Abstract

This project will look at an obscure part of many city economies: the taxicab industry. In US cities, government regulations often directly or indirectly restrict the number of taxicabs allowed on the streets. Thus, taxicab supply in these places is essentially a controlled number. The goal is to use a model of taxicab demand to determine how well government actions focus on taxicab demand factors when regulating taxicab supply; for a sample of 259 US cities. This addresses the question of how governments act within the taxicab industry, and what they act for. Regression analysis is used to reveal the taxicab demand factors that are statistically significant in explaining the government regulated number of taxicabs. Interpretations of these results provide quantitative evidence to qualitative analysis that shows governments do not necessarily act with ‘public interest’ in mind when establishing regulations within the taxicab industry.
# Table of Contents

Acknowledgements .............................................................................................................. 1

The Issues ............................................................................................................................ 2

Past Developments ............................................................................................................ 7

Concepts, Methods and Data ............................................................................................. 11

Results ............................................................................................................................... 18
  Additional Remarks and Results ..................................................................................... 23

Interpretation of Results ................................................................................................... 32

Concluding Remarks ........................................................................................................ 38

Appendix .......................................................................................................................... 41

Works Cited ....................................................................................................................... 47
Acknowledgements

Without the help of others, this project would not have been completed. I would like to thank my advisor, Professor Sean Gailmard, for steering my thoughts in the best direction and for the encouragement to look at things in an alternative way. Much gratitude goes to my parents and family, for their unending support and encouragement. And lastly, thank you to the dozens of taxicab drivers willing to engage in friendly (and sometimes not so friendly) banter about their lives, their problems, their jobs that provided me the initial interest to study an aspect of an industry that few have chosen to look at.
“The road to hell is thick with taxicabs.”
-Don Herold

The Issues

The taxi industry seems simple enough; stick out your hand and someone, for a fee, stops to take you to wherever it is you may need to go. Many daily cab users would be surprised to learn that this industry, which seems to offer service at its most basic level; has actually evolved into a highly complex, highly regulated (in most cases) and often times – highly controversial and highly political segment of city economies. In today’s industry, far more than a car and an entrepreneurial spirit are required to start a taxicab business and begin driving a taxi.

A major topic regarding taxis that faces nearly every city’s transportation commission is the issue of availability. For many, it is not difficult to think of a time when it appeared as if every taxi within a mile radius was occupied, or conversely when multiple cabs caused traffic logjams fighting for commuter business. A crucial task directly linked to this issue that city officials encounter is determining the proper number of taxicabs that their respective cities should allow, in other words - regulation. An undersupply of taxis may “create lengthy waits for cab service and sometimes prevent customers from obtaining service at all.”\(^1\) Likewise, an oversupply “can lead to service problems such as aging and ill-kept cabs and high turnover among underpaid and poorly qualified drivers.”\(^2\)

---

\(^1\) Bruce Schaller. *A Regression Model of the Number of Taxicabs in U.S. Cities*. Schaller Consulting. (November 15, 2004), p. 2.

\(^2\) Ibid.
Actual regulation guidelines that are enacted vary substantially from city to city. In a city like New York, a hard cap is placed on the number of cab licenses that are issued. A more indirect regulation type, like in many Ohio cities, would be minimum requirements on taxi fleet size or disallowing individual ownership of taxis. Regulation can also come in the form of health/safety requirements, price levels set by officials, or minimum quality standards. Even these types of regulations limit the number of taxicabs by self-selecting only those that pass the guidelines. The number of taxicabs in any given city is directly or indirectly related to the regulatory principles set in place by government.

Looking back at the history of taxicab regulation gives a clearer idea as to some of the issues with regulation. The *Economics of Regulation and Antitrust* uses the taxicab industry as a case study in regulation and says that “prior to the 1920’s the taxicab industry was largely unregulated.” It was through the massive unemployment during the US Great Depression that hordes of individuals took to driving taxicabs as a job alternative and “as a result between 1929 and 1937 most cities restricted entry into the taxicab industry.” These regulatory programs, however, “resulted in minimal entry for most taxicab markets.” What ended up happening in fact is that in many major cities, regulation entry restrictions made it such that taxicab numbers stagnated for long periods, often times for decades. This has happened in major taxicab markets such as Chicago, Boston and New York.

---

5 Ibid.
6 Ibid.
Essentially, the number of taxicabs in many of these major cities were set by regulators and left that way for decades following the initial restrictions established during the Great Depression. For instance, it was only recently that New York City increased the number of taxicab licenses it allowed, up from 11,787. This original number had been in place for over 50 years before the city decided to issue additional licenses via first bid, sealed auction. In other cities, regulators claim to base their allotments on metrics such as population and the economy. A few cities, such as Portland, review market trends based on pre-established factors.\(^7\) This does not seem to be the standard throughout US cities.

What seems to be the typical trend in taxicab industries is that regulations are set with little fanfare until further problems from them arise.

An issue that comes out of all these concerns with regulations is one that looks from the government’s perspective in establishing regulations. Rather than debate about the positive or negative consequences of regulations, a different angle would be to attempt to analyze some of the factors that influence these regulations in the first place. Generally, a market is efficient when supply meets demand. At its most basic level, the taxicab industry should be no different; except that in most of these cases supply is not freely determined in the market place. While it would be the most straightforward to create a model that can be used to project the best number of taxicabs for this government regulated supply to meet demand, current data constraints make it difficult to quantitatively measure taxicab demand in US cities; making such a model nearly impossible to accurately create.

What can be done instead is to develop a model to measure the factors that are most explanatory of the government regulated taxicab supply. Specifically, to see what characteristics of cities are relevant for governments when setting whatever type of entry restrictions they choose for their taxicab markets. There has been past expert research conducted on the conceptual ideas behind taxicab demand. By representing these conceptual ideas with hard data and actual numbers, analysis may be able to be conducted that quantitatively answers the type of relationship, or whether or not there is one, between the government regulated number of taxicabs and factors of taxicab demand.

Such a task would not give an end-all ‘x’ number of taxicabs as the precise amount that would resolve supply and demand issues. Instead, the idea is to introduce a quantitative way of looking at the regulations behind the taxicab industry, with the expectation that results will provide additional insight on how governments act. Often times the reasons given to explain government actions, not only in the taxicab industry, are vague and unfulfilling. There is always the lingering question, or the lingering assumption, on ulterior motives behind the restrictions that governments place in markets. In New York, “the brokers and lenders who support the [taxicab] industry are among the biggest political givers in mayoral and City Council races,” but there is yet to be empirical evidence showing that government regulations have little or no consideration of actual market demand. Hopefully this research project can provide some novel approaches in

---

beginning to answer these types of questions concerning government actions; previously looked at primarily through qualitative analysis.
Past Developments

Much of the other extensive taxicab industry research has been conducted by Bruce Schaller. Schaller has done studies of taxicab demand that provide much of the theory behind this project’s conceptual framework of taxicab demand.  

In his study, Schaller analyzes the forces behind taxicab demand for US cities with populations over 100,000. He organizes the determinants of taxi demand into seven factors:

1. City Size
2. Availability and cost of privately owned autos
3. Use of complement to taxicabs
4. Cost of taxi usage
5. Taxi service quality
6. Competing modes
7. Special populations

Through his own analysis, he ultimately comes to the conclusion that taxicab demand in major cities can be explained by the independent variables airport trips, transit ridership and no vehicle households. It is especially important to note that Schaller recognizes that factors other than population are crucial for regulators to consider in determining taxicab numbers.

The original plans for this research project were to employ a similar approach to answer the appropriate questions regarding taxicab demand. However, some of the inherent characteristics behind Schaller’s approach are a cause for concern. The main issue is that

\[9\] Schaller, p. 8.
the data he used considers the number of taxicabs in a city and taxicab demand in that city to be the same amount. With the knowledge of government restrictions in taxicab markets, this seems to be an inappropriate measure of taxicab demand; especially since restrictions to the data set are only limited by population size (only 100,000 or greater), and not by regulation type. Surveys have shown that only small fractions of cities in the US have deregulated taxicab industries, and that the vast majority, nearly 90%, have some form of restricted entry.\textsuperscript{10} If Schaller were to include only the cities with deregulated taxicab markets in his study, there could be claims that taxicab numbers reflect taxicab demand. By including the taxicab numbers of both regulated and deregulated cities in his data set, it seems as if it would be difficult to measure the true demand for taxicabs and that bias may be present in results. Even so, Schaller’s conceptual ideas behind the primary factors of taxicab demand are extremely relevant, and will heavily influence the taxicab demand model used for this research project.

The majority of past and current literature regarding the taxicab industry focuses on the unending debate of regulation and deregulation. The rationale behind many of the arguments reveals the issues that are especially important to industry personnel and in particular, customers.

There are two distinct sides in the regulation versus deregulation debate in current literature. Generally, proponents of regulation focus on the fact that regulation helps maintain a standard of quality, helps control congestion and helps to keep a price/ service

\textsuperscript{10} Kramer and Mellor, p. 5.
balance within the taxicab market.\textsuperscript{11} Establishing quality standards forces suppliers to focus on more than just providing a point-to-point ride; consumer valued characteristics such as cleanliness and friendliness also have to be considered. In many cities, congestion is a significant issue that causes problems for drivers, commuters and everyone in between. Limiting taxicab numbers controls for the increase in taxis that is generally seen with deregulation. Without regulation, not only would taxicab numbers fluctuate, the fares offered would fluctuate as well. Regulators have argued that the oversupply of taxicabs under deregulation would lead to increased fares to reduce the loss of profits from a reduced occupancy rate.\textsuperscript{12}

Conversely, advocates of deregulation focus on the fact that regulation limits competition and gives power to a few, eventually promoting corruption within the industry. Regulations that limit the number of cabs in a city, or give fleet minimums to start a taxicab business favor existing drivers. Ironically, arguments have also been made that with deregulation the competitive environment would be such that “fares are more likely to fall than to rise.”\textsuperscript{13} Increased competition, it is claimed, would provoke drivers to find new and creative ways to attract business within the existing industry framework. Finding better ways to control costs would be one of the alternatives. Often times however, the greatest concern that supporters of deregulation have is simply the fact that “grounds for restricting entry are generally not well articulated.”\textsuperscript{14} A study of the Australian taxicabs market resulted in conclusions that it was ambiguous as to why

\textsuperscript{12} Ibid.
\textsuperscript{13} Productivity Commission. \textit{Regulation of the Taxi Industry}. Commonwealth of Australia. (1999), p. XI.
\textsuperscript{14} Ibid.
restrictions on the markets were made, as the regulations seemed to have no identifiable benefits to the community.\textsuperscript{15}

Understanding some of the common themes behind the regulation and deregulation arguments shows just how complex the regulatory framework of the taxicab industry has become. A number of other papers allude to the idea that, “consumer welfare has been the principal stated objective of taxi regulation…but its remit has broadened over time.”\textsuperscript{16} In much of the literature reviewed, many seemed to question the effects that various regulations had, yet no one seemed to focus on the underlying purposes and whether or not the number of taxicabs actually “closely mirror users’ needs.”\textsuperscript{17} These debates over regulation provide an initial idea of the concepts that influence consumer demand for taxicabs. In particular, service quality, balanced fares and traffic are concerns that people have when looking at taxicabs as a transportation option. Also seen are debates over the effectiveness of regulation, and whether or not regulation is actually beneficial for the community. This further introduces the idea of looking at not only the regulation itself, but what influences certain regulatory measures.

\textsuperscript{15} Productivity Commission. \textit{Regulation of the Taxi Industry}. Commonwealth of Australia. (1999), p. XI.
\textsuperscript{16} Benedikt Koehler. \textit{Regulating Supply in Taxi Markets}. City University Department of Economics. (September 2004), p. 5.
\textsuperscript{17} Productivity, p. XI.
To refresh, the goal here is to empirically analyze the relationship between the government regulated number of taxicabs and taxicab demand. The theory behind the approach is that by using a dependent variable that is essentially controlled, indirectly or directly, by the government, it is possible to analyze how government actions relate to certain independent variables. In particular, independent variables that reflect an explanatory model of taxicab demand. This model itself is based off of factors that draw from a number of different sources from past research and encompasses the current views in literature on market demand. Admittedly, there are data constraints that limit the explanatory power of the model. Specifically, there is no data available on many service characteristics such as cleanliness, driver courtesy, speed, etc. that undoubtedly affect demand. Thus, the model employed for this research project can and will only include variables for which accurate data is readily accessible.
Based on the Transportation, Limousine and Paratransit Association’s (TLPA) annually distributed Fact Book, the metric average fares across US cities can be included as an independent variable. The TLPA provides the drop charges, additional mile charges, etc. from which an average five-mile fare can be calculated for all the cities in the study. The theory behind cost and demand is straightforward. Since there is no standard guideline for fare, it would not be surprising to see areas with higher fares experience less demand for taxicabs, as alternative transportation options are more attractive and perhaps more affordable.

Two of the measures borrowed from Schaller include people without vehicles and public transportation users. Transportation theory alludes to the fact that people without vehicles are one of the groups considered to be heavy users of taxicabs.\textsuperscript{18} Public transportation, it has been shown, is typically a strong complement to taxicab usage. Here, the variable encompasses all forms of public transportation, and not just the use of transit (subway). Data for both these variables are borrowed from the US Census year 2000 and are measured on a city basis as the number of people without vehicles and the number of daily users of public transportation.

Climate is a more obscure measurement that is also going to be considered, and is also briefly alluded to by Schaller. There seems to be a trend that links taxicab usage with varying climates. In particular, cold weather temperatures incite a greater desire to travel via taxicab versus other means which increase exposure to the elements. Climate will

\textsuperscript{18} Schaller, p. 4.
focus particularly on places with colder average temperatures. A dummy variable for cities with a lowest monthly average below 32 degrees (freezing) will be used. A direct temperature measure is not used because it seems unlikely that variations in temperature, when the temperature is already below a certain threshold, will have the same effects on taxicab demand. Essentially, the assumption is that below the freezing point, being even colder will not influence taxicab demand. Data is taken from weather.com.

**Age** is a factor that can be related to taxicab demand. As age increases past a limit, it seems as if the ability to drive as safely decreases. Evidence of this is seen in the fact that after a certain age, the likelihood to be involved in a traffic accident increases dramatically.\(^{19}\) Transportation users falling into this category may be more likely to turn to taxicabs as a smarter form of transportation. The threshold that is going to be considered is the number of people living in a city age 65 and above, from the 2000 US Census.

**Population density**, rather than simply population, is going to be used as a metric. It seems as if the concentration of people in a specific area would be more relevant in measuring demand in taxicabs. If population changes due to an expansion of the geographic boundaries for a given location, there is no reason that the demand for taxicabs would change along with it. Looking at Census data for land area, city sizes change from Census-to-Census. Using population density ensures that this inaccuracy will not occur. The 2000 Census provides the pertinent data.

\(^{19}\) Kramer and Mellor, p. 8.
An interesting metric that the 2000 Census also takes is the **average travel time to/from work** for people living across the US. The effect of travel time on taxicab demand seems relevant but ambiguous. A higher average travel time may discourage people to drive themselves to work as it reflects congestion in the city and a general inconvenience with driving to work. However, a higher average travel time could also discourage people from taxicabs as a transportation option due to higher costs and better options elsewhere.

**Lower income households** are generally thought of to be more reliant on taxicabs as a mode of transportation. Areas consisting of higher populations of lower income households generally lack transportation options other than personal vehicle ownership, especially in smaller-sized cities.\(^{20}\) The number of people in the labor force with incomes below $35,000 per year is used to reflect lower income households. Data is from the 2000 Census.

A measurement that has been seen numerous times in past research is **airport taxi trips**. In some US cities, the taxicab trips to and from airports are a significant part of the taxicab economy and can comprise up to one third of business.\(^{21}\) Thus, when considering taxicab demand, the number of airport trips is a crucial factor that needs to be included. Unfortunately, the importance of the factor is not reflected with straightforward and easy-to-obtain data; as there is no specific data set that details the number of taxi trips to and from airports for a large sample of US cities and airports. The closest metric available is from the Federal Aviation Administration (FAA) and measures overall airport traffic,\(^{20}\)Allred, Saltzman and Rosenbloom. *Factors Affecting the Use of Taxicabs by Lower Income Groups*. Transportation Research Record. (1978), p. 23.
\(^{21}\)Schaller, p. 4.
specifically, enplanements for a wide range of US airports. The assumption is that airport traffic and taxicab users to and from airports is directly related. Because of the difficulty in allocating accurate airport traffic quantities to the proper cities, it was decided that the best method is to simplify the entire approach and use the relative size of airports as the important measurement. Only cities in the direct vicinity of the airports listed by the FAA will be considered. For cities in direct vicinity of airports with between a half million and three million enplanements in the year 2000, the dummy variable is 1. For cities with over three million enplanements, the dummy variable is 2. All other cities have dummy variable 0. The logic is that such a system recognizes relative amounts of airport traffic and that cities with larger relative sizes have more airport traffic and hence, more taxicab demand.

The only dependent variable is the number of taxicabs in a city, from the Transportation, Limousine and Paratransit Association’s (TLPA) annually distributed Fact Book 2000. The major cities that are known to have deregulated taxicab industries are omitted from the data set. There is a possibility that some of the smaller, deregulated cities will still be included; however, it is unlikely that inclusion of such figures will have much effect. As mentioned before, studies have shown that nearly 90% of cities with taxicabs also have entry restrictions in place.

Overall, there are 259 cities in the sample size. California is the most represented state, with 24 cities, whereas Vermont is the least represented, with none. The year 2000 was chosen because it is the most recent year with Census data. All of the data was combined
first in an Excel spreadsheet, before being transferred over to STATA for the purposes of analysis.

Regression analysis is the best way to find the relationship between the dependent variable \((taxis)\) and independent variables \((avg\_cost, \ avg\_trv\_tim, \ lo\_avg\_temp, \ age\_65p, \ pub\_tr\_n, \ veh\_\_n, \ pop\_d\_\_s, \ hhi\_lo\_35 \ and \ enplane)\). Results from regressing the dependent variable on the independent variables will reveal which variables are statistically significant in explaining the government regulated number of taxicabs. Looking at the two sided alternative at a 5% significance level, a t-statistic greater than 1.96 or less than -1.96 would be significant.

**Variable Descriptions:**

\textbf{taxis}: the number of taxis for cities included in sample (measured in taxis)

\textbf{avg\_cost}: average cost of a 5 mile fare (measured in dollars)

\textbf{lo\_avg\_temp}: lowest monthly average low for given city, dummy variable 1 if below freezing point, 0 if above freezing point

\textbf{age\_65p}: the number of people age 65 and over (measured in thousands)

\textbf{pub\_tr\_n}: the number of people taking public transportation (measured in thousands)

\textbf{veh\_\_n}: the number of people without personal vehicles (measured in thousands)

\textbf{pop\_d\_\_s}: population divided by land area (measured in thousands per square mile)

\textbf{hhi\_lo\_35}: people with household income below $35,000 per year (measured in thousands)

\textbf{enplane}: the number of people using airport, dummy variable 2 if more than 3 million, 1 if between 500,000 and 3 million, 0 otherwise
## Summary Statistics:

<table>
<thead>
<tr>
<th></th>
<th>259 Obs.</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxis</td>
<td>271.98</td>
<td>902.48</td>
<td>2.00</td>
<td>12187.00</td>
</tr>
<tr>
<td>Average Cost (avg_cost)</td>
<td>11.25</td>
<td>2.83</td>
<td>4.75</td>
<td>27.75</td>
</tr>
<tr>
<td>Airport Enplanements (enplane)</td>
<td>2434.26</td>
<td>6086.75</td>
<td>2.70</td>
<td>40905.42</td>
</tr>
<tr>
<td>Population Density (pop_dens)</td>
<td>2.10</td>
<td>1.81</td>
<td>0.01</td>
<td>16.41</td>
</tr>
<tr>
<td>Lowest Monthly Temperature (lo_avg_temp)</td>
<td>25.35</td>
<td>13.70</td>
<td>-13.00</td>
<td>65.00</td>
</tr>
<tr>
<td>People Age 65+ (age65p)</td>
<td>24.88</td>
<td>68.48</td>
<td>0.04</td>
<td>937.86</td>
</tr>
<tr>
<td>Public Transportation Users (pubtran)</td>
<td>13.56</td>
<td>107.44</td>
<td>0.00</td>
<td>1684.85</td>
</tr>
<tr>
<td>Average Travel Time to Work (avgtrvtim)</td>
<td>20.91</td>
<td>5.01</td>
<td>8.90</td>
<td>40.00</td>
</tr>
<tr>
<td>People with Household Income below $35K (hhi_lo_35)</td>
<td>39.46</td>
<td>106.44</td>
<td>0.25</td>
<td>1400.92</td>
</tr>
<tr>
<td>People without Vehicles (vehicleno)</td>
<td>18.00</td>
<td>108.00</td>
<td>0.05</td>
<td>1682.95</td>
</tr>
</tbody>
</table>
The regression of the government regulated number of taxicabs \((taxis)\) on the variables of taxicab demand \((avg\_cost,\ avgtrvtim,\ lo\_avg\_temp,\ age65p,\ pubtran,\ vehicleno,\ pop\_dens,\ hhi\_lo\_35\ and\ enplane)\) shows that five of the nine independent variables are significant. As can be seen from abbreviated results in the table below (full results for all regressions can be found in the appendix, abbreviated results are rounded), average travel time to work \((avgtrvtim)\), public transportation users \((pubtran)\), people without vehicles \((vehicleno)\), people with household income below $35,000 \((hhi\_lo\_35)\) and airport enplanements \((enplane)\) all have t-statistics greater than 1.96 or less than -1.96, indicating statistical significance.

<table>
<thead>
<tr>
<th>Taxis</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Cost ((avg_cost))</td>
<td>-3.055</td>
<td>6.080</td>
<td>-0.50</td>
</tr>
<tr>
<td>Airport Enplanements ((enplane))</td>
<td>107.713</td>
<td>27.947</td>
<td>3.85</td>
</tr>
<tr>
<td>Population Density ((pop_dens))</td>
<td>19.474</td>
<td>13.568</td>
<td>1.44</td>
</tr>
<tr>
<td>Lowest Monthly Temperature ((lo_avg_temp))</td>
<td>-42.194</td>
<td>40.544</td>
<td>-1.04</td>
</tr>
<tr>
<td>People Age 65+ ((age65p))</td>
<td>-0.061</td>
<td>2.989</td>
<td>-0.02</td>
</tr>
<tr>
<td>Public Transportation Users ((pubtran))</td>
<td>11.888</td>
<td>2.748</td>
<td>4.33</td>
</tr>
<tr>
<td>Average Travel Time to Work ((avgtrvtim))</td>
<td>16.884</td>
<td>4.302</td>
<td>3.92</td>
</tr>
<tr>
<td>People with Household Income below $35K ((hhi_lo_35))</td>
<td>3.975</td>
<td>1.754</td>
<td>2.27</td>
</tr>
<tr>
<td>People without Vehicles ((vehicleno))</td>
<td>-8.278</td>
<td>3.052</td>
<td>-2.71</td>
</tr>
<tr>
<td>Constant</td>
<td>-293.025</td>
<td>119.643</td>
<td>-2.45</td>
</tr>
</tbody>
</table>

It seems as if this would infer that government decisions in regulating taxicab numbers for the cities in the sample can be explained by the city characteristics: airport enplanements, public transportation users, average travel time to work, low income households and people without vehicles. It appears from these results that governments, whether directly or indirectly, consider a substantial number of taxicab demand factors when allocating taxicabs for use in their cities. However, it seems odd that the coefficient on the people without vehicles \((vehicleno)\) variable is actually negative, indicating that based on current model results, for every 1,000 people without
automobiles governments tend to decrease taxicabs by approximately 8. This contrasts with the fact that coefficients show positive relationships between low income persons and taxicab numbers and public transportation users and taxicab numbers. This result differs from the previously established theory on taxicab demand. However, it is possible that this shows that government restrictions ignore the number of no vehicle persons in their decision making process and statistical significance was a matter of coincidence.

Before coming to such a bold conclusion, there are other diagnostic tests that can be run to ensure the accuracy of the regression results. In particular, multicollinearity seems as if it would be a factor within our independent variables. This is when there is high correlation between the independent variables such that it can affect estimation power.

The variance inflation factor (VIF) is a common diagnostic test for multicollinearity. If results show a high VIF, or likewise a low tolerance (1/VIF) for any independent variable, it is indicative of multicollinearity.

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
<th>1/VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>People without Vehicles (vehicleno)</td>
<td>388.57</td>
<td>0.003</td>
</tr>
<tr>
<td>Public Transportation Users (pubtran)</td>
<td>311.94</td>
<td>0.003</td>
</tr>
<tr>
<td>People Age 65+ (age65p)</td>
<td>149.86</td>
<td>0.007</td>
</tr>
<tr>
<td>People with Household Income below $35K (hhi_lo_35)</td>
<td>124.63</td>
<td>0.008</td>
</tr>
<tr>
<td>Population Density (pop_dens)</td>
<td>2.17</td>
<td>0.461</td>
</tr>
<tr>
<td>Airport Enplanements (enplane)</td>
<td>1.67</td>
<td>0.598</td>
</tr>
<tr>
<td>Average Travel Time to Work (avgtrvlim)</td>
<td>1.66</td>
<td>0.602</td>
</tr>
<tr>
<td>Lowest Monthly Temperature (lo_avg_temp)</td>
<td>1.20</td>
<td>0.831</td>
</tr>
<tr>
<td>Average Cost (avg_cost)</td>
<td>1.05</td>
<td>0.949</td>
</tr>
</tbody>
</table>

High VIF indicates that people without vehicles (vehicleno), public transportation users (pubtran), people age 65+ (age65p) and people with household income below $35,000 (hhi_lo_35) are at likelihood for multicollinearity. Looking at the correlations between these four independent variables shows that people without vehicles (vehicleno) and public transportation users (pubtran) (.9964), people age 65+ (age65p) and people with
household income below $35,000 ($hhi_lo_35$) (.9935) are pairs of variables that are very highly correlated – they essentially measure the same thing (Figure 1c). Of the other permutation pairs, public transportation users ($pubtran$) and people with household income below $35,000 ($hhi_lo_35$) are the least correlated. An option to reduce multicollinearity would be to reconstruct the model by dropping certain independent variables. Here the best candidates to omit would be people age 65+ ($age65p$) and people without vehicles ($vehicleno$), as they are fundamentally represented by people with household income below $35,000 ($hhi_lo_35$) and public transportation users ($pubtran$), respectively, in the model. This leaves us with the least correlated pair: people with household income below $35,000 ($hhi_lo_35$) and public transportation users ($pubtran$).

### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Cost (avg_cost)</td>
<td>-2.944</td>
<td>6.138</td>
<td>-0.48</td>
</tr>
<tr>
<td>Airport Enplanements (enplane)</td>
<td>105.195</td>
<td>28.057</td>
<td>3.75</td>
</tr>
<tr>
<td>Population Density (pop_dens)</td>
<td>23.621</td>
<td>13.551</td>
<td>1.74</td>
</tr>
<tr>
<td>Lowest Monthly Temperature (lo_avg_temp)</td>
<td>-59.035</td>
<td>40.011</td>
<td>-1.48</td>
</tr>
<tr>
<td>Public Transportation Users (pubtran)</td>
<td>4.670</td>
<td>0.410</td>
<td>11.39</td>
</tr>
<tr>
<td>Average Travel Time to Work (avgtrvtim)</td>
<td>16.744</td>
<td>4.345</td>
<td>3.85</td>
</tr>
<tr>
<td>People with Household Income below $35K ($hhi_lo_35$)</td>
<td>2.682</td>
<td>0.442</td>
<td>6.07</td>
</tr>
<tr>
<td>Constant</td>
<td>-288.048</td>
<td>120.926</td>
<td>-2.38</td>
</tr>
</tbody>
</table>

In the reconstructed model, $taxis$ is regressed on seven independent variables ($avg_cost$, $avgtrvtim$, $lo_avg_temp$, $pubtran$, $pop_dens$, $hhi_lo_35$ and $enplane$). Results show statistical significance for airport enplanements ($enplane$), public transportation users ($pubtran$), average travel time to work ($avgtrvtim$) and people with household income below $35,000 ($hhi_lo_35$). The t-statistics are much stronger in this model. Checking again for multicollinearity, VIF is substantially lower, and does not indicate a high risk for multicollinearity (Figure 2b). R-squared values for the two models have high explanatory power and are approximately the same, .9146 for the first and .9120 for the reconstructed model.
Looking at a graph of the residuals from the regression of $taxis$ on the independent variables ($avg\_cost$, $avgtrvtim$, $lo\_avg\_temp$, $pubtran$, $pop\_dens$, $hhi\_lo\_35$ and $enplane$) makes it appear as if there are variations in the residuals as well as outliers in the observations. The residuals are concentrated at low fitted values and then become more and more dispersed at higher fitted values.

Unequal variance in the residuals is typically indicative of heteroskedasticity, and is confirmed using the Breusch-Pagan/ Cook-Weisberg test (Figure 2c). One method to correct for heteroskedasticity would be to use robust standard errors.

After correcting for heteroskedasticity, the independent variable people with household income below $35,000$ ($hhi\_lo\_35$) is no longer significant. Instead, statistical significance is only seen in airport enplanements ($enplane$), public transportation users ($pubtran$) and average travel time to work ($avgtrvtim$). The t-statistics for airport enplanements ($enplane$) and average travel time to work ($avgtrvtim$) are stronger. Essentially, after correcting for both multicollinearity and heteroskedasticity, the only measures of taxicab demand that are significant in the
government regulated taxicab numbers are the number of enplanements, public transportation users and average travel time to work.

A startling observation from the graph of residuals is the outlying observation of New York City, seen at the far bottom right of the graph. Comparing summary statistics for the data set with New York City and without New York City shows that the impact of including NYC is large. Less New York City, the mean and standard deviation of nearly every variable decreases, even with a sample size of over 250 US cities. For example, the mean of public transportation users (pubtran) drops from 13.56 to 7.09, while its standard deviation drops from 107.44 to 38.11. In an attempt to correct for any biases that may exist with the inclusion of New York City in the data, a regression will also be run without the New York City observations.

<table>
<thead>
<tr>
<th>Taxis</th>
<th>Coefficient</th>
<th>Robust Std. Error</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Cost (avg_cost)</td>
<td>-3.602</td>
<td>3.360</td>
<td>-1.07</td>
</tr>
<tr>
<td>Airport Enplanements (enplane)</td>
<td>110.598</td>
<td>44.419</td>
<td>2.49</td>
</tr>
<tr>
<td>Population Density (pop_dens)</td>
<td>-10.746</td>
<td>20.169</td>
<td>-0.53</td>
</tr>
<tr>
<td>Lowest Monthly Temperature (lo_avg_temp)</td>
<td>-65.249</td>
<td>42.106</td>
<td>-1.55</td>
</tr>
<tr>
<td>Public Transportation Users (pubtran)</td>
<td>12.756</td>
<td>3.504</td>
<td>3.64</td>
</tr>
<tr>
<td>Average Travel Time to Work (avgtrvim)</td>
<td>15.497</td>
<td>3.438</td>
<td>4.51</td>
</tr>
<tr>
<td>People with Household Income below $35K (hhi_lo_35)</td>
<td>0.769</td>
<td>1.359</td>
<td>0.57</td>
</tr>
<tr>
<td>Constant</td>
<td>-174.035</td>
<td>75.848</td>
<td>-2.29</td>
</tr>
</tbody>
</table>

Regressing taxis on the seven independent variables (avg_cost, avgtrvim, lo_avg_temp, pubtran, pop_dens, hhi_lo_35 and enplane) with
robust standard errors gives the same three significant independent variables as when New York City was included. However, in this case, the coefficients and t-statistics for the variables are stronger overall, while overall the standard errors also decrease. Results show that for every 1,000 public transportation users, there 12.8 taxicabs present. Every minute of travel time to work adds 15.5 taxicabs. For cities with between 500,000 and 3,000,000 enplanements per year, there are 110.6 taxicabs added, and for cities with over 3,000,000 enplanements per year there are 221.2 more taxicabs.

**Additional Remarks and Results**

Addressing the issues of heteroskedasticity, multicollinearity and outliers seems as if it would correct for the more obvious biases inherent in the model. However, there is another observation that can be made from the fitted versus residuals plot (Figure 2d) that must not be so easily overlooked. Besides the unequal variances in the residuals that are indicative of heteroskedasticity, an argument can be made that the residuals also exhibit a noticeable downward pattern. That is, the greater the fitted value, the lower the residual.

The issue can be deduced logically. The fundamental definition of a residual is that it is the difference between the actual observation and the observation predicted (fitted) by the explanatory independent variables. If the dependent variable is very well explained, the residuals would generally be scattered randomly about zero, and not in a pattern. A pattern in the residuals typically means that there may be missing independent variables in the model that would eliminate the pattern, by ‘explaining away’ some of the residual. The problem with missing independent variables is that they may cause biases in the regression results.
The most straightforward approach to try and eliminate biases due to omitted variables is to simply add more independent variables to the model. The hope is that they will reduce some of the pattern in the residuals. This method will be taken here to reduce the potential biases of missing variables. The conceptual, generally agreed upon factors of taxicab demand were already included in the original regression model. Thus, the additional independent variables to be included are largely of the author’s own volition.

In all, seven more ‘experimental’ independent variables will be added to the reconstructed model. Because this regression takes on a more experimental approach rather than simply testing factors provided by past research, the effects of multicollinearity on the regression model will be put aside. The ‘experimental’ independent variables and their theoretical reasons for affecting taxicab demand are explained below. Following along the lines from our previous method, these factors, although not nearly as reviewed in current taxicab literature, are assumed to have an influence on taxicab demand. The goal then is to see whether or not governments consider these experimental variables when setting taxicab numbers via regulations. All data is from the Census year 2000.\textsuperscript{22}

Three of the chosen factors, \textbf{people who drive to work alone, people who bike/walk for transportation} and \textbf{car owners} take into consideration the alternative forms of transportation that are available in cities. Since a large percentage of taxicab trips are taken to and from work,\textsuperscript{23} people who drive to work alone would have a negative effect

\textsuperscript{22} Sioux Falls, Midland, San Angelo and Paris are omitted due to lack of data availability
\textsuperscript{23} Schaller, p. 4.
on taxicab demand. A person who drives to work may also indicate less of likelihood that taxicabs are used as the transportation means in other activities outside of work. Car owners may still use taxicabs on occasion, but it is likely that the purchase of a car means that the opportunity to use other forms of transportation such as taxicabs is reduced. A large number of walkers/bikers may show that taxicabs are not a desired form of transportation, or conversely it could indicate that the city is compact enough that taxicabs may be a relatively inexpensive form of transportation.

Another factor looks at the type of worker found in cities. It is believed that white collar workers are heavier users of taxicabs. These types of jobs are typically located in more urban environments where parking may be unavailable and driving may be inconvenient. The number of blue collar workers is the metric that is going to be used to measure worker type. This encompasses those employed in the agriculture, forestry, hunting, mining, manufacturing, construction and transportation industries. The nature of blue collar jobs makes it less likely that these types of workers would rely on taxicabs as a form of transportation. The types of industries for blue collar workers are generally in environments where it is easier to drive a car, rather than use other forms of transportation.

A higher number of seasonal households may indicate that a greater portion of the population is impermanent. Visitors or tourists are more likely to rely on alternative forms of transportation as their main mode and taxicabs would typically be higher in demand.
Studies have also shown that minorities are heavily dependent on taxicabs.\textsuperscript{24} Areas with higher minority populations generally have less access to public transportation and less vehicle ownership, leading to greater demand for taxicabs as a form of transport. This metric will be measured by the number of non-Caucasian residents in any given city.

The type of city, whether urban or rural, is another factor that can be considered when looking at an expanded model of the factors of taxicab demand. Urban areas generally have greater dependence on taxicabs, as other forms of transportation can prove to be both expensive and inconvenient. Characteristics of rural areas such as open, expansive land area make public transportation and taxicabs impractical in most cases. Urban environments typically consist of entertainment, shopping, and office buildings in a compacted area, which are the locations taxicabs are most often driven to and from.\textsuperscript{25} As measured by the urban population of a city, there should be more taxicabs in the areas with higher urban populations.

It is important to remember that these additional factors chosen by the author to be included in the expanded regression model were chosen not only for relevance, but also for data availability. Taxicab demand is also heavily influenced by factors that have not been empirically measured, particularly service and quality type aspects. The hope still is that by including these additional measures the bias from omitted variables can be reduced.

\textsuperscript{24} Kramer and Mellor, p. 8.
\textsuperscript{25} Schaller, p. 6.
Variable Descriptions:

**drv_alone**: the number of people who drive alone to work (measured in thousands)

**bike_walk**: the number of people who bike/walk for transportation (measured in thousands)

**blue_collar**: blue collar workers (measured in thousands)

**_seas_hh**: the number of seasonal households (measured in households)

**urban_pop**: urban population (measured in thousands)

**car_van**: the number of car or van owners (measured in thousands)

**no_white**: the number of people non-Caucasian (measured in thousands)

Summary Statistics:

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>255 Obs.</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drive to Work Alone (** drv_alone**)</td>
<td></td>
<td>61.12</td>
<td>111.63</td>
<td>0.37</td>
<td>982.74</td>
</tr>
<tr>
<td>Bike/Walk for Transportation (** bike_walk**)</td>
<td></td>
<td>5.51</td>
<td>23.28</td>
<td>0.01</td>
<td>347.29</td>
</tr>
<tr>
<td>Blue Collar Workers (** blue_collar**)</td>
<td></td>
<td>39.00</td>
<td>96.79</td>
<td>0.50</td>
<td>1141.80</td>
</tr>
<tr>
<td>Season Households (<strong>_seas_hh</strong>)</td>
<td></td>
<td>686.11</td>
<td>2041.92</td>
<td>0.00</td>
<td>28157.00</td>
</tr>
<tr>
<td>Urban Population (** urban_pop**)</td>
<td></td>
<td>216.04</td>
<td>616.31</td>
<td>0.00</td>
<td>8008.28</td>
</tr>
<tr>
<td>Car Owners (** car_van**)</td>
<td></td>
<td>73.24</td>
<td>138.25</td>
<td>0.45</td>
<td>1203.14</td>
</tr>
<tr>
<td>Non-Caucasian (** no_white**)</td>
<td></td>
<td>45.17</td>
<td>236.96</td>
<td>0.14</td>
<td>3245.16</td>
</tr>
</tbody>
</table>

The expanded model regressed the dependent variable (** taxis**) on 14 different independent variables (** avg_cost, enplane, pop_dens, lo_avg_temp, pubtran, avgtrvtim, hhi_lo_35, drv_alone, bike_walk, blue_collar, _seas_hh, urban_pop, car_van, no_white **), combining
the reconstructed model with seven new independent variables. The results were a bit startling. Of the 14 independent variables, only average cost (avg_cost), lowest monthly temperature (lo_avg_temp) and public transportation users (pub_tran) were found to be statistically insignificant. The others all had t statistics either greater than 1.96 or less than -1.96.

Using robust standard errors for heteroskedasticity reduces the number of statistically significant variables. Here, average travel time to work (avgtrvtim), drive to work alone (drv_alone), bike/walk for transportation (bike_walk), blue collar workers (blue_collar), seasonal households (_seas_hh), urban population (urban_pop) and car owners (car_van) are all significant. Airport enplanements (enplane) has a t-statistic of 1.94 and is barely insignificant when robust standard errors are used. The expanded model shows that governments consider a number of population characteristics when determining the restrictions on taxicab supply.

More importantly, the expanded model exhibits a residuals-versus-fitted plot that has much less of a pattern than the plot generated by the previous models. There is much less
of a downward sloping trend, and the residuals seem to be more scattered about zero.

Unequal variance in the residuals is still seen, and is corrected by using robust standard errors. Overall, there is a noticeable improvement in the residuals plot when including a wider variety of variables into the previously used regression models. This, however, does not mean that biases have been eliminated. The plot provides evidence that biases resulting from omitted variables may have been reduced and that future research may need to include an even greater variety of variables. However, with the current problems behind data constraints, it is not possible to include all relevant variables, simply more than before. Yet, it is still imperative to consider this expanded model in addition to the previous models, as there has been improvement made on the problems arising from missing taxicab demand variables.

The initial model indicated that five of the nine factors of taxicab demand were statistically significant when considered against the government regulated number of taxicabs for the 259 cities in the sample. Further testing indicated, however, that this preliminary model suffered from multicollinearity. By eliminating highly correlated independent variables, the issue was resolved. An observation of a graph of residuals indicated that the reconstructed model, after dropping people age 65+ (age65p) and people without vehicles (vehicleno), may have unequal variance in its residuals – heteroskedasticity. The use of robust standard errors corrected for this problem and
eliminated people with household income below $35,000 (hhi_35_lo) as a significant variable. Eliminating the New York City observation from the data did not change the significant variables, but made coefficients and t-statistics stronger overall. It was eventually determined that airport enplanements (enplane), public transportation users (pubtran) and average travel time to work (avgtrvtim) were the only statistically significant variables from the taxicab model of demand that helped explain the government regulated number of taxicabs in US cities.

Further observation of the residuals-versus-fitted plot led to the conclusion that the downward pattern in the residuals was a probable result of missing independent variables. The expanded model exhibited a residuals-versus-fitted plot that showed much improvement. The residuals seem more scattered and less patterned, indicating that some of the previous pattern may have been ‘explained away’ by the inclusion of additional relevant variables.

Results from the expanded model showed that many of the newly included independent variables were statistically significant to the government restricted number of taxicabs. After correcting for heteroskedasticity, these included average travel time to work (avgtrvtim), drive to work alone (dry_alone), bike/walk for transportation (bike_walk), blue collar workers (blue_collar), seasonal households (_seas_hh), urban population (urban_pop) and car owners (car_van). From the previous model, only average travel time to work (avgtrvtim) is still statistically significant – for every minute increase, there are 8.9 taxicabs. As the number of people who drive alone to work and bike/walk for transportation increases by 1,000, the number of taxicabs decreases by 26.5 and 39.3,
respectively. Likewise, for every 1,000 of those who live in urban areas, there are 15.3 more taxicabs and for every 1,000 blue collar workers, 11.5 less taxicabs. For every one seasonal household in a city, there are .073 more taxicabs. The only odd result is with car owners, as results show that for every 1,000 car owners, there are 11.4 more taxicabs. However, it is possible that governments see increases in car ownership as demand for more transportation in general and thus alter taxicab numbers accordingly. Such an reason could also be used to explain why more time spent in traffic to and from work results in increases in the number of taxicabs. Perhaps governments are predicting greater future taxicab demand and correcting for it in the present; as increases in car ownership and traffic congestion may lead to a greater desirability of transportation options with more convenience and less responsibility. This is simply a theory as to why the coefficient on car owner (\textit{car\_van}) is not what is expected. Beyond the results it is impossible to know the actual reasons why some of the independent variables are significantly related to the government restricted taxicab supply and why some, which may seem to be important, are not.
Interpretation of Results

The uniqueness of this research project comes from interpretation of the results. Most typical regression models are set up such that their results reveal how the independent variables influence the dependent variables. The basic model employed here takes this rudimentary idea behind regressions a step further. By choosing a dependent variable that is ‘controlled’ by a third party, in this case the government, and by selecting independent variables that form an encompassing model – taxicab demand, the influence of the independent variables on the dependent variable reveals whether or not the third party acts in accordance with the encompassing model. Specifically, the regressions can be used to infer how governments act; from the relationship (or lack thereof) between the government controlled metric and chosen independent variables. It is an indirect way to quantify the reasoning behind how governments establish certain restrictions or regulations. Applying the results found in the previous section will make this concept clearer.

The results section revealed that three independent variables from the model of taxicab demand were statistically significant to the government regulated number of taxicabs in US cities. The three independent variables from the reconstructed model were airport enplanements (enplane), public transportation users (pubtran) and average travel time to work (avgtrtlim). These results of significance, unfortunately, do not say anything about taxicab demand. One caveat of the model is that the independent variables are automatically assumed to have effects on taxicab demand, based on past theory. The dependent variable, the number of taxicabs, is a directly, or indirectly controlled factor.
that does not necessarily reflect taxicab demand. Instead, it simply reflects the number of
taxicabs in a managed atmosphere. Thus the significant independent variables are not
what affects taxicab demand, but instead are the taxicab demand factors that are
significant in the government establishing the number of taxicabs in a city. So overall for
the 259 US cities in the sample size, airport traffic, public transportation users and traffic
congestion are statistically significant factors to the government controlled number of
taxicabs.

Less of a straightforward interpretation of the expanded model can be taken. The nature
of the model makes the assumption that the factors of taxicab demand – the independent
variables, actually have an influence on taxicab demand. In the original model the
independent variables consisted of the concepts agreed upon in transportation literature.
The expanded model added variables of the author’s preference and are only theorized to
have an effect on taxicab demand. And while these variables seem to have reduced some
of the bias resulting from omitted factors, whether they actually influence taxicab
demand cannot be known for sure.

However, with the assumption that the added independent variables are actually factors
of demand, results reveal that in the expanded model average travel time to work
(avgtrvtim), drive to work alone (drv_alone), bike/walk for transportation (bike_walk),
blue collar workers (blue_collar), seasonal households (_seas_hh), urban population
(urban_pop) and car owners (car_van) were related to the government regulated number
of taxicabs. None of the statistically insignificant variables from the previous model
were found to be significant in the expanded model. Of note is that the measure for people non-Caucasian was also found to be insignificant.

One of the underlying hopes was that the atypical approach to this research project would expose unique results. In particular, results that look at ideas on what drives government decisions and who exactly these decisions are made for; essentially to see whether or not governments serve public interest. An issue like this is literally impossible to answer on a macro level. This is where the taxicab industry is relevant. Of primary interest is the fact that, for the most part, the taxicab industry in the US is highly regulated by the government. Thus the number of taxicabs in most US cities is also a regulated number. By comparing this to some measure of ‘public interest,’ it may reveal whether or not governments, on a micro-level, act for the social welfare of its constituents. This takes the previous analysis a step further. Before, the plan was to see how government regulated taxicab numbers compared to measures of taxicab demand; now the idea is to see how it measures up to ‘public interest.’ With how the original model was developed, as much can be said about what factors were found to be insignificant as can be said about the significant variables.

The Theory of Social Dependency introduces a way to measure the effectiveness of service based industries in serving ‘public interest.’ The key point of this theory is that the social welfare of service industries is gauged by how well they measure up to the social groups most reliant, or most dependent on them. For taxicabs, the most dependent social groups are by design a subset of the original factors of taxicab demand. Of the
nine independent variables from the original model, it seems as if 5 of them are representative of groups of people that are dependent on taxicabs. For now, the independent variables from the expanded model will not be included, as those measures have not yet been widely reviewed in literature. As a Theory, it seems as if the more agreed upon factors should take precedence.

Logical connections can be made with four of the five social groups. It is not difficult to see why people without vehicles, people going to/from the airport, people using public transportation or people of a certain age may strongly depend on taxicabs. These categories are typical of the standard taxicab customers in any town. A primary concern people will have with this theory is in including lower income people as socially dependent on taxicabs. Taxicabs are generally thought of as a convenience for the middle and upper class. Because of this, lower income people are not thought of to be heavy taxicab users, are not socially reliant and seem as if they cannot be included in a measurement of social welfare. Research has shown that this viewpoint is not necessarily
accurate. For example, “a study commissioned by the Urban Mass Transit administration determined that by every measure, low-income individuals ‘rely more heavily on taxicabs than do higher income individuals.’”26 Similarly, “in many taxi markets, consumption of taxi rides per capita is higher for low-income people.”27 These past studies make a direct connection of the dependency that lower income people have on taxicabs. Such research should supplant the notion that lower income people are not considered customers of taxicabs.

Looking at regressions results shows the statistical significance of airport traffic, public transportation and no vehicle in relation to taxicab numbers. It also shows the lack of significance of lower-income people and people aged 65 and over. This provides empirical evidence that shows that government regulations of the taxicab market may only somewhat serve public interest. There have been qualitative observations saying that “poor…and elderly customers are hit especially hard by [taxi] regulations [and] regulations impose a disproportionate burden on low income people.”28 To some degree, there may also now be quantitative analysis, as provided in this research project, to show that government regulations on taxicab numbers do not serve the interests of the low-income and the elderly, groups that rely on taxicab services.

The results from the previous models all have this same underlying tone. For governments, the interests of the overall population still seem to be important – as measures such as urban population and blue collar workers are significant. Likewise,

---

26 Kramer and Mellor, p. 8.
27 Ibid.
28 Ibid.
there was also a strong relationship of the government regulated number of taxicabs with the alternative forms of transportation in a city. Depending on the model, measures such as people who use public transportation, bike/walk for transportation, own cars, and drive to work alone were all relevant for governments. Measures for low income people, people age 65+ and minorities, however, were all insignificant.

Unfortunately, the nature of the model does not allow further explanation. It is not possible to answer the question of ‘why?’ given the regression results. Theories, without much empirical support, can be made as to why governments, in setting taxicab restrictions, choose to focus only on the overall urban population and not the minority population, or why certain city characteristics are significant as opposed others. However, for this research project, the goal was to simply deduce a ‘yes or no’ type answer – which resulted in finding that while certain factors of taxicab demand are significant, a number of groups dependent on taxicabs seem to be largely ignored when restrictions on taxicab numbers are decided.
Concluding Remarks

This project set out to explore an obscure area of city economies that few ever think about and even fewer conduct research on. The initial idea was based on the fact that the number of taxicabs in taxicab markets in the United States is often restricted by the government. With ‘supply’ at restricted numbers through direct or indirect government action, a natural question that arose was how this controlled taxicab ‘supply’ related to factors of taxicab demand. These factors of demand were constructed from past research on similar topics.

From this, a basic regression model was created with the number of taxicabs across 259 US cities as the dependent variable and factors of demand across the same cities as the independent variables. After manipulating the appropriate data it was determined that three primary factors of the original model were statistically significant in explaining the government regulated taxicab numbers – the amount of airport traffic in a given city, the number of public transportation users, and traffic congestion as measured by the time it took to get to work.

In order to correct for statistical discrepancies, an expanded model was also created to see its affect on biases from missing variables. Including a wider variety of independent variables visibly reduced the pattern in the model’s residuals-versus-fitted plot. The expanded model also yielded a different set of significant variables with average travel time to work, people who drive to work alone, people who bike/walk for transportation, blue collar workers, seasonal households, urban population and car owners as significant
variables. While the specific significant variables were different, the overall result insinuated the same outcome about government actions.

The way in which the model was developed offered a perspective on how governments act, via the taxicab industry. It was also determined that within the taxicab industry governments acted, whether directly or indirectly, with some public interest in mind. This is due to the fact that there are statistically significant variables from our taxicab social dependency model that have power in explaining the number of taxicabs governments allow. However, it was also interesting to note the insignificance of the number of lower income persons in a city and the number of elderly in a city, groups both heavily reliant on taxicabs, on the number of taxicabs a government allowed. Inferences may be made from this about what types of public interest taxicab regulations actually serve.

It is important to take this opportunity to make mention of some of the downfalls of the model used in this research project. One of the biggest is that the factors of demand used as independent variables are based on the assumption that they have influence taxicab demand. It is difficult to determine the actual degree of influence they have without a good measurement for taxicab demand currently present. In addition, the hard data associated with the factors of demand were simply chosen from the most available and based on the author’s determination of what measures seemed most relevant. As indicated by the issues with missing independent variables and the biases that they cause, it is very likely that there are many more variables that influence taxicab demand that should be included, but that simply cannot. While the existing model does provide a
general idea of what some of the crucial factors may be, additional data with a wider
variety of independent variables would help to eliminate a lot of the bias that arises from
leaving out important variables in the regression model. Within the regression model
itself, the only question that can be addressed through results is *how* governments act.
Unfortunately, after inferences about how, it remains difficult to answer the question of
*why* governments actually acted in a certain way.

Much of the value of this research, however, is in the approach. Before, the way that
governments acted was always looked at qualitatively. This way of looking at regression
analysis uses something on a micro-level scale, the taxicab industry, to attempt to
quantitatively determine the more macro-level question of how governments act. By no
means is this an open and shut case. Instead the hope is that it will provide new ideas,
spark new debate and continue pushing the bounds of answering the seemingly
unanswerable questions surrounding the enigma behind government actions.
Appendix

Full Regression Results

Figure 1:

1a.

```
. reg  taxis avg_cost enplane pop_dens lo_avg_temp age65p pubtran avgtrvtim hhi_lo_35 vehicleno
Source | SS       df       MS
-------------+------------------------------ F(  9,   250) = 294.98
Model | 191434324     9  21270480.4           Prob > F      = 0.0000
Residual | 17882661.3   250 72107.5052           R-squared     = 0.9146
-------------+------------------------------ Adj R-squared = 0.9115
Total | 209316985   259 814462.976           Root MSE      = 268.53
------------------------------------------------------------------------------
taxis | Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
avg_cost |  -3.054758   6.080455  -0.50   0.616    -15.03067    8.921158
enplane |   107.7131   27.94714     3.85   0.000     52.66905    162.7571
pop_dens |   19.47398   13.56769     1.44   0.152    -7.248613    46.19658
lo_avg_temp |  -42.19416   40.54356    -1.04   0.299   -122.0478    37.65946
age65p |  -.0602501   2.989044    -0.02   0.984    -5.947399    5.826899
pubtran |   11.88783   2.748301     4.33   0.000     6.474849    17.30082
avgtrvtim |   16.88369   4.302125     3.92   0.000      8.41033    25.35705
hhi_lo_35 |   3.975113   1.753817     2.27   0.024     .5208378    7.429387
vehicleno |  -8.277736   3.051628    -2.71   0.007    -14.28815   -2.267325
_cons |  -293.0248   119.643    -2.45   0.015    -528.6707   -57.37893
------------------------------------------------------------------------------
```

1b.

Test for Multicollinearity

```
Variable | VIF       1/VIF
-------------+----------------------
vehicleno | 388.57    0.002574
pubtran | 311.94    0.003206
age65p | 149.86    0.006673
hhi_lo_35 | 124.63    0.008024
pop_dens | 2.17      0.461359
enplane | 1.67      0.598410
avgtrvtim | 1.66      0.601754
lo_avg_temp | 1.20      0.831202
avg_cost | 1.05      0.948523
-------------+----------------------
Mean VIF | 109.20
```

1c.

Analysis of Correlation between Independent Variables

```
| vehicleno  pubtran  age65p  hhi_lo_35
-------------+--------------------------------------------------
```
Figure 2:

Eliminates independent variables vehicleo, age65p to correct for high correlation among independent variables.

2a

```
reg  taxis avg_cost enplane pop_dens lo_avg_temp pubtran avgtrvtim hhi_lo_35 
Source |       SS       df       MS              Number of obs =     259 
-------------+------------------------------ F(  7,   252) =  370.13 
Model |   190897168    7    27271024           Prob > F      =  0.0000 
Residual |  18419816.7   252 73679.2669           R-squared     =  0.9120 
-------------+------------------------------ Adj R-squared =  0.9095 
Total |   209316985   259 814462.976     Root MSE      =  271.44 
-------------+------------------------------ 
              |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval] 
-------------+---------------------------------------------------------------- 
 avg_cost |   -2.94415   6.138291    -0.48   0.632     -15.0335    9.145204 
enplane |   105.1945   28.05722     3.75   0.000     49.93586    160.4532 
pop_dens |   23.62118   13.55118     1.74   0.083    -3.067842     50.3102 
lo_avg_temp |  -59.03496   40.01173    -1.48   0.141     -137.838    19.76807 
pubtran |   4.670055    .410143    11.39   0.000     3.862279    5.477831 
avgtrvtim |   16.74423   4.344751     3.85   0.000     8.187247    25.30121 
 hhi_lo_35 |   2.681866 .44208     6.07   0.000      1.81119    3.552542 
 _cons |  -288.0484   120.9258    -2.38   0.018    -526.2115   -49.88528 
-------------+---------------------------------------------------------------- 
```

2b.

Test for Multicollinearity

```
Variable |       VIF       1/VIF 
-------------+---------------------- 
 hhi_lo_35 |      7.75    0.129038 
pubtran |       6.80    0.147080 
pop_dens |       2.12    0.472565 
avgtrvtim |       1.66    0.602865 
enplane |       1.65    0.606665 
lo_avg_temp |       1.15    0.872049 
 avg_cost |       1.05    0.951020 
-------------+---------------------- 
 Mean VIF |      3.17 
```

2c.

Test for Heteroskedasticity

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity 
 Ho: Constant variance 
 Variables: fitted values of taxis 
```
chi2(1) = 316.00
Prob > chi2 = 0.0000

2d.
Graph of Residuals for Regression

![Graph of Residuals](image)

2e.
Correct for Heteroskedasticity with Robust Standard Errors

```
reg  taxis avg_cost enplane pop_dens lo_avg_temp pubtran avgtrvtim hhi_lo_35, robust
```

Regression with robust standard errors

- Number of obs = 259
- F(  7,  252) = 148.03
- Prob > F = 0.0000
- R-squared = 0.9120
- Root MSE = 271.44

|          | Coef. | Robust Std. Err. | t    | P>|t|    | [95% Conf. Interval] |
|----------|-------|------------------|------|--------|----------------------|
| avg_cost | -2.94415 | 3.711184         | -0.79 | 0.428 | -10.25332             |
| enplane  | 105.1945  | 47.2503          | 2.23  | 0.027 | 12.13512          |
| pop_dens | 23.62118  | 16.50757         | 1.43  | 0.154 | -8.890453            |
| lo_avg_temp | -59.03496 | 44.4271          | -1.33 | 0.185 | -146.534            |
| pubtran  | 4.670055   | 1.093417         | 4.27  | 0.000 | 2.516573            |
| avgtrvtim | 16.74423  | 3.674152         | 4.49  | 0.000 | 9.50799            |
| hhi_lo_35 | 2.681866  | 1.402252         | 1.91  | 0.057 | -0.0798665          |
| _cons   | -288.0484  | 90.91743         | -3.17 | 0.002 | -467.1101          |

Figure 3:

Data Set does not include New York City to correct for outlier.
Correct for Heteroskedasticity with Robust Standard Errors

```
.reg taxis avg_cost enplane pop_dens lo_avg_temp pubtran avgtrvtim hhi_lo_35, robust
```

Regression with robust standard errors

```
Number of obs =     258  
F(  7,   251) =   21.80  
Prob > F      =  0.0000  
R-squared     =  0.7770  
Root MSE      = 244.57
```

|          | Coef.  | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|----------|--------|-----------|------|-----|----------------------|
| avg_cost | -3.602397 | 3.36009 | -1.07 | 0.285 | -10.22022, 3.015423 |
| enplane  | 110.5977  | 44.41968 | 2.49  | 0.013 | 23.11155, 198.0839 |
| pop_dens | -10.7459  | 20.16907 | -0.53 | 0.595 | -50.46963, 28.97783 |
| lo_avg_temp | -65.24969 | 42.1062 | -1.55 | 0.122 | -148.1794, 17.68002 |
| pubtran  | 12.75566  | 3.504262 | 3.64  | 0.000 | 5.853891, 19.65744 |
| avgtrvtim | 15.49671  | 3.438054 | 4.51  | 0.000 | 8.72534, 22.26809 |
| hhi_lo_35 | .7686028 | 1.358737 | 0.57  | 0.572 | -1.90748, 3.444685 |
| _cons    | -174.0352 | 75.84786 | -2.29 | 0.023 | -323.4204, -24.65008 |

Test for Multicollinearity

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
<th>1/VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>hhi_lo_35</td>
<td>3.92</td>
<td>0.255314</td>
</tr>
<tr>
<td>pubtran</td>
<td>3.63</td>
<td>0.275840</td>
</tr>
<tr>
<td>pop_dens</td>
<td>1.82</td>
<td>0.549733</td>
</tr>
<tr>
<td>enplane</td>
<td>1.63</td>
<td>0.613714</td>
</tr>
<tr>
<td>avgtrvtim</td>
<td>1.57</td>
<td>0.637925</td>
</tr>
<tr>
<td>lo_avg_temp</td>
<td>1.15</td>
<td>0.872976</td>
</tr>
<tr>
<td>avg_cost</td>
<td>1.05</td>
<td>0.950915</td>
</tr>
</tbody>
</table>

Mean VIF | 2.11

Figure 4:

Summary Statistics for Variables (with New York City, less New York City)
Figure 5:

Additional independent variables included

5a.

. reg taxis avg_cost enplane pop_dens lo_avg_temp pubtran avgtrvtim hhi_lo_35 > drv_alone bike_walk blue_collar _seas_hh urban_pop car_van no_white

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>196248432</td>
<td>14</td>
<td>14017745.2</td>
</tr>
<tr>
<td>Residual</td>
<td>12803441.9</td>
<td>241</td>
<td>54022.9618</td>
</tr>
</tbody>
</table>

| taxi | Coef.  | Std. Err.  | t     | P>|t|  | [95% Conf. Interval] |
|------|--------|------------|-------|------|-------------------|
| avg_cost | -4.529651 | 5.320336 | -0.85 | 0.395 | -15.01084 - 5.951538 |
| enplane | 81.62356 | 26.85699 | 3.04 | 0.003 | 28.71464 - 134.5325 |
| pop_dens | 32.06371 | 13.41004 | 2.39 | 0.018 | 5.645618 - 58.4818 |
| lo_avg_temp | -5.159751 | 37.05383 | -0.14 | 0.889 | -78.15667 - 67.83717 |
| pubtran | -4.593288 | 3.043937 | -1.51 | 0.133 | -10.58992 - 1.403342 |
| avgtrvtim | 8.901635 | 3.941805 | 2.26 | 0.025 | 1.136184 - 16.66709 |
| hhi_lo_35 | -4.418161 | 1.420015 | -3.11 | 0.002 | -7.215625 - 1.620697 |
| drv_alone | -26.48795 | 3.812289 | -6.95 | 0.000 | -33.99825 - 18.97765 |
| bike_walk | -39.31374 | 7.162525 | -5.49 | 0.000 | -53.42409 - 25.20339 |
| blue_collar | -11.52691 | 2.882727 | -3.99 | 0.000 | -17.21688 - 5.836947 |
| _seas_hh | 0.0731336 | 0.0183141 | 3.99 | 0.000 | 0.0370544 - 0.1092128 |
| urban_pop | 15.34812 | 2.872568 | 5.34 | 0.000 | 9.689089 - 21.00714 |
| car_van | 11.36587 | 1.710526 | 6.64 | 0.000 | 7.996097 - 14.73565 |
| no_white | 1.888154 | 0.3951392 | 4.78 | 0.000 | 1.10972 - 2.665898 |
| _cons | -156.9105 | 107.4246 | -1.46 | 0.145 | -368.5395 - 54.71844 |

5b.

Correct for Heteroskedasticity with Robust Standard Errors

. reg taxis avg_cost enplane pop_dens lo_avg_temp pubtran avgtrvtim hhi_lo_35 > drv_alone bike_walk blue_collar _seas_hh urban_pop car_van no_white, robust

Regression with robust standard errors

| taxi     | Coef.     | Std. Err.   | t     | P>|t|    | [95% Conf. Interval] |
|----------|-----------|-------------|-------|--------|---------------------|
| avg_cost | -4.529651 | 3.453027    | -1.31 | 0.191  | -11.3322 - 2.272895 |
| enplane  | 81.62356  | 42.14188    | 1.94  | 0.054  | -1.396955 - 164.6441 |
| pop_dens | 32.06371  | 22.84679    | 1.40  | 0.162  | -12.94501 - 77.07243 |
| lo_avg_temp | -5.159751 | 39.46663    | -0.13 | 0.896  | -82.90995 - 72.59045 |
| pubtran  | -4.593288 | 6.003687    | -0.77 | 0.445  | -15.78095 - 6.594864 |
5c.

Graph of Residuals for Regression

6.

Summary Statistics for Additional Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drive to Work Alone (drv_alone)</td>
<td>61.12</td>
<td>111.63</td>
<td>0.37</td>
<td>982.74</td>
</tr>
<tr>
<td>Bike/Walk for Transportation (bike_walk)</td>
<td>5.51</td>
<td>23.28</td>
<td>0.01</td>
<td>347.29</td>
</tr>
<tr>
<td>Blue Collar Workers (blue_collar)</td>
<td>39.00</td>
<td>96.79</td>
<td>0.50</td>
<td>1141.80</td>
</tr>
<tr>
<td>Season Households (_seas_hh)</td>
<td>686.11</td>
<td>2041.92</td>
<td>0.00</td>
<td>28157.00</td>
</tr>
<tr>
<td>Urban Population (urban_pop)</td>
<td>216.04</td>
<td>616.31</td>
<td>0.00</td>
<td>8008.28</td>
</tr>
<tr>
<td>Car Owners (car_van)</td>
<td>73.24</td>
<td>138.25</td>
<td>0.45</td>
<td>1203.14</td>
</tr>
<tr>
<td>Non-Caucasian (no_white)</td>
<td>45.17</td>
<td>236.96</td>
<td>0.14</td>
<td>3245.16</td>
</tr>
</tbody>
</table>
Works Cited


Data Sources


