2-1-2019

The Effect of Public Transportation Quality on Economic Outcomes

John Safy
CUNY Hunter College

The Effect of Public Transportation Quality on Economic Outcomes

by

John Safy

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Economics
Hunter College, City University of New York

2018

Thesis sponsor:

January 3, 2019	Dr. Randall K. Filer
Date	First Reader
January 3, 2019	Dr. Partha Deb
Date	Second Reader

Contents

1 Abstract

There exists a large body of research on the economic effects of creating and investing in public transportation infrastructure. There is however a comparatively small literature on the economic effects of the quality of existing infrastructure. Quality can be measured in a number of ways, but we focus on measures of on-time performance. Using New York City economic data and on-time performance from 2010-2016 we can examine the relationship between public transportation quality and economic outcomes, in this case zip code level payroll, number of businesses, and number of employees. We use fixed effects, random effects, and finite mixture models, and find statistically significant effects on all three outcomes, with increasing on-time performance increasing the number of businesses but decreasing the number of employees and total payroll.

2 Introduction

2.1 Background

Public transportation systems provide multiple benefits to the communities they serve. By giving residents greater mobility riders are better able to choose where to live and work as befits their skills and preferences, enhancing both their own utility and society's. This choice brings a host of economic, environmental, and social benefits, collectively termed option value, which can be viewed as a positive externality of investment in public transportation. Option value in this case may include greater workplace satisfaction, a better work-life balance, reduced carbon emissions from not using a car to commute, etc.

New York City has often been considered a poster child for public transportation systems. While New York City has multiple forms of public transportation this paper looks only at the New York City subway system. New York City's subway system is the ninth oldest in the world, has the most stations and lines in the world, and runs 24 hours a day, every single day of the year (MTA 2018). It serves 424 stations with over 850 miles of track and an average weekday ridership of 5.6 million riders. These unique circumstances have proven challenging to the maintenance of the subway as the average on-time-performance has decreased every year since 2010 as seen in Figures 1 and 2.

Another challenge is the subway's budget structure. New York City's Metropolitan Transportation Agency (MTA), through a series of agreements and charters, has ended up with the state Governor, 150 miles away in Albany, in charge of the majority of the MTA's budget. While there have been budget problems before, the efficacy of the MTA's budget agreements was called into question when the city and state fought over their contribution to the MTA budget, with Mayor Rudolph Giuliani cutting the city's contribution by 400 million dollars, prompting Governor George Pataki to follow suit. This set a precedent of mayors and governors cutting MTA budgets to use the money elsewhere. The MTA turned to borrowing, and is now spending 17% of its budget on interest payments. This makes it more difficult for the MTA to expand, or even maintain, its current infrastructure. The NYC subway has the dubious honor of being the only major subway system with fewer miles of track in 2018 than during World War II. These challenges may have contributed to the system's deteriorating performance, which culminated in Governor Andrew Cuomo declaring an MTA state of emergency in 2017, pledging an extra billion dollars to infrastructure improvements as well as loosening bureaucratic procedures.

Poor performance negates the benefits of option value when trains are consistently late and riders are no longer able to rely on their train's schedule. This lack of certainty, trust, and accountability may lead to riders being forced to make suboptimal decisions due to the logistics of their commute. Similarly, deteriorating performance may affect other quantified gains that the subway provides, e.g. the effect of subway access to property prices (Falcocchio et al. 2018).

The subway system is resilient and has recovered from poor circumstances before. In the 1970s and 1980s the subway experienced some of its lowest ridership due to crime, neglect, and inefficiency. However, a joint effort by city and state officials led to what then-governor George Pataki termed a "transit renaissance." By evaluating the effects of public transportation performance, the long-term effects of aging infrastructure can be intelligently, cost-effectively treated.

2.2 Literature Review

Weisbrod et al. (2014), in a piece prepared for the American Public Transportation Association, used an economic impact model to analyze the effect of public policy and budget shifts towards public transportation. Their base case used current ridership and investment, while their hypothetical case added 14.2 billion dollars per year over 20 years. They found that 1 billion dollars invested yields a 3.7 billion dollar increase in GDP, while also creating more than 50,000 jobs, of which 46% are created due to long-term productivity increases and other benefits within the community. The productivity impacts they found include access to a broader labor market, broader customer bases, travel cost savings, which may lead to increased consumer spending, and reduced travel congestion. These results are cumulative over the 20 years, and suggest compounding growth over time. They also found that transit agencies throughout the country are facing budget deficits, with the majority of agencies forced to raise fares or cut services from 2009–2014.

Another way for transit systems to raise funds is by entering into contracts with local developers. Mathur and Smith (2013) examine the effect of joint development projects on transit agency revenues. They look at five different case studies, some of which were revenue sharing agreements and others cost sharing. These agreements can include the leasing or sale of land, charging fees for special rights, etc. They find that while the results vary from case to case, local real estate market conditions are perhaps the most important contributor to revenue gained from joint development projects. Although agencies cannot directly control real estate prices one thing agencies can do to increase their revenue include conducting thorough studies to capture the value increase that their presence provides.

Falcocchio et al. (2018) provide one example of such a study. They look at the value increase provided by the NYC subway to commercial properties. Commercial properties were chosen because they benefit heavily from reliable and efficient public infrastructure, and thus the authors feel should be asked to contribute towards the maintenance of said infrastructure. To determine a fair contribution they first estimated how much commercial properties benefitted from their implicit subsidizing by public transportation, which they found to be \$4.58 per square foot. Then they determined the average value lost to delays. By using data on per capita gross metropolitan product and on-time performance (the same metric which our study analyzes) they found that a 2% reduction in on-time performance yields an additional average wait time of two minutes, which results in lost wages of \$0.51 per square foot, based on square footage and worker density data. Using these two figures to guide their estimation of a fair contribution, they determine that additional revenues of \$332 million to \$664 million could be generated if commercial and office properties were made to contribute to the MTA budget, an amount far below the value gained from nearby public transportation and consistent with the lost wages figure. This corresponds to a 0.3 to 0.6 percentage point increase in property tax rates, and could help offset up to \$0.22 in fare increases which may disproportionately affect low income riders.

Glaeser et al. (2008) analyzed land use and development patterns within the context of poverty and public transportation. They posited that the traditional explanation of urban poverty clusters, i.e. the income elasticity of demand for land did not fully explain the how poverty in cities developed and examined alternate explanations. Using data from the 2001 National Household Transportation Survey (NHTS) in conjunction with a linear regression model to determine the time costs of public transportation compared to other forms of transportation as a means of explaining the preference of the better off to use automobiles, they found that in NYC public transportation is both slower and has higher fixed time costs than driving. Thus as income increases, agents are expected to switch from public transportation to driving once their opportunity cost of time exceeds the increased cost of driving. Using a theoretical model they also present evidence that public transportation explains almost 75% of the centralization of poverty around public transportation, with an especially strong effect in the initial formation of such clusters. These clusters then become ingrained over time as other factors take hold, e.g. housing prices, education, etc. These results are also consistent when looking at multiple cities. They find that similar clustering in London can be partially attributed to public transportation. They note, however, that Paris is an exception to this trend, which they attribute to Napoleon III who initiated a massive program of urban gentrification that made the city center increasingly appealing to the wealthy. Their research has important implications both for existing and new public transportation systems, and shows that public transportation has a profound effect on the formation of neighborhoods and economic health of those within.

Baek (2016) attempts to analyze the impact of public transportation on food insecurity. Food insecurity is the inability to consistently obtain a varied and sufficient diet and is an issue that disproportionately affects poorer neighborhoods. Baek analyzes the impact of quality public transportation on food insecurity using data from the Current Population Survey Food Security Supplement (CPS-FSS) and the National Transit Database, covering 2006–2009, and estimates a linear probability model with both a binary dependent variable representing general food security and a categorical measure which takes into account respondents' answers to the questions of the CPS-FSS. Baek finds that for every additional bus running per 10,000 people food insecurity decreases by 1.6%, noting that this effect is stronger in poorer neighborhoods as well as predominantly minority neighborhoods. These results are in line with Glaeser et al.'s findings that public transportation quality has a greater effect on poorer and more disenfranchised communities.

Dragu et al. (2013) studied measures of quality for public transportation systems. There are many different ways of measuring public transportation quality, of which Dragu et al.

considered eight: services offered, accessibility, availability of information, on-time performance, attention given to passengers, comfort, safety, and environmental impacts. These metrics are both quantitative and qualitative, and try to account for both objective measures of quality as well as public perceptions of quality, which can have an inordinate impact on ridership. Unfortunately, gathering information on all of these metrics can be difficult, costly, and time-consuming. Dragu et al. argue that on-time performance alone can accurately explain the public's behavior in choice of transportation. To examine their measure of quality they performed a case study using Bucharest's public transportation system. They analyzed the on-time performance of an above-ground bus line on several weekends, with the idea that these constraints should prove as beneficial as possible to the line's timeliness. They then surveyed riders about their experiences and found that they reported a decrease in quality as deviations from the schedule increased, in either direction. From a provider's perspective, they also found that on a per-vehicle basis delays tend to compound until the vehicle ends its route, which can lead to delays throughout the line, further affecting public perception. Dragu et al. also note that there exists a tradeoff between consumer expectations and other factors, such as economic feasibility or environmental impact that transportation agencies need to consider when designing and operating their systems. Due to their findings, this study uses on-time performance as the sole basis for its quality measures with a degree of confidence.

3 Methodology

This study used panel formatted data to analyze the effect of on-time performance on economic outputs, including annual payroll, number of employees, and number of establishments per zip code. Panel data was chosen because it allows one to control for variables which may be difficult or impossible to measure but that change slowly over time. For example, panel data can control for differences in the culture across zip codes. In econometric terms it accounts for individual heterogeneity.

The panel data models used in this study are fixed effects or random effects, depending on the Hausman test. Fixed effects models are mechanically interesting because they allow one to look at the effect of time-variant variables on the dependent variable while removing the effect of time-invariant variables. It follows that fixed effects models cannot be used to look at the effect of time-invariant variables on the dependent variable. The tradeoff is that fixed effects removes unobserved heterogeneity. On the other hand, random effects assumes there is no covariance between the unobserved heterogeneity term and the covariates, which allows for the analysis of time invariant variables. The Hausman test can be used to help decide whether to use fixed or random effects. Hausman (1978) formulated a test whose null hypothesis is that there is no correlation between the covariates and the unobserved heterogeneity term. If one fails to reject the null the random effects model is more appropriate, else the fixed effects model may be more well suited.

In addition to the more traditional panel data methods this study used finite mixture models as another means of accounting for unobserved heterogeneity. Finite mixture models classify observations into distinct, unobserved subgroups which are not present in the data, allowing one to analyze different groups without knowing how to exactly define a subgroup. They do this by estimating the probability of each observation belonging to a certain group using their distributions. Given a user specified n the model creates n different groups which together approximate the original distribution of the data, but individually may look very different from one another. The model then assigns each observation a group depending on its characteristics. Thus they are most useful when the available dataset does not contain important predictors due to the difficulty of defining such predictors or lack of data. In our case, one argument may be that some zip codes may have a disproportionately important line. For example, some lines may have no feasible alternative routes, even if there are other lines serving that zip code. Another argument may be that some zip codes are less often stopped at, instead only being commuted through on the way to a different zip code. Quantifying such cases is difficult, but by using finite mixture models to split the data into groups we can then analyze and infer what makes each group distinct to better understand the effects on these sub-populations.

Finite mixture models can also be used to look at the relationship between time-variant variables and the dependent variable, similar to fixed effects. They can achieve this by using the deviations from the within-group mean as the variables on which the model classifies each observation. In doing so the effects of the time-invariant variables drop out, providing the same benefits as a fixed effects model with the additional grouping of the finite mixtures model. One can use the Bayesian information criterion (BIC) metric in order to determine which model is more appropriate, as well as to determine the optimal n. The BIC is a technique which can aid in model selection, penalizing over-specified models which may overfit the available data. A lower BIC might indicate that a model is more appropriate.

4 Data

4.1 Collection

All data was made freely available by various city, state, and federal departments and can be downloaded from either its respective department's website or NYC Open Data, a website which aggregates NYC related datasets.

Data on subway stations was obtained from NYC Open Data [1] in the form of a shapefile. This data contains the station's name, latitude, longitude, and the lines that serve it. Using the latitude and longitude coordinates, the station's zip code was reverse geoencoded. We chose to omit Staten Island from this study because Staten Island is an anomaly in a few respects. The island has only one train line, the Staten Island Railway, which does not connect to any other line and does not provide a way off the island. The Staten Island Railway also operates slightly differently than the rest of the subway. It is operated by the Staten Island Rapid Transit Operating Authority, a subsidiary of the MTA. These differences are why it was decided that this study would focus on the rest of New York City, with its more representative subway system.

Data on business and economic characteristics was obtained from the Zip Codes Business Patterns dataset, collected by the United States Census Bureau yearly. This data can be found on the Census website [2]. The Zip Codes Business Patterns data includes annual and Q1 payroll (in thousands of dollars), number of employees, and the number of businesses in various size categories, in terms of employees: 1 to 4, 5 to 9, 10 to 19, 20 to 49, 50 to 99, 100 to 249, 250 to 499, 500 to 999, and over 1,000 employees. The Zip Codes Business Patterns dataset includes the option to obtain data for specific industries per zip code or for

[1] https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49
[2] https://www.census.gov/data/developers/data-sets/cbp-nonemp-zbp/zbp-api.2016.html

all industries in that zip code. The industry specific data is however limited to the number of businesses, lacking number of employees and payroll, and has a significant amount of missing data and thus this study used the numbers pertaining to all industries per zip code. This data is available per zip code per year from 1994–2016.

Lastly, data on train performance was obtained from the MTA developer portal [3]. This study used actual on-time performance per line per year as well as the MTA's projected on-time performance per line per year. The MTA defines on-time performance as the percent of the time a line arrives at its final stop within 5 minutes of its scheduled time. There also exist a host of other metrics, such as injury rates, mean distance between failures, etc. which could provide interesting further avenues of study. Data is available by month. This study uses the year-to-date measure of on-time percentage as measured in November. November was chosen instead of December due to the absence of December projected on-time performance values in 2010, which would have resulted in one year less data. This dataset limited the analysis period, as it only goes as far back as June 2009. Furthermore, MTA metric projections for the next year do not exist until 2010. Since these projections are an integral part of this study, it was decided that this study would focus on 2010–2016.

4.2 Processing

While the dependent variables all came from the Zip Codes Business Patterns dataset, several of the independent variables had to be generated. We used two quality measures, both of which were created for this study. The first measure was defined as:

$$\min\{L_1, L_2, \dots L_n\} \tag{1}$$

where L_i represents the on-time percentage of line i out of n serving a zip code in a year. It is called minimum performance in the regression tables.

This measure was chosen due to the idea that an organization is only as strong as its weakest link. If a business has workers from throughout the city, each taking different lines to get to work, one consistently delayed line may have an exaggerated impact on workplace efficiency as other workers may not be able to work at peak output until everyone is present.

[3] http://web.mta.info/developers/developer-data-terms.html#data

Another measure was:

$$\min\{(L_1^a - L_1^p), (L_2^a - L_2^p), \ldots (L_n^a - L_n^p)\} \tag{2}$$

where L_i^a refers to the actual uptime of line i in a zip code in a year and L_i^p represents the MTA's projected uptime for the same line in that year. It is called minimum disruption in the regression tables. Since the sign can be negative the smallest value for disruption represents the largest shock from the projections.

This measure was chosen as a way to account for public perceptions of a line's reliability. The MTA's projections are made the year before, and while they don't make their process public, it is likely a function of prior performance as well as ongoing and planned infrastructure improvements, ridership, number of trains, etc., using data which may not be available to the public, meaning that these are likely the most accurate predictions of future performance. These projections are publicly available on the MTA website and assuming that riders are rational, and update their expectations as new information is received their predictions should not systematically differ from the MTA's projections on average. It follows that riders will adjust their behavior in response to a train's on-time performance and its expected performance in the future, which this metric attempts to take into account by looking at the effects of shocks and disruptions. Tellingly of the state of the system, in seven years of data, among 119 zip codes, a total of 833 observations, only 23 observations or 2.76% performed better than their projections.

The final dataset was in panel format and encompassed 119 zip codes, with seven years of data available in total, from 2010–2016. Zip codes without subways were not included in the final dataset, nor were Staten Island zip codes for aforementioned reasons. In total there are 833 observations with 25 variables.

4.3 Summary Statistics

Tables 1–5 show the characteristics of subway performance and economic output per zip code. Table 1 displays summary statistics for all zip codes examined. There are no concerning values or outliers, such as a line with 0% uptime. Tables 2–5 break down the summary statistics by borough. Manhattan has by far the highest annual payroll, number of employees, and number of businesses, but also the widest fluctuation between subway performance, as well as deviations from the MTA's projected values. Many subway lines in Kings, Queens, and Bronx counties are the only lines serving their specific neighborhoods, meaning less crowded stations and tracks. However most lines run through Manhattan, often sharing stations, leading to congestion and widespread delays when a delay in one line affects other lines as well.

Figures 1 and 2 display the decline in the city's subway systems visually. It groups together the 119 zip codes into 4 counties for easier viewing. Looking at Figure 1 one can see that there has not been a single year in the studied time period where on-time performance actually increased. Performance either remains steady or decreases, with an especially sharp decline from 2013–2014. Interestingly, the ordering of boroughs never changes. Queens is consistently the best performing county, then Kings, then Bronx, then New York. In Figure 2 one can see that disruptions from the MTA's projected values also display a downward trend.

5 Results

Tables 6 and 7 report the results of the panel data methods. Both sets of variables were analyzed with fixed and random effects, and the Hausman test was used to determine which model was more appropriate. In almost all cases, the Hausman test indicated that the fixed effects model was more appropriate. A "*" denotes cases where the results from the random effects models were used. The Hausman test cannot be used with robust standard errors, so a separate set of regressions were run with cluster robust standard errors, clustered on the grouping variable, zip code. In all cases the robust result matched the sign of the non-robust result. However, some robust results were not statistically significant at the 5% level even if the non-robust was statistically significant at that level. These results are denoted with a "ˆ" Table 6 reports the relationship of (1) with the dependent variables Annual Payroll, Employees, and Establishments. Table 7 reports the relationship of (2) with the same dependent variables. The models include the number of trains in that zip code, the projected performance of the worst performing train for that zip code and year, NYC's GDP for that year, and the annual payroll, number of employees, and number of establishments in that zip code for that year.

In Table 6 one can see that the relationship between (1) and annual payroll, employees, and establishments is statistically significant at the 5% level for each coefficient. A 1% increase in lowest on-time percentage in a zip code corresponds to a \$5,742,300 decrease in annual payroll, and almost 36 fewer employees. A 1% increase in lowest on-time percentage also corresponds to about one more establishment per zip code.

The relationship between (2) and the same dependent variables can be seen in Table 7. Once again, the coefficient on annual payroll is statistically significant at the 5% level, and the coefficients for employees and establishments are significant at the 1% level. Again, the signs for annual payroll and employees are both negative. Since annual payroll represents the sum of every employee's pay for that year in that zip code, it makes sense that fewer employees results in a lower annual payroll.

Tables 8 and 9 break down the relationship of (1) and (2) respectively by county. This shows that the most significant results pertain to New York county. There the coefficients on annual payroll, employees, and establishments are significant to the 0.1% level. Once again, we see the same signs and correlation between annual payroll and employees when looking at New York county. Looking column-wise, one sees that the coefficient on establishments is statistically significant across all counties, whereas the coefficients on annual payroll and employees are only significant in New York county. These results imply that changes in behavior vary by county.

Tables 10 and 11 display the results of (1) and (2) finite mixture models. The Bayesian information criterion (BIC) were used to help determine the optimal n and whether or not to use the fixed effects version of the finite mixture model. In all cases the BIC indicated that the fixed effects model was more appropriate than the standard model. However in some cases the BIC indicated that the optimal number of groups was two and in other cases three. The differences in the BIC in the cases where $n = 3$ was determined to be optimal were less than 1%, and so this study uses only the results from setting n equal to two, for the sake of explanation.

Interpreting which observations compose components one and two requires some inference. In Table 10 one can see that the largest coefficients occur in component one, although the coefficients in component two are more statistically significant. In Table 11 the results are slightly less distinct, with component 1's coefficient on annual payroll exceeding that of component two, but the component 2's coefficients on employees and establishments being the larger of the two, and all three larger coefficients being statistically significant. Again the signs follow the same pattern. It should also be noted how similar these results are to the New York county specific results.

To better understand how the finite mixture model split the groups, we can generate the posterior probability of an observation belonging in a group. We generated these probabilities with respect to an observation belonging to the group with the larger absolute value coefficient, then analyzed these probabilities with an OLS regression, the results of which can be seen in Tables 12 and 13. These coefficients are either weakly correlated (less than 20% correlation) or inversely correlated with each other, meaning that the groupings for each regression differ strongly. Tables 12 and 13 show that the most statistically significant predictors of group membership are a dummy variable representing whether an observation is in New York county, the number of employees, and the number of establishments. Given the data available any inferences about the groupings rely partly on conjecture, but we posit that the groupings roughly correspond to business hubs, of which New York county has many, explaining the large coefficients in each model. Kings county has also seen rapid economic development during the years studied, with the average number of establishments per zip code increasing 18.6%, which is reflected in the coefficient on establishments.

One possible explanation for the negative relationship between on-time performance with annual payroll and employees is that the relationship between on-time performance and number of establishments may vary depending on the sizes of the establishments. We can see evidence of this in Tables 13 and 14, where the signs on the coefficients of employees and establishments are always inverted. Tables 14 and 15 look at this relationship more closely. We find that the coefficient on smaller businesses (1–4, 5–9, and 10–19 employees) is positive, but that the coefficient on larger businesses is negative; better subway performance is correlated with more small businesses. Table 16 looks at the relationship between annual payroll and establishments of different sizes, and shows that smaller establishments (again 1–4, 5–9, and 10–19 employees) are negatively related to annual payroll. These results partially explain the negative relationship between on-time performance with annual payroll and employees. As on-time performance improves more smaller employers open and these smaller businesses may not be able to pay their employees as well as a larger company might.

6 Conclusion

Public transportation is often called the skeleton of a city, and using that metaphor New York City has osteoporosis. NYC relies on its various public transportation systems to an almost unique degree, expecting it to work 24 hours a day, every day, and reach the most far flung corners of the city. This relationship had proven fruitful for many years, catapulting the city's wealth and standing, but along the way the incredible efforts in maintaining such an infrastructure began to be taken for granted as budget cuts and layoffs became more commonplace.

Both models suggest a relationship where declining subway performance has a significant impact both on economic output as well as rider behavior. The clearest result, which complements the work of Falcocchio et al. (2018) nicely is that increased subway performance results in an increased number of businesses per zip code. This can be due to a number of reasons, such as a larger customer base, increased access to talent, higher neighborhood desirability, etc., all of which lead to a zip code being able to support more businesses, *ceteris paribus*. These findings, in addition to Falcocchio et al.'s value of a $4.58 implicit subsidization per square foot for commercial properties show the scale to which landlords and commercial ventures benefit from access to quality public transit. The effects on annual payroll and number of employees are also significant, and move in lockstep. They suggest a story of employees and entrepreneurs changing their behavior by starting more small businesses, which tend to pay less, or by working in other zip codes when their home zip code's line(s) improve.

By showing the relationship between public transportation performance and economic output we add another angle for policymakers to consider, as well as encourage more research into this heretofore untapped field. While the problems facing public transportation agencies all over the country are daunting, there are ways to turn it around, and it starts with raising awareness of how far the problems extend.

References

[1] T. W. Anderson and Cheng Hsiao. Estimation of dynamic models with error components. *Journal of the American Statistical Association*, 76(375):598–606, 1981.

[2] Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

[3] Deokrye Baek. The effect of public transportation accessibility on food insecurity. *Eastern Economic Journal*, 42(1):104–134, Jan 2016.

[4] Han Cheng, Chao Mao, Samer Madanat, and Arpad Horvath. Minimizing the total costs of urban transit systems can reduce greenhouse gas emissions: The case of san francisco. *Transport Policy*, 66:40 – 48, 2018.

[5] Vasile Dragu, Eugenia Alina Roman, and Vlad Catalin Roman. Quality Assessment In Urban Public Transport. *Theoretical and Empirical Researches in Urban Management*, 8(3):32–43, August 2013.

[6] John Falcocchio, Awais Malik, and Constantine Kontokosta. A data-driven methodology for equitable value-capture financing of public transit operations and maintenance. *Transport Policy*, 1 2018.

[7] Edward Glaeser, Matthew E. Kahn, and Jordan Rappaport. Why do the poor live in cities? the role of public transportation. *Journal of Urban Economics*, 63(1):1–36, 11 2008.

[8] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[9] J. A. Hausman. Specification tests in econometrics. *Econometrica*, 46(6):1251–1271, 1978.

[10] Shishir Mathur and Adam Smith. Land value capture to fund public transportation infrastructure: Examination of joint development projects' revenue yield and stability. *Transport Policy*, 30:327 – 335, 2013.

[11] MTA. About us: The mta network: Public transportation for the new york region. `http://web.mta.info/mta/network.htm`, 2018. [Online; accessed 26-November-2018].

[12] Glen Weisbrod, Derek Cutler, and Chandler Duncan. Economic impact of public transportation investment. *American Public Transportation Association*, 2014.

Tables and Figures

Figure 1: Average value of minimum performance by county over time

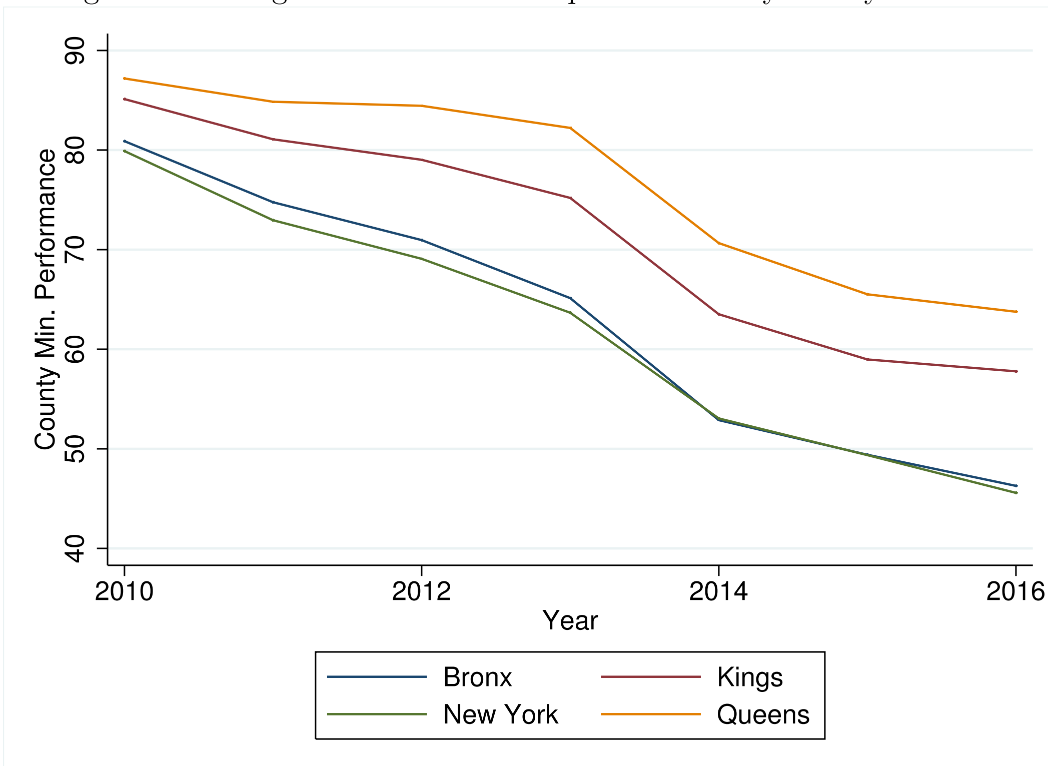

Figure 2: Average value of minimum disruption by county over time

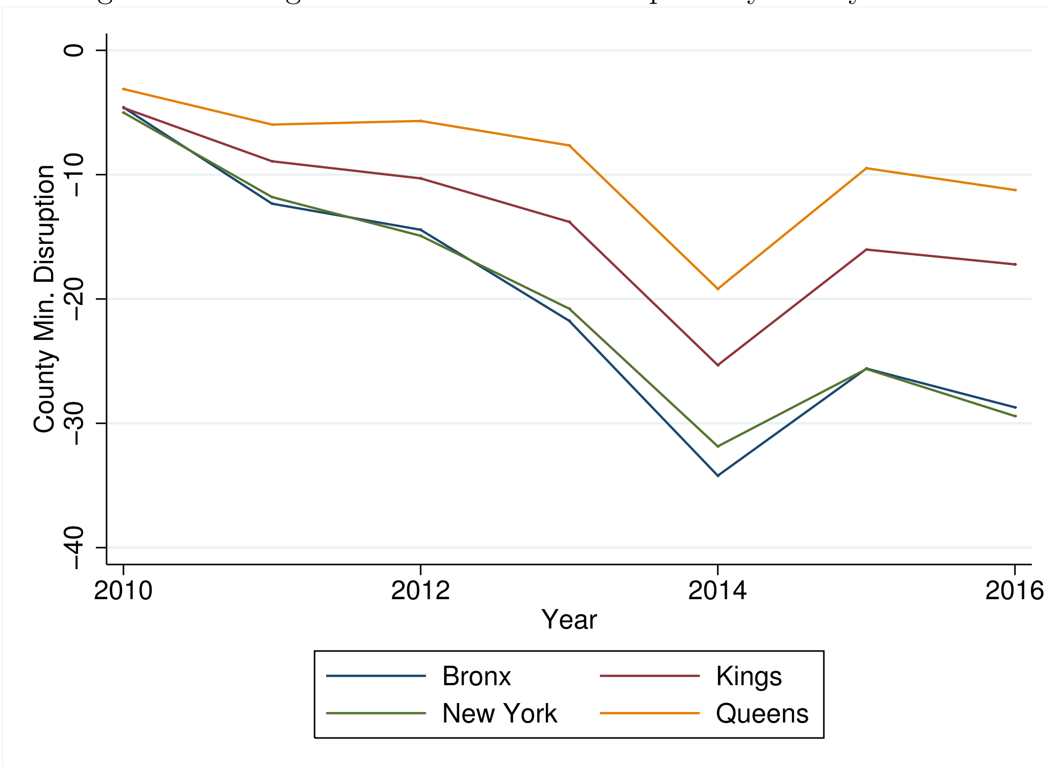

Table 1: Zip code descriptive statistics

	Mean	Variance	Std. Dev.	Min	Max
Min. Performance	68.358	270.912	16.459	36.570	97.600
MTA Projection	83.486	49.571	7.041	75.000	98.700
Min. Disruption	-15.609	151.462	12.307	-43.500	16.800
Annual Payroll	1747536.000	1.48e+13	3843672.000	30299.000	2.31e+07
Employees	22317.110	9.72e+08	31176.890	779.000	164531.000
Establishments	1500.011	1900672.000	1378.649	84.000	7373.000
Num. Trains	3.606	10.030	3.167	1.000	16.000
Num. Stations	3.807	6.752	2.599	1.000	15.000
GDP	1233.531	947.423	30.780	1188.749	1275.137
Num. 1–4	948.102	617015.500	785.503	47.000	4273.000
Num. 5–9	227.815	54204.630	232.819	11.000	1299.000
Num. 10–19	150.485	26823.150	163.778	4.000	856.000
Num. 20–49	106.215	16472.930	128.347	2.000	655.000
Num. 50–99	35.848	2819.007	53.094	0.000	281.000
Num. 100–249	20.642	1172.954	34.248	0.000	187.000
Num. 250–499	6.275	126.880	11.264	0.000	74.000
Num. 500–999	2.747	23.572	4.855	0.000	29.000
Num. 1000+	1.882	10.676	3.267	0.000	20.000
N	833				

Table 2: Zip code descriptive statistics for Brooklyn (King's County)

	Mean	Variance	Std. Dev.	Min	Max
Min. Performance	71.529	248.264	15.756	36.570	96.100
MTA Projection	84.738	60.996	7.810	75.000	96.800
Min. Disruption	-13.744	126.916	11.266	-43.500	16.800
Annual Payroll	550260.500	2.25e+11	474385.600	102585.000	2974770.000
Employees	14535.250	9.61e+07	9803.614	3506.000	55958.000
Establishments	1457.920	516674.800	718.801	364.000	3127.000
Num. Trains	3.559	7.809	2.794	1.000	13.000
Num. Stations	4.941	9.980	3.159	1.000	15.000
GDP	1233.531	950.278	30.827	1188.749	1275.137
Num. 1–4	1005.303	256795.200	506.750	269.000	2177.000
Num. 5–9	210.303	12792.730	113.105	31.000	545.000
Num. 10–19	126.214	4657.241	68.244	22.000	359.000
Num. 20–49	78.286	1918.045	43.796	18.000	275.000
Num. 50–99	20.954	244.399	15.633	0.000	106.000
Num. 100–249	10.697	68.659	8.286	1.000	51.000
Num. 250–499	3.479	11.820	3.438	0.000	20.000
Num. 500–999	1.651	3.713	1.927	0.000	10.000
Num. 1000+	1.034	1.864	1.365	0.000	9.000
N	238				

Table 3: Zip code descriptive statistics for Queens (Queen's County)

	Mean	Variance	Std. Dev.	Min	Max
Min. Performance	76.951	123.237	11.101	53.900	97.600
MTA Projection	85.638	55.069	7.421	75.000	98.700
Min. Disruption	-8.901	46.711	6.835	-30.200	8.100
Annual Payroll	481748.400	5.70e+11	755118.200	35819.000	4723439.000
Employees	10764.240	1.49e+08	12213.580	840.000	72657.000
Establishments	1006.408	585422.400	765.129	90.000	3586.000
Num. Trains	2.235	2.211	1.487	1.000	8.000
Num. Stations	2.964	5.204	2.281	1.000	12.000
GDP	1233.531	951.138	30.841	1188.749	1275.137
Num. 1–4	684.694	243743.600	493.704	55.000	2556.000
Num. 5–9	144.628	15486.860	124.446	13.000	551.000
Num. 10–19	87.036	6593.624	81.201	4.000	431.000
Num. 20–49	58.342	4535.354	67.345	2.000	387.000
Num. 50–99	17.612	646.382	25.424	0.000	150.000
Num. 100–249	9.561	173.899	13.187	0.000	85.000
Num. 250–499	2.709	15.756	3.969	0.000	22.000
Num. 500–999	1.036	2.3013	1.517	0.000	8.000
Num. 1000+	.791	1.705	1.306	0.000	8.000
N	196				

Table 4: Zip code descriptive statistics for The Bronx (Bronx County)

	Mean	Variance	Std. Dev.	Min	Max
Min. Performance	62.903	290.710	17.050	36.570	90.900
MTA Projection	82.825	43.212	6.574	75.000	94.200
Min. Disruption	-20.233	181.987	13.490	-43.500	1.200
Annual Payroll	364454.900	6.43e+10	253553.300	87657.000	1101432.000
Employees	9249.698	2.29e+07	4784.282	2367.000	20616.000
Establishments	739.937	57697.800	240.204	283.000	1297.000
Num. Trains	1.778	1.182	1.087	1.000	5.000
Num. Stations	3.444	1.593	1.262	1.000	6.000
GDP	1233.531	953.856	30.885	1188.749	1275.137
Num. 1–4	494.706	24968.000	158.013	200.000	808.000
Num. 5–9	100.206	1505.237	38.797	34.000	204.000
Num. 10–19	69.175	861.185	29.346	22.000	149.000
Num. 20–49	49.365	336.170	18.335	11.000	91.000
Num. 50–99	14.476	43.035	6.560	4.000	35.000
Num. 100–249	7.976	19.639	4.432	1.000	24.000
Num. 250–499	2.103	3.357	1.832	0.000	7.000
Num. 500–999	1.143	2.299	1.516	0.000	7.000
Num. 1000+	.786	.954	.977	0.000	4.000
N	126				

Table 5: Zip code descriptive statistics for Manhattan (New York County)

	Mean	Variance	Std. Dev.	Min	Max
Min. Performance	61.944	273.322	16.532	36.570	96.100
MTA Projection	81.156	28.781	5.365	75.000	95.700
Min. Disruption	-19.916	171.600	13.100	-43.500	16.800
Annual Payroll	4338430.000	3.45e+13	5875688.000	30299.000	2.31e+07
Employees	43426.780	2.10e+09	45786.970	779.000	164531.000
Establishments	2241.890	3920275.000	1979.968	84.000	7373.000
Num. Trains	5.476	15.331	3.916	1.000	16.000
Num. Stations	3.590	5.750	2.398	1.000	10.000
GDP	1233.531	949.765	30.818	1188.749	1275.137
Num. 1–4	1296.608	1207384.000	1098.810	47.000	4273.000
Num. 5–9	361.703	112071.100	334.770	11.000	1299.000
Num. 10–19	254.725	55481.520	235.545	7.000	856.000
Num. 20–49	191.172	34235.190	185.028	2.000	655.000
Num. 50–99	71.788	5984.999	77.363	1.000	281.000
Num. 100–249	43.114	2638.189	51.363	1.000	187.000
Num. 250–499	13.198	292.799	17.111	0.000	74.000
Num. 500–999	5.670	53.229	7.296	0.000	29.000
Num. 1000+	3.912	23.191	4.816	0.000	20.000
N	273				

Table 6: Panel data model estimates of min. performance on economic outcomes

	Annual Payroll	Employees ˆ	Establishments ˆ*
Min. Performance	-5742.3*	-35.77*	1.067*
	(-2.42)	(-2.43)	(2.09)
MTA Projection	3707.2	34.20*	-1.719**
	(1.36)	(2.02)	(-2.93)
Num. Trains	-106591.3	3810.0***	66.99***
	(-1.06)	(6.23)	(4.54)
GDP	922.5	-2.827	1.195***
	(0.93)	(-0.46)	(5.88)
Employees	88.36***		0.0127***
	(17.50)		(10.52)
Establishments	-1026.9***	9.384***	
	(-4.65)	(6.97)	
Annual Payroll		0.00342***	-0.0000194*
		(17.50)	(-2.49)
_cons	645407.0	-8391.5	-394.4
	(0.45)	(-0.94)	(-1.27)
N	833	833	833

t statistics in parentheses

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

Table 7: Panel data model estimates of min. disruption on economic outcomes

	Annual Payroll	Employees	Establishments ˆ
Min. Disruption	-4811.5*	-31.95**	0.936**
	(-2.54)	(-2.71)	(2.94)
Num. Trains	-100902.1	3825.2***	-74.58***
	(-1.00)	(6.28)	(-4.48)
GDP	1401.0*	-1.607	1.595***
	(2.17)	(-0.40)	(17.57)
Employees	88.81***		0.00685***
	(17.66)		(6.99)
Establishments	-1016.2***	9.405***	
	(-4.61)	(6.99)	
Annual Payroll		0.00344***	-0.0000287***
		(17.66)	(-4.61)
_cons	-149568.1	-10110.7*	-287.3*
	(-0.22)	(-2.35)	(-2.48)
N	833	833	833

t statistics in parentheses

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

Table 8: Panel data model estimates of min. performance on economic outcomes by County

	Annual Payroll	Employees	Establishments
Bronx Min. Performance	570.6	-21.09	1.304**
	(0.19)	(-1.12)	(2.80)
Kings Min. Performance	-3186.7	-4.126	-3.055***
	(-1.04)	(-0.22)	(-6.65)
New York Min. Performance	-12317.0***	-70.04***	1.717***
	(-4.51)	(-4.09)	(4.04)
Queens Min. Performance	-2366.6	-11.61	0.121
	(-0.67)	(-0.53)	(0.22)
MTA Projection	1645.5	19.51	-0.582
	(0.59)	(1.12)	(-1.34)
Num. Trains	-161944.0	3440.8***	-51.93***
	(-1.61)	(5.60)	(-3.36)
GDP	636.0	-4.832	1.302***
	(0.65)	(-0.79)	(9.02)
Employees	81.67***		0.00677***
	(15.75)		(7.52)
Establishments	-819.1***	10.97***	
	(-3.40)	(7.52)	
Annual Payroll		0.00319***	-0.0000197***
		(15.75)	(-3.40)
_cons	1168071.5	-5874.1	25.95
	(0.82)	(-0.66)	(0.12)
N	833	833	833

t statistics in parentheses

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

Table 9: Panel data model estimates of min. disruption on economic outcomes by County

	Annual Payroll	Employees	Establishments
Bronx Min. Disruption	2326.1	-9.533	1.880***
	(0.75)	(-0.49)	(3.75)
Kings Min. Disruption	-1007.5	6.530	-2.136***
	(-0.33)	(0.34)	(-4.33)
New York Min. Disruption	-13563.5***	-77.55***	2.128***
	(-5.30)	(-4.82)	(5.07)
Queens Min. Disruption	1523.6	-14.82	1.779*
	(0.35)	(-0.55)	(2.52)
Num. Trains	-140798.5	3536.5***	-66.07***
	(-1.42)	(5.82)	(-4.11)
GDP	1307.1*	-2.709	1.588***
	(2.03)	(-0.67)	(18.16)
Employees	81.19***		0.00705***
	(15.69)		(7.44)
Establishments	-859.2***	10.30***	
	(-3.78)	(7.44)	
Annual Payroll		0.00318***	-0.0000231***
		(15.69)	(-3.78)
_cons	37660.6	-8649.6*	-323.2**
	(0.05)	(-2.02)	(-2.90)
N	833	833	833

t statistics in parentheses

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

Table 10: Finite mixture model estimates of min. performance on economic outcomes

	Annual Payroll	Employees	Establishments
Component 1			
Min. Performance	-612.8***	1.197	-1.003***
	(-4.15)	(0.39)	(-4.67)
MTA Projection	-624.9***	1.936	0.214
	(-3.81)	(0.56)	(0.83)
Num. Trains	-1454.8	54.51	41.29**
	(-0.12)	(0.28)	(2.99)
GDP	-148.6*	0.610	0.486***
	(-2.44)	(0.51)	(5.52)
Employees	30.96***		0.00379***
	(27.01)		(6.99)
Establishments	175.1***	4.302***	
	(6.17)	(12.98)	
Annual Payroll		0.00957***	9.63e-08
		(47.63)	(0.02)
Component 2			
Min. Performance	-13892.4	-71.83	1.695
	(-1.27)	(-1.47)	(1.88)
MTA Projection	15567.6	68.83	-2.713**
	(1.05)	(1.12)	(-2.89)
Num. Trains	-362613.3	4519.1***	-135.4***
	(-1.62)	(3.72)	(-4.45)
GDP	4861.2	-0.916	2.426***
	(0.98)	(-0.04)	(6.70)
Employees	81.41***		0.00955***
	(7.26)		(4.12)
Establishments	-1872.7**	11.24**	
	(-2.87)	(3.18)	
Annual Payroll		0.00293***	-0.0000569***
		(8.41)	(-4.12)
N	833	833	833

t statistics in parentheses

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

Table 11: Finite mixture model estimates of min. disruption on economic outcomes

	Annual Payroll	Employees	Establishments
Component 1			
Min. Disruption	-15858.0**	-0.190	-0.695***
	(-2.74)	(-0.07)	(-3.86)
Num. Trains	90134.0	1973.4***	45.49***
	(0.58)	(7.76)	(3.45)
GDP	3257.1	-0.102	0.653***
	(1.63)	(-0.11)	(10.40)
Employees	80.11***		0.00375***
	(9.88)		(7.10)
Establishments	-1312.8**	5.530***	
	(-2.82)	(15.84)	
Annual Payroll		0.00663***	0.000000377
		(59.34)	(0.06)
Component 2			
Min. Disruption	-59.00	-86.65*	2.411***
	(-0.51)	(-2.01)	(3.56)
Num. Trains	-5300246.0***	3249.0*	-132.5***
	(-255.30)	(2.52)	(-4.44)
GDP	112.8**	0.136	2.658***
	(2.66)	(0.01)	(12.89)
Employees	32.36***		0.00963***
	(25.35)		(4.20)
Establishments	166.0***	12.97**	
	(7.25)	(3.26)	
Annual Payroll		0.00264***	-0.0000563***
		(6.78)	(-4.10)
N	833	833	833

t statistics in parentheses

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

Table 12: OLS estimates on probability of belonging to more affected group w.r.t. Table 10

	Prob. Annual Payroll	Prob. Employees	Prob. Establishments
Bronx	0	0	0
	(.)	(.)	(.)
Kings	0.00939	-0.0101	0.309***
	(0.32)	(-0.28)	(9.40)
New York	-0.270***	0.102**	0.116**
	(-8.51)	(2.66)	(3.23)
Queens	-0.0316	0.0236	0.104**
	(-1.09)	(0.67)	(3.20)
Num. Stations	-0.00313	0.0179**	0.00396
	(-0.61)	(2.89)	(0.69)
Num. Trains	-0.0153**	0.00304	-0.00684
	(-3.05)	(0.50)	(-1.20)
Annual Payroll	1.38e-08	-1.93e-08*	2.09e-08**
	(1.95)	(-2.24)	(2.60)
Employees	-0.00000981***	0.0000109***	-0.00000589***
	(-6.81)	(6.22)	(-3.62)
Establishments	0.0000577**	-0.0000608**	0.000183***
	(3.15)	(-2.73)	(8.84)
_cons	1.050***	0.0529	0.0647*
	(40.76)	(1.69)	(2.22)
N	833	833	833

t statistics in parentheses

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

Table 13: OLS estimates on probability of belonging to more affected group w.r.t. Table 11

	Prob. Annual Payroll	Prob. Employees	Prob. Establishments
Bronx	0	0	0
	(.)	(.)	(.)
Kings	-0.0125	-0.0273	0.311***
	(-0.36)	(-0.76)	(9.42)
New York	-0.286***	0.0660	0.111**
	(-7.64)	(1.69)	(3.10)
Queens	-0.142***	0.0316	0.106**
	(-4.15)	(0.89)	(3.24)
Num. Stations	0.00150	0.0165**	0.00418
	(0.25)	(2.62)	(0.72)
Num. Trains	-0.0155**	-0.00261	-0.00643
	(-2.60)	(-0.42)	(-1.13)
Annual Payroll	2.34e-08**	-2.29e-08**	2.00e-08*
	(2.79)	(-2.62)	(2.48)
Employees	-0.0000106***	0.00000873***	-0.00000571***
	(-6.24)	(4.92)	(-3.50)
Establishments	0.0000538*	-0.00000986	0.000180***
	(2.48)	(-0.44)	(8.67)
_cons	1.037***	0.0349	0.0703*
	(34.05)	(1.10)	(2.41)
N	833	833	833

t statistics in parentheses

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

Table 14: Panel data model estimates of min. performance on establishments by size

	Num. 1–4	Num. 5–9	Num. 10–19	Num. 20–49	Num. 50–99
Min. Performance	0.922**	0.350**	0.149	-0.163*	-0.0591
	(2.75)	(2.66)	(1.64)	(-2.51)	(-1.85)
MTA Projection	-1.249**	-0.361*	-0.256*	0.00653	0.0589
	(-3.25)	(-2.38)	(-2.42)	(0.09)	(1.62)
Num. Trains	20.41*	15.63***	10.75***	3.075	6.138***
	(2.03)	(5.53)	(6.08)	(1.13)	(4.59)
GDP	0.751***	0.227***	0.0954**	0.0804**	0.0228
	(5.65)	(4.34)	(2.63)	(3.12)	(1.80)
Annual Payroll	-2.27e-5***	-3.94e-6*	-8.44e-08	6.92e-7	3.39e-6***
	(-4.42)	(-2.02)	(-0.06)	(0.68)	(6.82)
Employees	0.00583***	0.00324***	0.00273***	0.000914***	0.000381***
	(7.32)	(11.02)	(13.74)	(5.72)	(4.85)
_cons	-101.3	-167.4*	-55.72	-15.04	-29.76
	(-0.50)	(-2.13)	(-1.02)	(-0.38)	(-1.55)
N	833	833	833	833	833

t statistics in parentheses

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

Table 15: Panel data model estimates of min. disruption on establishments by size

	Num. 1–4	Num. 5–9	Num. 10–19	Num. 20–49	Num. 50–99
Min. Disruption	0.846***	0.192*	0.183*	-0.0752	-0.0573*
	(3.86)	(2.26)	(2.50)	(-1.45)	(-2.25)
Num. Trains	-72.41***	-16.85***	10.90***	3.417	6.142***
	(-6.30)	(-3.79)	(6.21)	(1.26)	(4.61)
GDP	0.969***	0.307***	0.114***	0.119***	0.0236**
	(15.47)	(12.66)	(5.50)	(8.04)	(3.26)
Annual Payroll	-2.72e-5 ***	-6.68e-6***	-1.08e-7	7.67e-7	3.40e-6***
	(-6.35)	(-4.03)	(-0.08)	(0.76)	(6.85)
Employees	0.00272***	0.00103***	0.00274***	0.000930***	0.000385***
	(4.02)	(3.93)	(13.76)	(5.82)	(4.91)
_cons	13.79	-98.37**	-87.85***	-76.30***	-30.91***
	(0.17)	(-3.18)	(-3.39)	(-4.04)	(-3.33)
N	833	833	833	833	833

t statistics in parentheses

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

Table 16: Panel data model estimates of establishments on Annual Payroll

	Annual Payroll
Num. 1–4	-452.1
	(-1.23)
Num. 5–9	-1798.2
	(-1.72)
Num. 10–19	-346.7
	(-0.28)
Num. 20–49	2551.0
	(1.71)
Num. 50–99	19825.2***
	(6.94)
Num. 100–249	37511.8***
	(9.64)
Num. 250-499	34842.9***
	(4.96)
Num. 500–999	71860.9***
	(6.26)
Num. 1000+	82046.3***
	(4.06)
_cons	311539.3
	(1.19)
N	833

t statistics in parentheses

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