Macalester Journal of Economics

Volume 29

Spring 2019

Table of Contents

Published annually by the Macalester College Department of Economics and Omicron Delta Epsilon

Macalester Journal of Economics

Volume 29 *Spring 2019*

Economics Faculty and Staff

Paul Aslanian *Emeritus Professor*	*Samantha Çakir* *Visiting Assistant Professor*	*Amy Damon* *Associate Professor*	*Liang Ding* *Associate Professor*
Karl Egge *Emeritus Professor*	*Jeffery Evans* *Adjunct Professor*	*J. Peter Ferderer* *Edward J. Noble Professor*	*Felix Friedt* *Assistant Professor*
Gary Krueger *Cargill Professor of International Economics*	*Joyce Minor* *Karl Egge Professor*	*Karine Moe* *F. R. Bigelow Professor*	*Emily Richards* *Department Coordinator*
Mario Solis-Garcia *Associate Professor*	*Lucas Threinen* *Visiting Assistant Professor*	*Vasant Sukhatme* *Emeritus Professor*	*Sarah West* *G. Theodore Mitau Professor and Department Chair*

A Note from the Editors

During the spring of 2019, we had the pleasure of reading papers submitted by Macalester economics students for this edition of the Macalester Journal of Economics. The extraordinary level of scholarship present in the department is particularly evident in the five papers we have selected for publication. Moreover, these papers are a testament to the investigative and analytical skills of students of the entire Macalester student body and represent the clear and concise writing skills that all departments strive to impart in their students.

We would like to thank all of the faculty for their indispensable roles in the development of these research projects. We would also like to thank Emily Richards and Professor Mario Solis-Garcia for their support in the creation of this journal. We hope you find these papers indicative of the breadth and depth of research interests present in the Macalester Department of Economics.

Matthew Steele '19
Harrison Mitchell '19
Margot Robison '19
Duy (Jensen) Vu '19
Noah Zwiefel '19

Foreword

The Macalester College chapter of Omicron Delta Epsilon, the international honors society in economics, proudly edits the *Macalester Journal of Economics* every year. This year's editors – Matthew Steele '19 (Newton, MA), Harrison Mitchell '19 (Knoxville, TN), Margot Robinson '19 (Kansas City, MO), Duy (Jensen) Vu'19 (Hanoi, Vietnam), and Noah Zwiefel '19 (Rochester, MN) – have carefully selected five papers on a variety of important topics. These papers are a sample of the research that our students produced in the last academic year.

What drives rental housing prices in a city? That's the question that Joe Fish '19 (Oakland, CA) tackles in his paper. In particular, the city is San Francisco, where the tech industry is booming, rents are always rising, and gentrification is a constant. Using a structural vector autoregression framework, Joe finds that economic intuition is safe and sound: the booming tech sector and a lack of housing supply drive rental prices up, while high rental prices generates a high eviction ratio, which in turn helps gentrification move along. At the end of the day, this is a story about demand and supply, as a strong demand for a shrinking supply of housing is the main driver behind the soaring rentals in the Bay area.

Conventional wisdom suggests that increasing the minimum wage generates unemployment, yet conventional wisdom and the real world rarely agree on these issues. Matthew Yang '19 (Bethesda, MD) offers a novel take on this topic: he looks at the employment effects of the recent minimum wage increases in Minnesota using synthetic control methods, keeping a standard difference-in-differences approach for comparison purposes. Matthew finds that limited-service restaurant employment – the low-skilled employment that pundits claim will suffer the most as a consequence of a higher minimum wage – didn’t change as a result of the higher minimum wages in Minnesota. Let’s add another point to the real world’s score.

On a different topic, it’s hard to argue that wealthy people are often sizable donors when charity is involved. But are wealthy people naturally altruistic, or are they more willing to give to charity as long as their donations are observed by everyone? Alex Ramiller ’18 (Portland, OR) and Maddy Hillson ’18 (Minneapolis, MN) run an experiment to answer this question. They find that wealthy participants tend to donate lower amounts when their gifts remain anonymous but become more generous when they are told that their donations will be acknowledged by the public. This suggests that reputational effects may matter when designing a policy framework around charitable giving.

The Affordable Care Act (a.k.a. Obamacare) remains one of the most loved (or hated) government policies in the United States. One of its provisions is an expansion of coverage for low-income families via the Medicaid program. Arguably, this policy will achieve better health outcomes, yet it is surprising that several states did not sign up for this expansion. Elliot Cassutt ’18 (Minneapolis, MN), Khadidja Ngom ’19 (Kaolack, Senegal), and Duy (Jensen) Vu ’19 (Hanoi, Vietnam) explore the Medicaid expansion through the lens of their Economics of Public Policy course: they analyze the current state of the policy and offer recommendations for the future.

Finally, Alice Zhu ’19 (Nanjing, China) looks at the effects of codeshare agreements on non-stop flights in the continental U.S. Arguably, having a codeshare partner increases the number of ways a person can travel from point A to point B – which also happens to work like an increase in the supply of flights. Economic theory suggests that this should reduce the market price, and Alice’s paper shows that this is the case. Her results point to an average fall in prices of 8.3%, which besides being nontrivial, is robust to alternative specifications.

On behalf of my colleagues in the Economics Department, I am delighted to present the research of these talented students. I am confident that you will find it enlightening and be impressed by the value of a liberal arts education.

Mario Solis-Garcia
Associate Professor of Economics

Empirical Evidence of a Changing San Francisco

Joe Fish

Advanced Econometrics

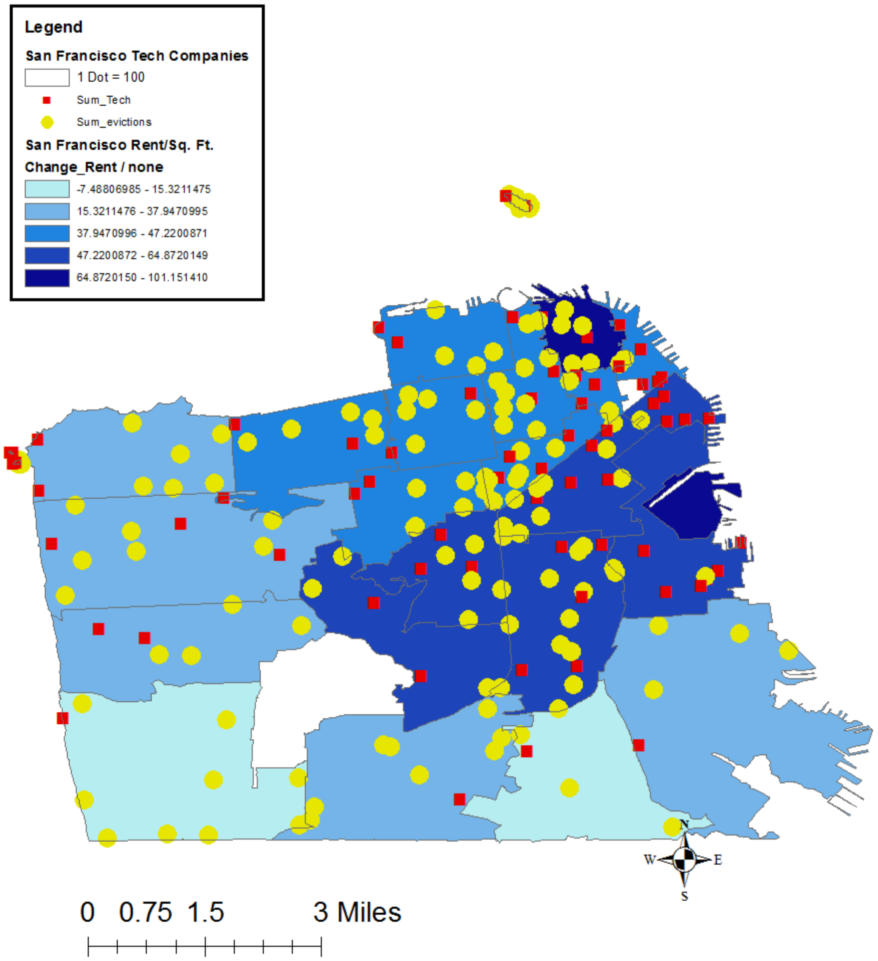
Abstract

While there has been an uptick on literature regarding rent control and gentrification, studies identifying the relationship between citywide policies, demographic changes, and changes in the labor force have been much sparser. Using a Structural Vector Autoregression (SVAR) this paper allows me to obtain temporal relationships between evictions, rent prices, the housing supply, and the number of tech companies in San Francisco. I find evidence that landlords are working to circumvent rent control through evictions and that the shrinkage of the housing supply per 1000 people has been the main driver of San Francisco rent.

I. Introduction

From 2010 to 2016, San Francisco's rent prices have increased from a monthly average of $2,693 to $3,855 (trulia.org, 2018, monthly evictions have increased by 64% (datasf.org, 2018), the number of tech locations has ballooned by over 7,000 (datasf.org, 2018) and the housing supply per 1000 people has shrunk from 452.3 to 433.0 (datasf.org, 2018). It's clear that these variables are related. As Figure 1 shows, tech growth has happened in the same neighborhoods as large rent increases, which occur in the same neighborhoods as large increases in evictions.

Figure 1.



Within San Francisco, local residents blame the influx of tech workers for rising rent prices, while tech workers decry wealthy residents for blocking any and all housing development, all while poor San Franciscans become increasingly rent burdened or are forced to move. However, it is difficult to establish a causal relationship between these variables. A dramatic increase in rent price is clearly a driver of evictions. However, there are clearly underlying drivers of this increase in rent price, which begs the question, "what is the driver of these increases in rent?" In this paper, I attempt to establish temporal relationships between displacement, increases in wealthy tech workers, changes in the housing supply, and rent prices.

Utilizing a structural VAR comprised of evictions, number of monthly registered technology businesses, housing units per 1000 people, and median market rent price, I find that the affordability crisis can largely be attributed to a housing supply that has failed to keep pace with a growing population. Furthermore, I find evidence that current attempts to address this affordability crisis, such as rent control, are being circumvented by landlords through evictions.

II. Literature Review

The vast majority of literature surrounding cities, affordability, and displacement concerns itself with the links between influxes of higher income households and the displacement or potential displacement of current residents. Within the literature, however, there is tremendous debate about what constitutes and contributes to displacement, affordability, and more broadly, what it means for a neighborhood or city to change.

Explicitly, displacement has been defined by a report by the U.S. Department of Housing and Development (HUD) (Grier and Grier, 1980) as

> "...when any household is forced to move from its residence by conditions which affect the dwelling or immediate surroundings, and which are beyond the household's reasonable ability to control or prevent…"

However, distinguishing "voluntary" from "involuntary" displacement is immensely difficult (Newman and Owen, 1982). Households who were forced to leave their rented home as a result of a landlord's actions but not evicted are observationally equivalent to households who decided to leave because they wanted to. In an effort to operationalize the term "displacement," researchers have proposed using evictions or significant increases in rent price as proxies (Atkinson, 2000). This research serves as the basis for my choice of evictions as a proxy for displacement.

Once a researcher chooses a definition of displacement and settles on a suitable proxy, they must determine what economic forces are driving displacement. Generally, this literature can be grouped into two categories: demand and supply side factors. Beginning with the demand side, Marcuse (1985) places the blame for displacement on higher income workers entering an area, creating additional demand for luxury housing. However, Marcuse makes no attempt to establish any form of causality and is primarily driven by anecdotal evidence. A more comprehensive panel study of household displacement in Philadelphia by Ding et al. (2016) found that non-mortgage holding residents in gentrifying neighborhoods were not more likely to move, but those that did move had higher risks of downward mobility. In a similar household-level panel dataset, Freeman and Braconi (2005) found that a neighborhood could decrease poverty by 18% over 10 years without any increase in displacement, indicating that a decrease in poverty is not necessarily a zero-sum game for residents.

The significant methodological flaw in these household level panel studies, however, is their lack of controls for housing supply. Gyourko et al. (2005) finds that in the presence of an increase of productivity, cities with more inelastic housing supplies absorb these shocks via changes in rent and home prices. The elasticity of housing could impact the occurrence of displacement, which raises serious questions about the efficacy of previous literature.

From the supply point of view, displacement is caused by cities creating market baskets of goods that are designed to attract a different type of resident (Metcalf, 2018). Public parks and transit, in particular, have both been found to be predictors of a gentrification[1] and eventual displacement (Citylab, 2015). This market basket approach can also be seen in the regulations and policies that cities put in place, such as, rent control and zoning regulations.

Rent control can be viewed as a city's decision to prioritize current residents over future ones. Rent control operates in the background of San Francisco housing economics by changing how the market is allowed to function. Researchers have generally found evidence of detrimental effects of rent control on a city. In a quasi-experimental variation in assignment of rent control in San Francisco and Cambridge, MA, Diamond et al. (2018) and Sims (2011) find that rent control reduced rental housing supplies by 15% and that, in San Francisco, rent control increased displacement. However, these economists have traditionally operated under the assumption of a well-functioning housing market. Arnott (1995) finds that "soft" rent control systems[2] operating in imperfectly competitive markets could produce outcomes that were more efficient than the

---

[1] Here, gentrification is defined as a significant increase, as defined by the researcher, of high-income residents.

[2] This refers to only allowing a certain percentage increase in rent each year.

market. These papers, however, do not put displacement and changes in demand for housing into a larger context, and are relatively narrow in their policy applications.

Similarly, in a review of housing literature, Gyourko and Molloy (2014) find that housing supply is constrained by the severity of local housing regulations. The effects of local conditions on the supply and price of housing are significant. In particular, Gyourko and Glaeser (2017) find that cities with relatively unconstrained housing development, such as Atlanta, produce housing at a price equal to minimal profitable production cost. In constrained markets the presence of barriers to entry such as regulations leads to prices well above the cost to make the home. Gyourko and Glaeser (2017) find that houses in San Francisco had a price-to-cost ratio of 2.84, indicating significant upcharges in the local real estate market. Finally, Gyourko et al. (2005) find that in the presence of an increase of productivity, cities with a more inelastic housing supply absorb these shocks through changes in rent and home prices. These studies, however, do not examine the role of housing supply restrictions on displacement.

There are thus two gaps in the current literature. First, the broadly focused displacement research neglects the effects of housing supply on levels of displacement and rent price. Next, the literature regarding rent control and housing markets narrowly focuses on one aspect of a much larger problem, which is affordability, In this paper, I aim to bridge this gap by presenting a model that considers the effects of housing supply and, implicitly, rent control, while maintaining a broad enough focus as to inform policy proposals.

III. San Francisco Context

San Francisco has a number of unique, city specific factors that affect how rent price, housing supply, tech growth, and evictions interact with each other. Beginning with rent price, San

Francisco has a system of rent control that encompasses 60% of all rental units (San Francisco Planning Commission, 2015). Rent control in San Francisco is “soft,” meaning that landlords of rent controlled units are allowed to raise annual rates by a certain percentage set by the city. Units are allowed to be set back to market rates in the event of a vacancy (i.e., tenant leaves, is evicted, or the property is substantially remodeled).

Secondly, San Francisco has a large number of tech workers relative to other cities (USA News, 2018). In 2012 San Francisco passed the “Twitter Tax,” which was a tax break specifically designed to encourage tech companies to stay, expand, or relocate to San Francisco, which attracted even more high-income tech workers to the city. This was accomplished by changing from a payroll tax to a gross-receipts one (this tax can be roughly thought of as a revenue tax) because tech companies have significantly high payrolls relative to their revenues. While it is unclear how effective this tax has been, it is certainly evidence towards San Francisco’s courtship of technology companies, and drive to attract tech workers.

Finally, San Francisco has some of the strictest zoning requirements in the nation and some of the lowest rates of development in the nation. According to Gyourko and Glaeser (2017), San Francisco has a price to minimal cost ratio of 2.8, which again is an indicator that there are significant barriers (both topologically and in the form of permits) to new building new construction in San Francisco. The result of this is that supply has not been allowed to keep pace with demand.

V. Theoretical Model

Rent prices, displacement, increases in the number of skilled workers, and housing supply are interrelated. However, there are still theoretical underpinnings that dictate how this ecosystem will function.

Evictions are driven by the cost of rent. As rent increases, there will be some portion of the population that is unable or unwilling to pay the adjusted rent price, which will then lead to evictions. However, because of the prevalence of rent control, which only allows resets to market rates when there is a vacancy (as would be the case in the event of an evictions), it is also possible that evictions predict future changes in rent price.

Next, the supply of housing may drive rent prices, as any change in the supply of housing will impact the rent price. However, rent price may also drive housing, both through population changes,[3] and because increased rents encourage developers to produce more rental units.

An increase in in the number of high-income workers in San Francisco should increase rent price, primarily under the assumption that high income workers prefer more expensive homes, such as luxury developments than their lower income counterparts.

The link between high-income workers and evictions and housing is less clear. High income workers affect eviction and housing supply through changes in rent. However, it is also possible that San Francisco, which promoted high income growth through its courtship of tech companies, would similarly promote changes in the housing supply in anticipation of tech growth. Similarly, it’s possible that landlords are forward-looking, and evict tenants in expectation of increases in tech workers.

---

[3] People leaving due to increased rent prices.

VI. Empirical Model

While the economics behind how each component of the housing market should affect the others are well grounded, significant endogeneity arises when one tries to model this system. Each of these components may be simultaneously determined. Consider the supply of rental housing. While it is certainly a determinant of the price of rent, the price of rent also determines the supply of rental housing. I address this through the usage of a vector autoregression (VAR) model. Popularized by Christopher Sims, a VAR model treats variables as if they operate in an ecosystem, where each variable affects each other variable in a cycle of harmonious endogeneity. By doing this, I do not impose the assumption of exogeneity that a normal ordinary least squares estimator or system of equations does. My VAR model treats rent, displacement, changes in the labor force, and housing supply as a set of endogenous variables. Following the matrix algebra proposed by Enders (2010), for a given variable, such as rent, I can write:

$$R_t = b_{10} + b_{11}H_t + b_{12}D_t + b_{13}L_t + b_{14}R_{t-1} + b_{15}H_{t-1} + b_{16}D_{t-1} + b_{17}L_{t-1}\,,$$

where R_t is Rent price at present time *t*, H_t indicates Housing supply at present time *t*, D_t indicates displacement at time *t*, and L_t indicates changes in the labor force, specifically tech growth, at time *t*. The subscript $t{-}1$ indicates a one-month lagged value of each of the variables.

Each variable in my model depends on lagged and present values of itself, as well as lagged and present values of itself and other endogenous variables. This VAR will allow me to produce impulse response functions (IRFs), forecast error variance decompositions, and Granger causality tables.[4]

I assume that the dependent and independent variables are stationary[5] with standard deviations σ, and that the errors are uncorrelated white noise disturbances. For an *n*-variable VAR with a single lag, I first write this system in equation form:

$$y_t = b_{10} + b_{11}x_t + \ldots + b_{1n}n_t + \gamma_{11}y_{t-1} + \gamma_{12}x_{t-1} + \ldots + \gamma_{1n}n_{t-1} + \varepsilon_{yt}$$

$$z_t = b_{20} + b_{21}y_t + \ldots + b_{2n}n_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \ldots + \gamma_{2n}n_{t-1} + \varepsilon_{zt}$$

$$\vdots$$

$$n_t = b_{n0} + + b_{n1}x_t + \ldots + b_{nn}(n-1)_t + \gamma_{n1}y_{t-1} + \gamma_{n2}x_{t-1} + \ldots + \gamma_{nn}n_{t-1} + \varepsilon_{nt.}$$

I next collect contemporaneous values of my endogenous variables and rewrite this system of equations in matrix form, where y denotes a current value of an endogenous variable, and x corresponds a lagged value of an endogenous variable:

$$\begin{bmatrix} 1 & b_{12} & b_{13} & \ldots & b_{1n} \\ b_{21} & 1 & b_{23} & \ldots & b_{2n} \\ b_{31} & b_{32} & 1 & \ldots & b_{4n} \\ . & . & . & \ldots & . \\ b_{n1} & b_{n2} & b_{n3} & \ldots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ \vdots \\ b_{n0} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \cdots & \gamma_{2n} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \cdots & \gamma_{3n} \\ . & . & . & \cdots & . \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \gamma_{nn} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \\ \vdots \\ x_{nt-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}$$

---

[4] Full definitions of these terms can be found in Appendix C. However, impulse response functions can be thought of as the change in variable y as a result of a one standard deviation shock in z, evaluated at each time unit. Forecast error variance decompositions represent the percent of variance within a variable that can be explained by a shock of itself or from another variable(s), evaluated at each time unit.

[5] Stationary implies no unit roots, or in other words, that the data are asymptotically bounded.

In compact form:

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t \, .$$

I obtain the multivariate generalization by premultiplying by B^{-1}:

$$y_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 x_{t-1} + B^{-1}\varepsilon_t$$

Let $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1 x_{t-1}$, and $\epsilon_t = B^{-1}\varepsilon_t$, so that

$$y_t = A_0 + A_1 x_{t-1} + \epsilon_t \, .$$

I take the observed values of ϵ_t and recover ε_t because $\epsilon_t = B^{-1}\varepsilon_t$, which is necessary to obtain impulse response functions and forecast error variance decompositions. However, this presents a problem. Consider a two-variable VAR where

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{21} \end{bmatrix} = \frac{1}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix} .$$

The structural errors, ε_{yt} and ε_{xt}, depend on the values of b_{12} and b_{21}, which makes my system underidentified and does not allow us to obtain impulse response functions and variance decompositions. In order to have a system that is exactly identified, I restrict $b_{12} = 0$. This is equivalent to saying that a contemporaneous shock in *Y* does not affect current values of *X*. The number of restrictions I must impose on the system is equal to $\frac{n^2 - n}{2}$. So, in the four-variable system proposed in this paper, I must impose 6 restrictions upon the system. In a structural VAR, I impose these restrictions based on the theoretical implications of contemporaneous shocks on other variables. These relationships have been outlined in the table below.

$$\begin{bmatrix} Variables & S.TechCompanies & S.RentPrice & S.Evictions & S.HousingPer1000 \\ TechCompanies & 1 & 0 & 0 & 0 \\ RentPrice & ? & 1 & 0 & 0 \\ Evictions & ? & ? & 1 & 0 \\ HousingPer1000 & ? & ? & ? & 1 \end{bmatrix}$$

In this case, I restrict current values of tech companies to be unaffected by contemporaneous shocks in evictions, rent, and housing permits. Similarly, I constrain the current median market price of rent to be unaffected by contemporaneous shocks in housing supply and evictions but allow it to be affected by changes in the number of technology companies. Finally, I restrict evictions to be unaffected by a shock in the number of housing permits per capita and I allow the number of housing permits per capita to be affected by contemporaneous shocks in all the other variables.

Our resulting B and B^{-1} matrices now contains zeros along the upper triangle and are exactly identified. Because this system is now exactly identified, I can extract the structural errors from the observed and thus generate meaningful impulse response functions and variance decompositions.

One of the assumptions in any time series model is that the variables are stationary, meaning they have means and variances that are constant over time. To ascertain the validity of this assumption, I computed the Dickey-Fuller statistic for each of the variables in my model under the null that the variables are not stationary. All of these variables were deemed stationary. The results of these tests can be seen in Appendix A.

Once I determined that my variables met the conditions for an SVAR, I estimated SVARs with differing lag lengths and chose the model that minimized AIC, which is a measurement of how well fit a model is, penalized for number of parameters. However, I will present two SVARs of reasonably similar quality to demonstrate that the choice of lag structure can have ramifications on the results of my Granger causality tests, IRFs, and SFEVDs.

Finally, impulse response functions, structural forecast error variance decompositions, and Granger causality results were obtained from both models.

VII. Data Description

My model measures the relationship between displacement, rent prices, housing supply, and changes in the labor force. Due to data constraints, however, I must proxy for some of these variables, in particular, displacement and changes in the labor force.

In this paper, I use monthly eviction filings in San Francisco as a proxy for displacement (datasf.org, 2018). It is important to note that I am examining eviction *filings* and not actual evictions. This is an important distinction because not all filings are successful and thus, we do not observe the true number of evictions.

Next, I use monthly data on rent price, as measured by rentjungle.com from 2011-2016. Because the website's methods are consistent over time, even if there is a persistent bias it should be present equally throughout the dataset, and thus not bias my results. It is important to note that this estimate is not indicative of what the average San Franciscan pays for rent, because so many units are subject to rent control. Indeed, 2010 Census estimates place the actual average rent paid at $1,385, versus the $2,600 value rentjungle.com provides. However, because units do not come onto the market at this rent-controlled rate, this number is representative of what a prospective new renter would pay each month.

Tech growth was modelled by tabulating the number of registered technology business locations in San Francisco each month (datasf.org, 2018). Tech growth was then seasonally adjusted by tabulating monthly dummy variables, regressing those dummy variables against tech growth, predicting residuals and adding those residuals to the mean of tech growth. This was done because a disproportionate number of tech companies are registered in December, presumably for tax purposes.

This metric is limited, given that it groups “professional and technical firms”; thus, it is not a direct measure of tech growth. Further, it relies on companies to register their closures. While it is reasonable to assume that the vast majority of, if not all, tech companies are registered, this count may overstate their existence by failing to exclude companies that closed but did not register their closing. Finally, this metric does not account for the size of tech companies. For instance, Twitter may have significantly more impact on the economy than a new local start-up, however, both are given equal weight in this metric.

Lastly, housing supply per capita was approximated by taking the total number of houses in San Francisco according to the 2010 census and then supplementing this with data from San Francisco’s housing supply report, which tracks the quarterly changes in San Francisco’s housing supply. Initially, the change in housing supply, as defined as the total amount of housing units, regardless of size, was a quarterly variable. This was then converted into monthly data by averaging out the change in housing supply across the 12 months. Then, yearly population from the American Community Survey was converted into monthly data by averaging out changes in population across the year. Housing supply was then divided by population to estimate the housing supply per 1000. Excusing the issue of wide standard of error estimates in ACS and housing data, the primary problem with this methodology is the assumption that housing and population change is equal across all months. Another significant limitation is that this housing stock does not account for changes in types of housing. For example, even if the housing stock has not changed, it is possible that the amount of affordable housing decreased while the amount of luxury condos increased. Obviously, this would have significant policy ramifications, however, this nuance is not captured in the data.

The largest limitation of this dataset is not measurement error or an inaccessibility to proper metrics. Instead, it is the fact that the data measure trends on a city, and not neighborhood, level. The significance of this can be seen in the map presented in the beginning of the paper. Both tech growth and rent prices tend to be clustered together, and by looking at the city as a singular unit these trends can potentially be obscured if they only affect certain segments of the city. Obviously, this is an important factor to be considered, however, it will be left to future research.

Table 1 shows the summary statistics of my dataset in 2011 and 2016, the beginning and end of the time period I examine.

Table 1.

VARIABLES	mean	sd	min	max
Evictions 2011	108	23.74	60	143
Evictions 2016	177	85.22	117	435
Rent 2011	2,693	265.8	2,426	3,174
Rent 2016	3854.9	85.21	3,716	4,040
Housing/1000 2011	452.3	2.25	448.9	455.77
Housing/1000 2016	433.01	.509	432.3	434.05
Tech/Month 2011	99.5	31.3	39.7	173.7
Tech/Month 2016	119.66	58.7	-52.3	195.7

Overall, we see sharp increases in citywide monthly evictions, rent price, and tech growth. However, we see sharp declines in the housing stock per 1000 people. It is important to note that the evictions filed variable is not normalized and is thus impacted by the fact that San Francisco's population grew during the course of the sample.

VIII. Theoretical Expectations

Predicted Granger Causalities

Granger causality tests are generated by regressing current values of a variable, x_t for example, against lagged values of itself and another dependent variable, with the null hypothesis that the coefficients on the lagged values are jointly zero. If the coefficients of the lagged dependent variable are collectively non zero then the dependent variable, y_t , is said to Granger-cause x_t .

The policy implications of these Granger causalities are important. While not direct causal inferences, they do provide evidence as to what variables are symptoms and not causes of other variables. Furthermore, they also allow us to indirectly measure the effects of certain policy decisions, specifically rent control, on landlord behavior.

Prior to running the SVAR, I expected the following Granger causality relationships to hold, with question marks representing relationships I did not have prior beliefs about, X's denoting relationships I did not expect to be significant, and plusses signifying relationships I expected to be significant.

Table 2.

.	Tech	Rent	Evictions	Housing Supply
Tech Growth	X	+	+	+
Rent Price	X	X	+	+
Evictions	X	?	X	?
Housing Supply	X	X	X	X

Here I expect tech growth to Granger-cause changes in rent under the hypothesis that rents are raised in response to a change in demand (brought about by an increase in high income tech workers). However, I assume that landlords are not forward looking and do not raise rents in expectation of future tech migration. Under this hypothesis, I would similarly expect evictions and housing supply to be predicted by lag values of tech growth, but not the other way around.

Previous months' rent prices should predict current values of evictions because increasing rent values results in a substitution away from San Francisco rent. Lagged values of rent should also predict current levels of housing because, assuming the costs of development have not changed tremendously, changes in housing should be preceded by a shift in demand.

I have no priors as to how evictions impact rent and housing supply. Sixty percent of units in San Francisco are rent-controlled, which is significant because landlords can only reset properties to market rates if they evict their tenant. Thus, under a situation where landlords are constrained by rent control, it is possible that evictions may proceed rent increases. As for housing supply, if the city deems evictions to be a public problem then it may decide to increase the supply of housing, either through laxer regulations or an increase in public housing.

Finally, I do not expect housing supply to Granger-cause any other variables. This is partly due to time constraints. The maximum lag length that I considered was 6 months, which is far under the amount of time it takes to build a new house, thus, I we should see no Granger causality of housing supply on any of my variables.

Predicted Impulse Response Functions

Following the previous format, I expect the following relationships to hold. However, because there are 16 possible relationships between the impulse response functions I have elected to only include those I think will be significant.[6] As a note, the value at any given time period of an IRF represents the change in that month's value relative to the previous month, while the integral of the IRF represents the cumulative effect of the shock.

Table 3.

Impulse	Response	Intercept	Equilibrium
Rent	Evictions	+	0
Rent	Housing	+	0
Housing Supply	Rent	-	0
Tech	Rent	+	0
Evictions	Rent	+	0

Beginning with the effects of a one-standard deviation increase in rent (about $300) on the number of monthly evictions, I expect that the intercept will be positive, however, I expect the shock to quickly dissipate because the market should quickly adjust and the remaining tenants should be able to pay the new rent price.

I expect the shock of rent on housing to exhibit a similar trend. An increase in rent should spur construction, however, this number should again quickly adjust. One possible problem with this hypothesis is that housing takes longer than a few months to develop, which would skew the effects of our shock.

Next, a one-standard deviation increase in the number of new tech companies per month should increase median rent, however, as with the previous predictions, the effect of the shock should only be felt in the first month or two.

---

[6] Some other relationships, such as the effect of tech on evictions, may be significant, however, per theory, they are significant through some other mechanism, like tech increasing rent which then increases evictions.

Finally, a shock in evictions on rent should follow the same pattern; there should be a positive intercept, followed by a sharp return to zero as equilibrium is reestablished. This shock, however, is dependent on the assumption that the model understands implicitly that rent control is affecting rent price.

Predicted Structural Forecast Error Variance Decompositions

Structural forecast error variance decompositions represent the percent of variance within a variable that can be attributed to shocks in itself and to shocks in other covariates. These variances are evaluated at each "step," month in this case. The more one variable is responsible for itself the more exogenous it is. If a variable is significantly impacted by a shock from another variable, then it is more endogenous and more dependent on other variables.

I have indicated which variable I expect to be the "dominant" variable, along with whatever variable I think will be the second most important by the 24th month. Here dominant and important are defined by the ranking of variance attributed to that variable. An "X" indicates I do not expect there to be a significant secondary predictor.

Table 4.

Variable	Dominant	Secondary
Tech	Tech	X
Housing	Housing	Rent
Rent	Rent	Housing
Evictions	Rent	Housing

I predict the most dominant predictor of tech is tech, however, I believe tech to be relatively exogenous to the system (80% or more of the variance in tech is attributable to tech). Tech companies are likely lured to the San Francisco Bay Area by *something*, but that something is not, to my belief, encapsulated within this model.

Next, I expect the supply of housing to be largely determined by shocks to itself and by shocks in rent. This is under the assumption that developers, regardless of how hard it is to build in San Francisco, do respond to changes in demand and thus build more.

As for rent, I expect that most of the variation in rent will be caused by shocks in rent. This can loosely be interpreted as changes in rent being caused by shifts in demand that are not related to population growth or the presence of tech companies.

IX. Results

Granger Causality

Table 5. SVAR with 1- and 2-month lags included

Equation	Excluded	Chi Sq	Df	Probability Chi Sq
Tech	Rent	2.292	2	0.318
Tech	Evictions	0.11598	2	0.944
Tech	Housing/1000	3.6003	2	0.165
Rent	Tech	3.3199	2	0.19
Rent	Evictions	7.192	2	0.027**
Rent	Housing/1000	10.7	2	0.005***
Evictions	Tech	6.4457	2	0.04**
Evictions	Rent	10.001	2	0.001***
Evictions	Housing/1000	23.645	2	0***
Housing/1000	Tech	3.5729	2	0.168
Housing/1000	Rent	3.4885	2	0.175
Housing/1000	Evictions	2.1642	2	0.339

In Figure 6, we see that rent Granger-causes evictions and housing, which is congruent with my prior expectations. However, tech growth does not Granger-cause rent prices. This could be because there are simply not enough tech workers for them to have a significant impact on rent price, or because tech growth is concentrated within certain neighborhoods and as such the impact is only visible on a neighborhood level. According to my estimates, evictions Granger-cause tech companies, rent, and housing supply. Evictions Granger-causing rent prices can be explained by landlords working against San Francisco's rent control policies. If for 60% of housing units, rent prices can only be raised when new tenants enter a residency, it would certainly follow that rent prices would be predicted by lagged values of evictions. With regards to housing and tech, it's possible that landlords are indeed forward looking and are evicting tenants in response to predicted tech growth and population change (which is encapsulated within the housing variable).

Table 6. SVAR with 1, 2, 4, and 5-month lags included.

Equation	Excluded	Chi Sq	Df	Probability Chi Sq
Tech	Rent	3.9199	4	0.407
Tech	Evictions	4.9062	4	0.297
Tech	Housing/1000	5.046	4	0.283
Rent	Tech	3.1674	4	0.530
Rent	Evictions	6.8162	4	0.146
Rent	Housing/1000	12.782	4	0.012*
Evictions	Tech	8.3497	4	0.08*
Evictions	Rent	15.553	4	0.005***
Evictions	Housing/1000	26.336	4	0***
Housing/1000	Tech	3.6153	4	0.461
Housing/1000	Rent	12.596	4	0.0.13*
Housing/1000	Evictions	2.412	4	0.660

My second SVAR produced similar results to the first. The third month lag has been omitted because it is insignificant, and the inclusion of insignificant lags can lead to inconsistent parameter estimates. There are a few notable changes, however. First, the relationships between evictions and tech employment growth, and evictions and rent are both noticeably weaker. We see a similar pattern with the relationship between rent and evictions. This is likely due to the fact that only the first two-month lags of evictions are predictive of contemporaneous values. In particular, a 5-month lag of rent is likely not predictive of current evictions because someone who has paid the last 5 months of rent is likely still able to continue paying. Secondly, in the second SVAR housing supply Granger-causes changes in rent. Here it is possible that five months is a sufficient amount of time for new housing to enter the market.

Overall, the significance of evictions on rent and tech growth suggests that landlords are circumventing San Francisco's rent control policies.

Impulse Response Functions

Impulse response functions (IRFs) show the results of a one-standard deviation shock on a variable on the values of one the other variables in my SVAR over time. As an example, if there is an IRF of rent price on evictions and an intercept of 8, that would mean that a one standard deviation increase in rent corresponds to 8 more evictions immediately. Here the importance of the specification of the structural restriction is evident. The results of the IRFs are dependent on the structure imposed on the system because the structure dictates whether current shocks in one variable affect current values of another**.** The left graphs represent the SVAR with a 1, 2, 4, and 5-month lag structure.

Figure 2.

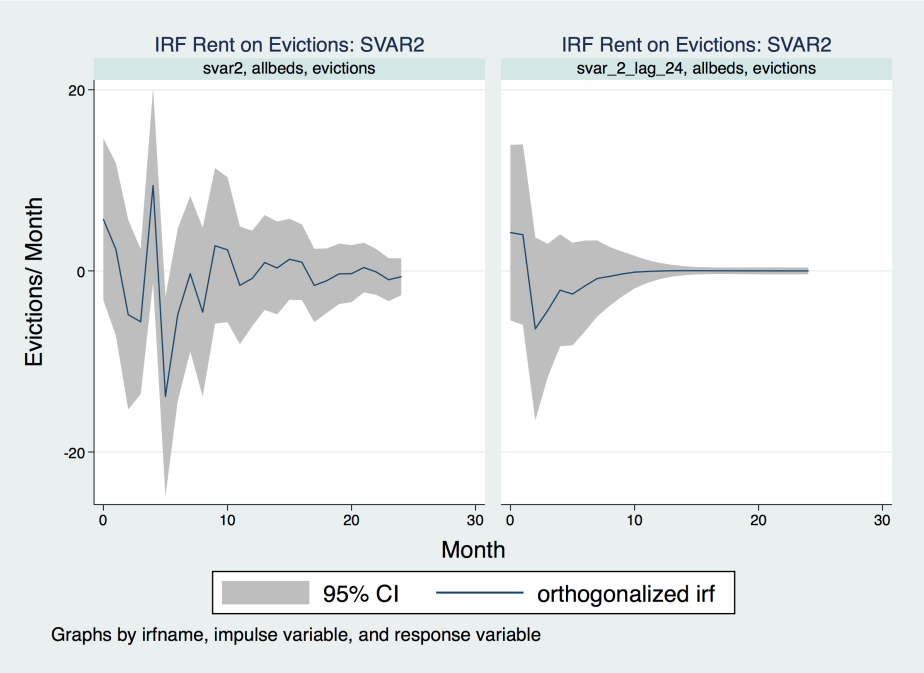


Figure 2 traces out the effect of a one standard deviation increase in rent on evictions. The intercept implies that the shock increases the number of evictions in San Francisco by about 8, afterword the number of monthly evictions falls to below the normal monthly average and eventually converges to zero, indicating that although evictions spike initially, there is no change in the long term number of evictions per month. Although the 95% confidence interval of the shock straddles zero, the effect and trend of a one standard deviation increase on the number of evictions per month is consistent with my theory. The intercept at time 0 is positive, which again is consistent with my theory, and the effect of the shock quickly diminishes as the market clears.

There are, however, differences between the two models as it pertains to the volatility of the shocks. The left graph is significantly more volatile than the right one, likely because the longer lag structure (erroneously or not) accounts for residents who were able to pay rent following the first two months but were eventually evicted. One notable aspect is that evictions immediately drop below equilibrium in the month following a rent increase. This is likely because there were a number of tenants who would have been evicted in the coming months under equilibrium conditions, but because of the shock, they were removed from the next month's pool of possible evictees, resulting in lower raw eviction numbers.

Figure 3.

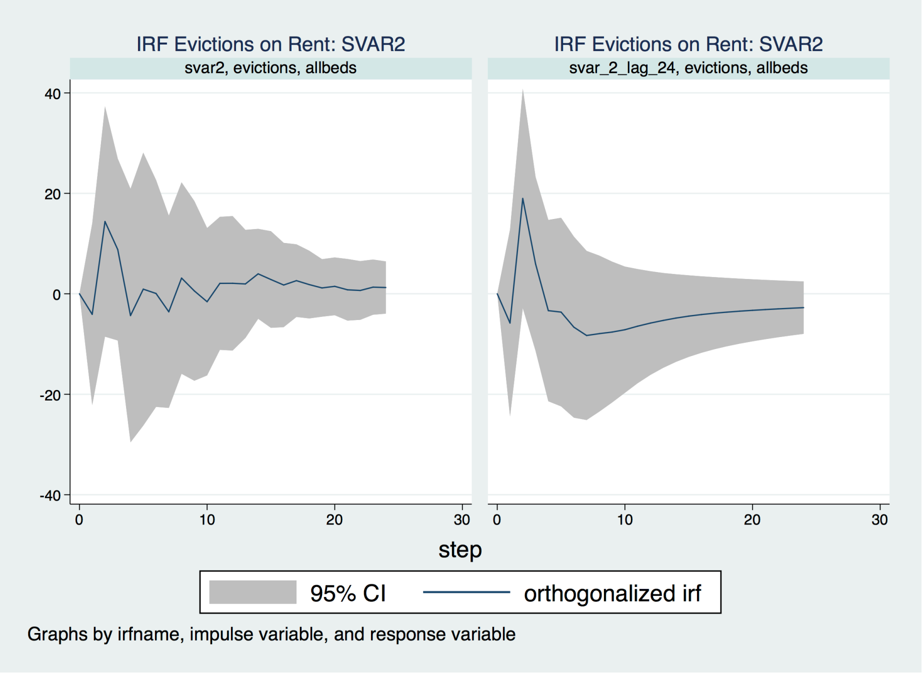
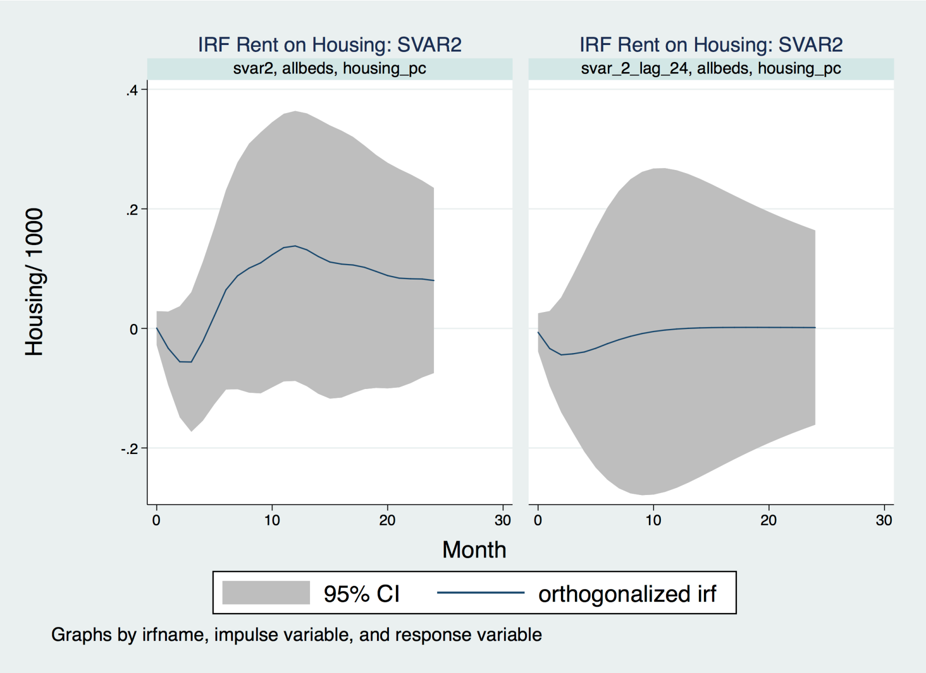


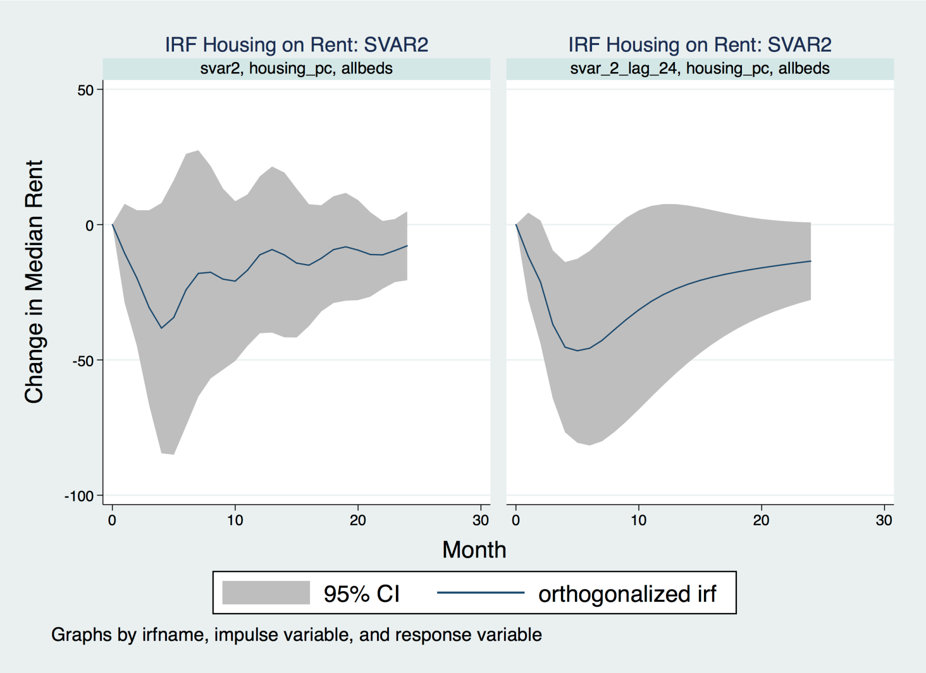
Figure 3 shows that an increase in evictions does seem to increase rent prices, at least in the short run. This is consistent with the notion that landlords are circumventing rent control by evicting tenants and then putting the property back on the market later. However, these results straddle zero, and as such are not technically significant.

Figure 4.



Interestingly, an increase of rent does not seem to impact the supply of housing. This could be because it takes years to build a house, and so we may not see the impact of an increase in rent until well after the 24-month time horizon. This is supported by the fact that, while again technically indistinguishable from zero, the model that includes a longer lag length has a longer and more pronounced increase over time.

Figure 5.



Again, the IRF, although straddling zero for the entirety of the first model and most of the second, is consistent with what my theory predicts as an increase in housing should lead to a decrease in rent. What is notable, however, is that while the impact of the shock continues decreasing over the entire time period of my IRF, it has not returned to the prior equilibrium at 24 months. The one standard deviation shock in housing per 1000 residents deserves particular attention because a one standard deviation shock in housing per 1000 implies either a constant population and an increasing housing supply, or a housing supply that increases with population growth. From a policy standpoint, this magnitude is an upper bound for what can be expected by an expansion of housing.

Figure 6.

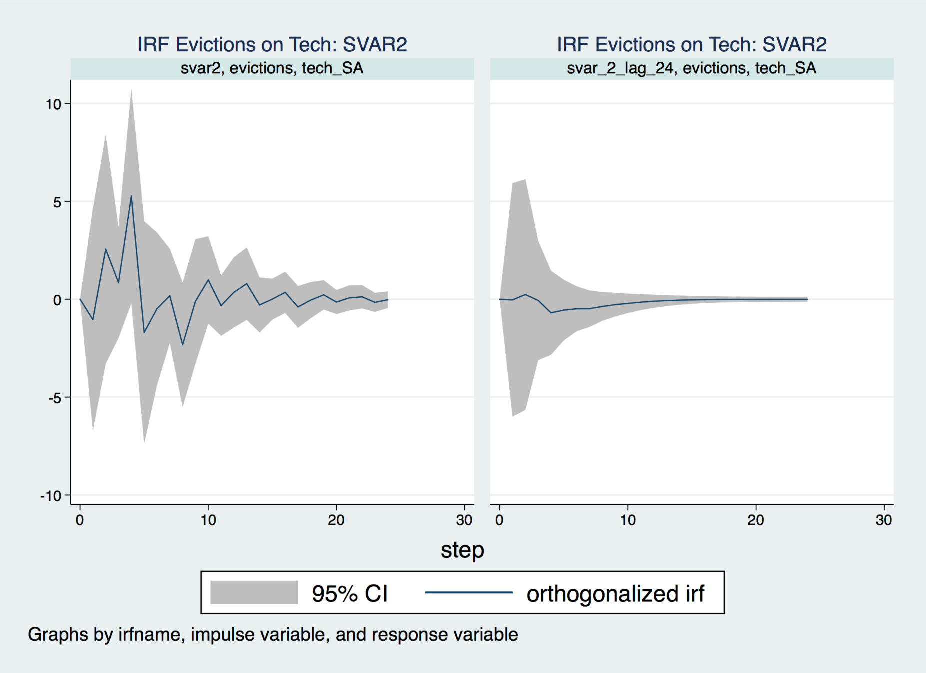
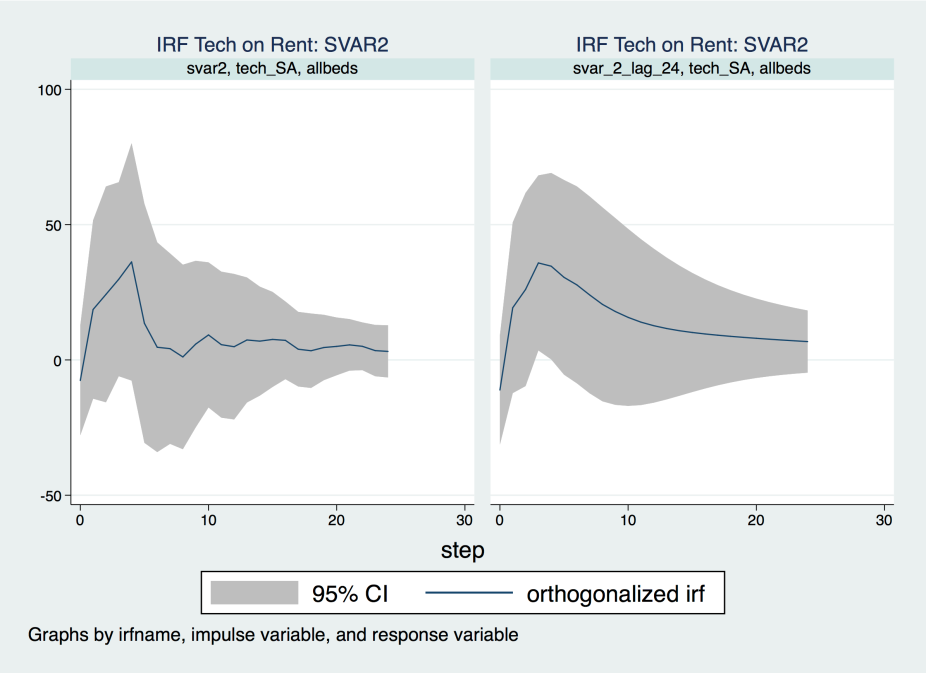


Figure 6 presents some evidence to the theory that landlords are forward looking and are evicting current tenants in expectation of an influx of wealthier potential tenants. While the one- and two-month lag SVAR shows no relationship between the two, the 1, 2, 4, and 5-month SVAR shows that there may be a relationship between evictions and tech growth.

Figure 7.

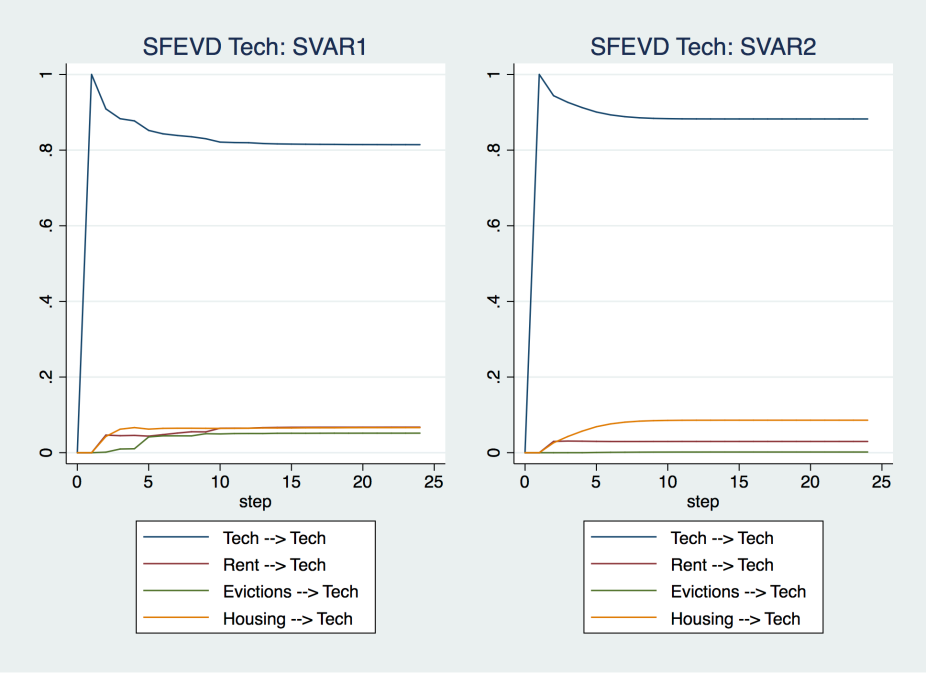


Finally, Figure 7 shows the effect of a shock in tech on rent. Especially in the 1- and 2-month lag SVAR, there is evidence of tech growth impacting rent prices, which is in exact accordance with what theory predicts.

Overall, while the majority of these IRFs have confidence intervals that straddle zero, and are thus technically insignificant, they do provide enough evidence to justify a further look into the relationships between these variables.

Forecast Error Variance Decomposition[7]

Figure 8.

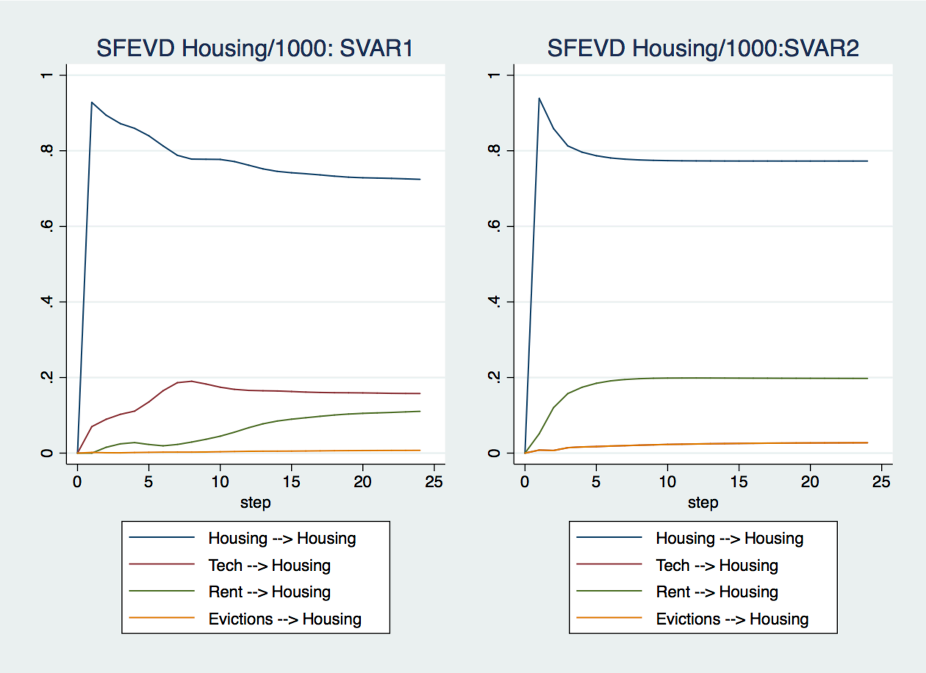


Beginning with Tech, I examine the percent of variance in each variable that is caused by shocks to itself and from rent, housing supply, and evictions. Here the results between the two presented SVARs are consistent with each other and show that change in tech growth is largely dependent upon a shock to tech. In practical terms, this lends evidence to tech growth being exogenous to the cost of rent, housing supply, and evictions. While this does not immediately lend itself to any policy recommendations, it does show that changes in population and productivity are largely exogenous to other housing market variables.

---

[7] Structural Forecast-Error Variance Decompositions Forecast can be read by reading the y axis of each line at a given time point. A value of .85 for a variable X at month 2 indicates that at month two 85% of the variance of some variable is caused by the shock toin X. In general, exogeneity increases as the percentage of variance in a variable that is caused by itself increase.

Figure 9.

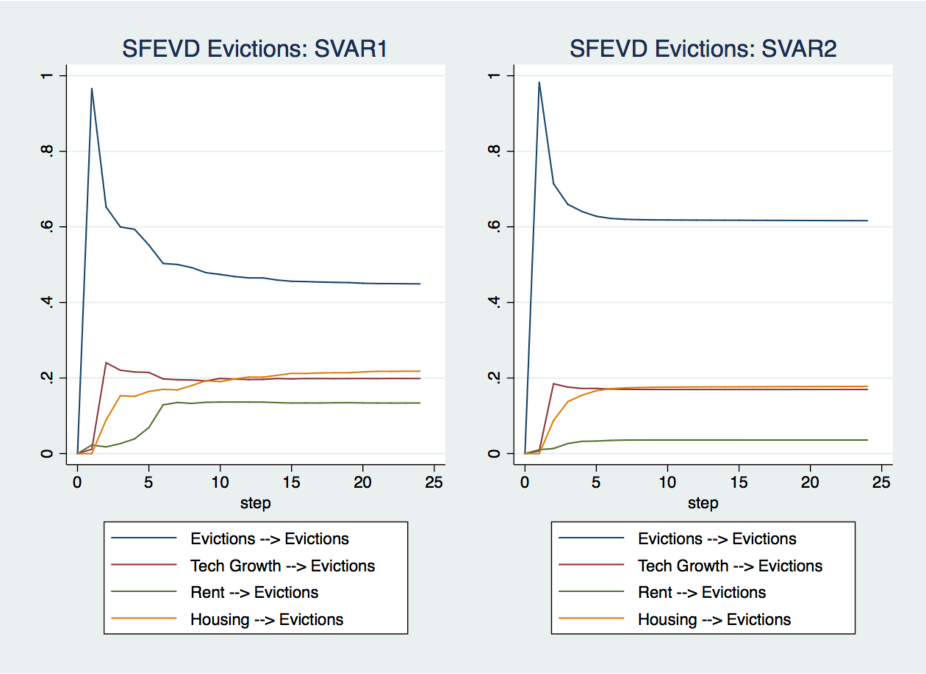


Next, the structural forecast error variance decomposition (SFEVD) of housing again shows a similar trend as tech with the strong majority of variance in housing being caused by shocks to housing. However, the two models do show differing conclusions for the impact of tech and rent on the supply of housing. The SVAR that utilizes a 1- and 2-month lag says that tech and rent have similar impacts on housing, with tech having a slight edge. However, the second SVAR with a 1, 2, 4, and 5-month lag structure shows tech having no impact on housing. Obviously, these results are very contradictory, and are caused by the fact that differing lag structures lead to differing coefficients, which lead to differing impacts of the shocks. Other than theory and criteria such as the Akaike Information Criterion (AIC), however, there are few ways to determine which lag structure is "right." Without questioning the validity of either SFEVD, it is notable that rent does appear to be a driver of housing, indicating that housing markets are responding to changes in demand. However, policy proposals are muddled because the variable is housing stock per 1000 residents, and as such housing may be exogenous, largely because migration is exogenous in these set of variables.

Figure 10.

Here my SFEVDs again give significantly different results. In the first SVAR, while the importance of rent as a driver of rent drastically decreases as time increases, it is still the most important variable in explaining the variance in rent. However, we do still see strong increases in the importance of housing supply. On the other hand, in the second model we see that the most important variable is housing supply. In both models we see tech have similar effects, while evictions are unimpacted. If the second SVAR is correct, then rental price in San Francisco would appear to be a supply-side issue, rather than a demand one. From this standpoint, San Francisco could do more to lower its rent price by expanding the city and allowing more housing than it could by imposing rent control ordinances.

Figure 11.



Finally, within evictions we again see somewhat contradictory results. Within both models there are similar significances between evictions and housing and tech, with each contributing to around 20% of the variance by the 20th month post-shock. However, between the two models there are very different estimates for the percent of variance caused by evictions and rent. The impact of tech and housing is notable, however, because it illustrates than tech growth and the shrinkage of housing supply do impact displacement, which is consistent with what prior literature suggests.

There is a common theme throughout all of these SFEVDs, however, and that is that evictions serve a minimal role in explaining the variance in any of the variables. If evictions are truly unimportant in forecasting the values of any other variables, it's possible that evictions are symptoms of some of the other variables, such as changes in rent price and the housing supply. It also raises questions about the results of the Granger Causality test and may indicate that evictions are acting as a proxy for some unknown variable which is driving changes in the other variables.

Residuals

The residuals for my SVARs, which can be seen in Figures 12-15 in Appendix A, represent ways to show how the models differ from each other and, more importantly, how they perform over the course of the time series.

As with all residuals, I am looking for straight lines clustered around zero, outliers, and heteroskedasticity. Beginning with tech, the residuals are consistent across both models and largely homoskedastic. Despite being seasonally adjusted, they are still incorrectly predicting values more in December-January each year, indicating that there may be a stronger seasonal effect in some years than others. Furthermore, there is a significant outlier in December of 2016, when there were significantly fewer new tech companies than predicted.

The residuals for rent suggest slightly more heteroskedasticity, with the residuals being slightly more clustered around zero at larger time values. The residuals also seem to be slightly serially correlated, with negative residuals preceding positive ones and vice versa.

The residuals for evictions are all consistently near zero and largely homoskedastic, with two obvious exceptions. In both June of 2012 and February of 2016, my models predicted significantly less evictions than actually occurred. This could be random noise in the data, however, there are large residuals for both tech and rent around this time period, which may indicate that there was a shock to the San Franciscan economy, possibly the tech tax, which was implemented around this time.

Finally, housing exhibits strong seasonal trends, with large positive residuals occurring yearly in November, followed by large negative residuals in January. This likely occurs because the seasonal adjustment is not entirely correctly calibrated, or that construction and change in population is not uniform across all months.

X. Conclusions

The majority of the results within my paper agree with both theory and intuition. Rent should drive evictions. Tech and housing should drive rent, while evictions, despite being predictive of rent should mostly be a symptom of increasing market price, and not actually a driver of it.

There are, however, some limitations of this study, as well as some avenues for future research. This research does not take place on the neighborhood level, and so localized neighborhood level relationships may be obscured. One could apply a similar methodology to neighborhood level data; however, this would pose a number of challenges. This would require the imposition of significantly more restrictions upon the model, to account for the inherent spatial correlation. This requires both a significant amount of technical expertise, as well as stronger theory and an additional year of monthly panel data to meet this restriction. There are additionally problems with using eviction filings, which obviously is not a perfect proxy for displacement, as well as the fact that I smoothed the housing per 1000 residents metric from quarterly data to monthly data, which suggests an inaccurate representation of how housing construction and migration actually occur.

Ultimately, this paper supplies evidence to the theory that markets must eventually reach equilibrium. While rent control can certainly play a part in keeping the current cost of rentals affordable, the evidence of evictions Granger Causing rent and tech growth suggests that landlords circumvent rent control policies. Furthermore, the seeming exogeneity of tech growth suggests that there is no way to prevent increases in high wage laborers into a city. If this is true, then a city must either increase housing or live with displaced residents.

References

Introduction to Analysis Constructing R from Q. Amherst College, www.amherst.edu/system/ files/ media/1526/Contruction%2520R%2520from%2520Q.pdf.

"Applied Econometric Time Series." *Applied Econometric Time Series - Walter ENDERS*, time-series.net.

Atkinson, Rowland. "The Hidden Costs of Gentrification: Displacement in Central London." *Journal of Housing and the Built Environment*, vol. 15, no. 4, 1 Dec. 2000, pp. 307–326, doi:10.1023/a:1010128901782.

"Dedekind Cut." *Wikipedia*, Wikimedia Foundation, 28 Oct. 2018, en.wikipedia.org/ wiki/Dedekind_cut.

Diamond, Rebecca, et al. "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco." 2018, doi:10.3386/w24181.

Ding, Lei, et al. "Gentrification and Residential Mobility in Philadelphia." *Regional Science and Urban Economics*, vol. 61, 2016, pp. 38–51, doi:10.1016/j.regsciurbeco.2016.09.004.

"Find Apartments in Your Area." *Rent Jungle: Apartment Search*, www.rentjungle.com/.

Freeman, Lance, and Frank Braconi. “Gentrification and Displacement New York City in the 1990s.” *Journal of the American Planning Association*, vol. 70, no. 1, 2004, pp. 39–52, doi:10.1080/01944360408976337.

Glaeser, Edward, and Joseph Gyourko. “Housing Dynamics.” 2006, doi:10.3386/w12787.

Glaeser, Edward, et al. “Urban Growth and Housing Supply.” 2005, doi:10.3386/w11097.

Grier, George, and Eunice Grier. “Urban Displacement: A Reconnaissance.” *Back to the City*, 1980, pp. 252–268, doi:10.1016/b978-0-08-024641-3.50022-2.

Killian, Lutz. *Chapter 4: Structural VAR Tools*. University of Michigan, www-personal.umich.edu/~lkilian/SVARch04.pdf.

Peter Marcuse, Gentrification, Abandonment, and Displacement: Connections, Causes, and Policy Responses in New York City, 28 Wash. U. J. Urb. and Contemp. L. 195 (1985).

Nevius, C.W. “Gabriel Metcalf Blog Hits Nerve on Housing.” *SFGate*, San Francisco Chronicle, 17 Oct. 2013, www.sfgate.com/bayarea/nevius/article/Gabriel-Metcalf-blog-hits-nerve-on-housing-4901915.php.

“Office of the Chief Data Officer | City and County of San Francisco.” *DataSF*, 26 Nov. 2018, datasf.org/.

“Planning Department.” *Planning Commission | Planning Department*, sf-planning.org/ planning-commission.

Sims, Christopher A. “Macroeconomics and Reality.” Econometrica, vol. 48, no. 1, 1980, pp. 1–48. JSTOR, www.jstor.org/stable/1912017.

Sims, David P. “Rent Control Rationing and Community Composition: Evidence from Massachusetts.” *The B.E. Journal of Economic Analysis and Policy*, vol. 11, no. 1, 2011, doi:10.2202/1935-1682.2613.

Sims, Peter. *Time Series*. Notre Dame, www3.nd.edu/~esims1/time_series_notes.pdf.

Suppe, Ryan. “Denver, the Midwest Emerge as New Hubs for Tech Workers.” *USA Today*, Gannett Satellite Information Network, 25 July 2018, www.usatoday.com/story/tech/news/2018/07/25/best-cities-tech-denver-midwest-emerge-new-hubs-tech-workers/819840002/.

Appendix A: Dickey-Fuller Tests and Residuals

Figure 12.

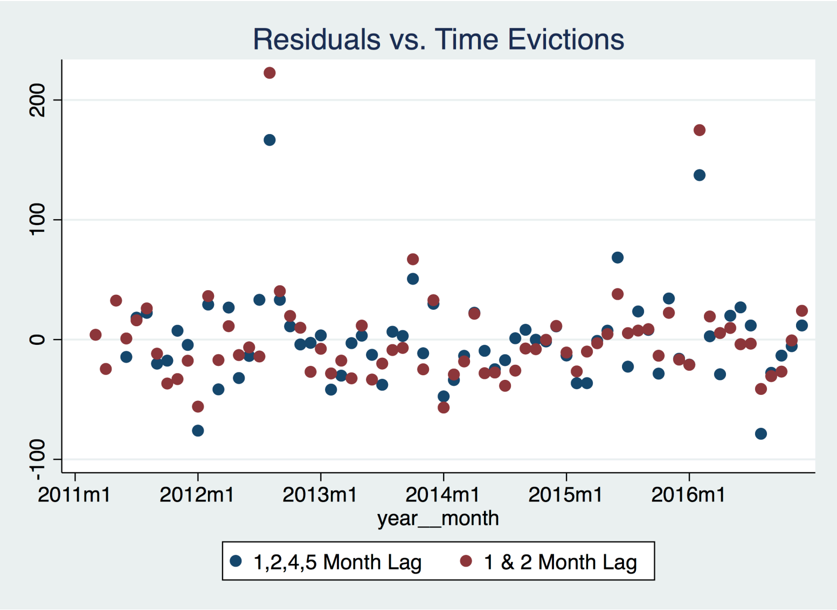


Figure 13.

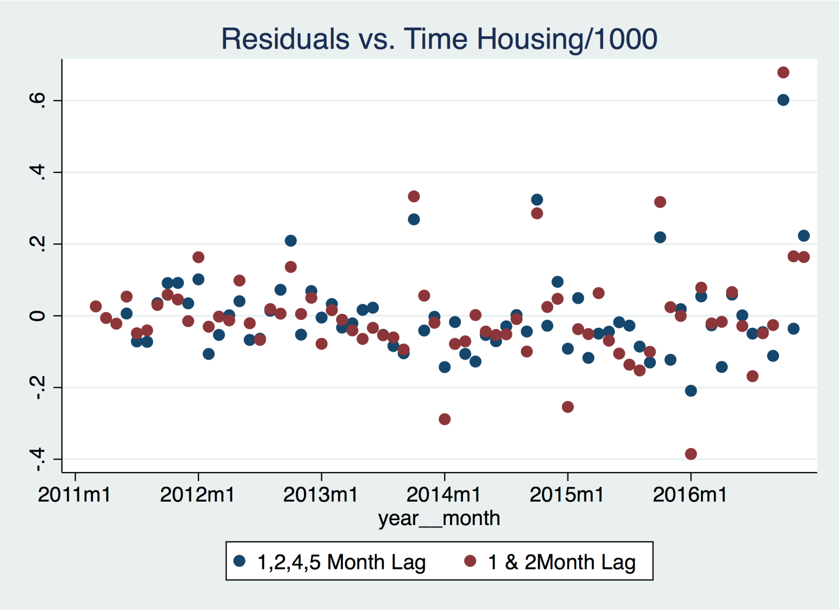


Figure 14.

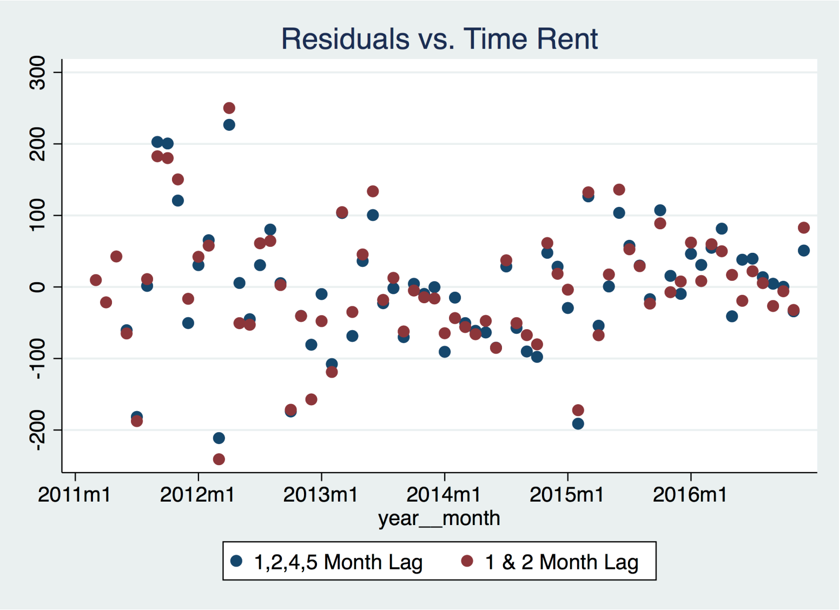


Figure 15.

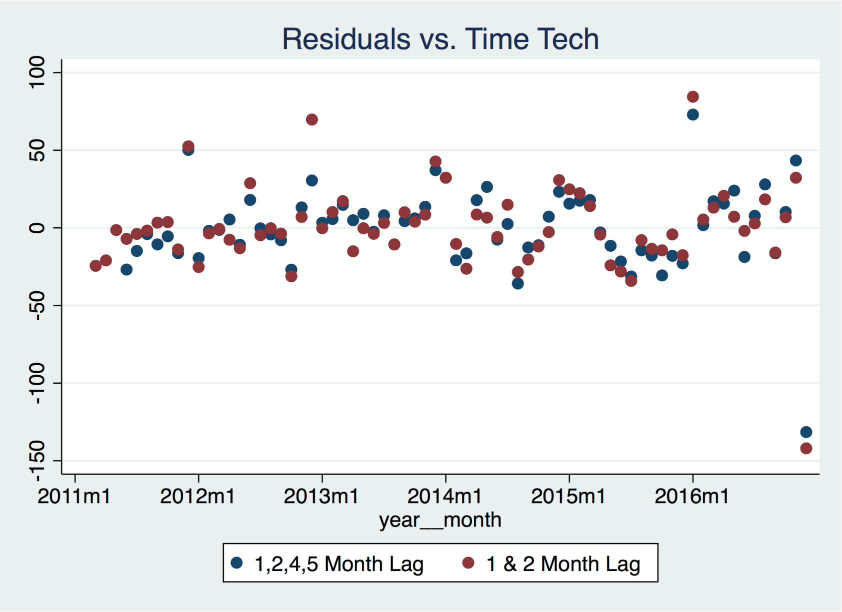


Table 7. Dickey-Fuller, Housing.

Housing	Coef	Std. Error	T	Probability t
L1.Housing	-.0155	.00599	-2.60	0.012
LD.	.8717	.1302	6.69	0
L2D.	-.0086	.1741	0.05	0.961
L3D.	-.795	.1911	2.79	0.007
L4.	.557	.1994	2.79	0.007
L5.	.00127	.2126	0.01	0.995
L6.	-.1839	.1722	-1.07	0.290
Constant	5.0117	1.735	2.89	0.05

MacKinnon Approximate P value: 0.0937

Table 8. Dickey-Fuller, Tech.

Tech	Coef	Std. Error	T	Probability t
L1.Tech	-1.605	.4223	-3.80	0.00
LD.	.2923	.3533	0.83	0.411
L2D.	0.0879	.2573	0.34	0.734
L3d.	0.0934	.1505	0.62	0.535
Trend	.129	.438	0.29	0.77-
Constant	179.7	46.57	3.86	0.000

MacKinnon Approximate P value: 0.00165

Table 9. Dickey-Fuller, Rent.

Rent	Coef	Std. Error	T	Probability t
L1.Rent	-.6454	.1662	-3.88	0.000
LD.	.1991	.1623	1.22	0.227
L2D.	.2872	.1496	1.92	0.060
L3D.	.3119	.1370	2.28	0.027
L4.	.0548	.1376	.1357	0.40
L5.	.016	.1321	0.12	0.904
L6.	-.025	.122	-.20	0.840
Trend	11.263	3.32	3.39	0.001
Constant	1765	430.61	4.10	0.000

MacKinnon Approximate P value: 0.0128

Table 10. Dickey-Fuller, Evictions.

Evictions	Coef	Std. Error	T	Probability t
L1.Evictions	-1.318	.350	-3.76	0.00
LD.	.2661	.304	0.87	0.386
L2D.	0.1281	.2573	0.50	0.622
L3d.	0.145	.1907	0.74	0.465
Trend	1.27	0.538	2.37	0.021
Constant	158.1	41.1	3.85	0.000

MacKinnon Approximate P value: 0.00185

Appendix B: Impulse Response Functions and Structural Forecast Error Variance Decompositions Definitions

Impulse Response Function: The change in the current and expected future values of random variable conditional on the realization of uncertainty at some point in time (Sims 2011). In math terms: $\text{IRF} = E_t x_{t+j} - E_{t-1} x_{t+j} \ \forall j \geq 0$.

Forecast Error Variance Decomposition: The percentage of the forecast error variance, or mean prediction squared error, of y_{t+h} at horizon $h = 0, 1, ..., H$ (Killian 2016).

Granger Causality: Granger causality tests are generated by regressing current values of a variable, x_t for example, against lagged values of itself and another dependent variable under the null that the coefficients on the lagged values are zero. If the coefficients of the lagged dependent variable are collectively non zero then the dependent variable, y_t , is said to *Granger-cause* x_t .

Evaluating Minnesota Minimum Wage Increases Using Synthetic Control Methods

Matthew Yang

Advanced Econometrics

Abstract

In this paper, I examine the employment effects of recent minimum wage increases in Minnesota using both synthetic control methods in addition to a conventional difference-in-difference approach. Across both methods, I find that the change in minimum wage resulted in no significant effects on employment rates. Additionally, an analysis of total quarterly wages in MN and WI shows that the policy did indeed increase total wages in MN. This implies that if there was any decrease in employment, it was outweighed by the increase in the minimum wage. Since employment in MN did not decrease after the minimum wage increase, the policy seems to be net beneficial. That is, employment did not fall, and the minimum wage increased, so total wages should increase.

I. Introduction

Economic theory predicts that in competitive labor markets, an increase in wages will decrease employment due to the higher cost of labor for employers. Nonetheless, advocates push to increase the minimum wage so that low-wage workers will be able to earn enough money to sustain a baseline quality of life. In this view, even if the minimum wage has a small negative effect on employment, the policy is still beneficial overall because it improves earnings for low-wage workers. But if the negative effect is large, would the higher wage of some workers still outweigh the lost employment for other workers? This leads to a central question in the minimum wage debate: By how much does an increase in the minimum wage decrease employment?

In this paper, I examine the employment effects of recent minimum wage increases in Minnesota. Up until 2014, MN did not have a binding state minimum wage. That is, the state minimum wage was below the federal minimum wage, so employers were required to pay the federal minimum wage, which is currently $7.25. In August of 2014, the MN state minimum wage was raised to $8 as part of the first phase of the state's plans to adopt a $15 minimum wage. The MN state minimum wage has increased in each subsequent year and is now $9.65.

In a 2018 policy brief from the Center for Research on the Wisconsin Economy (CROWE), Noah Williams compares employment levels in Wisconsin and Minnesota before and after the minimum wage increases in Minnesota. Williams's findings suggest that, overall, Minnesota's minimum wage increases have not been beneficial. He finds that while some workers received higher incomes after the wage increases, general employment in the state fell sharply. Moreover, Williams finds that restaurant prices increased significantly following the wage increases. This paper extends Williams' analysis by implementing different methods.

In any study of a policy intervention, one of the most important considerations is the construction of a control group. Previous studies of state-level minimum wages have chosen nearby states that kept the minimum wage constant during the time period of interest (such as Williams's choice of WI as a control for MN). This method of selecting the control group introduces arbitrariness into the analysis and can drastically affect results (Abadie et al., 2011). The synthetic control method, which is explained below, remedies this problem by using a data-driven algorithm to find the optimal control group.

The remainder of this paper is organized as follows. Section II reviews the relevant literature. Section III describes the econometric methods employed in the paper. Section IV discusses the data and implementation of the methods. Section V presents the results. Section VI provides concluding remarks.

II. Literature Review

There are a few central questions that drive research on the minimum wage. Neumark et al. (2014) summarize them succinctly: (1) How does a minimum wage affect employment? (2) Which workers are affected by a minimum wage? (3) How can we use econometrics to isolate the effects of the minimum wage? There is no conclusive answer in the literature to any of these questions. In this section, I will examine the major contributions made in the minimum wage literature and compare them in the context of these three questions.

The vast majority of papers on minimum wage can be classified into two groups. The first group of papers uses a two-way fixed effects estimator to identify trends in employment across states with different minimum wage levels. These papers employ panel data on employment across states and time and include state and period fixed effects in the model. This model enables the researcher to estimate the general employment effect from *all* of the minimum wage policies. The second group of papers attempts to study employment effects at a more granular level by examining individual case studies. That is, researchers match a specific place where there was variation in the minimum wage to a place where the minimum wage was constant, and then compare employment levels.

Neumark and Wascher (1992) were among the first to advocate for the use of a two-way fixed effects model. Prior to their paper, the majority of minimum wage studies used national-level time series data to identify employment effects. Neumark and Wascher add in state fixed effects to the model in order to exploit more variation in the minimum wage. The authors find employment elasticities ranging from -0.1 to -0.2 for teenagers and young adults, implying that a 10 percent increase in the minimum wage would result in a one to two percent decrease in employment. Many subsequent papers that used a two-way fixed effects estimator found similar disemployment effects. A main critique of this model is that is does not account for heterogeneous effects across states. That is, states may face different economic and social conditions overtime. Therefore, the two-way fixed effects model fails to account for time variant factors other than the minimum wage that may be confounding the employment effects.

Allegretto, Dube, and Reich (2011), hereafter ADR, remedy this critique of not accounting for spatial heterogeneity by controlling for both long-term growth differences among states and heterogeneous economic shocks. ADR control for regional differences among states by including "census division-specific time effects," which effectively compares a state only to those states within the same census division. This revised identification strategy yields employment elasticities that are statistically equal to zero. Other papers that have added spatial controls to the traditional fixed-effects model have found similar results.

Instead of comparing states within each census division, Dube, Lester, and Reich (2010), hereafter DLR, compare employment in every county pair that shares a state border. DLR find that higher minimum wages have no effect on employment, thereby confirming ADR's findings that controlling for spatial heterogeneity eliminates any significant disemployment effect. In addition to using this county matching strategy, DLR estimate employment effects using a traditional fixed effects specification and find negative employment elasticities. These findings are consistent with the pattern that has emerged in the minimum wage literature over the past several years: the traditional two-way fixed-effects model produces downwardly biased results. In his review of Neumark and Wascher's *Minimum Wages*, Dube (2010) writes, "Even simple regional controls and trends produce employment effects close to zero, as do more sophisticated approaches such as comparing contiguous counties across policy boundaries—which essentially embeds the 'case study' approach within panel data analysis." DLR's analysis demonstrates the importance of constructing adequate control groups.

Paramount to any study of an intervention on a specific group is the ability to estimate a counterfactual. In order to know the effect of a policy, we need to predict what would have happened in the absence of the policy. The case study approach attempts to estimate the unobserved counterfactual by using states or counties in close proximity to the treatment group as the control group. Case studies allow researchers to carefully choose a control group, eliminating the problem of spatial heterogeneity. Researchers often use a control group in close geographic proximity to the treatment group in order to account for regional shocks.

Card and Krueger (1994) offered one of the first case studies of the minimum wage. Much of the subsequent literature on the minimum wage in the last few decades are built on their work. The authors estimate employment effects in the fast-food industry by comparing two neighboring states: Pennsylvania and New Jersey. These two states offered a good case study because the minimum wage rose in NJ, while it stayed constant in PA. Researchers studying the minimum wage commonly focus on the fast-food or limited-service restaurant industries because many of the workers in these industries earn the minimum wage. Therefore, an analysis of employment in these sectors allows researchers to isolate the effect of the minimum wage on workers for which the minimum wage is binding. Card and Krueger use a difference-in-differences estimator to identify employment effects. They found no evidence that the rise in minimum wage reduced employment in NJ.

The synthetic control method offers a similar way to estimate the effect of an intervention on a treatment group, as compared to a control group. However, the synthetic control estimator chooses the control group in an empirical manner instead of leaving it up to the researcher's discretion, so there is less subjectivity in the construction of the control group (Abadie et al., 2007). Abadie, Diamond, and Hainmueller (2010) use the method to study the effect of a tobacco-tax program on cigarette consumption in CA. Instead of choosing a state as a direct comparison, the synthetic control method uses an optimization algorithm to compute a control group that is comprised of a weighted average of potential control states. Abadie et al. refer to these potential control states as "donor" states. The authors showed that the synthetic control method could provide a way to estimate a counterfactual as opposed to using a single control group.

A number of researchers studying the minimum wage have started turning to the synthetic control method as a way of comparing employment in states to a data-driven choice of control group. However, the application of the synthetic control method to minimum wages is not entirely straightforward. Minimum wage increases happen very frequently and in many places. As such, the states in the donor pool must be states that kept the minimum wage constant during the time period of interest, so the synthetic control does not experience any employment effects from a minimum wage increase. Neumark, Salas, and Wascher (2014), hereafter NSW, implement the synthetic control method to assess the validity of the assumptions of ADR. ADR showed that running the two-way fixed-effects model with added controls for census division and state employment patterns can produce unbiased employment elasticities. However, ADR implicitly assume that states within the same census division serve as good controls. NSW test this assumption by running the synthetic control algorithm on states in each census division and examining whether the algorithm chooses states within the same census division as controls. NSW find that, generally, it is not the case that states within the same census division provide a better control than other states do. Moreover, after comparing employment in states that increased the minimum wage to synthetic controls, NSW find evidence of disemployment effects, estimating employment elasticities for teenage workers around -0.15.

Reich, Allegretto, and Godoey (2017) implement an analogous experiment to NSW and arrive at the same conclusion. Specifically, Reich et al. employ the synthetic control method on county-level data to estimate employment effects of Seattle's minimum wage increases. The authors find that the optimal weighted control group selected by the algorithm contains counties outside Washington State, indicating that proximity does not always predict the adequacy of a control group. Indeed, the best method for constructing a control group and estimating a counterfactual continues to be a fundamental issue driving the minimum wage debate today.

Williams (2018) was the first author to examine Minnesota's recent minimum wage increases. After comparing limited-service restaurant employment in Minnesota and Wisconsin, Williams found that employment was significantly lower in Minnesota than in Wisconsin after the first minimum wage increase in 2014. Although Williams presents convincing evidence, Williams does not employ very robust statistical methods. Importantly, Williams does not present any inference for his results, so the significance of his findings is unclear. In contrast to Williams, Chinn and Johnston (2018) find no significant employment effects from the MN minimum wage increase. This paper will further explore the robustness of Williams' findings by implementing alternative methods.

III. Methods

In this paper, I employ two estimation techniques: the synthetic control method and a difference-in-differences estimator. For the former, I construct a control group that is a weighted average of states that kept the minimum wage constant. For the latter, I choose WI to serve as the control group for MN following Williams's analysis. Although Williams directly compares employment in MN to employment in WI, he does not fit any econometric models to test his hypotheses. Therefore, a thorough investigation of Williams's hypotheses is needed to evaluate his conclusions. Each of the techniques I employ has its advantages, so I use both to provide a more complete picture of employment effects in MN.

i. Synthetic Control Estimator

The first step in the synthetic control method is establishing "donor" units. In my analysis, I define the donor units as states for which the binding state minimum wage remained constant during the period of 2010 to the present. If we were to consider all states, including those that increased the minimum wage, we would be contaminating the control group with disemployment effects, and would therefore not obtain accurate results. In the following paragraphs, I refer to the increase in the minimum wage as the "intervention," with Minnesota serving as the "treated" state.

Assuming we have J donor states and one treated state, there are $i = 1, \dots, J + 1$ states in our analysis. We define the treated state to be the first state of the total units (i.e. $i = 1$). Furthermore, we can define the number of time periods for which we observe the states as $t = 1, \dots, T$, where T_0 is the period in which the intervention occurs, and $1 \leq T_0 < T$.

We are interested in comparing employment levels in the treated state post-intervention to the employment levels that would have been observed absent the treatment. Let Y_{it}^{I} represent the employment level in state i at time t if the state is exposed to the intervention in periods $T_0 + 1$ to T. We define the estimated treatment effect to be $\alpha_{it} = Y_{it}^{I} - Y_{it}^{N}$ for time periods post the initial intervention. Rearranging, we get,

$$Y_{it}^{I} = Y_{it}^{N} + \alpha_{it}$$

Since only one state in the sample is exposed to the intervention, we can generalize this relationship by introducing a dummy variable that is equal to one if the state is exposed to the intervention at time t and zero otherwise. Therefore, we have:

$$Y_{it} = Y_{it}^{N} + \alpha_{it} D_{it}$$

This relationship implies that the employment level is equal to Y_{it}^{N} for all states at all times except for the treated state after $t = T_0$.

All of the quantities in the previous equation are observed except for Y_{it}^{N} for unit $i = 1$ at $t = T_0 + 1, \dots, T$. In order to estimate the unobserved counterfactual, we need to construct a synthetic group that is a weighted average of the treated unit. We consider the following model (as described in Abadie et al. (2010)) to estimate Y_{1t}^{N} after the intervention:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}$$

In this model, the w_j's come from a $(J\ x\ 1)$ vector of weights W. All weights are nonnegative and must sum to one. Therefore, the vector W describes the weights that make up the synthetic group. The first term, δ_t, is a constant that is common among all units. Z_j is a $(r\ x\ 1)$ vector of observed explanatory variables, and θ_t is a vector containing the unknown coefficients. Furthermore, λ_t and μ_j describe an unknown common factor with different factor loadings across units, and ε_{jt} are the error terms.

A critical step in the synthetic control method is determining the vector W. As Abadie et al. (2010) describe, this is done by minimizing two quantities. First, we want to choose weights such that the synthetic group closely approximates the treated group in the pre-intervention period. If we can find a group that acts like the treatment group pre-intervention, then the simulated post intervention outcomes for the synthetic group become more believable. So, we have the following:

$$\sum_{j=2}^{J+1} w_j^* \overline{Y}_J^{\,K} = \overline{Y}_1^{\,K},$$

where w_j^* are the optimal weights and $\overline{Y}_J^{\,K}$ is the average employment in the pre-treatment period. Rarely does this equality hold exactly, but we can find the optimal weights by minimizing the error.

Second, the synthetic group should approximate the values of the observed covariates for the treatment group. Thus, we have the following:

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$$

By minimizing the error in these two equations, we can find the vector W and thereby estimate the treatment effect of the intervention. The following equation gives us the estimated treatment effect:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j * Y_{jt}$$

ii. Difference-in-Differences

A difference-in-differences estimator is a more traditional way of estimating a treatment effect from an intervention that has been employed numerous times in minimum wage studies. This model relies on the parallel trends assumption, which states that both the control group and treatment group must have similar trajectories prior to the intervention (Powell, 2018). The synthetic control method relaxes this assumption because it finds a synthetic control that matches the treatment group very closely prior to the intervention.

The treatment effect is then calculated by taking the difference in the differences between the two groups. That is, if we expect the difference in the outcome variable to remain constant across time between the two groups, then the change in that difference post-intervention is the estimated treatment effect.

IV. Data and Implementation

In this paper, I employ data from the Quarterly Census of Employment and Wages (QCEW). Specifically, I focus on limited-service restaurant (LSR) employment, as is common in the literature. The data is from 2010-2018, which allows us to have roughly equal periods before and after the minimum wage increase. Moreover, the QCEW contains information on various labor market characteristics, such as employment, wages, and number of LSR establishments. The main outcome variable in this study is employment, as I am interested in how much employment changed due to the increase in minimum wage. Additionally, I fit a model to estimate the effect that the minimum wage increase had on total wages.

First, I use the synthetic control method to estimate the magnitude of the treatment effect for employment. I fit several models using this method but will only present two in this paper. As discussed in the previous section, the construction of the Z matrix is important in the construction of a synthetic control. In the first model, I include average weekly wages, total taxable wages, and establishment count as covariates. Additionally, a much better synthetic control can be obtained when including specific lags of the dependent variable in the model. Therefore, I include 2010Q1, 2012Q1, and 2014Q1 employment as predictors. In order to test the sensitivity of the synthetic control algorithm, I take some predictors out of the model and compare the results.

The inference methods for the synthetic control method differ from those commonly performed with traditional regression analysis. The synthetic control method does not report standard errors, so I employ other techniques to understand the significance of the results. In this paper, I adopt the inferential methods put forth in Abadie and Gardeazabal (2003). More specifically, I fit placebo models to understand the ‘randomness’ of the results.

Next, I use a difference-in-differences estimator to compare the results obtained when using only WI as a control for MN, rather than a weighted average of donor states. This is implemented using the following regression model:

$$Employment = \beta_0 + \beta_1 MN + \beta_2 2014Q3 + \beta_3 MN * 2014Q3 + \varepsilon$$

where MN is a dummy variable that equals one if the state is MN and zero otherwise, and 2014Q3 is a dummy variable if the time period is after the third quarter of 2014. The interaction term between MN and 2014Q3 estimates the treatment effect.

Lastly, I measure the effect that the minimum wage increase had on total wages in MN using an analogous difference-in-differences model (with WI as the control), as shown here.

$$Total\ Quarterly\ Wages = \beta_0 + \beta_1 MN + \beta_2 2014Q3 + \beta_3 MN * 2014Q3 + \varepsilon$$

V. Results

i. Synthetic Control Method

The quality of the synthetic control can be measured by the root mean squared prediction error (RMSPE). The synthetic control model presented in Table 1 resulted in the best fit[8] (lowest RMSPE) of any of the models I ran. This model assigns positive weight to Illinois, Iowa, Maine, Missouri, and Wisconsin. Moreover, Figure 1 suggests that employment in synthetic MN becomes slightly higher than employment in actual MN after the initial minimum wage increase in August 2014, implying that the treatment effect is slightly negative. The placebo models[9] (shown in Figure 2) obtained after iteratively reassigning the treated unit show that the treatment effect for MN is not large relative to the other states. Thus, the results from this synthetic control model imply that there is not a very significant treatment effect for employment.

After taking the number of establishments out of the model and matching only on one time period pre-intervention (as opposed to three), I obtain markedly different results. Here, the

---

[8] The RMSPE in this model is 640.

[9] Each gray line represents a placebo model, and the orange line represents the magnitude of the treatment effect.

synthetic control algorithm chooses Idaho, Iowa, and North Carolina as part of the control group. The only control state in common with the previous model is Iowa, indicating that Iowa's LSR labor market exhibits many of the characteristics as that of MN. Figure 3 plots MN and synthetic MN employment. In contrast to Figure 1, this plot shows employment levels between the two groups diverging after the time of the intervention. However, the RMSPE for the second model is 1765, nearly triple the RMSPE for the first model. The large RMSPE in this model indicates that the synthetic control is a much worse approximate of MN than the corresponding synthetic control in the first model.

ii. Difference-in-Differences

Table 5 presents the difference-in-differences regression results for quarterly employment. The interaction term between MN and 2014Q, which represents the treatment effect, is positive but insignificant. Therefore, this model implies that the treatment effect is effectively zero. This result agrees with the conclusion reached from the synthetic control models that the minimum wage increase did not have a significant effect on employment. Additionally, the MN dummy variable is insignificant, implying that there is no difference in average LSR employment between the two states, regardless of time. Figure 5 supports this finding, as a simple plot of LSR employment overtime shows almost no difference between the two states over the entire period.

Since employment in MN did not decrease after the minimum wage increase, the policy seems to be net beneficial. That is, employment did not fall, and the minimum wage increased, so total wages should increase. The results from Table 6 confirm this, as the treatment effect on total wages is positive and significant, indicating that the policy increased total wages in MN relative to WI. Figure 6, which plots total quarterly wages in MN and WI over the entire period, shows the gap in total wages between the two states increasing after the intervention.

VI. Conclusion

This paper finds no evidence that the increase in the MN minimum wage decreased LSR employment. I present two main analyses to support this conclusion. The first synthetic control model discussed in Section V estimates a counterfactual that is close to the observed employment levels in MN after the time of the intervention. Moreover, the placebo study implied that the marginal disemployment effect was not significant. Although the second synthetic control model predicts that employment in the synthetic control (i.e. in the absence of the intervention) was significantly higher, the sensitivity of the synthetic control method alone is enough to invalidate the result. Furthermore, the synthetic control obtained in the second model had three times the error as the first model, indicating that the control group is less similar to MN.

Next, I compare LSR employment in MN and WI using a difference-in-differences model. I find that the treatment effect is slightly positive, but insignificant. Additionally, an analysis of total quarterly wages in MN and WI shows that the policy did indeed increase total wages in MN. This implies that if there was any decrease in employment, it was outweighed by the increase in the minimum wage.

There are a few main limitations of my analysis that may be affecting the results. In particular, the synthetic control method is extremely sensitive to the inputs included in the model. This could be the result of a few different factors. First, I employ non-seasonally adjusted data. The variability in employment may increase the error of the synthetic control method, as it might be harder to find a perfect fit. Next, the data contains a relatively short pre-intervention period of just under 5 years. Abadie et al. (2010) employ nearly 20 years of data prior to the intervention. Therefore, I may not have enough data to find an adequate synthetic control. The synthetic control method is difficult to implement in minimum wage studies for this reason—state minimum wage levels are constantly changing. So, when trying to isolate the effect of a minimum wage on a particular state, the number of potential control states is limited. If I were to extend the time period pre-intervention, there would be fewer states that held the minimum wage constant, which would limit the power of the synthetic control method.

As I have shown, implementing different methods than those presented in Williams's CROWE policy brief results in starkly different results. The synthetic control and difference-in-differences analyses presented in this paper provide evidence that the MN minimum wage did not reduce LSR employment and did increase total wages. Finding opposite or inconsistent results has become a common trend in the minimum wage literature. The variability in findings underscores the need for greater methodological refinement of control groups to be conducted. Only then will we be able to ascertain more clearly who are the winners and losers of a minimum wage policy.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 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, Alberto, Alexis Diamond, and Jens Hainmueller. 2011. Synth: an *R* package for synthetic control methods in comparative case studies. *Journal of Statistical Software.* 42(13): 1–17.

Abadie, Alberto and Javier Gardeazabal. 2003. The economic costs of conflict: a case study of the Basque country. *American Economic Review* 93(1): 112-132.

Allegretto, Sylvia A., Arindrajit Dube, and Michael Reich. 2011. Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. *Industrial Relations* 50(2): 205–40.

Card, David, and Alan B. Krueger. 1994. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* 84(4): 772– 93.

Chinn, Menzie, and Louis Johnston. 2018. The impact of minimum wage legislation in Minnesota vs. Wisconsin: contra Williams.

Chinn, Menzie. "Time Series Evidence on the Minimum Wage Impact in Minnesota vs. Wisconsin." *Econbrowser,* November 5, 2018, econbrowser.com/archives/2018/11/ time-series-evidence-on-the-minimum-wage-impact-in-minnesota-vs-wisconsin.

Doucouliagos, Hristos, and Tom D. Stanley. 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations.* 47(2): 406-428.

Dube, Arindrajit. 2011. Review of *Minimum Wages* by David Neumark and William L. Wascher. *Journal of Economic Literature* 49(3): 762–66.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. Minimum wage effects across state borders: Estimates using contiguous counties. *Review of Economics and Statistics* 92(4): 945–64.

Neumark, David, J. M. Ian Salas, and William Wascher. 2014. Revisiting the minimum wage–employment debate: Throwing out the baby with the bathwater? NBER Working Paper No. 18681. Cambridge, MA: National Bureau of Economic Research.

Neumark, David, and William Wascher. 1992. Employment effects of minimum and subminimum wages: Panel data on state minimum wage laws. *Industrial and Labor Relations Review* 46(1): 55–81.

Powell, David. 2018. Imperfect Synthetic Controls: Did the Massachusetts Health Care Reform Save Lives? Santa Monica, CA: RAND Corporation.

Reich, Michael, Sylvia Allegretto, and Anna Godoey. 2017. Seattle’s minimum wage experience 2015-16. Policy Brief. Center on Wage and Employment Dynamics, Institute for Research on Labor and Employment, UC Berkeley.

Williams, Noah. 2018. CROWE Policy Brief: Evidence on the Effects of Minnesota’s Minimum Wage Increases.

Tables and Figures

Table 1.

State	Weight
Alabama	0
Georgia	0
Idaho	0
Illinois	0.173
Indiana	0
Iowa	0.047
Kansas	0
Kentucky	0
Louisiana	0
Maine	0.337
Missouri	0.258
Nevada	0
New Hampshire	0
New Mexico	0
North Carolina	0
North Dakota	0
Oklahoma	0
Pennsylvania	0
South Carolina	0
Tennessee	0
Utah	0
Virginia	0
Wisconsin	0.185
Wyoming	0

Table 2.

Variables	Minnesota	Synthetic Minnesota
Avg. Weekly Wage	228	246
Establishment Count	3,665	3,627
Taxable Wages	1.74e+08	1.56e+08
Employment 2010Q1	59,274	59,151
Employment 2012Q1	61,869	61,831
Employment 2014Q1	64,973	65,069

Table 3.

State	Weight
Alabama	0
Georgia	0
Idaho	0.302
Illinois	0
Indiana	0
Iowa	0.347
Kansas	0
Kentucky	0
Louisiana	0
Maine	0
Missouri	0
Nevada	0
New Hampshire	0
New Mexico	0
North Carolina	0.352
North Dakota	0
Oklahoma	0
Pennsylvania	0
South Carolina	0
Tennessee	0
Utah	0
Virginia	0
Wisconsin	0
Wyoming	0

Table 4.

Variables	Minnesota	Synthetic Minnesota
Avg. Weekly Wage	228	228
Taxable Wages	1.74e+08	1.73e+08
Employment 2012Q1	61,869	61,923

Table 5.

Variables	
Minnesota	-554
	(874.7)
2014Q3	4,503***
	(770.8)
TreatEffect	285.8
	(1,160.2)
Constant	65,174***
	(576.2)
Observations	66
R-squared	0.50

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6.

Variables	
Minnesota	8.351e+06*
	(4.539e+06)
2014Q3	4.134e+07***
	(5.492e+06)
TreatEffect	1.818e+07**
	(8.812e+06)
Constant	1.831e+08***
	(3.010e+06)
Observations	66
R-squared	0.722

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 1.

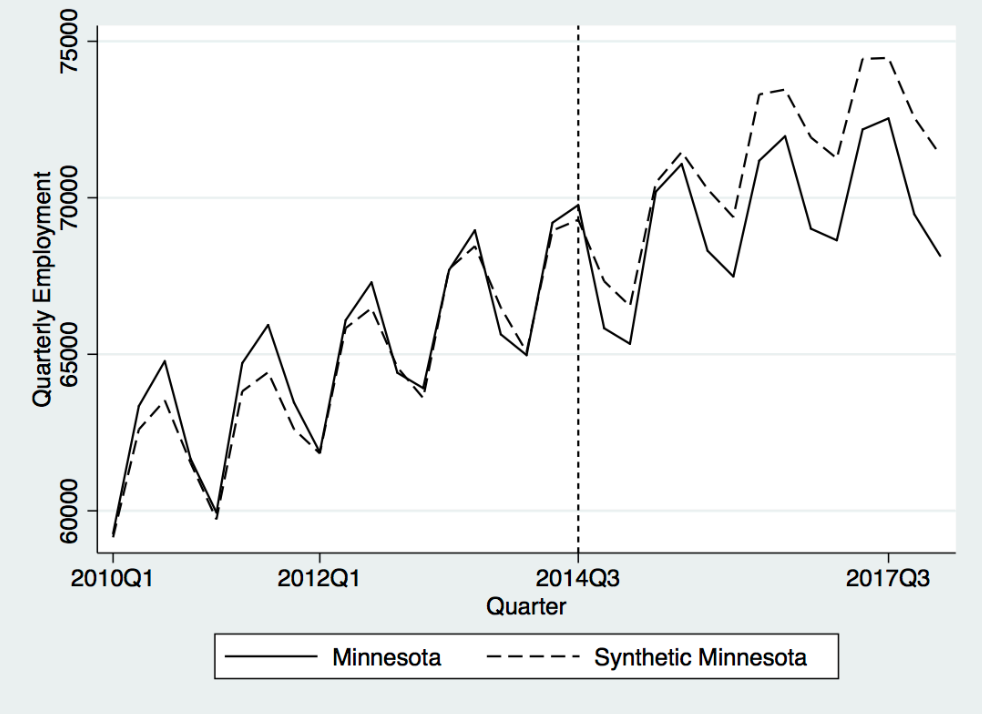


Figure 2.

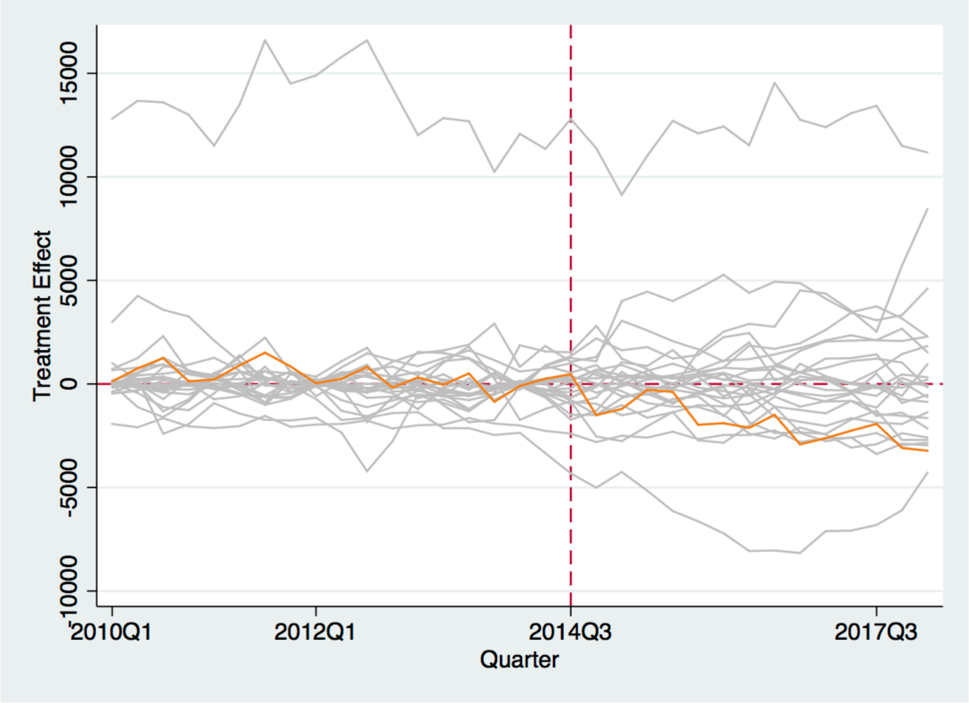


Figure 3.

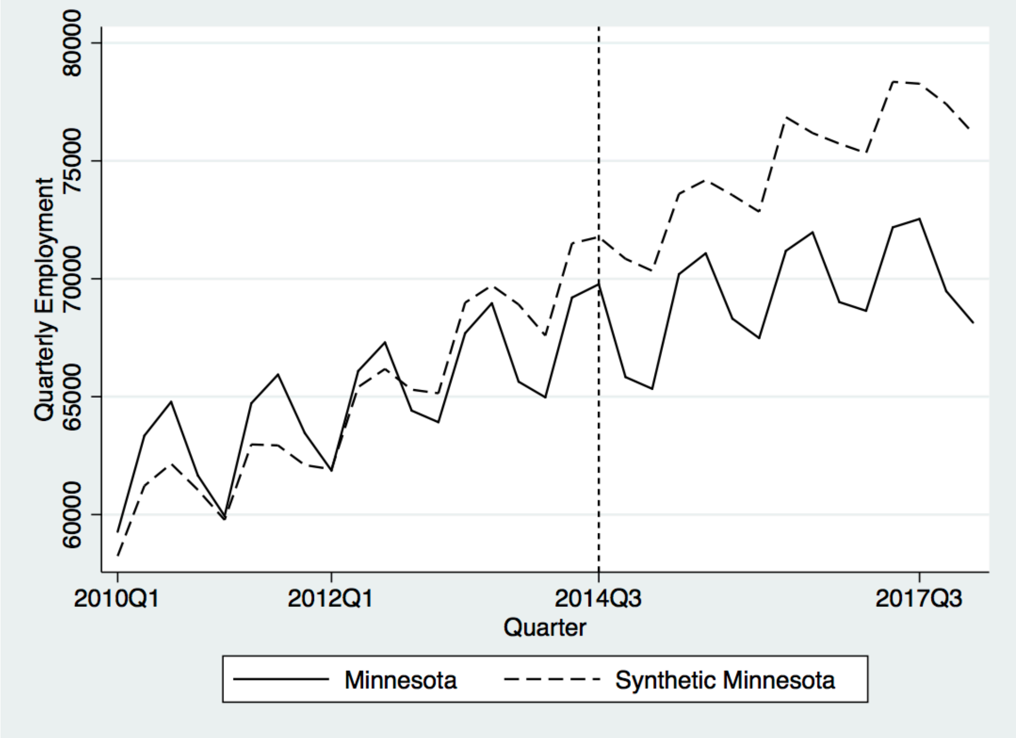


Figure 4.

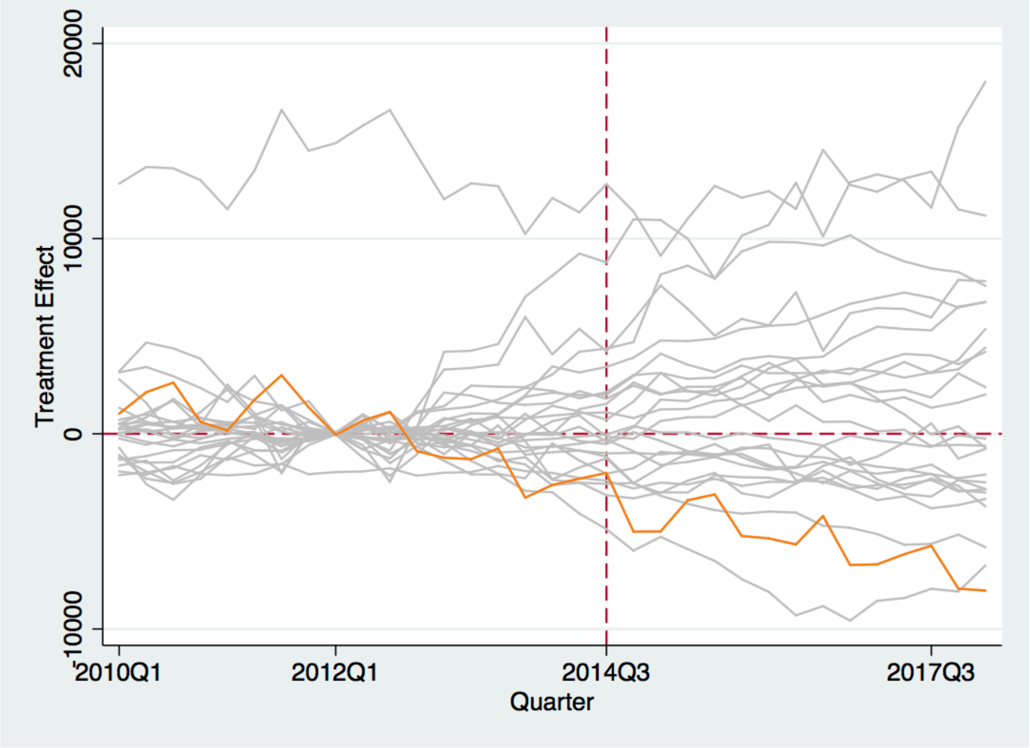


Figure 5.

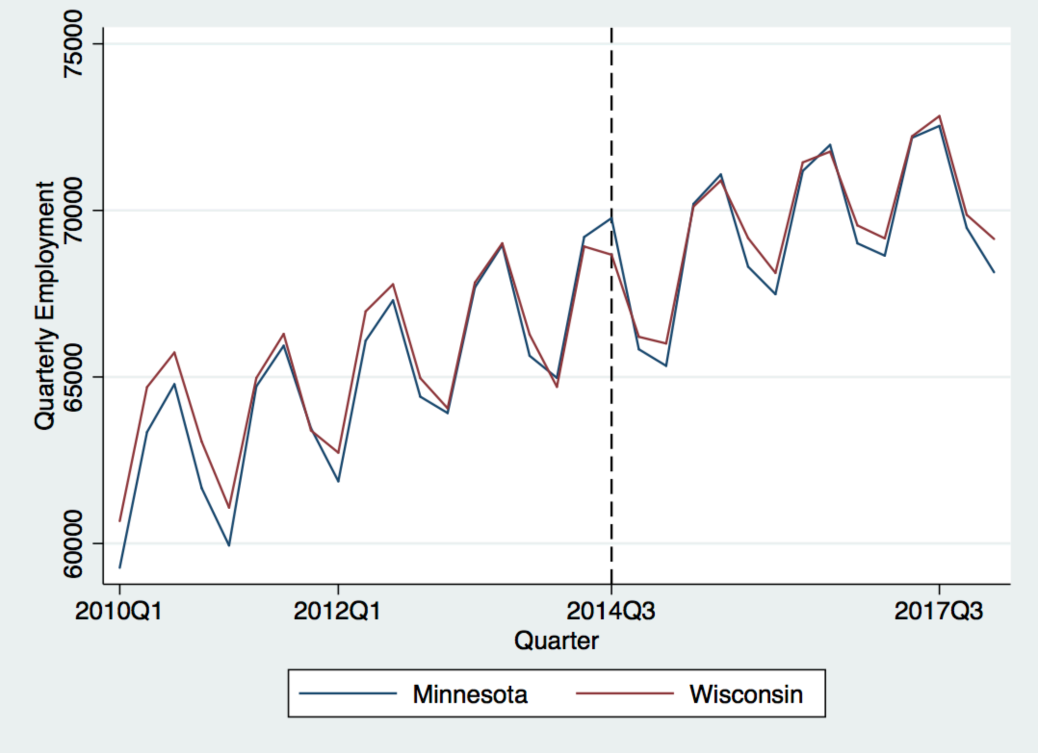
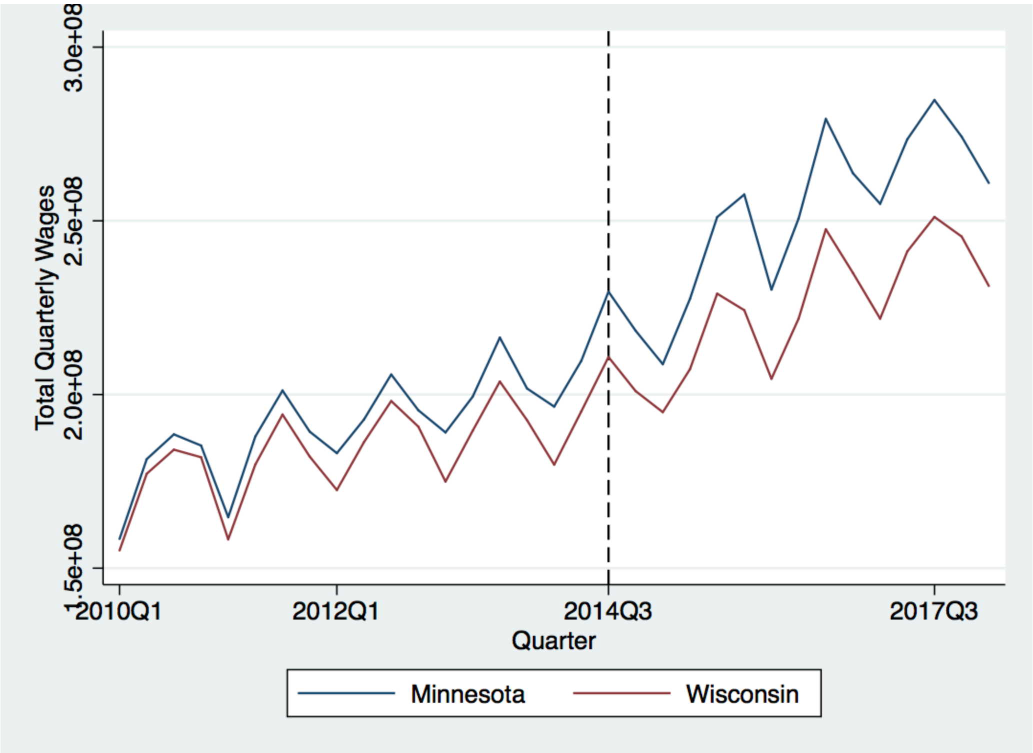


Figure 6.



Charitable Giving, Social Class, and Scrutiny: The Silent Poor and the Raucous Rich

Alex Ramiller and Maddy Hillson

Behavioral and Experimental Economics

Abstract

Charitable giving is a manifestation of prosocial behavior that may not always be motivated by pure altruism. Using a 3x2 experimental framework, this paper measures the effects of scrutiny and socioeconomic status on charitable giving. Notably, the paper finds that the level of charitable giving by wealthy individuals is significantly influenced by the visibility of their donations, suggesting that reputation-based altruism is a primary motivator for these donations.

I. Introduction

Charitable giving represents an important economic activity in the United States, with individuals, foundations, and corporations collectively donating $390 billion to charitable organizations in 2016. Individuals gave 71% of that number, making them the most highly charitable sector. Only 67% of households, however, gave anything in 2016, representing a key behavioral gap between households (The Philanthropy Roundtable, n.d.).

Researchers have long been interested in the motivations behind prosocial behavior, defined as behavior that benefits other people or society as a whole (Camerer and Thaler, 1995; Korndörfer, Egloff, and Schmukle, 2015; Kraus and Callaghan, 2016; List, 2007; Piff et al., 2010). In the realm of charitable giving, prosocial motivations are likely extremely varied – from genuinely wanting to positively impact the world to looking good to one's friends. A relatively recent dispute in the realm of charitable giving is about the income of charitable givers: are the rich more generous than the poor, or do poor people give more as a percentage of their income? Conflicting studies have emerged, finding that either the poor are more charitable (Piff et al., 2010), or that the rich are more charitable (Korndörfer, Egloff, and Schmukle, 2015). Others still have suggested that the real difference is motivation – that rich people are more motivated to give when their giving will be widely known (Kraus and Callaghan, 2016).

While much of the existing research pertains to interpersonal giving (Camerer and Thaler, 1995; Piff et al., 2010; Kraus and Callaghan, 2016), we are more interested in the factors that motivate charitable giving across socioeconomic statuses. Following the work of others, we hypothesize that affluent individuals are motivated to give more in public contexts, and that non-affluent individuals are motivated to give more in private contexts. We expect this to occur because affluent individuals are more reputation sensitive and will attempt to signal altruism through charitable giving in public contexts. We add a third context, researcher scrutiny, to ascertain whether studies that do not emphasize privacy are obtaining results more like the private or public contexts. This leads us to our research question: Does scrutiny affect charitable giving differently according to socioeconomic status?

II. Literature Review

Economic Theory

In order to address our research question, we must first turn to theories concerning the nature of human altruism. From an evolutionary perspective, there may seem to be no clear justification for altruism. Altruism entails an organism temporarily decreasing its own fitness in order to increase the fitness of another organism. Under the theory of natural selection, this action should reduce an organism's ability to survive and reproduce, which means that altruistic characteristics would eventually be eliminated from the population. The "kin selection" theory of altruism, however, which was first developed in the field of biology, refutes this idea and provides an evolutionary justification for altruism. Kin theory suggests that altruism *can* be evolutionarily beneficial if it means that those with altruistic genes share resources with their family members and relatives that also have altruistic genes. By supporting one another, those with altruistic genes therefore have greater fitness in the long-run (Okasha, 2013; Hamilton, 1964). While this may provide an evolutionary justification for the existence of altruism, however, altruistic behavior in human society has evolved beyond kin-driven altruism.

In human societies, the scope of altruistic behaviors has expanded to include not only those within an individual's immediate family but also with acquaintances and even strangers through a phenomenon known as "reciprocal altruism." Under reciprocal altruism, an individual chooses to act in a manner that decreases their own fitness while increasing the fitness of another, with the *expectation* that this act of kindness would later be reciprocated (Trivers, 1971). This type of altruism is beneficial both for individuals and for entire human societies, because it can allow for more complex social interactions and risk pooling. However, it also requires a high level of trust because it applies to interactions that are not between closely related individuals (Nowak and Sigmund, 2005). One motivation to engage in altruism in spite of this trust issue is the opportunity to gain reputational advantages from interacting with others, because doing so can make one appear more trustworthy and thereby increase one's fitness advantage. Reputation formation thus provides a clear incentive for an individual to engage in reciprocal altruism both because they can acquire future benefits both from the reciprocator and from others aware of their altruistic reputation (Fehr and Fischbacher, 2003).[10] Cultivating an altruistic reputation may also be desirable because it translates into higher social status (Hardy and van Vugt, 2006). Public donations may therefore be larger than private donations because they contribute to an altruistic reputation (Boehm, 2012; Flynn et al., 2006; Rand and Nowak, 2013; Sperber and Baumard, 2012). Meanwhile, those with higher socioeconomic status give may be more concerned with maintaining their social reputation because they have much more to lose in terms of status (Hart and Edelstein, 1992; Weininger and Lareau, 2009).[11] Therefore, both socioeconomic status and

---

[10] This theory is also presented in earlier works by Olson (1971) and Becker (1974).

[11] Another explanation for altruism comes from Andreoni (1990), who suggests that people derive utility from the act of giving. Moreover, altruism coefficients decline with income for all but the highest class, so that lower class and upper-class individuals give more than those in between. Andreoni notes that this could interact with scrutiny, because the mechanism by which people derive utility from altruism likely contains some component of publicity. Wealthier people might, therefore, simply derive more pleasure from giving publicly than poorer people.

the level of scrutiny that a donation is subject to will likely have an influence on the size of a charitable donation.

Previous Empirical Research

A number of empirical studies have attempted to explore the implications of the relationship between socioeconomic status and prosocial behavior. Piff et al. (2010) use four studies to show that individuals identifying with lower socioeconomic status tend to engage in more prosocial behaviors such as generosity, trust, and helpfulness than those identifying with a higher socioeconomic status. One of these studies focuses on the relationship between perceived socioeconomic class and charitable giving. This study first primed participants (81 undergraduate students) with questions about their social class relative to the United States population as a whole. Participants were then asked to complete a survey asking them how people should allocate their annual salaries, with charitable donations listed as one of the options. Their findings in this study confirm their overall hypothesis: those who perceived themselves to have a lower social class said on average that people should spend 4.65% of their income on charitable donations, while those who perceived themselves to be upper-class averaged 2.95%. While this study did not ask about each participant's actual charitable giving, it did explore the beliefs relating to charitable giving, which, presumably, are correlated.

The body of current evidence on the relationship between social class and prosocial behavior appears to confirm these findings, because individuals with higher socioeconomic status have greater feelings of control and power, which decreases their level of attention toward the welfare of others (Kraus et al., 2012; Piff et al., 2017). Therefore, individuals with higher socioeconomic status will tend to be less prosocial in their behavior unless doing so also benefits their own well-being through reputation enhancement. Conversely, lower-class individuals face greater daily uncertainty and consequently adopt a more "external, other-oriented focus," which means that they would theoretically be less affected by considerations of scrutiny (Piff et al., 2017, p. 6).[12]

This negative relationship between socioeconomic status and charitable giving is not consistent across the empirical literature, however. In fact, Korndörfer, Egloff, and Schmukle (2015) find exactly the opposite effect: using a number of longitudinal and panel data sets from institutional surveys in Germany and America, they find greater incidence of prosocial behaviors such as charitable giving among higher-class individuals. Korndörfer et al. argue that the reasoning behind a negative relationship between social class and prosocial behavior is problematic. They point out, for example, that previous examinations of charitable giving tend to ignore the large proportion of households that do *not* give to charity, instead focusing only on those that do. This suggests that finding negative relationship between charitable giving and social class is erroneous.

The differences in findings between Piff et al. (2010) and Korndörfer, Egloff, and Schmukle (2015) might be explained in part by context. In a paper comparing prosocial behavior in public and private settings, Kraus and Callaghan (2016) ran three experiments to ascertain which socioeconomic class engages in more prosocial behavior. In two of these experiments, respondents to an online survey participated in a "dictator game."[13] In a classic dictator game, one participant (the "dictator") is provided a certain amount of money and is allowed to determine how that money is distributed between them and another study participant who receives no money from the

---

[12] A large body of literature suggests an individual's socioeconomic status has a significant impact on their psychology and may influence their behavior towards others (i.e. Kraus et al., 2012; Kraus and Stephens, 2012)

[13] In the other study, Kraus and Callaghan analyzed Twitter posts signaling participation in in the "ALS ice bucket challenge" and found that Twitter users designated as "upper-class" (defined by income relative to national median income and educational attainment) tended to be more engaged in prosocial behavior in this public context.

researchers (List, 2007). In Kraus and Callaghan's study, survey participants were given the opportunity to share lottery tickets with a participant that would take the survey later. These participants were assigned to two conditions: a "public" condition in which the recipient of the gift would be told the name and home city of the donor, and a "private" condition in which the donor would remain anonymous to the recipient. Kraus and Callaghan found that lower-class people gave more than higher-class in a "private" context, whereas higher-class people gave more than lower-class people in a "public" context, thus corroborating the theory that socioeconomic status and reputation-driven altruism are closely associated. In particular, higher-class individuals may be more motivated by prestige and status compared to lower-class individuals.

Our study expands upon Kraus and Callaghan's original findings by experimentally examining the interaction of socioeconomic status and level of scrutiny in charitable giving. However, we depart from their approach in several key respects. Whereas Kraus and Callaghan use dictator games to explore prosociality in the form of giving money to another individual, our dictator game entails giving directly to a charitable organization. More fundamentally, whereas Kraus and Callaghan make a distinction between "public" and "private" giving, we explore *three* different types of scrutiny, making a distinction between giving to charity under scrutiny by a large group of people, under scrutiny only by the researchers conducting the experiment, and in a truly anonymous context. Therefore, our study makes an important contribution to the literature by differentiating the effect that researcher scrutiny in an experimental context has upon charitable giving.

Altruism and Researcher Scrutiny

The dictator game approach can be modified so that each participant is a "dictator" deciding how much money to keep and how much to donate for a charity. Carpenter, Connolly, and Myers (2008) provide a precedent for this approach by using a representative dictator game to measure the willingness of students and community members to give to a charity. Their experiment provides each participant with a hypothetical $100, and each was then asked how much of that money they would allocate to a charity of their choice. In order to give the decision real consequences, 10% of the participants were then randomly selected for whom their allocation decision would be enacted (both participant and designated charity would actually receive the amounts specified). Carpenter et al. test for the difference between students and community members and find that students are much less willing to give their endowments to charity. While this finding is interesting in its own right, the primary role the study serves in this context is to demonstrate that a dictator game involving charitable giving is viable.

The dictator game provides an optimal experimental strategy to examine our research question, as it has been established in previous research that researcher scrutiny can have a significant effect on individual actions in dictator games. Franzen and Pointner (2011), for example, suggest that the desire to appear fair to both other participants and to researchers will cause individuals to display other-regarding preferences in the context of dictator games. Participants will be more self-regarding, Frazen and Pointner suggest, if they are completely convinced that their decisions will remain completely anonymous. A "double-blind" method in which one researcher knows participant identities but not their donation decisions, and the other researcher knows donation decisions but not participant identities, has been proposed as an effective strategy to elicit true anonymity (including from researchers) in an experimental context

(Franzen and Pointner, 2011; Levitt and List, 2007). Therefore, the dictator game approach provides us with an effective strategy to address the existing gap in the literature concerning the differing effects that anonymity, researcher scrutiny, and public scrutiny have upon charitable giving.

III. Experimental Design

This study uses a 3x2 experimental design in order to explore the relationship between charitable giving, socioeconomic status, and the presence of scrutiny. For socioeconomic status, we use two groups representing higher and lower classes, termed "affluent" and "non-affluent." We interacted those two socioeconomic statuses with three treatment conditions that primed respondents for differing levels of scrutiny: complete anonymity, scrutiny by researchers, and scrutiny by the public. We used between-subject analysis, because each condition receives a unique message about the anonymity (or publicity) of their answers to the survey.

We used an online survey on the Qualtrics platform to collect data responses from participants. Links to the survey were shared within the online social networks of both researchers, through Facebook groups for Macalester College students, and through Facebook groups that one researcher had access to through her involvement in the Minnesota swing dance community. The survey consisted of three parts: demographic questions (some of which were intended to prime thoughts on socioeconomic status), descriptions of the experiment that primed participants for one of three possible conditions of a charitable donation game, and finally a charitable donation game where participants decided how much of a $20 endowment they will donate to charity. Realism was imposed by telling participants that there would be two random drawings for a $20 cash prize, where the cash would be distributed according to the winner's decision in the game.

Participants were randomly assigned to one of the three conditions for the charitable donation game. Participants in the anonymity condition were informed that a double-blind process was being used, such that nobody would ever know their donation decision and name. Participants in the researcher scrutiny condition were told that they would be contacted by the researchers if they won the random drawing, and that the researchers would remind them of their donation decision. Finally, participants in the public scrutiny condition were told that if they were to win the random drawing for the cash prize, their keep/donate decision may be shared with all other participants.

Subjects answered questions about their socioeconomic statuses. They were offered five options: “lower”, “lower-middle”, “middle”, “upper-middle”, and “upper”. We accepted participants’ assessments as fact. We also primed participants to think further about their socioeconomic status by asking if they considered their socioeconomic status as higher than, the same as, or lower than their peers. Finally, we asked if they had enough financial resources to live comfortably last year. This priming was intended to exaggerate any effects that socioeconomic status has on reactions to scrutiny and donation decisions.

Data

We collected 207 valid responses to our survey over a twelve-day period in April 2018. Given that our respondents were drawn largely from our own social networks, the demographics of the participant population reflect significant homogeneity. Nearly three quarters of participants (73%) identified themselves as current college students, which was reflected in the age distribution of the participant population. The average respondent age was 24.25, with a standard deviation of 9.21 years, a minimum age of 18, and a maximum age of 66. The age distribution of respondents was significantly right-skewed by the preponderance college students, with the vast majority of participants between the ages of 18 and 22 (see Figure 1). The gender distribution reflected the characteristics of the communities from which we solicited responses: 63% of our respondents identified as female, while 35% identified as male and 2% listed “other.” This statistic was almost identical for students and non-students and suggests that our participants almost exactly reflect the Macalester College population in terms of gender, as Macalester’s student body was 60% female as of 2018 (US News and World Report, 2018). Finally, the racial demographics of participants were also fairly homogeneous, with 72% of respondents identifying as white, a further 16% identifying as Asian, and only 2% identifying as black. The remaining 10% of respondents either listed their race as “other” or chose not to answer.

While our population is fairly homogeneous along these demographic dimensions due to the communities from which we solicited survey responses, we find that socioeconomic status is more normally distributed across participants. As noted previously, the survey gave respondents five options for their self-identified socioeconomic status: lower, lower-middle, middle, upper-middle, and upper. The “middle” category was the largest with 78 respondents, while “lower” and “upper” categories were the smallest with only 12 respondents in each group. Because we are primarily interested in the influence that higher socioeconomic status has on charitable giving, we group these five options into two primary categories: 1) the “non-affluent” category including those who responded “lower,” “lower-middle,” or “middle” and 2) the “affluent” category including those who responded with “upper-middle” or “upper.” Classified in this manner, there were 159 respondents in the non-affluent category and 46 in the affluent category (see Figure 2). Interestingly, there was no statistically significant difference in the size of charitable donation between these two groups, with non-affluent participants giving an average of $14.25 to charity while affluent participants gave an average of \$18.86.[14] In addition to this basic question about socioeconomic status, we also asked several other questions to ascertain each individual's perception of their economic well-being and prime them to think further about their socioeconomic status. Each respondent was asked, for example, if they believed that they had enough financial resources to live a comfortable life over the past year. The vast majority of participants (91%) said that they did have enough resources to live comfortably, while the remaining 9% said that they did not.

Participants were randomly assigned to one of the three treatment groups, with 66 respondents ultimately assigned to the anonymity condition, 78 to the researcher scrutiny condition, and 63 to the public scrutiny condition. Due to this random assignment, there should be no correlation between treatment assignment and any of the independent variables in our regression. Using balance tables to determine correlation between group assignment and independent variables, we can affirm this assumption for each of the three treatment groups (see figures 3-5). The only exception is a slight correlation between higher socioeconomic status and assignment to the public scrutiny group (see figure 6). However, this effect is relatively small and is only significant at the 10% level, which does not preclude the subsequent analysis.

While there did not appear to be systematic correlations between treatment group and our independent variable, there were substantial between-group differences in average donation. Those in the anonymous condition gave the least, contributing an average of only \$12.70 out of a possible \$20 to charity. The conditions with higher scrutiny yielded more generosity, with participants

---

[14] These average donations are much higher than those found in interpersonal dictator games (Kraus and Callaghan, 2016), but are similar in magnitude to those found by Carpenter, Connolly, and Myers (2008) in a charitable dictator game. It appears that participants are far more generous when giving to charity than when giving to other people.

giving an average of $14.60 in the researcher scrutiny condition and $15.14 in the public scrutiny condition. These averages suggest that people are, indeed, motivated by scrutiny. However, they do not reveal what effect, if any, the socioeconomic status of participants may have had on these results. In the following section, therefore, we explore the interaction between socioeconomic status and level of scrutiny in charitable giving decisions through the use of an OLS regression.

IV. Regression Results

We ran a linear regression on our data, using interaction terms between socioeconomic status and scrutiny condition to obtain our results. The regression we ran is as follows:

$$
\begin{aligned}
Donation_i = {} & \beta_0 + \beta_1\, Upperclass_i + \beta_2\, PublicCondition_i \\
& + \beta_3\, ResearcherScrutinyCondition_i + \beta_4\, Upperclass_i \text{ x } PublicCondition_i \\
& + \beta_5\, Upperclass_i \text{ x } ResearcherScrutinyCondition_i + \beta_5\, Age_i + \beta_6\, Student_i \\
& + \beta_7\, Male_i + \beta_8\, White_i + \beta_9\, ComfortableResources_i + \varepsilon_i
\end{aligned}
$$

Our variables of interest are β_1, β_4, and β_5, which are the effects from the interaction of condition and socioeconomic status. β_1 will show the difference in average donation between affluent and non-affluent participants, in the anonymous condition. β_4 and β_5 similarly show differences between affluent and non-affluent participants, but instead for the public scrutiny condition and the researcher scrutiny condition, respectively. We expect β_1 to be negative, such that affluent participants give less than non-affluent participants in the anonymous condition. We expect β_5 to be positive, such that affluent participants give more in the public condition. Finally, we expect β_4 to be somewhere between the two, since researcher scrutiny probably motivates participants less than public scrutiny but is likely still more motivating than anonymity.

Results from figure 7, column 1 show that our expectations were correct. In the anonymous condition, affluent participants gave on average $3.66 less than their non-affluent counterparts, significant at the 10% level. In the public condition, affluent participants gave $6.12 more than non-affluent participants, significant at the 5% level. Finally, in the researcher scrutiny condition, affluent participants gave $4.93 more than non-affluent participants, significant at the 10% level. A t-test reveals that there is no statistical difference between the affluent donations in public and researcher scrutiny conditions. Donation based on the average participant (a 25-year-old, white, female student) are displayed in figure 7. Our results show that non-affluent participants are unaffected by condition, but that affluent participants change their behavior when placed in a scrutiny condition (either public-scrutiny or researcher-scrutiny). These results are robust to alternate specifications: both a regression without controls (column 2) and a regression with a control for whether the participant had sufficient financial resources last year (column 3).

We find several other significant variables. Men are, on average, less generous than women by $1.60,[15] significant at the 10% level. Students give $4.25 less than those not in school, significant at the 1% level.[16] This may be because college is such a financial burden that it makes students thrifty. Meanwhile, those who feel that they had sufficient financial resources last year gave $4.14 more than those who did not, significant at the 5% level.

---

[15] This finding is corroborated by Korndörfer, Egloff, and Schmukle (2015), who find that women are more generous than men in prosocial games.

[16] This follows the findings of Carpenter, Connolly, and Myers (2008), who find that students are less generous than other community members.

We ran several other regressions with different groupings of socioeconomic statuses, to see which groupings were most revealing. We first tried a regression where, instead of grouping socioeconomic status into 2 levels, we used the 5 levels from our survey. As shown in figure 9, this specification gives us weak coefficients on all socioeconomic variables. The only consistently significant variable (at the 10% level) is the interaction of socioeconomic status and researcher scrutiny condition, which is positive - for every step up in socioeconomic status, participants in the researcher scrutiny condition gave $1.94 more. All other variables have the same sign as our primary specification, but lack significance. This shows that our findings are robust to alternate definitions of socioeconomic classes, giving us more certainty that our findings are valid.

Our second check of socioeconomic status groupings was to group “upper” and “upper-middle” class, as well as “lower” and “lower-middle” class (calling these wealthy and poor), and have middle class be a stand-alone class. We code these as independent dummy variables, to avoid enforcing a relationship between wealthy and middle-class participants’ behavior that may not otherwise exist. In figure 10, we show that middle class participants act in similar ways to poor participants, as shown by the coefficient on middle class being close to 0 and statistically insignificant. This justifies our original groupings of middle-class with the lower classes. This specification also finds that, in the researcher scrutiny condition, wealthy participants gave $4.86 more than lower class participants, significant at the 10% level. All other coefficients with socioeconomic status were insignificant, but of the same sign as our primary specification, once again reaffirming our confidence in our primary findings.

These results support the findings of Kraus and Callaghan (2016), that affluent participants give more in public contexts than in anonymous contexts. We find that in private contexts, affluent participants give significantly less than non-affluent participants. Conversely, in both researcher scrutiny and public scrutiny conditions, affluent participants give significantly more. We find a much stronger effect than Kraus and Callaghan, on the order of 8 times their magnitude. We assume this is due to differences in how we apply public scrutiny, and the type of prosociality being studied. Kraus and Callaghan study interpersonal giving (rather than charitable giving) and apply public scrutiny by telling participants that the recipient of their gift will know their name and home city. In contrast, we apply public scrutiny by telling participants that their donation amount and name may be shared with every participant, which would provide a much stronger incentive to protect one's reputation and motivate higher levels of giving. Our findings are predicted by altruism theory, which predicts that affluent people will respond more strongly to public context, due to the need to improve or maintain their reputation.

V. Conclusions

These results have fascinating implications for the study of charitable giving and social class, because they imply that charitable giving among wealthier individuals is influenced by the visibility of their generosity. Participants in our survey that identified themselves as being from the upper-middle or upper classes gave less than non-affluent participants when their donations were anonymous, which would seem to confirm the inverse relationship between socioeconomic status and prosocial behavior proposed by Piff et al. (2010; 2017). However, when those affluent participants were told that their donation decision would be acknowledged, either by the researchers or the entire community of survey participants, their behavior changed drastically. In fact, affluent participants became *more* generous relative to non-affluent participants, suggesting that they valued the reputational benefit of generosity much more highly.

There are several general conclusions that can be drawn from these findings. First, elucidating the relationship between socioeconomic status and scrutiny may explain the seemingly opposing findings in the literature about whether those of lower or higher socioeconomic status engage in greater prosocial behavior. In addition, this has substantial implications for future experimental research into charitable giving and prosociality in general, as it suggests that researcher scrutiny is statistically indifferentiable from public scrutiny in eliciting donations from affluent participants.[17] In dictator games run in-person (as opposed to online) or without clear enough statements of anonymity, experiments are likely picking up scrutiny effects similar to applying public scrutiny to participants. When socioeconomic status is not being studied, this effect would generate higher overall levels of giving. If socioeconomic status is being studied independently, this effect would make affluent people look more generous overall, like in the findings of Korndörfer, Egloff, and Schmukle (2015). To mitigate the differential effects of scrutiny, researchers should ensure that they apply no accidental scrutiny to participants. Therefore, these findings suggest that researchers should be prepared to take both of those factors into account in the future, by either mitigating bias in their experimental design, or by acknowledging bias in their analysis.

Limitations and Future Research

The significance of these findings suggests that further research should be conducted on the subject of socioeconomic status, scrutiny, and charitable giving. Replicating this study with a much larger

---

[17] This further emphasizes the point made by Levitt and List (2007) that laboratory experiments are likely to overestimate prosociality, due to the systematic differences between laboratory settings and real-world settings, especially in terms of scrutiny.

sample size and a more heterogeneous population would yield results that are both more balanced and more robust. The fact that the sample pool was largely composed of college students poses challenges for the generalizability of these results, for example, because college students do not necessarily possess the same economic independence from their families as working adults. Some college student participants may have thought primarily about their family's socioeconomic status as a proxy for their own when answering the survey. In addition, because we relied heavily on personal social networks to distribute this survey, it is likely that a number of our friends participated. If those friends were placed in the researcher scrutiny or public scrutiny conditions, the thought that we would be aware of their donation decision may have had a greater influence on them. Finally, it is important to reiterate that the treatment groups in this study were not perfectly balanced, as there was a slight preponderance towards higher socioeconomic status in the public scrutiny treatment group. While we do not believe that this affects the validity of our results, it does suggest that this study could benefit from a larger sample size. Overall, these findings open up a promising new direction in the study of prosociality, and we believe that the interaction between socioeconomic status and scrutiny is worth exploring further.

Acknowledgements

We would like to thank Pete Ferderer for helping us hone our question and experimental design, and for contributing financially to encourage participation. We would also like to thank Amy Damon for answering questions she didn't have to, and for being just a generally baller prof.

References

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, *100*(401), 464–477. doi.org/10.2307/2234133.

Becker, G. S. (1974). A Theory of Social Interactions. *Journal of Political Economy*, *82*(6), 1063–1093.

Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame* (1 edition). New York: Basic Books.

Camerer, C. F., and Thaler, R. H. (1995). Anomalies: Ultimatums, Dictators and Manners. *Journal of Economic Perspectives*, *9*(2), 209–219. doi.org/10.1257/jep.9.2.209.

Carpenter, J., Connolly, C., and Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics*, *11*(3), 282–298. doi.org/10.1007/s10683-007-9193-x.

Flynn, F. J., Reagans, R. E., Amanatullah, E. T., and Ames, D. R. (2006). Helping one's way to the top: self-monitors achieve status by helping others and knowing who helps whom. *Journal of Personality and Social Psychology*, *91*(6), 1123–1137. doi.org/10.1037/0022-3514.91.6.1123.

Franzen, A., and Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior and Organization*, *81*(1), 74–81. doi.org/10.1016/ j.jebo.2011.09.005.

Hamilton, W. D. (1964). The genetical evolution of social behavior. II. *Journal of Theoretical Biology*, *7*(1), 17–52.

Hardy, C. L., and Van Vugt, M. (2006). Nice Guys Finish First: The Competitive Altruism Hypothesis. *Personality and Social Psychology Bulletin*, *32*(10), 1402–1413. doi.org/10.1177/0146167206291006.

Hart, D., and Edelstein, W. (1992). The Relationship of Self-Understanding in Childhood to Social Class, Community Type, and Teacher-Rated Intellectual and Social Competence. *Journal of Cross-Cultural Psychology*, *23*(3), 353–365. doi.org/10.1177/0022022192233006.

Korndörfer, M., Egloff, B., and Schmukle, S. C. (2015). A Large-Scale Test of the Effect of Social Class on Prosocial Behavior. *PLOS ONE*, *10*(7), e0133193. doi.org/10.1371/journal.pone.0133193.

Kraus, M. W., and Callaghan, B. (2016). Social Class and Prosocial Behavior: The Moderating Role of Public Versus Private Contexts. *Social Psychological and Personality Science*, *7*(8), 769–777. doi.org/10.1177/1948550616659120.

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., and Keltner, D. (2012). Social class, solipsism, and contextualism: How the rich are different from the poor. *Psychological Review*, *119*(3), 546–572. doi.org/10.1037/a0028756.

Kraus, M. W., and Stephens, N. M. (2012). A Road Map for an Emerging Psychology of Social Class. *Social and Personality Psychology Compass*, *6*(9), 642–656. doi.org/10.1111/ j.1751-9004.2012.00453.x.

Levitt, S. D., and List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, *21*(2), 153–174. doi.org/10.1257/jep.21.2.153.

List, J. A. (2007). On the Interpretation of Giving in Dictator Games. *Journal of Political Economy*, *115*, 482–493.

Nowak, M. A., and Sigmund, K. (n.d.). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298.

Okasha, S. (2013). Biological Altruism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013). Metaphysics Research Lab, Stanford University. Retrieved from plato.stanford.edu/archives/fall2013/entries/altruism-biological.

Olson, M. (1971). *The Logic of Collective Action: Public Goods and the Theory of Groups, Second printing with new preface and appendix* (Revised edition). Cambridge, Mass.: Harvard University Press.

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., and Keltner, D. (2010). Having less, giving more: the influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784. doi.org/10.1037/a0020092.

Piff, P. K., and Robinson, A. R. (2017). Social class and prosocial behavior: current evidence, caveats, and questions. *Current Opinion in Psychology*, *18*, 6–10. doi.org/10.1016/j.copsyc.2017.06.003.

Rand, D. G., and Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. doi.org/10.1016/j.tics.2013.06.003.

Sperber, D., and Baumard, N. (2012). Moral Reputation: An Evolutionary and Cognitive Perspective. *Mind and Language*, *27*(5), 495–518. doi.org/10.1111/mila.12000.

The Philanthropy Roundtable. (n.d.). Statistics on U.S. Generosity. Retrieved May 2, 2018, from www.philanthropyroundtable.org/almanac/statistics.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

US News and World Report. (2018). How Does Macalester College Rank Among America's Best Colleges? Retrieved May 1, 2018, from www.usnews.com/best-colleges/macalester-college-2358.

Weininger, E., and Lareau, A. (2009). Paradoxical Pathways: An Ethnographic Extension of Kohn's Findings on Class and Childrearing. *Journal of Marriage and Family*, *71*(3), 680–695. doi.org/10.1111/j.1741-3737.2009.00626.x.

Appendix

Figure 1: Age distribution of survey participants.

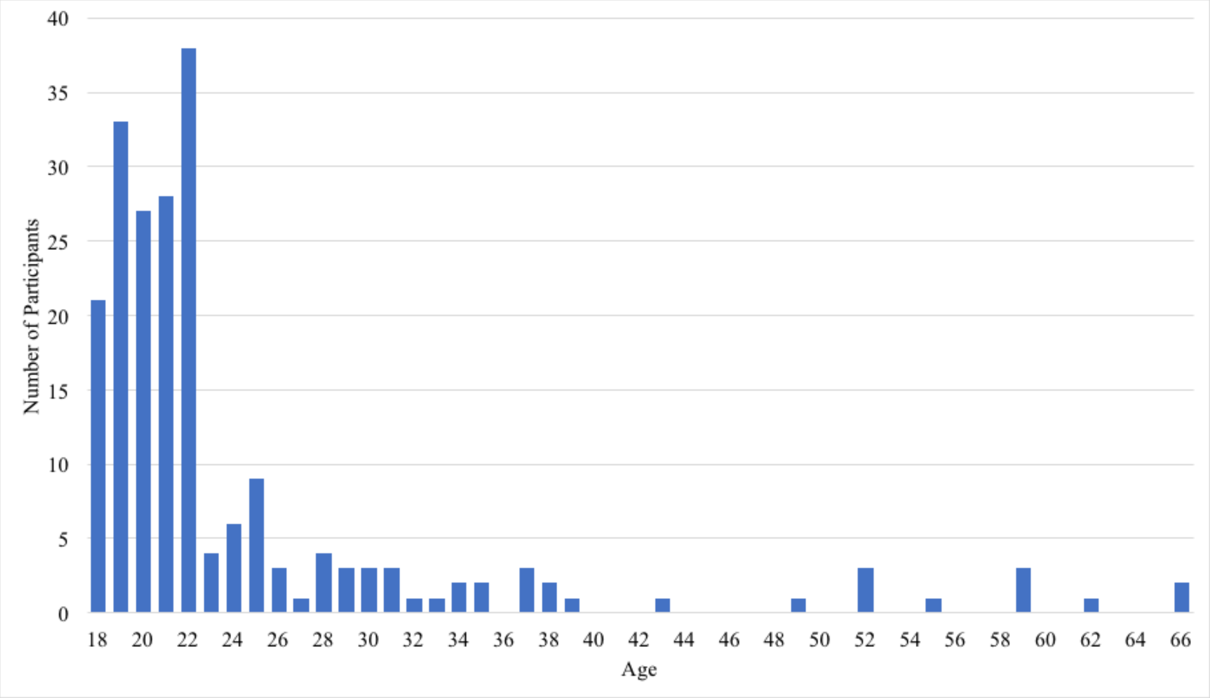


Figure 2: Distribution of survey participants by stated socioeconomic status.

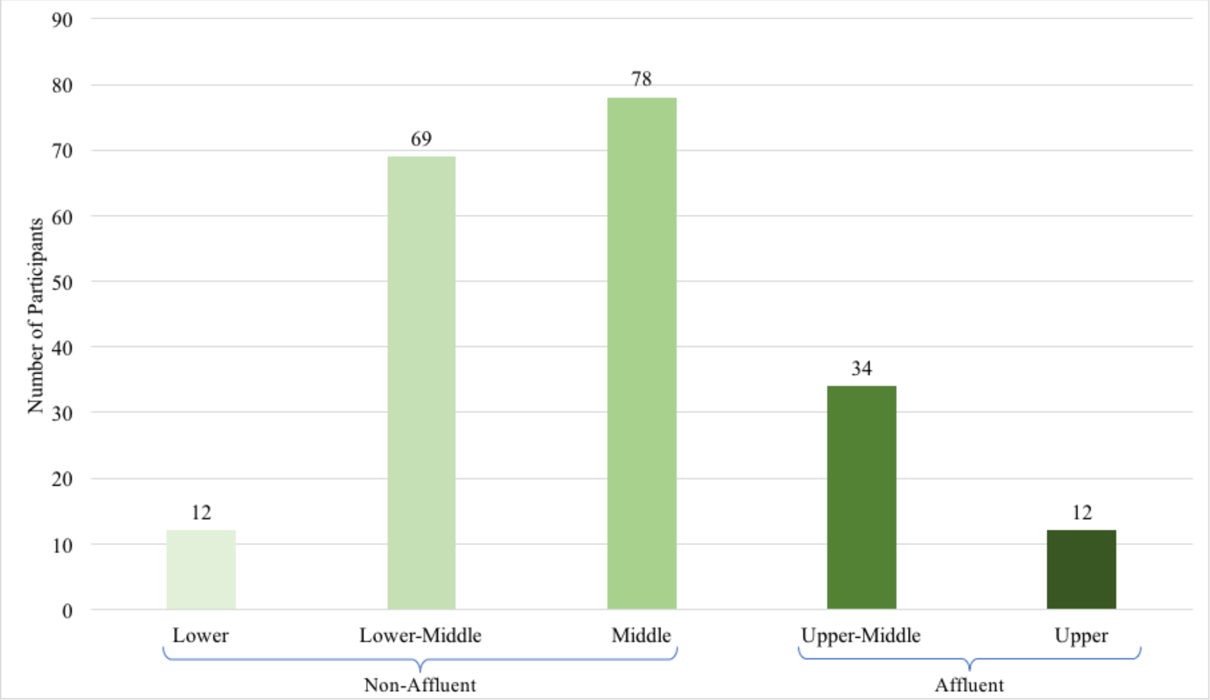
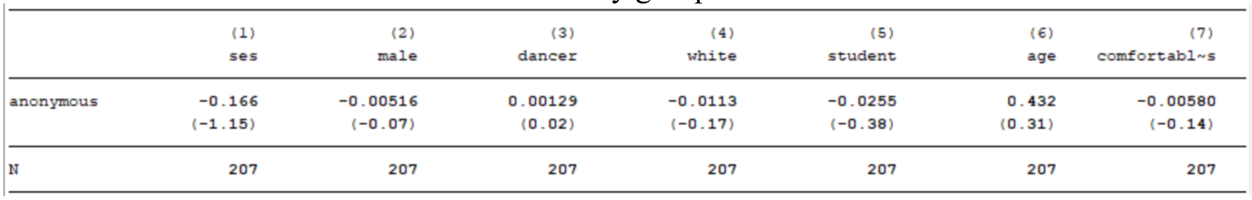
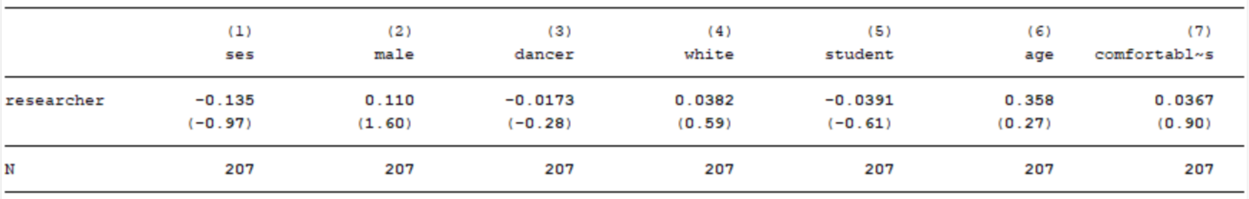


Figure 3: Balance table, comparing the anonymous group to the combined public and researcher-scrutiny groups.



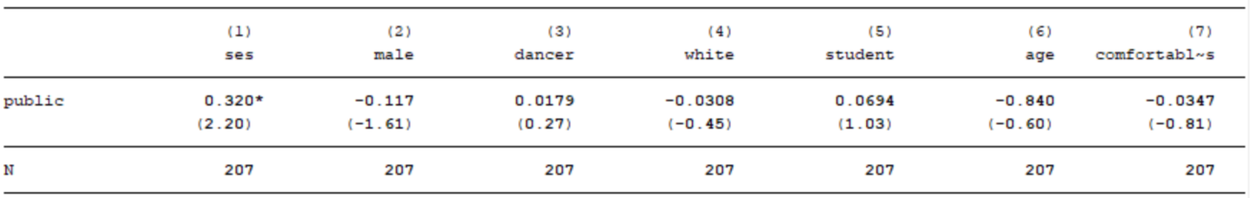
	(1) ses	(2) male	(3) dancer	(4) white	(5) student	(6) age	(7) comfortabl~s
anonymous	-0.166	-0.00516	0.00129	-0.0113	-0.0255	0.432	-0.00580
	(-1.15)	(-0.07)	(0.02)	(-0.17)	(-0.38)	(0.31)	(-0.14)
N	207	207	207	207	207	207	207

Figure 4: Balance table, comparing the researcher-scrutiny group to the combined public and anonymous groups.



	(1) ses	(2) male	(3) dancer	(4) white	(5) student	(6) age	(7) comfortabl~s
researcher	-0.135	0.110	-0.0173	0.0382	-0.0391	0.358	0.0367
	(-0.97)	(1.60)	(-0.28)	(0.59)	(-0.61)	(0.27)	(0.90)
N	207	207	207	207	207	207	207

Figure 5: Balance table, comparing the public group to the combined anonymous and researcher-scrutiny groups.



	(1) ses	(2) male	(3) dancer	(4) white	(5) student	(6) age	(7) comfortabl~s
public	0.320*	-0.117	0.0179	-0.0308	0.0694	-0.840	-0.0347
	(2.20)	(-1.61)	(0.27)	(-0.45)	(1.03)	(-0.60)	(-0.81)
N	207	207	207	207	207	207	207

Note: This balance table shows that the socioeconomic status of the public group is slightly higher than the socioeconomic status of the other groups, significant at the 10% level.

Figure 6: Distribution of survey participants by socioeconomic status in each treatment group, demonstrating a slight bias toward more affluent individuals in the public scrutiny treatment.

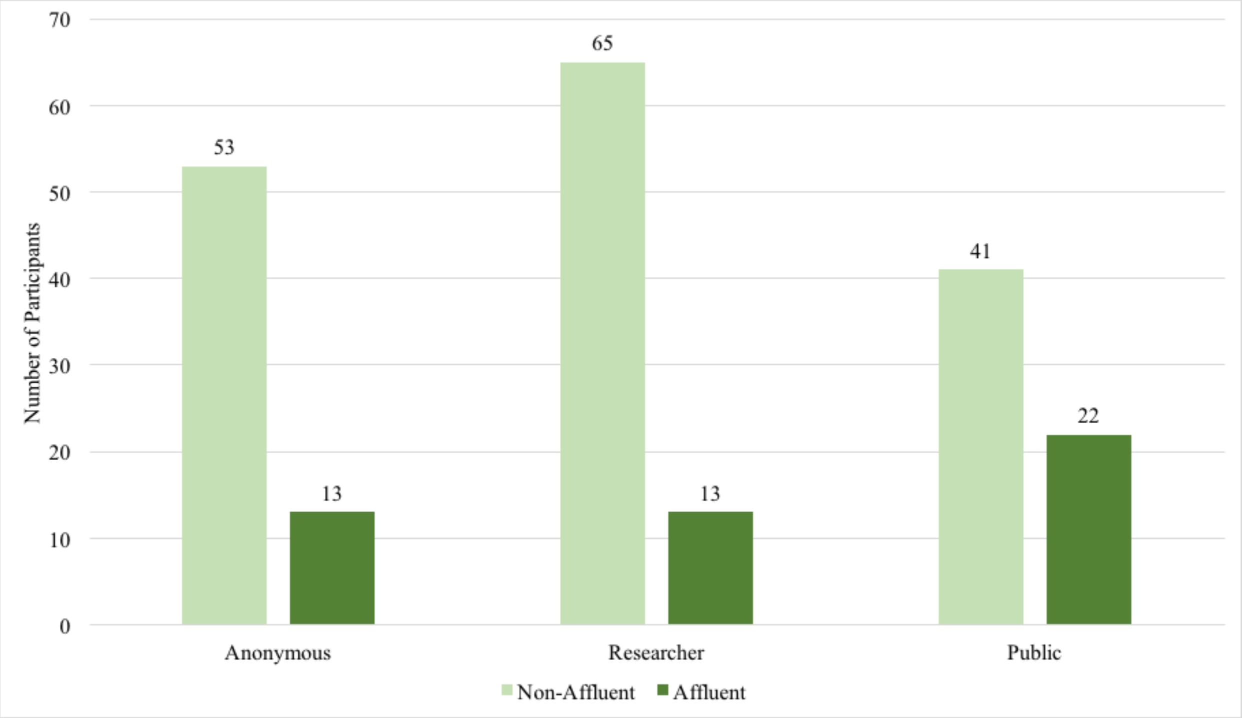
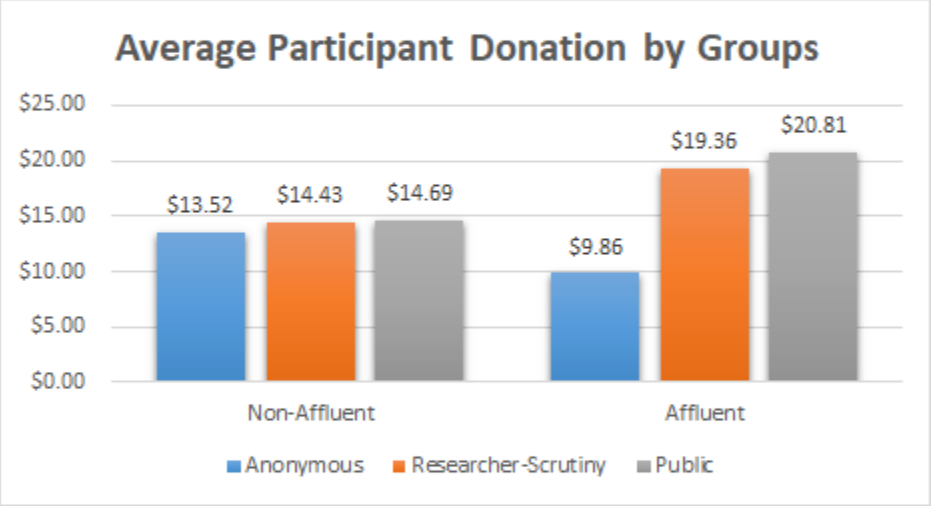


Figure 7: Regression output, primary specification, affluent versus non-affluent participants.

VARIABLES	(1) Primary Specification	(2) No Controls	(3) Resource Control
affluent	-3.659*	-4.795**	-4.132**
	(1.955)	(2.051)	(2.056)
public	1.175	1.017	0.802
	(1.283)	(1.378)	(1.370)
researcher	0.910	0.835	0.678
	(1.136)	(1.226)	(1.218)
affluent_public	6.122**	6.182**	6.731**
	(2.511)	(2.697)	(2.685)
affluent_researcher	4.926*	5.511*	5.961**
	(2.678)	(2.874)	(2.856)
age	0.0629		
	(0.0660)		
student	-4.250***		
	(1.359)		
male	-1.598*		
	(0.904)		
white	1.172		
	(0.964)		
Comfortable_resources	4.135**		3.807**
	(1.666)		(1.777)
Constant	10.89***	13.64***	10.05***
	(3.012)	(0.910)	(1.903)
Observations	207	207	207
R-squared	0.212	0.053	0.074

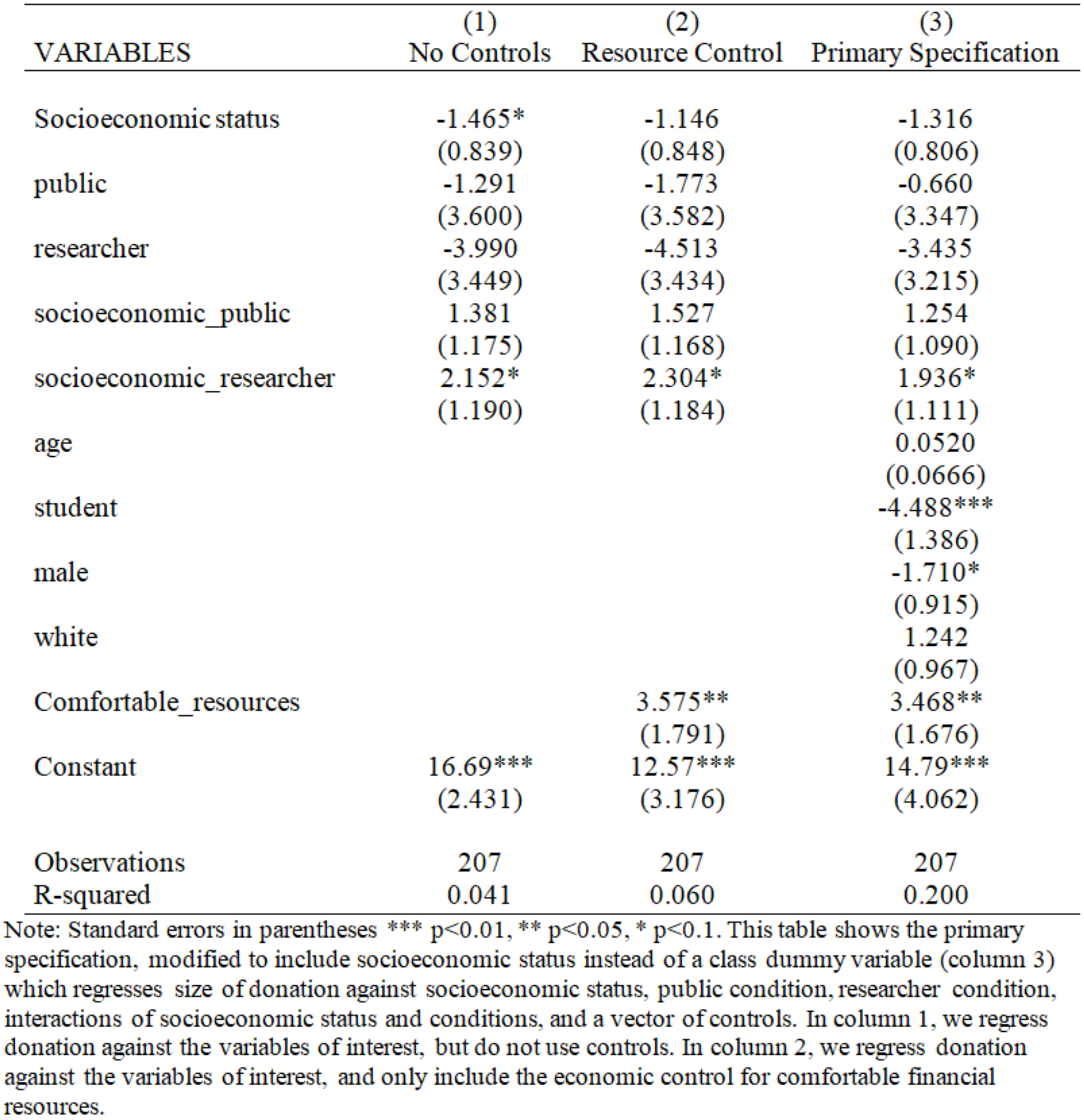
Note: Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$. This table shows the primary specification (column 1) which regresses size of donation against upper class, public condition, researcher condition, interactions of upper class and conditions, and a vector of controls. In column 2, we regress donation against the variables of interest, but do not use controls. In column 3, we regress donation against the variables of interest, and only include the economic control for comfortable financial resources.

Figure 8: Average donation by affluence and condition, predicted by primary specification.



Note: These are calculated by the average participant, a 25-year-old, white, female student.

Figure 9: Regression output, regressing all 5 levels of socioeconomic status.



VARIABLES	(1) No Controls	(2) Resource Control	(3) Primary Specification
Socioeconomic status	-1.465*	-1.146	-1.316
	(0.839)	(0.848)	(0.806)
public	-1.291	-1.773	-0.660
	(3.600)	(3.582)	(3.347)
researcher	-3.990	-4.513	-3.435
	(3.449)	(3.434)	(3.215)
socioeconomic_public	1.381	1.527	1.254
	(1.175)	(1.168)	(1.090)
socioeconomic_researcher	2.152*	2.304*	1.936*
	(1.190)	(1.184)	(1.111)
age			0.0520
			(0.0666)
student			-4.488***
			(1.386)
male			-1.710*
			(0.915)
white			1.242
			(0.967)
Comfortable_resources		3.575**	3.468**
		(1.791)	(1.676)
Constant	16.69***	12.57***	14.79***
	(2.431)	(3.176)	(4.062)
Observations	207	207	207
R-squared	0.041	0.060	0.200

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. This table shows the primary specification, modified to include socioeconomic status instead of a class dummy variable (column 3) which regresses size of donation against socioeconomic status, public condition, researcher condition, interactions of socioeconomic status and conditions, and a vector of controls. In column 1, we regress donation against the variables of interest, but do not use controls. In column 2, we regress donation against the variables of interest, and only include the economic control for comfortable financial resources.

Figure 10: Regression output, regressing three levels of socioeconomic status using dummy variables for wealthy and middle class.

VARIABLES	(1) No Control	(2) Econ Control	(3) Identity Control
middleclass	0.924	1.473	0.544
	(1.852)	(1.850)	(1.741)
wealthy	-4.412**	-3.482	-3.457
	(2.195)	(2.212)	(2.097)
public	2.642	2.642	3.274*
	(1.906)	(1.886)	(1.756)
researcher	0.742	0.742	0.978
	(1.702)	(1.684)	(1.570)
wealthy_public	4.557	4.910	4.035
	(3.006)	(2.979)	(2.774)
wealthy_researcher	5.604*	5.914*	4.856*
	(3.112)	(3.084)	(2.880)
middleclass_public	-3.348	-3.897	-4.202
	(2.782)	(2.765)	(2.572)
middleclass_researcher	-0.0380	-0.472	-0.243
	(2.483)	(2.465)	(2.305)
age			0.0566
			(0.0665)
student			-4.471***
			(1.383)
male			-1.732*
			(0.905)
white			1.214
			(0.965)
Comfortable_resources		4.031**	4.217**
		(1.796)	(1.676)
Constant	13.26***	9.228***	10.92***
	(1.193)	(2.150)	(3.250)
Observations	207	207	207
R-squared	0.062	0.085	0.227

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. This table shows the primary specification (column 3), modified to include two dummies for socioeconomic statuses – upper class and middle class - which regresses size of donation against socioeconomic status dummies, public condition, researcher condition, interactions of socioeconomic status dummies and conditions, and a vector of controls. In column 1, we regress donation against the variables of interest, but do not use controls. In column 2, we regress donation against the variables of interest, and only include the economic control for comfortable financial resources.

Medicaid Expansion: Does an Increase in Health Insurance Coverage Lead to Improved Health Care Outcomes?

Elliot Cassutt, Khadidja Ngom, and Duy (Jensen Vu)

Economics of Public Policy

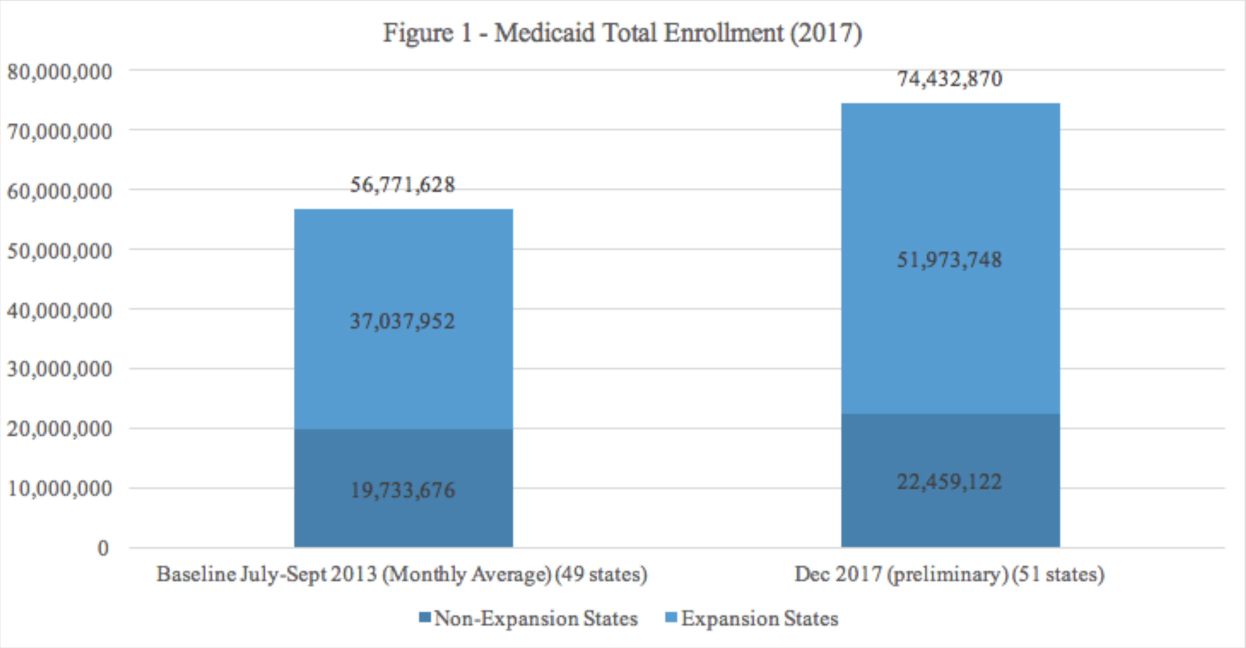
Abstract

One of the most widely debated provisions of the Affordable Care Act is the expansion of health coverage to low-income families through the Medicaid program. Despite its hopeful premise, the benefits of Medicaid expansion have been called into question, leading to 13 states not having expanded Medicaid by the end of 19. Our paper studies the effects of Medicaid expansion, based on three criteria: cost effectiveness, overall health outcomes, and the redistributional effects. We compile, synthesize, and critically evaluate the empirical methodologies of past economic research on the issue to form our conclusions and suggestions for future research.

I. Issue

Along with Medicare, Medicaid has been the largest source of healthcare funding in the U.S. since signed into law in 1965. A voluntary, federally-funded insurance program, Medicaid aims to provide low-cost medical and hospital care for all Americans who cannot afford such care. In 1982, Arizona became the last state to adopt Medicaid, making the program universally available in the U.S. for those eligible under its provisions.

At the end of 2013, about 57 million individuals were enrolled in Medicaid. Enrollment increased significantly the following year with the implementation of the Affordable Care Act (ACA). With the ACA, the Obama administration attempted to relax the eligibility requirement for Medicaid in order to expand its coverage. By 2017, 33 states have adopted the Medicaid expansion provisions, raising Medicaid enrollment by approximately 18 million, the majority of which come from expansion states.



Expanded coverage comes at a cost to the federal budget. In the 2016 fiscal year, spending for Medicaid totaled $574.2 billion, compared to the total amount of $460 billion in 2013 before the expansion. As with any other public redistributive program that accounts for such a large amount of public expenditure, Medicaid is a controversial topic among policymakers. In 2017, several pieces of legislations were proposed to roll back on the ACA provisions on Medicaid. Concerns about the program range from whether a federal program is more efficient in distributing healthcare resources than private solutions to whether Medicaid expansion actually leads to positive health outcomes for the population. These questions are empirical in nature, opening opportunities for future research.

This policy brief provides a summary of several studies conducted on the "Oregon experiment," a random selection of new households eligible to apply for Medicaid following an increase in funding for the state program. Using these past studies, the brief offers an insight into the empirical question of the effects of Medicaid expansion, which has strong implications for policy-making in public healthcare funding.

II. Evaluation Criteria

We evaluate Medicaid expansion based on three criteria: cost effectiveness, overall health outcomes, and the redistributional effects. The cost effectiveness of the program is measured by the relative effects on different forms of healthcare utilization. In theory, Medicaid coverage should decrease the number of emergency room visits because individuals will have access to primary and preventative care. Therefore, if we see a decrease in emergency room visits and an increase in preventative care, then we can conclude that Medicaid led to a more cost-effective outcome in the healthcare market. The second criterion, health outcomes, includes both physical and mental health measures such as blood pressure, cholesterol, and rates of depression. As long as those insured under Medicaid were not previously insured or had access to comparable public healthcare, we expect to see a positive effect on health outcomes. Lastly, redistributional effects will be measured by the prevalence of out-of-pocket catastrophic expenditures and the probabilities of individuals borrowing money to pay for medical expenses or of having bills sent to collection agencies. If Medicaid provides financial security to the newly insured, then we should observe a decrease in the types of financial strains listed above.

III. Isolating the Effects of Medicaid Expansion

To study the effect of the expansion of Medicaid on healthcare utilization, financial hardship and overall health outcomes, a large part of the literature compares states that have agreed to adopt Medicaid expansion to states that have not. The first group of expansionary states belong to the treatment group and the latter of non-expansionary states belong to the control group. This poses a few estimation challenges because, ideally, treatment and control groups should be identical except in their adoption of Medicaid expansion. However, in reality, expansionary and non-expansionary states may systematically differ on factors, both observable and unobservable, that also affect the outcomes of interest, a problem referred to as selection bias.

It is against this background that Oregon initiated in 2008 a limited expansion of its Medicaid program through lottery drawing of approximately 30,000 names from a waiting list of almost 90,000 persons. This lottery design created a unique opportunity to conduct a randomized control trial (RCT) to study the effects of Medicaid, allowing researchers to minimize the problem of selection bias and to isolate causal inference between insurance coverage and the outcomes of interest. The Oregon experiment is a series of studies that evaluate the impact of Medicaid expansion through this lottery distribution. Researchers study the outcomes of Medicaid on those selected by the lottery with those who were not selected as their control group.

Because of these differences in methodology, when summarizing findings, we systematically compare and contrast earlier studies of Medicaid expansion to the Oregon experiment as a benchmark. It is important to note that while the randomized nature of the Oregon experiment controls for many confounding variables, it does not eliminate all possible biases, and limits our ability to extrapolate the results to other contexts. First, the studies are not able to account for each individual's access to care prior to the enrollment period. Some enrollees may have already had access to health insurance or low-cost care through another public program, in which case the program's effect on overall health may be understated. Secondly, the randomization of the experiment mitigates the effects of adverse selection by removing systematic differences from the treatment and control groups. Because adverse selection persists in situations where Medicaid eligibility is not done by a lottery, these results cannot be generalized to other scenarios. Lastly, many of the studies that examined the experiment took place only one or two years following the rollout, which may have been too short of a time period to see the full effects of the program. Future research with broader time frames and larger datasets would allow for a more comprehensive analysis of the program's impact.

IV. Research Findings

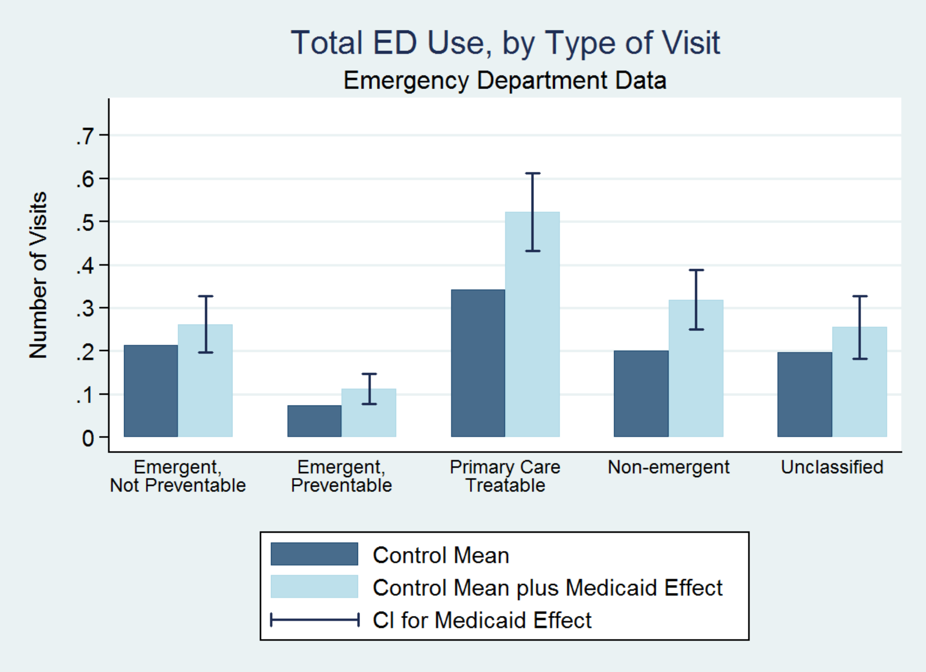
Effects on Healthcare Utilization

The first body of findings study the effects of expanding Medicaid on healthcare utilization. Knowing these effects would allow us to evaluate the program against our first evaluation criteria, cost effectiveness. Indeed, most empirical papers focus on the effects of insurance coverage on emergency department visits, hospital admissions, outpatient care as well as diagnosis and prescription rates while measuring the change in healthcare utilization. WeIt is our theory that a cost-effective expansion would allow people better access to primary and preventative care, thus reducing the need for emergency room visits. If Medicaid expansion leads a decrease in emergency room visits and an increase in preventative care, then we can say that Medicaid led to a more cost-effective outcome in the healthcare market.

Studies that do not use randomized data generally come to the conclusion that Medicaid expansion led to significant increases in healthcare utilization among expansionary states (Sommers et al., 2016; Ghosh et al., 2017; Nikpay et al., 2017). Ghosh et al. (2017) find that prescription drug utilization increased by 19 percent in Medicaid-expansion states in the first 15 months following the 2014 expansion, relative to states that did not adopt the expansion. Moreover, when isolating the heterogeneity in utilization by drug class, they find that largest increases occurred among drugs used for treating diabetes (24%) and cardiovascular diseases (21%) and smaller but significant increases occurred in acute diseases treatment. These results are corroborated by Sommers et al. (2016) who report an overall increase in prescription drug use by 12 percentage points associated with Medicaid expansion. The only discrepancy among the studies is in their estimated effects of Medicaid expansion on ED visits. While Sommers et al. (2016) report a 6% decrease, Nikpay et al. (2017) find that total ED use per 1,000 population increased by 2.5 visits more in Medicaid expansion states than in non-expansion states.

In comparison, studies from the Oregon experiment report similar trends in healthcare utilization. While Baicker et al. (2013), Finkelstein et al. (2012), and Finkelstein et al. (2016) all find that Medicaid coverage resulted in an increase in the number of hospital admissions, prescription drugs received and office visits made in the previous year, they also find that Medicaid increased ED visits by about 40%. Moreover, this increase persisted during the two-year span of the study, without evidence that Medicaid enrollees would substitute towards primary care. This is shown by the graph below and suggests that Medicaid expansion might not have led to the predicted cost-savings.

Figure 2. Total ED Use (Taubman, 2014).



Recall that cost effectiveness is achieved when Medicaid coverage decreases the number of emergency room visits because individuals will have access to primary and preventative care. Because of the contradictory findings in the literature, it is difficult to make a conclusion on the cost effectiveness of expanding Medicaid coverage.

Effects on Financial Hardship

The second outcome used to evaluate Medicaid expansion is its success or lack thereof to alleviate financial strains for the newly insured. Gauging this outcome provides an assessment of Medicaid expansion using its redistribution effect, our previously mentioned evaluation criteria.

Among studies comparing expansionary and non-expansionary states, Hu et al. (2016) and Brevoort et al. (2017) examine credit score, total debt, and the number of bills sent to collections. Both studies find that the Medicaid expansion that began in 2014 significantly reduced the number of unpaid bills and the amount of debt sent to collection agencies among people living in zip codes that are most likely affected by the expansions.

These results are similar to that of Finkelstein et al. (2012) of the Oregon experiment, who found that health insurance was associated with a decline in the probability of unpaid bills being sent to a collection agency by 4.8 percentage points, and a statistically significant reduction in out-of-pocket medical expenditures as well as the likelihood to borrow money by respectively 35% and 40%. Despite the estimates from these studies being different, their congruence in the direction of the relationship can be seen as verifying a natural generalization of the experimental result to the context of a large national reform. In sum, there is evidence that Medicaid expansion alleviated the financial burden on the newly insured.

Effects on Health

Lastly, Medicaid expansion is evaluated based on its effects on health outcomes. To study health outcomes, most empirical studies consider the impact on both clinical health outcomes and self-reported health.

Courtemanche et al. (2017) observe no statistically significant impacts on any of the risky behavior or health outcomes in either Medicaid expansion or non-expansion states. They do, however, find some evidence that the ACA improved self-assessed health among older non-elderly adults, particularly in expansion states.

In comparison, in the Oregon experiment, Baicker et al. (2013) estimate that Medicaid coverage led to 7.8 percentage points increase in the proportion of people who reported that their health was the same or better as compared with their health the previous year. Additionally, the study reports a drop in the number of people screening positive for depression by 9.2 percentage points, or a 30% reduction relative to the control group. Medicaid coverage was not shown, however, to have significant effects on physical health. Using a person's blood-pressure, cholesterol, and glycated hemoglobin levels as proxy for physical wellbeing, Baicker et al. (2013) find no significant effects of insurance coverage on these measures. However, the program was associated with a greater probability of receiving a diagnosis of diabetes and using medications by respectively 3.83 and 5.43 percentage points.

Because studying health outcomes involves using many proxies for health, we cannot conclude definitively that Medicaid expansion has a positive effect on health outcomes. It is important to note that, however, besides Baicker et al. (2013), the majority of studies referenced find at least some positive impact on health.

V. Policy Implications

Based on our research findings, we make the following recommendations in order to increase the effectiveness of Medicaid expansion:

First, Medicaid plans should have an adjusted copayment or coinsurance rate that incentivizes the utilization of low cost, preventative care services rather than emergency department use. It may also be beneficial to distribute information about primary care physicians to newly insured individuals. This would decrease the barriers to utilizing affordable healthcare services and shift stress away from costly emergency department visits for illnesses that can be treated elsewhere.

Furthermore, given the above recommendation and the findings that suggest insurance improves mental health and financial stability, Medicaid expansion should ensure that all uninsured, low-income individuals or households have access to enrollment. This would likely only be achievable through premium subsidies.

Lastly, we recommend that more research is done to explore the link between health insurance and health outcome. This may involve an extended analysis of the outcomes of the Oregon Health Experiment beyond the first few years.

Expanding Medicaid places a significant burden on taxpayers, and it is difficult to clearly weigh the benefits against the costs. The recommendations in this brief represent just a few steps to take that could improve the cost effectiveness of Medicaid and efficiency in the healthcare market.

References

Baicker, K., Taubman, S., Allen, H., Bernstein, M., Gruber, J., Newhouse, J.P., Schneider, E., Wright, B., Zaslavsky, A., Finkelstein, A. "The Oregon Experiment – Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine* (2013): 1713-1722.

Brevoort, K., Grodzicki, D., and Hackmann, M. "Medicaid and Financial Health." *National Bureau of Economic Research* (2017): 2-65.

Courtemanche, C., Marton, J., Ukert, B., and Yelowitz, A. "Early Effects of the Affordable Care Act on Health Care Access, Risky Health Behaviors, and Self-Assessed Health." *National Bureau of Economic Research* (2017): 1-71.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J., Group, T. O. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* (2012): 1057-1106.

Finkelstein, A., Taubman, S., Allen, H., Wright, B., Baicker, K. "Effect of Medicaid Coverage on ED Use – Further Evidence from Oregon's Experiment." *New England Journal of Medicine* (2016): 1505-1507.

Ghosh, A., Simon, K., Sommers, B. "The Effect of State Medicaid Expansions on Prescription Drug Use: Evidence from the Affordable Care Act." *National Bureau of Economic Research* (2017): 1-32.

Hu, L., Kaestner, R., Mazumder, B., Miller, S.,Wong, A. "The Effect of the Patient Protection and Affordable Care Act Medicaid Expansions on Financial Wellbeing." *National Bureau of Economic Research* (2016): 3-43.

Nikpay, S., Freedman, S., Levy, H., Buchmueller, T. "Effect of the Affordable Care Act Medicaid Expansion on Emergency Department Visits: Evidence From State-Level Emergency Department Databases." *Annals of Emergency Medicine* (2017): 215-225.

Sommers, B. D., Blendon, R. J., Orav, E. J., Epstein, A. M. "Changes in Utilization and Health Among Low-Income Adults After Medicaid Expansion or Expanded Private Insurance." *JAMA Internal Medicine* (2016): 1501-1509.

Taubman, S., Allen, H., Wright, B., Baicker, K., Finkelstein, A. "Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment." *Science* (2014): 263-268.

The Effects of Codeshare Agreements on Domestic Non-Stop Airfares

Alice Qingyu Zhu

Introduction to Econometrics

Abstract

This paper explores the effects of codeshare agreements on the U.S. domestic non-stop flight ticket prices. I construct a log-linear regression model with a market fixed-effect, include operating and ticketing carriers as control variables, and examine various routes from and to California. The regression results show an 8.3% decrease on average in airfares at the 1% significance level and is consistent throughout robustness checks.

I. Introduction

Code-sharing, officially named a codeshare agreement, is a business arrangement that allows tickets of the same flight to be marketed and sold under different airlines' ticket codes. Compared to the traditional behavior of operating and marketing a flight completely by one airline, code-sharing introduces a new pattern for different airlines to cooperate together. Airline companies benefit from such agreements by increasing their service frequency on served routes and gaining exposure in unserved markets without extra costs. For consumers, while code-sharing sometimes confuses them about the check-in process and the flight's operating carrier, it also brings more connectivity and provides more alternatives on schedule, which brings convenience especially for loyal customers of a specific airline. Despite these factors about convenience and customer experience, one crucial determinant of flight ticket purchase has always been ticket pricing. Since code-sharing introduces a new collaborating pattern involving different companies, it affects costs and revenues of both operating and marketing airlines, and thus, may also cause the companies to charge a different ticket price. So how do codeshare agreements affect airfares? I start by reviewing existing literature that focused on code-sharing in both international and domestic markets.

Over the last several decades, the rise of the airline industry along with its operating strategies have drawn increasing attention from economists. The influence of code-sharing on ticket prices and consumer welfare has been studied intensively in the international market, accompanied with a growing interest in the U.S. domestic market.

In international markets, where code-sharing was first developed as a strategy to expand airlines' networks and services, economists have theoretically and empirically examined airfares, consumer welfare and social welfare. While Hassin and Shy (2004) conclude that code-sharing is Pareto-improving, Park (1999) specifies that a complementary alliance, which runs traditional and vertical codeshare agreements,[18] increases overall social welfare while a parallel alliance, which runs virtual and horizontal agreements,[19] decreases it. Brueckner (2003) and Bilotkach (2005) make similar arguments that international alliances benefit interline passengers[20] significantly. Specifically, Brueckner (2003) measures a reduction of 8% to 17% in prices of international interline itineraries. Brueckner and Whalen (2000) measure the same fare reduction from a different perspective and conclude a 25% difference in interline fares between code-sharing partners and non-allied carriers. They attribute this difference to the internalization of double marginalization,[21] a problem that is likewise discussed in the context of antitrust immunity by Whalen (2007).

In domestic markets, economists have generated diverse results related to the practice of code-sharing. Czerny (2009) discusses the price discriminating purpose of complementary airline networks (i.e., vertical codeshare agreements) and casts doubts on the usefulness of such agreements to welfare improvement. Armantir and Richarg (2008) examine specifically the collaboration between Continental Airlines and Northwest Airlines and find an increase in average surplus for connecting passengers but a decrease for nonstop passengers. When considering travelers as a whole group, Gayle (2013) argues a non-trivial increase in consumer welfare due to elimination of double marginalization. He also suggests lower pricing in codeshare products. Similarly, Bamburger, Carlton and Neumann (2004) find a fall of 5-7% on average fares as a result of formation of two airline alliances. In terms of elasticity, Berry and Jia (2010) suggest that

---

[18] Traditional / vertical code-sharing: Each partner in the agreement flies part of the route and partners are able to market the entire route. Therefore, new products are generated through the practice.

[19] Virtual / horizontal code-sharing: One partner in the agreement flies the entire route and allows others in the agreement to market the route. No product is generated.

[20] Online itinerary: The operating carrier does not change during the entire route. Interline itinerary: The operating carriers change on the route.

[21] Double marginalization: Both operating and ticketing carriers impose their own price markups.

consumers were more price sensitive in 2006 than in 1999 and more strongly preferred nonstop flights.

In addition to the findings above, economists have also found evidence of how code-sharing affects flight ticket prices without specifying particular alliances or collaboration between airline companies. Shen (2017) and Ito and Lee (2007) find a negative effect of codeshare on airfares, while Gilo and Simonelli (2015) conclude an increase in airfares due to codeshare.

Shen (2017) distinguishes markets that include traditional code-sharing from those that don't because the new products generated from such practice cause extra competition in the market and extra fluctuation in ticket prices beyond code-sharing's impact. Shen develops a structural model following the classic market analysis of demand and supply. Consumers' demand is described in a discrete-choice model and companies' supply is an oligopoly. Considering both sides, Shen claims that codeshare may influence companies' pricing decisions and consumers' choices and, thus, the number of tickets purchased. After examining data from two markets before and after agreements start, Shen shows that codeshare products price 35-60% and 6-25% lower respectively in the two markets.

After finding that 85% of the U.S. domestic code-sharing is virtual, Ito and Lee (2007) construct a baseline model and a market fixed-effect model to examine the effects of virtual code-sharing. The baseline model takes market features as control variables, such as route distances, populations in origin cities and in the destination cities. The market fixed-effect generates dummies for all but one market, (i.e., directed routes) which take both observed and unobserved market features in account. Despite a difference in absolute numerical results, both models show negative signs of code-sharing consistently. In particular, the fixed-effect model suggests that domestic virtual codeshare products are 5-6% less expensive than their single-carrier substitutes.

In contrast, Gilo and Simonelli (2015) claim that the practice of code-sharing creates a round table effect and a double marginalization effect that both increase airfares. They believe codeshare partners raise airfares as they cooperate rather than compete (i.e., the round-table effect). Each partner is able to impose its own markup so overall, two markups raise the amount that draws out of consumers' pockets (i.e., the double marginalization effect). They empirically examine the relationship through a basic ordinary least squares regression, which includes a lot of control variables, such as the origin city per capita income as a market feature, miles flown as a route feature, carrier dummies as carrier features and booking class dummies as ticket features. Gilo and Simonelli (2015) find evidence of a 5% increase for each effect and conclude a total of 10% increase in airfares.

Overall, what are the effects of codeshare agreements on domestic airfares? From our review above, we see different specification of which codeshare patterns are considered in different papers. Moreover, even when specification in two papers is the same or close, for example, in Ito and Lee (2007) and Gilo and Simonelli (2015), the results can still be completely opposite. To address this contradiction, this paper conducts an empirical analysis based on the models in Shen (2017), Ito and Lee (2007) and Gil and Simonelli (2015) and examine whether the practice of code-sharing generates any influence on airfares, specifically on the U.S. domestic non-stop flights. As I only consider non-stop flights, I have simplified the situation–because only virtual code-sharing is possible for these flights, I no long worry about extra impact introduced by the presence of new products.

II. Economic Theory

Following Shen (2017), a market is defined as a directional origin and destination city pair. Each differentiated product in a certain market is determined by departure and landing time, transfer cities, operating carriers and ticketing carriers. Due to several unobserved factors in each market and on each product, such as purchasing time, frequent flyer discounts and luggage services, each product price is calculated as an average price of all tickets purchased for a certain product.

Given each market includes a set of differentiated products with some unobserved product characteristics, we consider a discrete-choice model for consumers' demand (Berry, 1994; Shen, 2017) and an oligopoly supply side with firms able to set their own prices for each product they operate. Accordingly, the ticketing carriers are not in charge of prices.

In the discrete-choice model for demand, I consider a consumer's utility is affected by product characteristics, product pricing, the codeshare status as well as other unobservable product characteristics (Shen, 2017). For the oligopoly supply side, I consider airline companies maximize profits by summing up profits from each product they are involved with as either an operating or ticketing carrier. For a codeshare product, the ticketing carrier receives part of profits from the operating carrier. Thus, the codeshare status is positively correlated with the ticketing carrier's profits but negatively correlated with the operating carrier's (Shen, 2017).

As a result, codeshare may directly affect consumers' utilities and, thus, the quantity purchased as well as the product price which is considered as an average price. Codeshare may also affect companies' marginal costs, and thus, their pricing decisions. Given the ambiguous effect of code-sharing on companies' profits overall, we are uncertain of the influence direction.

III. Empirical Theory

Ideally, in each market, a randomized control experiment would randomly assign the codeshare status to products and randomly assign a ticketing carrier different from its operating carrier to codeshare products. Then the pricing difference before and after the codeshare assignment could be explained by the codeshare status. We would gather ticket prices from all products in one market, calculate average prices of codeshare and non-codeshare products respectively, and compare the two averages to determine the codeshare effect. Otherwise, we could also run a univariate regression with a flight's codeshare status being the independent variable and the ticket price being the dependent variable. The coefficient for the independent variable would quantify the codeshare effect.

However, given diversity in markets, carriers and itineraries, such randomized control experiments would be extremely time-consuming, energy-consuming, and inefficient. To simulate results from such an experiment, we consider a market fixed-effect model (Ito and Lee, 2007) and take into account heterogeneity in market-specific, carrier-specific, and itinerary-specific characteristics (Gilo and Simonelli, 2015; Ito and Lee, 2007). Given the distribution of airfares is skewed to the left, as shown in Figure 1, we take the natural logarithm of airfares rather than use airfares directly. We regress the natural logarithm of airfare $ln(fare)_i$ on the codeshare status $D(codeshare)_i$ and several control variables, shown as the log-linear model in Equation (1). Itineraries are indexed by i and markets are indexed by j.

$$\ln(fare)_i = \beta_0 + \beta_1 D(codeshare)_i + \boldsymbol{\alpha}_1 \boldsymbol{Itin}_i + \boldsymbol{\alpha}_2 \boldsymbol{D}(\boldsymbol{op})_i + \boldsymbol{\alpha}_3 \boldsymbol{D}(\boldsymbol{tkt})_i + \boldsymbol{\gamma} \boldsymbol{D}(\boldsymbol{mkt})_j + \varepsilon_i \quad (1)$$

In Equation (1), control variables include itinerary characteristics $\boldsymbol{Itin}_i$, dummy variables for operating carriers $\boldsymbol{D}(\boldsymbol{op})_i$ and for ticketing carriers $\boldsymbol{D}(\boldsymbol{tkt})_i$. The dummy variables for carriers capture carrier-specific characteristics. The itinerary-specific characteristics which are summarized as $\boldsymbol{Itin}_i$, include $D(LCC)_i$, a dummy variable to indicate the involvement of a low-cost carrier (LCC) and $Orgshare_i$, the itinerary's market share in the origin city (Ito and Lee, 2007).

The fixed-effect control is revealed in $\boldsymbol{\gamma}\, \boldsymbol{D}(\boldsymbol{mkt})_j$, where $\boldsymbol{D}(\boldsymbol{mkt})_j$ consists of dummy variables representing all but one markets. As this term handles market-specific characteristics, here we provide several examples of such characteristics (Ito and Lee, 2007). Depending on the features of origin and destination city pairs, the price of an itinerary may be influenced by demographics and market competitiveness, such as the distance between two cities ($dist_j$), the population (potential maximal market size) of the origin city ($popori_j$), the population of the destination city ($popdest_j$), the collective share of low-cost carriers ($LCCshare_j$) and the number of all itineraries in the market ($Noitin_j$). Given the variety and complexity of market characteristics, for the sake of efficiency and accuracy, I manage them using a market fixed-effect model.

Further, I remove both variables for itinerary characteristics – $D(LCC)_i$, the dummy variable of Low Cost Carriers, and $Orgshare_i$, the itinerary's market share in the origin city – for following reasons. In markets that we consider, low-cost carriers operate and market their flights completely on their own, without any ticketing or operating overlap with other airlines. Since these low-cost tickets are all non-codeshare, including them does not help explain how codeshare status affects airfares at all. Hence, I omit them from our regression. For $Orgshare_i$, since the dataset I use only provides purchase records without specifying itineraries or flight time, I assume each purchase record represents an independent itinerary. Then as each record's market share is tiny, this variable does not affect regression much and thus, we take this variable out.

Therefore, the primary regression we adopt for the empirical analysis is as following:

$$ln\,(fare)_i = \beta_0 + \beta_1\, D(codeshare)_i + \boldsymbol{\alpha}_1\, \boldsymbol{D(op)}_i + \boldsymbol{\alpha}_2\, \boldsymbol{D(tkt)}_i + \boldsymbol{\gamma}\, \boldsymbol{D(mkt)}_{\boldsymbol{j}} + \varepsilon_i \qquad (2)$$

Although a market fixed-effect helps the regression take unobserved market factors into consideration, price determinants related to unobserved itinerary-specific factors and individual preferences may still cause omitted variable biases. For example, flight time, airline services and fare classes are all itinerary-specific factors that influence consumers' ticket purchase and thus, airfares. Though we are aware of these variables' importance, these measurements are unable to be included in the regression due to lack of data.

To identify the codeshare effect on airfares, we examine the value and the significance of the coefficient β_1 for the codeshare dummy variable $D(codeshare)_i$.

IV. Data Description

The data for our analysis is from the Airline Origin and Destination Survey (DB1B) provided by the U.S. Department of Transportation (DoT). This survey samples 10% of U.S. domestic airline tickets and includes three datasets respectively organizing information according to coupons, tickets and markets. My analysis draws 430,531 observations from the DB1B Market dataset, representing domestic flights from or to California in the second quarter of 2017. I choose this sample mainly because California has more than one of the largest airports in the U.S. flights arriving or departing there are not dominated by one specific airline or airport. This pattern of involving most airlines brings in variation in carriers and markets nicely. The second quarter has the least holiday seasons in all four quarters, so prices and consumers' purchasing behavior are least affected by holiday reasons.

The unit of observation in the dataset is a ticket purchased at a specific price. We consider 5 groups of variables in our analysis, including Airfare, Codeshare Status, Ticketing Carriers, Operating Carriers and Flight Markets as unique origin and destination city pairs. As shown in Table 1, we keep ticket prices that are higher than $25 and lower than $2,000 to omit existence of discounts and other privileged purchases. We then take the natural logarithm of Airfare as its distribution skewed to the left. Figure 2 shows the distribution of ln(Airfare), which is much closer to a normal distribution than that of Airfare as shown in Figure 1. In Table 2, we see the vast majority (over 95%) of flights are non-codeshare. For carriers that are summarized in Table 3 and 4, we recode regional carriers as their parent carriers and consider only domestic carriers for both Ticketing and Operating Carriers. The dataset contains 95 origin cities and 91 destination cities, as well as a total of 649 unique origin and destination city pairs, each of which is considered a market.

Table 5 examines two different codeshare patterns in the dataset. First, some carriers only operate flights but never market even their own operating flights. For example, KS and YV operate all their flights under names of AA, AS and UA. Second, some carriers share their operating flights with codeshare marketing partners and also market their partners' flights. For examples, AA and AS market each other's flights in addition to operating and marketing their own flights.

When conducting a t-test on average airfares for codeshare and non-codeshare products, we see a *t*-statistic of 68.50 and an average decrease of 131.81 dollars from non-codeshare to codeshare products. As the t-statistic implies a 100% difference between ticket prices in two groups, we expect to see a negative sign in front of the coefficient for codeshare in our regression.

V. Results

As shown in the first column of Table 6, our primary regression indicates that the codeshare status decreases airfares by 8.3% on average at a 1% significant level. Noticeably, the dummy control variable for the operating carrier DL is automatically omitted due to multicollinearity. In Table 5, we see that DL is the only carrier that operates and markets its flights without any interactions with other airlines. As the regression already includes a dummy variable for the ticketing carrier DL, another dummy variable for the operating carrier DL does not add any extra information to the regression and, thus, is eliminated.

To check robustness, in the second column of Table 6, we drop the carrier DL from the dataset and run the regression again. Despite some small changes in coefficients for carrier dummies, the result of interest is -8.2%, almost the same as from our primary regression. As I have discussed the two patterns of code-sharing at the end of the Data Description section, carriers KS and YV only operate but never market flights. I am also curious about how the results change when these two are removed. By removing these two carriers, I expect the results to show only the effects of when partners mutually market their flights. As shown in the third column of Table 6, the coefficient is still -8.3%, the same as in our primary regression. In this regression, the dummy variable of the operating carrier UA is also omitted, because after removing KS and YV, the carrier UA stands in the same situation as DL, operating and marketing flights entirely by itself. Finally, when we drop DL, KS, and YV altogether, as shown in the last column of Table 6, the result is -8.2%. As numbers from all these robustness checks are the same or quite similar to -8.3% from our primary regression, I believe this result is concrete no matter what codeshare patterns the dataset covers.

The R-squared for all four regressions are from 35.3 to 35.7%. I do anticipate some unobserved omitted variables, which might be flight time or time ranges, airplane models and travelers' main purposes of traveling. Also, as our dataset only consists of flights from or to California, we might have missed some patterns that are more dominant in other markets. Despite these potential biases, I conclude that my results are robust and implies an average decrease of 8.3% caused by domestic code-sharing.

VI. Conclusion

In this paper, I have explored the effects of codeshare agreements on the U.S. domestic non-stop flight ticket prices. Following existing discussion, I see an ambiguous codeshare influence on domestic airfares. While some argue that airfares decrease (Ito and Lee, 2007; Shen, 2017), others argue that airfares increase (Gilo and Simonelli, 2015). For economic theory, I adopt a discrete-choice demand model and an oligopoly supply model. Due to the unclear codeshare impact on companies' profits, I could not anticipate the effect direction. I construct a log-linear regression model with a market fixed-effect which includes carriers as control variables. I examine the relation on various routes from and to California. The regression shows an 8.3% decrease on average in airfares at the 1% significance level as the result of codeshare and is consistent throughout robustness checks.

References

Armantier, O., and Richard, O. (2008). Domestic airline alliances and consumer welfare. *The RAND Journal of Economics*, *39*(3), 875–904.

Bamberger, G. E., Carlton, D. W., and Neumann, L. R. (2004). An Empirical Investigation of the Competitive Effects of Domestic Airline Alliances. *The Journal of Law and Economics*, *47*(1), 195–222.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics, 25*(2), 242–262.

Berry, S., and Jia, P. (2010). Tracing the Woes: An Empirical Analysis of the Airline Industry. *American Economic Journal: Microeconomics*, *2*(3), 1–43.

Bilotkach, V. (2005). Price Competition between International Airline Alliances. *Journal of Transport Economics and Policy*, *39*(2), 167–189.

Brueckner, J. K. (2003). International Airfares in the Age of Alliances: The Effects of Codesharing and Antitrust Immunity. *The Review of Economics and Statistics*, *85*(1), 105–118.

Brueckner, J. K., and Whalen, W. T. (2000). The Price Effects of International Airline Alliances. *The Journal of Law and Economics*, *43*(2), 503–546.

Czerny, A. I. (2009). Code-sharing, Price Discrimination and Welfare Losses. *Journal of Transport Economics and Policy (JTEP)*, *43*(2), 193–212.

Gayle, P. G. (2013). On the Efficiency of Codeshare Contracts between Airlines: Is Double Marginalization Eliminated. *American Economic Journal: Microeconomics*, *5*(4), 244–273.

Gilo, D., and Simonelli, F. (2015). The price-increasing effects of domestic code-sharing agreements for non-stop airline routes. *Journal of Competition Law and Economics*, *11*(1), 69–83.

Hassin, O., and Shy, O. (2004). Code-sharing Agreements and Interconnections in Markets for International Flights. *Review of International Economics*, *12*(3), 337–352.

Ito, H., and Lee, D. (2007). Domestic Code Sharing, Alliances, and Airfares in the U.S. Airline Industry. *The Journal of Law and Economics*, *50*(2), 355–380.

List of the busiest airports in the United States. (2018, April 26). In *Wikipedia*. Retrieved from en.wikipedia.org/w/index.php?title=List_of_the_busiest_airports_in_the_United_States&oldid=838358019.

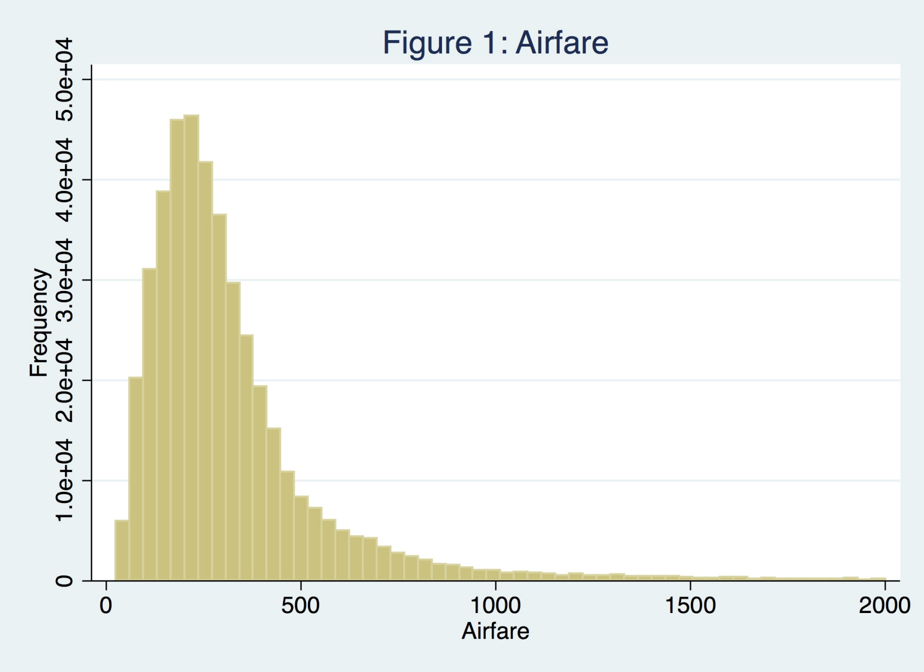
Park, J.-H. (1997). The effects of airline alliances on markets and economic welfare. *Transportation Research Part E: Logistics and Transportation Review*, *33*(3), 181–195.

Shen, C. (2017). The effects of major U.S. domestic airline code sharing and profit-sharing rule. *Journal of Economics and Management Strategy*, *26*(3), 590–609.

U.S. Department of Transportation. *The airline origin and destination survey (DB1B)*. Retrieved from www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=125&Link=0.

Whalen, W. T. (2007). A panel data analysis of code-sharing, antitrust immunity, and open skies treaties in international aviation markets. *Review of Industrial Organization*, *30*(1), 39–61.

Tables and Figures



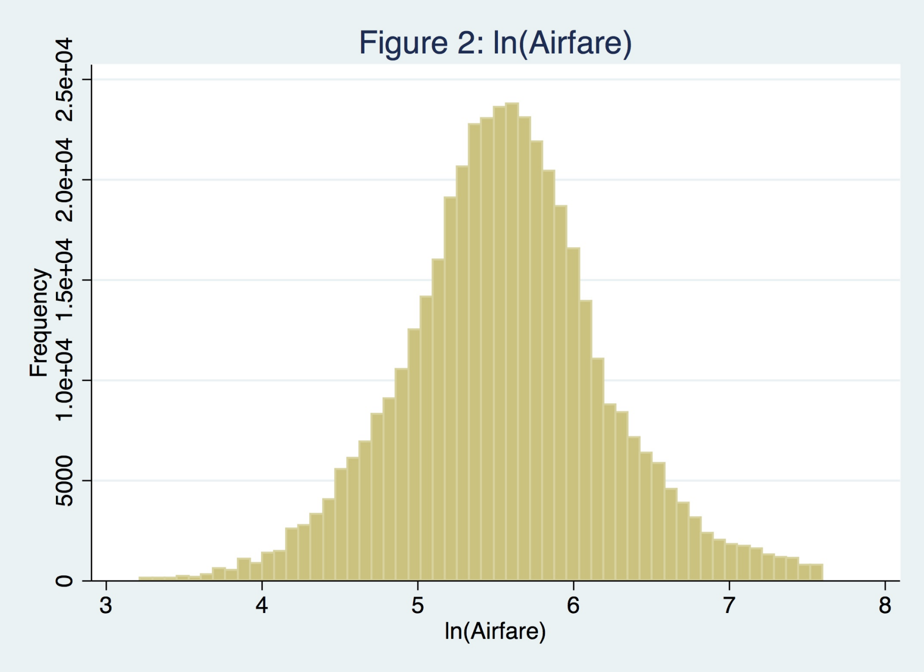


Table 1: Summary of Airfare.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Airfare	430,531	323.2	249.3	25	2000

The distributions of Airfare and ln(Airfare) are shown on the previous page.

Table 2: Summary of Codeshare Status.

Codeshare Status	Freq.	Percent	Cum.
Non-Codeshare	413,237	95.98	95.98
Codeshare	17,294	4.02	100.00
Total	430,531	100.00	

For Codeshare Status, 0 represents a non-codeshare flight and 1 represents a codeshare flight.

Table 3: Summary of Ticketing Carriers.

Ticketing Carrier	Freq.	Percent	Cum.
AA	95,652	22.22	22.22
AS	76,857	17.85	40.07
DL	80,165	18.62	58.69
HA	15,694	3.65	62.33
UA	126,841	29.46	91.80
VX	35,322	8.20	100.00
Total	430,531	100.00	

All six carriers both market and operate flights. See Table 4 for more explanation.

Table 4: Summary of Operating Carriers.

Operating Carrier	Freq.	Percent	Cum.
AA	91,526	21.26	21.26
AS	64,698	15.03	36.29
DL	80,165	18.62	54.91
HA	15,692	3.64	58.55
KS	24	0.01	58.56
UA	126,833	29.46	88.02
VX	46,824	10.88	98.89
YV	4,769	1.11	100.00
Total	29,688	100.00	

All carriers except KS and YV both market and operate flights. KS and YV only operate flights.
All flights operated by KS and YV are considered codeshare flights.
Detailed codeshare patterns are shown in Table 5.

Table 5: Detailed Codeshare Patterns.

	Ticketing Carrier						
Operating Carrier	AA	AS	DL	HA	UA	VX	Total
AA	90,709	817	0	0	0	0	91,526
AS	182	64,516	0	0	0	0	64,698
DL	0	0	80,165	0	0	0	80,165
HA	0	0	0	15,692	0	0	15,692
KS	0	24	0	0	0	0	24
UA	0	0	0	0	126,833	0	126,833
VX	0	11,500	0	2	0	35,322	46,824
YV	4,761	0	0	0	8	0	4,769
Total	95,652	76,857	80,165	15,694	126,841	35,322	430,531

Two different patterns of codeshare are revealed.
First, some operating carriers of codeshare flights do not market their flights.
For example, KS and YV only operate flights under names of AA, AS and UA.
Second, some carriers share their operating flights with codeshare marketing partners
and also help partners market their flights.
For examples, AA and AS both operate and market their own flights as well as market the other's flights.

Table 6: Log-linear Regressions with a Market Fixed Effect.

VARIABLES	(1) ln(Airfare)	(2) Without DL	(3) Without KS and YV	(4) Without DL, KS and YV
Codeshare Status	-0.083***	-0.082***	-0.083***	-0.082***
	(0.015)	(0.015)	(0.015)	(0.015)
Ticketing Carrier = 2, AS	-0.068***	-0.071***	-0.069***	-0.071***
	(0.015)	(0.015)	(0.015)	(0.015)
Ticketing Carrier = 3, DL	0.012***		0.012***	
	(0.004)		(0.004)	
Ticketing Carrier = 4, HA	0.154***	0.159***	0.154***	0.158***
	(0.025)	(0.027)	(0.025)	(0.027)
Ticketing Carrier = 5, UA	0.454***	0.438***	0.048***	0.032***
	(0.149)	(0.149)	(0.004)	(0.004)
Ticketing Carrier = 6, VX	0.180***	0.181***	0.179***	0.180***
	(0.015)	(0.015)	(0.015)	(0.015)
Operating Carrier = 2, AS	-0.074***	-0.106***	-0.073***	-0.105***
	(0.015)	(0.016)	(0.015)	(0.016)
Operating Carrier = 4, HA	-0.204***	-0.241***	-0.203***	-0.241***
	(0.025)	(0.027)	(0.025)	(0.027)
Operating Carrier = 5, KS	-0.222**	-0.221**		
	(0.099)	(0.099)		
Operating Carrier = 6, UA	-0.406***	-0.406***		
	(0.149)	(0.149)		
Operating Carrier = 7, VX	-0.297***	-0.321***	-0.297***	-0.320***
	(0.015)	(0.015)	(0.015)	(0.015)
Operating Carrier = 8, YV	0.099***	0.098***		
	(0.023)	(0.023)		
Constant	5.098***	5.099***	5.133***	5.136***
	(0.066)	(0.066)	(0.088)	(0.088)
Observations	430,531	350,366	425,738	345,573
R-squared	0.355	0.357	0.353	0.355

Market dummies for the market fixed-effect are hidden.
Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1