, Goldman Sachs, JP Morgan Chase, Morgan Stanley, State Street Corporation, Wells Fargo, 1st Financial Services Corporation, and Bank of Commerce Holdings. The public perception was that these banks were almost forced to take CPP funds as part of the government's bailout of the financial sector.

Individual banks applied for CPP funds through their federal regulator- the Federal Reserve, FDIC, Office of the Comptroller of the Currency, or the Office of Thrift Supervision.[1] Banks indicated a preferred level of stock purchase between one and three percent of the total risk-weighted Assets of the applicant up to $25 billion.

Federal regulatory agencies chose which banks received money and sent preferred set of applicants to the Treasury Department for final clearance. Duchin and Sosyura (2014) show that

[1]The application period lasted between October 3rd, 2008 to November 14th, 2008 for publicly held companies, December 8th for Privately held companies, and February 13th, 2008 for S Corporations. On May 20th, 2009, Timothy Geithner announced that for banks with assets less than $500 million would have a second window to apply for CPP funds for the following 6 months. https://www.treasury.gov/press-center/press-releases/Pages/tg139.aspx

of roughly 600 public firms, 416 firms (79.8%) applied, 329 (79.1%) were accepted, and that 278 (84.5%) accepted the funds but 51 (15.5%) declined. Among private banks that applied, applications that were rejected or withdrawn were not announced or publicly disclosed. All initial payments to participating banks were made before January 1st, 2010. However, there are clear spikes in lending. A large number of funds were dispersed in 2008, a slow down through the holidays, and another large group of funds were dispersed at the start of 2009 (Figure 2.2). Many counties had only a few banks receive funds, and even a smaller share of banks received multiple injections. Between 2008 and the end of 2010, the average county had 2.06 injections in total, often in separate banks (Figure 2.3).

The non-voting senior preferred shares required a 5% dividend for the first 5 years and 9% afterwards.[2] However previous research has indicated that these purchases were preferential for the banks. The Congressional Oversight Panel estimated that the Treasury gave out $254 billion in 2008 across all TARP programs, for which it received assets worth approximately $176 billion, a difference of $78 billion. Equivalently, Veronesi and Zingales (2010) estimate during the first 10 transactions of the CPP, the Treasury overpaid between $6-13 billion for financial claims.

Overall the CPP provided standardized amounts of capital to participating banks in one of two main treatment branches, at the very end of 2008 or the very beginning of 2009. We see bunching in funds per worker in Figure 2.4. Many banks applied, and few turned down funds after being accepted.[3] Since most counties only had a small number of banks receive CPP funds, treatment can be viewed through the lens of did a specific county have at least one bank receive CPP funds in either 2008 or 2009. This allows reduction of an otherwise complex problem with both continuous treatment assigning and treatment intensity as a more tractible problem with discrete treatment assignment and singular treatment intensities.

[2]Participating banks would also be able to receive future Treasury purchases of common stock up to 15% of the initial CPP investment for the following 10 years- allowing for additional buy in if the Treasury judged their initial purchase was not high enough.

[3]Official documentation guaranteed banks that applied and got turned down for funds did not get publicly announced. This makes extrapolation from the Duchin and Sosyura (2014) results difficult.

2.3 Data and Summary Statistics

The primary dependent variables of interest are county level establishment entry, establishment exit, employment expansion, and employment contraction from 1999 to 2015 provided by the Census Statistics of US Businesses & Business Information Tracking Series (SUSB).[4] Establishments are classified a single physical location in which business is conducted, where individual companies or enterprises can be spread across multiple establishments. Most importantly, each establishment has non-zero levels of employment, ruling out non-employee firms from the sample. Estimation of the average treatment on the treated, the average change caused by the CPP on entry, exit, employment expansion, or employment, covers a wide span of pass through activities from increased rates of lending.

Entrants are establishments with zero employment in the first quarter of the initial year, and positive employment in the subsequent year. Exiters have positive employment in the initial year and zero employment in the subsequent year. Expansions are establishments with positive first quarter employment in both the initial and subsequent years and increased employment during the time period between the first quarter of the initial year and the first quarter of the subsequent year. Contractions are establishments that have positive first quarter employment in both the initial and subsequent years and decrease employment during the time period between the first quarter of the initial year and the first quarter of the subsequent year. We exclude any county that had zero firm entry or exit, removing 161 counties (see Appendix 2.8.3).[5]

Figures 2.5-2.6 plot mean establishment entry and establishment exit by observed treated status in 2008 and 2009. When plotted at levels, there exist large differences in firm dynamics. Counties that received treatment in both 2008 and 2009 average more than 1500 new entrants/exits a year. Counties that received only one treatment tend to average around 500-700 new entrants and exists a year, and non-treated counties have barely any entry. However rescaling each time series subject

[4]The underlying files can be downloaded as https://www2.census.gov/programs-surveys/susb/.

[5]Moreover, as discussed later, our estimation strategy never picks up these counties when looking to create synthetic counties using either the level or rates of firm dynamics.

to within-group means and standard deviations prior to 2007 show considerable similarities in each group, and show that appropriate Difference-in-Differences techniques might be useful in creating valid counterfactuals for each treated sub-group.

The majority of firms are small. From the Census' County Business Patterns data, which tracks the total number of establishments in a given county, roughly 55% of firms have between one and four employees, 20% have between five and nine employees, and 12% have between 10 and nineteen employees. These numbers are very stable across all years in the sample. The majority of firms are small mom-and-pop set ups. The SUSB data does not disentangle firm size, but using this sample I assume that the majority of new entry is small. This is further supported by other studies, for example Mata and Portugal (1994); Bartelsman et al. (2005); Kaniovski and Peneder (2008). Most firms enter and fail within the first year or two.

There is strong evidence that in good times credit constraints do not impact the decision to enter into entrepreneurial activity given a lack of a relationship between wealth and entry into entrepreneurship (Hurst and Lusardi, 2004). Data from the 2003 National Survey of Small Business Finances show that among firms that had only opened after 2002, 25% of firms had 0 outstanding loans, and 50% had less than \$7000 in loans. Among those firms that had taken out capital leases, 25% of them owed less than \$4000 in principal, and 75% owed less than \$45,000.[6] The Federal Reserve's Small Business Credit Survey, in 2018, across the life cycle of firms, 25% to 35% of firms with employers had no outstanding debt. For debt, 46% of new firms did use a loan or line of credit as a regular source of external financing, while only 9% of new firms had outside equity financing. Over the life cycle the share of firms taking equity fell, while the share taking on loan increased. Almost half of firms between 0 and 15 years in business applied for financing in the previous year, most seeking between \$25,000 and \$100,000. Shane (2010) points roughly 48.4% start in residence-

[6] Of new firms that do not take out loans, most are in categories highly likely to fill consulting jobs, special trade contractors, miscellaneous manufacturing industries, personal services, and engineering and management services. Comparably among new entrants that did take out loans, they were more concentrated in restaurants, retail, business services, trucking, or durable storage.

such as home or garage, and an additional 40.64% in a rented or leased space. And that the typical median start-up in the US requires $24,000-30,000 in start up capital.

Most importantly there is large stability in the change in the number of establishments at different firm sizes. Figure 2.1 graphs the change in the share of establishments with different sets of employees, 1-4, 5-9, 10-19, 20-49, 50-99, 100-249. Both the number of new firm entrants in each county and the share of firms at different levels of employment are very stable. Firms with 1-4 employees consistently make up almost 55% of the change in establishments, firms with 5-9 employees has fallen slightly from being 20% of the change in establishments to 17%, and firms with 20-49 employees have increased from 0.8% to 1% of new entrants. These shifts are small, but follow general concerns about firm concentration, and a need for perceived higher capital constraints relative to the late 1990's.

Treatment status is defined as a county receiving CPP funds in a particular year. The Treasury Department updates the TARP Transaction Report that includes bank name, state name, and city name data. I directly attach Federal Reserve Replication Server System Database ID's (RSSD ID) using the 2008 and 2009 FFIEC Call Reports and Summary of Deposits.[7] From this we are able to calculate both a head quarty specific county treatment effect, and a bank network county treatment effect.

For concreteness, let $(i, j) \in \{1, \ldots, N_c\}$ index the number of counties, and $k, l \in \{1, \ldots, N_b\}$ index bank headquarters, and for each headquaker k we have $b_k \in \{1, \ldots, N_{b_k}\}$ as an index for the number of branch locations, and each bank k exists in some county i. Now there are two treatments, HQ treatment location, and bank wide (BW) treatment. Own treatment is defined as an indicator value on whether or not a county received any CPP funds during a given time period.

[7] The TARP Transaction Report. Similarly, the 2008 and 2009 FFIEC Call Reports can be found here: https://www.fdic.gov/regulations/resources/call/index.html. To match banks in the TARP Transaction Report to RSSDID's we first pick a bank-state-city group from the TARP Transaction Report, then condition the Call Report data on city, state, and only banks that contain the entirety of the bank from the Transaction Report (after removing REGEX and making both names lower case). This matches on roughly 630 of the 707 banks. The remaining share are added directly.

I separate own-treatment status into two groups, the first being receiving CPP funds in 2008, and the second being receiving CPP funds in 2009. The own-treatment variable takes the form,

$$Own_{i,t}^{HQ} = 1\{\exists k \in i \text{ s.t. } CPP_{k,t} > 0\}$$

$$Own_{i,t}^{BW} = 1\{\exists b_k \in i \text{ s.t. } CPP_{k,t} > 0\}$$

The term $CPP_{i,t}$ is the dollar amount of CPP funds given to banks in county i in period t. Using this definition, 63 counties received CPP funds only in 2008, 243 received CPP funds only in 2009, and 81 counties received CPP funds in both periods.

Similarly, credit markets may extend beyond county borders, implying that treating a county i may impact nearby counties. A neighbor is defined to be any county with centroid distance within 50 miles of a subject county i.[8] This metric is used as entrepreneurs have empirically traveled moderate distances trying to find beneficial loan deals, such that in Belgian banks the maximum loan distance is 50 miles (Degryse and Ongena, 2005), while in the US average bank applications come from 10 miles away (Agarwal and Hauswald, 2011), with a standard deviation of 21 miles, while accepted applications come from even closer to the bank (2.62 miles), with a smaller standard deviation (10.67 miles). Thus while most bank applications are local, applicants are willing to drive moderate distances in search of favorable loan contracts. Under this setting we define the neighbor treatment variable as

$$Neigh_{i,t}^{HQ} = 1\{\exists j \text{ adjacent to } i \; s.t. \; CPP_{j,t} > 0\}$$

$$Neigh_{i,t}^{BW} = 1\{\exists b_k \text{ adjacent to } i \text{ s.t. } CPP_{k,t} > 0\}$$

A major source of possible bias is that the largest banks in the US were perceived to be highly illiquid at the start of the Great Financial Crisis. These banks were effectively told to take CPP funds, and thus did not opt in to the program. Moreover, most of these banks paid back CPP loans quickly in order to remove requirements on executive pay and other conditions for the funds.

[8]Based on NBER County Distance Database restricted to county centroids within 50 miles of each other. http://www.nber.org/data/county-distance-database.html

The concern is that these banks sat on the funds rather than using them as part of regular bank operations, and adding in their responses might spoil results (see for example Li (2013)). The counties with the top 20 largest banks, and the communities immediately adjacent to them are removed from our sample. Moreover, we treat locations with branch locations as non-treated by the status of the headquarters.[9] The major reason for this assumption is that most of these banks had been caught with high credit risk due to investment activities, and not underlying weakness in branch location financial conditions.

Mean bank characteristics at the county level are calculated from FDIC call sheet data, Following Li (2013) we calculate troubled assets ratio, annualized Return on Assets, and loan-to-deposits ratio.[10] These proxy for local community bank health that the Federal Regulators may have observed when deciding which banks to accept into the CPP program.

Local labor market characters are provided through the BLS's Local Area Unemployment statistics on county level unemployment rates.[11] While taxes do impact firm entry decisions, often these impacts are economically small, and explain a very small fraction of the variation in firm entry decisions (Duncan, 2015). Instead a major driver of firm entry appear to be unobserved demand

[9] Most of these counties are bank holding companies. The FDIC call sheet data lists all downstream assets held by branches at the bank holding company's headquarters. The list of banks include, Goldman Sachs, J.P.Morgan Chase Bank, Keybank (Keycorp), PNC Bank, Fifth Third Bank, Bank of America, BB&T Bank (BB&T Corp), State Street, U.S. Bank (U.S Bancorp), Wells Fargo Bank, Suntrust Bank, Citibank, Capital One, Regions Bank, Bank of New York Mellon, Northern Trust Company, Comerica Bank, M&T Bank, Marshall&Ilsley Bank, and Morgan Stanley. In practice this excludes New York, NY; Charlotte, NC; Boston, MA; Minneapolis, MN; Cleveland, OH; Pittsburgh, PA; Cincinnati, OH; Atlanta, GA; McLean, VA; Birminham, AL; Chicago, IL; Dallas, TX; Buffalo, NY; and Milwaukee, WI.

[10] Values are calculated from call sheet data from 2008Q3. Tier 1 Ratio is calculated directly in the Call Sheets as RCON7206. Troubled Asset Ratio is loans 90 days past due/total capital. Troubled Assets are calculated as 90 Days Past Due C&I Loans (RCON5460) and All Other Loans Past Due 90 Days or More (RCON5460). For Total Capital are calculated as Total Assets (RCON2170) and subtracted Total Liabilities (RCON2948). Return on Assets was Net Income (RIAD4340) divided by Total Assets. Cash to Assets was Cash and Due From Depositors (RCON0010) divided by Total Assets. Loan to Deposits Ratio was Loans, Leases, Net Unearned Income (RCONB528) divided by Total Deposits (RCON2200).

[11] https://www.bls.gov/lau/

for products and agglomeration economies. Measures of upstream and downstream agglomeration economies are calculated from input-output tables. These take three forms, the first is industry cluster, measured as each industry's share of total employment in a county/year pair relative to the industry share in the nation as a whole. Upstream and downstream measures of connectedness are calculated from the Bureau of Economic Analysis' 1997 Standard Use Table. The share of workers providing inputs to each 2 digit NAICS code is calculated for in each county and year. Using this again the upstream and downstream measures is calculated by taking the share of workers providing inputs into each 2 digit NAICS code divided by total employment in each period. This is again normalized by the average across the United States. Measures of household financial health are provided by the FDIC experimental county level home price index, however The FDIC data exclude counties without enough mortgages to draw a consistent enough estimate of household financial wealth, this using only counties where the home price index exists excludes many rural counties.

Summary statistics for each of these variables is provided in Table 2.1. The first column, "PrGFC" is Pre-Great Financial Crisis, provides the mean across all counties and year from 1999 to 2007. The second column, "PoGFC" is Post-Great Financial Crisis, and reports the mean across all counties and years from 2008 to 2015. The third column, "Diff" reports the difference-in-means between the first and second column. As expected, firm entry and employment expansion went down, first exit and employment contractions went up. Unemployment rates went up, banks deleveraged and Troubled Asset Ratio's decreased, and return on assets increased. The average change in the Home Price Index (HPI) was negative over the Post-GFC time period. Columns four and five report the standard deviation of the pre and post financial crisis period, and the sixth reports the difference. Entry, exit, and employment expansion all feature less variation in the post-financial crisis era, while contractions variation increased.

Finally, a number of other policy drivers have studied determinants of firm entry, such as right to work laws (Holmes, 1998) or lower taxes (Rohlin et al., 2014; Duncan, 2015). Often specific research designs are used to estimate these effects and remove endogeneity of pro-business practices such as I exclude these variables due to fear of inducing larger biases in my estimates, especially given that they do not explain a large share of the overall variation in firm entry dynamics. In many of these cases the proposed models both explain a small share of the overall variation in firm entry, or show that the treatment effects have economically small coefficients.

2.4 Empirical Design

We are interested in recovering the direct and indirect Average Treatment on the Treated effect for a county having a bank receive CPP funds on future business dynamics. A major concern is that there is large heterogeneity in how communities were impacted by the 2008 financial crisis, and how local bank financial characteristics created pass through to local businesses and entrepreneurs. This creates ambiguity in what the appropriate counterfactual is to non-treated counties, and motivates the use of synthetic control methods.

A major source of confounding in my research design exists in credit market spillovers. Entrepreneurs are likely to travel moderate distances in order to acquire credit to start, expand, or stop bank failure on a business. As a result counties are not independent of each other, and instead rely on a both their own sources of productivity and access to credit, as well as those around them. We follow Huber and Steinmayr (2019) to utilize a potential outcome framework with own and neighbor treated status. The core assumptions being that changes in which neighbor is treated does not impact your potential outcome outside of either a neighbor being treated or the number of neighbors being treated, and no complementarities between own treatment and neighbor treatment status.

More formally, there are T time periods. From periods $0, \ldots, T_0 < T - 2$ all counties are untreated. In periods $T_1 = T_0 + 1$ each county can receive CPP treatment. In periods $T_2, \ldots, T$ no more treatment is assigned. Under this framework we have two treatments, own treatment $Own_{i,t} \in \{0, 1\}$ or neighbor treatment $Neigh_{it} \in \{0, 1\}$. Therefore individuals treatment status can be characterized in the set $S_{iT_1} = (Own_{iT_1}, Neigh_{iT_1}) \in \{(0,0),(1,0),(0,1),(1,1)\}$. Assume the simple structural model for untreated counties as,

$$y_{it}(0,0) = x_{it}\beta + \lambda_t'\mu_i + \epsilon_{it} \tag{2.1}$$

Then for treated counties we get the following series of equations,

$$y_{it}(1,0) = y_{it}(0,0) + \alpha_{it} \tag{2.2}$$

$$y_{it}(0,1) = y_{it}(0,0) + \gamma it \tag{2.3}$$

$$y_{it}(1,1) = y_{it}(0,0) + \alpha_{it} + \gamma_{it} \tag{2.4}$$

Under this factor structure λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, and the error terms ϵ_{it} are unobserved transitory shocks at the region level with zero mean. This structure is general and nests a number of common data generating processes.[12] Implicitly in this structure we assume no complementarities or substitution effects between treatment and neighbor treatment status. This allows estimation of average treatment effects through estimation of synthetic control on sample splitting. That is,

$$\alpha_{it}^{(1,0),(0,0)} = y_{it}(1,0) - E(y_{it}(0,0) \mid (1,0)) \tag{2.5}$$

$$\alpha_{it}^{(1,1),(0,1)} = y_{it}(1,1) - E(y_{it}(0,1) \mid (1,1)) \tag{2.6}$$

$$\gamma it^{(0,1),(0,0)} = y_{it}(0,1) - E(y_{it}(0,0) \mid (1,0)) \tag{2.7}$$

$$\gamma it^{(1,1),(1,0)} = y_{it}(1,1) - E(y_{it}(1,0) \mid (1,1)) \tag{2.8}$$

The aim is to construct a synthetic county out of linear combinations of counties with a different treatment status. Traditionally this was done through a convex hull assumption such as in Abadie

[12]It is common in the "synthetic control" literature to assume a shared time varying intercept for all counties in the sample, equivalently, the "panel data approach" assumes an county specific intercept. Both are special cases of the unconstrained fixed effects model. For example, while the model with the shared time varying intercept nests the differences-in-differences model when $\lambda_t = 1$, both models are nested when $\lambda_t' = [1\ \eta_t]'$, $\mu_i = [\theta_i\ 1]$.

and Gardeazabal (2003); Abadie et al. (2010, 2015); Ferman and Pinto (2016), where all weights are strictly positive and sum to one. This assumption was removed in Hsiao et al. (2012); Li and Bell (2017). The main difference between the two is that the "panel data approach" is an unconstrained regression, and the synthetic control method is a constrained regression. Similar approaches without constraints have started to implement LASSO and other regularization methods (Doudchenko and Imbens, 2016; Chernozhukov et al., 2017; Amjad et al., 2018; Carvalho et al., 2018). A comparison of these methods was conducted by Gardeazabal and Vega-Bayo (2017); Wan et al. (2018). With only a single treatment, synthetic control estimates county specific ATT's, but with two different treatment effects these estimates become a county specific total treatment effects, and parsing out average direct and spillover effects requires modifications.

Estimation of the LASSO-synthetic control estimator is carried out through minimizing penalized regression.

$$\begin{bmatrix} \hat{w}_i \\ \hat{\beta}_{i,0} \end{bmatrix} = \arg\min_{B_{i,0},w_i} \frac{1}{T_0}\sum_{t=1}^{T_0}\left(y_{it} - \beta_{i,0} - \sum_{j=1}^{N_0} w_{ij}y_{jt}\right)^2 + \phi \mid\mid w_i \mid\mid_2 \tag{2.9}$$

The first part of this equation is regular OLS as carried out in Hsiao et al. (2012), where we match a set of donor counties to a specific treated counties for all the pre-treatment time periods. However, since $N_B >> T_0$, we force the procedure to select only a subset of counties. Therefore the second term, $\phi \mid\mid w_i \mid\mid_2$ helps drive selection on fewer counties than pre-treatment time periods, where $\phi > 0$ determines the severity of the penalty for picking an additional county and is determined by cross validation, and $\mid\mid w_i \mid\mid_2 = \sum_j w_{ij}^2$. This structure is close to (Wan et al., 2018; Doudchenko and Imbens, 2016; Li and Bell, 2017). Without loss of generality, assume we are estimating $\alpha_{it}^{A,B}$, where A is a treated set, and B is a donor set. Then Equation 2.5 can be reformulated

$$\alpha_{it}^{A,B} = (y_{it} - w_i Y_{jt})$$
$$= \left(\alpha_{it} + (\gamma_{it} - \sum_{j \in B} w_{ij}\gamma_{jt}) + \lambda_t(\mu_i - \sum_{j \in B} w_{ij}\mu_j) + (\epsilon_{it} - \sum_{j \in B} \epsilon_{jt}) \right)$$

This estimator becomes unbiased under the following assumption

Assumption 2.4.1.

$$E[\epsilon_{it} \mid Own_i \ Neigh_i] = E[\epsilon_{it}] = 0$$

$$\exists w^* \in \mathbb{R}^{N_B} \mid (\mu_i - \sum_{j \in B} w_{ij}\mu_j)) = 0, E[\gamma_{it} - \sum_{j \in B} w_{ij}\gamma_{jt}] = 0$$

The first part of this assumption states that treatment can be correlated with the factor loading term, $\lambda_t' \mu_i$, but are uncorrelated with idiosyncratic shocks to a given county. The second requires that our pre-treatment fit provides a close approximation for the unobserved time-invariant county specific factor loadings, and that in the post treatment time period provide a mean zero approximation for the second treatment effect. This implies that the shared treatment effects γ_{it} all share common support across the target and donor pools.

A major concern is that the term $(\gamma_{it} - \sum_{j \in B} w_{ij}\gamma_{jt})$ varies over time. Assumption 2.4.1 claims in each period the treatments are random effects, such that $y_{it} = y_t + \upsilon_{it}$ and $upsilon_{it}$ is white noise.[13] In turn, we primarily focus on estimates of the mean effect,

$$\alpha_t^{A,B} = \frac{1}{N_A} \sum_{i \in A} (y_{it} - \hat{y}_{it}) \tag{2.10}$$

However, variation in treatment assignment can further be leveraged in estimation of effects. As discussed previously, there was an initial wave of payouts at the end of 2008, a slow down, followed by a second wave of dispersed funds at the start of 2009. Under this setup there are now more effects, and estimation assuming single effects leads to plausibly biased samples. Extending the previous treatment assignment description to include two periods of own treatment and neighbor

[13] An implicit implication of this is that individual counties should have no meaningful heterogeneity, and instead should be jumping around the mean. But this is often violated in practice.

treatment is fairly routine. As above, in periods $0, \ldots, T_0 < T-3$ all counties are untreated. In periods $T_1 = T_0 + 1$ and $T_2 = T_0 + 2$ each county can receive CPP treatment. In periods $T_3, \ldots, T$ no more treatment is assigned. Under this framework we now have two possible time periods where in each period one of two possible treatments can be received. In period T_1 individuals treatment status can be characterized as above. In period T_2 the nested outcomes generate sixteen potential outcomes. We index counties by their second period potential outcomes

$$(Own_{i,T_1}, Neigh_{i,T_1}, Own_{i,T_2}, Neigh_{i,T_2})$$

As above, assume the simple structural model for untreated counties,

$$y_{it}(0,0,0,0) = x_{it}\beta + \lambda_t'\mu_i + \epsilon_{it}$$

In period T_1 this generates the four possible outcomes in Equations 2.2. In period T_2 the potential framework becomes nested, where the four potential outcomes are repeated, conditional on treatment status from T_1. This leads to many cases similar to treated and neighbor treated in T_1, where there are many plausible individual specific parameters, but estimation of a single marginal effect (for example impact of first period treatment), now generates a large vector of nuissance parameters. First, under this framework, we can recharacterize the estimated treatment effect without loss of generality as,

$$\begin{aligned}
&(\alpha_{it}^{T_1} + (\gamma_{it}^{T_1} - \sum_{j \in B} w_{ij}\gamma_{jt}^{T_1}) + (\alpha_{it}^{T_2} I\{Own_{jT_2} = 1\} + \gamma_{it}^{T_2} I\{Neigh_{jT_2} = 1\}) \\
&\quad - \sum_{j \in B} w_{ij}(\alpha_{jt}^{T_2} I\{Own_{jT_2} = 1\} + \gamma_{jt}^{T_2} I\{Neigh_{jT_2} = 1\}) \\
&\quad + \lambda_t(\mu_i - \sum_{j \in B} w_{ij}\mu_j) + (\epsilon_{it} - \sum_{j \in B} \epsilon_{jt}))
\end{aligned}$$

The leading term $\alpha_{it}^{T_1}$ is assumed to be the estimated effect of interest. The second term is the difference between the spillover effect in the first time period. The third term represents omitted second treatment effects on the first period treated unit of interest. The fourth term reflects unaccounted for second period treatment effects within the donor pool. Without additional assumptions it is not possible to sign the difference between the third and fourth terms.

As above, the case of $Own_t = 1, Neigh_t = 1$ is hard to identify with the synthetic control method as the difference in secondary treatment effects not of interest creates a moving nuissance parameter. For example, consider the following set of possible potential outcomes in the two period, two treatments, framework.

$$y_{it} = \begin{cases} y_{iT_2}(0,0,0,0) + \alpha_{it}^{T_1} + \alpha_{it}^{T_2} & \text{if } Own_{T_1} = 1,\ Neigh_{T_1} = 0,\ Own_{T_2} = 1,\ Neigh_{T_2} = 0 \\ y_{iT_2}(0,0,0,0) + \gamma_{it}^{T_1} + \gamma_{it}^{T_2} & \text{if } Own_{T_1} = 0,\ Neigh_{T_1} = 1,\ Own_{T_2} = 0,\ Neigh_{T_2} = 1 \\ y_{iT_2}(0,0,0,0) + \alpha_{it} + \gamma_{it} & \text{if } Own_{T_1} = 1,\ Neigh_{T_1} = 0,\ Own_{T_2} = 0,\ Neigh_{T_2} = 0 \end{cases}$$

Without additional assumptions it is impossible to jointly identify $(\alpha_{iT_2}^{T_1}, \alpha_{iT_2}^{T_2})$, $(\gamma_{iT_2}^{T_1}, \gamma_{iT_2}^{T_2})$, nor $\{(\alpha_{iT_2}^{j}, \gamma_{iT_2}^{j})\}_{j \in \{T_1, T_2\}}$. As above I remedy this issue by conditioning on a given positive treatment regime, and targeting the specific average effect of interest. For example, if I am interested in $\alpha_t^{T_0}$, the donor pool becomes $A = (1,0,0,0)$, and the donor pool is $B = (0,0,0,0)$. Similarly, the target pool $A = (1,1,0,0)$ is paired with the donor pool $B = (0,1,0,0)$. The estimator is still unbiased under Assumption 2.4.1. This means all treatment effects- own treatment 2008, neighbor treatment 2008, own treatment 2009, neighbor treatment 2009, share common supports across all treated counties.

The advantages of this approach is reducing each equation down to the canon causal effects structure, with downside being the loss of data within each equation. For each observation in the treated branch we construct a synthetic control county using the donor pool, and the fit across the donor pools differs greatly. Counties that would be picked by selecting weights across the entire sample are often excluded due to treatment statuses outside of the comparison at hand. Cao and Dowd (2018) offer an alternative way to estimate this equation under an imposed symmetry for indirect effects of receiving treatment. Their method allows for using the full sample to estimate the set of weights for every county in the sample, but imposes a stronger structural assumption on the underlying causal framework.

Inference for synthetic control methods is carried out using a permutation test (Abadie et al. (2015)). For each group assume the null hypothesis of no-treatment effect. Then re-sample without replacement a new treated group of size N_A and estimate the mean LASSO-synthetic control estimator. This procedure is repeated 1000 times. This approximates the exact null distribution under the sharp null of no-treatment effect. Therefore, there exists a treatment effect when the point estimates for the observed treated group lies outside the 95% permutation test confidence interval.

2.5 Results

Results for levels and rates are plotted in Figures 2.7-2.14. The results are similar both for levels and rates. In both cases the synthetic control estimator described in Section 2.4 fail to reject the null hypothesis of no treatment effect. The pre-treatment fit is well within the 95% permutation confidence interval, and never crosses the permutation confidence interval in the post-treatment time period.

Point estimates are notable, in levels the direct impact on establishment entry is almost 50 fewer firms a year in the period immediately after treatment, or about a 1% lower entry rate, with both returning towards zero in the long run. The major difference between these two terms is that the average treated county has moderately higher average firm entry rates as discussed in Section 4.2. Indirect effects are much smaller on entry, with about 20 fewer firms, or about 0.1% lower firm entry rate. Establishment Exit shows a long run decline in both levels and longs, with 50 fewer exits a year, or about -.1% lower exit rate. The spillover here is of the same magnitude as the direct effect. A curious part here is that the entire 95% permutation confidence interval is declining over the post-treatment time period even among the untreated pool.

Establishment employment expansion shows almost zero mass in the 95% permutation interval above zero in the years immediately following treatment. In the long run, this rises to about 100 more establishment expansions per year for both the direct and indirect effects, however in rates this the point estimates are approximinately close to zero. Equivalently, there is almost zero permutation distribution below zero immediately following the 2008 financial crisis. In the long run levels return to zero, while contraction rates show moderate decreases in the rate of firm contractions in the entire sample.

These results indicate no-effect from counties receiving CPP both directly and indirectly on the counties around them. However, these graphs are misleading in two ways. Each mean effect pools the average of 12 different estimates. This generates variation in treatment timing, particularly as the majority of counties did not receive CPP funds until 2009. As a result, I plot both the mean response along with 90% quantiles for each effect in Figures 2.15 and 2.16.[14] However, each of these are subject to possible estimation error from residual components of their potential outcomes. As before, both direct and indirect effects of entry are centered around zero.

The most useful conclusion from these graphs is a sign of clear bridge loan pass through in establishment employment expansion and contraction. Estimated effects are almost uniformly negative (positive) in the case of employment expansion (contraction), implying that few firms were able to forgo impacts of cratering consumer demand on their own employment status.

Overall, results indicate no effect on local establishment dynamics after a counties bank received CPP funds. For both direct and indirect effects entry and expansions decreased, exit and contractions increased, however considerable heterogeneity exists in these effects on individual counties. Many counties saw considerable positive gains to firm entry for both direct and indirect effects. Direct effects on entry saw some counties saw large declines in firm exit, while spillover effects saw a large tail of counties that saw excess exit for years following treatment. Both direct and spillover effects on expansion were generally negative immediately following treatment, followed by either strong positive expansion starting in 2011 for the directly treated counties, and generally no effect for spillover counties. The direct and spillover effects of contractions saw a strong center directly on zero, with a large upper tail of excess contractions.

[14] Some of the models fit particularly poorly, and given the relatively few treated individuals in some groups, the 90% quantile is a good first-pass approximation to the distribution among treated units. Later I carry out permutation tests under the null of no effect, and can compare these quantiles.

2.6 Robustness Checks

This section provides a variety of robustness checks on our primary results. We provide Difference-in-Differences and Instrumental Variables Difference-in-Differences results, specifically talking about how pretrend violations occur and providing further evidence of the need of data driven methods of constructing counterfactuals. Generally we find violations of the pre-trend for Own treatment status across both model specifications, but Neighbor pretrends generally hold. However as above, there is still no discernable spillover effect.

Continued concerns about differing pretrends among treated and untreated individuals leads to estimation of interactive fixed effects difference-in-differences models Gobillon and Magnac (2016); Xu (2017). These explicitly estimate a factor loading model such as Equation 2.10 to constructing counterfactuals, implying a stronger structure than mainline synthetic control estimates require.

Finally, as noted in Section 4.2 the Tarp Transaction Report is tied to branch location that received funds. We construct a full network of counties with a treated bank's branch locations and reestimate interactive fixed effects difference-in-differences models.

2.6.1 Difference-in-Differences

The four and sixteen potential outcome framework discussed in Section 2.4 enables canonical difference-in-differences estimation now including a combination of own and neighbor treatment statuses. Recent work have helped decompose multiple time period or spillover effect difference-in-differences, most notably Imai and Kim (2014), and dealing with treatments in multiple time periods (Goodman-Bacon, 2018; Callaway and Sant'Anna, 2018).

We proceed by first estimating models with a single treatment period, and then two treatment periods with a full set of heterogeneous treatment effects.[15] For each model joint significance tests over the pre-trend are conducted using clustered standard errors are the state level. Finally, a

[15]Models are estimated using the plm package in R carrying out a within (individual FE) transformation with two way fixed effects. Heteroskedastic robust variance-covariance matrices are calculated with Arellano (1987) style standard errors with with county level clusters.

stepdown method is utilized to test whether or not there was an active policy duration. This provides a conservative test for exactly how long there was a policy effect from the CPP on local establishment dyanmics.

Recent work in difference-in-differences and event study methods have increasingly utilized policies that exhibit variation in treatment timing. Under these conditions it is common to generate pre and post treatment effects from time of initial treatment. Two concerns arise out of this. For counties treated in the second period treatment, the period prior to treatment is now subject to the Great Financial Crisis, something the treated in the first period group is not, therefore tests for differing pre-trend are carried out just on pre-Financial Crisis periods. The two-way fixed effects model, a saturated model with own and neighbor treatment effects, own and neighbor events for all years outside of $t = 2006$ to exclude the start of the financial crisis, county specific fixed effects, and time fixed effects, generates the estimated equation,

$$
\begin{aligned}
y_{it} = \beta_1 Own_i + \beta_2 Neigh_i + \beta_3 I\{t > T_0\} + \sum_{s=-9}^{7} \gamma_s Own_i I\{s = t - \min_k\{Own_{i,k+1} - Own_{i,k} = 1\}\} \\
+ \sum_{s=-9}^{7} \alpha_s Neigh_i\{s = t - \min_k\{Neigh_{i,k+1} - Neigh_{i,k} = 1\}\} + \Gamma X_{it} + \mu_i + \lambda_t + \epsilon_{it}
\end{aligned}
\tag{2.11}
$$

The term $I\{s = t - \min_k\{Own_{i,k+1} - Own_{i,k} = 1\}\}$ denotes the difference between the current time period and the first year a given county received treatment. This specification generates three different tests for pretrends of interest. The first is that all pretrends differ from zero, the second that only own pretrends differ from zero, and the third that only neighbor pretrends differ from zero.[16] Recent research has pointed out that by doing this, standard errors of post-treatment coefficients are often conservative (Roth, 2018; Kahn-Lang and Lang, 2019)- but generally this

[16]One concern is that since there is time-variation in treatment that occurs after the onset of the Great Financial Crisis that pre-trend tests might fail due to the large number of firms treated in period two and the pre-trend coefficient for γ_{-1} being strongly negative due to the GFC. To remedy this I actually report F-tests for a model which estimates $OwnTreated \times Year$ factors, and then imposes that there is no differing pre-trend from 2000 to 2007.

paper prefers a more conservative approach to estimating effects and does not carry out further corrections.

$$H_0^{ALL} = \gamma_{-9} \ \dots \ \gamma_{-1} \ \alpha_{-9} \ \dots \alpha_{-1} = 0$$
$$H_0^{OWN} = \alpha_{-9} \ \dots \alpha_{-1} = 0$$
$$H_0^{NEIGH} = \gamma_{-9} \ \dots \ \gamma_{-1} = 0$$

Even study style graphs of results are presented in Figure reffig:DIDPooled, and the resulting joint hypothesis tests on pretrend are presented in Table 2.2. Among own treatment effect, firm entry, firm exit, and employment contractions all grow leading up to the initial period of treatment. This visible difference in pre-trends (and levels) between treated and untreated counties in different treatment groups invalidates the use of the (mean) non-treated counties as a viable counter factual. Explicit discussion of the resulting effects generated by this estimation procedure might create poor policy conclusions. Comparably, the neighbor treated effect seems to more likely to follow a shared pretrend, even though it is still rejected in the joint test, but the resulting coefficients are close to zero.

As discussed in Section 2.4, there might have been meaningful choices in when Federal regulators and the Treasury decided to disperse funds to different banks or regions. As a result, the pooled estimator presented in Equation 2.11 does not capture the full heterogeneity in responses. Thus we estimated a fully differentiated model with differing pretrends and post treatment effects by each treatment group. This allows for heterogeneous responses within each own-treatment and neighbor-treatment couplet, and the resulting estimated equation then becomes.

$$
\begin{aligned}
y_{it} = {} & \beta_1 Own_i + \beta_2 Neigh_i + \beta_3 I\{t > T_0\} \\
& + \sum_{s=-9}^{7} \gamma_s^{10} Own_i^{10} I\{s = t - \min_k \{W_{k+1}^{10} - W_k^{10} = 1\}\} \\
& + \sum_{s=-9}^{7} \gamma_s^{10} Own_i^{01} I\{s = t - \min_k \{W_{k+1}^{01} - W_k^{01} = 1\}\} \\
& + \sum_{s=-9}^{7} \gamma_s^{10} Own_i^{11} I\{s = t - \min_k \{W_{k+1}^{11} - W_k^{11} = 1\}\} \\
& + \sum_{s=-9}^{7} \alpha_s^{10} Neigh_i^{10} I\{s = t - \min_k \{G_{k+1}^{10} - G_k^{10} = 1\}\} \\
& + \sum_{s=-9}^{7} \alpha_s^{10} Neigh_i^{01} I\{s = t - \min_k \{G_{k+1}^{01} - G_k^{01} = 1\}\} \\
& + \sum_{s=-9}^{7} \alpha_s^{10} Neigh_i^{11} I\{s = t - \min_k \{G_{k+1}^{11} - G_k^{11} = 1\}\} \\
& + \Gamma X_{it} + \mu_i + \lambda_t + \epsilon_{it}
\end{aligned}
\tag{2.12}
$$

Event study figures for results from Equation 2.12 are presented in Figures 2.19-2.20 in the Appendix 2.8.1. As above, joint tests on pretrends are carried out, where now this extends to all pretrends for each treatment subgroup. While visually the estimates appear to be much more centered around 0, most models still reject the hypothesis that there are no differing pre-trends among the different treatment groups. Allowing for additional heterogeneity shows that estimates for neighbor spillover effects tend to satisfy the shared pre-trend assumption. This implies that DID estimates for spillover effects are not invalidated, and that post-treatment estimates are supported by a valid counter factual.

As the financial crisis becomes less severe, capital is likely to ease nationally, and renormalization between treated and untreated counties may occur. Thus we develop explicit tests for policy effectiveness duration by conducting a multiple hypothesis tests using a step-down multiple hypothesis test outlined in section 2.6.1, based on a test for nested hypotheses proposed by Bauer and Hackl (1987). This test controls for family-wise error in trying to evaluate multiple p-values simultaneously. To motivate this problem imagine the set of hypotheses,

$$H_0^k: \; \gamma_s = 0 \quad \forall s \in [1, \ldots, k] \tag{2.13}$$

Then a level α-test for any null hypothesis H_0^k is given by the critical region $\min_{i \leq j \leq k} p_j \leq \alpha/(2(k-i+1))$, as under the null,

$$P(reject H_0^k) \leq \sum_{i=1}^{k} P(p_j \leq \alpha/(2(k-i+1))) \leq \alpha$$

By use of Bonferroni's inequality. This test then jointly controls for family wise error of multiple tests being conducted for the no treatment effect. This test is a worst-case bound for the existence of positive policy effective duration, and basically selects and carries out the appropriate joint hypothesis in an iterative fashion.[17] Tables for Stepdown tests of Equation 2.12 are presented in Tables 2.4-2.5 in Section 2.8.2.

Consistent with the results from the synthetic control methods there is no policy duration effect for spillover effects. There exist moderate effect durations for Own Treatment, but without accepting the shared pre-trend it is hard to argue what exactly the Difference-in-Differences estimator recovers.

2.6.2 Instrumental Variables Estimation

A concern about identification is that treatment is correlated with still unobserved shocks, even after conditioning on the interactive fixed effects. As noted in Li and Bell (2017), if Federal regulators and the Treasury picked areas for CPP funds with high latent demand for loans, then these estimates would overstate the CPP effectiveness, while comparably if they picked areas with low latent demand for loans, this might understate CPP effects. In turn we instrument own and

[17]There exist step down methods that use bootstrap methods to estimate dependence in the underlying tests to generate a less conservative tests. This test can also be augmented to explicitly test one sides hypothesis by using the appropriate t-values.

neighbor treatment using counties own or neighbors political ties, whether or not any bank in a given county had a board member serving as a branch Federal Reserve chair, whether or not the counties local House representative was serving on the banking and finance committee, the share of donations to the local representative coming from Financial, Investment, and Real Estate groups, and whether or not the local House representative was a democrat.

Following Ruonan Xu (2019) we estimate a bivariate Probit for for each year instrumenting using political connections of counties, where the outcomes are own and neighbor treated status. For treatment status in 2009 we further condition on whether or not a county or a neighbor received treatment in the previous time period. This generates six instruments, being the relative probabilities of own, neighbor, and both treatment status in both 2008 and 2009 from the two Probit models.

Sanderson and Windmeijer (2016)'s augmented F-test for multiple endogenous variables is carried out, where our instruments are strong using the Stock and Yogo (2005) tables. The generated conditional F-values are 27.9, 78.13, 13.6, and 28.6 for own treatment in 2008, neighbor treatment in 2008, own treatment in 2009, and neighbor treatment in 2009- respectively. Taking the norm bias of 10%, the relevant comparative critical value is 11.12. We then instrument each of our treatment statuses as a function of each of our instruments

$$Treat_{i,t} = \beta_0 + \beta_1 \hat{p}_{10}^{2008} + \beta_2 \hat{p}_{01}^{2008} + \beta_3 \hat{p}_{11}^{2008} + \beta_4 \hat{p}_{10}^{2009} + \beta_5 \hat{p}_{01}^{2009} + \beta_6 \hat{p}_{11}^{2009} + X_{it}'\Gamma + \epsilon_{it}$$

$Treat_{i,t}$ here includes $Own_{i,t}, Neigh_{i,t}$, and the timing-variants discussed for estimating Equation 2.12. Moreover, X_{it} includes county mean bank financial health, and local unemployment characteristics. Using these instrumented measured, we re-estimate Equations 2.11 and 2.12. As above, these models continue to reject the assumption of shared pre-trends presented in Table 2.6.

This is likely due to selection by central banks into which banks have members serve on the board, such that banks situated in larger areas where likely to be serving at the local Fed chair, and these regions were more likely to feature different trends in the build up to the Great Financial Crisis. Thus, even if the IV solves the issue of possible poaching by the Treasury into providing CPP funds to areas with disproportionately high or low latent credit demand, the IV might exacerbate issues underlying differing pre-trends among different counties in the US. As a result, I omit point estimates for treatment effects derived from the IV model.

2.6.3 Interactive Fixed Effects Differences in Differences

Instead of relying on a specific form of additively seperable individual and time specific fixed effects, the simple structural model presented in Equation 2.10 is built on interactive fixed effects, where r unknown time loading factors λ_t are interacted by county specific effects μ_i that determine how impactful certain shared shocks are on a given county.

This forces a more explicit structural model to be estimated than presented for the synthetic control model in Section 2.4, but enables a broader set of time-varying covariates. We follow the estimation processes outlined in Gobillon and Magnac (2016); Xu (2017), and results are presented in Figures 2.22-2.25. As before, results are often indistinguishable from zero for establishment entry and establishment exit. Comparably the long term the Interactive Fixed Effects DID models show long run improvements in employment expansion, and fewer employment contractions. However, these improvements become distinct from zero well after the counties received CPP funds. Therefore it is hard to know whether or not these results are coming from CPP treatment, or supplemental responses or policy changes happening in the long run.

2.6.3.1 Accounting for Downstream Counties

Results presented so far have relied on where the TARP Transaction Report said the receiving bank was located. Often this is tied to bank headquarters, where many of the banks that received TARP funds were publicly traded bank holding companies, or small regional branches. A concern about our earlier identification strategy is that banks might have passed CPP funds from the receiving headquarters location down to branches.

Identification of treatment effects here is difficult. By including all branch locations of the 10 largest banks, there is no identification to be had, and all remaining counties on our sample- almost

2500- become treated. However most of these banks were all but forced to take the money, and paid it back quickly to get out of requirements the CPP imposed on banks normal operation. As a result, we assume that these banks did not pass funds to downstream banks in their network. Instead this leaves about 1500 counties that received treatment as presented in Figures 2.26-2.28.

Previous robustness checks have cast consistent doubt on the presence of spillover effects, so rather than splitting the sample, we estimate only direct pooled treatment effects using the Interactive Fixed Effects Difference-in-Differences method. These results are presented in Figures 2.29-2.32, and generally confirm with out prior results, that is, an absence of treatment effects across the board. Results differ slightly though, there is a noticable drop off in the number of establishments expanding employment that is statistically different from zero. In the long run, this returns to zero. Secondly, long run trends in exits and employment contracting firms decrease in the long run- similar as they did for our non-downstream accounting for estimators earlier.

Since these trends are long after counties received CPP funds, it is hard to tie these improvements to the CPP. In all cases estimated effects are close to zero around the treatment window, or reflect negative policy outcomes.

2.7 Conclusion

The Capital Purchase Program provided over $200 billion to banks to shore up bank finances and ease credit constraints faced by credit worthy households and small businesses. In this paper we estimated how possible pass through of the CPP might have impacted county level establishment dynamics, including entry, exit, employment expansion, and employment contraction. This paper builds on the back of a borader corporate finance literature that found mixed evidence of whether or not banks generated more commercial or industrial loans, and that companies that borrowed from banks that received CPP funds generally did not put it towards R& D, employment, or new capital expenditures and instead changed around their underlying balance sheets.

Examining firm entry has several benefits over direct bank level responses. Relationship lending is a major driver of extending loans to new or existing entrepreneurs, and formally modeling the method by which banks extend these loans is difficult, leading to biased estimation by improper understanding of this mechanism. However, higher firm entry was still a preferred outcome of policy makers at this time as a way of encouraging new job growth and aiding recovery efforts.

The main estimation technique uses augments Hsiao et al. (2012) to include a LASSO penalty term, and leverages identification of marginal direct and spillover treatment effects for the treated using sample splitting techniques. This specification shows no evidence that counties receiving TPP funds having higher establishment entry, exit, employment expansion, or employment contraction both as a direct or indirect effect of the CPP. Most notably here is that bunching of both mean and county specific effects show the mass of treatment is above (below) zero for employment contractions (expansions), indicating that firms were unlikely to receive branch loans during periods directly following the CPP when counties and regional communities were most at risk for harsh contractions in consumer demand and business dynamics.

Robustness checks include a slew of more traditional Difference-in-Difference estimators that validate concerns about pre-trend violations among different treated groups after controlling for county specific and time fixed effects, as well as level of urbanization by time, and Federal Reserve branch area by time fixed effects. Instrumenting treatment status using political connections actively makes pretrend tests perform worse. Using a fully saturated model with different treatment effects based around two periods of treatment with two treatment statuses in each period (own and neighbor), Difference-in-Differences results satisfy pretrend assumptions but confirm with prior no spillover effect from our mainline specifications.

Two final robustness checks confirm our preferred LASSO-synthetic control estimates. We explicitly estimate Equation 2.1 and develop an interactive fixed effects difference-in-differences estimator that satisfies pretrends for all models, but continues to show no effects across entry, exit, employment expansion, and employment contraction. The last check incorporates all branch locations of treated banks outside the largest 20 banks in the country. These results mirror previous interactive fixed effects difference-in-differences estimates, where treated counties showed long term improvement in firm exit, employment expansion, and employment contraction, but generally occurred well after the program started, and are hard to tie explicitly to just county's CPP treatment.

These results closely mirror previous results showing no effect on bank level lending behavior following receiving CPP funds. If banks did not actively ease credit constraints to local firms, then new entrepreneurs and existing businesses would have continued to face the brunt of negative credit and consumer demand shocks unassisted. Given the large outlay of government funds to promote business lending, and poaching by Federal regulators and the Treasury to give money to predominately healthier banks, casts doubt on the use of such programs in the future.

2.8 Appendix

2.8.1 Figures

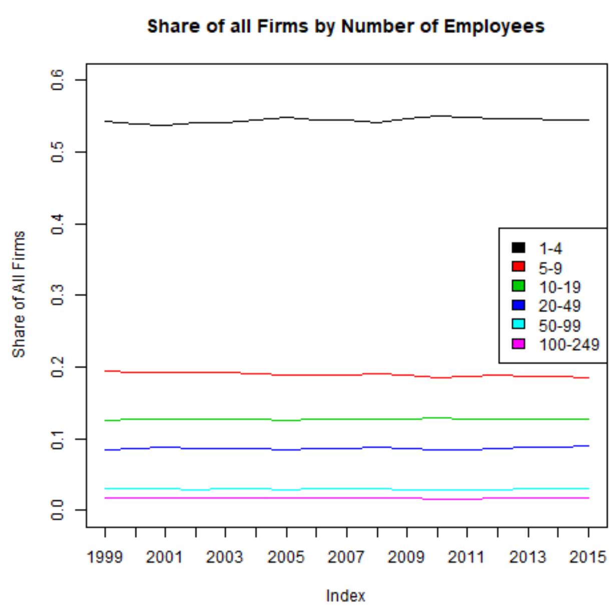

Figure 2.1 Share of Firms by Number of Employees

Data compiled from Census's County Business Patterns. Data shows share of establishments at different sizes from 1999 to 2015.

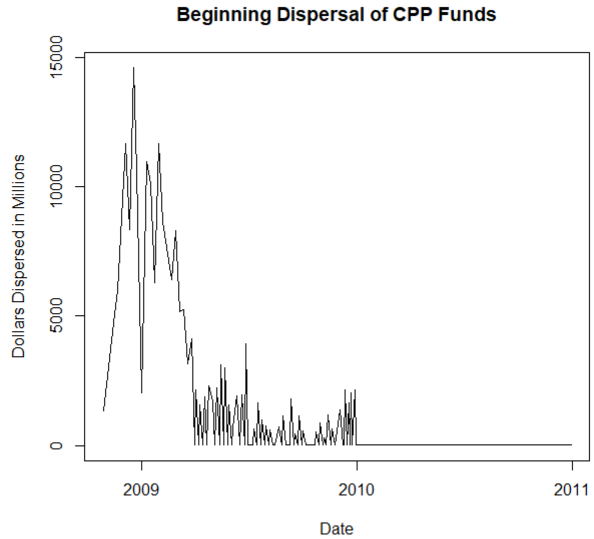

Figure 2.2 Dispersal of CPP Funds 2008-2009

This figure shows the dispersal of CPP funds to banks across the US by date of Treasury to bank transaction listed in the TARP Transaction Report.

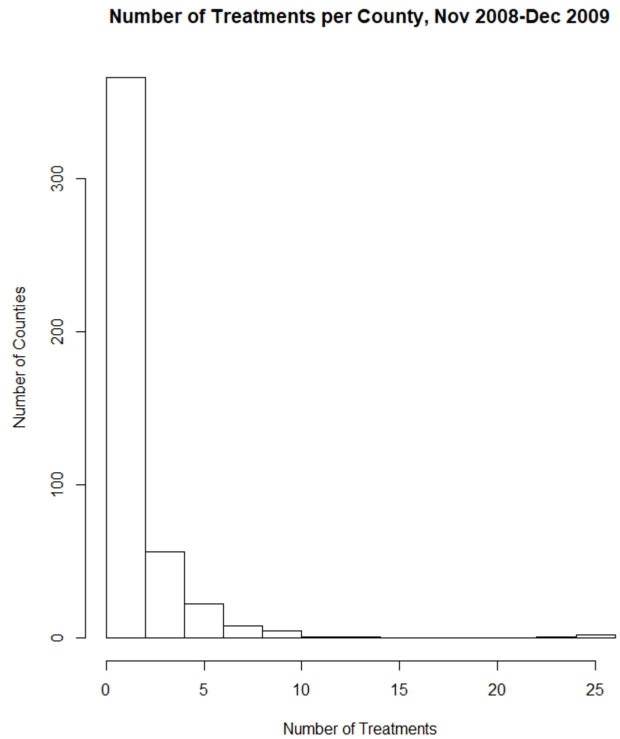

Figure 2.3 Number of Banks that Received CPP Funds Among Counties that Received CPP Funds

Data compiled from Treasury CPP Transaction Reports. Shows among counties how many banks in a given county received treatment. The presence of New York City, New York, is a clear outlier, from otherwise highly bunched few-treatments-per-county among the reamining sample.

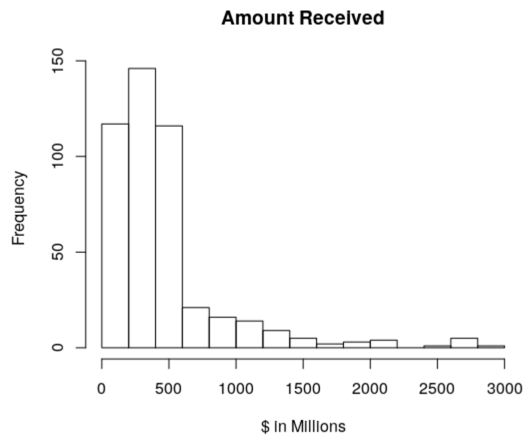

Figure 2.4 Amount Received Per Worker

Total CPP funds per county divided by 2008 labor force compiled from Treasury CPP Transaction Reports and BLS local area unemployment statistics. Does not exclude counties that had a Bank Holding Company headquarters.

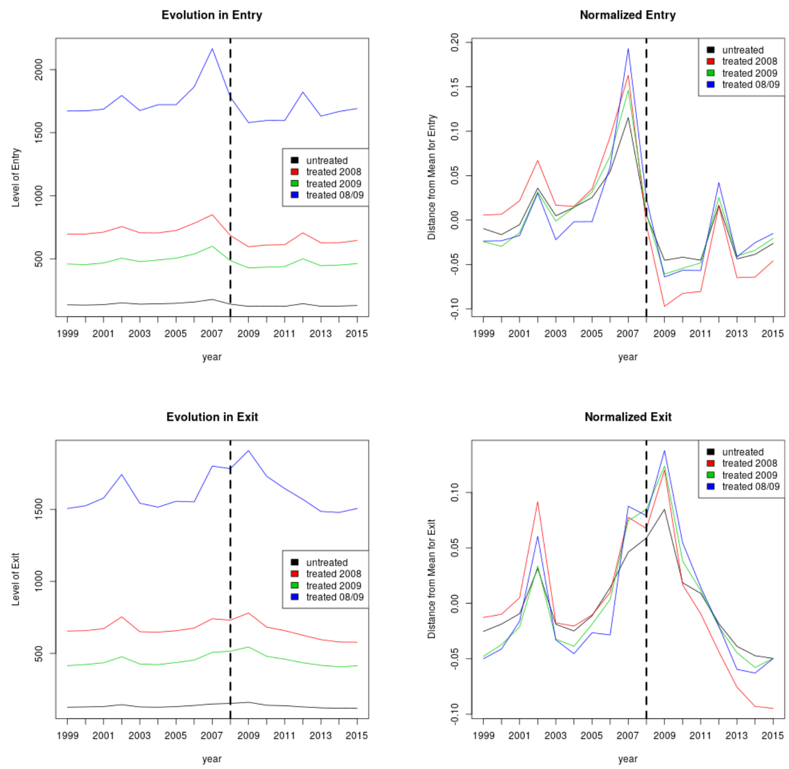

Figure 2.5 Subgroup Pre-Trends: Entry and Exit

The left column charts trends in establishment entry levels by treatment group- receiving treatment in both 2008 and 2009, receiving treatment in only 2008 or 2009, and not receiving treatment. The right hand column normalizes the series by pre-treatment group means and variances.

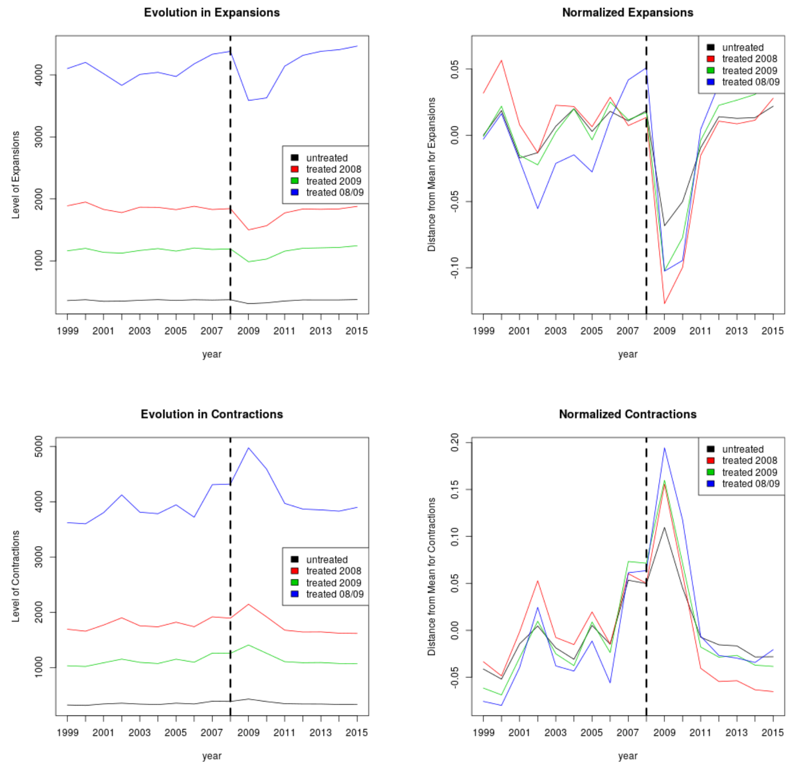

Figure 2.6 Subgroup Pre-Trends: Employment Expansion and Contraction

The left column charts trends in establishment entry levels by treatment group- receiving treatment in both 2008 and 2009, receiving treatment in only 2008 or 2009, and not receiving treatment. The right hand column normalizes the series by pre-treatment group means and variances.

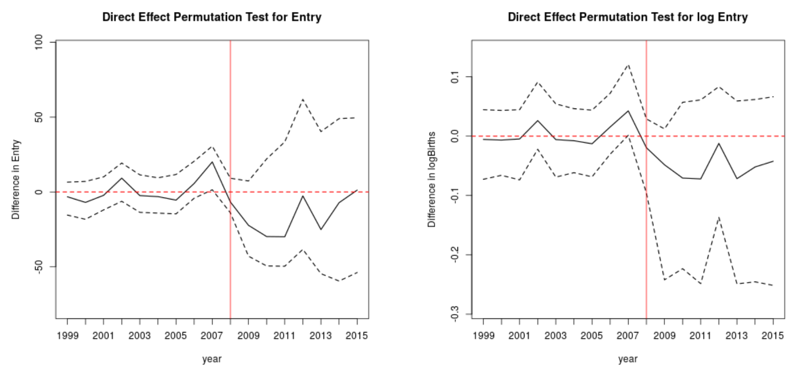

Figure 2.7 Direct Effect Establishment Entry

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

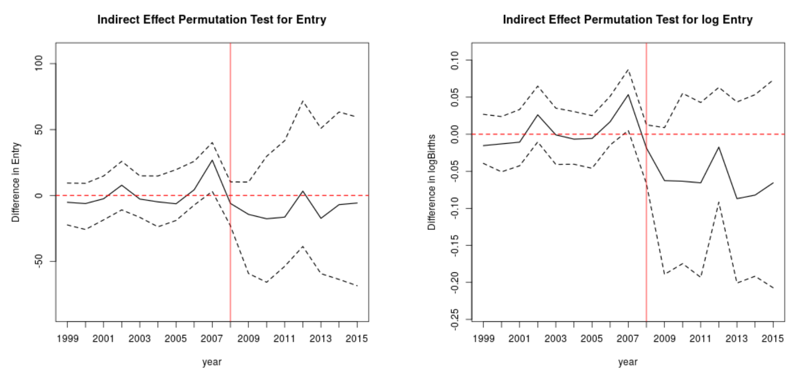

Figure 2.8 Indirect Effect Establishment Entry

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

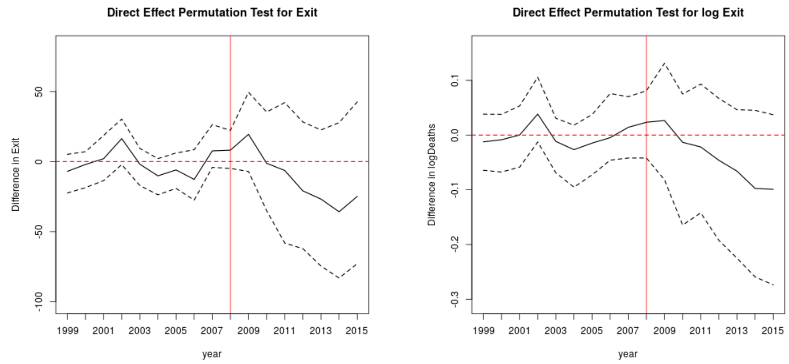

Figure 2.9 Direct Effect Establishment Exit

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

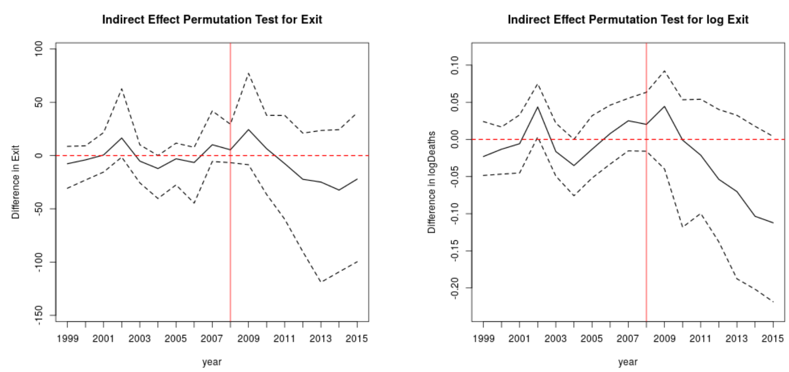

Figure 2.10 Indirect Effect Establishment Exit

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

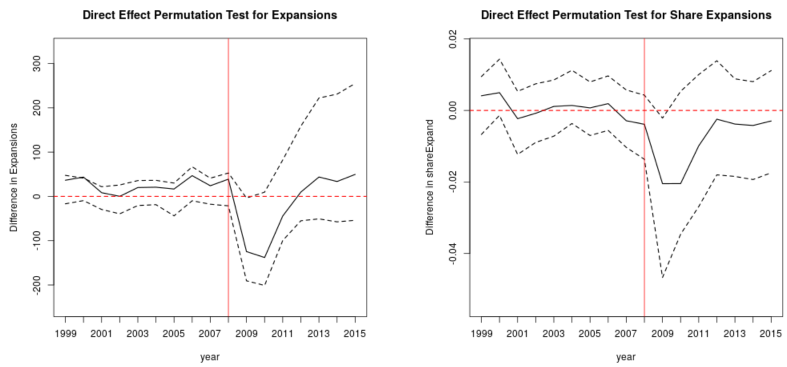

Figure 2.11 Direct Effect Employment Expansion

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

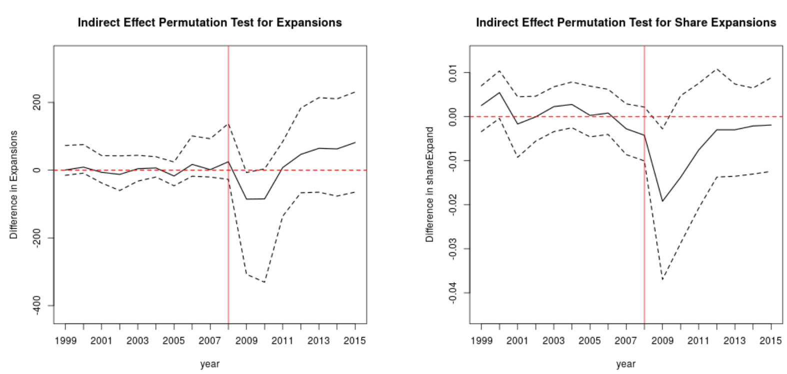

Figure 2.12 Indirect Effect Employment Expansion

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

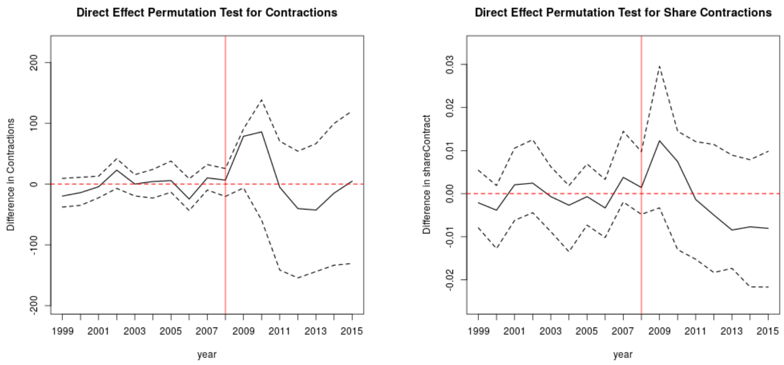

Figure 2.13 Direct Effect Employment Contraction

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

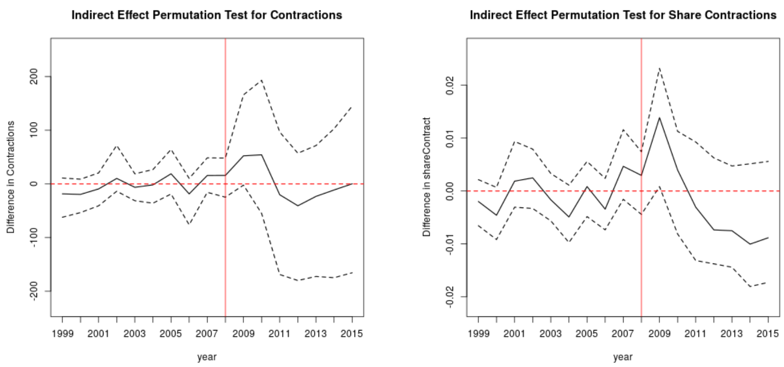

Figure 2.14 Indirect Effigfect Employment Contraction

LASSO-synthetic control estimates for the pooled effect of receiving treatment in either 2008 or 2009 in both levels and logs. Black line is the estimate for the the empirically observed set of treated counties, and the dashed black lines represent the 90% permutation test confidence intervals under the null hypothesis of no treatment.

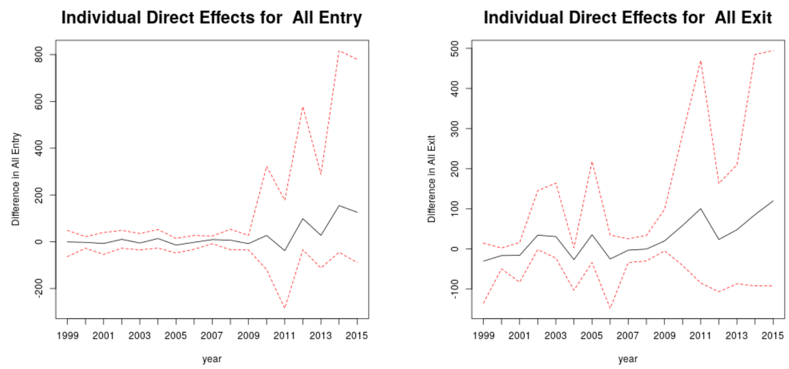

Figure 2.15 Heterogeneous Impacts: Entry & Exit

Black line is the mean effect among the empirically observed treatment group. Dashed black lines represent the 95% confidence interval among treated responses.

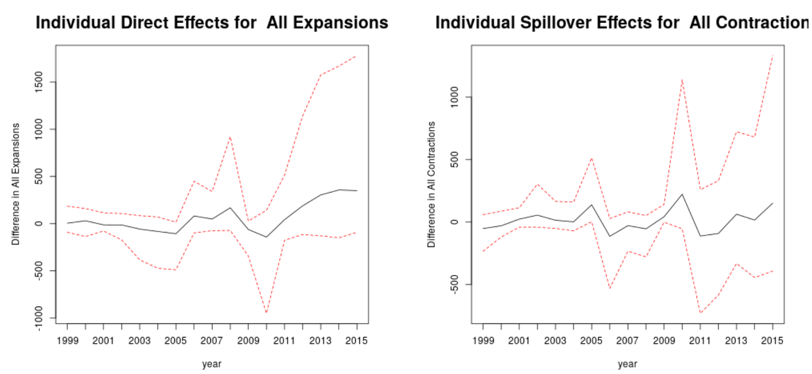

Figure 2.16 Heterogeneous Impacts: Expansions & Contractions

Black line is the mean effect among the empirically observed treatment group. Dashed black lines represent the 95% confidence interval among treated responses.

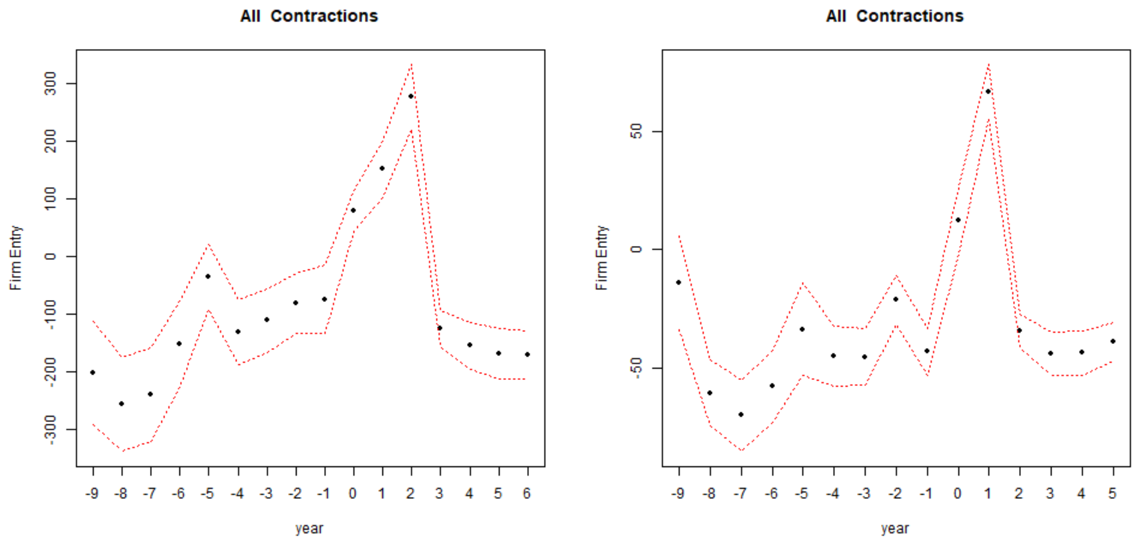

Figure 2.17 DID Own & Neighbor Treatment Status

Event study plot of pre-trends and post-treatment effects for a Difference-in-Differences two-way fixed effects regression with level of urbanization by time and Federal Reserve branch by time effects and shared treatment effect across time-of-treat subgroups.

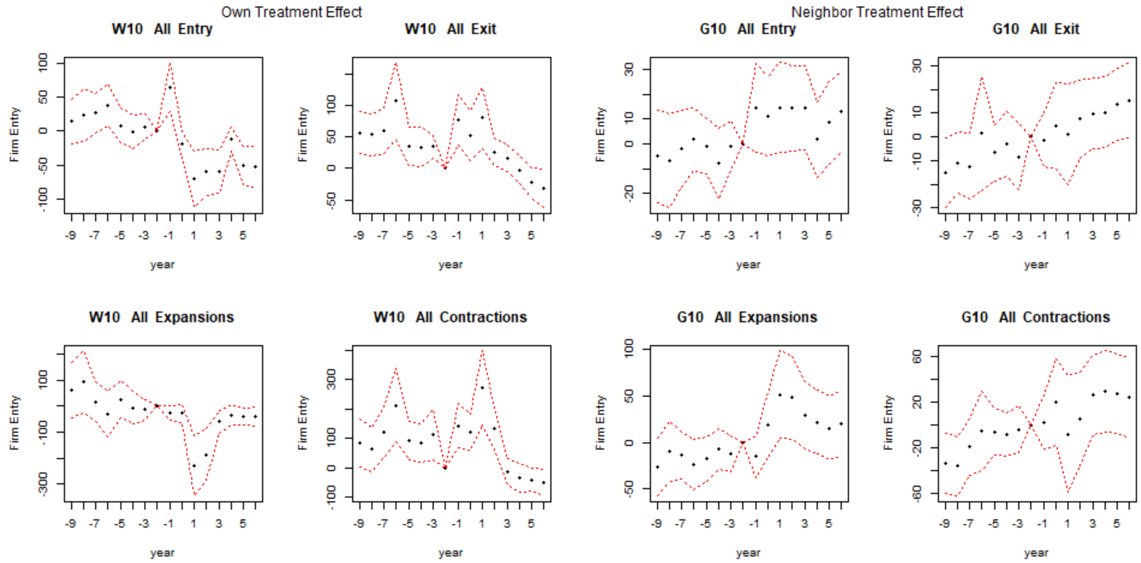

Figure 2.18 Own(1,0) & Neigh(1,0) Treatment Status

Event study plot of pre-trends and post-treatment effects for a Difference-in-Differences two-way fixed effects regression with level of urbanization by time and Federal Reserve branch by time effects and shared treatment effect for individuals who only received treatment, or have a neighbor receive treatment in 2008

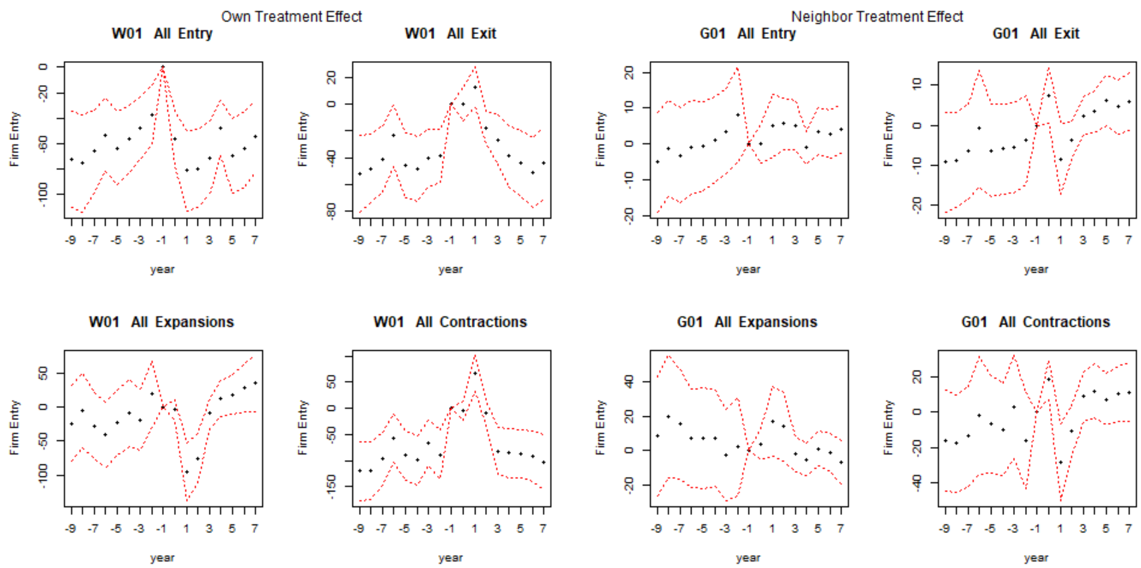

Figure 2.19 DID Own(0,1) & Neigh(0,1) Treatment Status

Event study plot of pre-trends and post-treatment effects for a Difference-in-Differences two-way fixed effects regression with level of urbanization by time and Federal Reserve branch by time effects and shared treatment effect for individuals who only received treatment, or have a neighbor receive treatment in 2009

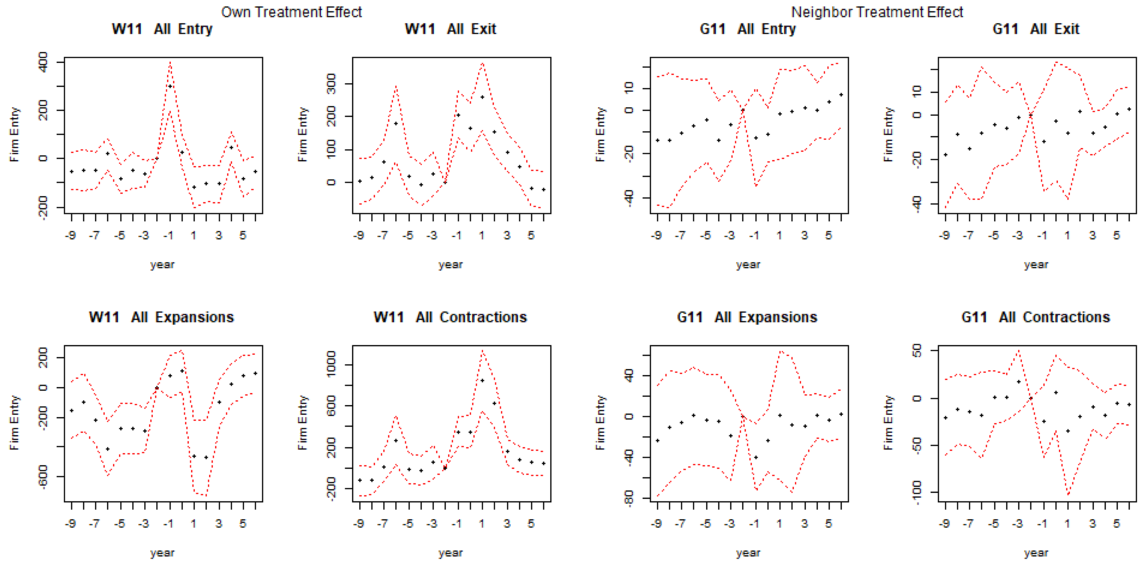

Figure 2.20 DID Own(1,1) & Neigh(1,1) Treatment Status

Event study plot of pre-trends and post-treatment effects for a Difference-in-Differences two-way fixed effects regression with level of urbanization by time and Federal Reserve branch by time effects and shared treatment effect for individuals who only received treatment, or have a neighbor receive treatment in both 2008 and 2009

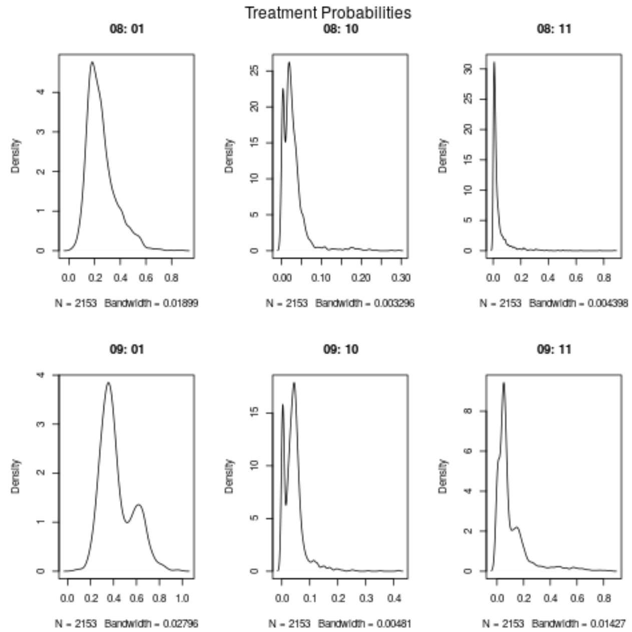

Figure 2.21 Bivariate Probit Propensity Scores

Each row from left to right is the probability of only Own Treatment, only Neighbor Treatment, or Both Treatment in either 2008 (top row) or 2009 (bottom row) based on estimating bivariate probits in 2008 and 2009 on a set of 4 instruments of county level political connections plus additional exogenous variables.

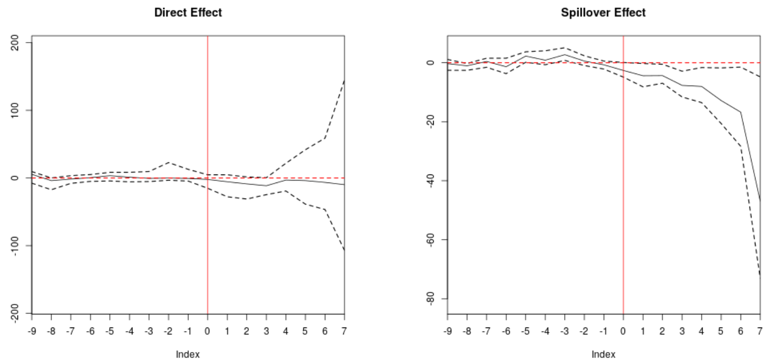

Figure 2.22 Interactive Fixed Effects Difference-in-Differences Entry

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009.

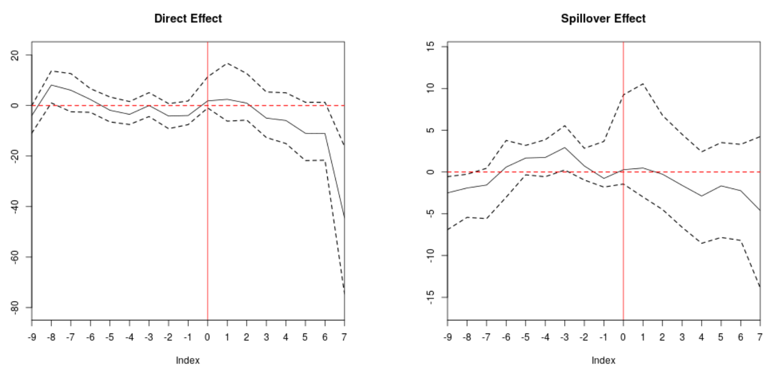

Figure 2.23 Interactive Fixed Effects Difference-in-Differences Exit

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009.

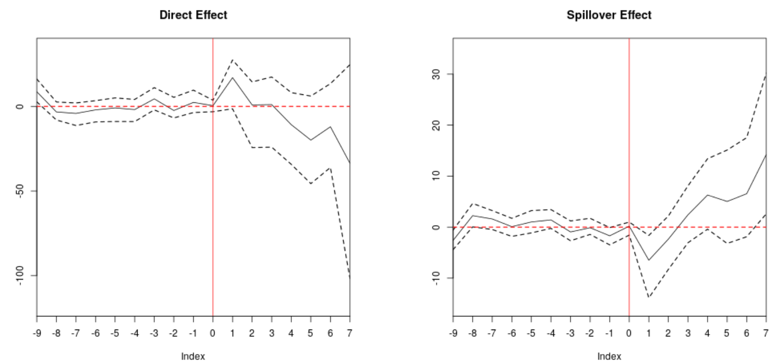

Figure 2.24 Interactive Fixed Effects Difference-in-Differences Expansions

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009.

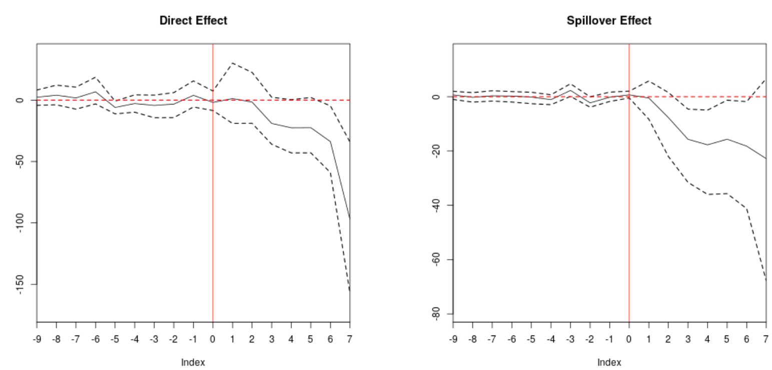

Figure 2.25 Interactive Fixed Effects Difference-in-Differences Contractions

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009.

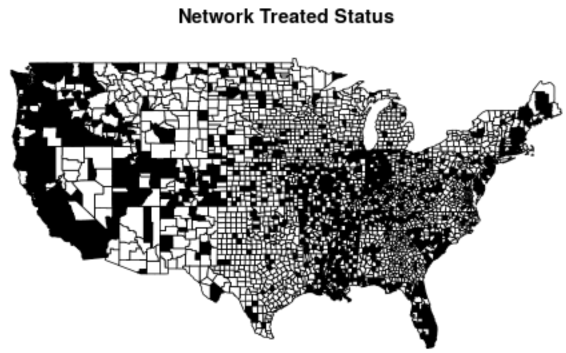

Figure 2.26 Treated Downstream Counties

Map of all counties with a branch location of a bank that received CPP funds in either 2008 or 2009.

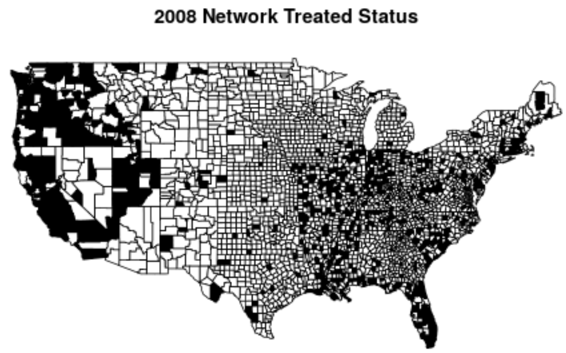

Figure 2.27 Treated Downstream Counties 2008

Map of all counties with a branch location of a bank that received CPP funds in 2008

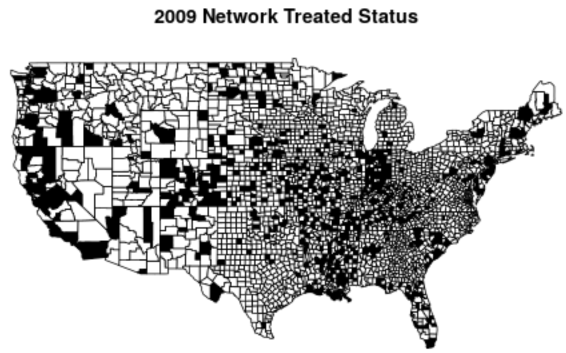

Figure 2.28 Treated Downstream Counties 2009

Map of all counties with a branch location of a bank that received CPP funds in 2009.

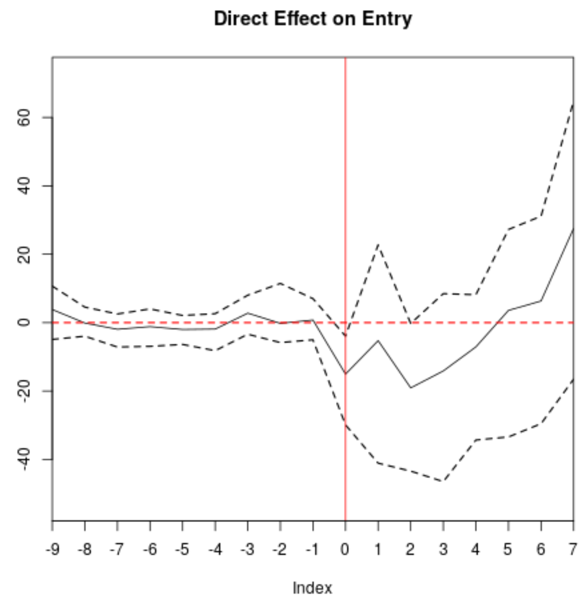

Figure 2.29 Interactive Fixed Effects Difference-in-Differences Network Entry ATT

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009. Identifies all counties that had a bank from all branch locations of treated bank treated.

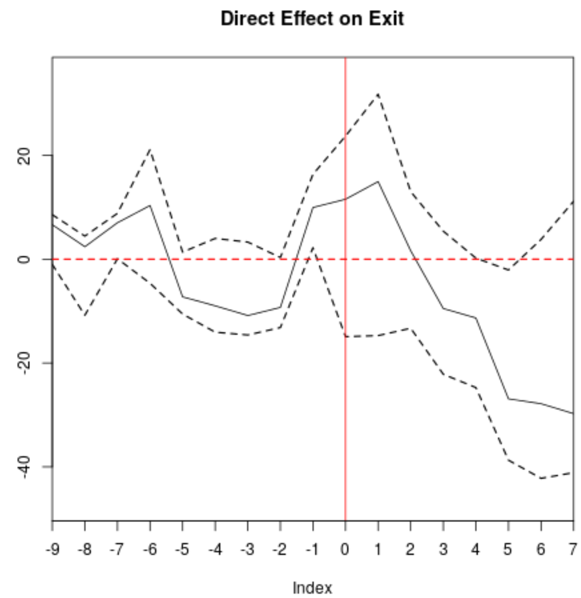

Figure 2.30 Interactive Fixed Effects Difference-in-Differences Network Exit ATT

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009. Identifies all counties that had a bank from all branch locations of treated bank treated.

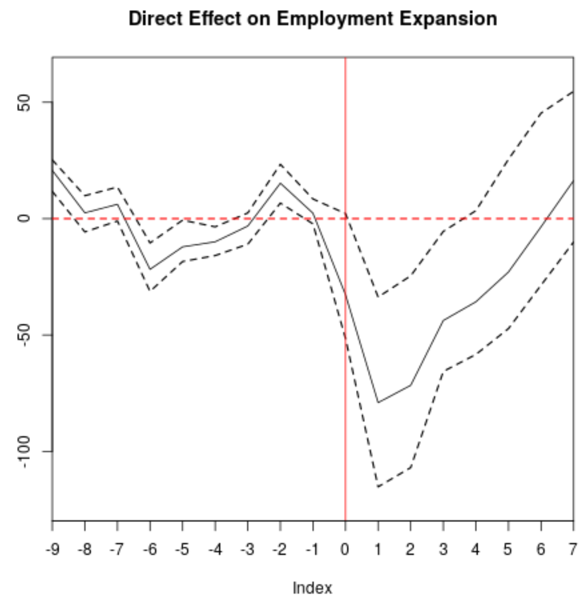

Figure 2.31 Interactive Fixed Effects Difference-in-Differences Network Employment Expansions ATT

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009. Identifies all counties that had a bank from all branch locations of treated bank treated.

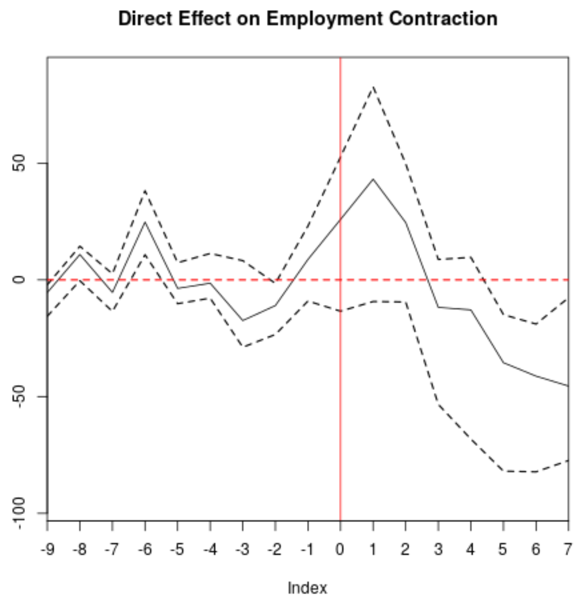

Figure 2.32 Interactive Fixed Effects Difference-in-Differences Network Employment Contractions ATT

Treatment effect for time-from-treated. Estimating using Interactive Fixed Effects Difference-in-Differences model based on pooled treatment effect across counties that received treatment in 2008 or 2009. Identifies all counties that had a bank from all branch locations of treated bank treated.

2.8.2 Tables

Table 2.1 Summary Statistics of Data

	PrGFC	PoGFC	Diff	prGFC SD	PoGFC SD	SD Diff
Firm Entry	266.356	240.067	-26.289	730.369	681.613	-48.756
Firm Exit	238.954	242.725	3.771	659.948	659.921	-0.027
Emp. Expansion	638.102	630.029	-8.072	1,614.693	1,575.101	-39.592
Emp. Contraction	604.360	637.556	33.196	1,539.482	1,565.769	26.287
Unemp. Rate	5.088	7.483	2.394	1.769	2.754	0.985
Neighbor Unemp. Rate	5.157	7.549	2.392	1.470	2.487	1.016
Troubled Asset Ratio	0.028	0.018	-0.009	0.076	0.056	-0.020
Neigh. Troubled Asset Ratio	0.029	0.019	-0.010	0.046	0.028	-0.017
Return on Assets	0.457	0.554	0.097	0.523	4.689	4.166
Neigh. Return on Assets	0.444	0.561	0.117	0.330	2.180	1.850
Loans to Deposits	52.320	60.262	7.942	49.362	38.405	-10.957
Neigh. Loans to Deposits	50.671	58.118	7.447	34.230	20.979	-13.252
HPI Change	4.820	-0.496	-5.317	4.465	4.623	0.159
HPI	228.467	255.897	27.430	125.742	132.322	6.579

Table 2.2 Wald Tests for Model 1 and NAICS code ..

Pretrend	Significant
Entry All Treated	No Shared Pretrend
Entry Own Treated	No Shared Pretrend
Entry Neigh Treated	No Shared Pretrend
Exits All Treated	No Shared Pretrend
Exits Own Treated	No Shared Pretrend
Exits Neigh Treated	No Shared Pretrend
Expansions All Treated	No Shared Pretrend
Expansions Own Treated	No Shared Pretrend
Expansions Neigh Treated	No Shared Pretrend
Contractions All Treated	No Shared Pretrend
Contractions Own Treated	No Shared Pretrend
Contractions Neigh Treated	No Shared Pretrend

No shared pretrend implies a p-value less than 0.005

Table 2.3 Step Down Tests for Non-Zero ATT Following 10 Treatment

stepDownNames	own.diff.sig	neigh.diff.sig
Entry	Effect for 5 Time periods	No Effect
Exit	Effect for 2 Time periods	No Effect
Expansions	Effect for 2 Time periods	No Effect
Contractions	Effect for 7 Time periods	No Effect

Table 2.4 Step Down Tests for Non-Zero ATT Following 01 Treatment

stepDownNames	own.diff.sig	neigh.diff.sig
Entry	No Effect	No Effect
Exit	Effect for 2 Time periods	No Effect
Expansions	Effect for 2 Time periods	No Effect
Contractions	Effect for 2 Time periods	No Effect

Table 2.5 Step Down Tests for Non-Zero ATT Following 11 Treatment

stepDownNames	own.diff.sig	neigh.diff.sig
Entry	No Effect	No Effect
Exit	No Effect	No Effect
Expansions	No Effect	No Effect
Contractions	Effect for 7 Time periods	Effect for 3 Time periods

Table 2.6 Wald Tests for IV Pretrend

Pretrend	Significant
Entry Own Treated 2008	No Shared Pretrend
Entry Neigh Treated 2008	No Shared Pretrend
Entry Own Treated 2009	No Shared Pretrend
Entry Neigh Treated 2009	No Shared Pretrend
Entry All	No Shared Pretrend
Exit Own Treated 2008	No Shared Pretrend
Exit Neigh Treated 2008	No Shared Pretrend
Exit Own Treated 2009	No Shared Pretrend
Exit Neigh Treated 2009	No Shared Pretrend
Exit All	No Shared Pretrend
Expansions Own Treated 2008	No Shared Pretrend
Expansions Neigh Treated 2008	No Shared Pretrend
Expansions Own Treated 2009	No Shared Pretrend
Expansions Neigh Treated 2009	No Shared Pretrend
Expansions All	No Shared Pretrend
Contractions Own Treated 2008	No Shared Pretrend
Contractions Neigh Treated 2008	No Shared Pretrend
Contractions Own Treated 2009	No Shared Pretrend
Contractions Neigh Treated 2009	No Shared Pretrend
Contractions All	No Shared Pretrend

No shared pretrend implies a p-value less than 0.005

2.8.3 Dropped Counties

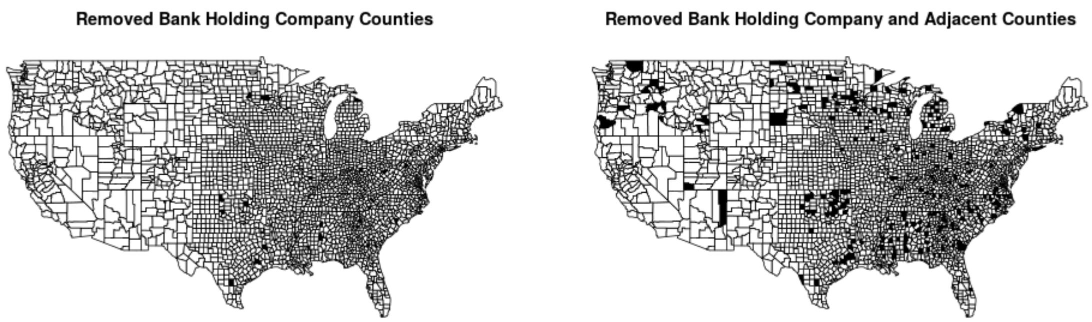

Figure 2.33 Removed Bank Holding Company Counties

Left map are counties that had the top 20 largest banks or bank holding companies in them. The right map are all counties that had a county centroid within 50 miles of a county that had one of the largest banks or bank holding companies. All these counties are dropped from our sample.

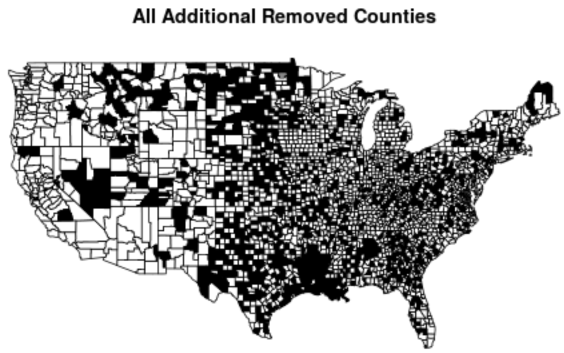

Figure 2.34 Additional Removed Counties

All counties that are dropped for a variety of reasons. This includes being an unbalanced panel in our data set, not having enough loans to register in the FHFA's county level home price index, or having zero new establishment entrants or establishment exits for at least one period from 1999 to 2015. These counties are only dropped in our robustness checks that require additional covariates.

2.9 References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1):113–132.

Agarwal, S. and Hauswald, R. B. H. (2011). Distance and Private Information. *SSRN Electronic Journal*, 23(7):2757–2788.

Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of Machine Learning Research*, 19:1–51.

Arellano, M. (1987). PRACTITIONERS' CORNER: Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, 49(4):431–434.

Bartelsman, E., Scarpetta, S., and Schivardi, F. (2005). Comparative analysis of firm demographics and survival: Evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14(3):365–391.

Bassett, W., Demiralp, S., and Lloyd, N. (2017). Government support of banks and bank lending. *Journal of Banking and Finance*, 7(0):25–0.

Bauer, P. and Hackl, P. (1987). Multiple Testing in a Set of Nested Hypotheses. *Statistics*, 18(3):345–349.

Berger, A. N. (2018). The Benefits and Costs of the TARP Bailouts: A Critical Assessment. *Quarterly Journal of Finance*, 8(2):1850011.

Berger, A. N., Makaew, T., and Roman, R. A. (2019a). Do Business Borrowers Benefit from Bank Bailouts?: The Effects of TARP on Loan Contract Terms. *Financial Management*, 48(2):575–639.

Berger, A. N. and Roman, R. A. (2016). Did TARP Banks Get Competitive Advantages? *Journal of Financial and Quantitative Analysis*, 50(6):1199–1236.

Berger, A. N. and Roman, R. A. (2017). Did Saving Wall Street Really Save Main Street? The Real Effects of TARP on Local Economic Conditions. *Journal of Financial and Quantitative Analysis*, 52(5):1827–1867.

Berger, A. N., Roman, R. A., and Sedunov, J. (2019b). Did TARP reduce or increase systemic risk? The effects of government aid on financial system stability. *Journal of Financial Intermediation*.

Berger, A. N. and Udell, G. F. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *Economic Journal*, 112(477):F32–F53.

Black, L. K. and Hazelwood, L. N. (2013). The effect of TARP on bank risk-taking. *Journal of Financial Stability*, 9(4):790–803.

Blau, B. M., Brough, T. J., and Thomas, D. W. (2013). Corporate lobbying, political connections, and the bailout of banks. *Journal of Banking and Finance*, 37(8):3007–3017.

Calabrese, R., Degl'Innocenti, M., and Osmetti, S. A. (2017). The effectiveness of TARP-CPP on the US banking industry: A new copula-based approach. *European Journal of Operational Research*, 256(3):1029–1037.

Callaway, B. and Sant'Anna, P. H. C. (2018). Difference-in-Differences With Multiple Time Periods and an Application on the Minimum Wage and Employment. *SSRN Electronic Journal*.

Calomiris, C. W. and Khan, U. (2015). An Assessment of TARP assistance to financial institutions. *Journal of Economic Perspectives*, 29(2):53–80.

Cao, J. and Dowd, C. (2018). Inference in Synthetic Controls with Spillover Effects. *Working Paper*.

Carow, K. A. and Salotti, V. (2014). The U.S. treasury's capital purchase program: Treasury's selectivity and market returns across weak and healthy banks. *Journal of Financial Research*, 37(2):211–241.

Carvalho, C., Masini, R., and Medeiros, M. C. (2018). ArCo: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of Econometrics*, 207(2):352–380.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2017). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls.

Clementi, G. L. and Palazzo, B. (2016). Entry, exit, firm dynamics, and aggregate fluctuations. *American Economic Journal: Macroeconomics*, 8(3):1–41.

Cole, R. A. (2012). How Did the Financial Crisis Affect Small-Business Lending in the U.S.? *SSRN Electronic Journal*.

Contessi, S. and Francis, J. L. (2011). TARP beneficiaries and their lending patterns during the financial crisis. *Federal Reserve Bank of St. Louis Review*, 93(2):105–125.

Cowan, A. R. and Salotti, V. (2015). The resolution of failed banks during the crisis: Acquirer performance and FDIC guarantees, 2008-2013. *Journal of Banking and Finance*, 54:222–238.

Davis, S. J. and Haltiwanger, J. C. (2019). Dynamism Diminished: The Role of Housing Markets and Credit Conditions.

Degryse, H. and Ongena, S. (2005). Distance, lending relationships, and competition. *Journal of Finance*, 60(1):231–266.

Di Stefano, R. and Mellace, G. (2020). The inclusive synthetic control method.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. Technical report, National Bureau of Economic Research, Cambridge, MA.

Duchin, R. and Sosyura, D. (2014). Safer ratios, riskier portfolios: Banks' response to government aid. *Journal of Financial Economics*, 113(1):1–28.

Duncan, K. D. (2015). *Impacts of Taxes on Firm Entry along State Borders A Pseudo-Regression Discontinuity Approach Table of Contents*. PhD thesis, Iowa State University, Digital Repository, Ames.

Ferman, B. and Pinto, C. (2016). Revisiting the Synthetic Control Estimator.

Gardeazabal, J. and Vega-Bayo, A. (2017). An Empirical Comparison Between the Synthetic Control Method and HSIAO et al.'s Panel Data Approach to Program Evaluation. *Journal of Applied Econometrics*, 32(5):983–1002.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. Technical Report 25018, National Bureau of Economic Research, Cambridge, MA.

Harris, O., Huerta, D., and Ngo, T. (2013). The impact of TARP on bank efficiency. *Journal of International Financial Markets, Institutions and Money*, 24(1):85–104.

Holmes, T. J. (1998). The effect of state policies on the location of manufacturing: evidence from state borders. *Journal of Political Economy*, 106(4):667–705.

Hsiao, C., Steve Ching, H., and Ki Wan, S. (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with Mainland China. *Journal of Applied Econometrics*, 27(5):705–740.

Huber, M. and Steinmayr, A. (2019). A Framework for Separating Individual-Level Treatment Effects From Spillover Effects. *Journal of Business and Economic Statistics*, pages 1–39.

Hurst, E. and Lusardi, A. (2004). Liquidity constraints, household wealth, and entrepreneurship. *Journal of Political Economy*, 112(2):319–347.

Imai, K. and Kim, I. S. (2014). On the Use of Linear Fixed Effects Regression Estimators for Causal Inference.

Jang, K. Y. (2017). The effect of TARP on the propagation of real estate shocks: Evidence from geographically diversified banks. *Journal of Banking and Finance*, 83:173–192.

Kahn-Lang, A. and Lang, K. (2019). The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications. *Journal of Business and Economic Statistics*, pages 1–14.

Kaniovski, S. and Peneder, M. (2008). Determinants of firm survival: A duration analysis using the generalized gamma distribution. *Empirica*, 35(1):41–58.

Lee, Y. and Mukoyama, T. (2015). Productivity and employment dynamics of US manufacturing plants. *Economics Letters*, 136:190–193.

Li, K. T. and Bell, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, 197(1):65–75.

Li, L. (2013). TARP funds distribution and bank loan supply. *Journal of Banking and Finance*, 37(12):4777–4792.

Mata, J. and Portugal, P. (1994). Life Duration of New Firms. *The Journal of Industrial Economics*, 42(3):227.

Parker, S. C. (2005). The economics of entrepreneurship: What we know and what we don't. *Foundations and Trends in Entrepreneurship*, 1(1):1–54.

Riding, A. L. and Haines, G. (2001). Loan guarantees: Costs of default and benefits to small firms. *Journal of Business Venturing*, 16(6):595–612.

Rohlin, S., Rosenthal, S. S., and Ross, A. (2014). Tax avoidance and business location in a state border model. *Journal of Urban Economics*, 83:34–49.

Roth, J. (2018). Should We Adjust for the Test for Pre-trends in Difference-in-Difference Designs? *arXiv Working Paper arXiv:1804.01208v2*.

Ruonan Xu (2019). Weak Instruments with a Binary Endogenous Explanatory Variable.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.

Shane, S. A. (2010). *The illusions of entrepreneurship: The costly myths that entrepreneurs, investors, and policy makers live by.* Yale University Press.

Sheng, J. (2015). The Real Effects of Government Intervention: Firm-Level Evidence from TARP. *SSRN Electronic Journal*.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in Linear Iv regression. Technical report, National Bureau of Economic Research, Cambridge, MA.

Veronesi, P. and Zingales, L. (2010). Paulson's Gift. *journal of financial economics*, 97(3):339–368.

Wan, S. K., Xie, Y., and Hsiao, C. (2018). Panel data approach vs synthetic control method. *Economics Letters*, 164:121–123.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

CHAPTER 3. LINEAR HYPOTHESIS TESTS OVER FIXED EFFECTS WITH SERIALLY CORRELATED PANELS

Kevin D. Duncan

Iowa State University

Modified from a manuscript to be submitted to *Quantitative Economics*

Abstract

This paper develops a joint hypothesis test over fixed effects in large N small T panel data models with symmetric serial correlation within cross sectional observations. The enables joint hypothesis tests over the traditional varying intercept model as well as models with individual specific slope coefficients. I establish two different set of assumptions where feasible tests exist. The first assumption requires that individual errors follow a stationary $AR(p)$ process, allowing for individual specific restrictions. The second tests individuals to be in a known grouping structure under the null. Examples of these tests arise in wanting to establish latent panel structure, such as unobserved grouping of individuals, wanting to compare different models of teacher or firm value added against each other, or testing whether or not fixed effects can be approximated by Mundlak-Chamberlain devices.

3.1 Introduction

Consider the following panel data model;

$$y_{it} = x_{it}'\beta + z_{it}'\gamma_i + \epsilon_{it}; \quad i = 1, ..., n; \; t = 1, ..., T$$

$$\mathrm{E}(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i' \mid \mathbf{X}_i \; \mathbf{Z}_i) = \Omega \quad \forall i; \boldsymbol{\epsilon}_i = [\epsilon_{i1} \ldots \epsilon_{iT}]$$

$$\mathrm{E}(\epsilon_{it}\epsilon_{js} \mid \mathbf{X}_i \; \mathbf{Z}_i \; X_j \; Z_j) = 0 \quad \forall j \neq i; \; \forall t, s = 1, ..., T$$

where x_{it} is a $K \times 1$ vector of regressors whose coefficients are shared across individuals, and z_{it} is a $L \times 1$ vector of regressors whose coefficient values vary across individuals, often called fixed effects. This model is quite flexible, and includes both the usual additive fixed effects model when $z'_{it} = 1$ for all individuals, as well as additional regressors, including either individual specific time trends (Hansen, 2007; Wooldridge, 2005) or stochastic regressors (Arellano and Bonhomme, 2012). In this paper I develop a Wald test for joint hypotheses over the fixed effects γ_i in large N small T panel data models when the error process features symmetric serial correlation among individuals in the sample. This test is among the first to be robust to both first stage estimation error and serial correlation in the underlying error process.

When T remains small hypothesis tests on individual fixed effects γ_i cannot be carried out, but joint restrictions that grow with the sample size, such as $H_0 : \gamma_i = 0$, provide a meaningful way to explore underlying heterogeneity in short T panels. This paper proves the asymptotic normality of a feasible centered Wald tests using OLS residuals under two different assumptions. The first set of assumptions generate a valid test for whether or not individual fixed effects are the same for all individuals in a group. Under a known group structure and shared covariate values across individuals there exist a feasible test under an unknown non-stationary serial correlation.[1] The second set of assumptions allow for a general set of linear hypotheses with individual and time varying covariates if the errors follow a stationary $AR(p)$ process.

Hypotheses of either form arise in many economic models, such as testing for homogeneity of returns to education (Heckman and Vytlacil, 1998), teacher value added (Kane et al., 2008; Chetty et al., 2017), or country production functions (Durlauf et al., 2001). Many researchers estimate models without allowing for parameter heterogeneity, and the broadest class of models-where parameters vary both across individuals and time cannot be estimated without auxiliary assumptions.[2] Single coefficient models are misspecified when the underlying population features

[1]The test is indifferent to the source of knowledge on the group coefficient. This enables the test to be used with both fixed number of groups or growing number of groups, as long as the number of individuals per group is greater than 2.

[2]In the teacher value added literature, Chetty et al. (2017) assume that teacher value added in period t can be estimated from class mean test scores in the previous periods. This allows estimation of individual and time varying teacher value added.

individual or group varying effects, and ordinary least squares may not recover the mean effect (Verbeek and Nijman, 1992; Heckman and Vytlacil, 1998).

A few notable examples appear here. First, researchers might be interested in testing latent panel structure. In this framework, researchers might believe that there is a set of groups, $g = 1, \ldots, G$, and know through auxiliary information, or theory, that for every individual in the group g, $\gamma_i = \gamma_g$. As shown later in the paper, the proposed test does not require a consistent estimator for γ_g. In large N, T panels, recent work has allowed researchers to explore latent panel structure using augmented LASSO methods (Phillips and Moon, 1999; Lin and Ng, 2012; Su et al., 2014). This allows for joint estimation of group assignment and group-specific fixed slopes. My test provides a heuristic way of testing latent panel structure without large T asymptotics. A second example lies in testing if individual varying coefficients are equivalent to an auxiliary model. For example, the Mundlak device defines $\gamma_i = c + \bar{X}_i{}'\alpha + u_i$ (Mundlak, 1978). Traditionally, the focus has been on testing whether or not $\alpha = 0$, however, tests for whether or not this approximation is correct was previously impossible.[3]

To estimate population moments, I develop two sets of assumptions under which consistent estimators exist. The first set of assumptions allows for a general set of linear restrictions with the errors following a stationary $AR(p)$ process.Hansen (2007) develops bias correction methods for FGLS estimation for panel data when errors follow a stationary $AR(p)$ process under both fixed and increasing T panels. The second set of assumptions assumes that there is a known grouping of individuals, such that population moments can be estimated from taking the difference between the estimated errors for two individuals in the same group. Hausman and Kuersteiner (2008) show consistent estimation of a non-stationary variance covariance matrix for panels with additive unobserved heterogeneity. This technique has been used elsewhere, but most notably in matching estimators Hanson and Sunderam (2009). Cattaneo et al. (2018b) discuss general issues in estimation of general variance-covariance matrices without symmetry across individuals when

[3]The overall use of the Mundlak device is still applicable even if the functional form is violated. Rejecting the null that $\alpha = 0$ is still a sign of correlation between X_{it} and time invariant unobservables. Policy and welfare extensions may still rely on the functional form being correct.

the number of parameters is growing. Much of the previous literature has focused on estimation of β as the dimension of additional regressors grows, such as Stock and Watson (2009); Belloni et al. (2013); Cattaneo et al. (2018a).

Historically γ_i were treated as "nuisance parameters" due to the inability to calculate consistent estimators (Nerlove, 1971; Nickell, 1981) and swept away by the traditional within transformation or first differencing. Instead this paper refocuses efforts to test hypotheses over these individual fixed effects (Wooldridge, 2010). Comparably, joint hypothesis over fixed effects lead to a non-negligible number of regressors relative to the number of observations. Swamy (1970) develops tests for random coefficient models, Hashem Pesaran and Yamagata (2008) and Blomquist and Westerlund (2013) extends this to a more general class of large N, T panels. Boos and Brownie (1995) study the behavior of ANOVA tests when the number of levels increases with the sample size. Akritas and Papadatos (2004) extend this framework to allow for heteroskedasticity, non-normality with fixed effects (Bathke, 2004). For regression models, Calhoun (2011) develops an F test for non-normal but homoskedastic data as the number of regressors grows with the sample size. Equivalently, Anatolyev (2012) explores the behavior of the F, LM, and LR tests with homoskedastic Gaussian errors. Orme and Yamagata (2006, 2014) develop F-tests for whether or not the fixed effects for the linear additive model are equal to the pooled OLS constant with either homoskedastic, or heteroskedastic but still serially uncorrelated errors.

The usefulness of the test is seen by how it compares to previous work in the literature.Orme and Yamagata (2006, 2014) explicitly test the null hypothesis that $\gamma_i = \gamma$, the later being the pooled OLS constant. Comparably my test is applicable for a more general set of plausible linear hypotheses, including group specific effects, or alternative models of individual effects. For the later, examples include the Mundlak device, or alternative models of teacher value added, such as developed by Chetty et al. (2014). The resulting class of models that my test can be applied to is also more general, and common among the Correlated Random Coefficients literature (Wooldridge, 2005; Arellano and Bonhomme, 2012).

Secondly, comparable tests do not allow for serial correlation (Calhoun, 2011; Anatolyev, 2012). These tests focus on the specific case when the dimension of β grows, relative to incidental parameters that might vary across individuals, and use homoskedastic properties of the errors to develop asymptotitic corrections to recover the true parameters. My method rests on similar methodology, but requires new techniques to estimate population moments because of singularities induced in their methods under common hypotheses of interest. Secondly, in panel settings, considerable attention has been paid to the development of methods that are robust to serial correlation (Bertrand et al., 2004; Hansen, 2007), but there exist no joint hypothesis test over incidental parameters that is robust to serial correlation. Kline et al. (2018) offers some discussion of estimation of variance components with a leave-out estimator, but this is focused on the variance of the covariates, not hypothesis tests.

The paper proceeds in the following manner. Section 3.2 outlines notation and assumptions, and defines the centered Wald test. Section 3.3 develops feasible estimates for the unknown population moments under two sets of assumptions and first stage estimation error. Section 3.4 proves the asymptotic normality of the feasible centered Wald test. Section 3.5 provides Monte Carlo evidence of the size and power of our proposed tests. Section 3.6 concludes.

3.2 Assumptions and Notation

This section describes the estimating equation, what hypothesis researchers are interested in testing, and a baseline set of assumptions required to establish the asymptotic behavior for a centered Wald test. For any $B \times T$ matrix b define the within transformation to be, $\ddot{b} = (I_T - Z_i(Z_i'Z_i)^{-1}Z_i')b$, and define $P_A = A(A'A)^{-1}A'$, such that $P_{A,ij,ts} = A_{i,t}(A'A)^{-1}A_{j,s}'$ is 1×1. Individuals data generating process stacked by time period follows the equation,

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\gamma_i + \boldsymbol{\epsilon}_i; i = 1, \ldots, n$$

where $\mathbf{X}_i$ is a $T \times K$ vector of covariates whose coefficient is shared across all individuals, $\mathbf{Z}_i$ is a $T \times L$ vector of covariates whose coefficient varies across individuals. We assume that for each individual's errors process is symmetric, such that $\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid X_i \; Z_i) = \Omega$ for all individuals. Further stacking individual equations into a sample matrix representation we get,

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{\gamma}_n + \boldsymbol{\epsilon} \tag{3.1}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{Z}_n \end{bmatrix}$$

and $\boldsymbol{\gamma}_n = [\gamma_1 \; \gamma_2 \; \dots \; \gamma_n]$. The OLS estimators for β and $\boldsymbol{\gamma}_n$ are,

$$\hat{\beta} = (\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{Y}}$$

$$\hat{\gamma_n} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

The joint hypothesis of interest takes the form,

$$H_0 : \; R_n[\beta \; \boldsymbol{\gamma}_n] = r_n \tag{3.2}$$

Where R_n is a $q_n \times (K + nL)$ matrix, and r_n is $q_n \times 1$. Usually the tests impose testing a single linear restriction for each fixed effect estimate in a sample, e.g.

$$R_n = \begin{bmatrix} 0_{k,n} : & I_n \end{bmatrix} \quad r_n = [0 \dots 0]'$$

With outside information (r.e. model restrictions, latent panel structure, etc), r_n can be non-zero, or follow a more general set of linear restrictions imposed by R_n. The test of interest can be characterized as,

$$W_{n,OLS} = \frac{1}{q}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}^2_{OLS,it,0} - \hat{\epsilon}^2_{OLS,it}\right) \tag{3.3}$$

This test is similar to the regular Wald test, using the residual sum of squares from an unconstrained OLS regression of equation (3.1), and after imposing the null in Equation (3.2). To carry out this analysis, an independence across individuals assumption is imposed.

Assumption 3.2.1. $\{(\mathbf{X}_i\ \mathbf{Z}_i),\ \boldsymbol{\epsilon}_i\}$ *are i.i.d. across i.* $\mathrm{E}(\boldsymbol{\epsilon}_i \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = 0$.

The imposed assumptions are standard in the fixed effects literature (see for example Wooldridge (2010) section 11.7.2). This assumption implies independence across individuals along with strict exogeneity of our regressors, and rules out dynamic panel settings. No restrictions are placed on the relationship between (X_i, Z_i, γ_i), placing the analysis in a "fixed-effects" or "Correlated Random Coefficients" effects setting. Traditional higher order moment conditions are now imposed,

Assumption 3.2.2. $\mathrm{Rank}(\sum_{t=1}^{T} \mathrm{E}[\ddot{x}_{it}\ddot{x}_{it}'])= \mathrm{Rank}(\mathrm{E}[\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i]) = K, \quad \mathrm{Rank}(Z_i'Z_i) = L$ *with probability* 1 *and there exists a constant* Δ *such that* $\mathrm{E}[x_{ith}^4] \leq \Delta < \infty$ *for all* t, h*, and* $E[z_{ith}^4] \leq \Delta < \infty$ *with probability 1 for all* t, h. $\mathrm{E}(|\epsilon_{it}|^{4+r}) < \infty$

The first part of this assumption requires that $J < T$. For the usual additive intercept this implies $T \geq 2$, and with a time trend implies $T \geq 3$. The rank condition on $\mathbf{Z}_i$ is due to the fixed T asymptotics. Under small T we need that matrices and moments hold with probability 1, rather than in expectation. The rank condition in Assumption 3.2.2 implies that there can only be a single time-invariant regressor among $(\mathbf{X}_i,\ \mathbf{Z}_i)$. Finally, a general assumption on the error variance structure is imposed.

Assumption 3.2.3.

$$\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \Omega$$

$$\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \mid \mathbf{X}\ \mathbf{Z}\ \boldsymbol{\gamma}_n) = \Sigma = I_n \otimes \Omega$$

$$\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \mu_{stuv}$$

With $\mathrm{E}(\epsilon_{it}\epsilon_{is} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i) = \sigma_{t,s}$

This assumption has two parts. The first is that the variance-covariance matrix is symmetric across individuals in the sample. Further restrictions on the variance-covariance matrix are required to generate a feasible test under later assumptions, but for now no further structure on the second moments is required. The second is dealing with the serial correlation present in the higher moments. Throughout the paper define $\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i)$ as the fourth cross-product moments. Under stronger parametric assumptions the exact behavior of this term is known[4], however throughout the form of conditional serial correlation is also allowed to be unconstrained.

Matrix algebra provides an expression for the Wald statistic in Equation (3.3) as the difference in the residual sum of squares, which generates quadratic forms,

$$W_{n,OLS} = \frac{1}{q}(R_n[\beta\ \boldsymbol{\gamma}_n] - r_n)'(\mathbf{W}'\hat{\Sigma}^{-1}\mathbf{W})(R_n[\beta\ \boldsymbol{\gamma}_n] - r_n)$$

$$=^{H_0} \frac{1}{q}\boldsymbol{\epsilon}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\epsilon}$$

Where $=^{H_0}$ denotes that the equality holds under the null. Finding a limiting distribution requires us to first understand the mean and norm of these quadratic terms.

Lemma 3.2.1. *Under Assumptions 3.2.1-3.2.3 are met. Define,*

$$P_n^* = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$$

Then, for a balanced panel with N individuals observed T time periods,

$$\mathrm{E}(W_{n,OLS} \mid \mathbf{X}\ \mathbf{Z}) = \sum_{t,s}\sigma_{t,s}\frac{\sum_i P^*_{n,ii,ts}}{q}$$
$$= tr(\Sigma P_n^*)/q$$

[4]When $\boldsymbol{\epsilon}_i \sim N(0,V)$, Isserlis' Theorem states that $\mathrm{E}[\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\ \gamma_i] = \mathrm{E}[\epsilon_{it}\epsilon_{is} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]\,\mathrm{E}[\epsilon_{iu}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i] + \mathrm{E}[\epsilon_{it}\epsilon_{iu} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]\,\mathrm{E}[\epsilon_{is}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i] + \mathrm{E}[\epsilon_{it}\epsilon_{iv} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]\,\mathrm{E}[\epsilon_{is}\epsilon_{iu} \mid \mathbf{X}_i\ \mathbf{Z}_i\gamma_i]$.

$$\begin{aligned}\mathrm{Var}(W_{n,OLS} \mid \mathbf{X}\ \mathbf{Z}) = q^{-2}\sum_{tsuv}\mu_{tsuv}\sum_{i}P^*_{n,ii,ts}P^*_{n,ii,uv} \\ + q^{-2}\sum_{tsuv}\sigma_{t,s}\sigma_{u,v}\sum_{i,j\neq i}P^*_{n,ii,ts}P^*_{n,jj,uv} \\ + q^{-2}\sum_{tsuv}\sigma_{t,u}\sigma_{s,v}\sum_{i,j\neq i}P^*_{n,ij,ts}P^*_{n,ij,uv} \\ + q^{-2}\sum_{tsuv}\sigma_{t,v}\sigma_{s,u}\sum_{i,j\neq i}P^*_{n,ij,ts}P^*_{n,ji,uv}\end{aligned}$$

Note that the tests rely on the true population parameters $\sigma_{t,s}$ and μ_{tsuv} for all pairs of t, s, u, v. Generating a feasible version of this test statistic requires explicit estimation of all cross-product second and fourth moments for the symmetric error process across individuals. This differs from much of the traditional fixed effects literature where it is sufficient to rely on the moments from the within-transformed OLS residuals (Stock and Watson (2008)). Naive estimates for $\sigma_{t,s}$ based on the residuals directly from either Pooled OLS or the Within Transformation do not recover the population term, and instead of subject to asymptotic bias. See Example (3.2.1).

Example 3.2.1. *Consider the panel data model with just varying intercepts.*

$$y_{it} = \mu_i + \epsilon_{it}$$

$$\hat{\epsilon} = y_{it} - \hat{\mu}_i$$

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

Since T is fixed, the resulting estimators are unbiased but inconsistent.

$$\hat{\mu}_i = \mu_i + \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}$$

As a result,

$$\hat{\epsilon}_{it} = y_{it} - \mu_i - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}$$
$$= \epsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}$$

Under this fixed T framework, with "general" regularity conditions, the estimator

$$\begin{aligned}
n^{-1}\sum_i \hat{\epsilon}_{it}\hat{\epsilon}_{is} &= n^{-1}\sum_i(\epsilon_{it} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it})(\epsilon_{is} - \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}) \\
&= n^{-1}\sum_i\left(\epsilon_{it}\epsilon_{is} - \frac{1}{T}\sum_{u=1}^{T}\epsilon_{iu}(\epsilon_{it}+\epsilon_{is}) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\epsilon_{it}\epsilon_{is}\right) \\
&\to^p \sigma_{ts} - \frac{1}{T}\sum_u(\sigma_{ut}+\sigma_{us}) + \frac{1}{T^2}\sum_{u,v}\sigma_{uv}
\end{aligned}$$

Creating a pivotal statistic requires estimation of $\sigma_{t,s}$ and $\mu_{t,s,u,v}$ for all t,s,u,v. The next section develops asymptotic theory in order to estimate these terms under two different sets of assumptions.

3.3 Estimating Population Moments

The traditional estimators for σ_{ts} and μ_{tsuv} lead to biased moments under fixed effects due to first stage estimation error as shown in Example 3.2.1. This section develops estimators for these moments that control for the impacts of this first stage estimation error to construct consistent estimators for $\sigma_{t,s}$ and $\mu_{t,s,u,v}$ for all t,s,u,v.

This is carried out until two plausible feasible strategies. The first assumes a known number of groups $g = 1, ..., G$, where each group has at least two individuals, while allowing Ω to be fully non-stationary across individuals. As a downside this assumption requires that Z_i is the same across individuals. Under this structure, researchers are able to test the hypothesis that the individual coefficient is equal to the group coefficient, that is, $\gamma_i = \gamma_g \forall i \in g, \forall g$. These assumptions fit many applications of fixed effects testing for latent panel data structure, where researchers utilize individual specific intercept and time trends. With exogenous group descriptors, these assumptions allow researchers to test similarities between individuals in a group.

The second set of assumptions allows for a more general matrix of linear restrictions R, and Z_i that vary across individuals. Under this framework errors are assumed to be a mean zero, covariance stationary $AR(p)$ process. This parametric assumption provides closed form solutions for the the second and cross-product fourth moments that can be estimated in a two-stage fashion-first estimate the $AR(p)$ parameters, and then carrying out an asymptotic correction. These set of assumptions allow for z_{it} to vary across individuals and time, and to have linear restrictions and r_n to be unique for each individual in the sample.

These two methods are not the only way to create feasible estimators, but they cover two major cases in joint hypothesis tests for panel data methods. The main downside of both is the requirement that even within a known group structure individuals remain independent of each other ruling out cluster robust estimation at different hierarchical levels. Implicitly both methods rely on pooling information across individuals in the sample, where unobserved interdependence renders this technique inadequate.

3.3.1 Feasible Estimation under Known Group Structure

Most studies that impose fixed effects assume that fixed effects enter as an additive varying intercept for each individual, and occasionally a time trend. This structure is used routinely in both the correlated random coefficient (Wooldridge (2005)) and fixed effect (Hansen (2007)) literature. Under this shared-regressor framework, for each N if researchers assume a known group structure, such that $\gamma_i = \gamma_g$ for all individuals i in group g, then we can construct a cleaned model under the null,

$$y_{it} - y_{jt} - (x'_{it} - x_{jt})'\beta =^{H_0} \epsilon_{It} - \epsilon_{jt} \tag{3.4}$$

Since z_{it} is the same across individuals, and γ_i is the same for everyone in a particular group. This has several useful properties under independence across individuals.

$$\mathrm{E}(\epsilon_{it} - \epsilon_{jt} \mid \mathbf{W}\ \boldsymbol{\gamma}_n) = 0$$

$$\mathrm{Var}(\epsilon_{it} - \epsilon_{jt} \mid \mathbf{W}\ \boldsymbol{\gamma}_n) = 2\sigma_{tt}$$

$$\mathrm{Cov}(\epsilon_{it} - \epsilon_{jt}, \epsilon_{is} - \epsilon_{js} \mid \mathbf{W}\ \boldsymbol{\gamma}_n) = 2\sigma_{ts}$$

$$\mathrm{E}((\epsilon_{it} - \epsilon_{jt})(\epsilon_{is} - \epsilon_{js})(\epsilon_{iu} - \epsilon_{ju})(\epsilon_{iv} - \epsilon_{jv}) \mid \mathbf{W}\ \boldsymbol{\gamma}_n) = (\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{W}\ \boldsymbol{\gamma}_n) + \sigma_{t,s}\sigma_{u,v} + \sigma_{t,u}\sigma_{s,v} + \sigma_{t,v}\sigma_{s,u})$$

This difference preserves the dynamic relationship in the errors and motivates simple to implement sample analogs. Moreover, this structure allows for a growing number of groups. The assumptions for feasible estimation under this asymptotic framework can now be formalized.

Assumption 3.3.1. *Let the data be generated by Equation (3.1). We then make the following assumptions,*

(a) $\mathrm{E}(\epsilon_{it} \mid \mathbf{W}\ \boldsymbol{\gamma}_n) = 0$, $\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid \mathbf{W}\ \boldsymbol{\gamma}_n) = \Omega \quad \forall i$*, is positive definite, where the* t, s *term is denoted* $\sigma_{t,s}$*. There exists an* $r > 0$ *such that* $\mathrm{E}(|\epsilon_{it}|^{4+r}) < \infty$ *for all* i, t*.*

(b) $\{X_i, \boldsymbol{\epsilon}_i\}_{i=1}^n$ *are i.i.d.* Z_i *is non-stochastic and shared across everyone in the sample.*

(c) $\mathrm{Rank}(\sum_{t=1}^T \mathrm{E}[\ddot{x}_{it}\ddot{x}_{it}']) = \mathrm{Rank}(\mathrm{E}[\ddot{X}_i\ddot{X}_i']) = K$ *and* $\mathrm{Rank}(Z_i'Z_i) = L, \forall i$*.*

(d) There exists a constant Δ *such that* $\mathrm{E}[x_{ith}^4] \leq \Delta < \infty$*.*

(e) For all N *, there exist a number of groups* G_n*, where* g_i *denotes the group assignment of individual* i*, such that group size is denoted* n_g*. As* $n \to \infty$*,* $\sum_g n_g \to \infty$*.*

Most of these assumptions are standard. The first describes the errors as being mean zero, with the same variance-covariance matrix for all individuals, and sufficiently large enough moments for our application. The second imposes independence across individuals, and reiterates that Z_i is shared across everyone in the sample. Common examples of this is when $z_{it} = 1$ and the model is the traditional linear additive fixed effects model, or $z_{it} = (1, t)$ and includes individual specific time trends. In this case z_{it} is both non-random and shared across all individuals in the sample.

This assumption is invalidated in cases where z_{it} is a stochastic regressor on whether or not an individual was a smoker, years of education, or other covariate that represents

The third and fourth are common assumptions in the fixed effect literature to ensure the appropriate law of large number limits exist. The last requirement allows for either researchers to impose a fixed number of groups, and assume $\min_g n_g \to \infty$, or to let $G_n \to \infty$ as long as $\min_g n_g \geq 2$. The null assumes the individual's DGP is the equation,

$$y_{igt} = x_{it}'\beta + z_{it}'\gamma_g + \epsilon_{it} \tag{3.5}$$

And the hypotheses of interest is,

$$H_0: \ \gamma_i = \gamma_g \quad \forall i \tag{3.6}$$

That is, individual effects are actually the same as some deterministic grouping mechanism. To calculate the relevant moments individuals within each group are matched. This generates $\lfloor n_g/2 \rfloor$ pairs. For $i_1, i_2 \in g$, define

$$\hat{\zeta}_{g,i_1,i_2,t} = y_{i_1gt} - y_{i_2gt} - (x_{i_1t} - x_{i_2t})'\hat{\beta}^{FE} \tag{3.7}$$

Matched pairs can be generated via sampling without replacement, such that $\hat{\zeta}$ are i.i.d. under assumption 3.3.1. Sampling without replacement guarantees each individual is picked at most once, since no This implies every individual is picked at most once, since no additional bootstrap weights are imposed on the difference between residuals.

Proposition 3.3.1. *Let assumptions 3.2.1, 3.2.2, 3.2.3 and 3.3.1 hold. Moreover, n_g is even for all g. For each g, t, pick two individuals without replacement to generate $i_g = 1, \ldots, n_g/2$ pairs. For each g, i_g, define*

$$\hat{\zeta}_{g,i_1,i_2,t} = y_{i_1gt} - y_{i_2gt} - (x_{i_1t} - x_{i_2t})'\hat{\beta}^{FE}$$

Where we surpresss g is the subscript of i for ease. Then,

$$\frac{1}{2}\left(\sum_g n_g/2\right)^{-1}\sum_{g=1}^{G}\sum_{i_1=1}^{n_g/2}\hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s}\rightarrow^p \sigma_{t,s} \tag{3.8}$$

$$\frac{1}{2}\left(\sum_g n_g/2\right)^{-1}\sum_{g=1}^{G}\sum_{i_1=1}^{n_g/2}\hat{\zeta}_{i_1,i_2,t}\hat{\zeta}_{i_1,i_2,s}\hat{\zeta}_{i_1,i_2,u}\hat{\zeta}_{i_1,i_2,v}$$
$$\rightarrow^p 2(\mu_{tsuv}+\sigma_{t,s}\sigma_{u,v}+\sigma_{t,u}\sigma_{s,v}+\sigma_{t,v}\sigma_{s,u})$$

Moreover,

$$\frac{1}{2}\left(\sum_g \lfloor n_g/2\rfloor\right)^{-1}\sum_{g=1}^{G}\sum_{i_1=1}^{\lfloor n_g/2\rfloor}\hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s}=\frac{1}{2}\left(\sum_g \lfloor n_g/2\rfloor\right)^{-1}\epsilon'[I_{ts,\sum_g\sum_{n_g}}+K_{ts,1}+K_{ts,2}]\epsilon$$

Where $I_{ts,\sum_g\sum_{n_g}}$ is a diagonal matrix with 1's corresponding to observations that were picked in a particular sampling process. $K_{ts,1}$ is an upper diagonal matrix with 1's corresponding to each first matched pair, and $K_{ts,2}$ is a lower diagonal matrix with 1's corresponding to each second matched pair.

This subsection has shown that under Proposition 3.3.1 there exist consistent estimators for Σ and $\mathrm{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv}\mid \mathbf{W})$. Importantly, the final estimator allows us to represent our test statistic in Equation 3.3 as a function of a matrix independent of ϵ directly allowing for estimates of the variance following Proposition 3.2.1.

3.3.2 General Joint Hypothesis and Varying Z_i

The method developed in Section 3.3.1 is used for estimating latent panel structure in fixed T panels. This method fails in two cases. First, if Z_i vary across individuals, such as heterogeneous returns to schooling, and secondly if researchers are interested in a general joint hypothesis instead of just testing latent panel structure. This subsection develops an alternative feasible test that allows for a more general set of joint restrictions as suggested in Equation (3.2) at the cost of stationary assumptions on the error process. The following assumption clarifies the structural assumption on the error's data generating process.

Assumption 3.3.2. *$\epsilon_{it} = \epsilon_{it}^{-\prime}\alpha + \eta_{it}$ where $\epsilon_{it}^{-\prime} = [\epsilon_{i(t-1)}, \ldots, \epsilon_{i(t-p)}]$, and η_{it} is strictly stationary in t for each i, $\mathrm{E}[\eta_{it}^2] = \sigma_\eta^2$, $\mathrm{E}(\eta_{it}^4) = \mu_{4,\eta}$, $\mathrm{E}[\eta_{it}\eta_{i\tau}] = 0$ for all $t \neq \tau$, $\mathrm{E}(\eta_{it}^{4+r}) < \infty$ for some $r > 0$, and the roots of $1 - \alpha_1\xi - \alpha_2\xi^2 - \cdots - \alpha_p\xi^p = 0$ have absolute value greater than 1. We also have $T > p/2$ such that*

$$\mathrm{E}(\boldsymbol{\epsilon}_i \mid X_i\ Z_i\ \boldsymbol{\gamma}_n) = 0,\ \mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid X_i\ Z_i\ \boldsymbol{\gamma}_n) = \Gamma(\alpha)$$

and $MA(\infty)$ representation,

$$\epsilon_{it} = \sum_{d=0}^{\infty} \psi_d \eta_{i(t-d)}$$

This assumption states that the errors follow a mean zero $AR(p)$ process, where the innovation process has $4 + r$ moments, and that the individual error process is "block homogeneous" in time. The stationary assumption guarantees the existence of an invertible $MA(\infty)$ representation of each individual's error process with absolutely summable coefficients. The following Lemma shows that the $AR(p)$'s covariances can be represented as a function of α and σ_η^2, and the cross-product fourth moments as a function of α, σ_η^2, $\mu_{4,\eta}$.

Lemma 3.3.1. *Let Assumption 3.3.2 hold. Then, for any t, and $j, k, l \in \mathbb{Z}$ we have*

$$\mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}) = \sigma^2 \sum_{d=0}^{\infty} \psi_d \psi_{d+j}$$

and

$$\begin{aligned}
\mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}\epsilon_{i(t-k)}\epsilon_{i(t-l)}) = (\mu_{4,\eta} - 3\sigma_\eta^4) &\sum_{d=0}^{\infty} \psi_{d+|l|}\psi_{d+|l-j|}\psi_{d+|l-k|}\psi_d \\
&+ \sigma_\eta^4 \sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty} \psi_{d+|k|}\psi_{c+|l-j|}\psi_c\psi_d + \sigma_\eta^4 \sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty} \psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d \\
&+ \sigma_\eta^4 \sum_{d=0}^{\infty}\sum_{b\neq d}^{\infty} \psi_{b+|i|}\psi_b\psi_{d+|k-j|}\psi_d
\end{aligned}$$

This lemma shows that creating a estimates of $\sigma_{t,s}$ and μ_{tsuv} requires creating a consistent estimator for α, σ^2_η, and $\mu_{4,\eta}$. The three estimators proceed in similar, but slightly different methods. In all three cases, first generate an estimator for α, then construct an unbiased estimator for σ^2_η and $\mu_{4,\eta}$ using an asymptotic expansion, then construct an estimator for the cross product fourth moments. These constructions depend heavily on the estimator for α concstructed in Hansen (2007), and relaxation of non-stochastic regressors mostly leads to a notation change in their proofs (see Appendix). Define the OLS estimator for α to be,

$$\hat{\alpha} = \left(\frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p+1}^{T} \hat{\epsilon}_{it}^{-} \hat{\epsilon}_{it}^{-\prime} \right)^{-1} \left(\frac{1}{n(T-p)} \sum_{i} \sum_{t=p+1}^{T} \hat{\epsilon}_{it}^{-} \hat{\epsilon}_{it} \right) \tag{3.9}$$

Where $\hat{\epsilon}_{it}^{-\prime} = [\hat{\epsilon}_{i(t-p)}, \ldots, \hat{\epsilon}_{i(t-1)}]$. Proposition (3.8.2) in the appendices characterizes the asymptotic distribution of this estimator. It shows the proposed method generates a consistent estimation of α without any distributional assumptions on the innovation process outside of the usual exclusion restrictions for OLS. Most importantly it requires that the polynomial lag degree of α is known by the researcher. Now note that

$$\begin{aligned} y_{it} - & x_{it}'\hat{\beta} - z_{it}\hat{\gamma}_i - \hat{\alpha}(y_{i(t-1)} - x_{i(t-1)}'\hat{\beta} - z_{i(t-1)}\hat{\gamma}_i) \\ &= \eta_{it} + (x_{it} - \alpha x_{i(t-1)})'(\beta - \hat{\beta}) + (z_{it} - \alpha z_{i(t-1)})'(\gamma_i - \hat{\gamma}_i) \end{aligned}$$

Under the assumptions, the usual fixed effects estimator $\hat{\beta}^{FE}$ converges in probability to β. Therefore asymptotically this approximate $AR(p)$ correction implies

$$y_{it} - x_{it}'\hat{\beta} - z_{it}\hat{\gamma}_i - \hat{\alpha}(y_{i(t-1)} - x_{i(t-1)}'\hat{\beta} - z_{i(t-1)}\hat{\gamma}_i) \to^p \eta_{it} + (z_{it} - \alpha z_{i(t-1)})'(\gamma_i - \hat{\gamma}_i)$$

Consistent estimation of the underlying moments follows from an asymptotic expansion around this convergence, and generates the following estimator.[5]

[5]The appendix includes proofs that include the fourth moment, but underlying dependence in the meat of the sandwich generated by this estimator make it not useful for estimating the variance of the estimator defined in Equation (3.3).

Proposition 3.3.2. *Let assumptions 3.3.2 hold. Define,* $P^L_{Z_i,st} = \hat{\phi}(L)P_{Z_i,st} = (\phi(L')z_{it})'(Z_i'Z_i)^{-1}z_{is})$*. Then,*

$$(N(T-p))^{-1}\sum_i\sum_{t>p}\hat{\phi}(L)\hat{\epsilon}_{it}\hat{\phi}(L')\hat{\epsilon}_{it}/\omega_1 \rightarrow^p \sigma^2_\eta$$
$$\omega_1 = \left(1 - 2\frac{\sum_{i,t>p}\sum_s P^L_{Z_i,st}\hat{\psi}_{t-s}}{n(T-p)} + \frac{\sum_{i,t>p}\sum_{s,u}P^L_{Z_i,st}P^L_{Z_i,ut}\hat{\sigma}_{us}}{N(T-p)}\right)$$

Moreover,

$$\begin{aligned}\mathrm{tr}(\hat{\Sigma}P^*_n) &= \sum_{s,u}\hat{\sigma}_{s,u}\sum_{i,j}P_{ij,su}\\ &= \hat{\sigma}\sum_{s,u}\phi_d\phi_{d+|s-u|}\sum_{i,j}P_{ij,su}\\ &= \epsilon'\left((I_{nt}-K_{nt})*(n*t)^{-1}\sum_{s,u}\sum_{d=0}^{\infty}\hat{\phi}_d\hat{\phi}_{d+|s-u|}\sum_{i,j}P_{ij,su}\right)\epsilon\end{aligned}$$

With $\hat{\psi}_{-j} = 0\ \forall\ j > 0$*.*

Proposition 3.3.2 generates a consistent estimator for σ^2_η using the asymptotic expansion around $y_{it} - x'_{it}\hat{\beta} - z_{it}\hat{\gamma}_i - \hat{\alpha}(y_{i(t-1)} - x'_{i(t-1)}\hat{\beta} - z_{i(t-1)}\hat{\gamma}_i)$ under both the alternative and null hypothesis. As noted in Calhoun (2011), the null usually has fewer parameters and using the resulting projection matrix is the preferred method for estimating the sample moments.

However, compared to the grouping mechanism, the final line shows that the resulting estimator for $\hat{\Sigma}$ cannot be interpreted as a matrix $\epsilon' P_{ar}\epsilon$, where P_{ar} does not depend on ϵ. This is happens as $\hat{\phi}_d$ is a function of ϵ for all d, thus $\mathrm{E}[\epsilon'[(I_{nt}-K_{nt})*(n*t)^{-1}\sum_{s,u}\sum_{d=0}^{\infty}\hat{\phi}_d\hat{\phi}_{d+|s-u|}\sum_{i,j}P_{ij,su}]\epsilon \mid X] \neq \mathrm{E}[tr(\epsilon'\epsilon[(I_{nt}-K_{nt})*(n*t)^{-1}\sum_{s,u}\sum_{d=0}^{\infty}\hat{\phi}_d\hat{\phi}_{d+|s-u|}\sum_{i,j}P_{ij,su} \mid X]$. To estimate this term consistently a bootstrap procedure is carried out.

The next section characterizes the limiting distribution of the test statistic in Equation 3.3 using the population moments estimated in this section.

3.4 A Feasible Joint Hypothesis Test

This section characterizes the asymptotic behavior of two joint hypothesis tests for when the number of restrictions grows with the sample size. Under Proposition 3.2.1 and Assumptions 3.3.2 or 3.2.1, the test statistic (3.3) has two different representations,

$$\hat{G}_{gr} =^{H_0} \frac{1}{q}\frac{\boldsymbol{\epsilon}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\epsilon}}{\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1}\epsilon'[I_{ts,\sum_g \sum_{n_g}} + K_{ts,1} + K_{ts,2}]\epsilon} - 1$$

$$\hat{G}_{ar} =^{H_0} \frac{1}{q}\frac{\boldsymbol{\epsilon}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}R_n'[R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n']^{-1}R_n(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\epsilon}}{\epsilon'[(I_{nt} - K_{nt}) * (n*t)^{-1}\sum_{s,u}\sum_{d=0}^{\infty}\hat{\phi}_d\hat{\phi}_{d+|s-u|}\sum_{i,j}P_{ij,su}]\epsilon} - 1$$

The first test statistic, $\hat{G}_{n,gr}$ is the matrix representation under Assumptions 3.3.1. Under this framework, both the numerator and denominator take on quadratic form representations, $\boldsymbol{\epsilon}'P_n\boldsymbol{\epsilon}$ where P_n is non-random. The second test statistic is a matrix representation under assumptions 3.3.2, where the numerator has a quadratic form $\boldsymbol{\epsilon}'P_n^*\boldsymbol{\epsilon}$ in the numerator, however the denominator is now also a random matrix that depends on ϵ.

This transformation is useful since now each test is mean zero. This helps generate the following main result,

Theorem 3.4.1. *Let Assumptions 3.2.1-3.2.2 hold and* $\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid X_i \; Z_i) = \Sigma$ *is known. Then,*

$$\frac{q^{1/2}}{\nu_j^{1/2}}\hat{G}_j \Rightarrow N(0,1), \;\; j \in \{ar, gr\} \tag{3.10}$$

Where

$$\nu_{gr} = q^{-2}\sum_{tsuv}\mu_{tsuv}\sum_{i}P^{*}_{gr,ii,ts}P^{*}_{gr,ii,uv}$$
$$+ q^{-2}\sum_{tsuv}\sigma_{ts}\sigma_{uv}\sum_{i,j\neq i}P^{*}_{gr,ii,ts}P^{*}_{gr,jj,uv} + q^{-2}\sum_{tsuv}\sigma_{tu}\sigma_{sv}\sum_{i,j\neq i}P^{*}_{gr,ij,ts}P^{*}_{gr,ij,uv}$$
$$+ q^{-2}\sum_{tsuv}\sigma_{tv}\sigma_{su}\sum_{i,j\neq i}P^{*}_{gr,ij,ts}P^{*}_{gr,ji,uv}$$
$$\nu_{ar} = \left(\frac{1}{q^2}\epsilon^{*\prime}_{b}\left((I_{nt}-K_{nt})(nt)^{-1}\sum_{s,u}\sum_{d=0}^{\infty}\hat{\phi}^{*}_{d}\hat{\phi}^{*}_{d+|s-u|}\sum_{i,j}P^{*}_{ij,su}\right)\epsilon^{*}\right)^{2}$$

This theorem follows as ν_j^2 is the variance of the quadratic form that makes up the numerator, and the denominator converges in probability to the mean. The representation for ν_{gr}^2 follows directly from Proposition's 3.2.1 and 3.3.1. The representation for ν_{ar}^2 is the bootstrap estimator for the variance induced by Proposition 3.3.2.

3.5 Monte Carlo

This section compares various feasible implementations of the test statistics under conditions where both the grouping assumptions and the parametric assumptions of the error process are valid. For each test, data is generated from models of the form

$$y_{it} = \gamma_i + \epsilon_{it}$$

and test the null

$$H_0: \ \gamma_i = 0, \forall i$$

Implicit to this null, individuals are in the same group, and that group coefficient should is zero. Moreover, in each specification ϵ_{it} follows a covariance-stationary $AR(1)$ process, with different ρ and and innovation structures and therefore both Assumptions 3.3.1 and 3.3.2 are satisfied. For each monte carlo draw three statistics are reported,

$$\hat{G}_{tr} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}^2_{OLS,it,0} - \hat{\epsilon}^2_{OLS,it}}{\text{tr}(\hat{\Sigma}P_n^*)} - 1$$
$$\hat{G}_{ar} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}^2_{OLS,it,0} - \hat{\epsilon}^2_{OLS,it}}{\text{tr}(\hat{\Sigma}_{ar}P_n^*)} - 1$$
$$\hat{G}_{gr} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}^2_{OLS,it,0} - \hat{\epsilon}^2_{OLS,it}}{\text{tr}(\hat{\Sigma}_{gr}P_n^*)} - 1$$

The only novel test statistic here is G_{tr}. This test is calculated assuming that the researcher knows the errors data generating process, and estimates a million draws from the underlying innovation process, calculates $\hat{\Sigma}$ without first stage estimation error, and then calculates the remaining of the statistic as usual. The asymptotic variance in this case is calculated directly from Proposition 3.2.1, and converges to a standard Normal distribution under the same limit theory as presented in Theorem 3.4.1.

Figure 3.1 graphs the limiting distribution of the three tests for three different specifications. In each I$n = 500, t = 3$ with one of three different data generating processes for the error term. The left hand most plot coincides with homoskedastic normal errors, the middle with normal errors following an $AR(1)$ process with $\rho = 0.5$, and the right hand plot Student's T distributed innovations in an $AR(1)$ process with $\rho = 0.5$ and 8 degrees of freedom. These plots are based on 5000 monte carlo sims of the test under each DGP. The resulting size for each test is roughly $4.5 - 5$. The plots are generated by 5000 replications of the test statistic.

Next is a more general set of size tests.[6] Throughout only the three tests are presented. Comparable naive Wald tests with clustered standard errors *always* rejected.[7] Other tests designed for when the number of hypotheses grows with the number of observations, such as Calhoun (2011) and Orme and Yamagata (2006) are incompassed in my test in the case of homoskedastic errors,

[6]They take a while to run! Was having coding problems that I just fixed this weekend

[7]Each model was estimated with the plm package in R. The joint hypothesis is tested with the linearHypothesis function all with a robust variance-covariance matrix that clustering on the id and Newey-White standard errors to allow for serial correlation. The level of misspecification was so large as to always reject.

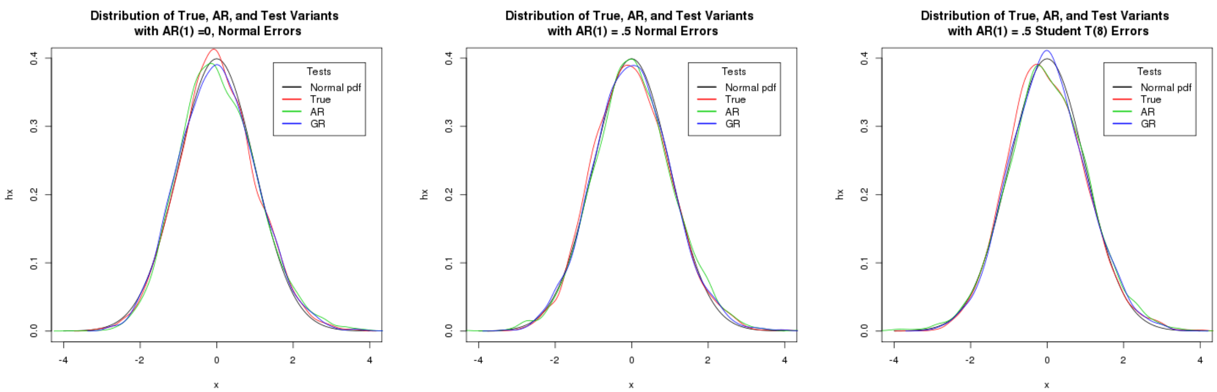

Figure 3.1 Monte Carlo Distributions for Three Different Designs

Table 3.1 Simulated Size of Test Statistics

			$t(8)$			$t(30)$			Exponential		
n	T	α	$\hat{G}_{tr}$	$\hat{G}_{ar}$	$\hat{G}_{gr}$	$\hat{G}_{tr}$	$\hat{G}_{ar}$	$\hat{G}_{gr}$	$\hat{G}_{tr}$	$\hat{G}_{ar}$	$\hat{G}_{gr}$
500	3	0	4.24	3.36	4.12	4.48	4.2	5	4.64	4.04	4.24
500	3	0.5	4.2	5.08	3.48	3.76	4.32	3.72	4.4	5.52	4.12

Simulated size for a nominal 5% test statistic with no regressors, based on 2500 simulated with innovations following either a Standard normal distribution with no serial correlation, standard Normal innovations generating a AR(1) = .5 process, Student's T distribution with 8 degrees of freedom and either no serial correlation, or generating an AR(1) = .5 process. The null hypothesis of each test is that all fixed effects are zero. Each column contains the size for a given test statistic and error distribution.

and Orme and Yamagata (2014) is when errors still exhibit no serial correlation, or exhibit time series heteroskedasticity with serial independence. This makes comparable baseline tests hard to find in the existing literature.

Overall, these tests are broadly close to, if not slightly under an appropriate 5% test, however approximate the preferred test size given the sample size and number of simulation runs quite well.

3.6 Conclusion

This paper develops a joint hypothesis tests over fixed effects for large N small T panels when errors are serially correlated. Two different feasible centered Wald style joint hypothesis tests are developed under different sets of assumptions. The first assumes a known grouping structure, where regressors that hypotheses are being constructed over are the same for each individual in the group. Researchers in this framework are interested in tests for the individual fixed effects being equal to the group fixed effect. Importantly the asymptotics here do not require consistent estimation of the group effect, allowing for both many groups with at least two members in them, as well as pooled group effects.

The second set of assumptions allows for individual-time varying covariates and a generalized set of joint hypotheses at the cost of required error to follow a stationary $AR(p)$ distribution. This setup is useful for situations where auxiliary information or alternative models might imply an alternative method for estimating individual fixed effects, and needing a method to compare these two models against each other. Prime examples of this are found in estimating differing models of teacher value added, or explicit tests if individual fixed effects are approximated by Mundlak-Chamberlain devices. As always, these applications are left up the reader!

It is shown that the test relies on estimating moments that the naive estimators for are biased due to first stage estimation error. This error is so systemic that naive tests using Newey-White serial correlation robust fixed effects estimators as a proxy still always reject under even mild serial correlation. Under the proposed assumptions, alternative unbiased estimators for the Variance-Covariance matrix is constructed that allow the creation of Wald style tests that are pivotal and converge to a standard Normal distribution. Formal size analysis of the tests are carried out and show that the test behaves well even under small to moderate panels.

3.7 References

Akritas, M. and Papadatos, N. (2004). heteroscedastic one-way ANOVA and lack-of-fit tests. *Journal of the American Statistical Association*, 99(466):368–382.

Anatolyev, S. (2012). Inference in regression models with many regressors. *Journal of Econometrics*, 170(2):368–382.

Arellano, M. and Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies*, 79(3):987–1020.

Bathke, A. (2004). The ANOVA F test can still be used in some balanced designs with unequal variances and nonnormal data. *Journal of Statistical Planning and Inference*, 126(2):413–422.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?

Blomquist, J. and Westerlund, J. (2013). Testing slope homogeneity in large panels with serial correlation. *Economics Letters*, 121(3):374–378.

Boos, D. D. and Brownie, C. (1995). ANOVA and rank tests when the number of treatments is large. *Statistics and Probability Letters*, 23(2):183–191.

Calhoun, G. (2011). Hypothesis testing in linear regression when k n is large. *Journal of Econometrics*, 165(2):163–174.

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018a). Alternative asymptotics and the partially linear model with many regressors. *Econometric Theory*, 34(2):277–301.

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018b). Inference in Linear Regression Models with Many Covariates and Heteroscedasticity. *Journal of the American Statistical Association*, 113(523):1350–1361.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2017). Measuring the impacts of teachers: Reply. *American Economic Review*, 107(6):1685–1717.

Durlauf, S. N., Kourtellos, A., and Minkin, A. (2001). The local Solow growth model. *European Economic Review*, 45(4-6):928–940.

Hansen, C. B. (2007). Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, 140(2):670–694.

Hanson, S. G. and Sunderam, A. (2009). The Variance of Average Treatment Effect Estimators in the Presence of Clustering. *The Review of Economics and Statistics*, 94(4):1–23.

Hashem Pesaran, M. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1):50–93.

Hausman, J. and Kuersteiner, G. (2008). Difference in difference meets generalized least squares: Higher order properties of hypotheses tests. *Journal of Econometrics*, 144(2):371–391.

Heckman, J. and Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model. Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, 33(4):974–987.

Jöckel, K. and Sendler, W. (1981). A central limit theorem for generalized discounting. *Series Statistics*, 12(4):605–608.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6):615–631.

Kline, P., Saggio, R., and Sølvsten, M. (2018). Leave-out estimation of variance components.

Lin, C.-C. and Ng, S. (2012). Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown. *Journal of Econometric Methods*, 1(1):42–55.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69.

Nerlove, M. (1971). Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections. *Econometrica*, 39(2):359.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417.

Orme, C. D. and Yamagata, T. (2006). The asymptotic distribution of the F-test statistic for individual effects. *Econometrics Journal*, 9(3):404–422.

Orme, C. D. and Yamagata, T. (2014). A Heteroskedasticity-Robust F-Test Statistic for Individual Effects. *Econometric Reviews*, 33(5-6):431–471.

Phillips, P. C. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Stock, J. H. and Watson, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1):155–174.

Stock, J. H. and Watson, M. W. (2009). Forecasting in Dynamic Factor Models Subject to Structural Instability. *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*.

Su, L., Shi, Z., and Phillips, P. C. B. (2014). Identifying Latent Structures in Panel Data. *SSRN Electronic Journal*, 84(6):2215—-2264.

Swamy, P. A. V. B. (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, 38(2):311.

Verbeek, M. and Nijman, T. (1992). Testing for Selectivity Bias in Panel Data Models. *International Economic Review*, 33(3):681.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 87(2):385–390.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

3.8 Appendix

3.8.1 Fixed Effect Estimation Lemmas

Lemma 3.8.1. *Let $\hat{\beta}$ be the ordinary least squares estimate of β. Then if the assumptions hold, $\hat{\beta} - \beta \to^p 0$ and $\sqrt{N}(\hat{\beta} - \beta) \to^d N(0, \mathrm{E}(\ddot{X}_i'\ddot{X}_i)^{-1}\ddot{X}_i'\Omega\ddot{X}_i\,\mathrm{E}(\ddot{X}_i'\ddot{X}_i)^{-1})$. Moreover,*

$$\hat{\mathbf{Y}} - \mathbf{Y} = \hat{\ddot{\mathbf{Y}}} - \ddot{\mathbf{Y}} = (1 - \mathbf{P}_{\ddot{\mathbf{X}}_n})\ddot{\epsilon} = \mathbf{M}_{\ddot{\mathbf{X}}_n}\mathbf{M}_{\mathbf{Z}}\boldsymbol{\epsilon}$$

Proof of Lemma 3.8.1. The proof is almost identical to Hansen (2004) Lemma 1.8.1 except that the Z_i's are allowed to vary across individuals.

We know $||AB|| \leq ||A||||B||$ by the Cauchy-Schwarz inequality, and $I - Z_i(Z_i'Z_i)^{-1}Z_i'$ is positive semi-definite for all i,

$$\begin{aligned}
\mathrm{E}\,||\ddot{X}_i{}'\ddot{X}_i|| &\leq \mathrm{E}\,||\ddot{X}_i||||\ddot{X}_i|| = \mathrm{E}\,||\ddot{X}_i||^2 \\
&= \mathrm{E}(\mathrm{tr}(X_i'X_i - X_i'Z_i(Z_i'Z_i)^{-1}Z_i'Z_i)) \\
&\leq \mathrm{tr}(\mathrm{E}(X_i'X_i)) = \mathrm{tr}(\sum_{t=1}^{T}\mathrm{E}(x_{it}x_{it}')) < \infty
\end{aligned}$$

Next, define $\ddot{\boldsymbol{\epsilon}}_i = \boldsymbol{\epsilon}_i - Z_i(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i$. Then, $||\ddot{X}_i{}'\ddot{\boldsymbol{\epsilon}}_i|| \leq (\mathrm{E}(||\ddot{X}_i||^2)\,\mathrm{E}(\ddot{\boldsymbol{\epsilon}}_i)^2)^{1/2}$, and the same arguments as before hold. By the weak law of large numbers,

$$\frac{1}{nT}\sum_i \ddot{X}_i{}'\ddot{X}_i \to^p \mathrm{E}(\ddot{X}_i{}'\ddot{X}_i)$$

and

$$\frac{1}{nT}\sum_i \ddot{X}_i{}'\ddot{\boldsymbol{\epsilon}}_i \to^p 0$$

Then $\hat{\beta} - \beta \rightarrow^p 0$.

For asymptotic normality, we know by 3.2.2 $\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i$ is iid and has mean zero. Then,

$$\mathrm{E}(||\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i\ddot{\boldsymbol{\epsilon}}_i'\ddot{X}_i||) \leq (2\,\mathrm{E}(||X_i||^4)\,\mathrm{E}(||\boldsymbol{\epsilon}_i||^4)^{1/2} < \infty$$

by the Cauchy-Schwarz inequality, and

$$\begin{aligned}\mathrm{E}(||\ddot{X}_i') &= \mathrm{E}((\mathrm{tr}(X_i'X_i))^2 - 2\,\mathrm{tr}(X_i'X_i)\,\mathrm{tr}(X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i) + (\mathrm{tr}(X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i))^2) \\ &\leq \mathrm{E}(2\,\mathrm{tr}(X_i'X_i)^2) = 2\,\mathrm{E}(||X_i||^4)\end{aligned}$$

where in the inequality follows from $X_i'X_i$, $X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i$ and $I_T - Z_i(Z_i'Z_i)^{-1}Z_i$ positive semi-definite. It then follows from the Lindberg-Levy CLT that $\frac{1}{\sqrt{N}}\sum_{i=1}^N \ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i \rightarrow^d N(0,\Omega)$ since $\mathrm{E}(\ddot{X}_i'\ddot{\boldsymbol{\epsilon}}_i\ddot{\boldsymbol{\epsilon}}_i'\ddot{X}_i) = \mathrm{E}(\ddot{X}_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'\ddot{X}_i) = \mathrm{E}(\ddot{X}_i'\Gamma(\alpha)\ddot{X}_i)$, from which $\sqrt{N}(\hat{\beta} - \beta) \rightarrow^d N(0, M^{-1}\Omega M^{-1})$ is obtained. □

Lemma 3.8.2. *Let Assumption 3.3.2 hold. Then, for any t, and $j,k,l \in \mathbb{Z}$ we have*

$$\mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}) = \sigma^2 \sum_{d=0}^{\infty} \psi_d\psi_{d+j}$$

and

$$\begin{aligned}\mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}\epsilon_{i(t-k)}\epsilon_{i(t-l)}) &= (\mu_{4,\eta} - 3\sigma_\eta^4)\sum_{d=0}^{\infty}\psi_{d+|l|}\psi_{d+|l-j|}\psi_{d+|l-k|}\psi_d \\ &+ \sigma_\eta^4\sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty}\psi_{d+|k|}\psi_{c+|l-j|}\psi_c\psi_d + \sigma_\eta^4\sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty}\psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d \\ &+ \sigma_\eta^4\sum_{d=0}^{\infty}\sum_{b\neq d}^{\infty}\psi_{b+|i|}\psi_b\psi_{d+|k-j|}\psi_d\end{aligned}$$

Proof of Lemma 3.3.1. Since Y_t is a covariance stationary pth-order autoregressive process, there exists a $MA(\infty)$ representation of the form,

$$\epsilon_{it} = \alpha(L)\eta_{it} \tag{3.11}$$

where $\alpha(L) = (1 - \alpha_1 L - \cdots - \alpha_p L^p)^{-1}$, where the inverted $MA(\infty)$ representation generates the infinite sequence $\{\psi_j\}_{j=1}^{\infty}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Then, for any t, s, u, v, such that we can rewrite this as $s = t - i,\ u = t - j,\ v = t - k$ with $i, j, k \in \mathbb{Z}$. Then, we have

$$\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}) = \mathrm{E}(\epsilon_{it}\epsilon_{i(t-j)}\epsilon_{i(t-k)}\epsilon_{i(t-l)}) \tag{3.12}$$

$$= \mathrm{E}(\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{d=0}^{\infty} \psi_a\psi_b\psi_c\psi_d \eta_{i(t-a)}\eta_{i(t-j-b)}\eta_{i(t-k-c)}\eta_{i(t-l-d)}) \tag{3.13}$$

$$= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{d=0}^{\infty} \psi_a\psi_b\psi_c\psi_d\, \mathrm{E}(\eta_{i(t-a)}\eta_{i(t-j-b)}\eta_{i(t-k-c)}\eta_{i(t-l-d)}) \tag{3.14}$$

Now we know that η_{it} is an iid process. Such that,

$$\mathrm{E}(\eta_{i(t-a)}\eta_{i(t-j-b)}\eta_{i(t-k-c)}\eta_{i(t-l-d)}) = \begin{cases} \sigma^4 & t-a = t-j-b \neq t-k-c = t-l-d \\ & t-a = t-k-c \neq t-j-b = t-l-d \\ & t-a = t-l-d \neq t-j-b = t-k-c \\ \mu_4 & t-a = t-j-b = t-k-c = t-l-d \\ 0 & \text{otherwise} \end{cases}$$

Carrying out expectations, we get,

$$\begin{aligned} \mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}) &= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{c=0}^{\infty}\sum_{d=0}^{\infty} \psi_a\psi_b\psi_c\psi_d\, \mathrm{E}(\eta_{i(t-a)}\eta_{i(t-i-b)}\eta_{i(t-j-c)}\eta_{i(t-k-d)}) \\ &= (\mu_4 - 3\sigma^4)\sum_{d=0}^{\infty} \psi_{d+|k|}\psi_{d+|k-i|}\psi_{d+|k-j|}\psi_d \\ &+ \sigma^4 \sum_{d=0}^{\infty}\sum_{c\neq d}^{\infty} \psi_{d+|k|}\psi_{c+|j-i|}\psi_c\psi_d \\ &+ \sigma^4 \sum_{d}\sum_{c} \psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d \\ &+ \sigma^4 \sum_{d}\sum_{b} \psi_{b+|i|}\psi_b\psi_{d+|k-j|}\psi_d \end{aligned}$$

□

3.8.2 Moment Estimation

3.8.2.1 GR

Proposition 3.8.1. *Let assumptions 3.3.1 hold. For each g, t, pick two individuals without replacement. And define the term*

$$\hat{\zeta}_{g,i_1,i_2,t} = y_{i_1gt} - x'_{i_1t}\hat{\beta}^{FE} - y_{i_2gt} + x'_{i_2t}\hat{\beta}^{FE}$$

$$\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \sum_{g=1}^{G} \sum_{i_1=1}^{\lfloor n_g/2 \rfloor} \hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s} \to^p \sigma_{t,s} \tag{3.15}$$

$$\begin{aligned}\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \sum_{g=1}^{G} \sum_{i_1=1}^{\lfloor n_g/2 \rfloor} & \hat{\zeta}_{i_1,i_2,t}\hat{\zeta}_{i_1,i_2,s}\hat{\zeta}_{i_1,i_2,u}\hat{\zeta}_{i_1,i_2,v} \\ & \to^p 2(\mathrm{E}(\epsilon_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{Z}\ \mathbf{X}) + \sigma_{t,s}\sigma_{u,v} \\ & + \sigma_{t,u}\sigma_{s,v} + \sigma_{t,v}\sigma_{s,u})\end{aligned}$$

Moreover,

$$\frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \sum_{g=1}^{G} \sum_{i_1=1}^{\lfloor n_g/2 \rfloor} \hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s} = \frac{1}{2}\left(\sum_g \lfloor n_g/2 \rfloor\right)^{-1} \epsilon'[I_{ts,\sum_g \sum_{n_g}} + K_{ts,1} + K_{ts,2}]\epsilon$$

Where $I_{ts,\sum_g \sum_{n_g}}$ is a diagonal matrix with 1's corresponding to observations that were picked in a particular sampling process. $K_{ts,1}$ is an upper diagonal matrix with 1's corresponding to each first matched pair, and $K_{ts,2}$ is a lower diagonal matrix with 1's corresponding to each second matched pair.u

Proof. Let n, G, n_g all be even. For each g, t pick two individuals without replacement. Where each pair is indexed by $i_{gt} = 1, \ldots, n_g/2$. Definine $\sum_g n_g/2 = N_G$ and

$$\hat{\zeta}_{i_{1,gt},i_{2,gt}} = y_{i_1gt} - x'_{i_1t}\hat{\beta}^{FE} - y_{i_2gt} + x'_{i_2t}\hat{\beta}^{FE}$$

Then,

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}\hat{\zeta}_{g,i_1,i_2,t}\hat{\zeta}_{g,i_1,i_2,s} = \frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(y_{i_1gt} - x'_{i_1t}\hat{\beta}^{FE} - y_{i_2gt} + x'_{i_2t}\hat{\beta}^{FE})(y_{i_1gs} - y_{i_2gs} - (x_{i_1t} - +x'_{i_2t})\hat{\beta}^{FE})$$

$$= \frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(\epsilon_{i_1gt} - \epsilon_{i_2gt} - (x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE}))(\epsilon_{i_1gs} - \epsilon_{i_2gs} - (x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE}))$$

$$= \frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}\epsilon_{i_1gt}\epsilon_{i_1gs} + \epsilon_{i_2gt}\epsilon_{i_2gs} + \epsilon_{i_1gt}\epsilon_{i_2gt} + \epsilon_{i_1gs}\epsilon_{i_2gs}$$

$$- \frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gs} - \epsilon_{i_2gs}) - (x_{i_1s} - x_{i_2s})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gt} - \epsilon_{i_2gt})$$

$$+ \frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\beta - \hat{\beta}^{FE})'(x_{i_1s} - x_{i_2s})$$

Under Assumption 3.3.1,

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}\epsilon_{i_1gt}\epsilon_{i_1gs} \to^p \sigma_{ts}$$

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}\epsilon_{i_2gt}\epsilon_{i_2gs} \to^p \sigma_{ts}$$

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}\epsilon_{i_1gt}\epsilon_{i_2gt} \to^p 0$$

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}\epsilon_{i_1gs}\epsilon_{i_2gs} \to^p 0$$

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gs} - \epsilon_{i_2gs}) \to^p 0$$

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(x_{i_1s} - x_{i_2s})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gt} - \epsilon_{i_2gt}) \to^p 0$$

$$\frac{1}{N_G}\sum_{j=1}^{G}\sum_{i=1}^{n_g/2}(x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\beta - \hat{\beta}^{FE})'(x_{i_1s} - x_{i_2s}) \to^p 0$$

The first four directly follow from the traditional LLN under independence. The remaining three require slightly more care.

$$\beta - \hat{\beta}^{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}\ddot{\epsilon}$$

Then

$$\begin{aligned}
E((x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gs} - \epsilon_{i_2gs})) &= E((x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gs} - \epsilon_{i_2gs}) \mid \mathbf{W}) \\
&= (x_{i_1t} - x_{i_2t})'(\ddot{X}'\ddot{X})^{-1}\ddot{X}E(\ddot{\epsilon}(\epsilon_{i_1gs} - \epsilon_{i_2gs}) \mid \mathbf{W})
\end{aligned}$$

Where, with some mild notation abuse

$$E(\ddot{\epsilon}_{it}(\epsilon_{i_1gs} - \epsilon_{i_2gs}) \mid \mathbf{W}) = \begin{cases} (\sigma_{ts} - \operatorname{tr}(\Omega P_Z) & i = i_1 \\ -(\sigma_{ts} - \operatorname{tr}(\Omega P_{\ddot{X}}) & i = i_2 \\ 0 & \text{otherwise} \end{cases}$$

Define

$$\sigma = [\sigma_{11} - \operatorname{tr}(\Omega P_{\ddot{X}}), \sigma_{12} - \operatorname{tr}(\Omega P_{\ddot{X}}), \dots]$$

Then,

$$\begin{aligned}
E((x_{i_1t} - x_{i_2t})'(\beta - \hat{\beta}^{FE})(\epsilon_{i_1gs} - \epsilon_{i_2gs})) &= (x_{i_1t} - x_{i_2t})'((\ddot{X}'\ddot{X})^{-1}\ddot{X}\operatorname{Vec}(E(\ddot{\epsilon}(\epsilon_{i_1gs} - \epsilon_{i_2gs}) \mid \mathbf{W})) \\
&= (x_{i_1t} - x_{i_2t})'((\ddot{X}'\ddot{X})^{-1}\ddot{X}(E(\ddot{\epsilon}(\epsilon_{i_1gs} \mid \mathbf{W})) - (x_{i_1t} - x_{i_2t})'((\ddot{X}'\ddot{X})^{-1}\ddot{X}(E(\ddot{\epsilon}\epsilon_{i_2gs}) \mid \mathbf{W})) \\
&= (x_{i_1t} - x_{i_2t})'((\ddot{X}'\ddot{X})^{-1}\ddot{X}_i\sigma - (x_{i_1t} - x_{i_2t})'((\ddot{X}'\ddot{X})^{-1}\ddot{X}_{i_2}\sigma \\
&= (x_{i_1t} - x_{i_2t})'(\ddot{X}'\ddot{X})^{-1}(\ddot{X}_{i_1} - \ddot{X}_{i_2})\sigma
\end{aligned}$$

□

3.8.2.2 AR

Lemma 3.8.3. *Let the assumptions hold. Then define* $\hat{\alpha} = \left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\tilde{e}_{it}^{-}\tilde{e}_{it}^{-\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T}\tilde{e}_{it}^{-}\tilde{e}_{it}'\right)$ *be the least squares estimate of* α *using the lest squares residuals* $\tilde{e}_{it}$ *from estimating* β_1*. Then,*

$$\hat{\alpha} = \left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T} \ddot{e}it^{-}\ddot{e}_{it}^{-\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{n}\sum_{t=p+1}^{T} \ddot{e}_{it}^{-}\ddot{e}_{it}^{\prime}\right) + o_p(N^{-1/2})$$

Proof of 3.8.3. Same as Hansen (2007) □

Lemma 3.8.4. *Define $\tilde{e}_{it}$ to be the residual from least squares regression, i.e. $\tilde{e}_{it} = y_{it} - x_{it}'\beta - z_{it}'\gamma_i = \epsilon_{it} - x_{it}'(\hat{\beta}_1 - \beta) - z_{it}'(\hat{\gamma}_i - \gamma_i)$, where $\hat{\beta}$ and $\hat{\gamma}_i$ are least squares estimates of β and γ_i. Then under our Assumptions, $N^{-1}\sum_i \sum_{t=p+1}^{T} \tilde{e}_{it}^{-}\tilde{e}_{it}^{-\prime} = N^{-1}\sum_i \sum_{t=p+1}^{T} \ddot{e}_{it}^{-}\ddot{e}_{it}^{-\prime} + o_p(N^{-1/2})$ and $N^{-1}\sum_i \sum_{t=p+1}^{T} \tilde{e}_{it}^{-}\tilde{e}_{it}' = N^{-1}\sum_i \sum_{t=p+1}^{T} \ddot{e}_{it}^{-}\ddot{e}_{it}' + o_p(N^{-1/2})$*

Proof of 3.8.3. Same as Hansen (2007) □

Proposition 3.8.2. *Suppose $\alpha_T(\alpha)$ is continuously differentiable in α and that the derivative matrix of $\alpha_T(\alpha)$ in α, $H = D\alpha_T(\alpha)$, is invertible for all α such that Assumption 3.3.2 is satisfied, and $D\alpha_T(\alpha)$ is the derivative matrix of $\alpha_T(\alpha)$ in α . Then, $\hat{\alpha}^{\infty} - \alpha \to^p 0$ and*

$$\sqrt{N}(\hat{\alpha}^{\infty} - \alpha) \to^d \frac{1}{T-p} H^{-1}(\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_{\gamma(\alpha)})^{-1}\chi$$

where $\chi = N(0, \Xi_T)$, and

$$\Xi_T = \mathrm{E}[\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T} \ddot{e}_{it_1}\ddot{\mu}_{it_1}\ddot{\mu}_{it_2}\epsilon\ddot{_{it_2}}]$$

and

$$\ddot{\mu_{it}} = \ddot{e}_{it} - \ddot{e}_{it}^{-\prime}\alpha_T(\alpha)$$

Under these asymptotics, the proposition shows that show $\hat{\alpha} \to^p \alpha_T(\alpha) = (\Gamma(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha) + \frac{1}{T-p}\Delta_A(\alpha))$, with $A(\alpha) = [\gamma_1 \ \ldots \ \gamma_p]'$, Δ_Γ is a $p \times p$ matrix with

$$\begin{aligned}[\Delta_\Gamma(\alpha)]_{k,j} &= \mathrm{tr}(\Gamma(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_i(Z_i'Z_i)^{-1}Z_{i,-k}'Z_{i,-j}(Z_i'Z_i)^{-1}Z_i') \\ &\quad - \mathrm{tr}(\Gamma_{-k}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-j}(Z_i'Z_i)^{-1}Z_i') - \mathrm{tr}(\Gamma_{-j}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-k}(Z_i'Z_i)^{-1}Z_i')\end{aligned}$$

and $\Delta_A(\alpha)$ is a $p \times 1$ matrix with

$$[\Delta_A(\alpha)]_{i,1} = \mathrm{tr}(\Gamma(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_i(Z_i'Z_i)^{-1}Z_{i,-k}'Z_{i,-0}(Z_i'Z_i)^{-1}Z_i')$$
$$- \mathrm{tr}(\Gamma_{-k}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-0}(Z_i'Z_i)^{-1}Z_i') - \mathrm{tr}(\Gamma_{-0}(\alpha)\frac{1}{n}\sum_{i=1}^{n} Z_{i,-k}(Z_i'Z_i)^{-1}Z_i')$$

with $\Gamma_{-k}(\alpha) = \mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i,-k}' \mid \mathbf{W})$, $\boldsymbol{\epsilon}_{i,-k}' = [\boldsymbol{\epsilon}_{i(p+1-k)},\ \boldsymbol{\epsilon}_{i(p+2-k)}, \ldots,\ \boldsymbol{\epsilon}_{i(T-k)}]$, and $Z_{i,-k}$ defined equivalently. Without estimation error the OLS estimator would just be equal to $\Gamma(\alpha)^{-1}A(\alpha)$, and these remaining terms represent the role of first stage estimation error on the expected value of the estimator. Note that the resulting value $\alpha_T(\alpha)$, a function of the true underlying parameter. Therefore in sufficiently large samples this implies if $\alpha_T(\alpha)$ is invertible, we can generate a consistent estimator for α by taking the inverse of α_T^{-1} around $\hat{\alpha}$. That is,

$$\hat{\alpha}^{\infty} = \alpha_T^{-1}(\hat{\alpha}) = \alpha_T^{-1}(\alpha_T(\alpha)) = \alpha \tag{3.16}$$

Proof of Lemma 3.3.2. Define $\phi(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$, where L is the lag operator. Under Assumptions 3.3.2-3.2.2,

$$\phi(L)\epsilon_{it} = \eta_{it} \tag{3.17}$$

Since $\phi(L)$ has roots outside the unit circle, there exists an invertible MA(∞) representation,

$$\epsilon_{it} = \sum_{j=0}^{\infty} \psi_j \eta_{i(t-j)} \tag{3.18}$$

Such that $\mathrm{E}(\epsilon_{it}\epsilon_{i(t-s)} \mid \mathbf{W}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|s|}$. For $\boldsymbol{\epsilon}_i = [\epsilon_{i1} \ \ldots \ \epsilon_{iT}]'$, we know

$$\mathrm{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i' \mid \mathbf{W}) = \sigma^2\Omega \tag{3.19}$$

Then, for $\ddot{\epsilon}_{it} = \epsilon_{it} - z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i$,

$$\phi(L)\ddot{\epsilon}_{it} = \eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i = \eta_{it} - \phi(L)z_{it}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i \tag{3.20}$$

Then,

$$\begin{aligned}
\mathrm{E}[\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it} \mid \mathbf{W}] &= \mathrm{E}((\eta_{it} - \phi(L)z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)(\eta_{it} - \phi(L')z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)' \mid \mathbf{W}) \\
&= \mathrm{E}(\eta_{it}^2 \mid \mathbf{W}) - 2\phi(L)z'_{it}(Z'_iZ_i)^{-1}Z'_i\,\mathrm{E}(\boldsymbol{\epsilon}_i\eta_{it} \mid \mathbf{W}) \\
&\quad + \mathrm{E}((\phi(L)z_{it})'(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}\phi(L')z_{it} \mid \mathbf{W})
\end{aligned}$$

Now note that

$$\mathrm{E}(\epsilon_{is}\eta_{it} \mid \mathbf{W}) = \begin{cases} 0 & t > s \\ \psi_{|t-s|}\sigma^2 & \text{otherwise} \end{cases}$$

Such that we can define $\boldsymbol{\psi}_t = [1_{t\le 1}\psi_{|1-t|} \dots 1_{t\le T}\psi_{|T-t|}]$. Thus, $\mathrm{E}(\boldsymbol{\epsilon}_i\eta_{it} \mid \mathbf{W}) = \sigma^2\boldsymbol{\psi}'_t$. Thus,

$$\begin{aligned}
\mathrm{E}[\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it} \mid \mathbf{W}] &= \sigma_\eta^2 - 2\sigma_\eta^2\phi(L)z'_{iT}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\psi}'_t \\
&\quad + \phi(L)\,\mathrm{E}(\mathrm{tr}((\phi(L)z_{iT})'(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}\phi(L)z_{iT} \mid \mathbf{W}) \\
&= \sigma_\eta^2 - 2\phi(L)\sigma_\eta^2 z'_{iT}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\psi}'_t + \sigma^2\,\mathrm{tr}(\Omega Z_i(Z'_iZ_i)^{-1}\phi(L')z_{it}(\phi(L)z_{it})'(Z'_iZ_i)^{-1}Z'_i) \\
&= \sigma_\eta^2(1 - 2\phi(L)z'_{iT}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\psi}'_t + \mathrm{tr}(\Omega Z_i(Z'_iZ_i)^{-1}\phi(L')z_{iT}(\phi(L)z_{iT})'(Z'_iZ_i)^{-1}Z'_i) \\
&< \infty
\end{aligned}$$

Then the results follow follow from the Khintchine's LLN as,

- $\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it}$ are iid across individuals by Assumption 3.2.1
- $\mathrm{E} \mid \eta_{it}\eta_{it} \mid < \infty$
- $\mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i\eta_{it}| \mid \mathbf{W}) < \left(\mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|\eta_{it}^2| \mid \mathbf{W})\right)^{1/2}$
- $\mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}z_{it}| \mid \mathbf{W}) \le \left(\mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2\mathbf{W})\right)^{1/2}$

With

$$
\begin{aligned}
\mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2\mathbf{W}) &= \mathrm{E}(|z'_{it}(Z'_iZ_i)^{-1}Z'_iZ_i(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i|^2\mathbf{W}) \\
&\leq ||z'_{it}(Z'_iZ_i)^{-1}Z'_i||^2\,\mathrm{E}(||z'_{iT}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i||^2 \mid \mathbf{W}) \\
&= \mathrm{tr}(z_{it}(Z'_iZ_i)^{-1}z_{it}) \times \mathrm{E}(\mathrm{tr}(\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i) \mid \mathbf{W}) \\
&\leq L\,\mathrm{E}(\mathrm{tr}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}'_i) \mid \mathbf{W}) = L\,\mathrm{E}(\sum_t^T \eta_{it} \mid \mathbf{W}) < \infty
\end{aligned}
$$

The final inequality follows as $\mathrm{tr}(z'_{it}(Z'_iZ_i)^{-1}z_{it}) < \mathrm{tr}(Z_i(Z'_iZ_i)^{-1}Z'_i) = L$ and $\mathrm{tr}(\boldsymbol{\epsilon}'_iZ_i(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i) \leq \mathrm{tr}(\boldsymbol{\epsilon}'_i\boldsymbol{\epsilon}_i)$.

For the second set of error corrections, Define $P_{ts} = \phi(L)z_{it}(Z'_iZ_i)^{-1}z'_{is}$

$$
\begin{aligned}
\mathrm{E}[(\phi(L)\hat{\epsilon}_{it})^4 \mid \mathbf{W}] &= \mathrm{E}((\eta_{it} - \phi(L)z_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)(\eta_{it} - \phi(L)z_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)'\times \\
&\qquad (\eta_{it} - \phi(L)z_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)(\eta_{it} - \phi(L)z_{it}(Z'_iZ_i)^{-1}Z'_i\boldsymbol{\epsilon}_i)' \mid \mathbf{W}) \\
&= \mu_4 - 4\sum_s \mathrm{E}(\eta_{it}^3\epsilon_{is} \mid \mathbf{W})P_{ts} \\
&\quad + 6\sum_{s,u} \mathrm{E}(\eta_{it}^2\epsilon_{is}\epsilon_{iu}) \mid \mathbf{W})P_{ts}P_{tu} \\
&\quad - 4\sum_{s,u,v} \mathrm{E}(\eta_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv}) \mid \mathbf{W})P_{ts}P_{tu}P_{tv} \\
&\quad + \sum_{s,u,v,w} \mathrm{E}(\epsilon_{is}\epsilon_{iu}\epsilon_{iv}\epsilon_{iw}) \mid \mathbf{W})P_{ts}P_{tu}P_{tv}P_{tv}
\end{aligned}
$$

As above, we can now generate the expectations of each term. Let $\psi_{-j} = 0, \forall j > 0$. This simplifies the notiation quite a bit, and allows us to express the expectation as

$$\mathrm{E}(\eta_{it}^3\epsilon_{is} \mid \mathbf{W}) = \mathrm{E}(\eta_{it}^3 \sum_{d=0}^{\infty} \psi_d \eta_{i(s-d)} \mid \mathbf{W})$$

$$= \begin{cases} \psi_{|s-t|}\mu_4 & s \geq t \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}(\eta_{it}^2\epsilon_{is}\epsilon_{iu} \mid \mathbf{W}) = \mathrm{E}(\eta_{it}^2 (\sum_{d=0}^{\infty} \psi_d \eta_{i(s-d)})(\sum_{d=0}^{\infty} \psi_d \eta_{i(u-d)}) \mid \mathbf{W})$$

$$= \begin{cases} \sigma^4(\sum_{d=0}^{\infty} \psi_d \psi_{d+|s-u|} - \psi_{u-t}\psi_{|u-t|+|s-u|}) & s \geq u \geq t \\ + \mu_4 \psi_{u-t}\psi_{|u-t|+|s-u|} & \\ \sigma^4 \sum_{d=0}^{\infty} \psi_d \psi_{d+|s-u|} & s > t > u \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}(\eta_{it}\epsilon_{is}\epsilon_{iu}\epsilon_{iv} \mid \mathbf{W}) = \mathrm{E}(\eta_{it} (\sum_{d=0}^{\infty} \psi_d \eta_{i(s-d)})(\sum_{d=0}^{\infty} \psi_d \eta_{i(u-d)})(\sum_{d=0}^{\infty} \psi_d \eta_{i(v-d)}) \mid \mathbf{W})$$

$$= \begin{cases} \mu_4 \psi_{v-t}\psi_{u-t}\psi_{s-t} & \\ + \sigma^4(\psi_{s-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-v|+f} - \psi_{v-t}\psi_{u-t}\psi_{s-t}) & \\ + \sigma^4(\psi_{u-t} \sum_{d=0}^{\infty} \psi_d \psi_{|s-v|+d} - \psi_{v-t}\psi_{u-t}\psi_{s-t}) & s \geq u \geq v \geq t \\ + \sigma^4(\psi_{v-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-s|+d} - \psi_{v-t}\psi_{u-t}\psi_{s-t}) & \\ \psi_{s-t} \sum_d \psi_d \psi_{d+|u-v|} + \psi_{u-t} \sum_d \psi_d \psi_{d+|s-v|} & s > u > t > v \\ \psi_{s-t} \sum_d \psi_d \psi_{d+|u-v|} & s > t > u > v \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}(\epsilon_{is}\epsilon_{iu}\epsilon_{iv}\epsilon_{iw} \mid \mathbf{W}) = (\mu_4 - 3\sigma^4) \sum_{d=0}^{\infty} \psi_{d+|k|}\psi_{d+|k-i|}\psi_{d+|k-j|}\psi_d + \sigma^4 \sum_{d=0}^{\infty} \sum_{c \neq d}^{\infty} \psi_{d+|k|}\psi_{c+|j-i|}\psi_c\psi_d$$

$$+ \sigma^4 \sum_d \sum_c \psi_{c+|j|}\psi_{d+|k-i|}\psi_c\psi_d + \sigma^4 \sum_d \sum_b \psi_{b+|i|}\psi_b\psi_{d+|k-j|}\psi_d$$

$$= \mu_4 \pi_{1,suvw} + \sigma^4 \pi_{2,suvw}$$

$$
\begin{aligned}
\mathrm{E}[(\phi(L)\hat{\epsilon}_{it})^4 \mid \mathbf{W}] &= \mu_4 - 4\mu_4 \sum_{s\geq t} \psi_{s-t} P_{ts} \\
&\quad + 6 \sum_{s,u} (\sigma^4 \sum_{d=0}^{u-t-1} \psi_d \psi_{d+|s-u|} + \mu_4 \psi_{u-t} \psi_{|u-t|+|s-u|}) P_{ts} P_{tu} \\
&\quad - 4 \sum_{s,u,v} (\mu_4 \psi_{v-t} \psi_{u-t} \psi_{s-t} + \sigma^4 \psi_{s-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-v|+f} \\
&+ \sigma^4 \psi_{u-t} \sum_{d=0}^{\infty} \psi_d \psi_{|s-v|+d} + \sigma^4 \psi_{v-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-s|+d}) P_{ts} P_{tu} P_{tv} \\
&\sum_{s,u,v,w} (\mu_4 \pi_{1,suvw} + \sigma^4 \pi_{2,suvw}) P_{ts} P_{tu} P_{tv} P_{tw} \\
&= \mu_4 (1 - 4 \sum_{s} \psi_{s-t} P_{ts} + 6 \sum_{s,u} \psi_{u-t} \psi_{|u-t|+|s-u|} P_{ts} P_{tu} \\
&\quad\quad - 4 \sum_{s,u,v} \psi_{v-t} \psi_{u-t} \psi_{s-t} P_{ts} P_{tu} P_{tv} + \sum_{s,u,v,w} \pi_{1,suvw}) \\
&\quad + \sigma^4 (6 \sum_{s,u} \sum_{d=0}^{u-t-1} \psi_d \psi_{d+|s-u|} P_{ts} P_{tu} \\
&\quad\quad - 4 \sum_{s,u,v} (\psi_{s-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-v|+f} + \psi_{u-t} \sum_{d=0}^{\infty} \psi_d \psi_{|s-v|+d} + \psi_{v-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-s|+d}) P_{ts} P_{tu} P_{tv} \\
&\quad\quad + \sum_{s,u,v,w} \pi_{2,suvw} P_{ts} P_{tu} P_{tv} P_{tw})
\end{aligned}
$$

Similarly, as above, we know the results follow follow from the Khintchine's LLN as,

- $\phi(L)\ddot{\epsilon}_{it}\phi(L')\ddot{\epsilon}_{it}\phi(L'')\ddot{\epsilon}_{it}\phi(L''')\ddot{\epsilon}_{it}$ are iid across individuals by Assumption 3.2.1
- $\mathrm{E} \mid \eta_{it}\eta_{it} \mid < \infty$
- $\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\eta_{it}| \mid \mathbf{W}) < \left(\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|\eta_{it}^2| \mid \mathbf{W})\right)^{1/2}$
- $\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}z_{it}| \mid \mathbf{W}) \leq \left(\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\,\mathrm{E}(|z_{it}'(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2\mathbf{W})\right)^{1/2}$

Then, we can calculate the terms

$$\omega_2 = \frac{1}{n(T-p)} \sum_{i,t>p} (6 \sum_{s,u} P_{ts} P_{tu} (\sum_{d=0}^{\infty} \psi_d \psi_{d+|s-u|} - \psi_{u-t} \psi_{|u-t|+|s-u|})$$

$$- 4 \sum_{s,u,v} P_{ts} P_{tur} P_{tv} (\psi_{s-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-v|+f} + \psi_{u-t} \sum_{d=0}^{\infty} \psi_d \psi_{|s-v|+d} + \psi_{v-t} \sum_{d=0}^{\infty} \psi_d \psi_{|u-s|+d} - 3\psi_{v-t}\psi_{u-t}\psi_{s-t})$$

$$\omega_3 = \frac{1}{n(T-p)} \sum_{i,t>p} (1 - 4 \sum_{s} \psi_{s-t} P_{ts} + 6 \sum_{s,u} P_{ts} P_{tu} \psi_{u-t} \psi_{|u-t|+|s-u|}$$

$$- 4 \sum_{s,u,v} \psi_{v-t} \psi_{u-t} \psi_{s-t} P_{ts} P_{tu} P_{tv} + \sum_{s,u,v,w} \pi_{1,suvw})$$

As a result, we get the following system of equations,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1/\omega_1 \end{bmatrix} \begin{bmatrix} (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^4 \\ (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^2 \end{bmatrix} \to^p \begin{bmatrix} w_2 & w_3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_4 \\ \sigma^4 \end{bmatrix}$$

$$\begin{bmatrix} 1/\omega_2 & -\omega_3/\omega_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/\omega_1 \end{bmatrix} \begin{bmatrix} (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^4 \\ (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^2 \end{bmatrix} \to^p \begin{bmatrix} \mu_4 \\ \sigma^4 \end{bmatrix}$$

$$\begin{bmatrix} 1/\omega_2 & -\omega_3/(\omega_1\omega_2) \\ 0 & 1/\omega_1 \end{bmatrix} \begin{bmatrix} (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^4 \\ (n(T-p))^{-1} \sum_{i,t>p} \hat{\epsilon}^2 \end{bmatrix} \to^p \begin{bmatrix} \mu_4 \\ \sigma^4 \end{bmatrix}$$

□

3.8.3 Error Correction

Lemma 3.8.5. *Under our Assumptions,*

$$N^{-1} \sum_{i} \frac{1}{T-p} \sum_{t=p+1}^{T} \ddot{e}_{it}^{-} \ddot{e}_{it}^{-\prime} = \mathrm{E}(\sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-} \ddot{\epsilon}_{it}^{-\prime}) = (T-p)(\Gamma_p(\alpha) + \frac{1}{T-p} \Delta_\Gamma(\alpha))$$

$$N^{-1} \sum_{i} \sum_{t=p+1}^{T} \ddot{e}_{it}^{-} \ddot{e}_{it} = (T-p)(A(\alpha) + \frac{1}{T-p} A_\Gamma(\alpha))$$

where $\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha)$ *is a* $p \times p$ *matrix with* k, j *element*

$$\mathrm{E}(\frac{1}{T-p}\sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-}\ddot{\epsilon}_{it}^{-\prime}) = \gamma_{|i-j|}(\alpha) - \frac{1}{T-p}\mathrm{tr}(\Gamma_{-k}(\alpha)Z_{i,-j}(Z_i'Z_i)^{-1}Z_i)$$
$$- \frac{1}{T-p}\mathrm{tr}(\Gamma_{-j}(\alpha)Z_{i,-k}(Z_i'Z_i)^{-1}Z_i) + \frac{1}{T-p}\mathrm{tr}(\Gamma(\alpha)Z_i(Z_i'Z_i)^{-1}Z_{i,-k}'Z_{i,-j}(Z_i'Z_i)^{-1}Z_i')$$

$A(\alpha) + \frac{1}{T-p}\Delta_A(\alpha)$ *is a* $p \times 1$ *vector with ith element*

$$\mathrm{E}(\frac{1}{T-p}\sum_{t=p+1}^{T} \ddot{\epsilon}_{it}^{-}\ddot{\epsilon}_{it}') = \gamma_i(\alpha) - \frac{1}{T-p}\mathrm{tr}(\Gamma_{-k}(\alpha)Z_{i,-0}(Z_i'Z_i)^{-1}Z_i)$$
$$- \frac{1}{T-p}\mathrm{tr}(\Gamma_{-0}(\alpha)Z_{i,-k}(Z_i'Z_i)^{-1}Z_i) + \frac{1}{T-p}\mathrm{tr}(\Gamma(\alpha)Z_i(Z_i'Z_i)^{-1}Z_{i,-k}'Z_{i,-0}(Z_i'Z_i)^{-1}Z_i')$$

Lemma 3.8.6. *Let* $\hat{\alpha} = (\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\hat{\epsilon}_{it}^{-}\hat{\epsilon}_{it}^{-\prime})^{-1}(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\hat{\epsilon}_{it}^{-}\hat{\epsilon}_{it})$ *be the least squares estimate of* α *using the least squares residuals,* $\hat{\epsilon}_{it}$ *from estimating* β*. If the Assumptions hold,*

$$\hat{\alpha} = (\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{\epsilon}_{it}^{-}\ddot{\epsilon}_{it}^{-\prime})^{-1}(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{\epsilon}_{it}^{-}\ddot{\epsilon}_{it}) + o_o(N^{-1/2})$$

Lemma 3.8.7. *Define* $\ddot{\mu}_{it} = \ddot{\epsilon}_{it} - \ddot{\epsilon}_{it}^{-\prime}\alpha_T(\alpha)$*. If the Assumptions hold,*

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{\epsilon}_{it}^{-}\ddot{\mu}_{it} \to^{D} N(0, \Xi)$$

where $\Xi = \mathrm{E}(\sum_{t_1=p+1}^{T}\sum_{t_2=p+1}^{T}\ddot{\epsilon}_{it_1}\ddot{\mu}_{it_1}\ddot{\mu}_{it_2}\ddot{\epsilon}_{it_2}]$

Proposition 3.8.3. *If the Assumptions hold then* $\hat{\alpha} \to^{p} \alpha_T(\alpha)$*, where*

$$\alpha_T(\alpha) = \mathrm{E}\left[\ddot{\epsilon}_{it}^{-}\ddot{\epsilon}_{it}^{-\prime}\right]\mathrm{E}\left[\sum_{t=p+1}^{T}\ddot{\epsilon}_{it}^{-}\ddot{\epsilon}_{it}\right] = (\Gamma_p(\alpha) + \frac{1}{T-p}\Delta_\Gamma(\alpha))^{-1}(A(\alpha) + \frac{1}{T-p}\alpha_A(\alpha))$$

Proof of Lemma 3.8.5. Again, the proof is almost identical to Hansen (2004) Lemma 1.8.3.

$$
\begin{aligned}
\left[N^{-1}\sum_i \sum_{t=p+1}^{T} \ddot{e}^{-}_{it}\ddot{e}^{-\prime}_{it}\right]_{k,j} &= N^{-1}\sum_i \frac{1}{T-p}\sum_{t=p+1}^{T} \epsilon_{(t-k)}\epsilon_{i(t-j)} \\
&\quad + N^{-1}\sum_i \frac{1}{T-p}\sum_{t=p+1}^{T} \epsilon_{i(t-k)} z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i \\
&\quad + N^{-1}\sum_i \frac{1}{T-p}\sum_{t=p+1}^{T} \epsilon_{i(t-j)} z'_{i(t-k)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i \\
&\quad + N^{-1}\sum_i \frac{1}{T-p}\sum_{t=p+1}^{T} z'_{i(t-k)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}z_{i(t-j)} \\
&= N^{-1}\sum_i \frac{1}{T-p}\sum_{t=p+1}^{T} \epsilon_{(t-k)}\epsilon_{i(t-j)} \\
&\quad + N^{-1}\sum_i \frac{1}{T-p} Z'_{i,-j}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i,-k} \\
&\quad + N^{-1}\sum_i \frac{1}{T-p} Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_{i,-j} \\
&\quad + N^{-1}\sum_i \frac{1}{T-p} Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}Z_{i,-j} \\
&= \sigma_{k-j} + ((N(T-p))^{-1}\sum_i Z'_{i,-j}(Z_i'Z_i)^{-1}Z_i')\Gamma_{-k}(\alpha) \\
&\quad + ((N(T-p))^{-1}\sum_i Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i')\Gamma_{-j}(\alpha) \\
&\quad + \mathrm{tr}(\Gamma(\alpha)(N(T-p))^{-1}\sum_i Z_i(Z_i'Z_i)^{-1}Z_{i,-j}Z'_{i,-k}(Z_i'Z_i)^{-1}Z_i')
\end{aligned}
$$

Denote E^* to be the conditional expectation given X_i and Z_i. Then the results follow from the Khinchin Law of Large Numbers, repeated application of the triangle and Cauchy-Schwarz inequalities since,

- $\left[\sum_{t=p+1}^{T} \ddot{e}^{-}_{it}\ddot{e}^{-\prime}_{it}\right]_{k,j}$ is i.i.d across individuals under Assumption 3.2.1.
- $\mathrm{E}^*[\epsilon_{(t-k)}\epsilon_{i(t-j)}] < \infty$ for all k, j.
- $\mathrm{E}^*[\epsilon_{i(t-k)} z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i] \leq \left(\mathrm{E}^*(|z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2)\,\mathrm{E}^*(|\epsilon_{i(t-k)}|^2)\right)^{1/2}$

- $\mathrm{E}^*(z'_{i(t-k)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}z_{i(t-j)}) \leq \left(\mathrm{E}^*(|z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2)\,\mathrm{E}^*(|z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i|^2)\right)^{1/2}$

-

$$\begin{aligned}
\mathrm{E}^*(|z'_{i(t-j)}(Z_i'Z_i)^{-1}\boldsymbol{\epsilon}_i|^2) &= \mathrm{E}^*(|z'_{i(t-j)}(Z_i'Z_i)^{-1}(Z_i'Z_i)(Z_i'Z_i)^{-1}Z_i\boldsymbol{\epsilon}_i|^2)\\
&\leq ||z'_{i(t-j)}(Z_i'Z_i)^{-1}Z_i||^2\,\mathrm{E}^*(||Z_i(Z_i'Z_i)^{-1}Z_i\boldsymbol{\epsilon}_i||^2)\\
&= trace(z_{i(t-j)}(Z_i'Z_i)^{-1}z_{i(t-j)}) \times \mathrm{E}^*[trace(\boldsymbol{\epsilon}_i'Z_i(Z_i'Z_i)^{-1}Z_i'\boldsymbol{\epsilon}_i)]\\
&\leq L\,\mathrm{E}^*(\mathrm{tr}(\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i)) < \infty
\end{aligned}$$

Then, by the Law of Iterated Expectations, it follows that, $\mathrm{E}\left[\frac{1}{T-p}\sum_{t=p+1}^{T}\ddot{e}_{it}\ddot{e}_{it}'\right] < \infty$

□

Proof of Lemma 3.8.6. Same as Hansen (2004) Lemma 1.8.4. □

Proof of Lemma 3.8.7. Same as Hansen (2004) 1.8.5. □

3.8.4 Wald Test

Lemma 3.8.8. *Let Assumptions 3.3.2-3.2.2 hold. Then, for a balanced panel with N individuals observed T time periods,*

$$\mathrm{Var}(q^{-1/2}\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}) = q^{-1}\sum_{s,t,u,v}K_{stuv} - q^{-1}\left(\sum_i\left(\sum_{t,s}P^*_{ii,tt}P^*_{ii,ss} + 2\sum_{t,s}P^{*2}_{ii,ts}\right)\right) + 2(1+d_n)$$

Proof of Lemma 3.8.8. The usual variance formula is,

$$\mathrm{Var}(\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}) = \mathrm{E}\left((\boldsymbol{\epsilon}^{*\prime}A_n^*\boldsymbol{\epsilon}^*)^2 \mid \mathbf{W}\right) - \mathrm{E}\left(\boldsymbol{\epsilon}^*A_n^*\boldsymbol{\epsilon}^{*\prime} \mid \mathbf{W}\right)^2$$

Since $d_n = q/(n(T-L)-K)$,

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}\right) &= \mathrm{tr}(P^*I_{nT})\\
&= \mathrm{tr}(P^*) = \mathrm{tr}(P^*_{C^{-1\prime}\mathbf{W}'} - d_n(I_{nT} - P_{C^{-1\prime}\mathbf{W}'}))\\
&= \mathrm{tr}(P^*_{C^{-1\prime}\mathbf{W}'}) - d_n\,\mathrm{tr}(I_{nT} - P_{C^{-1\prime}\mathbf{W}'}) = q_n - d_n(n(T-L)-K) = q_n - q_n = 0
\end{aligned}$$

By symmetry of the errors across individuals,

$$\begin{aligned}
\mathrm{E}\left((\boldsymbol{\epsilon}^{*\prime} P^* \boldsymbol{\epsilon}^*)^2 \mid \mathbf{W}\right) &= \sum_i \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} P_{ii}^* \boldsymbol{\epsilon}_i^* \boldsymbol{\epsilon}_i^{*\prime} P_{ii}^* \boldsymbol{\epsilon}_i^* \mid \mathbf{W}) + 2 \sum_{i=2}^{n} \sum_{j<i} \mathrm{E}(\boldsymbol{\epsilon}_i^* P_{ij}^* \boldsymbol{\epsilon}_j^{*\prime} \boldsymbol{\epsilon}_i^* P_{ij}^* \boldsymbol{\epsilon}_j^{*\prime} \mid \mathbf{W}) \\
&= \sum_{t,s,u,v} \mathrm{E}(\epsilon_{it}^* \epsilon_{is}^* \epsilon_{iu}^* \epsilon_{iv}^* \mid \mathbf{W}) \sum_i P_{ii,ts}^* P_{ii,uv}^* + \sum_{i,j<i} \left(\sum_{t,s} P_{ii,tt}^* P_{jj,ss}^* + 2 \sum_{t,s} P_{ij,ts}^{2*} \right) \\
&= \sum_{t,s,u,v} \mathrm{E}(\epsilon_{it}^* \epsilon_{is}^* \epsilon_{iu}^* \epsilon_{iv}^* \mid \mathbf{W}) \sum_i P_{ii,ts}^* P_{ii,uv}^* - \left(\sum_i \left(\sum_{t,s} P_{ii,tt}^* P_{ii,ss}^* + 2 \sum_{t,s} P_{ii,ts}^{*2} \right) \right) \\
&\quad + \mathrm{tr}(P^*)^2 + 2\,\mathrm{tr}(P^{*2}) \\
&= \sum_{t,s,u,v} \mathrm{E}(\epsilon_{it}^* \epsilon_{is}^* \epsilon_{iu}^* \epsilon_{iv}^* \mid \mathbf{W}) \sum_i P_{ii,ts}^* P_{ii,uv}^* - \left(\sum_i \left(\sum_{t,s} P_{ii,tt}^* P_{ii,ss}^* + 2 \sum_{t,s} P_{ii,ts}^{*2} \right) \right) \\
&\quad + 2(1 + d_n)
\end{aligned}$$

Dividing by q gives the result. □

Lemma 3.8.9. *Let Assumptions 3.3.2-3.2.2 hold, then as $n, q \to \infty$, T fixed, then,*

$$\frac{\boldsymbol{\epsilon}^{*\prime} P_{C^{-1\prime}\mathbf{W}'}^* \boldsymbol{\epsilon}^* - q}{\sqrt{\mathrm{Var}(\boldsymbol{\epsilon}^{*\prime} P_{C^{-1\prime}\mathbf{W}'}^* \boldsymbol{\epsilon}^* \mid \mathbf{W})}} \Rightarrow N(0,1)$$

$$\boldsymbol{\epsilon}^{*\prime}(I_{nT} - P_{C\mathbf{W}})\boldsymbol{\epsilon}^*/(n(T-L)-K) \to^p 1$$

3.8.4.1 GLS

A previous version of this paper generated many of the primary results under a GLS transformation. Under this setting, $\hat{\Sigma}$ is estimated, and then a cholesky decomposition is taken, $\Sigma = CC'$, $\hat{\Sigma} = \hat{C}\hat{C}'$. Under this criteria, we can define $es = \hat{C}\boldsymbol{\epsilon}$, and the resulting projection matrix becomes based around $\hat{C}'\mathbf{W}$. This transformation provides a homogenization of the variance-covariance that ties results to previous work by Calhoun (2011).

Lemma 3.8.10. *Under Assumptions 3.3.2-3.2.2 are met. Then for any N , T,*

$$\mathrm{E}(F_n \mid \mathbf{W}) = 1$$

Proof of Proposition 3.8.10. I first prove results for the GLS Wald statistic, $W_{n,GLS}$. The numerator and denominator are independent as.

$$\begin{aligned}
&C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime}(I_{nT}-C^{-1}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1}) \\
&\quad = C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1} \\
&\quad - C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1}C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1} \\
&\quad = C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1} \\
&\quad - C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1} \\
&\quad = 0
\end{aligned}$$

We know $\mathrm{E}(\boldsymbol{\epsilon}^*) = 0$, by our GLS parameterization the resulting error structure is equivalent to the identity matrix, $\mathrm{E}(\boldsymbol{\epsilon}^*\boldsymbol{\epsilon}^{*\prime}) = C^{-1}(I_n \otimes \Gamma(\alpha))C^{-1\prime} = I_T$, and by usual expectations of a quadratic form, we have

$$\begin{aligned}
&\mathrm{E}(\boldsymbol{\epsilon}^{*\prime}C^{-1}\mathbf{W}(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime}\boldsymbol{\epsilon}^*/q_n \mid \mathbf{W}) \\
&\quad = \mathrm{tr}(C^{-1}\mathbf{W}(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n'(R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime})/q_n \\
&\quad = \mathrm{tr}((R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}'C^{-1\prime}C^{-1}\mathbf{W}(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')/q_n \\
&\quad = \mathrm{tr}((R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')^{-1}R_n(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}R_n')/q_n = 1
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\epsilon}^{*\prime}(I_{nT}-P_{C\mathbf{W}})\boldsymbol{\epsilon}^*/(n(T-L)-K) \mid \mathbf{W}) &= \mathrm{tr}(C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1})/(n(T-L)-K) \\
&= \mathrm{tr}((C^{-1\prime}\mathbf{W}'(\mathbf{W}'\Sigma^{-1}\mathbf{W})^{-1}\mathbf{W}C^{-1})I_n \otimes \Gamma(\alpha))/(n(T-L)-K) \\
&= 1
\end{aligned}$$

□

Lemma 3.8.11. *Let Assumptions 3.3.2-3.2.2 hold. Then, for a balanced panel with N individuals observed T time periods,*

$$\text{Var}(q^{-1/2}\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}) = q^{-1}\sum_{s,t,u,v} K_{stuv} - q^{-1}\left(\sum_i\left(\sum_{t,s} P^*_{ii,tt}P^*_{ii,ss} + 2\sum_{t,s} P^{*2}_{ii,ts}\right)\right) + 2(1+d_n)$$

Proof of Lemma 3.8.8. The usual variance formula is,

$$\text{Var}(\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}) = \text{E}\left((\boldsymbol{\epsilon}^{*\prime}A_n^*\boldsymbol{\epsilon}^*)^2 \mid \mathbf{W}\right) - \text{E}\left(\boldsymbol{\epsilon}^* A_n^*\boldsymbol{\epsilon}^{*\prime} \mid \mathbf{W}\right)^2$$

Since $d_n = q/(n(T-L)-K)$,

$$\begin{aligned}
\text{E}\left(\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^* \mid \mathbf{W}\right) &= \text{tr}(P^* I_{nT}) \\
&= \text{tr}(P^*) = \text{tr}(P^*_{C^{-1\prime}\mathbf{W}'} - d_n(I_{nT} - P_{C^{-1\prime}\mathbf{W}'})) \\
&= \text{tr}(P^*_{C^{-1\prime}\mathbf{W}'}) - d_n\,\text{tr}(I_{nT} - P_{C^{-1\prime}\mathbf{W}'}) = q_n - d_n(n(T-L)-K) = q_n - q_n = 0
\end{aligned}$$

By symmetry of the errors across individuals,

$$\begin{aligned}
\text{E}\left((\boldsymbol{\epsilon}^{*\prime}P^*\boldsymbol{\epsilon}^*)^2 \mid \mathbf{W}\right) &= \sum_i \text{E}(\boldsymbol{\epsilon}_i^{*\prime}P^*_{ii}\boldsymbol{\epsilon}_i^*\boldsymbol{\epsilon}_i^{*\prime}P^*_{ii}\boldsymbol{\epsilon}_i^* \mid \mathbf{W}) + 2\sum_{i=2}^n\sum_{j<i}\text{E}(\boldsymbol{\epsilon}_i^* P^*_{ij}\boldsymbol{\epsilon}_j^{*\prime}\boldsymbol{\epsilon}_i^* P^*_{ij}\boldsymbol{\epsilon}_j^{*\prime} \mid \mathbf{W}) \\
&= \sum_{t,s,u,v}\text{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W})\sum_i P^*_{ii,ts}P^*_{ii,uv} + \sum_{i,j<i}\left(\sum_{t,s}P^*_{ii,tt}P^*_{jj,ss} + 2\sum_{t,s}P^{2*}_{ij,ts}\right) \\
&= \sum_{t,s,u,v}\text{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W})\sum_i P^*_{ii,ts}P^*_{ii,uv} - \left(\sum_i\left(\sum_{t,s}P^*_{ii,tt}P^*_{ii,ss} + 2\sum_{t,s}P^{*2}_{ii,ts}\right)\right) \\
&\quad + \text{tr}(P^*)^2 + 2\,\text{tr}(P^{*2}) \\
&= \sum_{t,s,u,v}\text{E}(\epsilon^*_{it}\epsilon^*_{is}\epsilon^*_{iu}\epsilon^*_{iv} \mid \mathbf{W})\sum_i P^*_{ii,ts}P^*_{ii,uv} - \left(\sum_i\left(\sum_{t,s}P^*_{ii,tt}P^*_{ii,ss} + 2\sum_{t,s}P^{*2}_{ii,ts}\right)\right) \\
&\quad + 2(1+d_n)
\end{aligned}$$

Dividing by q gives the result. □

Lemma 3.8.12. *Let Assumptions 3.3.2-3.2.2 hold, then as $n, q \to \infty$, T fixed, then,*

$$\frac{\boldsymbol{\epsilon}^{*\prime}P^*_{C^{-1\prime}\mathbf{W}'}\boldsymbol{\epsilon}^* - q}{\sqrt{\text{Var}(\boldsymbol{\epsilon}^{*\prime}P^*_{C^{-1\prime}\mathbf{W}'}\boldsymbol{\epsilon}^* \mid \mathbf{W})}} \Rightarrow N(0,1)$$

$$\boldsymbol{\epsilon}^{*\prime}(I_{nT} - P_{C\mathbf{W}})\boldsymbol{\epsilon}^*/(n(T-L)-K) \to^p 1$$

Proof of Lemma 3.8.9. The two proofs are nearly identical, so for brevity we show only the first one. By symmetry of the orthogonal projection matrix we have,

$$q^{-1/2}(\boldsymbol{\epsilon}^{*\prime}P^*_{C^{-1\prime}\mathbf{W}'}\boldsymbol{\epsilon}^*-q) = q^{-1/2}\sum_i(\boldsymbol{\epsilon}_i^{*\prime}P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*-\sum_t P^*_{C^{-1\prime}\mathbf{W}',ii,tt})+2q^{-1/2}\sum_{i=2}^n\sum_{j<i-1}\boldsymbol{\epsilon}_i^{*\prime}P^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^* \tag{3.21}$$

The first and second summation are mean zero processes, and by Assumption 3.2.1, the two are uncorrelated. Therefore, we show each part follows a central limit theorem separately. The first term is a mean zero process as,

$$q^{-1/2}\sum_i \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime}P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*-\sum_t P^*_{C^{-1\prime}\mathbf{W}',ii,tt}) = \mathrm{tr}(P^*_{C^{-1\prime}\mathbf{W}'})-\mathrm{tr}(P^*_{C^{-1\prime}\mathbf{W}'}) = q-q=0$$

Define the Frobenious Norm of a $nT\times nT$ matrix A to be $||A|| = \sqrt{\mathrm{tr}(AA')}$. Therefore, by applying Cauchy-Schwarz and definition of the Frobenious Norm we have,

$$\begin{aligned}|\boldsymbol{\epsilon}_i^{*\prime}P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*| &\le ||\boldsymbol{\epsilon}_i^*||\,||P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*|| \\ &\le ||\boldsymbol{\epsilon}_i^*||\,||P^*_{C^{-1\prime}\mathbf{W}',ii}||\,||\boldsymbol{\epsilon}_i^*|| = ||\boldsymbol{\epsilon}_i^*||^2||P^*_{C^{-1\prime}\mathbf{W}',ii}||\end{aligned}$$

We know $P^*_{C^{-1\prime}\mathbf{W}',ii}$ is a sub-matrix of $P^*_{C^{-1\prime}\mathbf{W}'}$, which is an orthogonal projection matrix, which by construction has Frobenious norm 1. The Frobenious norm is also weakly increasing in the number of cell blocks, such that $||P^*_{C^{-1\prime}\mathbf{W}',ii}|| \le ||P_{C^{-1\prime}\mathbf{W}'}|| < 1$. This implies $|\boldsymbol{\epsilon}_i^{*\prime}P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*| \le ||\boldsymbol{\epsilon}_i^*||^2$. Then by Minkowski's Inequality and Assumption 3.3.2

$$\mathrm{E}\,|\boldsymbol{\epsilon}_i^{*\prime}P^*_{C^{-1\prime}\mathbf{W}',ii}\boldsymbol{\epsilon}_i^*|^{2+r} \le [\sum_t\{\mathrm{E}(|\boldsymbol{\epsilon}_{it}^{*2}|)^{2+r}\}^{\frac{1}{2+r}}]^{2+r} < \infty$$

And the first term converges to a standard Normal distribution by the Lindeberg-Feller CLT. For the second term, we define Λ_n to be a block diagonal matrix with the $T\times T$ sub-matrices along $P^*_{C^{-1\prime}\mathbf{W}'}$'s diagonal. Then, let $\tilde{P}^*_{C^{-1\prime}\mathbf{W}'} = P^*_{C^{-1\prime}\mathbf{W}'}-\Lambda_n$. We define the following martingale difference series.

$$U(n) = \sigma(n)^{-1} \sum_{i \leq n} \sum_{j \leq n} \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*$$

With $\sigma(n)^2 = \sum_{ij} \mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)^2 \mid \mathbf{W})$. From Jöckel and Sendler (1981) Proposition 3.2 it suffices to show that the following terms are all lower order than $\sigma(n)^4$.

$$DJ_I = \sum_{1 \leq i < j \leq n} \mathrm{E}[(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)^4]$$

$$DJ_{II} = \sum_{1 \leq i < j < k \leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)^2 (\boldsymbol{\epsilon}_i^{*} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik} \boldsymbol{\epsilon}_k^*)^2 + \mathrm{E}(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ji} \boldsymbol{\epsilon}_i^*)^2 (\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',jk} \boldsymbol{\epsilon}_k^*)^2$$
$$+ \mathrm{E}(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ki} \boldsymbol{\epsilon}_i^*)^2 (\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',kj} \boldsymbol{\epsilon}_j^*)^2$$

$$DJ_{IV} = \sum_{1 \leq i < j < k < l \leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik} \boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk} \boldsymbol{\epsilon}_k^*)$$
$$+ \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',il} \boldsymbol{\epsilon}_l^*)(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',kj} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_k^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',kl} \boldsymbol{\epsilon}_l^*)$$
$$+ \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik} \boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',il} \boldsymbol{\epsilon}_l^*)(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',jk} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_j^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',jl} \boldsymbol{\epsilon}_k^*)$$

We now show that DJ_I and DJ_{II} are of lower order than $\sigma(n)^4$. Let us have some sequence $K(n)$ where as $n \to \infty, K(n) \to \infty$. We first show that these terms follow a Lindberg-style condition.

$$\max_{1 \leq i < j \leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| > K(n) \sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W}))$$
$$= \max_{1 \leq i < j \leq n} \mathrm{E}(\mathrm{tr}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_i^{*\prime}) I_{|\,\mathrm{tr}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_i^{*\prime})| > K(n) \sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W}))$$
$$= \max_{1 \leq i < j \leq n} \mathrm{tr}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \, \mathrm{E}(\boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_i^{*\prime} I_{|\,\mathrm{tr}(\boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_i^{*\prime})| > K(n)} \mid \mathbf{W}) / (\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts})) \to^p 0$$

Where convergence in probability comes from Assumption 3.3.2 such that $\mathrm{E}(\boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_i^{*\prime} I_{|\,\mathrm{tr}(\boldsymbol{\epsilon}_j^* \boldsymbol{\epsilon}_i^{*\prime})| > K(n)} \mid \mathbf{W}) \to^p 0$. This condition ensures that a tail-truncated series converges to the full sequence in L^2.

$$\mathrm{Var}(\sum_{ij} \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* - \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| < K(n) \sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})$$

$$\leq \mathrm{Var}(\sum_{ij} \boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| > K(n) \sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})$$

$$\leq \sum_{ij} \sigma_{ij}^2 (\max_{1 \leq i < j \leq n} \sigma_{ij}^{-2} \, \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| > K(n) \sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W}))$$

$$= o(\sigma(n))$$

Since the truncated sequence converges to the full one in L^2, it suffices to show that $DJ_I',\ DJ_{II}',\ DJ_{III}'$ are all lower order than $\sigma(n)^4$. From above we know

$$\max_{1 \leq i < j \leq n} \mathrm{E}(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^* I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| > K(n) \sum_{t,s} P^{*}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \rightarrow^p 0$$

Which in turn implies

$$\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)^2 I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| \leq K(n) \sum_{t,s} P^{*}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \leq K(n) (\sum_{t,s} P^*_{C^{-1\prime}\mathbf{W}',ij,ts})^2$$

But this implies

$$\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)^4 I_{|\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*| \leq K(n) \sum_{t,s} P^{*}_{C^{-1\prime}\mathbf{W}',ij,ts}} \mid \mathbf{W})) \leq K(n)^2 (\sum_{t,s} P^{*2}_{C^{-1\prime}\mathbf{W}',ij,ts} hu)^4$$

But this implies through the Cauchy-Schwarz theorem, that $DJ_I',\ DJ_{II}'$ are of lower order than $\sigma(n)^4$. Finally, for DJ_{IV}' we know,

$$\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik} \boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj} \boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_l^{*\prime} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk} \boldsymbol{\epsilon}_k^*) \mid \mathbf{W}) =$$

$$\sum_{t,s,u,v,w,x,y,z} \mathrm{E}(\epsilon_{it}^* \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij,ts} \epsilon_{js}^* \epsilon_{iu}^* \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik,uv} \epsilon_{kv}^* \epsilon_{lw}^* \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj,wx} \epsilon_{jx}^* \epsilon_{ly}^* \tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk,yz} \epsilon_{kz}^* \mid \mathbf{W})$$

Where,

$$\mathrm{E}(\epsilon^*_{it}\epsilon^*_{i\tau} \mid \mathbf{W}) = \begin{cases} 1 & t = \tau \\ 0 & \text{otherwise} \end{cases}$$

Implies

$$\begin{aligned} &\mathrm{E}((\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_i^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik}\boldsymbol{\epsilon}_k^*)(\boldsymbol{\epsilon}_l^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj}\boldsymbol{\epsilon}_j^*)(\boldsymbol{\epsilon}_l^{*\prime}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk}\boldsymbol{\epsilon}_k^*) \mid \mathbf{W}) \\ &= \sum_{t,s,u,v} \tilde{P}^*_{C^{-1\prime}\mathbf{W}',ij,ts}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',ik,tu}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lj,vs}\tilde{P}^*_{C^{-1\prime}\mathbf{W}',lk,vu} \end{aligned}$$

But then again we know that the eigenvalues of $\lambda_{\max}(\tilde{P}^*_{C^{-1\prime}\mathbf{W}'}) \le |\lambda_{\max}(P^*_{C^{-1\prime}\mathbf{W}'}) - \lambda_{\min}(\Lambda)| \le 1$, and the matrix being idempotent implies that again these sums are bounded by similar logic as above. As a result we know these sums are bounded, and are lower order than $\sigma(n)^4$. Thus G'_{IV} is of lower order than $\sigma(n)^4$, completing the proof.

□

Proof of Theorem 3.4.1. From Lemma (3.8.8) we know that the numerator is conditionally mean zero, with variance η^2. Lemma (3.8.9) shows that the numerator is asymptotically normal, and that the denominator converges in probability to 1, completing the proof. □

CHAPTER 4. DO NUDGES INDUCE SAFE DRIVING? EVIDENCE FROM DYNAMIC MESSAGE SIGNS

Sher Afghan Asad and Kevin D. Duncan

Iowa State University

Modified from a manuscript to be submitted to *Eastern Economics of the Journal*

Abstract

Behavioral economics has transformed the way we think about policy problems of our age. Governments all over the world are using nudges, one of the tools from behavioral economics, to direct people's behavior towards socially desirable outcomes. However, it is not clear what kind of nudges are most effective, if at all. In this research, we look at the traffic-related messages such as "drive sober," "x deaths on roads this year," and "click it or ticket," displayed on major highways, on reported near-to-sign traffic accidents. This provides estimates of the impact of different types of nudges on road safety behavior. To estimate the causal effect of these nudges, we build a new high-frequency panel data set using the information on the time and location of messages, crashes, overall traffic levels, and weather conditions using the data of the state of Vermont over a three year time period. We estimate models that control for endogeneity of displayed messages, or allow for spillover effects from neighboring messages. We find that all nudges are at best ineffective in reducing the number of crashes. Our findings are robust to many different specifications and assumptions.

4.1 Introduction

Egan (2017) define nudges as "choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives." These benign behavioral interventions have become increasingly popular with the researchers and the governments all over the world to address various policy problems. There have been hundreds of studies which show that nudges are effective in influencing behavior ranging from donating organs (Johnson and Goldstein, 2003), reducing energy consumption (Allcott and Mullainathan, 2010), and saving more money (Thaler and Benartzi, 2004). While there is considerable evidence on the effectiveness of nudges, there is not much known about what kind of nudges are most effective and if poorly designed nudges can backfire. In this paper, we study the effectiveness of different nudges on road safety behavior and examine whether all nudges are created equal and if not which ones are most effective in ensuring road safety.

Nudges have been used all over the world for decades as a part of road safety management and to give useful information to drivers. Recently governments are increasingly adopting intelligent transportation systems (ITS) to provide real time information to the drivers on the road. As a part of ITS, different state governments in the United States have installed dynamic message signs (DMS) on various highways. These dynamic signs have become a regular medium through which the government provide information, such as updated time to destination or road conditions, or behavioral nudges, such as reminders to wear a seat belt or how many individuals have died on the road this year, to the drivers. The purpose of DMS is to reduce driver anxieties related to commutes, and to encourage safer driving. These signs generally face broad public approval (Benson (1997) Tay and De Barros (2008)), however despite this popularity, whether or not these signs actively encourage safer driving is unknown. Even small changes to driver behavior can have large effects societal welfare through decreasing road injuries and deaths. In 2019, over 37,000 people died in road accidents in the United States, and another 2.35 million individuals were injured or disabled due to road incidents. Since 2010 motor vehicle accidents have ranked 11th overall as a cause of death, and 6th in terms of years of life lost.[1]

In this paper, we identify the causal impact of two types of nudges on near-to-sign reported traffic incidents. In particular we categorize each message on the DMS into two kinds of nudges, behavioral and informational. Behavioral Nudges constitute messages which are aimed at encouraging drivers to drive safe without any concrete information on driving conditions, such as "Buckle

[1]https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812203

Up", "Drive Sober", and "Click It or Ticket." Informational Nudges consist of messages which provide concrete information on driving conditions, such as information on road conditions, weather conditions, upcoming events, and road diversions. The main focus of this paper is to examine whether behavioral nudges and/or informational nudges are effective at reducing traffic incidents.

The distinction between behavioral nudges and information nudges has its roots in the psychology literature, where different message types may trigger a different response from drivers. Behavioral nudges are supposed to encourage individuals to take precautionary measures and drive safer, however poorly thought out nudges can have opposite effects. Nudges can be interpreted as condescending and may invoke negative reaction (Dholakia, 2016). Individuals may feel that they don't like to be told what to do or how to drive and may indulge in over-speeding or more reckless driving. Informational nudges are more direct and are aimed at providing information relating to traffic, weather, or excess road risk. This may cause drivers to drive more carefully in light of information, or drive less carefully if the information changes their prior on riskiness in the opposite direction.

We estimate the causal impact of nudges on near to sign accidents on either the mile, or quarter mile, directly after a DMS using a Poisson fixed effects model. The main challenge in estimating the causal effect is to address the joint determination of nudges and crashes. To account for this we further include models with site specific trends, or nest a Difference-in-Difference approach that indexes post DMS treatment effects to be relative to the road area just before a DMS. Finally,, informational nudges (such as about weather and crashes on the road) are selectively displayed when conditions are precarious while behavioral nudges are selected in otherwise less risky conditions. To control for plausible contemporaneous assignment of messages to road conditions, we motivate a model a Poisson model with sequential exogeneity using generalized method of moments. Estimation is carried out using a new high frequency panel data set that covers the population of displayed DMS message signs, including exact display times, and reported traffic accidents from January 2016 to December 2018 in the state of Vermont. This further pools geocoded information on hourly traffic, weather, and temperature data. We map each reported crash to the potential

DMS using information on road network incorporating driving distance, driving time, and number of turns between signs and the location of the crash. This allows pairing the location and the timing of the crash and the nudge to evaluate the treatment assignment of each driver before getting into a crash.

Our results show that behavioral nudges have either no to small negative effects on traffic accident rates on the road area just after a DMS. Across a variety of specifications message content of behavioral nudges do not meaningfully alter reported accident rates immediately after the sign. Our mainline specifications imply a decrease in accidents ranging from a 9-40% decrease in near-to-sign accidents thanks to behavioral nudges, and a 35-150% increase in accidents caused by Informational Nudges- though across both specifications some of these effects are indistinguishable from zero. In our preferred specification, Behavior Nudges have a 40% decrease in accidents, and Information Nudges do not impact near-to-sign accidents, which amounts to about 22 fewer traffic accidents a year.Over the three year period, this amounts to roughly 52 fewer property-only accidents, 12.5 fewer injuries, and .66 fewer deaths an additional hours worth of behavioral nudges are displayed every day. We run alternate specifications that decompose results into heterogeneous effects by message sub type, and show that there exists moderate heterogeneity between message types. Many of the behavioral messages remain statistically insignificant. Comparably, results for informational messages are shown to be a mix of a strong, endogenous response caused by "Crash Ahead" messages, but that "Other Caution" messages do provide information to drivers and decrease near-to-sign accidents. Across other informational messages, we find no effects.

This paper adds to the literature in a few meaningful ways. First, we provide new estimates of the impacts of DMS message content on near-to-sign traffic incidents outside of the initial reason for sign roll out. Many previous studies have focused on evaluating a single message type, while often deployed DMSs display a rotating array of messages. Secondly, our estimates are tied directly to when messages are displayed. By pairing over 600 reported traffic incidents in the mile after a DMS and known start and end times of different messages, we are able to more accurately tie treatment status at time of crash to the displayed message at the time. Finally, within the observational setting of our data, we discuss ways to accommodate the endogenous selection of displayed message to traffic and other time varying hazards that might impact accident rates immediately following a message board. Combined, these provide precise estimates of revealed driver behavior in response to the small nudges provided by DMS systems.

Due to the proclivity of many states to set up, maintain, and use DMS networks to provide message content to drivers, there is strong interest in understanding the impacts of DMS nudges on driver behavior. Setting up, running, and maintaining these systems is not cost-free, and observational work on the effects of DMS networks often comes to conclusions that displayed messages do not improve driver safety. Hall and Madsen (2020) estimate the impacts of Death Toll reminders in Texas on near-to-sign traffic incidents. Using exogenous roll out of Death Toll reminders, they find strong increases in the number of traffic incidents even 10 kilometers after a DMS. Their identification strategy relies on the relative increase in Death Toll reminders as a share of total displayed messages, versus causally identified with times when Death Toll reminders are actually up. Secondly, the lack of effectiveness of behavioral nudges is consistent with at-least one previous study that indicate that message signs might cause additional collisions. Song et al. (2016) and Erke et al. (2007) show reading messages on DMSs may lead to a slowing down and speeding up effect among drivers, potentially making roads more dangerous around the signs. Norouzi et al. (2013) show no treatment effect using both on/off analysis, and comparing downstream traffic incidence to near-to-sign traffic incidents. Fallah Zavareh et al. (2017) examine how people respond to DMSs with road risk ratings. Risky behavioral adaptations were observed under low and medium risk messages during night time. The effects of high risk messages were consistently related to safety adaptations. The effects of messaging on rear-end collisions were significant only in the fast lane at night time. Overall, observational studies of the impacts of DMSs on near-to-sign traffic incidents often indicate that these signs have detritus effects on driver behavior and lead to an increase in such incidents.

The broader literature on the effectiveness of nudges in the transportation literature generally uses a mixture of simulation and stated preference surveys, or observational data comparing before and after using simulated traffic and accident data (see Mounce et al. (2007)). Compared to the observational work discussed above, simulation and stated preference surveys often find quite strong and positive evidence on message boards on driver behavior (Benson (1997); Peng et al. (2004); Hassan et al. (2012); Xu et al. (2018); Tarry and Graham (1995); Bonsall (1992)). Recent work by Choudhary et al. (2019) combines experimental design and revealed observational driving changes from interventions. They randomly given driving quality feedback messages on driver's smartphone, showing that personalized nudges generally improve driving performance compared to the control group.

Behavioral and informational responses on roads have been studied in many other contexts. Changing incentives for risky driving is common through examining how budget shortfalls and resulting decreases in police staffing impact safe driving (Makowsky and Stratmann (2011),DeAngelo and Hansen (2014)), that reduction in accidents following texting bans are short-lived Abouk and Adams (2013), and that scaling DUI punishments associated with how far over the legal limit a driver registers impact recidivism and future driving behavior Hansen (2015). Understanding and improving driver responsiveness to DMSs may lead to moderate reductions in traffic accidents, injuries, and fatalities. Alternatively, towns may be overstating belief in DMS value to drivers. De Borger and Proost (2013) show that the city government over-invest in externality reducing infrastructure whenever this infrastructure increases the generalized cost of through traffic. We can therefore expect an excessive number of speed bumps and traffic lights, but the right investment in noise barriers. In turn, we would expect higher rates of DMSs to exist along roads than socially optimal, and understanding these effects might help governments and public policy groups set more socially optimal levels of message signage. While in our study most of the DMSs are fixed, new hazards alternatively might cause additional road incidents, in this setting Xu and Xu (2020) evaluates how the introduction of new fracking wells is associated with near-to-well fatal car accidents.

The remainder of the paper proceeds as follows. Section 4.2 provides background information on the DMS system in Vermont and details on data and their sources. Section 4.3 describes the empirical strategy and series of models to be estimated. Section 4.4 provides the estimation results of our models. Section 4.5 provides various robustness check and finally Section 4.6 concludes.

4.2 Data

This section provides detailed information on how Dynamic Message Signs (DMSs) location and message content as decided upon by the Vermont Agency of Transportation (VTrans). This is collected from a series of primary documents and direct communication with the agency. We further describe our accident, traffic, and weather data, and how these data are combined. Finally, we provide basic descriptions of the final variables that we use in our estimation procedures.

4.2.1 Dynamic Message Sign Location

The installation of DMSs are a part of VTrans' effort under the Intelligent Transportation System (ITS) to facilitate drivers with updated and timely information on traffic and road conditions.[2] VTrans initially deployed these boards with portable installations with the aim to eventually phase in permanent installations. The message boards covered in this study are all portable installations - called portable variable message signs (PVMS).[3] The signs are typically mounted on trailers or pads, often with the wheels removed and secured in place for longer duration of use. Typically, PVMS run on solar power or battery. The PVMS have the ability for an adjustable display rate, which is typically set to allow for the message to be read at least twice at the posted speed limit.

The location of the message board is determined based on multiple factors including frequency of crashes and weather related incidents on a road segment. The detailed plan of location choice is provided in Vanasse Hangen Brustlin (2007). Broadly, the general location is determined by identi-

[2] In particular, the DMSs are primarily aimed at providing information on i) road conditions, ii) adverse weather notifications, iii) incident management, iv) in-route emergency evacuation information, iv) national missing and exploited children alert system - amber alerts, v) special events, vi) flight, train, and bus schedules in transportation terminals, vii) congestion management, viii) construction information/detours, ix) road closures, and x) special messages (such as variable speed limits, etc).

[3] Throughout our sample, the location of individual PVMS are fixed.

fying areas where it warrants weather notification to the drivers of hazardous conditions, advanced notification of substandard roadway conditions and upcoming “chain up” areas can be provided to the truck drivers, notification of construction and planned events can be provided to avoid congestion on relevant roads, or notifications can complement counties’ transportation management plans involving traffic and roadside safety. According to officials at VTrans, “*our goal was just to place them (DMSs) in high traffic areas and close to RWIS (Road Weather Information System). The placement of RWIS was based on high crash areas.... Going forward the goal was decided to place DMS before on/off ramps on interstates and close to major intersections on secondary highways.*” This suggests that these message boards are installed in the areas which are more susceptible to crashes.

The specific location of the message board is determined considering horizontal and vertical alignment of the message board. Typically, PVMS is visible from approximately 0.5 miles (or 2,500 feet) under both day and night conditions. The message is legible from a minimum distance of an 1/8th of a mile (or 650 feet). When possible, the PVMS signs are placed behind guardrail sections or outside the clear zone for errant vehicles. PVMS are mounted in such a way that the bottom of the message sign panel is minimum of seven feet above the roadway. Once the location of the DMS is determined, the next issue is about the content of messages that needs to be displayed on a particular DMS.

4.2.2 Message Data

Based on the conversations with officials at VTrans, the choice of message is determined based on risk factors such as road and weather conditions. For example, if the road conditions are more susceptible to accidents because of icy roads then drivers will be cautioned about the slippery conditions of the road. Behavioral nudges (such as nudging individuals to drive sober or notifying traffic death counts) are considered low priority messages and are only displayed when there is no other important information that needs to be conveyed. Some of the nudge messages such as “Click It and Ticket”, “Drive Sober or Get Pulled Over”, etc are based on national campaigns run by National Highway and Traffic Safety Administration (NHTSA). NHTSA run regular campaigns countrywide to raise awareness on drunk driving, seat belts etc. Just like other nudge messages, the campaign messages are displayed if the message boards are not being used for more important informational messages such as construction, crashes, winter weather, etc.

The data on messages is obtained from VTrans from June 2016 to December 2018. Messages were displayed on 67 unique sites during this time period. Table 4.1 presents the number and duration of various messages during the time period. During this period, there were total of 10,409 messages spanning 308,800 hours. Figure 4.1 shows the durations during which the message boards were active. It's clear that there is considerable heterogeneity in the activity and duration of messages across these message boards.

As shown in Table 4.1 and Figure 4.1 we categorize each message into different groups. "Death toll reminder" provide information about the number of people who have died that year from traffic accidents. "Seat belt reminders," "Texting reminders," "Speed reminders," and "Drinking reminders" aim to encourage seat belt use, no texting, staying under the speed limit, and sober driving, respectively. "Road condition message" displays information about the road characteristics such as gravel road. "Weather message" display the current weather conditions, and are most frequent in winter season due to snow and icy roads along with precipitation levels. "Traffic message" display information about traffic congestion or delays. Comparably, "Work zone message," "Road closure message," and "Crash message" inform about the upcoming work areas, upcoming road closures, or if there is a crash ahead. Finally, "Other caution message" is any other message which does not fall into the above category such as "Drive Safe" or "Better late than sorry". Some messages can have overlapping content between these types, for example, "40 Lives Lost in 2017 Buckle Up" is categorized in both "Death related message" and "Seat belt related message."

For our analysis we further categorize each message type into two kinds of nudges, behavioral and informational. Behavioral Nudges are nudges which are aimed at encouraging drivers to drive safe without any concrete information on driving conditions. Death, seat belt, texting, drinking, and speed are categorized as Behavior Nudges. Information Nudges are nudges which provide concrete information on driving conditions to the drivers. Road condition, weather, traffic, work zone, road closure, crash ahead, other caution, and other messages are categorized as Information Nudges, which broadly categorizes each of the message types into either reminders, or informational messages.

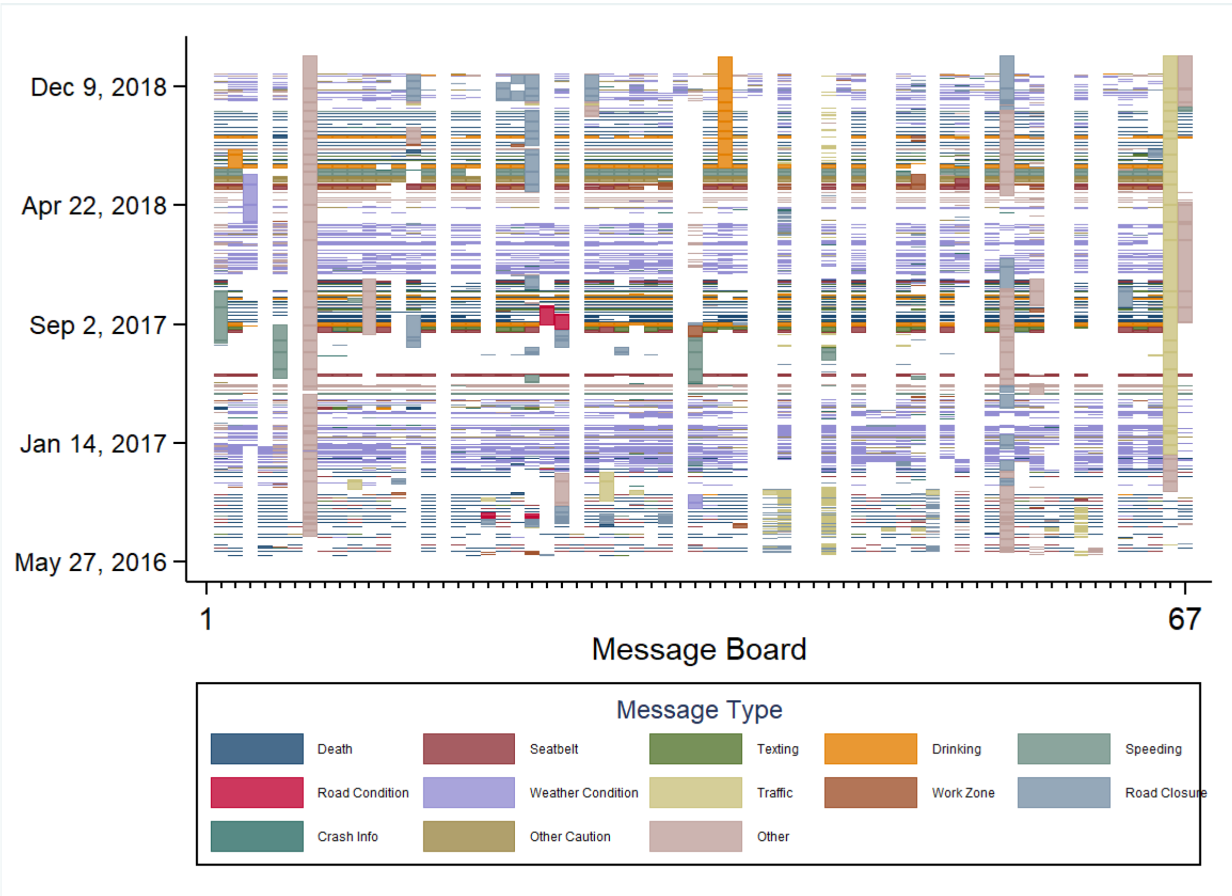

Figure 4.1 Message Boards Activity

Notes: The figure shows the time periods during which any message was displayed on the message board. Each bar represents a message board, with white areas indicating no message during the time period.

Table 4.1 Summary of Messages

	(1) Proportion of Number of Messages	(2) Proportion of Duration of Messages (in hours)
Death related message	0.19	0.11
Seatbelt related message	0.03	0.07
Phone related message	0.02	0.04
Drinking related message	0.02	0.09
Speed related message	0.16	0.14
Road condition message	0.05	0.03
Weather message	0.48	0.17
Traffic message	0.06	0.08
Work zone message	0.02	0.03
Road closure message	0.05	0.09
Crash message	0.01	0.00
Other caution message	0.03	0.04
Other message	0.09	0.25
Total	10409.00	308799.85

Notes: The table presents the number and duration of various messages during the time period from June 2016 to December 2018 in Vermont.

4.2.3 Crash Data

Data on crashes between the periods January 2016 and December 2018 is also obtained from VTrans. This data set reports wide set of details from the police reports about the crash including location, time and date, road conditions, weather conditions, driver details and condition, vehicle details, number and nature of injuries, and number of passengers involved. There were total of 35,554 police reports, that involved 64,027 vehicles, of crashes during this time period in the state of Vermont.

Since the data on crashes come from police reports in a busy field setting of a crash site, the spatial location of each crash may not always be accurate. For our purposes, the exact location of a crash is crucial to be able to map the crash to a potential message that may have been seen by the driver before getting into the crash. Here we describe the measures that we have taken to validate the spatial location of each crash.

First, VTrans has taken steps to geocode precise crash location for the recent data in their efforts to improve the quality of data for traffic safety and analysis purposes.[4] We use this data to get precise spatial location of most (35,202) crash sites during the said time period. Second, we are able to update geographic location of 5,999 police reports using spatial location of overlapping subset of crash data provided on VTrans Public Query Tool.[5] Third, few of the geographic coordinates in our data are using State Plane Coordinate System (rather than GPS coordinate system), we convert those to GPS coordinates and are able to update spatial location of 31 crash sites. Forth, there are cases for which crash location is provided in the text fields but with missing coordinates. We use ArcGIS® to geocode these addresses and are able to update spatial location of 140 crash sites.

To check for the validity of the coordinates from the above sources we reverse geocode the GPS coordinates using ArcGIS® and find that coordinates of 94 crash sites are either not street addresses or fall outside the county (within Vermont) in which they are supposed to lie (as determined on the basis of county of crash site). We then, once again, geocode the addresses for which either GPS coordinate is missing, not a street address, or found to lie outside the respective county and are able to find locations of 42 crash sites. The above measures leave us with missing or incorrect location for 347 crash sites which are manually looked at using information on various address fields.[6] Given the information, we remain unable to manually locate 82 crash sites, i.e., overall, we are able to locate 35,472 crashes (99.9 percent) with reasonable degree of accuracy.

[4] The data set is available at https://geodata.vermont.gov/datasets/

[5] This data set can be extracted from http://apps.vtrans.vermont.gov/CrashPublicQueryTool/

[6] We use information on street address and distance from intersecting street to manually locate the crash location on the Google Maps. In case of missing information about the street address the nearest intersecting street information is used to approximate the location of the crash. We use Google Map's measurement feature to measure the offset from the intersection based on the information provided, for example, 100 feet south of 1st St. and 1st Ave. We also use the measurement feature to locate addressed based on mile markers, such as, I-89 South, Mile Marker 65.3. We find base mile markers by using a map of Vermont's interstate exits and rest areas which is then located on Google Maps to get a reference mile marker and a measure of specified distance to the target mile marker.

Out of the located crash sites, 8 police reports lack information on the date and time of crash and therefore we drop them. Our final data set of crashes has 35,472 crash reports majority of which (approximately 79%) involved "property damage" only while the rest of them constitute injuries (approximately 20%) and fatalities (approximately 1%). The geographical location of each crash along with message board location in visually presented in Figure 4.2. As is typical of the collision data, the crashes are clustered around each other. The value of the nearest neighbor index is 0.11 ($z = -426.66$) which represents high degree of clustering of crashes around each other (Clark and Evans, 1954).

The contributing circumstances for the crashes as recorded by the VTrans are presented in Table 4.2. Factors such as fast driving, failure to yield, failure to keep in proper lane, following too closely, and inattention are some of the major factors contributing to the crashes.

Table 4.2 Contributing Circumstances to the Crash

	(1) Proportion
No improper driving	35.62
Inattention	12.47
Other improper action	11.06
Driving too fast for conditions	9.42
Failed to yield right of way	8.23
Failure to keep in proper lane	7.77
Other	5.97
Followed too closely	5.74
Under the influence of medication/drugs/alcohol	1.86
Visibility obstructed	1.52
Other Activity- Electronic Device	0.33
Distracted	0.02
Total	100.00

Notes: This table presents the share of each contributing factor in a crash as recorded by the VTrans.

4.2.4 Combining Message and Crash Data

The purpose of this study is to assess the impact of a particular nudge on the probability of a crash. To analyze that, we restrict our analysis to the regions where a message board is installed for

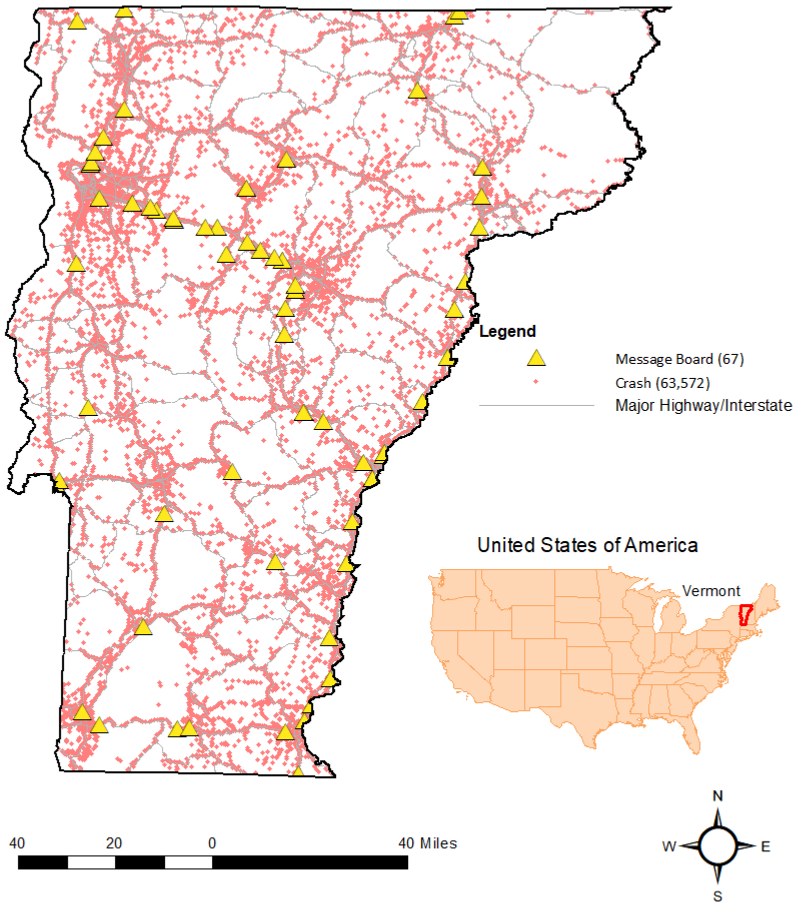

Figure 4.2 Map of Message Boards and Crashes

Notes: This map of Vermont represents crashes and message boards throughout the period between June 2016 and December 2018.

at-least some duration during the study time period. This implies that we use 67 locations/message-boards around which we focus our analysis. We map crashes to these 67 locations using ArcGIS® 'Find Closest Facility' tool. This tool finds one or more crashes that are closest from a message board based on travel distance (and travel time), and outputs the driving directions between the message board and the crash. When finding closest crashes, we specify to find closest crashes within a 10 mile distance to or from a message board and then restrict to crashes which are maximum of two turns away from the message board. We restrict to a maximum of two turns to be reasonably confident of a driver having read the message before getting into the crash. We also adjust for the message read time by adjusting the time of crash by the travel time from the message board to the location of the crash. We also assign the status of "pre" or "post" to each crash to determine whether the crashed occurred before the mapped message board or after. We use driving directions along with direction of travel of a vehicle to determine the pre/post status. Approximately 1,700 traffic accidents are mapped across the mile bandwidth directly before, and after all of the message boards in the sample. The mile just prior to all of our DMSs experiences 1,034 accidents, while the mile after has 627 reported traffic accidents over the sample time period.

The average number of crashed vehicles by different message types is presented in Figure 4.3. Of note, a clear case of endogeneity can be seen using crash info message which is displayed when there is a crash ahead Secondly, across the remaining message types, there remains large heterogeneity in the number of crashes naively correlated when those messages are displayed. Death, drinking, and speeding reminders all seem to have about equivalent average crashes/hour associated with them. Of the Behavior Nudges, Texting and Seat belt Reminders have about half of the average crashes. Similar stratification exists among Information Nudges. Multiple, Other, Other Caution, Road Closure, and Work Zone messages similarly have equivalently naive road hazard associated with them. Displayed Weather messages have the highest, likely related to being displayed when weather related hazards on the road are the most perilous. Comparably, Road Condition and Traffic messages have very low perceived correlation between being displayed, and number of accidents at the time.

Generally these signs go up endogenously in response to other perceived, or lack of, road hazards. Therefore, comparison of means is not a sufficient measure to judge their impacts on near-to-sign crashes, and more robust models and estimators need to be employed.

4.2.5 Traffic and Weather Data

Traffic on a particular road is considered to be one of the crucial factors that can effect the probability of a crash. The traffic data is obtained from the VTrans which has installed traffic counters on various highways in the state of Vermont. This data covers hourly road volume counts

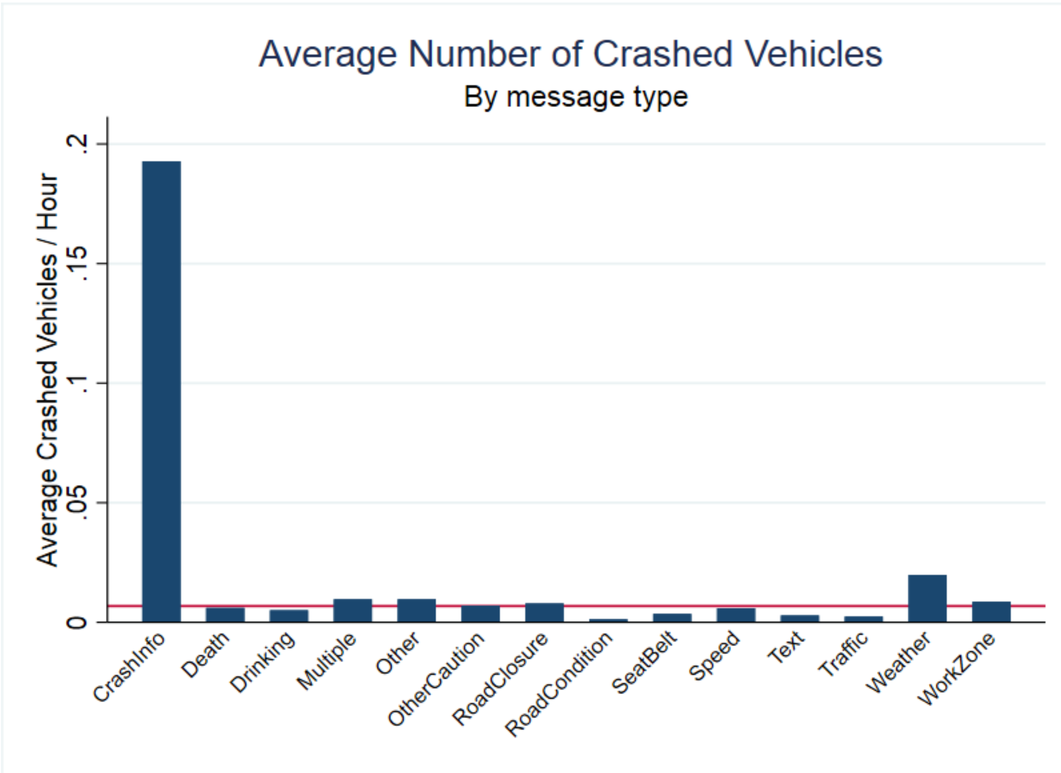

Figure 4.3 Average Number of Crashed Vehicles by Message Type

Notes: This figure presents the average number of crashed vehicles by message type within 10 miles post the message sign. Red line indicate the average crashed vehicles per hour when there was no message displayed at the time the driver may have passed the location of the message board.

across 86 sites in Vermont over the entire duration of our traffic and message board data. The traffic volume on an average day follows the typical seasonality with traffic peaking during rush hours and returning to low volume during off-peak hours.

We map the traffic information to the message boards by once again using ArcGIS® 'Find Closest Facility' tool. In most instances, the closest traffic monitoring station is found on the same road as the message board, and when there is no traffic monitoring station on the road of the message board, we use the nearest traffic monitoring station to map traffic information to the message board. We also account for the direction bound of the road in mapping the stations to the message boards. This gives a local approximation to local traffic trends, and is generally a good approximation as both volume counters and DMSs tend to be placed on busier roads.

We further obtain hourly weather data from National Oceanic and Atmospheric Administration's National Centers for Environmental Information's Local Climatological Data, which provides daily and hourly summaries for approximately 11 Vermont locations, including Automated Surface Observing System and Automated Weather Observing System stations. They provide us daily data on snowfall and snow depth, as well as hourly data on dew point temperature, precipitation, wind condition, sky condition (cloudy, overcast, etc), weather condition (snowing, raining, drizzle, hail etc), and visibility.[7]

The exact definition of variables used is provided in Table 4.3. As a final note, our sample does not include road construction datawhich might be a relevant piece of information that influences crashes around the DMS.Correspondence with VTrans concluded that obtaining this data would be costly, and empirical design is ideally robust to this missing information due to the overall fixed location of signs during our period of observation, and overall methods described in Section 4.3.

4.3 Empirical Model

The question we are answering is as follows: is the change in the number of crashes happening because of nudges, or would the crashes be lower (higher) anyway at the time these nudges are displayed, perhaps because these are displayed during times when conditions are relatively safe (unsafe) for driving. To answer this question, we model the number of crashes as a count process following a Poisson distribution. We take this modeling approach for several reasons. First, individual probabilities of accidents on a given road segment at a given time are Bernoulli trials, so the probability of observing a certain number of crashes on a given location follows a Binomial distribution, for which as traffic volume gets large, converges to the Poisson distribution. Secondly, as shown in Hausman et al. (1984); Chamberlain (1987); Wooldridge (1999) location specific fixed effects fall out of the Poisson distribution. Generally though in our setting this is not a problem, since we have roughly 62 different message boards, but almost 27,000 observations per site. To get around an incidental parameters problem with respect to the time dimension, our fixed effects

[7] Link to data.

follow a Year-by-Month structure. Specifically, we adopt the following baseline specification

$$E[Y_{it}|\alpha_i, \lambda_t, \boldsymbol{X_i}, \boldsymbol{T_i}] = exp(\alpha_i + \lambda_t + X'_{it}\beta + T'_{it}\rho) \tag{4.1}$$

where Y_{it} equal the number of crashes on road segment i at time t, X_{it} is a vector of traffic and weather conditions, T_{it} is a vector of Behavior and Information Nudge treatment status on road segment i at time t,and α_i is a vector of unobserved but fixed confounders that influence near to sign road hazard. For instance, road segments with specific features (e.g. rough road, curved road, junctions, merging roads) are more probable to have one nudge or the other. λ_t represents time fixed effects to control for year and seasonal effects that impact road hazard.. Similar to the canonical Within-Transform of the linear additive fixed effects model, this transformation removes both individual effects, as well as other time-invariant factors from Equation 4.1 (Wooldridge (2010) Section 18.7.4), such that much of the remaining variation in accidents is coming from time varying covariates such as traffic, weather, and message board status.

Endogeneity arises when $E[\epsilon_{it}|\alpha_i, \lambda_t, \boldsymbol{X_i}, \boldsymbol{T_i}] \neq 0$. This may arise from three sources. First, via location choice of message boards. As the message boards are strategically installed on high risk roads, it is likely that the number of crashes are higher on these sites as compared to sites without message boards. Second, when the unobserved effects that may influence crashes on the site are varying time (i.e. α_i is actually time varying). In that case, ρ will capture the wrong effect when T_{it} is correlated withchanges in these unobserved factors. Finally, if the nudges are jointly determined with the crashes. This is likely to occur given that the messages are not displayed randomly and therefore there is a selection bias from message choice. Most importantly, Informational Nudges include displayed warnings of crashes ahead, which respond to accidents either contemporaneously or that just happened. Ignoring these potential sources of endogeneity may lead to inconsistent estimates in the above specification.

To deal with endogeneity from the location choice of message boards, we restrict the sample to only sites with message boards. This is possible because we know the precise location of each crash and the message board. The assumption here is that sites with message boards during the study period are quite similar to each other. To control for unobserved time varying factors we follow

Angrist and Pischke (2009) and introduce site-specific time trends to the list of controls in 4.1, i.e., we estimate

$$E[Y_{it}|\alpha_i, \lambda_t, \boldsymbol{X_i}, \boldsymbol{T_i}] = exp(\alpha_o + \alpha_i t + \lambda_t + X_{it}'\beta + T_{it}'\rho) \tag{4.2}$$

where α_i is the site-specific trend coefficient multiplying the time trend variable, t. This allows treatment and control sites to follow different trends. Our interest is in examining whether the estimated effect ρ changes by the inclusion of these site specific trends.

A remaining concern is that messages are displayed when even conditional on traffic and weather data there might already be additional average hazards on the road on the entire segment around the sign. Implicitly this accommodates variants of our mainline specification where α_i is changing, or, perhaps switching states, or accommodating unobserved heterogeneity in when different message types go as decided by VTrans. Under this setting the entire road segment, both immediately before and after the sign, might face excess, or lower, probabilities of individual drivers getting into accidents. To address this concern we estimate the following fixed effects specification:

$$E[Y_{itr}|\alpha_{ir}, \lambda_t, \boldsymbol{X_i}, \boldsymbol{T_i}] = \exp(\alpha_{ir} + \lambda_t + X_{it}\beta_r + T_{it}'\rho + 1\{r = 1\}T_{it}'\rho_1) \tag{4.3}$$

Here r indexes relative distance to a DMS. The accidents for both $r = -1$ (the mile before a DMS) and $r = 1$ (the mile after a DMS) are combined together, allowing for estimation of level shifts in mean road hazards when messages of a given time are displayed. α_{ir} is now a fixed effect for each tranche relative to a DMS, such that $\alpha_{i1} = \alpha_i$ as in Equation 4.1, but allowing the mile immediately before a sign to have it's own fixed effects, and allowing identification of whether or not there is excess road hazard locally around a DMS when signs of a particular type are up. For this model, the key parameters of interest is ρ_1. Similar to traditional Differences-in-Differences, ρ is the pre-treatment effect when a sign is active in the mile before a sign. This tries to capture level differences in near-to-sign hazards that exist on average when messages of a given type are displayed. Indexing in this fashion implies the causal interpretation that, ρ_1 is the impact of a particular message type on the mile wide bin after a sign while controlling for excess hazard in the full region.[8]

As with Equation 4.2, we can use this stacked structure to estimate a model with time varying location specific factors by creating an ID variable that indexes observations by DMS ID and time, and has a creates a secondary index based on relative distance to the DMS. This model has multiple desirable features. First, it removes all covariates that are invariant over the two distances, including our traffic and weather variables, which due to matching might have meaningful measurement error. Secondly, it allows for temporary or other time varying changes to near-to-sign road hazard, such as temporary construction work, or seasonal variation in mean road hazard that might not just be captured by weather variables that do not fit into the linear trend presented in Equation (4.2).

$$E[Y_{itr}|\alpha_{it}, \lambda_r, \boldsymbol{X_i}, \boldsymbol{T_i}] = \exp(\alpha_{it} + \lambda_r + T'_{it}\rho + 1\{r = 1\}T'_{it}\rho_1) \tag{4.4}$$

The above models is robust to initial placement of DMSs, time varying location specific effects, and unobserved risk factors on the entire road segment around the DMS, they do not account for concerns about endogenous response of messages, in particular informational nudges, to crashes happening after a DMS. The fixed effect estimation require a core assumption of strict exogeneity:

$$E[\epsilon_{is}|\alpha_i, \lambda_t, X_{it}, T_{it}] = 0 \ \forall \ s, t \tag{4.5}$$

This assumption forbids current value of ϵ_{is} to be correlated with past, present, and future values of T_{it}. Comparably, in the presence of reverse causality this assumption is necessarily violated, i.e., if crashes in current time period (Y_{it}) effect choice of nudge in the next time period (T_{it+1}), then ϵ_{it}

[8]This can be thought of as a pseudo Regression Discontinuity Design. Since we do not observe actual traffic data at the DMS, and traffic accidents are very rare, canonical Regression Discontinuity Design estimation strategies are unavailable to us. This approach enables two different ways of trying to proxy for the at-sign hazard rate. Allowing for different hazard rates on either side of the DMS through relative distance level fixed effects controls time invariant factors that might be correlated with the initial sign placement. The variable ρ reflects the mean hazard in the mile immediately before and after a DMS under each of the different signs.

is correlated with T_{it+1}. This violation leads to biased estimates using the fixed effect estimation. We relax 4.5 to instead accommodate the data generating process,

$$E[Y_{it}|\alpha_i, X_{i1}, ..., X_{it}, T_{i1}, ..., T_{it}, \lambda_1, ..., \lambda_t] = exp(\alpha_i + \lambda_t + X'_{it}\beta + T'_{it}\rho) \tag{4.6}$$

Under this setting future displayed messages can be correlated with past levels of realized crashes. To accommodate for this we follow Wooldridge (1997); Chamberlain (1992); Windmeijer (2000) to estimate fixed effect Poisson models that exhibits sequential exogeneity. We take quasi first-difference to eliminate the fixed effects by using the following transformation,[9]

$$\Delta Y_{it} = \frac{\mu_{i,t-1}}{\mu_{it}} Y_{it} - Y_{i,t-1}$$

where $\mu_{it} = exp(\lambda_t + X'_{it}\beta + T'_{it}\rho)$. Iterated expectations shows that

$$E[X_{it-s}\Delta Y_{it}|X_{i1}, ..., X_{it}, T_{i1}, ..., T_{it}, \lambda_1, ..., \lambda_t] = 0$$

for all $s \geq 2$ Windmeijer (2000),[10] allowing for estimation using generalized method of moments.[11] This model offers protection both unobserved time-invariant heterogeneity but also from reverse causality, but it does not offer protection against time varying heterogeneity of discussed in models 4.2 and 4.3. Similar methods are further discussed in Allison (2012); Colin and Pravin (2013).

4.4 Main Results

In this section we present our main results. First, we present the baseline estimates on the effect of nudges within one mile from the legibility of the message boards. We then put these estimates through various specifications (as outlined in Section 4.3) to address the endogeneity of nudges and crashes. Finally, we examine the effect of these nudges as the distance from the

[9]The difference appears as under 4.6 $Y_{it} = exp(\lambda_t + X'_{it}\beta + T'_{it}\rho)u_{it} = exp(\lambda_t + X'_{it}\beta + T'_{it}\rho)\phi_i\epsilon_{it}$. From this we get $\phi_i\epsilon_{it} = \frac{Y_{it}}{\exp(\lambda_t + X'_{it}\beta + T'_{it}\rho)} = \frac{Y_{it}}{\mu_{it}}$.

[10]A major concern is that due to the pooling information at the hour level, X_{it} dependent on ϵ_{it}, to get around this we use a twice lagged set of covariates as instruments.

[11]This moment condition is equivalent to $E[X_{t-s}\mu_{it-s}\Delta u_{it}] = 0 \; \forall \; s \geq 2$.

message board location decreases.Throughout we report Incident Rate Ratios. These provide a standardize measure of understanding the relative rates of accidents occurring given a 1 unit (1 hour) increase in a given message type, which is equivalent to reporting $\exp(\beta)$ for a given effect of interest. Comparably, the percentage increase in accidents can be calculated as $(\exp(\beta) - 1)x100$ is the percent change in crashes caused by a one hour increase in a given message type's exposure time. In turn, we report the incidence rate ratio, along with the mean number of accidents for the a given tranche side following a DMS. The benchmark values for the number of accidents are 0.0004 accidents happen on average in a given hour, and 0.0000387 crashes happen on average in the quarter mile after a sign.

To get an idea of how rare crashes of any type are, across our 35,000 accidents, we have roughly 67 message boards and 26,208 hourly observations per board. Across the entire state, the total number of accidents is about 1.35 per hour. Thus, even across the whole state of Vermont, accidents are rare, let alone in just the mile long tranche immediately following a DMS. As a result, even a 100% increase in the rate of accidents implies the mean number of accidents happening in the mile after a DMS rises to only 0.0008 accidents per hour, or about 7 accidents per a year. Similarly, a 100% increase in the mean number of accidents happening in the mile after a DMS rises to 0.0000774 accidents per hour, or about 0.678 accidents per year. The multiplicative effect of the incidence rate ratio makes the actual impacts on number of crashes dependent on the tranche size, and likely that otherwise "large" effects will exist.

Table 4.4 presents the set of estimates that captures the effect of nudges within one mile from the installation of message boards. In the baseline specification (column 1), the parameter estimate for behavioral nudge is statistically zero while information nudge is positive and significant. Controlling for site-specific time trends (column 2) do not change the parameter estimates much. In both cases though the implied increase in crashes caused by the presence of informational nudges is about 145%. After conditioning on road hazard correlated with the presence of signs in the mile before a DMS when a given message type goes up, implies either only a 30% increase in near-to-sign accidents, or a result that is statistically indistinguishable from zero.

Similarly, when looking at just the Post Mile results, Behavior Nudges have no impact on near-to-sign road accidents. After conditioning on pre-trends, Behavioral Nudges causally decrease near-to-sign traffic incidents by about a 40%. These results indicate that both Behavior and Information Nudge messages go up when there is already excess risk on the whole region around the DMS relative to time periods without a displayed message.

A remaining concern is that the mile-long bandwidths before and after a DMS capture too much road surface area not attributable to a particular sign. To test this hypothesis we estimate our preferred specifications using data on accidents that occurred in just the quarter mile before, and the quarter mile after, a DMS. Ideally this better captures near to sign determinants of sign content, as well as a stronger share of drivers who actively saw the sign's message. Throughout the measures of informational nudges increase by 390-580% increase in accidents, but as before, results that condition on road hazard immediately around a DMS removes those effects. Compared to above, across models the coefficients related to behavioral nudges remain statistically insignificant.

Both these models share fundamentally similar results. Across models, behavioral nudges have no impact on near-to-sign accidents. Poisson regressions on just the mile or quarter-mile tranche after a DMS, shows strong effects of informational nudges increase in the relative rate of accidents. By controlling for road hazard associated with signs of a given type being displayed, or controlling for endogeneity in displayed message, reduce these effects to zero.

4.5 Robustness Checks

In this section we provide two sets of robustness checks over our mainline specification in Section 4.4. .Each robustness check estimatesvariants of Equations 4.1, 4.2, 4.3, and 4.4. Throughout they focus on disaggregated estimates of DMS message content, breaking behavioral nudges into 5 categories- death toll, drinking and driving, seat belt, speeding, and texting reminders- and breaking informational nudges into eight different reminders- road, weather, traffic, work zone, road closure, crash info, other caution, or other message conditions. The first test re-estimates our mainline set of models with these disaggregated displayed message content. The second robustness check further tries to control for whether or not spatial spillovers might be biasing results. This concern comes from two points of view. The first is an implicit test for whether or not DMSs might be creating long-run changes in driving behavior. If spatial spillovers appear in the our results, the induced better, or worse, driving behavior might continue throughout the DMS network. Secondly, many DMSs are placed relatively close to each other, and results from Section 4.4 might be confounded by the presence of upstream messages. This check for spatial spillovers is provided through estimating two different models. The first includes whether or not there was an upstream behavioral or informational nudge within 5 miles, and the second conditions on signs that are at least 5 miles away from any upstream neighbor.

4.5.1 Heterogeneous Message Type Effects

As noted above, we estimate impacts of heterogeneous message content. Using detailed message content provided by VTrans, we split the messages into 13 different categories. The aim here is to understand where plausible sources of endogeneity are coming from, for example Crash Ahead should always be contemporaneously correlated with an accident occurring ahead at some interval. Secondly, drivers might be responding to different message types in a heterogeneous fashion. Seat belt reminders might not elicit a response for safer driving since individuals are already often buckled up, while death toll reminders might elicit momentary feelings of remorse and changes to safer driving behavior. The model that controls for plausible sequential exogeneity of regressors is omitted, since the likelihood is too flat with respect to our parameters of interest.

Effects for the mile wide tranche are discussed in Table 4.6. All of the disaggregated measures of behavioral message content are again statistically insignificant from zero, but many indicate a decrease in the number of near-to-sign accidents by about 30-50%. One concern is that the point estimates across the two class of models vary wildly, to the point where one much choose which class of specification they believe in. Comparably there are large heterogeneous effects across informational content provided by DMSs. As expected of our indexing approach, road condition and weather messages disappear from being significant due to these being reflected in excess road hazard both before and after the sign, and can only be displayed due to fixed nature of the signs in the sample. Moreover, Crash Ahead messages lead to huge increases in the probability of a crash, but are also clearly endogenous to the displaying of such messages- a reason why we estimated a model with sequential exogeneity in our mainline specifications.

Changing to the quarter mile bandwidth paints and entirely different story. Under this specification there are many signs that now have strong, statistically significant, negative impacts on near-to-sign accidents. Drinking, Speed, Work zone, and Other Caution Messages share sign, and often magnitudes, across specifications. These effects range from a 30-100% decrease in the number of near to sign crashes. Texting, Traffic, and Crash Info messages do not have crashes associated with them on this interval, so are dropped from the IDxTime fixed effects model.

4.5.2 Spillover from neighboring signs

A remaining concern is that upstream signs might have impacts on downstream driving behavior. Our final set of robustness checks explicitly models this in two ways, first we include an indicator for a behavioral or informational nudge from upstream signs, and secondly we subset our sample to include signs that have at least 5 miles between them and any upstream neighbor. This is important since

in the presence of non-zero treatment effects from upstream signs, drivers might be already driving differently than they would have in the presence of no prior treatment assignment.

To model this effect we condition on upstream neighbor sign status within miles of each subject sign. To do so, define variables

$$SpillBehavior_{it} = 1\{\text{Upstream sign within 5 miles of i has an Behavioral Nudge message up}\}$$

$$SpillInformation_{it} = 1\{\text{Upstream sign within 5 miles of i has an Informative Nudge message up}\}$$

In the case where there is no spatial dependence, the coefficients on these terms converges to zero and this becomes a regular Poisson fixed effects model presented in Equation 4.1, while if they're non-zero incorporates downstream sign information. The coefficients for sign effects are almost identical to Table 4.4, indicating that while in some models these spillover effects are meaningful, they are seemingly independent of drivers response to new message content even a few miles downstream. While for the Poisson fixed effects estimator on just the mile after has positive effect, after accounting for road hazard in the entire region, these upstream messages have no impact on downstream driving.

Alternatively, we create a sub sample of signs with no upstream neighbor within 5 miles. Under local effects this implies that plausible spillover effects of neighboring signs is zero, and might alternatively remove impacts of upstream message content on downstream signs if upstream response is also heterogeneous as shown in Section 4.5.1. The downside to this approach is that it further restricts the sample to rural or otherwise isolated areas, and away from urban and high traffic areas. Even after carrying out this sub setting, the story remains identical.

4.6 Conclusion

In this paper we study the impact of behavioral and informational nudges on near-to-sign traffic incidents. We generate a large geospatial panel data set using hour level information on weather, traffic, crashes, and the content of messages displayed on dynamic message signs. We estimate several variations of Poisson fixed effects models, including a baseline model on just the (quarter) mile after a DMS, allowing for cite specific trends, two difference-in-differences model with Message Board ID by Time or Message Board ID by Distance fixed effects, and a model with sequential exogeneity to identify the effect of the nudges.

Our results show that without conditioning on near-to-sign excess road hazard that is correlated with variant message signs going up, or the causal relationship between different message types and previous periods or contemporaneous accidents, can lead to spuriously believe that informational nudges have a causal relationship on near-to-sign accidents.After accounting for these relationships, there exists minor evidence that behavioral nudges might improve driving behavior, while informational nudges have no impact on near-to-sign accidents. After disaggregated results there exists large heterogeneity in driver response to individual signs. Many of our behavioral nudges have economically meaningful, but statistically insignificant, negative impacts on near-to-sign traffic incidents, while reductions in accidents caused by Other Caution Messages are meaningfully swamped out by the existence of endogeneity between Crash Ahead and contemporaneous traffic behavior. Models that try to explicitly control for spillover effects further cast doubt on claims about short lived duration of positive or negative effects of DMSs on near-to-sign traffic incidents.

From the policy perspective, our results indicate that behavioral nudges are not an effective way to reduce the number of traffic incidents. Drivers seem to ignore these signs messages, or not change their driving behavior in response to them. On the other hand, informational nudges, being direct useful information, may causally reduce the number of crashes, but continued issues with Crash Ahead and messages muddle this response in many models. Therefore, we suggest the use of informational nudges while avoiding behavioral nudges. More research is needed to examine which exact kind of behavioral nudges are effective and which may lead to negative reaction from the drivers.

The effects that we study in the paper are local to the sign location. We acknowledge that these nudges might be causing long run changesin overall driving behavior which are not captured in the immediate vicinity of the message board location. It is possible that these nudges have time varying effects, for example, people may react to these nudges when they are first displayed but over time drivers become immune to them and start ignoring them, or worse start to get annoyed by them, thereby petering out the effect. How rare crashes are in our matched message sign and near to sign crashes make such models hard to estimate. Instead we restrict ourselves to models where drivers are effectively naive, and do not experience long run changes in driving behavior caused by displayed message treatment.

4.7 References

Abouk, R. and Adams, S. (2013). Texting bans and fatal accidents on roadways: Do they work? Or do drivers just react to announcements of bans? *American Economic Journal: Applied Economics*, 5(2):179–199.

Allcott, H. and Mullainathan, S. (2010). Behavior and energy policy. *Science*, 327(5970):1204–1205.

Allison, P. (2012). *Fixed Effects Regression Models*. SAGE Publications.

Angrist, J. D. and Pischke, J.-S. (2009). Parallel Worlds: Fixed Effects, Differences-in-Differences, and Panel Data. In *Mostly Harmless Econometrics*, pages 221–247. Princeton University Press.

Benson, B. G. (1997). Motorist attitudes about content of variable-message signs. *Transportation Research Record*, 1550(1550):48–57.

Bonsall, P. (1992). The influence of route guidance advice on route choice in urban networks. *Transportation*, 19(1):1–23.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334.

Chamberlain, G. (1992). Comment: Sequential moment restrictions in panel data. *Journal of Business and Economic Statistics*, 10(1):20–26.

Choudhary, V., Shunko, M., Netessine, S., and Koo, S. (2019). Nudging Drivers to Safety: Evidence from a Field Experiment.

Clark, P. J. and Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, 35(4):445–453.

Colin, C. A. and Pravin, T. (2013). *Regression analysis of count data, Second edition*. Cambridge University Press.

De Borger, B. and Proost, S. (2013). Traffic externalities in cities: The economics of speed bumps, low emission zones and city bypasses. *Journal of Urban Economics*, 76(1):53–70.

DeAngelo, G. and Hansen, B. (2014). Life and death in the fast lane: Police enforcement and roadway safety. *American Economic Journal: Economic Policy*, 6(2):231–257.

Dholakia, U. M. (2016). Why Nudging Your Customers Can Backfire. *Harvard Business Review*.

Egan, M. (2017). *Nudge: Improving decisions about health, wealth and happiness*.

Erke, A., Sagberg, F., and Hagman, R. (2007). Effects of route guidance variable message signs (VMS) on driver behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(6):447–457.

Fallah Zavareh, M., Mamdoohi, A. R., and Nordfjærn, T. (2017). The effects of indicating rear-end collision risk via variable message signs on traffic behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, 46:524–536.

Hall, J. D. and Madsen, J. (2020). Can behavioral interventions be too salient? Evidence from traffic safety messages.

Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *American Economic Review*, 105(4):1581–1617.

Hassan, H. M., Abdel-Aty, M. A., Choi, K., and Algadhi, S. A. (2012). Driver behavior and preferences for changeable message signs and variable speed limits in reduced visibility conditions. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 16(3):132–146.

Hausman, J., Hall, B. H., and Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica*, 52(4):909.

Johnson, E. J. and Goldstein, D. (2003). Do defaults save lives? *Science*, 5649(302):1338–1339.

Makowsky, M. D. and Stratmann, T. (2011). More tickets, fewer accidents: How cash-strapped towns make for safer roads. *Journal of Law and Economics*, 54(4):863–888.

Mounce, J. M., Ullman, G. L., Pesti, G., Pezoldt, V., Institute, T. T., of Transportation, T. D., and Administration, F. H. (2007). Guidelines for the Evaluation of Dynamic Message Sign Performance. Technical Report 2.

Norouzi, A., Haghani, A., Hamedi, M., and Ghoseiri, K. (2013). Impact of Dynamic Message Signs on occurrence of road accidents.

Peng, Z. R., Guequierre, N., and Blakeman, J. C. (2004). Motorist response to arterial variable message signs. *Transportation Research Record*, 1899(1899):55–63.

Song, M., Wang, J.-H., Cheung, S., and Keceli, M. (2016). Assessing and Mitigating the Impacts of Dynamic Message Signs on Highway Traffic. *International Journal for Traffic and Transport Engineering*, 6(1):1–12.

Tarry, S. and Graham, A. (1995). The role of evaluation in ATT development. IV: Evaluation of ATT systems. *Traffic Engineering and Control*, 36(12):688–693.

Tay, R. and De Barros, A. G. (2008). Public perceptions of the use of dynamic message signs. *Journal of Advanced Transportation*, 42(1):95–110.

Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(1):S164–S187.

Vanasse Hangen Brustlin, I. (2007). Dynamic Message Sign Study. Technical report, Vermont Agency of Transportation.

Windmeijer, F. (2000). Moment conditions for fixed effects count data models with endogenous regressors. *Economics Letters*, 68(1):21–24.

Wooldridge, J. M. (1997). Multiplicative Panel Data Models Without the Strict Exogeneity Assumption. *Econometric Theory*, 13(5):667–678.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Xu, M. and Xu, Y. (2020). Fraccidents: The impact of fracking on road traffic deaths. *Journal of Environmental Economics and Management*, 101:102303.

Xu, W., Zhao, X., Chen, Y., Bian, Y., and Li, H. (2018). Research on the Relationship between Dynamic Message Sign Control Strategies and Driving Safety in Freeway Work Zones. *Journal of Advanced Transportation*, 2018:1–19.

Table 4.3 List of Variables

Variable	Source	Description
Dependent Variable (Y_{it}^{r})		
Crashes	VTrans	Number of crashes r miles from site i at time t
Time Varying Covariates (X_{it})		
Traffic	VTrans	Hourly traffic volume from 24 traffic counters in Vermont
Dew Point Temperature	NOAA	Hourly dew point temperature in Fahrenheit.
Precipitation	NOAA	Hourly amount of precipitation in inches to hundredths.
Humidity	NOAA	Hourly relative humidity given to the nearest whole percentage
Visibility	NOAA	Hourly horizontal distance an object can be seen and identified given in whole miles.
Sky Conditions	NOAA	Hourly report of cloud layer with options clear, partly cloudy, and mostly cloudy.
Wind Speed	NOAA	Hourly speed of the wind at the time of observation given in miles per hour (mph).
Snow Depth	NOAA	Daily amount snow depth in inches.
Snowfall	NOAA	Daily amount of snowfall in inches
Message Data (T_{it})		
Behavioral Nudge	VTrans	Dummy variable which takes a value 1 if either of the death, seat belt, texting, drinking, speeding, or other caution was active for at-least some time during the hour t on site i.
Informational Nudge	VTrans	Dummy variable which takes a value 1 if either of the road condition, weather condition, traffic, work zone, road closure, crash info, or other message was active for at-least some time during the hour t on site i.

Table 4.4 Effect of nudges on crashes within 1 mile from message board

	Baseline (1)	Site Trends (2)	IDxDist (3)	IDxTime (4)	Sequential Exog (5)
BehaviorNudge	0.826	0.910	0.585**	0.611*	0.277
	(0.191)	(0.218)	(0.106)	(0.133)	(2.839)
InformationNudge	2.316***	2.451***	1.509	1.349*	1.906
	(0.498)	(0.580)	(0.385)	(0.203)	(74.48)
Weather Controls	Yes	Yes	Yes	No	Yes
Traffic Controls	Yes	Yes	Yes	No	Yes
Site-Specific Trends	No	Yes	No	No	No
Observations	1473017	1473017	2919730	2386	1472891

*Notes: The table presents the estimates for the effect of nudges on crashes within one mile from the legibility of message board. Column 1 presents the result of baseline fixed-effect Poisson regression. Column 2 control for the site-specific time trends in the baseline specification. Column 3 presents an event study style estimator indexed to the mile before a DMS and an indexing variable of Message Board ID and relative distance. Column 4 presents an event study style estimator indexed to the mile before a DMS with an indexing variable of Message Board ID and Date. Column 5 presents results of a Poisson fixed effects regression under sequential exogeneity. Clustered robust standard errors in parenthesis. * for p ¡ 0.05, ** for p ¡ 0.01, and *** for p¡ 0.001*

Table 4.5 Effect of nudges on crashes within 1/4 mile from message board

	Baseline (1)	Site Trends (2)	IDxDist (3)	IDxTime (4)
BehaviorNudge	0.791	2.058	0.461	0.447
	(0.634)	(1.617)	(0.361)	(0.370)
InformationNudge	4.941***	6.818***	0.767	1.640
	(1.796)	(2.417)	(0.388)	(0.807)
Weather Controls	Yes	Yes	No	Yes
Traffic Controls	Yes	Yes	No	Yes
Site-Specific Trends	No	Yes	No	No
Observations	447166	447166	516	1525627

*Notes: The table presents the estimates for the effect of nudges on crashes within a quarter mile from the legibility of message board. Column 1 presents the result of baseline fixed-effect Poisson regression. Column 2 control for the site-specific time trends in the baseline specification. Column 3 presents an event study style estimator indexed to the quarter mile before a DMS and an indexing variable of Message Board ID and relative distance. Column 4 presents an event study style estimator indexed to the quarter mile before a DMS with an indexing variable of Message Board ID and Date. Clustered robust standard errors in parenthesis. * for p ¡ 0.05, ** for p ¡ 0.01, and *** for p¡ 0.001*

Table 4.6 Effect of Message Types on crashes within 1 mile from DMS

	Baseline (1)	Site Trends (2)	IDxDist (3)	IDxTime (4)
Death Toll	1.253	1.325	0.718	0.693
	(0.383)	(0.419)	(0.327)	(0.316)
Seatbelt Reminder	0.762	0.775	2.259	2.440
	(0.423)	(0.435)	(1.859)	(1.705)
Texting Reminder	0.434	0.436	1.995	2.181
	(0.292)	(0.298)	(1.903)	(1.903)
Anti Drinking Reminder	1.021	1.118	0.440	0.452
	(0.354)	(0.393)	(0.194)	(0.243)
Speeding Reminder	0.918	0.996	0.574	0.615
	(0.292)	(0.326)	(0.211)	(0.215)
Road Condition	0.444*	0.430*	1.140	1.006
	(0.145)	(0.142)	(0.705)	(0.518)
Weather Condition	3.211***	3.317***	1.302	1.281
	(0.907)	(1.004)	(0.427)	(0.270)
Traffic Condition	1.513	1.461	3.949	5.907
	(1.707)	(1.599)	(5.977)	(8.302)
Work Zone	0.982	0.972	0.429	0.356
	(0.575)	(0.580)	(0.333)	(0.282)
Road Closure	1.131	0.996	1.714	1.475
	(0.364)	(0.454)	(0.603)	(0.599)
Crash Info	24.47***	24.70***	1.2950e+157***	7.81365e+18***
	(19.65)	(20.24)	(5.4940e+158)	(2.58189e+19)
Other Caution	0.177	0.209	0.0571**	0.0493**
	(0.188)	(0.222)	(0.0620)	(0.0542)
Other Message	1.357	1.375	1.503	1.176
	(0.736)	(0.845)	(0.888)	(0.317)
Weather Controls	Yes	Yes	Yes	No
Traffic Controls	Yes	Yes	Yes	No
Site-Specific Trends	No	Yes	No	No
Observations	1473017	1473017	2919730	2386

*Notes: Column headings mirror Equations (1)-(4). * for p ¡ 0.05, ** for p ¡ 0.01, and *** for p¡ 0.001*

Table 4.7 Effect of Message Types on crashes within 1/4 mile from DMS

	Baseline (1)	Site Trends (2)	IDxDist (3)	IDxTime (4)
Death Toll	4.741	5.617*	2.109	1.888
	(4.018)	(4.873)	(1.862)	(1.989)
Seatbelt Reminder	1.489	2.185	1.433	0.913
	(1.601)	(2.517)	(1.870)	(0.835)
Texting Reminder	6.43e-113***	6.31e-100***	1.51e-15	1
	(3.56e-111)	(2.08e-98)	(1.29e-13)	(.)
Anti Drinking Reminder	4.71e-54***	5.23e-50***	7.66e-62***	0.000000114***
	(3.89e-53)	(5.45e-49)	(7.07e-61)	(8.56e-08)
Speeding Reminder	8.47e-195***	3.14e-193***	7.63e-233***	1.90e-09***
	(3.79e-193)	(1.37e-191)	(3.22e-231)	(3.52e-09)
Road Condition	0.133	0.158	0.117	0.107*
	(0.166)	(0.191)	(0.170)	(0.122)
Weather Condition	3.280	4.662*	0.766	0.459
	(2.068)	(3.110)	(0.589)	(0.348)
Traffic Condition	6.54e-195***	4.93e-166***	4.97536e+38	1
	(4.11e-193)	(2.69e-164)	(3.55679e+40)	(.)
Work Zone	7.21e-36***	7.20e-24***	5.59e-40***	0.000000320***
	(6.95e-35)	(3.17e-23)	(5.04e-39)	(0.000000326)
Road Closure	9.607***	6.033***	10.09***	1.978
	(2.799)	(1.514)	(3.603)	(2.265)
Crash Info	2.53e-77***	8.31e-90***	4.7396e+128***	1
	(3.78e-76)	(1.43e-88)	(3.1588e+130)	(.)
Other Caution	1.91e-186***	3.68e-175***	4.59e-220***	8.62e-17***
	(6.11e-185)	(1.29e-173)	(1.34e-218)	(2.60e-16)
Other Message	13.82***	22.03***	18.53***	0.900
	(9.436)	(18.74)	(13.20)	(0.732)
Weather Controls	Yes	Yes	Yes	No
Traffic Controls	Yes	Yes	Yes	No
Site-Specific Trends	No	Yes	No	No
Observations	447166	447166	1525627	516

*Notes: Column headings mirror Equations (1)-(4). * for p ¡ 0.05, ** for p ¡ 0.01, and *** for p¡ 0.001*

Table 4.8 Effects on crashes within 1 mile from DMS with spillovers

	Baseline (1)	Site Trends (2)	IDxDist (3)	IDxTime (4)
Death Toll	1.215	1.285	1.513**	0.667
	(0.366)	(0.392)	(0.198)	(0.307)
Seatbelt Reminder	0.771	0.785	0.700	2.484
	(0.426)	(0.439)	(0.307)	(1.730)
Texting Reminder	0.462	0.463	2.253	2.182
	(0.314)	(0.318)	(1.866)	(1.936)
Anti Drinking Reminder	0.965	1.034	1.992	0.428
	(0.348)	(0.378)	(1.952)	(0.235)
Speeding Reminder	1.012	1.088	0.424	0.605
	(0.362)	(0.392)	(0.199)	(0.222)
Road Condition	0.451**	0.433**	0.592	1.048
	(0.137)	(0.133)	(0.250)	(0.550)
Weather Condition	2.433**	2.526**	1.129	1.210
	(0.796)	(0.856)	(0.701)	(0.267)
Traffic Condition	1.392	1.361	1.212	5.605
	(1.551)	(1.470)	(0.447)	(7.755)
Work Zone	0.976	0.946	3.975	0.361
	(0.580)	(0.578)	(6.045)	(0.287)
Road Closure	1.054	0.971	0.429	1.424
	(0.347)	(0.416)	(0.339)	(0.579)
Crash Info	22.03***	22.10***	1.684	7.10121e+18***
	(18.68)	(19.20)	(0.610)	(2.34758e+19)
Other Caution	0.131	0.151	1.4289e+156***	0.0449**
	(0.143)	(0.165)	(5.9744e+157)	(0.0495)
Other Message	1.132	1.171	0.0522**	1.112
	(0.564)	(0.658)	(0.0577)	(0.305)
SpillBehavior	1.002	0.977	1.408	1.066
	(0.164)	(0.158)	(0.800)	(0.214)
SpillInformation	1.706**	1.705**	1.002	1.122
	(0.340)	(0.341)	(0.199)	(0.157)
Observations	1473017	1473017	2919730	2386

Table 4.9 Effects on crashes within 1 mile from DMS with no upstream neighbor

	Baseline (1)	Site Trends (2)	IDxDist (3)	IDxTime (4)
Death Toll	1.726	1.845	0.930	0.958
	(0.708)	(0.773)	(0.621)	(0.576)
Seatbelt Reminder	0.214	0.231	0.396	0.399
	(0.208)	(0.225)	(0.464)	(0.285)
Texting Reminder	0.756	0.762	1.567	1.604
	(0.498)	(0.516)	(1.505)	(1.399)
Anti Drinking Reminder	1.530	1.661	0.464	0.444
	(0.389)	(0.457)	(0.246)	(0.305)
Speeding Reminder	0.722	0.801	0.421	0.457
	(0.254)	(0.282)	(0.187)	(0.232)
Road Condition	0.517*	0.500*	2.067	1.827
	(0.165)	(0.157)	(1.432)	(1.114)
Weather Condition	3.356***	3.594***	1.325	1.354
	(0.982)	(1.074)	(0.427)	(0.376)
Traffic Condition	2.31e-80***	7.85e-68***	4.63535e+09	1
	(1.15e-78)	(1.44e-66)	(3.91606e+11)	(.)
Work Zone	0.762	0.740	0.268	0.210
	(0.577)	(0.584)	(0.273)	(0.225)
Road Closure	0.759	0.892	2.301	2.231
	(0.419)	(0.516)	(2.132)	(1.490)
Crash Info	1.36e-207***	3.95e-171***	4.64e-17	1
	(8.97e-206)	(2.54e-169)	(4.16e-15)	(.)
Other Caution	0.195	0.236	0.226	0.169
	(0.212)	(0.256)	(0.251)	(0.219)
Other Message	1.043	1.201	0.777	1.044
	(1.171)	(1.694)	(0.932)	(0.360)
Weather Controls	Yes	Yes	Yes	No
Traffic Controls	Yes	Yes	Yes	No
Site-Specific Trends	No	Yes	No	No
Observations	868025	868025	1709746	1228

CHAPTER 5. GENERAL CONCLUSION

In this dissertation I have explored how government policies impact regional dynamics, whether they be establishment dynamics or accidents, and found that in both cases the proposed policy does not have the desired effects. In both cases these are meaningful, since large expenditures or continued costs are levied on taxpayers in order to pay for these policies, or, as unexplored, they may have bad long-term impacts on community health. In the third paper, I developed a novel joint hypothesis tests over fixed effects, showed it's asymptotic properties, and developed a feasible implementation under a set of reasonable joint tests of interests and assumptions on the error process.

The first paper showed that despite the massive outlay of funds from the government to local and regional banks, little of these funds got passed through to help new entry or improve employment, nor as a part of a loss aversion strat to avoid excess firm exit or employment contraction. Instead, consistent with previous literature, funds appeared to either not have been lent out, or lent out in largely non-productive capacities, where large firms simply changed their own leverage. In turn, evaluation of the CPP appears to be that outside of choosing banks that were likely to pay back the funds so that it didn't cost the tax payers money, the program was mostly a failure. Most of the banks that received funds from the CPP were already predisposed to do well, and the lack of an improvement in firm dynamics might have long term consequences on increases in local-regional market power and other long term undesirable effects.

The second paper developed a new joint hypothesis test over fixed effects in short panels with serial correlation. I showed under two assumptions, either the errors following a stationary $AR(p)$ process, or wishing to test a shared grouping structure, with same covariate values for all individuals in a group, a feasible test exists that converges to a standard normal distribution. I show the test has good small sample size using a sparse, but useful, set of monte carlo experiments.

The third paper creates a new data set linking over 15,000 crashes to each of the 67 Dynamic Message Boards in the state of Vermont over a three year time spam. We then categorize what messages were displayed in a given hour between behavior and information nudges, and showed that after accounting for multiple sources of endogeneity, namely either plausible simultaneous roll out of sign message to near to sign crashes, or messages being displayed when total hazard is raised around the sign, such as with weather, that most of the effects of Dynamic Message Boards disappears.