Analysis of the Efficacy of Development Subsidy Policy in the United States

Zachary Kobban
Gambier, Ohio 43022
(510) 575-2537

Dear Selection Committee,

In the coming summer, I propose to work with Professor Daniel Koliner on a research project with the goal of determining, empirically, the efficacy of development subsidy policy in the United States. The phenomenon of states competing to attract large companies through various incentives, including tax breaks, loans, land and infrastructure, has been a prominent feature of our economic development policy. A key example of this is $2 billion in subsidies given to Intel for the development of a chip factory in New Albany, Ohio. This taxpayer money was given to Intel under the assumption that this factory would attract economic growth to the area and benefit Ohio through more jobs and increased standards of living. However, questions remain. What are the actual observed short and long run effects? Who are the recipients of such benefits? Do these types of policies benefit people of color? Do most of the benefits of these subsidies benefit wealthy households?

In order to determine the actual impact of development subsidy policy, I plan on utilizing the difference in difference regression analysis with synthetic control groups. This methodology was developed by Abadie in their paper *The Economic Costs of Conflict: A Case Study of the Basque Country*. It was then popularized by *Minimum Wage Increases, Wages, and Low-Wage Employment: Evidence from Seattle*. Difference in Difference (DID) works generally by comparing the before-and-after changes in outcomes between a group that receives the subsidies (treatment group) and one that does not (control group). This strategy is useful because it allows researchers to control for potential factors that could influence measurements of the success of the program but are not caused by the program. For example, if a given policy was instituted in 2007, without more evidence, it would be incorrect to state that the policy led to the great recession of 2008. The method relies on the assumption that the treatment and control groups follow parallel trends. In other words, if neither group received the treatment (in this case the subsidy), both groups would have followed similar paths over time. This assumption is critical because it enables us to attribute differences between the group outcomes to the treatment being studied (the subsidy). By analyzing the "difference in differences," we can isolate the effect of development subsidies on the treated group after accounting for baseline changes in the control group.

In order to create the differences in differences (DID) analysis we need data. To start, we will access Good Jobs First's subsidy tracker. This will provide me with a list of development subsidies, their value, zip codes, year, company and type. Traditionally, in order to satisfy the parallel trends assumption, researchers would match areas with generally similar characteristics and growth trends. Because areas near each other are likely to have similar characteristics, trends, laws, ect, this is often done by matching areas located in close physical proximity to the area directly affected by the policy. This would not work in our case because of a risk of policy spillover. Policy spillover is highly likely in our case because subsidies that impact determinants of economic success are likely to also impact the economy of nearby areas. In order to address this problem, we will use the concept of synthetic control groups. To accomplish this, we will create a procedure to generate areas with comparable characteristics and trends, which we will call synthetic controls. These are made by taking different percentages of existing areas to create a sort of “franken-"area.” These synthetic controls will then be compared to the areas that received subsidies to measure the effects of the policy.

In order to correctly match subsidy areas with proper synthetic controls we will need demographic data. This data will be taken from the LEHD's Origin-Destination (OD), Resident area characteristic (RAC) and Workplace area characteristic (WAC) datasets. This data is publicly available from 2002 to 2021. The OD data includes data on the origins and destinations of workers. This facilitates a comprehensive analysis of commuting patterns. The RAC data is useful because it gives us stronger and more specific information about where and how individuals live. This allows for an in-depth analysis of the residential populations of specific areas, helping to identify which communities may be underserved or disproportionately affected by economic policy. RAC data will most likely be used to create synthetic controls and verify the parallel trends assumption. WAC data gives us insight into the employment characteristics in specific areas. Lastly, the American Community Survey will be used to access additional demographic data that is only available in yearly increments. We will also primarily rely on the LEHD data because it will allow us to view specific counts for jobs, worker age, earnings, industry sector, race and educational attainment at the neighborhood level. The Census Bureau defines a neighborhood to contain ~1500 people. Neighborhood level data is also important in this case because it gives us a very large number of observations and therefore a stronger statistical model. After running the DID regression, the benefit of the small geographical area of the neighborhood is that it allows us to see with greater accuracy who is impacted and where the benefits or consequences of a particular subsidy are realized. Data at the neighborhood level allows us to determine if the benefits of these subsidies travel to low-income and / or mostly BIPOC areas. This allows us to answer questions of equity concerns from both a race and class perspective.

Research is fluid, but this is what we currently believe will be included in the statistical model:

$$share\_5miles\_away_{bt} = \lambda_{gt} + \lambda_b + \beta SubsidyValue_{bt}$$

The subscript t represents time, b represents neighborhood and g represents group. In our model, λ_b accounts for the fact that different neighborhoods are likely to have different economic baselines even if they follow similar trends. λ_{gt} controls local trends that vary across time but that are unrelated to the treatment. In other words, this controls for natural economic fluctuations that occur in both the synthetic control and subsidy recipient. The outcome of interest is Share_5miles_away. This can be thought of as the share of workers that live more than 5 miles from the project site. The exact

number of miles can be adjusted further down the line. The term β is the coefficient of interest. The value of this coefficient determines the outcome of the DID analysis.
For a subsidy value in the millions, we might expect the β coefficient to be 0.05. That would mean that an additional million dollars in subsidy will lead to an 5% increase in the share of commuters from 5 miles away or more.

All of the costs associated with this project are for software and program licenses. Because all of the census demographic data is publicly available, this is free. The subsidy tracker created by the Good Jobs First watchdog costs $45 for the amount of data I would need to access. Because of the estimated number of observations that this project needs to work with, I am also requesting access to a yearlong Stata/SE license. This costs $179 and therefore total costs will be $224.

I want to do this research over the summer for many reasons. First, I am incredibly interested in looking at economic policy from an empirical perspective. I have always been motivated to study economics in order to attempt to understand how we can create more efficient and equitable policy and this research project is a perfect way to do this. Second, the work I have done recently at the Knox County Area Development Foundation (Knox ADF) has brought me face to face with specific questions and issues surrounding development subsidy policy. As an Intern, I attempted to predict how the New Albany Intel plant would impact the demographic changes and amenity preferences of Knox County Residents. The goal of this research is to make sure Knox County is in the best position to attract the benefits of the already approved development subsidy. When doing this work, I often found myself asking questions as to the efficacy of this program specifically, and this policy in general. How does this benefit or harm the current residents of Knox county? Summer Scholars offered me a perfect opportunity to do real work to answer this specific question. Third, the type of work that I am proposing to do this summer is what I would like to do for my career. After Kenyon, my ultimate goal is to work for the Federal Reserve Bank system. There I would hopefully be working under PhD Economists doing similar data analysis projects. From there, I have in mind getting a PhD in Economics. This would enable me to continue asking questions about the efficacy of our current policy and providing actionable proposals for improvements. Summer Scholars will give me the experience and confidence to pursue this career and life goal.

Task	Deadline
Data Collection	5/20/24
Data Cleaning/Wrangling	5/24/24
Implement Distance Function	5/31/24
Select Final Regression	6/7/24
Paper First Draft	6/21/24
Paper Final Draft	7/5/24
Submit to Journals	7/9/24

Note: All dates are dependent on the start date of the program and thus, this serves as a general outline.

The final goal of this project is to submit this work to the Regional Science and Urban Economics as well as the Journal of Policy Analysis and Management, John Wiley and Sons. The opportunity to possibly co-author in an economic journal is something that attracted me greatly to the Summer Scholars program.

Thank you for your time and consideration,
Zachary Kobban

Mentor Contact Information:
Daniel Kolliner
kolliner1@kenyon.edu

Economics Academic Advisor Contact Information:
Jaret Treber

treberj@kenyon.edu

Mathematics Academic Advisor Contact Information:

Bob Milnikel
milnikelr@kenyon.edu