THE CORRELATES OF CONGESTION: INVESTIGATING THE LINKS BETWEEN CONGESTION AND URBAN AREA CHARACTERISTICS

by

Milton Gregory Fields

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography and Urban Regional Analysis

Charlotte

2014

Approved by:

____
Dr. Jean-Claude Thill

____
Dr. Tyrel Moore

____
Dr. Harrison Campbell

____
Dr. Diane Brockman

ABSTRACT

MILTON GREGORY FIELDS. The correlates of congestion: investigating the links between congestion and urban area characteristics. (Under the direction of DR. JEAN-CLAUDE THILL)

Traffic congestion is a major quality of life issue, as well as being a major drain on productivity and urban competitiveness. This exploratory research seeks to identify the set of urban characteristics that are most correlated with traffic congestion. It considers just the background in which congestion occurs and does not consider causal relationships. After a review of the literature concerning congestion and urban areas, three dependent congestion variables representing the three dimensions of congestion (intensity, extent and duration) and 52 potential predictor variables are identified for 100 cities in the United States, using predominantly 2010 data. Variables are analyzed using multiple methods: simple correlation, partial least squares (PLS) regression and chi-square automatic interaction detection (CHAID) decision trees. Of the 52 predictor variables, 19 are determined to be important in all three dimensions of congestion, 13 are important in some dimensions, but not in others, and 20 are not important in any of the three dimensions. Fifteen of the 19 important variables have effects in the expected direction, including per capita freeway and arterial mileage, population density, per capita income, network intersections on the upper level system, workers per upper level network mile, jobs-housing balance, the level of sprawl, housing affordability, and urban area size (both in footprint and in population). Four more important variables (the density of transit service, level of poly-centricity, percentage of commuters driving alone, and per capita number of special events) have effects in the opposite direction as expected, which indicates that additional research is needed to clarify their relationship with congestion.

The analysis concludes that, although no causal relationships are determined, efforts to improve congestion levels through adding supply are reasonable, particularly if the focus is on both arterial and freeway capacity; efforts to improve congestion levels by decreasing demand are also reasonable, although the results of some strategies might not have the expected results and results may diminish as cities become large and very large.

DEDICATION

To my wife, Ellis, who allowed me the freedom to indulge myself in this pursuit of knowledge and provided unflagging support along the long journey. Also to my parents, Milton and Lea, who provided the background and support through the years for me to be all I could be – and also for the unique weight-loss incentive.

ACKNOWLEDGMENTS

This study was conducted under the direction of Professor Jean-Claude Thill and I would like to express my sincere appreciation to him for his guidance and mentorship, as well as for his reviews and critiques of the many drafts of this dissertation. He continually challenged my thinking and pushed me to a higher level of understanding.

I would like to thank my committee members, Professors Tyrel Moore, Harrison Campbell and Diane Brockman, for their thoughtful questions and suggestions. Of particular benefit was Professor Campbell's detailed reading of the proposal and final drafts, along with insightful comments, which made me pause and reconsider some of my initial concepts. All input helped me keep the paper clear and understandable, enhancing the end product considerably.

I would also like to thank the researchers at the Texas Transportation Institute, and in particular Drs. Tim Lomax and David Shrank, for their efforts in producing the annual Urban Mobility Report. Their perspective and detailed understanding of congestion measurement was an invaluable resource on which my effort was based.

Finally, I would like to thank Emeritus Professor David Hartgen for his longtime mentorship and support. Time spent over the past ten years working with him on a variety of transportation studies has been both a challenging and a rewarding experience – and the "grasshopper" has learned much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	American Community Survey
AWSC	all-way stop control
CART	classification and regression tree
CBD	central business district
CHAID	chi-square automatic interaction detection
CTPP	Census Transportation Planning Package
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
GIS	geographical information system
GWR	geographically weighted regression
HOT	high occupancy toll
HPMS	Highway Performance Monitoring System
ITE	Institute of Transportation Engineers
LOS	level of service
LU	land use
MPO	Metropolitan Planning Organization
NHTSA	National Highway Traffic Safety Administration
OLS	ordinary least squares (regression)
PLS	partial least squares (regression)
POV	privately owned vehicle
SEM	structural equations methodology
SOV	single occupant vehicle

TAZ	traffic analysis zone
TDM	transportation demand management
TSM	transportation system management
TTI	travel time index
TWSC	two-way stop control
UA	urban area
UMR	Urban Mobility Report
UTPP	Urban Transportation Planning Process
V/C	volume/capacity
VHT	vehicle hours of travel
VMT	vehicle miles of travel
VIP	variable importance in the projection

CHAPTER 1:   INTRODUCTION

The Texas A&M Transportation Institute (TTI) estimates that traffic congestion costs each automobile commuter in the nation's 498 urban areas about $818 annually in 2011 (based on time and fuel savings). This translates into about $121 billion annually nationwide, 2.9 billion gallons of wasted fuel, and 56 billion pounds of additional carbon dioxide emissions. (Schrank, Lomax and Eisele 2012). This extraordinary cost suggests that congestion is a major quality of life issue, as well as being a major drain on productivity and urban competitiveness. But is it? There are other views about congestion that differ from the TTI's widely accepted assessment. For example, Littman (2013, p. 2) finds that congestion is a "modest cost overall, increasing total travel time and fuel costs at most by 2%." Balaker and Staley (2006) argue that the real costs of congestion are less in the time and fuel savings than in the loss of accessibility to the destinations that matter most to us (employment, recreation, romance, etc.) which is associated with a lower level of income and wealth. Dumbaugh (2012) finds that congestion and productivity go hand-in-hand (an increase in one is associated with an increase the other) and argues that efforts to eliminate or reduce congestion are misguided, noting that cities are not adversely affected by congestion. Others (Downs 2012; Duranton and Turner 2011) add that regardless of the costs of congestion, there is little we can do about it except learn to live with it. Downs (2012, p. 20) further notes

"Traffic congestion is not essentially a problem. It's the solution to our basic mobility problem."

As these examples indicate, there is a wide range of beliefs about the nature of congestion, beliefs which impact on the solutions to and remediation of the congestion problem. There are supply-oriented solutions, such as adding highway capacity; there are demand-oriented solutions, such as reducing demand for the road network by decreasing the need to travel by car; and, there are flow-oriented solutions, such as improving intersection mobility. While all of these remedial actions have their place, there seems to be no approach that has "buy-in" from all quarters. One's perspective about the nature of the city, as well as regional cultural preferences, helps shape the approaches to tackling the congestion issue. For example, Joel Kotkin (2013) observed a disconnect between urban area planners and residents:

> "Under almost any imaginable scenario, we are unlikely to see the creation of regions with anything like the dynamic inner cores of successful legacy cities such as New York, Boston, Chicago, or San Francisco. For better or worse, demographic and economic trends suggest our urban destiny lies increasingly with the likes of Houston, Charlotte, Dallas-Ft. Worth, Raleigh, and even Phoenix. The critical reason for this is likely to be missed by those who worship at the altar of density and contemporary planning dogma. These cities grow primarily because they do what cities were designed to do in the first place: help their residents to achieve their aspirations—and that's why they keep getting bigger and more consequential, in spite of planners who keep ignoring or deploring their ascendance."

Given this interrelationship between urban areas and congestion, it seems imperative that the link be more fully explored. Are there characteristics of the urban area that seem to be associated with congestion? In this exploration, it is not necessary to take sides or adopt a preferred approach to congestion remediation. It is only necessary to use solid data and rigorous statistical analysis to disentangle these relationships.

1.1 Statement of Research

This research seeks to uncover the underlying conditions that are associated with traffic congestion. It does not deal with the proximate causes of congestion, which at its base is simply a supply and demand problem; demand exceeds supply at a point in time and space. Instead, the focus is on identifying the urban characteristics that comprise the background in which congestion occurs. These conditions are not necessarily the causes of congestion, but they are conditions that are present in many, if not most, cases where congestion is problematic. Given that congestion is extremely complex and multi-faceted, these conditions and their interactions must necessarily be disentangled using a number of complementary measures examined separately or in conjunction with one another. It is the nature and strength of the interrelationships between the underlying urban conditions and traffic congestion that are the target of the research, and not a determination of causality.

Causal variables are extremely difficult to determine, especially in the realm of human behavior, and congestion is a behavioral issue. To begin to explore congestion from a cause and effect perspective would require a micro-analysis of one or just a few urban areas, and even then any causes of congestion would be suspect. As long as people have the freedom to act in a manner of their choosing, human behavior will be difficult to predict. Collective behavior is statistically easier to predict than individual behavior; indeed there are behavioral models of travel that are used widely. Congestion, however, is a time/space phenomenon that occurs at the confluence of individual travel behavior and system supply, so identifying these confluences to a degree useful to planning officials even for one city would likely be difficult. Moreover, extending the findings for one city to others, that by nature will have different circumstances, culture and people, can be seen as reaching.

This research circumvents the causality issue altogether. Instead, it explores the background in which congestion occurs, seeking to identify the urban characteristics that are present when congestion is also present. Knowing these characteristics should lead to a better understanding of urban congestion and perhaps a better ability to predict its occurrence without getting bogged down in questions of causality. Additionally, knowing which urban characteristics are more important allows follow-on research in areas such as causality to be more soundly grounded.

1.2 Research Questions

Before we can assess congestion, we need to know what it is and how it is measured. This understanding will allow us to identify characteristics of urban areas that might have an impact on regional or localized traffic congestion. Much research has already been done in defining congestion and developing metrics to assess its effects. It is important that we identify those metrics that will best enable the exploration of the links between congestion and the urban landscape. With this understanding at hand, we can investigate the following research questions:

1.2.1 What set of urban area characteristics impact traffic congestion?

There are a wide number of measures that are available to describe the urban landscape. From population counts of various types, to employment statistics of various aggregations, to highway statistics, to social measures, the available metrics are well in the hundreds. Many of these measures will overlap and extracting ones that are relatively independent from one another will be difficult, if not impossible. (It should be noted here that urban areas can be considered analogous to living organisms that grow according to their own demands, with each new "growth spurt" impacting on the whole. In this light, most urban area characteristics will have some level of interrelationship with the others.) While most of the focus would be on policy-type variables (those characteristics that could be affected by changes in policy), other variables may emerge that are interesting and offer insight to our understanding of the urban area-congestion dynamic. (It could be argued, of course, that most variables could be controlled, at least somewhat, by policy, depending on the level of governmental control. This study considers policy variables from a US perspective.)

To identify those urban characteristics that might be used as variables, it seems prudent to start with a review of the theories and models that have been developed over the years that address congestion, the urban condition, and urban travel. With these theories and models in hand, we can determine the points of intersection where potential variables of interest may reside. Once identified, we can hypothesize their potential effects by magnitude and direction.

1.2.2 How do these urban area characteristics correlate with traffic congestion?

Once characteristics that might impact traffic congestion are identified, we can analyze them to determine if they do indeed have an impact and the extent of that impact. Moreover, we can identify which are the most important. This is the crux of the study. It is this knowledge that should enhance policy makers' ability to develop sound strategies to lessen the negatives associated with congestion.

CHAPTER 2: THE CONGESTION-URBAN AREA LINK

2.1 Defining Congestion

The verb “congest” is defined in the Merriam-Webster Dictionary as “to obstruct by overcrowding” (1997, p. 170). So congestion is the obstruction caused by overcrowding, which suggests three underlying causes: inadequate supply, excessive demand and/or poor flow. Traffic congestion then, is the obstruction (or delay in vehicular travel) caused by the presence of other vehicles, with the three underlying causes being inadequate highway capacity, excessive travel demand and poor traffic flow. These underlying causes do not generally apply to the entire transportation network, but just to specific geographic points at specific times. For example, most congestion is limited to a few links in the network during peak hours of travel. Most streets are not congested during peak hours and virtually no streets are congested at night. So traffic congestion is a time-space issue.

Traffic congestion results in the delay of travel as compared to travel during times of “free flowing” traffic. Vehicles may encounter stoppages or stop-and-go traffic on roadways, wait for more than one cycle to pass through signalized intersections, see travel times that are slower (and in some cases much slower), and experience variations in travel times that make trip planning difficult. Congestion has three main components, or dimensions: intensity (the severity of the congestion present), extent (the portion of the network with the congestion problem), and duration (the number of hours in the day when congestion is a problem).

Traffic congestion has been increasing over time with increasing urbanization, as more and more people crowd into areas with transportation networks that for a number of reasons to be explored below cannot or do not keep pace with the growth in vehicular demand. Interestingly, as traffic congestion has become more common, people have adapted somewhat and become more tolerant of travel delays and even more skilled at negotiating congested roadways. With this adaptation, congestion is often considered simply a cost of living in crowded areas – what passed for severe congestion in 1990 might be considered only moderate congestion in 2010.

2.2 Causes of Congestion

The Federal Highway Administration has identified seven root causes of traffic congestion, grouped into three categories: traffic-influencing events [traffic incidents, work zones, and weather]; traffic demand [fluctuation in normal traffic and special events]; and, physical highway features [traffic control devices and physical bottlenecks (capacity)]. (FHWA 2008) These causes routinely interact with one another to compound their individual effects and make it more difficult for traffic managers and planners to devise remedial actions.

2.3 Types of Congestion

The above root causes of congestion can be grouped into two basic types: recurrent and non-recurrent. Recurrent congestion is that congestion that is experienced on a routine basis and is due to the general conditions of the transportation network, while non-recurring congestion occurs when specific non-routine incidences serve to reduce the number of vehicles that a segment of the network can handle.

From the public's perspective, non-recurrent congestion is perhaps more understandable (though often highly frustrating) and often beyond the control of transportation planning officials, while the recurrent congestion falls within the planning purview and many feel could be fixed if the political will were there and sufficient funds allocated. Although non-recurrent congestion accounts for over half of total congestion delay,[1] it is the recurrent congestion that is arguably the more irritating. Accidents do happen, weather events do occur, and special events are one of the attractive amenities of the urban experience, but they are sporadic and often come with "advance notice" that allows some prior planning. Recurrent congestion happens daily and is "in ones face" during large portions of the urban drive. Hence, it tends to be more memorable.

2.3.1 Recurrent Congestion. Causes of recurrent congestion include:

- Physical Bottlenecks ("Capacity"): The maximum number of vehicles a given segment or link in the transportation network can handle under normal prevailing operational conditions is called its capacity. Capacity is dependent largely on the physical characteristics of the roadway (number of lanes, widths of lanes and shoulders, the geometries of the roadbed (curvature and slope), and the level of access or number of entry exit points). Inadequate capacity equates to inadequate supply in the supply-demand relationship which yields traffic congestion.

[1] The Federal Highways Administration (FWHA) estimates that 55 percent of congestion is non-recurrent, broken out as follows: incidents: 25 percent, bad weather: 15 percent, work zones/construction: 10 percent, and special events: 5 percent. The remaining 45 percent of congestion is recurrent (bottlenecks: 40 percent and poor traffic signal timing: 5 percent) (FHWA 2012b).

- Traffic Control Devices: As congestion is at its heart a flow issue, anything impeding flow can cause congestion. Traffic control devices (e.g. stop signs, stop lights, and railroad crossing gates) by design impede traffic flow and cause the existing traffic to "bunch up", effectively reducing the highway supply available to those vehicles. If travel demand is sufficiently high, then congestion is generated.
- Fluctuations in Normal Traffic: Travel demand varies by season, day of the week and time of day. Since it is generally not cost-effective to design roads for the worse case situations, demand can regularly exceed supply at points in time, such as during the proverbial rush hours on weekday mornings and evenings, and Saturdays in the summer when beach houses turn over. When demand exceeds supply, congestion follows.
- Poor or Distracted Driving: Congestion is a flow issue and disruptions to the smooth flow of traffic cause it. Vehicle operators who are distracted (texting, talking on a cell phone, reading office papers, eating, fiddling with radios/GPS devices, disciplining children in the back seat, etc.) or of diminished capacity (intoxicated, sleepy, under the influence of prescription medication, overly timid or unsure of themselves, etc.) often drive erratically or slower than the prevailing traffic flow, which can reduce throughput and if demand is sufficiently high, contribute to congestion.

2.3.2 Non-recurrent Congestion. Causes of non-recurrent congestion include:

- Traffic Incidents: Events such as traffic accidents, vehicle breakdowns, and debris on the roadways are known as traffic incidents, and can contribute to congestion in two ways. They can reduce the available capacity (supply), as well as impeding the flow by reducing speeds and decreasing throughput.

- Work Zones: Construction/maintenance activities long the roadway, whether marked as a work zone or not, often cause congestion when demand is sufficiently high. Similar to traffic incidents, work zones may reduce available capacity (by reducing speed limits, closing lanes and shoulders, and reducing lane widths) and impede flows (by traffic shifts, lane diversions, and distracting drivers).
- Weather: Weather conditions can significantly affect flow as drivers slow down in response to limited visibilities caused by precipitation and fog and the adverse roadway conditions caused by rain, ice, and snow. Additionally, the weather can reduce the available capacity by narrowing and even closing some lanes to traffic.
- Special Events: Special events (e.g. sporting events, concerts, high school graduations, funerals, etc.) can cause congestion in and around the event. This happens in three ways: demand can increase in the vicinity of the event, capacity can decrease in the same area as lanes may be blocked by event attendees or by police officers controlling traffic, and flow can be inhibited by driver uncertainty and distraction.

2.4 Measures of Congestion

There are a number of ways to measure congestion; one can focus on a link (road segment) or node (intersection) in the transportation network or one can focus on the network as a whole. The link/node measures are essential when micro-analyzing the transportation grid so that specific, often highly localized improvements can be made. Network-wide measures are better for the macro-analysis of the region as a whole as one tries to get a handle on the extent and scope of the congestion problem so that decisions can be made in the allocation of resources to provide congestion relief.

2.4.1 Measuring Congestion on a Link. There are two widely used ways to describe congestion on a link, both of which address the supply/demand connection and both of which are useful in identifying specific road segments that need attention.

- Level of Service (LOS). This measure is widely used throughout transportation planning and assessment.[2] For highway links, the LOS is a primarily visual measure of supply and demand, where traffic flows are rated from A to F, with LOS A being smooth free-flowing traffic and LOS F showing severe congestion. Littman (2012) provides a nice summary of the levels of service, derived from the description in the Highway Capacity Manual (TRB 2010):

Table 1: Descriptions of the levels of service for network links

LOS	**Description**
A	Traffic flows at or above the posted speed limit and all motorists have complete mobility between lanes
B	Slightly congested, with some impingement of maneuverability. Two motorists might be forced to drive side by side, limiting lane changes.
C	Ability to pass or change lanes is not assured. Most experienced drivers are comfortable, and posted speed is maintained, but roads are close to capacity. This is often the target for urban highways.
D	Typical of an urban highway during commuting hours. Speeds are somewhat reduced, motorists are hemmed in by other cars and trucks.
E	Flow becomes irregular and speed varies rapidly, but rarely reaches the posted limit. On highways this is consistent with a road over its designed capacity.
F	Flow is forced; every vehicle moves in lockstep with the vehicle in front of it, with frequent drops in speed to nearly zero mph. A road for which the travel time cannot be predicted.

Source: Littman (2012), TRB (2010)

- Volume/capacity (V/C) ratio. This measure is a straightforward measure of supply and demand. The supply, or capacity, is derived from parameters in the

[2] For example, the Highway Capacity Manual (TRB 2010) applies the Level of Service concept to most modes of traffic and often to several conditions within a single mode.

Highway Capacity Manual (TRB 2010) and indicates the maximum amount of traffic a road segment can carry at the posted speed limit based on its design and geometry. Modern freeways have capacities of about 2400 vehicles per lane per hour (or one vehicle passing a roadside observer every 1.5 seconds), while modern highways and older freeways have lower capacities (1800-2200 vehicles per lane per hour); other roads and streets may have even lower capacities because of their design characteristics and cross-street flow. The demand, or volume, is the actual number of vehicles on that road segment per unit time. Both volume and capacity are typically measured in one-hour increments, and routinely grouped in peak (morning and evening rush hours) and non-peak (mid-day and night) periods.

Values for the V/C ratio can range from 0 to 1.0+, with 0 denoting a road segment devoid of any traffic whatsoever and 1.0 denoting a segment with a flow rate at maximum capacity. It should be noted that a road segment can carry more than its maximum capacity for brief periods, although congestion when V/C ratios are near or above 1.0 is quite severe. Moderate to heavy congestion occurs when V/C ratios are in the 0.7 – 0.9 range.

2.4.2 Measuring Congestion at a Node. Congestion also occurs at nodes, or intersections, in the network where the smooth flow of traffic is disrupted by design. Most intersections have signage and/or traffic signals that constrict the flow in some manner to enhance safety in and around the intersection and to prevent vehicles from occupying the node at the same time, so there is some delay at virtually all intersections. This delay does not include the reduction in speed upon approach to the intersection or the reduction in speed while resuming the posted speed limit; only the wait times are considered. This delay is typically assessed by the average delay per vehicle as measured in seconds of delay and assigned a level of service, as noted in the table below. Today at signalized intersections, severe congestion occurs when the delay exceeds 80 seconds per vehicle, which happens when the vehicle takes more that one timing cycle to pass through the light. Interestingly, the delay needed to rate as severe has increased over time as congestion has become more ubiquitous and drivers more tolerant. Two decades ago, a delay of 60 seconds was rated as severe (LOS F).

Table 2: Average delay by level of service for network nodes

LOS	**Signalized Intersection**	**All-way-stop-control (AWSC)**	**Two-way-stop-control (TWSC)**
A	0-10 sec	0-10 sec	0-10 sec
B	> 10-20 sec	> 10-15 sec	> 10-15 sec
C	> 20-35 sec	> 15-25 sec	> 15-25 sec
D	> 35-55 sec	> 25-35 sec	> 25-35 sec
E	> 55-80 sec	> 35-50 sec	> 35-50 sec
F	> 80 sec	> 50 sec	> 50 sec

Source: Highway Capacity Manual (TRB 2010).

2.4.3 Measuring Congestion in the Network. Any area, urban or rural, could have specific links or nodes in the transportation network that are congested. For example, in rural areas near a high school, the links in and around the school may become congested when school is dismissed in the afternoons. Nonetheless, in these areas, congestion is not necessarily a problem. To assess the congestion in a region, a regional measure is needed, and there are many. Given that congestion has three dimensions (intensity, extent, and duration), these regional measures attempt to address one or more of these dimensions. It should be noted that these measures are average measures, which can effectively “hide” real congestion problems on specific links and intersections. Even cities with acceptable regional congestion ratings may have location and time-specific congestion issues, so care must be taken in their application.

- Travel Time Index (TTI). One of the more prominent regional congestion measures is the travel time index (TTI), developed by the Texas Transportation Institute (also abbreviated as TTI) and used in its annual Urban Mobility Report. Although originally developed for larger cities, where congestion is greatest and also where it makes the most sense to allocate the needed funds to make the necessary measurements, TTIs have been calculated for all cities in the nation with populations of 50,000 and above (Hartgen and Fields, 2006).

  The TTI is an indirect measure of supply and demand. Instead of focusing on the road capacities and traffic volumes, TTIs focus on travel times, which addresses the supply/demand problem as it impacts on traffic flows. The TTI is the ratio of travel in the peak hours (rush hours in the mornings and evenings) to travel in the off-peak (mid-days and nights):

$$\text{TTI} = \frac{\text{Average travel time in peak hour}}{\text{Average travel time in off-peak hours}}$$

  For example, a TTI of 1.17 means that it will take 17 percent longer to travel the same routes in the peak hours as compared to off-peak (free-flow) hours. While there are no current TTI congestion standards, interpolations based on the Highway Capacity Manual suggest that a TTI of 1.18 indicates severe congestion (TRB 2010). The TTI is predominantly an intensity measure although it does get at the extent issue as well. Since the peak hour period is undefined, the TTI does not deal with the duration problem. Although TTI could be calculated for individual links in the traffic network, it is most commonly used as a regional measure of congestion.

- VMT (vehicle miles of travel) traveled in congested conditions. This measure assesses the extent of the congestion problem; that is, how much of the network has supply/demand issues. This metric may be expressed in total congested miles during peak hours or congested miles as a percent of the total mileage.
- VHT (vehicle hours of travel) traveled in congested conditions. This measure also assesses the extent of the congestion problem, but from a time perspective. This metric may be expressed in total travel time over congested links during peak hours or congested travel times as a percent of the total travel time.
- Percent of lane miles that are congested. This measure is another way to assess the extent of the congestion problem and considers congestion from the network's perspective rather than the individual's experience on the network, as the two previous metrics do.
- Peak Hour Length. This measure estimates the length of the morning and evening "rush hours" to assess the duration of the congestion problem. It does not address the severity of congestion or the extent within the region.
- Average Traffic Speeds. A key difference in the flow of traffic during congested and uncongested periods is that of travel speed. Comparing average speeds to average free flow speeds is a metric that indicates the intensity and extent of the congestion problem.
- Hours of Delay. A comparison of congested speeds with free-flow speeds can lead to calculations of the delay caused by congestion. Measures of delay (annual

hours of delay, delay per capita, delay per commuter, etc.) are common ways to make the costs of congestion less "geeky" and more meaningful to highway users. While these type metrics do a good job of addressing the intensity, extent and duration of congestion collectively, they are too general to use for targeting remedial efforts to improve the system.

- Costs of Congestion. Similar to the measures of delay, these metrics attempt to quantify costs to the highway users of congestion. Costs, which can include the value of delays, the value of the excess fuel consumed, and the value of additional vehicle operating costs, are often expressed as component costs or total costs, or as costs per capita, per driver, per commuter, etc. These measures also tackle the intensity, extent and duration aspects of congestion, but they too are too general to assist with targeted corrective strategies.

2.5 Defining Urban Areas and Cities

While people often use the terms "urban areas" and "cities" interchangeably, there is a difference. The various definitions of cities include terms such as "centers of population, commerce and culture", "large", "densely populated", "important", "historical", "permanent", "socially heterogeneous", and "with self-government". Most descriptions of cities seem to denote a long-term structured entity that serves a large, concentrated number of people. Urban areas, on the other hand, are described in geographical and population terms without regard to jurisdictional boundaries or administrative structure. Since the descriptions of urban areas and cities are different and approach the idea of the concentration of people from different perspectives, an exploration of both would be useful in uncovering some of the characteristics potentially associated with urban congestion.

2.5.1 Cities. Louis Wirth (1938), of the highly regarded "Chicago School" of urban sociology, included four traits in his definition of cities: relatively large (population-wise), dense, permanent, and socially heterogeneous. He regarded a city as being something more than its physical structure and regarded urban living as a "machine-based" style of living (vs. the "nature-based" style in rural areas). He further noted that large populations were associated with increasing differentiation between people and that high population densities were associated with increasing specialization of employment, goods, and services. This sociologically-focused definition is compared with others that are more characteristic-based. The Demographia website (Demographia 2013) discusses cities as municipalities, as metropolitan areas, and as urban areas, noting that interpretations of the term city differ among countries and cultures. Municipalities are generally the smallest entities of the three, while metropolitan areas typically include multiple municipalities. A city's urban area would include the core municipality and the adjacent suburbs, and a metropolitan area's urban area would include all the area of continuous urban development.

2.5.2 Urban Areas. A simple definition of urban area might be the city area plus the continuous built-up surrounding areas, irrespective of local body administrative boundaries. Demographia (2013) observes that an urban area will unlikely ever be the same as a municipality (some urban areas might be larger or smaller than the cities with which they are associated) and suggests that an urban area "might be thought of as defined by the lights seen from an airplane on a clear night." Demographia further notes that metropolitan areas mean labor markets, or the areas from which the region's employees come, and because of this the metropolitan area will always be larger than the urban area since many of the metropolitan employees will come from the rural regions beyond the urban development.

The Census Bureau differentiates urban and rural areas based largely on population density and land uses. For the 2010 Census, an urban area comprised a "densely settled core of census tracts and/or census blocks that meet minimum population density requirements, along with adjacent territory containing non-residential urban land uses as well as territory with low population density included to link outlying densely settled territory with the densely settled core" (Census 2010). The Census Bureau subdivides urban areas into urbanized areas (centers with populations of 50,000 or more) and urban clusters (centers with populations of 2,500 or more, but less than 50,000). Rural areas would include anything not characterized as an urbanized area or urban cluster.

2.6 Theoretical Underpinnings of Urban Congestion

2.6.1 Key Theories and Concepts. Urban agglomeration and urban travel involve both individual and collective decisions. More specifically, both are the aggregated result of thousands of individual decisions based on individual needs and preferences, which are often affected by normalizing forces of their social networks. These decisions are made at the various stages in aggregation/travel process and are infinitely complex. Urban agglomeration involves decisions about where to work, where to live and how to live; while urban travel involves why to go, when to go, where to go, and how to go. Multiple factors impact these decisions, to include education, work credentials, financial resources, time, destination alternatives, time of day, and end of trip activities. Academics and planning practitioners have studied both aspects of human behavior in an urban setting and have developed any number of theories to explain it (some examples are discussed below). Most of these theories are more valuable as explanations of behavior rather than predictors of it. For example, a rational actor may well act to maximize his/her utility, but it is extremely difficult to determine this utility until after the fact. Short term goals compete with long term goals and the eventual utility that is maximized cannot be easily predicted. Still there are several key theories and concepts that merit mention in exploring the link between congestion and urban characteristics.

- Central Place Theory. Developed by Walter Christaller in the 1930s while studying urbanized areas in southern Germany, this theory sought to explain how towns and cities evolve in relation to one another; how many would arise, how big would they grow, and how far apart would they be (Christaller 1966). Much of central place theory revolved around geometric shapes and topographic relationships and has been criticized heavily through the years, but its notion of a city as a distribution center of goods and services to the surrounding populace remains a core characteristic of cities to this day. The number and variety of the goods and services available in the urban area, together with the number and density of the people, are the pre-determinants of urban congestion.
- Structural functionalism. Based on Herbert Spencer's theory of functionalism in the mid-1800s, which had been expanded by Emile Durkheim in the following decades, structural functionalism is an underlying theory of self-sustaining social interaction. Championed by Talcott Parsons in the 1950s-60s, this theory holds

that society, as a system, more so than any particular actor, develops those systems needed to support and sustain itself (Ritzer 1996). (These systems, or structures, are functional, hence the name.) In keeping with this theory, cities and transportation networks have arisen and evolved to meet some universal need(s). These needs could be evidence of the failure or shortcomings of some other organization entrusted with a particular responsibility, of the lack of an organization to address the needs, or of the lack of agreement of what the universal needs should be. A working system becomes entrenched and part of the culture and as people are educated and socialized to fit into society, they are trained in their roles within the new system. So the opportunities and amenities afforded in the urban environment have become an accepted part of the landscape and many individuals have come to expect and desire them. Likewise, the urban transportation grid allows these individuals to partake of the urban offerings. The structural functionalist would argue that the degree to which cities and their transportation networks are integrated into local society and meet the needs of the people would determine the levels of aggregation and travel, respectively.

- Land-rent Theory. This theory holds that the most valuable land in an urban area is in the central business district (CBD). As one moves farther away from the CBD and the economic activity that occurs there, one loses some of the advantages that occur with geographic proximity and face-to-face interaction. The land becomes less valuable as there is less competition for it and it is used less intensively (Forkenbrock, Mathur and Schweitzer 2001). (It should be noted here, that distance is not only Euclidean distance, but network distance, as well, measured in both distance and travel time. It is the quickness of access to the CBD and other key destinations within the city than affect land rents.) This most often translates into residential areas located “out” from the commercial regions, although changes in accessibility can change land value and land use. Land-rent theory offers a good explanation of the link between the transportation grid and city form. Both transportation costs and land costs are factors in location decisions; the more accessible the land is the more valuable land. The better the transportation network, the less variability in land accessibility and hence prices, and the more decentralized development tends to be. Accessibility follows the street network.

- Circuit-switched networks vs. packet-switched networks. In the world of data transmission, there are two basic methods of transmitting data. Circuit-switched data requires a dedicated circuit from the point of origin to the destination. The whole data set travels along this route without deviation. Packet-switched data, on the other hand, takes advantage of any alternate routes that the network offers. Here the data set is broken into packets at the origin, with each packet moving independently to the destination, to be reassembled at that point in time into the proper configuration. Packet-switching has proven superior to circuit-switching; especially as computers have become more and more powerful (packets require far more power than circuits) and information networks have become denser. This concept has been applied to the transportation world in the discussion of transit vs. automobiles (Fleming 2007). Although the analogy is not exact, there is merit in the idea that the more alternate routes that are available, the more quickly the packets can arrive at their destinations.

- Changing Urban Needs over Time. The needs of an urban area will naturally change over time with advances in technology and changing demographics. Legacy cities, those cities that came into primacy before the street car era (prior to about 1890), needed a different city structure and transportation grid than did cities that developed in conjunction with street car capability. Street car cities, in turn, had different requirements than cities that came to fore in the 1920s when the automobile democratized urban travel and made cheap land available to large segments of the urban populace. Cities developing after the Second World War and the rise of the freeway had still different needs. The size, density, robustness and connectivity of the street networks were different in each era and the land use varied accordingly. As needs evolved, efforts were made to adapt the street network. Often this required reconstructing urban areas that were adapted for the previous paradigm, which sometimes proved to be too expensive to attempt. The result is that many of the legacy cities are ill-designed to accommodate today's mobility preferences and would likely have different structures had they come into prominence more recently.
- The Demand for Transportation is Derived. The demand for transportation is not generally a demand for transportation itself, but a demand for mobility and access to get to a good or service. Although there is some travel just for the sake of travel (e.g., a drive through the country to look at the fall leaves, a quick drive around the neighborhood to put the baby to sleep, and a conspicuous cruise

through the teen-gathering areas to make one's presence known), the vast majority of travel is undertaken to reach a particular destination. Derived demand is compared to direct demand where the demand is for that particular good or service. Here, the supply-demand curve works well; as the price of the good/service decreases, demand increases. For derived demand, the demand for that good/service does not necessarily increase as price decreases. This is particularly true for transportation, in which cheap travel does not necessarily mean more travel. Most travelers need somewhere to go to make the travel worthwhile, regardless of transportation cost. There is, of course, some impact on travel due to costs. For example, decreases in costs (e.g., the cost of gas) may make some worthwhile trips affordable when they were previously not. Still, travelers need destinations to drive their travel.

- Rational choice theory. This theory, predicated on the assumption that human beings are rational creatures with free will and have perfect knowledge of the characteristics of alternatives, holds that when faced with a decision of any type, people will consider their options, weigh the pros and cons, and make the decision that is in their own individual best interests. Rational choice theory, first developed in the field of economics (Zafirovski 2001) where individual decisions are fundamental considerations in most aspects of the discipline, has wide-ranging applications and is now used to explain individual behavior in all types of circumstances. Gill and Gain (2002) describe this "rational" decision-making behavior using five key aspects: utility (an outcome that provides the individual some relative benefit), purposefulness (the decision will lead to an increase in utility), certainty (higher certainty is preferred over lower certainty), sincerity (the tendency to choose what one believes is best, as compared to strategic decisions-making or "gaming"), and comparability (alternatives can be compared). In making decisions about moving to an urban area or going somewhere, rational choice theorists maintain that people would attempt to maximize their individual utility (which, in traveling, is generally taken to mean minimizing their travel times), regardless of how that decision might affect the public as a whole.

- Social exchange theory. A key sociological theory about urban life seems to be social exchange theory, which holds that people (or organizations) maintain relationships that benefit themselves (Cook and Rice 2001). Reciprocity is the key factor in this theory and systems will tend to fail unless there is some mutual benefit. City dwellers and the businesses they support (and that support them) enjoy a symbiotic relationship that is mutually beneficial. Social exchange theory holds that as long as the exchange is balanced, the system will continue. It is clear that this theory, while very reasonable and readily defendable, avoids a key question about the exchange between agents within the city context: what constitutes a benefit to the individual?
- Reference group theory. Reference group theory takes a different tack and explains individual decisions within a group context. Sociologists have studied the impacts of such reference groups extensively, focusing primarily along two lines of thought: the Lewinian approach (because of face-to-face interactions between group members, individuals behave in a manner consistent with the

norms of the group[3]) and the social identity model (individuals behave in a manner consistent with the norms of the group simply because they identify with the group, whether they have face-to-face interactions with other group members or not) (Koch 1995). So the reference groups included in this theory can be formal and tight-knit, where people interact with one another directly, or open-ended with no defined membership, as long as one identifies with the group and will follow the normative behavior of the group.

- Interest group pluralism. Two of the cornerstones of the American constitutional system are freedom of speech and freedom of association. These two rights together mean that people can associate by forming organizations that pursue activities with different objectives. Such objectives include making a living, maintaining social networks, enhancing the personal enjoyment of life, and achieving social goals. All routinely involve mobility and access to the city's transportation network. How one uses this mobility and access is often shaped by the interest groups with which one is affiliated.
- Tragedy of the Commons.[4] When individuals act in their own enlightened self-interest to a portion of a commonly shared resource, with little regard to the compounding effects of other individuals acting in a similar manner, the end result is an overuse or even depletion of the shared resource. Road congestion is an example of such a tragedy of the commons; roads are seen as essentially free,

[3] Derived from Kurt Lewin's dynamic approach rule, which states that the "elements of any situation should be regarded as parts of a system" (Neumann 2005).
[4] Although the problem of the commons had been known in the time of Aristotle, it was ecologist Garrett Hardin (1968) who coined the term "tragedy of the commons" in a journal article of that name (Ostrom 1990).

and there is little incentive to avoid overuse. This is especially true when selecting the time and routes of travel.

- Queuing Theory. Queuing theory considers the competition for the use of a shared, but limited, resource. This theory depends heavily on mathematical relationships and formulas to describe the flow of a service, whether it be at the checkout counter, on the assembly line, or at a telephone switchboard. In the field of transportation, queuing theory is used to analyze traffic flows at traffic bottlenecks, such as places with lane reductions and signalized intersections. Here, the traffic volumes may exceed the capacity of the roadway/signal and interrupt the stability of the flow causing traffic to back up to wait their turn; the more the delay, the increasingly more the back-up. Queuing theory helps to explain one of the issues of concentration that plague the transportation system. The varying levels of service can provide acceptable performance in the upper ranges, but once they become degraded, very little additional traffic is needed to produce gridlock. The speed curves in the Highway Capacity Manual (TRB 2010) are testament to this idea. Queuing theory suggests that moderate congestion can become problematic very quickly and with few additional vehicles added to the mix.
- Loss aversion. Loss aversion is the idea that people are more sensitive to the value of something they lose (or may lose) than they are to something of similar value that they gain (or may gain). That is, people will work harder to avoid a loss than they will to make an equivalent gain. Behavioral studies have found that losses have about twice the power over us that gains do (Tversky and Kahneman

1981). Loss aversion affects travel behavior, especially in congested conditions, as drivers try to protect their "turf", follow too closely, and engage in rude and inconsiderate driver behavior, which can lead to stop and go traffic, rather than smoothly flowing, but slower moving traffic. Loss aversion also plays a role in the resistance drivers have to certain congestion remediation efforts, such as congestion pricing and high occupancy toll (HOT) lanes. People tend to resist paying fees and tolls, especially where none has been charged before (FHWA 2009).

2.6.2 Models of Urban Travel. Travel demand forecasting began in the early 1960s with area-wide transportation studies in Chicago and Detroit. The initial motivation for the development of travel demand models was fundamental: to provide an objective tool to evaluate major transportation projects and develop long-range regional transportation plans (Martin and McGuckin 1998). The early models were crude and cumbersome by today's standards, but with the advent of the personal computer and the steady and dramatic improvements in computer software, they have evolved into fairly sophisticated tools. The four-step, trip-based Urban Transportation Planning Process (UTPP), first developed in the 1950s (Weiner 1997), remains the framework for most of the current travel demand models.

There is some major effort to develop activity, tour-based models[5] and travel simulation models.[6] Although these new approaches are now beginning to bear fruit,

[5] Tour or activity-based models treat travel differently than trip-based models. Instead of building the model upon individual trips from point A to point B, tour-based models combine multiple trip legs into tours or trip chains. For example, a parent might pick up a child at soccer practice and stop at the grocery store on the way home from work. These trip legs, which are considered separately in the UTPP, are linked together into tours, which are then modeled as a whole. When so considered, the modeled travel behavior of this parent could easily be different than if all of these trip legs were considered separately.

especially in transportation related air quality issues (Beckx et al. 2009, Hatzopoulou and Miller 2010), many in the transportation field believe that our underlying understanding of travel behavior is currently inadequate to support such models, which rely heavily on combining the various trips made by individuals into some coherent, predictable behavior. Research and development efforts are continuing and the advent of new and improved model paradigms are likely. Meanwhile, the four-step method is the paradigm of choice and used by all the major models currently in widespread use. The four steps are:

- Trip Generation. Trip generation is the process in which the amount and type of trips in the planning region are calculated, based on the use of the land and the preferences and needs of the people making the trips, as well as the various socio-economic and employment data that impact on these two factors. Both the number of trips "produced" in a zone (based on travel surveys and socioeconomic data and categorized into a number of trip purposes) and "attracted" by a zone (based on land use data and ITE tables[7]) are determined for all zones in the region. It is these productions and attractions that generate the demand side of the supply-demand function of urban traffic.

---

[6] Travel simulation models attempt to model individual travel behavior through the use of decision matrices at specific points in the trip. These matrices are based on situations and routinely include some probability functionality to reflect the notion that many stops along the trip route, or even the trip routes themselves, are often unplanned at the outset. Because of the vast numbers of decision points in a single trip, travel simulation models are best suited for modeling behavior at specific points in the network, such as intersections and on and off ramps.

[7] The Institute of Transportation Engineers (ITE) publishes periodically a Trip Generation Manual (currently in its 9th edition) that contains instructional material, a recommended practice on the use of this resource, and data on land use descriptions, trip generation rates, equations, and data plots. Transportation planners use this information to determine the number of trips "attracted" by a business establishment, recreational facility, or other destination.

- Trip Distribution. This step in the four-step process matches the productions with the attractions and a “gravity” model (based on the idea that people (productions) tend to go to the closest attraction that will meet their need) is almost universally used to do so. To reflect the amount of difficulty to move between zones, some type of impedance measures are included in the model. These impedance measures can be a single “friction” factor, a look-up table of friction factors, or some form of travel time decay function. Once trips are distributed, trip lengths are, in practice, checked against the household survey or Census data. And finally, to ensure the external trips (those trips with one end of the trip outside of the planning area) are properly integrated, external overlays are generally used, particularly in smaller cities. Trip distribution adds a spatial component to the demand side of urban traffic. Congestion has a spatial component and the proximity of the productions and attractions would seem to have some impact on its formation.
- Mode Choice. This step splits the trips into the various modes so that they can be assigned to the traffic network. Transit is the most common mode, but other modes include walking, bicycling, and carpooling. In most areas, solo automobile use is so dominant that the mode choice step is often skipped. In the larger cities, and especially those with large percentages of pedestrians and transit riders, this step is essential for realistic traffic assignment. To the extent that mode choice moves users off the street network, the numbers of trips will be reduced and traffic flow positively affected.

- Traffic Assignment. As the name implies, in this step the trips are assigned to the traffic network. The desktop computer has allowed for far more sophisticated assignment techniques than in times past, and "equilibrium" methods dominate. These methods use an iterative process to continually adjust the trips assigned to each link based on the traffic volumes. Traffic assignment is typically based on rational actor models, where drivers have perfect knowledge of road conditions and use that knowledge to minimize their travel times. Here the supply side of the supply-demand function comes into play, and the size, density, connectivity, and "thickness" of the network is a key component of urban congestion.

The four-step method, though the paradigm of choice, is not without its shortcomings. Several of these are problem areas and the manner in which they are being addressed is noted below:

- Time variations. Travel behavior varies by time (of day, of week, of month, and of year) and this variation can often be lost in the aggregated nature of the trip generation process. For such time variations to be captured in the modeling process, the specific time periods of interest must be isolated and the transportation behavior patterns within them considered separately. Time of day considerations for the smaller MPOs are especially important, since often congestion problems are limited to certain hours of the day, such as peak hours or factory shift change. In areas near recreational attractions such as beaches and ski resorts, transportation issues would likely be more seasonal with congestion problems becoming more acute during the summer or winter months. So it is important for the trip generation process to incorporate time variations if the situation so demands and this is best handled by modeling the various time periods separately.

- Low use trips/modes. Low use trips/modes can also be lost in the aggregated nature of trip generation. Trip generation tables and travel surveys can often hide or overlook those trips that are rare for the individual but numerous enough to support a successful business in the whole. (Examples might include trips to make donations to the Salvation Army or to Boy Scout carwashes hosted by the local McDonalds.) Such trips may not ‘fit’ anywhere in the process, but can, or may, still be influential on traffic patterns. These type trips are difficult to isolate in the UTPP and most often are not fully accounted for. It would seem likely that the more diversified and specialized a city, the more such low use trips and modes would be in evidence.
- Multi-mode trips. In a similar fashion, multi-mode trips are not routinely accounted for. The trip generation process considers origins and destinations and for multi-modal trips to be fully incorporated, these origins and destinations would need to include mode change points, such as park and ride sites and van pool pick-up points. This could be done with more expansive trip generation procedures, but is most often omitted. Modes of travel other than the automobile are important primarily in the larger MPOs, but most often are not a major factor in the smaller regions. In larger regions, though, they can have a significant impact, with more multi-mode trips likely in areas with more transit or bike-pedestrian options.

- Transportation System Management/Transportation Demand Management (TSM/TDM) efforts. Travel behavior can be affected by external attempts to control or influence individual access to and use of the transportation network. Such attempts (e.g. encouraging the use of mass transportation modes or telecommuting during high ozone days, regulating access to freeways through ramp controls, and the use of tolling and other pricing mechanisms) can affect the trip generation, trip distribution, and the mode split processes, and can be captured in normal modeling procedures only by micro-analyzing the issues of concern. In other words, modelers would need to develop specific parameters for TSM/TDM activities and model them separately. As this would routinely require the allocation of additional resources, the costs and benefits would need to be considered carefully. Likely, this effort would not be worth the cost (especially for the small MPOs) except in the study of specific, and perhaps somewhat unique, policy issues.
- Zone structures. Because of the vast numbers of potential travel origins and destinations even in the smallest MPOs, such Os and Ds are typically grouped into zones called Traffic Analysis Zones (TAZs). All travel in and out of these zones is said to originate from the TAZ centroids, a notion that obscures the nature of the travel within a particular zone. When the zones are small, this may be only a minor problem, but when the zones are large, the problem can be quite substantial. Currently, TAZs, which are typically based on census tracts or portions thereof, are subdivided as traffic volumes increase to make the zonal size

less problematic.[8] Moreover, traffic congestion within TAZs is typically not a problem.

- Land use feedback. Most regions have land use zoning restrictions of some type to control the development of the urbanized area. Travel between TAZs regularly occurs between zones of different land uses (e.g., people leave home (areas zoned residential) to go to work (areas zone commercial or industrial)), but the resulting traffic patterns do not seem to figure into the zoning decisions made by municipal agencies. Indeed, the link between land use and the transportation grid is one of the weakest links in the urban planning process. This is widely recognized and there are many land use models that attempt to feed into travel demand models and strengthen this connection. Thus far, however, a good linking process remains elusive. Still, the variation and distribution in the land uses may be a factor that impacts on urban congestion.
- Behavior choices. Individual travel preferences form the basis for all travel, and capturing these preferences is difficult. Travel surveys, which try to identify the trips a population sample takes in a given time period, and trip generation tables, which identify the average number of trips a type of land use attracts, are the instruments of choice in codifying travel behavior. But these methods are lacking: what folks say they do and what they actually do are different, and land uses do not uniformly draw the same visitors over time and across geographies.

[8] It is possible, however, to eliminate this problem altogether. Today's PC technology allows the distribution of trips to and from their actual origins and destinations, without the use of TAZs. In his assessment of traffic patterns in Pocatello, ID, Horner (1998) found this technique to smooth out the traffic flow within TAZs and better capture the use of the transportation network. This approach does, however, require additional resources, and may not be worth the additional expense, especially in the smaller MPOs, where TAZs are fully adequate.

Additionally, the same individuals often exhibit different travel activity patterns depending on the combination of trips (trip chains) they are making. So travel behavior estimates are quite problematic and provide a shaky foundation for the entire travel demand modeling process. Still, surveying techniques are getting better, trip generation tables becoming more sophisticated, and household activity algorithms becoming more reflective of actual behavior. But there is clearly more improvement needed in this area.

2.6.3 Models of Urban Development. There are many models of urban development, whether they be predicated on the idea that cities formed for trade purposes before the advent of agriculture (Jacobs 1969) or required the development of agriculture to allow large numbers of people to gather and thrive in one location (Bairoch 1988). Three basic models that have since been improved upon in the detail and complexity that defines today's urban areas are briefly discussed below:

- Concentric Zone Model. Ernest Burgess of the famed Chicago School introduced the Concentric Zone Model (Dreidger 1991) where cities grew around a central business district (CDB) in rings, much like a tree. Rings tend to be business or class-based, with business occupying the inner rings and residential areas the outer rings. As the city's size increases, the rings push outward through the process of invasion and succession, with the best land going to the more dominant group (i.e., the group that controlled more wealth). Additionally, as the rings push outward, heterogeneity increases along with increased segregation and class differentiation.

- Sector Theory. Homer Hoyt noted that transportation advancements seemed to allow the bypass of Burgess's concentric zones and developed a Sector Theory (Dreidger 1991), where "pie slices" along the high speed transportation corridors were integrated into the concentric zone concept, effectively mixing up the city structure. The CBD was still the key economic sector but the other rings (light manufacturing and residential) followed the access offered by transportation technologies instead of remaining in their rings. Again, the best land, which commanded the highest "rent", went to the more dominant group.
- Multi-nuclei theory. Chauncey Harris and Edward Ullman attempted to improve on Hoyt's sector theory by suggesting that as a city grew, it diversified. This diversification engendered a diversification of land use as well, as people and businesses sought to take advantage of increased transportation options and cheaper land at the city's periphery. This resulted in a city that had more than one center of economic activity, a city with multiple nuclei (Dreidger 1991). While the CBD remained the dominant economic driver, other outlying business districts developed, often with their own concentric rings or supporting sectors.

The city structure has an impact on the supporting transportation network. No major urban areas are purely concentric zone or sector models, nor are they exactly how Harris and Ullman envisioned them in their Multi-Nuclei model; with increasing size, they generally exhibit traits of all three. The predominance of any one structure, or the spread of the three, is a factor of the underlying city's history, geography, and dynamics. The degree of centralization and spread of the economic drivers affect the resulting street network and the traffic flows over that network. Research by Meijers and Burger (2010) found that poly-centricity is associated with higher labor productivity, which may be because poly-centricity appears to be a function of urban growth and urban growth leads to knowledge spillover, which in turn enhances productivity. This suggests that it may be advantageous for businesses, over time, to locate outside the CBD to take advantage of this increased productivity. This, in turn, supports the observation that city organization gets more complex as the city ages.

2.6.4 Models of Land Use. The city structure is also affected by the use of the land, which is largely determined by the people and the free market, in conjunction with the zoning authority. There are a number of models used to project land use into the future, all of which begin with existing population and employment totals, the existing zoning structure and a knowledge of recent growth patterns. While future land use is important to planners and will shape city development, it is the current land use that is most important in this analysis. It seems likely that zoning patterns are correlated at least to some degree to congestion.

2.7 Points of Intersection between Theories and Concepts, Travel and Urban Structure

The theories and concepts that may affect travel behavior, the urban transportation models, and the underlying urban structure are the three circles in a Venn diagram, with the intersection of all three offering some insights into the understanding of urban congestion. These are discussed below within the template of supply, demand and flow.

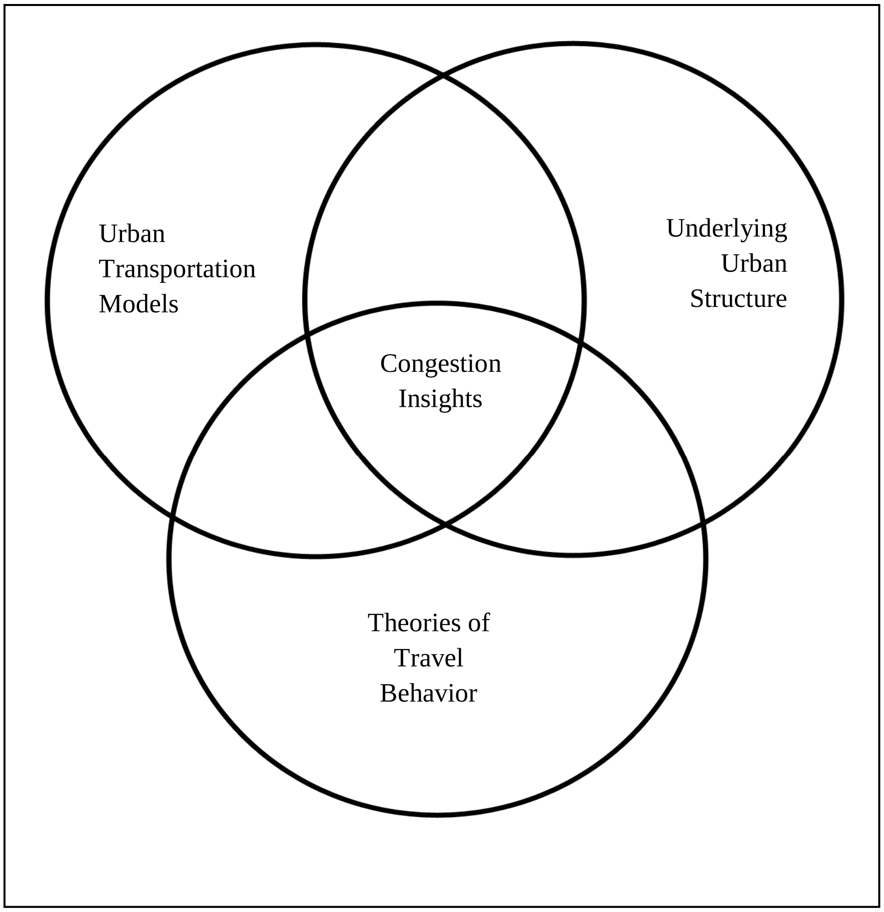


Figure 1: Relationships between travel behavior, urban transportation models and urban structure

2.7.1 Supply. Central place theory provides a basis for the rise of urban areas and structural functionalism provides a basis for their organization. Given that the needs of the city are fundamental, it seems clear that a transportation network would arise to support those needs, needs which are likely to be at least somewhat different in different urban areas depending on culture, expectations of city services and responsibilities, and city wealth. Land-rent theory further shaped urban area and the underlying street grid, with the supply of streets and roads directly related to the dimensions of the network, its size, density, robustness and connectivity. These dimensions, in turn, are affected by the structure and layout of the urban area. Cities following the concentric zone model commonly have a hub-and-spoke, radial network with all major arterials leading to the CBD. Sector zone and multi-nucleic cities have more complex networks with major arterials leading to the key sectors and the other nuclei in addition to the CBD. Networks with more intra-connectivity offer more routes of travel from origins to destinations and according to packet-switching network theory should offer faster and more consistent travel times. City age also has a role to play in both the urban form and the resulting transportation network. Needs change over time and older cities may have street networks that are inadequate to support the way the populace wishes to travel, but have limited options for improvement, primarily because of cost and a need for extensive, and perhaps excessive, use of eminent domain.

The supply of transportation is measured by the size, density, and robustness of the street network, as well as its ability to effectively link people with where they want to go.

- Network size is characterized by descriptive numbers without regard to the underlying population, the underlying area, or any other underlying feature. Measures of size include such metrics as miles of freeway, lane miles of freeway, miles of local streets, and numbers of links and nodes. Size measures are somewhat limited in their value when exploring congestion since congestion is a spatial and temporal phenomenon that typically occurs only in limited areas and during specific time periods. Size measures do not “get at” the concentration of vehicles in space-time. Nonetheless, one would expect cities with larger networks to have more congestion, simply because of the theory of structural functionalism. The cities built these networks because they were needed and the lag time between identifying the need and adding to supply usually means that urban areas are continually in a state of playing "catch up".

- Network Density is the size of the network in relation to another underlying parameter (population, commuter, area, etc.). Density can reflect a number of different circumstances, such as age of the city, structure of the city, use of the land, and under- or over-building of the highway network.
- Network Structure is the nature of the network in relation to itself. Given that the transportation network will follow the urban area network, this measure is concerned less with network layout (radial, grid, or natural) and more with the proportional sizes of the various functional classes of highway. One measure along these lines is the size of the freeway system in relation to the network as a whole. Different network structures will be able to carry different loads of traffic and have differing susceptibilities to congestion.
- Network Robustness is the carrying capacity of the network; i.e., the number of lanes, the speed limits, and the level of access of the roads in the grid. The more robust the network, the more traffic it can carry. This is particularly true of the freeway system as it tends to carry the most traffic, especially from the outskirts of the urban area where land rent theory posits people will choose to live because of housing costs.
- **N**etwork intra-connectivity is a measure of the number of nodes in the network, which allow the movement of vehicles from one link to another. The more connected the network, the more routes that "packets" have from origin to destination.

- Network inter-connectivity is a measure of the ability for people outside the network area (often taking advantage of cheaper or more attractive places to live) to access the attractions in the urban area and also the ability of people inside the network to access attractions outside the urban area. Higher speed roads connecting the urban area with the hinterland (surrounding area and nearby towns) are essential to good inter-connectivity.

2.7.2 Demand. The driving principle behind transportation demand is that it is derived. While there is the occasional joyride, the vast majority of travel is to go a destination. The destinations of choice are explained, at least partially by a number of behavioral theories. Rational choice theory posits that travelers make rational decisions, thereby maximizing their utility. Social exchange theory would explain the basis for the utility that is being maximized. Both reference group theory and interest group pluralism suggest that the social networks to which travelers belong can sometimes override their rational decision-making or at least alter what travelers may consider rational. These last three theories are sociologically-based and are affected by cultural norms and likely lead to different outcomes in different urban areas and in different parts of the country. Regardless the motivation, however, people will make their travel decisions to best fit their needs. Unfortunately, these decisions are commonly made without full consideration of the external costs of using the network at that point in time and along the chosen route. Hence, demand can overwhelm supply at points in time and space and result in congestion, which is a tragedy of the commons.

As noted, the demand for transportation is derived, so it would be affected by the characteristics of the travelers as well as by the characteristics of the places they go. The four-step model estimates this demand by identifying productions and attractions; their numbers are important but their nature may matter even more, because it may cause the distribution of the trips be different than might be expected if all productions and attractions were uniformly “vanilla”. People may choose “more exotic flavors” of destinations that are farther from their origins than “more vanilla” destinations nearby.

- Internal Productions are the points of origin for the trips generated within the area covered by the transportation network. These trips generated trips are derived – travelers have to want to go somewhere. Their desire and ability to go, however, is driven by the characteristics of the travelers, which include the numbers of travelers, the density of travelers (in relation to the area or the network), the incomes of travelers (wealthier people tend to travel more)(Balaker and Staley 2006), the levels of automobile ownership (absolute or in relation to the area or the network), and the levels of solo automobile commuting (absolute or in relation to the area or the network). Internal productions are also affected by the numbers of people who may not travel at all because of low income or age. These traveler/non-traveler characteristics are influenced by the reference groups and interest groups with which the individuals identify, which in turn have a regional component.
- Internal Attractions are the points of destination for the trips generated within the area covered by the transportation network. These are driven by the characteristics of the attractions; e.g., the size of the employment centers, restaurants, and malls (which indicates their “pull” within the gravity model), the number and types of specialty shops (which may indicate the degree of variety in

the attractions which may propel people to travel farther than the gravity model would suggest), and the model of urban development (which might cause people to travel farther to get to work, school or shopping than in other models of development). These metrics can be assessed as absolute measures or in relation to population, area or the network. Like the internal productions above, the attractions also have cultural and regional components – attractions can be more or less attractive in different parts of the country.

- External Productions are the points of origin for the trips generated outside the area covered by the transportation network and are similar in nature to internal productions. These are measured by the inflows of people from outside the urban area; e.g., people from surrounding areas commuting to employment, attending local schools, or shopping at local businesses.
- External Attractions are the points of destination for the trips generated outside the area covered by the transportation network and are similar in nature to internal attractions. These measured by the outflows of people from the urban area; e.g., people commuting to outside employment, attending outside schools, or shopping at outside businesses.
- Mode Split determines the number of trips that is assigned to the street network and is affected by the numbers of travelers who choose not to travel by car (users of transit and bicycles, pedestrians, and telecommuters). Key determinants of this choice is car ownership (individuals without access to an automobile are more likely to use transit), age (older and younger people tend to have less access to cars), and the penetration and extend of the available transit services.

2.7.3 Flow. As people move out onto the network (the supply) to go to their choice of destinations (the demand), they move into the flow where their choices will interact with the choices of the other users of the network to determine the level of service (LOS) for the facility. If they behave as good stewards of the facility, they can minimize the adverse effects their driving behavior will have on flow. If they are inattentive, or engage in overly aggressive or loss averse behavior, they can negatively affect throughput and unnecessarily increase congestion. Regardless of their behavior, there will be interruptions in flow caused by other factors that may cause queues to develop, which will need to be resolved per queuing theory. These other factors may be characteristics of the drivers and vehicles that use the network, the network itself, or the non-recurring interruptions of flow **(**traffic incidents, work zones, weather, and special events) discussed in Section 2.3.2 above.

- Trucks affect the flow by slowing it down and reducing throughput, so the number of trucks using the network (absolute or in relation to the population, the area or the network) would make a difference.
- Distracted drivers impede the flow of traffic in a number of ways: by failing to maintain pace with the flow; by delaying unnecessarily at intersections; and, by driving erratically, which forces other drivers out of a smooth traffic flow.
- Intersections with traffic signals and stop-controlled signage are designed to interrupt the flow of traffic to allow access from other links in the network. In crowded conditions, this can lead to queues and congestion. Poorly timed signals or ill-conceived signage can add to this congestion, often significantly. However, even well-designed signals and signage can have a detrimental impact on travel

speeds, which is a key reason freeways, which do not have intersections, typically allow for faster travel times.

- Traffic Incidents impede the smooth flow of traffic and the more accidents or breakdowns the more the impedance. Even with identical accident rates, urban areas with more vehicle miles of travel (VMT) will have more accidents, which will, in turn, lead to increased congestion.
- Work Zones impede the smooth flow of traffic and the more work zones the more impedance.   Even with similar construction rates, urban areas with more VMT will have more construction and hence a larger impact on congestion.
- Weather may impede the smooth flow of traffic, but bad weather is fairly widespread and travelers routinely adapt to the prevailing conditions. Still urban areas with more adverse weather may experience a larger impact on congestion that areas with more benign weather patterns.
- Special Events impede the smooth flow of traffic and the more special events the more impedance.   Urban areas with more special events will have a larger impact on congestion.

CHAPTER 3: THE RESEARCH IN CONTEXT

3.1 Congestion Research to Date

Congestion is a practical problem, and also an expensive one as noted above. Consequently, most recent congestion research has tended to focus along three axes: impacts, mechanics and remediation. These perspectives are all practical and lend themselves to developing or supporting public policy. Theoretical aspects of congestion have received less attention. There has been some additional work to describe congestion further and examine its causes, but this has occurred generally within the context of standardizing terminology and immediate causation. Congestion researchers have sometimes explored the causes of congestion from their own perspective, but they generally have limited themselves to the immediate causes (demand exceeding supply for whatever reason) and often have begun with the reality that congestion already exists. To date no one seems to have taken a holistic approach to the underlying urban dynamics that result in congestion.

There has, however, been research on aspects of urban travel that are related to congestion, such as travel behavior and vehicle usage. These factors have been weighed against various characteristics of the urban arena (density, education, income, sustainability, etc.) to assess relationships. Since these studies often suggest or infer impacts on traffic congestion, they are included in the review of the literature.

3.1.1 Congestion Impacts. A key focus of transportation researchers has been to quantify the impacts of traffic congestion. These impacts are known to be large and their quantification is likely to draw attention to the congestion problem and beg some resolution. Consequently, the lead agencies in this research have been the government, universities and privately funded think tanks. The Texas Transportation Institute (TTI) is one of the key players in this area and publishes a frequently cited annual Urban Mobility Report. The 2012 report (Schrank, Eisele, and Lomax 2012) uses measured traffic data for selected urban areas (provided by INRIX, a private company and a leading provider of traffic information) to calculate the impacts of congestion in a variety of ways (wasted time, wasted fuel, wasted money, additional CO_2 emissions, and cost to shipping companies) and from several perspectives (total cost, cost per urban area, and cost per commuter). The Federal Highway Administration (FHWA) (2013b) provides aggregate data annually on the annual change in hours of congestion, travel time index (average congestion), and planning time index (worst-day congestion).

While the preceding research served to assess the generalized costs of congestion on urban regions as a whole, other research has focused on assessing the effects of congestion on specific aspects of the urban condition. Such aspects include regional productivity (Hartgen and Fields 2009; Prud'homme and Lee 1999), greenhouse gas emissions (Kelly 2012; Hartgen, Fields and Moore 2011), and sustainability (Li et al. 2012; Ramani et al. 2011). Finally, Llewelyn-Davies (2004) reviewed several hundred sources, many of which concerned congestion, for their assessment of the link between transportation and city competitiveness.

3.1.2 Congestion Mechanics. Much of the research on congestion impacts is at the macro level, which considers congestion as a regional phenomenon. Other research is conducted at the micro level and explores the mechanics of how congestion works. This tends to be the world of traffic engineers and proponents of specific policy issues. The various sections of the Highway Capacity Manual (TRB 2010) that are concerned with traffic flows and congestion are based on such research. Other authors have worked on details, exploring the links between congestion and two-way (vs. one-way) street networks (Gayah 2012; Gunay 2009), the transportation of school children (Wang, Campbell, and Parsons 2010), rain events (Watkins and Hallenbeck 2010), various non-recurring traffic events (Chin et al. 2002a and 2002b), cell phones (Yager 2013; Holden 2009; Strayer, Drews and Crouch 2006), roundabouts (Dahl and Lee 2012; Uddin 2011), ramp metering (Shen and Zhang 2010; Varaiya 2008), left turn assessments (Yu and Prevedouros 2013; Chowdury et al. 2005), and signalization (Aziz and Ukkusuri 2012; Wu and Liu 2011).

3.1.3 Congestion Remediation. Most, but not all, of the research on congestion impacts and mechanics tends to set the stage for suggestions for congestion remediation. Moreover, there are additional documents that propose methods to alleviate congestion that do not include discussions of the impacts or mechanics, but instead build upon those that do. The congestion problem is an expensive one and there is no shortage of potential solutions. Remediation efforts are discussed in terms of supply, demand and flow.

- Supply-oriented solutions generally involve adding new capacity or using exiting capacity more efficiently. Cox and Pisarski (2004) and Hartgen and Fields (2006) argue for more roads and Zimmerman et al. (2011) make the case for increased network connectivity. Supply solutions are not without their detractors. A number of researchers (e.g., Duranton and Turner 2011; Downs 2004; Cervero 2003a; Cervero 2003b, Littman 2001, and Hansen 1995) find that generated traffic, also known as induced traffic, will eat up much of the increased capacity offered by the new construction. This is not necessarily a bad thing – induced travel can also lead to induced growth and induced investments, which can add to the urban area's vitality and attractiveness. Still, this does not help the congestion remediation efforts.

- Demand-oriented solutions generally involve decreasing demand for travel through pricing strategies (Glaister and Graham 2006), changes in travel behavior (Strickland and Berman 1995; Viegas 2001), increased modal alternatives (Crampton 2000; Aftabuzzaman, Currie and Sarvi 2010; Aftabuzzaman 2011), changes in urban design (Zhao, Luë, and de Roo 2010; Boarnet 2008; Buliung and Kanaroglou 2006; Crane and Chatman 2003; Cervero 2001) or some combination of these (Smart Growth America 2012; Littman 2012).
- Flow-oriented solutions generally involve improved traffic operations (Hensher 2003; FHWA 2012a) and improved public information to allow drivers to make more informed decisions (Ogunbodede 2007).

These various strategies are summed up in good detail by Cambridge Systematics (2005) and the Federal Highway Administration (2013).

3.1.4 Congestion Description and Discussion. There has, of course, been other congestion research with a focus other than impacts, mechanics and remediation efforts. Topics are varied and include comparing various mobility and congestion measures

(Bertini 2005), developing improved measures of congestion (Maitra, Sikdar and Dhingra 1999; Mallinckrodt 2010), differentiating between recurrent and non-recurrent congestion (McGroarty 2010; Skabardonis, Varaiya, and Petty 2002), exploring the link between congestion and economic development (OECD/ECMT 2007), and changing how travel behavior research is done (Gärling, Gillholm, and Gärling 1998).

Finally, there is research to support the notion that congestion is here to stay (Duranton and Turner 2011). Indeed, Downs (2004, p. 20) says that "Traffic congestion is not essentially a problem. It's the solution to our basic mobility problem." He further asserts (p. 21) that "Peak-hour congestion is the balancing mechanism that makes it possible for Americans to pursue goals they value, such as working while others do, living in low-density settlements, and having many choices of places to live and work." If this is true, then perhaps most research on congestion is an academic drill with little practical application.

3.1.5 Studies Related to Congestion. There have also been other studies that have looked at travel behavior and vehicle usage in relationship to characteristics of travelers and their environment. These studies do not address congestion directly, but since travel behavior and vehicle usage are key components of congestion, they merit a quick look. It seems to be commonly accepted that increased car ownership and family wealth lead to increased travel. Martin and McGuckin (1998) use the level of car ownership as a factor in estimating trips and Balaker and Staley (2006) argue that the increases in wealth that generate more travel is a good thing. Both assessments find links with increased travel, which could lead to increased congestion. Baum-Snow (2007) investigated the relationship between highway development and intercity growth and concluded that "one new highway passing through a central city reduces its population by about 18 percent" (p. 2.), suggesting that freeways promote an exodus from the city center, which has implications (both positive and negative) for congestion. Research in sustainable development regularly considers the relationships between development/city form/land use and travel. One such study (Boarnet and Crane 2001) found that while land use can affect the price of travel, the evidence of a link with increased travel is mixed; i.e., sustainable development practices cause some people to travel more and others to travel less.

3.2 The Contribution of this Research to Understanding Congestion

Regardless of whether or not congestion research serves a practical purpose, it can serve an academic one. It is hoped that this research will extend the understanding of the underlying foundations of congestion. While the research to date has focused on the immediate causes of congestion and the necessary steps to resolve the congestion issue, none has considered the underlying urban characteristics that are present when congestion arises. This study seeks to identify these correlates of congestion and assess their relative importance.

CHAPTER 4: RESEARCH DESIGN

4.1 Overview

The purpose of this study is to uncover the set of urban characteristics that are correlated with traffic congestion. This involves identifying measures of congestion for assessment, specifying a study area and time frame for which there are a wide variety of potential correlates, selecting key datasets from which to gather the potential correlates, isolating a number of urban characteristics from the multitude of possible variables available based on theories and the literature, and identifying methods of analyses that can handle this number of variables, which are likely to be correlated with one another, often highly so. Of these steps, the first, and most fundamental, is identifying measures of congestion, so we begin with this.

4.2 Measures of Network Congestion

There are two basic types of data: primary data, which are collected first hand by individual researchers/research teams for use in their own studies, and secondary data, which are collected by other people or organizations and then used by individual researchers/teams in their studies. Primary data can be quite expensive, often prohibitively so, and moreover can be extremely difficult to collect. These issues steer researchers/teams to use secondary data, which have their own set of problems. Often, secondary data were not collected for the purpose(s) of the study, so care must be taken to ensure that the data are appropriate. Care must also be taken to ensure that data come from a single source so that measurement is consistent across the dataset. In this light, existing studies of congestion that have a national scope and a wide variety of city sizes are reviewed in search of measures of congestion; specifically, measures of the three dimensions of congestion – intensity, extent and duration. All of these measures are found in the Urban Mobility Report (UMR) using INRIX data, published annually since 1992 by the Texas Transportation Institute. (It should be noted that the UMR uses roadway inventory data from the Federal Highway Administration's Highway Performance Monitoring System (HPMS) in most of its calculations. Beginning in 2011, private sector traffic speed data from INRIX data was incorporated into the mobility performance measures.)

The UMR contains a wealth of data for the years 1992-2011 and will serve as the foundation for this analysis. (Given this 20-year period of data availability, 2010 is selected as the base year of the study in anticipation of using published 2010 census data for at least some of the potential predictor variables.) In addition to the three measures of congestion, data include two key demographic metrics (population and number of commuters) and key network metrics (lane miles and daily vehicle miles of travel (VMT) for both freeways and arterials). There are other data concerned with calculations of delay, fuel costs, and value of time that are not used in this assessment.

Several potential measures of congestion are identified in Section 2.4.3. Of these, the UMR uses three network-wide measures, which characterize the three dimensions of congestion. These measures are summarized in Table 3 and discussed in the ensuing paragraphs. The 2010 data taken directly from the UMR are used for each of these measures.

Table 3: Measures of network congestion

VARIABLE		EXPECTED		
CONCEPTUAL	**OPERATIONAL**	**EFFECT**	**SOURCE**	**JUSTIFICATION**
Congestion – Intensity	Travel Time Index (TTI)	NA - Dependent variable	UMR	Measure of congestion intensity Used in UMR
Congestion – Extent	Portion of lane miles that are congested	NA - Dependent variable	UMR	Measure of congestion extent Used in UMR
Congestion – Duration	Length of peak periods	NA - Dependent variable	UMR	Measure of congestion duration Used in UMR

- Congestion – Intensity (Travel Time Index (TTI)). The TTI is a solid measure of the intensity of traffic congestion (see Section 2.4.3). Moreover, it is widely used and accepted. Specific details of its calculation are in the UMR (Schrank, Eisele and Lomax 2012), but in general, vehicular speeds are measured at specific points on the network at various times in the day and then generalized across the network. Finally, travel times in the peak hours are compared to travel times in the off-peak hours to derive the TTI (recall that the TTI is the ratio of travel time in the peak to travel time in the off-peak).
- Congestion – Extent (Portion of Lane Miles that are Congested). This metric is a clear and understandable measure of the extent and scope of the congestion problem. Additionally, its calculation is relatively straightforward as explained in the UMR (Schrank, Eisele and Lomax 2012). In general, the UMR uses measured vehicular speeds, generalized across the network, to calculate the percent of the freeway and arterial lane miles that are congested, with the idea that once free flow speeds are reduced to a certain point (depending on the road classification), then the road is congested. The UMR, however, calculates the *percent* of lane miles that are congested. The *portion* of lane miles that are congested (or percent

in decimal form) is used in this analysis instead, so that that the decision tree analysis works properly.

- Congestion – Duration (Length of Peak Periods). The duration of the congestion problem is perhaps best measured by the congested hours data recorded in the quarterly Urban Congestion Report (hours: minutes of congested travel per weekday). Unfortunately, these data are available only for a small number of cities. In lieu of this metric, the length of the peak travel periods (specifically, the number of rush hours) in the UMR is used (Schrank, Eisele and Lomax 2012). This calculation is explained in detail in the UMR, but in general, TTIs are calculated for each hour in the day, and rush hours are then derived from the TTIs.

4.3 The Selected Urban Areas

The UMR is selected as the foundation for this analysis because of the availability of the three congestion variables. This will necessarily limit the studied urban areas to those included in the UMR.

The 2012 UMR identifies 498 urban areas in the United States (Schrank, Eisele, and Lomax 2012). Of these 498, the UMR includes hard, measured data for 101, which are grouped based on population: 15 very large (more than 3 million people); 32 large (1 million to 3 million); 33 medium (500,000 to 1 million); and, 21 small (less than 500,000). The hard data include the three congestion variables used as dependent variables in the analysis, which restricts the selected urban areas to these 101. Recognizing that congestion is a function of the supply, the demand and the flow and that these traits are fundamentally affected by culture and personal preferences, this study will focus on US cities only; specifically those in the contiguous 48 states, Alaska and Hawaii. Cities in US territories will not be considered, so the territorial city of San Juan, Puerto Rico is dropped from the list. The selected urban areas, now an even 100, include a range of city sizes and are geographically dispersed.

This set of 100 urban areas is not a random sample, but instead is "top heavy." All the very large and large urban areas in the US are included in the study set (except San Juan, PR) (100% of the very large and large cities in the 50 states and the District of Columbia), as are 33 of the 36 mid-size urban areas (92%). Of the 415 remaining small urban areas only 21 (5%) are represented and these are in the upper half in population. Boulder, CO, with a population of about 150,000, is the smallest city in the study area. This over-representation of the more populated areas is not expected to be a problem; however, as it is in the more populated areas where congestion is a real problem and a random sample could hinder the identification of factors of congestion. In a national assessment of the urban congestion problem, Hartgen and Fields (2006) estimated the costs of the additional highway capacity needed to eliminate severe congestion, which was defined as a Travel Time Index (TTI) of at least 1.18. To arrive at this cost estimate, they calculated the TTI for all urbanized areas in the US for 1995, 2003 and 2030. When cross-referencing these TTIs with city populations, severe congestion was most always present in cities with populations of 700 thousand or more, commonly present in cities with populations of 300-700 thousand, and rarely present in cities with populations of 150-300 thousand. For cities smaller than 150,000, severe congestion was never present or estimated to be present in 2030. While the TTI is but one measure of congestion and it is a regional measure that would not preclude severe congestion occurring at specific points in the network and at specific times, the Hartgen-Fields study seems to support the concentration of generalized severe congestion in the larger urban areas. Moreover, as the goal of the study is not to study congested cities in and of themselves, but rather to uncover the urban characteristics that are most linked to congestion, it is essential that the larger urban areas where congestion is a problem be well represented in the sample. It is most likely that a random sample would include fewer of the larger cities and thus make the identification of the targeted urban characteristics much more difficult. In essence, it seems, to achieve the goal of the study, the sample must be "top-heavy."

The selected urban areas are shown on the map (highlighted by population grouping) and in the table (grouped by population size) below.

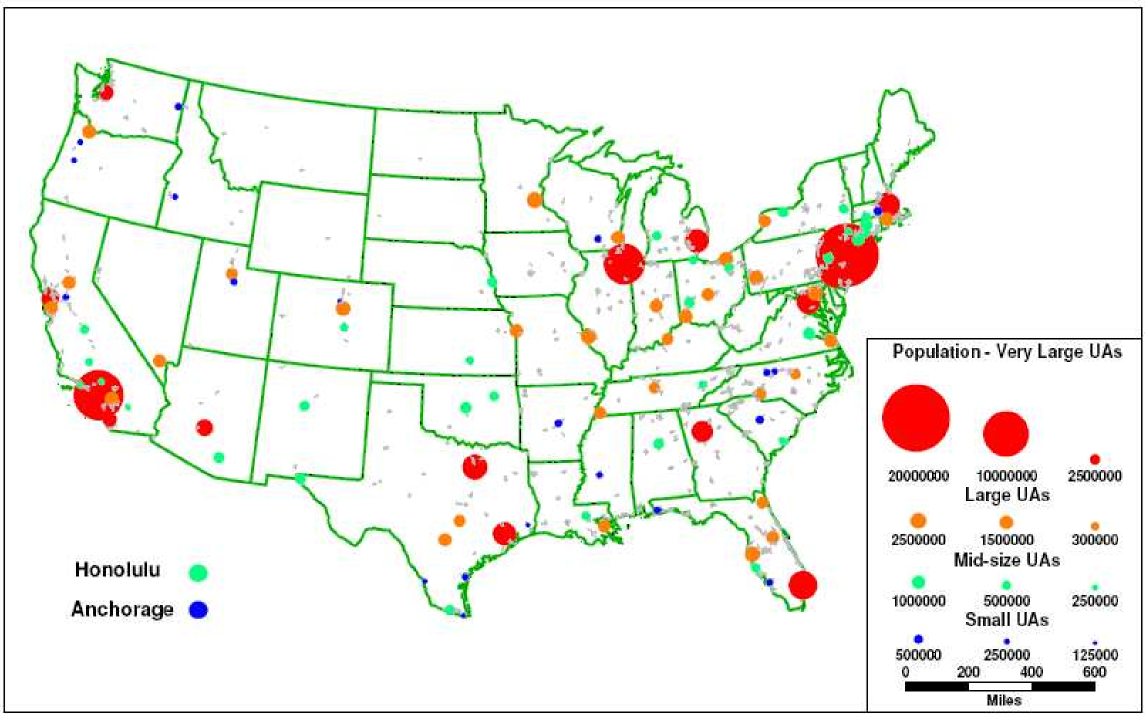


Figure 2: Selected 100 urban areas by population group

Table 4: Selected 100 urban areas by population group

Very Large Urban Areas (15 areas)	**Large Urban Areas (31 areas)**	**Mid-size Urban Areas (33 areas)**	**Small Urban Areas (21 areas)**
Atlanta GA	Austin TX	Akron OH	Anchorage AK
Boston MA-NH-RI	Baltimore MD	Albany NY	Beaumont TX
Chicago IL-IN	Buffalo NY	Albuquerque NM	Boise ID
Dallas-Fort Worth-Arlington TX	Charlotte NC-SC	Allentown-Bethlehem PA-NJ	Boulder CO
Detroit MI	Cincinnati OH-KY-IN	Bakersfield CA	Brownsville TX
Houston TX	Cleveland OH	Baton Rouge LA	Cape Coral FL
Los Angeles-Long Beach-Santa Ana CA	Columbus OH	Birmingham AL	Columbia SC
Miami FL	Denver-Aurora CO	Bridgeport-Stamford CT-NY	Corpus Christi TX
New York-Newark NY-NJ-CT	Indianapolis IN	Charleston-North Charleston SC	Eugene OR
Philadelphia PA-NJ-DE-MD	Jacksonville FL	Colorado Springs CO	Greensboro NC
Phoenix-Mesa AZ	Kansas City MO-KS	Dayton OH	Jackson MS
San Diego CA	Las Vegas NV	El Paso TX-NM	Laredo TX
San Francisco-Oakland CA	Louisville KY-IN	Fresno CA	Little Rock AR
Seattle WA	Memphis TN-MS-AR	Grand Rapids MI	Madison WI
Washington DC-VA-MD	Milwaukee WI	Hartford CT	Pensacola FL-AL
	Minneapolis-St. Paul MN	Honolulu HI	Provo-Orem UT
	Nashville-Davidson TN	Indio-Cathedral City-Palm Springs CA	Salem OR
	New Orleans LA	Knoxville TN	Spokane WA-ID
	Orlando FL	Lancaster-Palmdale CA	Stockton CA
	Pittsburgh PA	McAllen TX	Winston-Salem NC
	Portland OR-WA	New Haven CT	Worcester MA-CT
	Providence RI-MA	Oklahoma City OK	
	Raleigh-Durham NC	Omaha NE-IA	
	Riverside-San Bernardino CA	Oxnard CA	
	Sacramento CA	Poughkeepsie-Newburgh NY	
	Salt Lake City UT	Richmond VA	
	San Antonio TX	Rochester NY	
	San Jose CA	Sarasota-Bradenton FL	
	St. Louis MO-IL	Springfield MA-CT	
	Tampa-St. Petersburg FL	Toledo OH-MI	
	Virginia Beach VA	Tucson AZ	
		Tulsa OK	
		Wichita KS	

Tables 5 and 6 show the urban areas grouped by UMR size category and census region, respectively. In both tables, urban areas are sorted within the groupings by population. Congestion variables are shaded based on the quartile, with darker shades indicating higher measures of congestion. These tables are good reference points when considering the findings from this analysis. Knowing where cities are and relatively how large they are in their neighborhoods aids in the understanding and interpretation of the results. Note that the averages based on population grouping show that the larger urban areas are worse than the smaller urban areas in each dimension of congestion (a detail reinforced by the quartile shading). The averages based on geographic groupings, however, show a more mixed result (also reinforced by the shading). For TTI, the Northeast region has the highest average and the Mid-west the lowest; for PortLMCong, the West region has the highest and the Mid-west the lowest; and, for PkHrs, the Northeast region has the highest and the South the lowest. So in general, it appears that if all other factors are equal, it pays to be a smaller city in the Mid-west region, as far as congestion is concerned. All other factors, however, are rarely equal, but which factors matter?

Table 5: Urban area congestion measures by UMR size category, sorted by population, 2010

Urban Area	**TTI**	**PortLM Cong**	**PkHrs**
Very Large Urban Areas			
New York-Newark	1.33	0.52	6.75
Los Angeles-L. Beach-S. Ana	1.37	0.62	8.00
Chicago	1.25	0.70	5.25
Miami	1.25	0.80	5.00
Philadelphia	1.26	0.54	5.00
Dallas-Fort Worth-Arlington	1.25	0.43	5.00
Washington	1.31	0.68	7.00
Atlanta	1.24	0.58	5.00
Boston	1.28	0.39	5.00
San Francisco-Oakland	1.22	0.58	6.00
Houston	1.26	0.48	5.75
Detroit	1.18	0.47	5.00
Phoenix-Mesa	1.18	0.51	5.00
Seattle	1.26	0.46	6.00
San Diego	1.18	0.58	5.00
Large Urban Areas			
Minneapolis-St. Paul	1.21	0.34	5.00
Baltimore	1.23	0.57	4.25
Tampa-St. Petersburg	1.20	0.65	4.00
St. Louis	1.14	0.25	4.00

Urban Area	**TTI**	**PortLM Cong**	**PkHrs**
Mid-size Urban Areas			
Oklahoma City	1.15	0.36	2.50
Richmond	1.11	0.36	2.50
Bridgeport-Stamford	1.27	0.40	5.00
Hartford	1.18	0.29	3.50
Birmingham	1.19	0.39	3.25
Rochester	1.13	0.18	2.50
Dayton	1.11	0.24	2.50
El Paso	1.21	0.25	3.50
Honolulu	1.36	0.51	4.25
Tucson	1.16	0.58	2.50
Tulsa	1.12	0.37	2.50
Oxnard	1.10	0.42	2.50
Fresno	1.08	0.38	2.50
Sarasota-Bradenton	1.12	0.56	2.50
Omaha	1.11	0.39	2.50
Allentown-Bethlehem	1.17	0.42	2.50
Springfield	1.13	0.24	2.50
Albuquerque	1.10	0.33	2.50
Akron	1.12	0.19	2.50
New Haven	1.17	0.29	3.25

Urban Area	TTI	PortLM Cong	PkHrs
Denver-Aurora	1.27	0.58	5.25
Riverside-San Bernardino	1.23	0.51	4.00
Portland	1.28	0.50	4.50
Sacramento	1.20	0.60	4.00
San Jose	1.24	0.62	6.00
Pittsburgh	1.24	0.34	4.00
Cincinnati	1.20	0.35	4.00
Cleveland	1.16	0.21	4.00
Kansas City	1.13	0.23	4.00
Virginia Beach	1.20	0.44	4.00
San Antonio	1.19	0.45	4.00
Milwaukee	1.15	0.26	4.00
Orlando	1.20	0.74	4.00
Las Vegas	1.20	0.58	4.00
Austin	1.31	0.48	5.50
Columbus	1.18	0.36	4.00
Providence	1.16	0.34	4.00
Indianapolis	1.17	0.56	4.00
Nashville-Davidson	1.23	0.48	4.00
Raleigh-Durham	1.14	0.51	4.00
Louisville	1.18	0.49	4.00
Jacksonville	1.14	0.50	4.00
Charlotte	1.20	0.51	4.00
Memphis	1.18	0.30	4.00
New Orleans	1.20	0.36	4.00
Buffalo	1.17	0.21	4.00
Salt Lake City	1.14	0.53	4.00
Averages			
Very Large Urban Areas	1.25	0.56	5.65
Large Urban Areas	1.20	0.45	4.21
Mid-size Urban Areas	1.14	0.34	2.88
Small Urban Areas	1.12	0.33	1.67
Top Quartile			
2nd Quartile			
3rd Quartile			
Bottom Quartile			

Urban Area	TTI	PortLM Cong	PkHrs
Albany	1.16	0.32	2.50
Grand Rapids	1.09	0.33	2.50
Baton Rouge	1.22	0.39	5.00
Lancaster-Palmdale	1.08	0.32	2.50
Indio-Cath.City-Palm Springs	1.08	0.37	2.50
McAllen	1.16	0.40	2.50
Colorado Springs	1.13	0.27	3.00
Poughkeepsie-Newburgh	1.12	0.37	2.50
Bakersfield	1.11	0.30	2.50
Charleston-North Charleston	1.15	0.50	4.25
Toledo	1.13	0.23	2.50
Wichita	1.09	0.09	2.50
Knoxville	1.16	0.32	2.50
Small Urban Areas			
Columbia	1.11	0.39	1.50
Provo-Orem	1.14	0.35	1.50
Cape Coral	1.15	0.39	2.25
Little Rock	1.07	0.32	2.00
Worcester	1.13	0.28	1.50
Jackson	1.10	0.22	1.50
Stockton	1.10	0.31	1.50
Madison	1.11	0.30	1.50
Winston-Salem	1.11	0.23	1.50
Spokane	1.12	0.15	1.75
Pensacola	1.11	0.37	1.50
Greensboro	1.10	0.28	1.50
Corpus Christi	1.04	0.16	1.50
Boise	1.06	0.48	2.00
Anchorage	1.18	0.26	1.50
Eugene	1.08	0.23	1.50
Salem	1.14	0.29	1.50
Beaumont	1.10	0.15	1.50
Laredo	1.14	0.23	1.50
Brownsville	1.18	0.24	1.50
Boulder	1.18	0.22	3.00

Table 6: Urban area congestion measures by census region, sorted by population, 2010

Urban Area	TTI	PortLM Cong	PkHrs
Urban Areas in the South Region			
Miami	1.25	0.80	5.00
Dallas-Fort Worth-Arlington	1.25	0.43	5.00
Washington	1.31	0.68	7.00
Atlanta	1.24	0.58	5.00

Urban Area	TTI	PortLM Cong	PkHrs
Urban Areas in the Northeast Region			
New York-Newark	1.33	0.52	6.75
Philadelphia	1.26	0.54	5.00
Boston	1.28	0.39	5.00
Pittsburgh	1.24	0.34	4.00

Urban Area	TTI	PortLM Cong	PkHrs
Houston	1.26	0.48	5.75
Baltimore	1.23	0.57	4.25
Tampa-St. Petersburg	1.20	0.65	4.00
Virginia Beach	1.20	0.44	4.00
San Antonio	1.19	0.45	4.00
Orlando	1.20	0.74	4.00
Austin	1.31	0.48	5.50
Nashville-Davidson	1.23	0.48	4.00
Raleigh-Durham	1.14	0.51	4.00
Louisville	1.18	0.49	4.00
Jacksonville	1.14	0.50	4.00
Charlotte	1.20	0.51	4.00
Memphis	1.18	0.30	4.00
New Orleans	1.20	0.36	4.00
Oklahoma City	1.15	0.36	2.50
Richmond	1.11	0.36	2.50
Birmingham	1.19	0.39	3.25
El Paso	1.21	0.25	3.50
Tulsa	1.12	0.37	2.50
Sarasota-Bradenton	1.12	0.56	2.50
Baton Rouge	1.22	0.39	5.00
McAllen	1.16	0.40	2.50
Charleston-North Charleston	1.15	0.50	4.25
Knoxville	1.16	0.32	2.50
Columbia	1.11	0.39	1.50
Cape Coral	1.15	0.39	2.25
Little Rock	1.07	0.32	2.00
Jackson	1.10	0.22	1.50
Winston-Salem	1.11	0.23	1.50
Pensacola	1.11	0.37	1.50
Greensboro	1.10	0.28	1.50
Corpus Christi	1.04	0.16	1.50
Beaumont	1.10	0.15	1.50
Laredo	1.14	0.23	1.50
Brownsville	1.18	0.24	1.50
Urban Areas in the Midwest Region			
Chicago	1.25	0.70	5.25
Detroit	1.18	0.47	5.00
Minneapolis-St. Paul	1.21	0.34	5.00
St. Louis	1.14	0.25	4.00
Cincinnati	1.20	0.35	4.00
Cleveland	1.16	0.21	4.00
Kansas City	1.13	0.23	4.00
Milwaukee	1.15	0.26	4.00
Columbus	1.18	0.36	4.00
Indianapolis	1.17	0.56	4.00
Dayton	1.11	0.24	2.50

Urban Area	TTI	PortLM Cong	PkHrs
Providence	1.16	0.34	4.00
Buffalo	1.17	0.21	4.00
Bridgeport-Stamford	1.27	0.40	5.00
Hartford	1.18	0.29	3.50
Rochester	1.13	0.18	2.50
Allentown-Bethlehem	1.17	0.42	2.50
Springfield	1.13	0.24	2.50
New Haven	1.17	0.29	3.25
Albany	1.16	0.32	2.50
Poughkeepsie-Newburgh	1.12	0.37	2.50
Worcester	1.13	0.28	1.50
Urban Areas in the West Region			
Los Angeles-L. Beach-S.Ana	1.37	0.62	8.00
San Francisco-Oakland	1.22	0.58	6.00
Phoenix-Mesa	1.18	0.51	5.00
Seattle	1.26	0.46	6.00
San Diego	1.18	0.58	5.00
Denver-Aurora	1.27	0.58	5.25
Riverside-San Bernardino	1.23	0.51	4.00
Portland	1.28	0.50	4.50
Sacramento	1.20	0.60	4.00
San Jose	1.24	0.62	6.00
Las Vegas	1.20	0.58	4.00
Salt Lake City	1.14	0.53	4.00
Honolulu	1.36	0.51	4.25
Tucson	1.16	0.58	2.50
Oxnard	1.10	0.42	2.50
Fresno	1.08	0.38	2.50
Albuquerque	1.10	0.33	2.50
Lancaster-Palmdale	1.08	0.32	2.50
Indio-Cath. City-Palm Sprngs	1.08	0.37	2.50
Colorado Springs	1.13	0.27	3.00
Bakersfield	1.11	0.30	2.50
Provo-Orem	1.14	0.35	1.50
Stockton	1.10	0.31	1.50
Spokane	1.12	0.15	1.75
Boise	1.06	0.48	2.00
Anchorage	1.18	0.26	1.50
Eugene	1.08	0.23	1.50
Salem	1.14	0.29	1.50
Boulder	1.18	0.22	3.00
Averages			
Northeast Region	1.19	0.34	3.63
South Region	1.17	0.42	3.34
Mid-west Region	1.15	0.32	3.51
West Region	1.17	0.43	3.47

Urban Area	TTI	PortLM Cong	PkHrs
Omaha	1.11	0.39	2.50
Akron	1.12	0.19	2.50
Grand Rapids	1.09	0.33	2.50
Toledo	1.13	0.23	2.50
Wichita	1.09	0.09	2.50
Madison	1.11	0.30	1.50

Urban Area	TTI	PortLM Cong	PkHrs
Top Quartile			
2nd Quartile			
3rd Quartile			
Bottom Quartile			

4.4 Potential Predictor Variables

The selection of the most appropriate factors (predictor variables) is key to the validity of the analysis. The variables to be used with their expected effect, source, and justification are reflected in the tables below and discussed further in ensuing paragraphs. Most variables have a national scope to facilitate comparisons between urban areas; locally developed variables are avoided to the extent possible. Variables are linked where applicable to the discussion in Section 2.7 (Points of Intersection between Theories and Concepts, Travel and Urban Structure), with additional “wild card” variables added in an effort to address the congestion issue from additional angles.

As noted above, congestion is often strongly influenced by the size of the urban area, in that larger cities are more prone to congestion. With this in mind, variables are selected and expressed to control for the size effect. For the most part, variables are expressed in ratio terms in order to explore the underlying structural foundations of congestion. Still size is an issue and its importance in relation to structure is assessed with specific population and size variables.

The use of only US urban areas in this study allows the use of excellent sources of secondary data, to include US Census data in all its many forms (to include the American Community Survey), Federal Highway Administration (FHWA) highway statistics, and

TransCAD[9] GIS and network data. Data from each of these sources are reliable, well tested and almost universally accepted and are the basis for most of the selected variables. The specific data sources used in the calculations are noted for each variable.

There are some differences in the data years for some of the variables. The year 2010 is considered the base year of the study and data of that year are the target of the data collection effort, but there are many cases where 2010 data are not available. In these situations, the nature of the variables (i.e., that they are expressed in ratio terms) should serve to mitigate the problem. While many measures will change with the growth of a city over time, they often will change in tandem, so the ratios of the measures will likely show less variation.

4.4.1 Variables Impacting Supply. Table 7 summarizes the independent variables impacting supply, with each variable discussed in the ensuing paragraphs. The justification column includes references to theories and concepts discussed in Section 2.7 above.

Table 7: Variables impacting supply

VARIABLE		**EXPECTED**		
CONCEPTUAL	**OPERATIONAL**	**EFFECT**	**SOURCE**	**JUSTIFICATION**
Network* Size	Percent change in population 2000 to 2010	High/Negative	UMR	Measure of inadequacy of network size -Structural functionalism -Lag-time concept
	Political party control in 2000 (political affiliation of mayor)	Low/Negative	City Records World-statesmen website	Measure of transportation investments -Structural functionalism -Political party trends
Network* Density	Network miles per square mile	Mod/Positive	FHWA Census	Measure of network ability to accommodate demand -Structural functionalism

[9] TransCAD© is a geographical information system (GIS)-based transportation analysis platform produced by Caliper Corporation, Newton, MA (http://www.caliper.com).

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
	Freeway miles per square mile	Mod/Positive	FHWA Census	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
Network* Structure	Freeway lane miles per network lane mile	Mod/Positive	UMR FHWA	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
Network* Robustness	Freeway lane miles per thousand commuters	High/Positive	UMR	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
	Freeway miles + arterial miles per capita	Mod/ Positive	FHWA UMR	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
	City Age (Census urban area reached 50k in population (decades before 2010))	Low/Negative	Census Wikipedia city pages	Measure of network ability to accommodate demand -Structural functionalism -Changing urban needs over time
Network* intra-connectivity	Network nodes / Network links	Mod/Positive	TransCAD Census	Measure of available alternate routes -Structural functionalism -Packet-switching network theory

*Unless otherwise noted, the network includes all streets and highways in the urban area; i.e., those that one would find on Mapquest or Google Maps

- Network Size (Percent change in population from 2000 to 2010). As an urban area grows, there is a need for additional network capacity to accommodate this growth. More people require more roads. Unfortunately, there is a time lag between the identification of this growth need and the provision of the additional capacity. When an area grows very rapidly, any unused capacity tends to be used up rather quickly and congestion worsens before new capacity can be brought on line. The speed of growth, then, may be a factor affecting congestion, with more rapidly growing areas having a larger congestion problem. Calculation: the 2000 and 2010 populations for each urban area as listed in the UMR are compared to determine the percent change.

- Network Size (Political party control in 2000 as indicated by the political affiliation of the mayor). With population growth, additional transportation network capacity is needed. The mode of this additional capacity, however, can come in several flavors: additional highways and streets, increased public transportation, enhanced bicycle and pedestrian pathways, or some combination of these. Since most people choose to travel by car, and most municipal and private-provider services are delivered by cars and trucks (police protection, fire and emergency response, garbage collection, meter readings, mail/package delivery, lawn services, etc.), there is a need for additional highway and street mileage. Unfortunately, there are limited municipal resources, and this need for streets bumps up against the need for increases in the other modes of transportation. This allocation of resources issue is resolved in the political arena. Some anecdotal evidence suggests that Republicans tend to favor adding capacity to the street/highway network to allow people to follow their revealed preference for the automobile, while Democrats tend to favor the other modes in a bid to improve the sustainability of the urban area. Since traffic congestion is by definition a street/highway problem and the modal share of non-auto travel is quite small, the control of the municipal government by Republicans would seem to favor local congestion relief. Given the time lag in the identification of network needs to additional supply, party control of the system would likely lag as well. Even if the anecdotal evidence is incorrect, it is still of interest to determine if political control has any correlation with the levels of congestion. Still it should be noted here that not all local roads are controlled locally. Indeed, most of the higher functional classes of roads and highways are maintained with funds from outside the municipal government, e.g. from county, state and federal agencies. Calculation: the city websites are reviewed to determine the mayor in 2000 and his/her party affiliation at that time. Where the political affiliation is not readily apparent, additional websites (individual sites, collective sites such as World Statesmen (worldstatesmen.org), newspaper sites, etc.) are consulted to make this determination. Republican mayors are assigned a value of 0; Democratic mayors a value of 1.

- Network Density (Network miles per square mile). The ability to move about by car within an area is affected by the supply of streets and highways. A denser network, then, would seem to favor easier movement and less congestion. Calculation: total urban area network mileage from the Federal Highway Administration's (FHWA) Highway Statistics series (2010 Table HM-71) is divided by urban area square mileage from the 2010 census urban area list.
- Network Density (Freeway miles per square mile). As measured by vehicle miles of travel (VMT), automobile and truck travel occurs disproportionately on the upper level system, the arterials and freeways. In other words, more traffic is concentrated on the freeways than the local streets. A denser freeway network, like the network as a whole, would seem to favor easier movement and less congestion. Calculation: total urban area freeway mileage from the Federal Highway Administration's (FHWA) Highway Statistics series (2010 Table HM-71) is divided by urban area square mileage from the 2010 census urban area list.

- Network Structure (Freeway lane miles per network lane mile). The make-up of the street grid is an important factor in the ability to move from one point to another. A typical automobile trip might involve movement from a local street to a collector to an arterial to a freeway and back down the ladder to a local street. Since VMT is disproportionately on the freeways, the ratio of freeway lane miles to total network lane miles may be linked to congestion, with a higher ratio associated with lower congestion. Calculation: 2010 freeway lanes miles come directly from the UMR. Urban area freeway and arterial miles from the FHWA's Highway Statistics series (2010 Table HM-71) are subtracted from total network miles to get total non-freeway/arterial mileage, which is then multiplied by 2 to get non-freeway/arterial lane-mileage (assuming two lanes for all lower classes of streets). This total is added to urban area freeway lane-mile and arterial lane-mile totals from the 2010 UMR to get total network lane miles. Freeway lane miles are then divided by total network lane miles.
- Network Robustness (Freeway lane miles per thousand commuters). The robustness of the street grid is an indicator of how well it does its job, which is to allow vehicular movement. Since congestion tends to be more problematic during the morning and evening "rush hours," the ease of commuter movement would seem to be an important factor in the congestion issue. More freeway lane miles per commuter would likely be associated with less congestion. Calculation: 2010 urban area freeway lane-miles from the UMR are divided by 2010 urban area commuters from the UMR.

- Network Robustness (Freeway miles + arterial miles per capita). Another measure of the robustness of the street grid is the size of the upper level system relative to the population. More upper level system miles per person should translate into lower congestion levels. Calculation: freeway plus arterial miles from the Federal Highway Administration's (FHWA) Highway Statistics series (2010 Table HM-71) are divided by 2010 UMR population.
- Network Robustness (City Age – census year when the urban area reached 50,000 in population, measured in decades before 2010). The ability and political will to develop and maintain a robust street network are affected by a variety of factors. Perhaps most important is cost. Legacy cities, those cities that came into primacy before the street car era (prior to about 1890), have inherited an infrastructure and built-up area that were not constructed for the automobile. Costs, in both dollars and cultural destruction/change, to retrofit these cities for the automobile can be quite high and are often prohibitive. Newer cities have not had this level of constraint and city officials have had more latitude to respond to the increased infrastructure demands of cars. City age, then, may impact network robustness and congestion. While public transit, which legacy cities were designed for, will remove some of the traffic from the streets, it is not a major player in most cities, especially the smaller ones. Commuters tend to prefer automobiles for the commute (Cambridge Systematics 2005). With respect to city age, older cities would be expected to have more congestion. Calculation: census records, city records and other websites are checked to determine the census year when the city reached a population of 50,000; that year is compared to 2010 to calculate the

decades before 2010 measure. For example, Akron reached at least 50,000 in the 1910 census, 10 decades before 2010.

- Network intra-connectivity (Network links / Network nodes). The ability to move from one point another is in large part a factor of the number of routes available. The more connected a network is, the more routes one has to reach a particular destination and to avoid congested thoroughfares. Intra-connectivity can be measured by the ratio of nodes to links, with a higher ratio being associated with increased connectivity and lower congestion. While most VMT occurs on the upper system (freeways and arterials), the intra-connectivity of the entire system is used since additional routes on the collector and local street system provide drivers flexibility in when and where they access the upper level system. Calculation: The 2010 census urban area boundary layers are exceedingly complex and prove to be difficult to manipulate in a GIS, so this analysis is carried out using 2010 census tracts. All tracts with some portion falling in the urban area boundary are selected and mapped in TransCAD. The 2006 TransCAD street layer (the most recent available) is added and a clipping operation performed to find all the streets that fall into the collective tracts. Links are then divided by nodes.

4.4.2 Variables Impacting Demand. Table 8 summarizes the independent variables impacting demand, with each variable discussed in the ensuing paragraphs. The justification column includes references to theories and concepts discussed in Section 2.7 above.

Table 8: Variables impacting demand

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	**OPERATIONAL**			
Internal Productions	Commuters per square mile	Mod/Negative	UMR Census	Measure of commuter productions -Social exchange theory -4-Step Urban Travel Demand Model
	Persons per square mile	Mod/Negative	UMR Census	Measure of total productions -Social exchange theory -4-Step Urban Travel Demand Model
	Cars per household	Mod/Negative	Census ACS12-1	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
	Income per capita	Mod/Negative	Census ACS12-1	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
Internal Attractions	Employment per capita	Mod/Negative	Census ACS12-1	Measure of commuter attractions -Transportation demand is derived -Rational choice theory -Various sociological theories
	Persons per restaurant	Mod/Positive	US Census Economic Census 2007	Measure of other attractions -Transportation demand is derived -Rational choice theory -Various sociological theories
External Productions	In-commuting flows per worker (Jobs in UA tracts - Workers in UA tracts)	Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of external productions -Social exchange theory -Land rent theory -Transportation demand is derived
Trip Distribution	Average commuting time in minutes	Mod/Negative	Census ACS12-1	Measure of time on network -Rational choice theory -Land rent theory
Mode Split	Percent of commuters in single occupant vehicles (SOV)	Low/Negative	Census ACS12-1	Effects of decreasing highway demand -Rational choice theory -Various sociological theories
	Transit vehicle revenue miles per square mile	Low/Positive	NTD Census	Effects of decreasing highway demand -Rational choice theory -Various sociological theories
Variations in Demand between Urban Areas	Dummy variables based on city size (population)	Mod/Negative	UMR	Measure of variations in demand -Rational choice theory -Various sociological theories

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
	Dummy variables based on geography (US region)	Low/Unknown	Census	Measure of variations in demand -Rational choice theory -Various sociological theories

- Internal Productions (Commuters per square mile). The demand for transportation is derived and a key source of demand is from commuters, especially given that congestion is more common during the principal commuting periods in the mornings and evenings. As the number of commuters per square mile increases, one would expect congestion to worsen. Calculation: the 2010 urban area commuters from the UMR are divided by urban area square mileage from the 2010 census urban area list.
- Internal Productions (persons per square mile). Using reasoning similar to that above, the demand for transportation is also affected by the concentrations of people other than just commuters. These people derive benefits from transportation and so would generate trip productions. As population densities increase, one would expect congestion to increase as well. Calculation: the 2010 urban area population from the UMR is divided by urban area square mileage from the 2010 census urban area list.
- Internal Productions (Cars per household). Jane Jacobs famously noted that traffic congestion is “caused by vehicles, not by people in themselves” (Jacobs 1961). People do, however, choose to drive vehicles, especially when they are readily available. Several studies have found that increased car ownership is associated with increased travel (e.g. Aftabuzzaman 2011). It therefore seems likely that increased car ownership would also be linked to increased congestion.

Indeed, "autos per household" is a factor in trip generation in the four-step model (Martin and McGuckin 1998). Calculation: the urban area data from the ACS 2012-1, Table DP-04 includes the total number of households and the numbers of households with 0, 1, 2, 3, 4+ vehicles, respectively. The total vehicles is calculated (assuming that no households had more than four vehicles) and then divided by the total households.

- Internal Productions (Income per Capita). Income is also a factor in the trip generation calculations (Martin and McGuckin 1998), with higher incomes being linked to increased travel. Increased travel seems likely linked to increased congestion, so cities with higher incomes are likely to have worse congestion. Calculation: Income per capita is taken directly from the urban area data in ACS-2012-1, Table DP-03.
- Internal Attractions (Employment per capita). People travel to get somewhere and a key attraction is work, especially since congestion is largely a "rush hour" problem, especially in the smaller urban areas. Clearly, more jobs available for each man, woman and child generate more commuters, which in turn increases the potential for congestion. Calculation: Total employment data from the urban area data in ACS-2012-1, Table DP-03 is divided by population data from the same table.
- Internal Attractions (Persons per restaurants). People also travel to take advantage of the many amenities that lie within the urban area. Restaurants are used here as a proxy for these amenities. In the US, as in much of the developed world, there are three times of the day when meals are usually consumed. Two of

these meal times (breakfast and supper) coincide with the morning and evening "rush hours" and the third (lunch) occurs during the mid-day traffic spike. Restaurant customers frequenting these facilities during these time periods can add to the traffic burden. Since restaurants are generally opened in response to demand (and demand has risen with the rise of two-income households), and one might expect the greater the number of restaurants per capita, the greater the congestion. Conversely, the greater number of persons per restaurant would likely be associated with lower congestion. Calculation: The metropolitan area population from the ACS-2007-1, Table DP-05 is divided by the number of restaurants for metropolitan areas (urban area data is not available) from the 2007 Economic Census (the most recent year available). While this geography may not be the same as the rest of the study (metropolitan areas vs. urban areas), the ratio of people to restaurant seems likely to be less varied between the two areas.

- External Productions and Attractions (In-commuting flows per worker: Jobs in UA tracts - Workers in UA tracts). The hinterlands of an urban area serve two key purposes: markets for finished goods and services and sources for raw materials and workers. With today's globalized markets resulting in large part to decreased transportation costs, it is the source of workers that is arguably the hinterlands' most important purpose. These in-bound commuters can add significantly to the network burden, especially on the upper level system which they typically use most often. (In this regard, in-commuting flows might be a measure of network inter-connectivity, as well.) One might reasonably expect that the higher the share of commuters from outside the urban area, the greater the congestion. Calculation: The UA census tract workers (workers by place of residence) from the CTPP 5-Year ACS 2006-2010 are subtracted from the UA census tract jobs (workers by place of work) from the same source to determine the net flow into the UA from the hinterlands. (This assumes one worker per job and no out-commuting. If one worker has more than one job, in-commuting would decrease; if there are some out-commuters, in-commuting would increase. It seems likely that these assumptions would at least partially offset one another.) This net in-flow is then divided by the UA census tract workers.

- Trip Distribution (Average commuting time in minutes). The time it takes people to travel to work is a straight-forward measure of the extent they use the network. The longer they travel, the longer they are on the streets and freeways and the more they add to the burden on the network, which if heavy enough, becomes congestion. Calculation: Mean travel time to work is taken directly from urban area in the ACS-2012-1, Table DP-03.
- Mode Split (Percent of commuters in single occupant vehicles (SOV)). The vast majority of commuters drive to work alone. There have been extensive efforts to encourage the public to commute in other than SOVs (transit, carpools, vanpools, bicycles, etc.), but as yet, these efforts have not had much success. The number of commuters using SOVs has a direct impact on the number of cars on the roads and hence, a direct impact on congestion. It seems likely that a larger share of commuters using SOVs would be associated with higher levels of congestion. Calculation: Workers commuting to work who drove alone from urban area data

in the ACS-2012-1, Table DP-03 is divided by the total workers commuting to work from the same source.

- Mode Split (Transit vehicle revenue miles per square mile). Not all travelers use the street/highway network in their own vehicles; some use public transportation. While this percentage is quite small in most urban areas, it is likely to have some impact on congestion. The number of riders is affected by the availability (and quality) of transit service; the more that transit is able to satisfy the derived demand for transportation, the more likely it is to attract riders. One manner of assessing transit availability is using the number of miles transit vehicles travel to provide their service per square mile. It seems likely that more densely packed transit vehicle revenue miles would be associated with increased ridership and lower congestion. Calculation: 2010 vehicle revenue miles for all types of transit from the National Transit Database are divided by urban area square mileage from the 2010 census urban area list.
- Variations in Demand (Dummy variables based on city size and geography). There are likely to be variations in demand based on the cultural aspects of the urban area's populace. Interest groups and reference groups help shape demand and these associations vary. Two basic ways to assess any differences in travel demand are by city size (by population) and city location. People in larger cities with perhaps larger congestion problems might adapt and travel less, which might offset some congestion. Larger populations, however, are likely to have more aggregate travel, outstripping any such offsets, which could reasonably lead to worse congestion. People in different parts of the country might also travel

differently and there may be differences in the urban planning cultures and city amenity expectations in different regions. The nature of these impacts on congestion, while expected to be small, are unknown. Calculation: Four variables are created for each of the two categories (city size and geography). Recognizing that the use of all four variables of the same category in a single analysis leads to perfect collinearity and that some analytical methods cannot handle perfect collinearity, the number of dummy variables used will be reduced as needed. This will be discussed in the methods section below. The city size dummies are based on the four population categories of the 2010 UMR (small cities, medium cities, large cities and very large cities). The geographic dummies are based on city location within the four 2010 census regions (South, Northeast, Midwest, and West).

4.4.3 Variables Impacting Flow. Table 9 summarizes the independent variables impacting traffic flow, with each variable discussed in the ensuing paragraphs. The justification column includes references to theories and concepts discussed in Section 2.7 above.

Table 9: Variables impacting flow

VARIABLE		**EXPECTED**		
CONCEPTUAL	**OPERATIONAL**	**EFFECT**	**SOURCE**	**JUSTIFICATION**
Trucks	Percent of trucks on freeways	Mod/ Negative	FHWA	Measure of truck impact on flow -Differences in truck-car nimbleness
Distracted Driving	Percent of population 16-24 plus percent of population 65 and over	Low/ Negative	Census ACS12-1	Measure of flow interruptions due to distracted drivers -Consequences of human interaction -Loss aversion
Intersections with traffic signals and stop-controlled signage	Nodes per network mile (upper level system only)	Mod/ Negative	TransCAD Census	Measure of flow interruptions due to signals/signage -Queuing theory

VARIABLE		EXPECTED		
CONCEPTUAL	OPERATIONAL	EFFECT	SOURCE	JUSTIFICATION
Road Condition	Pavement condition (percent in poor condition)	Low/ Negative	TRIP Urban Roads Report	Measure of decreases in flow caused by lower speeds due to poor pavement -Human nature and driving skills
Traffic Incidents	Accident rate x VMT per capita	Low/ Negative	NHTSA FHWA UMR	Measure of flow interruptions due to traffic accidents -Consequences of human interaction -Loss aversion
Weather	Annual precipitation	Low/ Negative	NCDC	Measure of flow interruptions due to bad weather -Mother nature and geography
Special Events	Number of upper level sports teams	Low/ Negative	Various Internet Websites	Measure of flow interruptions due to special events -Structural functionalism -Social exchange theory

- Trucks (Percent of trucks on freeways). The flow of traffic can be adversely affected by vehicles that are less nimble than the norm. Trucks fit this description; moreover they often present obstacles to the line of sight of nearby drivers impacting their ability to react in traffic. Together these truck characteristics can reduce throughput and contribute to congestion. Since truck traffic largely uses the upper level system and in particular, the freeways, the focus of this measure is there. A larger percentage of trucks would likely have a larger negative impact on congestion. Calculation: The percent of each state's urban VMT comprised by trucks is published in the Federal Highway Administration's (FHWA) Highway Statistics series (2008 Table PS-1, the most recent available). These data are allocated to urban areas based on their primary state.
- Distracted Driving (Percent of drivers 16-25 plus percent of drivers over 65). The flow of traffic can also be affected by distracted, inattentive, or less responsive

drivers. In general, younger drivers and older drivers are more at risk for distracted driving and being slow to respond to highway conditions. If this is true, then the more of these drivers on the network, the more likely congestion will be present, as well. Ideally, driver data by age by urban area would be used for this variable. Unfortunately, these are not available and census age data are used as a surrogate. (This approach does assume that the percentages of drivers in each age group are uniform across the nation. This may not always be the case, however, especially in areas where alternative transportation modes are readily available.) Calculation: This measure is calculated from the urban area sex and age data in the ACS 2012-1, Table DP-05.

- Intersections with traffic signals and stop-controlled signage (Nodes per network mile for the upper level system only). The smooth flow of traffic is interrupted by design at many intersections in the network to allow increased access. These interruptions can significantly reduce throughput and decrease the available capacity. An increased number of intersections per mile is likely associated with increased congestion. Calculation: The 2010 census urban area boundary layers are exceedingly complex and prove to be difficult to manipulate in a GIS, so this analysis is done using 2010 census tracts. All tracts with some portion falling in the urban area boundary are selected and mapped in TransCAD. The 2006 TransCAD street layer (the most recent available) is added and a clipping operation performed to find all the streets that fall into the collective tracts. This street file includes TIGER line file data on the types of roads and streets. The primary, secondary, and connecting roads are segregated, node counts and link lengths are summed, and total nodes are then divided by total length.

- Road Condition (percent of pavement in poor condition). Traffic flows faster and more smoothly on good pavement than on poor pavement. Pavement in poor condition (i.e., is rough and bumpy) can cause drivers to reduce speeds and thereby reduce throughput. Calculation: data is taken from the 2013 TRIP[10] Urban Roads Report, which is based on a 2011 FHWA survey of state transportation officials on the condition of major state and locally maintained roads and highways (Interstates, freeways, and other arterial routes). Pavement condition is determined from a uniform pavement rating index.
- Traffic Incidents (Accident rate x VMT per capita). Traffic incidents are perhaps the key source of total travel delay, but they are non-recurrent (although often common). Nonetheless, delays associated with the traffic incidents are typically captured in the congestion calculations in the Urban Mobility Report.[11] Calculation: daily VMT miles from the Federal Highway Administration's (FHWA) Highway Statistics series (2010 Table HM-71) are multiplied by 365 to determine annual VMT, which is then multiplied by the national crash rate calculated from the 2010 Data Summary from the Fatality Analysis Reporting System (FARS) General Estimates System (published by the National Highway

[10] TRIP is a Washington, DC-based nonprofit organization that researches, evaluates and distributes economic and technical data on highway transportation issues.

[11] Current TTI methodology calls for real-time measures of traffic speeds, which will capture travel from any origin and to any destination as long as it happens during the measured periods. These speeds are an annual average of traffic speeds for each section of road for every 15 minutes of each day for a total of 672 day/time period cells (24 hours x 7 days x 4 periods per hour) (Schrank, Eisele, and Lomax 2012).

Traffic Safety Administration (NHTSA)). Finally, this crash total is divided by the 2010 UMR UA population.

- Weather (Annual precipitation). Weather issues are ubiquitous; all urban areas have them, but some areas may be more impacted than others. Annual precipitation is used to represent all weather issues since life goes on in the rain in a way that it does not during hurricanes, tornados, and snow and ice storms. Since rain tends to slow down traffic and reduce throughput, one would expect that higher annual precipitation totals would be associated with greater congestion. Calculation: Monthly precipitation data from the 2010 Annual Climatological Summaries (published by the National Climate Data Center and providing historical monthly temperature and precipitation data for reporting stations throughout the United States) is summed to get annual totals. One station is selected for each urbanized area; when there are multiple stations available in an area, primacy is given to those stations adjacent to major airports and those with complete data. In one case (Riverside-San Bernardino), complete data is not available for 2010. In this case, the average annual precipitation for Riverside-San Bernardino from The Weather Channel website is used instead.
- Special Events (Number of upper-level sports teams). Special events can be a key cause of non-recurrent congestion as they become a point of convergence on the traffic grid. There are numerous special events in a given urban area, too many to assess in a macro-analysis like this one. Moreover, detailed data on special events by urban area are not available. As a surrogate for all special events, the number of upper level (NCAA Division I, Minor League, and Major League) sports teams

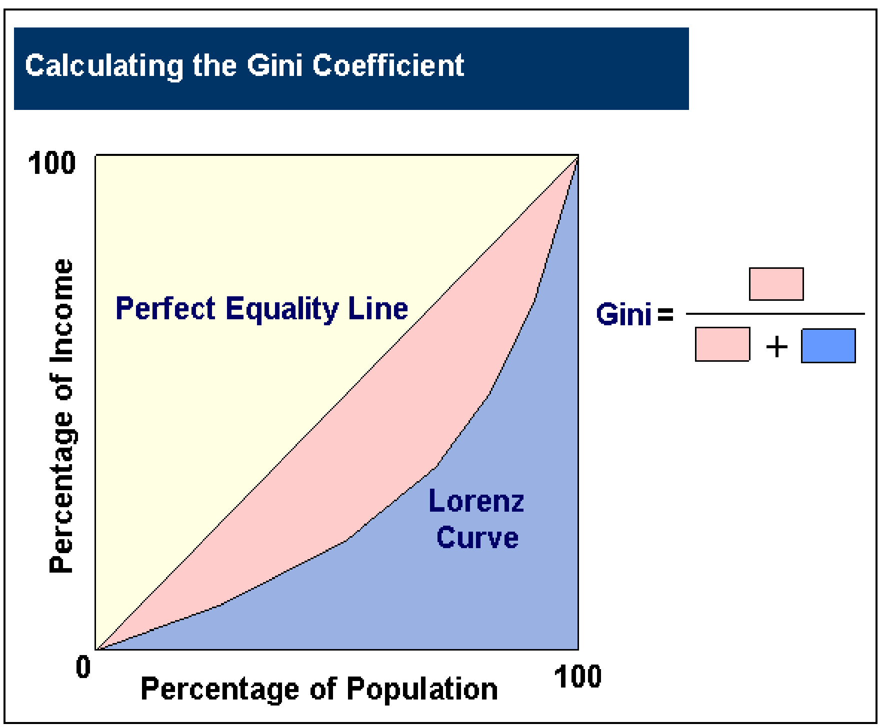
is used, with the acknowledgement that this surrogate is imperfect. Still, sporting events of such teams routinely occur, at least in part, during the evening commute where any traffic delays associated with the event are captured in the congestion calculations in the Urban Mobility Report.[12] Calculation: base data from the website 50States.com (listing the sports teams in each state by location) is crosschecked against city websites and lists of professional sports teams, minor league baseball teams, NASCAR racetracks, and NCAA Division I institutions in Wikipedia.

4.4.4 Variations across Urban Areas of Variables Potentially Impacting Congestion (Measures of Spread). The above variables are single variables representing an entire urbanized area. It is almost certain that there is some variation in these variables across the urban area, variations that may have some impact on congestion. This variation can be thought of as an unequal distribution of the measure in question and a method is needed to assess unequal distribution. There are several measures that could be used here, to include the variance, the standard deviation and the Gini coefficient. The first two are related (the standard deviation is the square root of the variance) and measure the variation around the mean. Small variances and standard deviations indicate that the data dispersion is close to the mean, while high variances indicate that the data are more dispersed from the mean. In the former case, the data points are closer to one another than in the latter case. Small and high variances, however, are relative and their meanings are not always straightforward.

---

[12] Current TTI methodology calls for real-time measures of traffic speeds, which will capture travel from any origin and to any destination as long as it happens during the measured periods. These speeds are an annual average of traffic speeds for each section of road for every 15 minutes of each day for a total of 672 day/time period cells (24 hours x 7 days x 4 periods per hour) (Schrank, Eisele, and Lomax 2012).

On the other hand, the Gini coefficient conveys meaning in the coefficient itself. Developed by and named for Italian sociologist Corrado Gini, the Gini coefficient is a dimensionless measure of statistical dispersion most commonly used to assess the equal distribution of income and wealth. Coefficients range from 0 to 1, with 0 indicating perfect equality (all measured units have an equal share) and 1 indicating perfect inequality (one measured unit has all). The Gini coefficient is the ratio of the area between the Lorenz curve (developed by economist Max Lorenz to assess wealth distribution) and the perfect equality line to the area under the perfect equality line (Figure 3.) Since Gini coefficients are between 0 and 1, comparisons can be readily made between variables.



Source: Spagnoli 2008
Figure 3: Calculating the Gini coefficient

Calculation: Online calculators are used to determine the Gini coefficients. To ensure correct calculations, two different calculators are used and the results are crosschecked. The first calculator offers extensive explanatory tables and graphs (Wessa 2014), while the second is much easier to use (Had2Know.com 2014). In general, for each of the 100 urban areas, eight variables are calculated for all census tracts inside or partially inside the urban area boundary, and then the Gini coefficient are calculated for each of the eight variables. As noted above, the Gini coefficient is dimensionless, so there are no units of measure, but it is based on census tract data and the numbers of tracts in urban areas vary widely from 4,454 in New York to 32 in Boulder.

Table 10 summarizes these Gini coefficient variables, with each variable discussed in the ensuing paragraphs. The justification column includes references to theories and concepts discussed in Section 2.7 above.

Table 10: Measures of spread across urban area census tracts

VARIABLE		**EXPECTED EFFECT**	**SOURCE**	**JUSTIFICATION**
CONCEPTUAL	**OPERATIONAL**			
Network Layout	Gini coefficient of population per network mile	Low/Negative	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of network layout efficiency -Structural functionalism -Changing urban needs over time
	Gini coefficient of workers per upper level network mile	Low/Negative	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of network layout efficiency -Structural functionalism -Changing urban needs over time
Internal Productions	Gini coefficient of car ownership (aggregate vehicles per HH)	Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
	Gini coefficient of median income per HH	Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
	Gini coefficient of workers per capita	Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of time on network -Level of mixed land use -Land rent theory

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
Urban Spatial Structure	Gini coefficient of employment (jobs) density	Mod/Negative	CTPP 5-Year ACS 2006-2010	Measure of degree of monocentricity -Central place theory -Land rent theory
	Gini coefficient of jobs/HH balance	Mod/Negative	CTPP 5-Year ACS 2006-2010	Measure of degree of monocentricity -Central place theory -Land rent theory
	Gini coefficient of jobs/worker balance	Mod/Negative	CTPP 5-Year ACS 2006-2010	Measure of degree of monocentricity -Central place theory -Land rent theory

- Network Layout (Gini coefficient of population per network mile). The network layout itself is important in the flow of people and goods and access to the network essential to fully participate in what the urban area has to offer. The equality of this access may have an impact on congestion, with a larger disparity of access (higher Gini) being associated with more congestion. Calculation: Population data from the Census CTPP 5-Year ACS 2006-2010 is divided by network mileage from the 2006 TransCAD Street Layer.
- Network Layout (Gini coefficient of workers per upper level network mile). The upper level network is important for the commute and its access may make commuting faster and more convenient. The upper level system (freeways, expressways and major arterials and connectors) carries the major portion of commuters, who presumably use it because it affords travel time advantages. Less access to this system may be associated with longer commutes and more congestion and less uniform access (higher Gini) may be likewise associated. Calculation: Employment by place of work data from the Census CTPP 5-Year

ACS 2006-2010 is divided by upper level network mileage in the 2006 TransCAD Street Layer (selected using Census TIGER codes).

- Internal Productions (Gini coefficient of car ownership (aggregate vehicles per HH)). As noted above, car ownership is a factor in the trip generation process – the more cars a household has, the more trips it makes. An unevenly distributed number of cars (higher Gini) across the network might make the network unbalanced (depending, of course, on the network layout) and lead to more congestion. Calculation: aggregate vehicles per household are divided by the number of households, both from the Census CTPP 5-Year ACS 2006-2010.
- Internal Productions (Gini coefficient of median income per HH). In a similar manner as car ownership, higher median incomes are associated with increased trip-making, and unevenly distributed income might lead to an unbalanced network and higher congestion levels. Calculation: median household income data is taken directly from the Census CTPP 5-Year ACS 2006-2010.
- Internal Productions (Gini coefficient of workers per capita). Like the two above variables, an unequal distribution of workers might lead to an unbalanced network and higher congestion. Calculation: workers by place of residence are divided by the population, both from the Census CTPP 5-Year ACS 2006-2010.
- Urban Spatial Structure (Degree of mixed land use (LU)) (Gini coefficient of employment density). Most urban theorists hold that city design is a key factor in the demand for travel in general and travel by car in particular. Well-designed cities with mixed LU allow people to travel less often and less far. The degree of mixed LU can be measured by the spread of employment across the city

landscape. While this measure is not perfect (it does not differentiate between types of use), it does allow ready comparisons across urban areas using easily accessible data. If these urban theorists are correct, lower levels of mixed use development (higher Ginis) would be associated with higher congestion. Calculation: Calculation: workers by place of work from the Census CTPP 5-Year ACS 2006-2010 are divided by the area from the TransCAD census tract layer.

- Urban Spatial Structure (Degree of mixed LU) (Gini coefficient of jobs-household balance). Another measure of mixed LU is the ratio of jobs within a census tract to households and how this measure is spread out though the urban area. Lower levels of mixed use development (higher Ginis) would be associated with higher congestion. Calculation: workers by place of work are divided by the number of households, both from the Census CTPP 5-Year ACS 2006-2010.
- Urban Spatial Structure (Degree of mixed LU) (Gini coefficient of jobs-worker balance). Similar to the jobs-household balance, the jobs-worker balance gets at the degree of mixed LU issue. An even balance across the city (lower Gini) would be good for congestion. Calculation: workers by place of work are divided by workers by place of residence, both from the Census CTPP 5-Year ACS 2006-2010.

4.4.5 Other Variables Potentially Impacting Congestion. Table 11 summarizes other independent variables of interest, with each variable discussed in the ensuing paragraphs. The justification column includes references to theories and concepts discussed in Section 2.7 above.

Table 11: Other variables potentially impacting congestion

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	**OPERATIONAL**			
Centrality	Percent of Employment in Job-Rich, Job-Dense Tracts	Mod/Negative	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of the spread of employment -Central place theory -Land rent theory
Sprawl	Percent of Population in Job-Poor Tracts	Mod/Negative	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of the spread of population -Central place theory -Land rent theory
Urban Spatial Structure	Degree of poly-centricity (higher more poly)	Mod/Positive	Lee and Gordon (2007)	Measure of degree of monocentricity -Central place theory -Land rent theory
Land Costs	Housing affordability	Mod/Negative	Int'l Housing Afford. Survey	Measure of the bid-rent function -Land rent theory
Government Employment	Percent of employees working for government	Low/Negative	Census ACS12-1	Measure of private-public employment split -Degree of peak hour participation
8-hour work day	Percent of employment in retail	Low/Positive	Census ACS12-1	Measure of employees not working a standard 8-hour day -Rational choice theory -Degree of peak hour participation
Creativity	Patents per 1000 workers	Low/Positive	Brookings Institute	Measure of participation in the status quo -Transportation demand is derived -Rational choice theory -Various sociological theories
Activity Density	Real GDP per VMT	Mod/Negative	BEA UMR FHWA	Measure of city density of activity -Structural functionalism -Social exchange theory
Size	Urban area square miles	Mod/Negative	Census	Measure of city size -Rational choice theory -Various sociological theories
Size	Urban area population	Mod/Negative	Census	Measure of city size -Rational choice theory -Various sociological theories

- Centrality (Percent of Employment in Job-Rich, Job-Dense Tracts). A principal characteristic of city formation is centrality. Indeed, the centralizing of people and economic interests are the essence of city development and the central business district (CBD) is commonly synonymous with the city as a whole, often providing the primary source of employment for the entire urban area. Unfortunately, the CBD is an uncertain geography and its nature can vary from city to city. Moreover, as cities grow over time, the CBD often ceases to be the primary source of employment; multiple nuclei arise and employment is centralized at multiple nodes throughout the urbanized region. Regardless of whether there is a single employment center or multiple centers, commuter traffic will tend to converge on these centers, generating congestion, especially during the morning and evening commutes. The degree of concentration of employment in these employment centers should vary with congestion, with a greater concentration being associated with greater congestion. If this idea is true, then defining these concentrations becomes the challenge. When identifying the CBD (or any other employment center), how far out does the boundary extend? Should non-contiguous areas be included? Most often, designations of employment centers include some measures of employment density, but there is no industry standard. Employment centers are identified here as census tracts in terms of job richness and job density. Calculation: workers by place of work (jobs) are compared to workers by place of residence, both from the Census CTPP 5-Year ACS 2006-2010, to identify job-rich tracts (tracts with at least twice as many jobs as workers). Workers by place of work are divided by the area to determine employment density by census tract, which is compared to average employment density for all tracts to identify job-dense tracts (tracts with at least five times the average jobs per square mile). The jobs in tracts that are both job-rich and job-dense are then summed and compared to the total jobs in all tracts. Note that job-richness and job-density are relative measures: job-richness is a function of the jobs-to-worker ratio and job-density is a function of the tract jobs per square mile relative to the urban area average jobs per square mile. Relative measures will, of course, vary from city to city, but so too will the notion of centrality – what is considered “central” in Laredo might be considered by a New Yorker to be just another neighborhood.

- Sprawl (Percent of Population in Job-Poor Tracts). While the definition of sprawl varies among the term’s users, most will agree that sprawl involves some form of low density development. It is thought by many that sprawl leads to increased congestion by increasing the driving needed to overcome the longer distances between productions and attractions (Newman and Kenworthy 1999). If this is true, then greater sprawl should be associated with greater levels of congestion. Like the employment centers above, sprawl is difficult to define, much less measure. It is likely that the effects of sprawl are at least partially included in some of the demand-focused and spread-focused variables above. The net in-commuting flows, in particular, provide a measure of sprawl, although they would likely include some people from outside the limits of sprawling development. For this study, sprawl is defined in terms of the percent of the population living in job-poor census tracts (those without sufficient jobs for the workers who live in those tracts). Calculation: workers by place of work (jobs) are compared to workers by place of residence, both from the Census CTPP 5-Year ACS 2006-2010, to identify job-poor tracts (tracts with at least twice as many workers as jobs). The population in tracts that are job-poor are then summed and compared to the total population in all tracts.

- Urban Spatial Structure (Degree of poly-centricity). Akin to the ideas of centrality and sprawl is the nature of the urban spatial structure; whether it is mono- or polycentric. This harkens to the discussion of the models of urban development (Section 2.6.3), which notes that the degree of centralization and spread of the economic drivers affect the resulting street network and the traffic flows over that network. There has been some research done on developing measures of dispersion and centralization. One notable effort is by Lee and Gordon (2007), who ranked the larger US metropolitan areas using measures of dispersion, decentralization and poly-centricity. Their poly-centricity metric is a ratio of the employment in all the primary employment centers (less the CBD) to the employment in the all primary employment centers (to include the CBD). The CBD and employment centers were identified using a geographically weighted regression (GWR) procedure that identified peaks in the employment density surfaces across census tracts, with peaks having to have 10,000 jobs or more to qualify as employment centers. Since congestion is a concentration issue, it is expected that the higher this metric, the more poly-centric the city is and the better the congestion problem will be. Calculation: This measure is taken directly from Lee and Gordon (2007), which is based on 2000 data. Although this data is ten years older than the 2010 base year data for this study, it should not be too troublesome; urban spatial structure should be slow to change, especially in the larger cities, and any changes that might occur in a ten year period would likely be towards increased poly-centricity and away from concentration.

- Land Costs (Housing affordability). A sprawling form of development tends to follow the availability of cheaper land for development. Cheaper land typically leads to cheaper, more affordable housing. This can also be accompanied by an increase in the consumption of land and housing, which can in turn lead to more sprawl. This is a key reason that propels urban areas to seek to control sprawl (and development along with it) by enacting various urban containment regulations. A side effect of such regulations is the increase in the cost of land and the decrease in the affordability of housing (Cox, 2013). Urban containment regulations are often associated with increased planning activities designed to steer urban development towards a particular vision rather than allowing the city to grow and develop on its own. Such increased planning activities are in turn associated with non-auto focused transportation solutions, suggesting that network supply may not be expanding as much as needed. It thus seems reasonable that decreased housing affordability is associated with increased traffic congestion. Calculation: The affordability measure used is the median multiple (the ratio of the median housing cost to the median household income) and is provided in the 7th Annual Demographia International Housing Affordability Survey (2010: 3rd Quarter) (Cox 2011).
- Government Employment (Percent of employees working for local government). Is there a difference in the congestion generation potential of the private and public sectors? Government workers are viewed by many as nine-to-five clock

punchers. If this assessment is correct, then one might expect that a larger percent of employees working for government would be associated with greater levels of congestion since the government workers would all commute during prime rush hours. It is recognized that this assessment might not be correct and that the average government worker might works odd hours. This might be especially true with emergency, law enforcement, and military personnel. It is tested as noted, however, and any contrary findings are addressed in the results section below. (It is also recognized that there are a number of state capitals, as well as the national capital, included in the studied urban areas and that these cities are likely to have a larger percentage of government workers. As the study is concerned with the background characteristics linked with congestion and is not comparing cities head-to-head, this should not be a problem. If the percentage of government workers is an important correlate, then the congestion in these capitals would reflect this.) Calculation: This measure is calculated from urban area class of worker data in the ACS 2012-1, Table DP-03.

- 8-hour work day (Percent of employment in retail). Much of the congestion problem centers around the morning and evening "rush hours" as commuters move back and forth to work. These peak travel periods exist largely because of the standard 8-hour workday. Not all workers have these hours, however, and many travel during "shoulder periods" (those adjacent to the peak) or non-peak periods. These workers are often in retail, so the percentage of employment in the retail sector may be linked to congestion, with a higher percentage of retail employees linked to lower levels of congestion. Calculation: This measure is calculated from the urban area employment by industry data in the ACS 2012-1, Table DP-03.

- Creativity (Patents per 1000 workers). There has been extensive research on the creative class, a socioeconomic class developed and popularized by economist and social scientist Richard Florida in a number of publications beginning in 2002 with *The Rise of the Creative Class*. Florida believes this group to be a primary propellant of technological advancement. Urban areas with a large and well-established creative class will edge ahead of cities with a less developed creative class. While the traits of the creative class are as many and varied as they are with people in general, a key characteristic seems to be the ability to think creatively, outside the box, to tackle problems and develop solutions. This involves a certain degree of challenging the status quo. The creative class might be expected to do things some differently, such as working a non-standard work week (odd hours or working outside the office), favoring less commonly used modes of transportation (transit, bike, or walking), and having a less economic focus on their productive working hours. With these notions in mind, one might expect that a more developed the creative class would be associated with lower congestion levels. Calculation: The creativity measure used is the patents per 1000 workers and is provided in a Brookings Institution study on patenting and innovation in metropolitan America (Rothwell et al. 2013).
- Activity Density (Real GDP per VMT). People travel to go someplace. More destinations provide more opportunities and larger urban areas tend to have disproportionately more destinations; i.e., urban areas enjoy benefits from

agglomeration. Increased levels of urban activity, as measured by money available to spend, are likely associated with increased congestion. Sivak (2013) examined the relationship between economic activity (GDP) and the amount of driving (VMT) for the 50 states and the District of Columbia using the metric GDP per VMT. He found that in 2011, GDP/VMT for states ranged from $30.04/mile in DC to $2.51/mile in Mississippi, with a US median value of $4.66/mile. This study uses this same metric at the urbanized area level. Calculation: 2010 real GDP per capita for each city's metropolitan statistical area from the Bureau of Economic Analysis is expanded to the urbanized area using the population data from the UMR to get total real GDP. Real GDP is then divided by annual VMT (daily VMT data from the FHWA Highway Statistics series (2010 Table HM-71) multiplied by 365).

- Size (Square miles of Urban Area Footprint). Congestion is a problem of concentration – too many cars using too little capacity at a point in time. It seems likely that a larger "driver shed" (the area from which the drivers come) would have a greater potential for cars to concentrate. If this is true, then a larger city would likely have worse congestion than a smaller city simply based on the city footprint. Calculation: The urban area square mileage is taken directly from the 2010 census.
- Size (Urban Area Population). In a like manner, it would seem that a more populous "driver shed" would also have a greater potential for cars to concentrate. If this is true, then a more populous city would likely have worse congestion than a smaller city simply based on the numbers of people. A dummy variable is

already being used to address the demand aspects of the population issue, with a focus on the size delineations used in the UMR. This variable addresses the population issue without regard to category. Calculation: The urban area population is taken directly from the 2010 UMR.

4.5 Variable Roll-up

Table 12 lists the 55 variables (3 dependent and 52 independent) along with selected descriptive statistics. Shaded variables have some missing observations.

Table 12: Study variables with selected descriptive statistics

Variable Description	**Variable Code**	**n**	**Minimum**	**Maximum**	**Mean**	**Std. deviation**
Dependent Variables						
Travel Time Index (TTI)	TTI	100	1.04	1.37	1.17	0.07
Portion of lane miles that are congested	PortLMCong	100	0.09	1.23	0.40	0.17
Length of Peak Periods	PkHrs	100	1.50	8.00	3.45	1.45
Independent Variables Impacting Supply						
Percent change in population 2000-2010	PctPopCh	100	-5.49	71.21	17.66	13.03
Political party control in 2000	Rep-Dem	97	0.00	1.00	0.69	0.46
Network miles per square mile	NetMi_SqMi	100	4.59	43.45	11.00	4.06
Freeway miles per square mile	FwyMi_SqMi	100	0.08	1.67	0.32	0.17
Freeway lane miles per network lane mile	FwyLM_NetLM	100	0.02	0.17	0.07	0.02
Freeway lane miles per 1000 commuters	FwyLM_KCmtr	100	0.25	2.99	1.36	0.50
Freeway and arterial miles per capita	FwyArtMi_Cap	100	287.89	1627.00	755.35	256.57
City Age (decade before 2010 when city reached 50k in population)	DecBeforeNow	99	0.00	21.00	9.83	4.36
Network links per Network node	Links_Node	100	1.11	1.45	1.27	0.06
Independent Variables Impacting Demand						
Commuters per square mile	Cmtr_SqMi	100	602.53	4531.34	1470.35	717.56
Persons per square mile	Pers_SqMi	100	1150.29	8681.54	2875.88	1443.76
Cars per Household	Veh_HH	100	1.17	2.09	1.68	0.12
Income per Capita	Inc_Cap	100	13391.00	46808.00	27377.82	5425.16
Employment per capita	Empl_Cap	100	0.35	0.57	0.46	0.04
Persons per restaurant	Pers_Rest	94	541.95	1094.78	710.99	83.40
In-commuting flows per worker	Inflows_Wkr	100	-0.21	0.54	0.05	0.09
Average commuting time in minutes	AvgCmtTime	100	17.60	35.20	23.90	3.64
Percent of commuters in single occupant vehicles (SOV)	PctSOV	100	48.96	87.23	77.85	6.46

Variable Description	Variable Code	n	Minimum	Maximum	Mean	Std. deviation
Transit vehicle revenue miles per square mile	VRM_SqMi	100	0.00	278438.26	41741.24	45728.54
Population dummy variable-Small	PopSm	100	0.00	1.00	0.21	0.41
Population dummy variable-Mid-size	PopMed	100	0.00	1.00	0.33	0.47
Population dummy variable-Large	PopLg	100	0.00	1.00	0.31	0.46
Population dummy variable-Very large	PopVLg	100	0.00	1.00	0.15	0.36
Geographic dummy variable - Northeast	GeoNE	100	0.00	1.00	0.15	0.36
Geographic dummy variable - South	GeoS	100	0.00	1.00	0.39	0.49
Geographic dummy variable - Midwest	GeoMW	100	0.00	1.00	0.17	0.38
Geographic dummy variable - West	GeoW	100	0.00	1.00	0.29	0.46
Independent Variables Impacting Flow						
Percent of trucks on freeways	PctTrks	100	3.45	17.45	7.94	2.49
Percent of population young (16-24) and old (65+)	PctOldYng	100	21.14	38.84	25.92	3.09
Nodes per network mile, upper level system only	Nodes_UpNetMi	100	4.53	10.33	7.24	1.16
Percent pavement in poor condition	PctPrPvmt	95	1.00	64.00	26.41	14.987
Crashes per 1000 persons	Crashes_Kcap	100	6.38	25.12	15.42	3.64
Annual precipitation in inches	YrPrecipIn	100	5.90	65.10	35.13	14.02
Number of professional sports teams or NCAA Division I colleges per million people	SpTms_Mcap	100	0.00	9.54	3.39	1.93
Variations across Urban Area Census Tracts						
Gini of population per network mile	GPop_NetMi	100	0.20	0.99	0.35	0.16
Gini of workers per upper network mile	GWkr_UpNetMi	100	0.55	0.99	0.82	0.11
Gini of car ownership per household	GVeh_HH	100	0.08	0.32	0.13	0.03
Gini of median income per household	GMedInc_HH	100	0.16	0.35	0.23	0.03
Gini of workers per capita	GWkr_Cap	100	0.06	0.16	0.10	0.02
Gini of employment density	GJobs_SqMi	100	0.43	0.78	0.64	0.07
Gini of jobs-household balance	GJobs_HH	100	0.44	0.89	0.66	0.10
Gini of jobs-worker balance	GJobs_Wkr	100	0.47	0.88	0.65	0.09
Other Variables Potentially Impacting Congestion						
Percent of Employment in Job-Rich, Job-Dense Tracts	PctJobsJRDTcts	100	17.09	54.35	35.26	8.39
Percent of Population in Job-Poor Tracts	PctPopJPTcts	100	29.14	62.19	46.07	6.33
Degree of poly-centricity	LeePoly	72	0.00	91.00	38.69	25.15
Housing affordability	Med_Mult	100	2.02	8.45	3.51	1.14
Percent of employment in government	PctGovtEmp	100	8.48	23.26	14.60	3.82
Percent of employment in retail	PctRetEmp	100	8.26	16.85	11.86	1.36
Patents per 1000 workers	Pat_KWkrs	98	0.02	10.29	0.94	1.33

Variable Description	Variable Code	n	Minimum	Maximum	Mean	Std. deviation
Real GDP per VMT	GDP_VMT	98	2.03	14.52	5.32	2.29
Size - Area	UASqMi	100	32.49	3450.20	569.23	582.37
Size - Population	UAPop-K	100	150.00	18852.00	1696.34	253.94

4.5.1 Missing Values. For each of the 55 variables, there are 100 potential observations, one for each urban area. While most variables have values for all cities, there are seven independent variables that do not. These missing observations are spread across 30 urban areas. Deleting these observations would either reduce the usable independent variables to 48 or the usable urban areas to 70. Since the former would eliminate some urban characteristics from consideration and the latter would significantly reduce the size of the study group, the decision is made to estimate the missing values so that all variables may be used. Details on the missing values follow:

- Political party control in 2000. There are three missing values for this variable. In each of these cases, the elections were non-partisan and while the mayor in 2000 is identified, extensive Internet searches on his/her party affiliation are inconclusive. Since the other values are either 0 (Republican) or 1 (Democrat), a value of 0.5 is assigned for the three unknowns.
- City Age (decade before 2010 when city reached 50k in population). There is one missing value for this variable and it is the result of cities with populations below 50k being combined into a larger urbanized area. The missing value is estimated by comparing growth patterns of the two cities involved and the growth in other urban areas in the vicinity.
- Persons per restaurant. There are six missing values for this variable, four for small urban areas and two for mid-sized urban areas. All large and very large

urban areas are accounted for. The number of restaurants includes both full service and limited service facilities. The cities are segregated by size and two regression equations (one for full service and one for limited service) are developed for the two sizes of urban areas with the missing values (small and mid-sized). Since the restaurant data is from the 2007 Economic Census, 2007 population data from the American Community Survey is used as the independent variable. The regression equations are then used to calculate the missing number of restaurants and the ACS population data to determine the persons per restaurant.

- Percent pavement in poor condition. There are five missing values for this variable, all for small cities. Regressions against population and network miles, together or separately, do not yield a model with good fit to the data, so the average value of percent poor pavement for small cities is used as the missing value for each urban area.
- Degree of poly-centricity. There are 28 missing values for this variable, 17 for small cities, eight for mid-sized cities and three for large cities. All very large cities are all accounted for. Regressions against population do not yield a model with good fit to the data, so the average poly-centricity measure is used for each urban area based on the size category.
- Patents per 1000 workers. There are two missing values for this variable, both for mid-sized cities. The average patents per 1000 workers for mid-sized urban areas is used for the two missing values.

- Real GDP per VMT. There are two missing values for this variable, both for mid-sized cities. The average GDP per capita for mid-sized urban areas is used for the two missing GDP per capita values. The population data from the UMR is then used to determine total real GDP and the annual VMT from FHWA Highway Statistics to determine the Real GDP per VMT.

4.5.2 Revised Variable Roll-up. Table 13 lists the 55 variables, with the selected descriptive statistics updated to reflect the 47 missing values. The variables with the missing data are again shaded for easy reference.

Table 13: Study variables with selected descriptive statistics with estimates for the missing values

Variable Description	**Variable Code**	**n**	**Minimum**	**Maximum**	**Mean**	**Std. deviation**
Dependent Variables						
Travel Time Index (TTI)	TTI	100	1.04	1.37	1.17	0.07
Portion of lane miles that are congested	PortLMCong	100	0.09	1.23	0.40	0.17
Length of Peak Periods	PkHrs	100	1.50	8.00	3.45	1.45
Independent Variables Impacting Supply						
Percent change in population 2000-2010	PctPopCh	100	-5.49	71.21	17.66	13.03
Political party control in 2000	Rep-Dem	100	0.00	1.00	0.68	0.46
Network miles per square mile	NetMi_SqMi	100	4.59	43.457	11.00	4.06
Freeway miles per square mile	FwyMi_SqMi	100	0.08	1.67	0.32	0.17
Freeway lane miles per network lane mile	FwyLM_NetLM	100	0.02	0.17	0.07	0.02
Freeway lane miles per 1000 commuters	FwyLM_KCmtr	100	0.25	2.99	1.36	0.50
Freeway and arterial miles per capita	FwyArtMi_Cap	100	287.89	1627.00	755.35	256.57
City Age (decade before 2010 when city reached 50k in population)	DecBeforeNow	100	0.00	21.00	9.83	4.34
Network links per Network node	Links_Node	100	1.11	1.45	1.27	0.06
Independent Variables Impacting Demand						
Commuters per square mile	Cmtr_SqMi	100	602.53	4531.34	1470.35	717.56
Persons per square mile	Pers_SqMi	100	1150.29	8681.54	2875.88	1443.76
Cars per Household	Veh_HH	100	1.17	2.09	1.68	0.12
Income per Capita	Inc_Cap	100	13391.00	46808.00	27377.82	5425.16
Employment per capita	Empl_Cap	100	0.35	0.57	0.46	0.04
Persons per restaurant	Pers_Rest	100	541.95	1094.78	711.04	82.20
In-commuting flows per worker	Inflows_Wkr	100	-0.21	0.54	0.05	0.09

Variable Description	Variable Code	n	Minimum	Maximum	Mean	Std. deviation
Average commuting time in minutes	AvgCmtTime	100	17.60	35.20	23.90	3.64
Percent of commuters in single occupant vehicles (SOV)	PctSOV	100	48.96	87.23	77.85	6.46
Transit vehicle revenue miles per square mile	VRM_SqMi	100	0.00	278438.26	41741.23	45728.54
Population dummy variable-Small	PopSm	100	0.00	1.00	0.21	0.41
Population dummy variable-Mid-size	PopMed	100	0.00	1.00	0.33	0.47
Population dummy variable-Large	PopLg	100	0.00	1.00	0.31	0.46
Population dummy variable-Very large	PopVLg	100	0.00	1.00	0.15	0.36
Geographic dummy variable - Northeast	GeoNE	100	0.00	1.00	0.15	0.36
Geographic dummy variable - South	GeoS	100	0.00	1.00	0.39	0.49
Geographic dummy variable - Midwest	GeoMW	100	0.00	1.00	0.17	0.38
Geographic dummy variable - West	GeoW	100	0.00	1.00	0.29	0.46
Independent Variables Impacting Flow						
Percent of trucks on freeways	PctTrks	100	3.45	17.45	7.94	2.49
Percent of population young (16-24) and old (65+)	PctOldYng	100	21.14	38.84	25.92	3.09
Nodes per network mile, upper level system only	Nodes_UpNetMi	100	4.53	10.33	7.24	1.16
Percent pavement in poor condition	PctPrPvmt	100	1.00	64.00	26.23	14.61
Crashes per 1000 persons	Crashes_Kcap	100	6.37	25.12	15.42	3.64
Annual precipitation in inches	YrPrecipIn	100	5.90	65.10	35.13	14.02
Number of professional sports teams or NCAA Division I colleges per million people	SpTms_Mcap	100	0.00	9.54	3.39	1.93
Variations across Urban Area Census Tracts						
Gini of population per network mile	GPop_NetMi	100	0.20	0.99	0.35	0.16
Gini of workers per upper network mile	GWkr_UpNetMi	100	0.55	0.99	0.82	0.116
Gini of car ownership per household	GVeh_HH	100	0.08	0.32	0.13	0.03
Gini of median income per household	GMedInc_HH	100	0.16	0.35	0.23	0.03
Gini of workers per capita	GWkr_Cap	100	0.06	0.16	0.10	0.02
Gini of employment density	GJobs_SqMi	100	0.43	0.78	0.64	0.07
Gini of jobs-household balance	GJobs_HH	100	0.44	0.89	0.66	0.10
Gini of jobs-worker balance	GJobs_Wkr	100	0.47	0.88	0.65	0.09
Other Variables Potentially Impacting Congestion						
Percent of Employment in Job-Rich, Job-Dense Tracts	PctJobsJRDTcts	100	17.09	54.35	35.26	8.39
Percent of Population in Job-Poor Tracts	PctPopJPTcts	100	29.14	62.19	46.07	6.33
Degree of poly-centricity	LeePoly	100	0.00	91.00	33.47	23.18
Housing affordability	Med_Mult	100	2.02	8.45	3.52	1.14

Variable Description	Variable Code	n	Minimum	Maximum	Mean	Std. deviation
Percent of employment in government	PctGovtEmp	100	8.48	23.26	14.60	3.82
Percent of employment in retail	PctRetEmp	100	8.26	16.85	11.86	1.36
Patents per 1000 workers	Pat_KWkrs	100	0.02	10.29	0.94	1.32
Real GDP per VMT	GDP_VMT	100	2.03	14.52	5.418	2.36
Size - Area	UASqMi	100	32.49	3450.20	569.23	582.37
Size - Population	UAPop-K	100	150.00	18852.00	1696.34	253.94

4.6 Methods

When selecting a method or methods of analysis, it is important to keep in mind the nature of the research, of which there are several types. Explanatory research seeks to uncover the causal relationships between variables to achieve a better understanding of the studied phenomena. Predictive research seeks to develop models that allow the prediction of the studied phenomena without necessarily understanding the causal relationships involved. Confirmatory research seeks to confirm proposed hypotheses, and so combines somewhat the explanatory and predictive approaches. All three of these research types begin with a sound understanding of the studied subject. If this sound understanding is not yet in hand, descriptive or exploratory research may be needed. The former seeks to describe the population or phenomenon being studied, while the latter seeks to develop the understanding of the population or phenomenon more fully through the exploration of variable relationships. Both descriptive and exploratory research can lead to hypothesis development, which can then be studied using explanatory, predictive, or confirmatory approaches.

This research is exploratory in nature. It does not seek to predict congestion, nor does it seek to uncover *causal* relationships between congestion and urban area characteristics; it has no formal hypotheses to test. Instead, it seeks to identify those urban area characteristics that are linked with congestion so that follow-on research into causes and effects might be pursued.

In determining those urban characteristics that are linked with congestion, it is prudent to state how this determination is to be made. While there may be many characteristics that are linked with congestion (indeed, it may be that all urban characteristics are so linked), the interest here is only in the ones that are most important and influential. Important and influential are vague terms and call for some, perhaps arbitrary, thresholds. These thresholds are made within the context of the particular method being used and generally consider the size of the effect and the importance within the model. Characteristics with tiny effects are not considered important, nor are those characteristics that are statistically insignificant or unimportant in model development. Also of limited importance are those variables that are not useful in differentiating between the observations (i.e., urban areas). If every city, those with low levels of congestion and those with high levels, has the same "degree" of a particular variable, than using that variable to distinguish between urban areas becomes problematic. Importance is discussed in each part of the results section.

One of the challenges in this analysis is the nature of the data – characteristics of urbanized areas are commonly interrelated in complex ways, with each having some effect on or being affected by one or more of the others. This precludes, or at least makes more difficult, the use of one of the most common analytical methods, ordinary least squares (OLS) multiple regression. There are other approaches that are less sensitive to multicollinearity and get around this issue. One such approach is a form of linear regression, partial least squares (PLS) regression, which has the added benefits of working well with a larger number of independent variables and being well-suited for exploratory research (Garson 2014). This method, does, however, assume a linear relationship between the variables, and that may not always be the case.

Another approach that handles both the highly correlated variables and the linearity issues is the method of decision tree induction, which involves the sequential subdivision of observations on the basis of the discriminating power of independent variables in accounting for their relationship with the dependent variable, which in this case, is congestion. This relationship can be derived from a number of measures depending on the particular method, to include correlation, covariance, least square deviation, and minimum likelihood ratio. The resulting "decision trees" are then interpreted within the constraints of the study. There are several widely used decision tree algorithms, to include Chi-square Automatic Interaction Detection (CHAID), exhaustive CHAID, Classification and Regression Tree (CART), and QUEST. QUEST deals with categorical data and is inappropriate for this study. It has been suggested that CART is more useful for prediction while CHAID is better for analysis (Shmueli 2007). Since this research is exploratory, then it seems that either CHAID or exhaustive CHAID are the better alternative.

Both methods (PLS and CHAID) can assist in identifying the key correlates of congestion, and while the general approaches to these two methodologies are linked to the methodology itself, the specific algorithms used in the calculations are often dependent upon the software package being used.

4.6.1 Analytical Software Platforms. There are a number of software packages that offer these types of data analysis approaches, to include packages from SAS, SPSS, XLSTAT, StatSoft, and others. The XLSTAT package is selected based on cost, ease of use and background support materials. This platform has the added benefit of being an MS Excel add-on, so data do not have to be exported/imported between software packages. This platform supports partial least squares regression (PLS), four decision tree methods (CHAID, exhaustive CHAID, CART, and QUEST) as well as a variety of other methods, to include Pearson correlations and ordinary least squares regression (OLS).

4.6.2 Partial Least Squares (PLS) Regression. Introduced by Wold in the 1980s after two decades of development (Sanchez 2013), PLS is an extension of linear modeling that combines features of principal components analysis (PCA) and multiple regression. It is a dimension reduction technique where a large number of independent variables are analyzed to create a reduced number of components and then an ordinary least squares (OLS) regression step is used to predict values of the dependent variable. Unlike PCA, which develops components based on just the relationships among the independent variables, PLS finds the set of components that explains as much of the covariance between the dependent and independent variables as possible (Maitra and Yan 2008). As these components are orthogonal and non-overlapping (Garson 2014), PLS is less restrictive than OLS and is well-suited for situations where there are a large number of independent variables that are likely correlated. (Like OLS, however, PLS cannot handle perfect collinearity between variables and so care must be taken with dummy variables to ensure that only k-1 variables are used to represent k conditions (Garson 2014).) PLS also has an advantage over OLS in its ability to model more than one dependent variable at a time. Including the three congestion variables in the same model allows covariances between them to be captured in the calculations and adds another dimension to the results.

Unlike OLS, PLS is often characterized as a distribution-free technique in that no particular distribution of the data is assumed, which has some advantages and disadvantages. On the minus side, the lack of a common distribution among the data can affect the calculations of the sizes of the effects and also precludes significance testing, which limits the ability to generalize the results (Garson 2014). On the plus side, the lack of a need for a common distribution gives tremendous flexibility with the data. As this study is exploratory in nature and aimed at identifying the most important correlates of congestion rather than predicting specific effects or explaining causal relationships, the inability to generalize the results is less of a concern. Also of less concern is the exactness of the sizes of the effects; general magnitudes should be adequate.

4.6.3 Chi-square Automatic Interaction Detection (CHAID). CHAID was developed by Kass (1980) and then extended by Biggs, Ville and Suen (1991) to include the exhaustive CHAID method. The former allows decision trees with splits of more than two branches, while the latter involves an additional repetitive sub-routine that always produces a binary tree. Both methods allow the use of nominal, ordinal and ratio data and involve three steps: splitting, merging and stopping. In the splitting step, the chi-square test for independence is used to assess whether splitting a node improves the purity by a significant amount, with a goal of maximizing the variance between nodes while minimizing the variance within nodes (Ratner 2007). Independent variables are analyzed and the one with the lowest criterion value is selected as the split variable (as long as the criterion value is lower than user-defined threshold). The criterion for quantitative dependent variables is the p-value and for qualitative dependent variables is either Tschuprow's T or the maximum likelihood ratio (as defined by the user). In the merging step, similar categories are merged into common sub-nodes by comparing Tschuprow's T or the maximum likelihood ratio to a user-defined threshold (if the maximum value is greater than the threshold, the two groups are merged). In exhaustive CHAID, merging continues until only two categories remain. Splitting and merging continue recursively until the stopping criteria are met. The stopping criteria are primarily used-defined (maximum tree depth, minimum size for a parent-node, and minimum size for a child-node), but also include reaching a pure node (a node containing only objects of one category or one value of the dependent variable), which cannot be further split. Variables are considered individually, so it is important that dummy variables for all conditions are included in the mix, i.e. k dummy variables vs. k-1 dummy variables. While it is possible to make the k-1 dummy variable approach work, it is inefficient and requires the development of multiple trees (Shmueli 2014).

When considering which of the CHAID options to use, it is important to remember that one reason for using decision trees is to avoid the linearity issues associated with forms of regression. With this in mind, it seems that an algorithm that permits more than two branches off a single node would better allow the uncovering of non-linear relationships between the variables. Regular CHAID allows this.

4.6.4 Multiple Methods. Upon considering the data, together with the various methods, there appears to be no single best method. Given that the research is exploratory in nature and that the goal is to identify the most important and influential urban characteristics that are correlated with congestion, multiple methods are used. Each method is used to identify the important characteristics within the limitations of its strengths and weaknesses. The important characteristics in each analysis are then compared with the results from the other analyses to determine the characteristics that are common across all methods. It is these commonly important variables that are likely to be the most influential. Four methods are used: Pearson correlations, PLS regression, and CHAID decision trees, which are used in two ways: the decision tree as a whole and the first split in the decision tree. All four methods are discussed in detail in Chapter 5.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Overview

First, the simple correlations between the variables are examined, since as noted in Chapter 4, urban characteristics are likely to be correlated, often highly so. Next, the relationships between the variables are investigated using PLS and the linearity assumption is explored. Then the relationships between the variables are investigated using CHAID, both in whole trees and in a first split analysis. Next, the results are summarized, compared across all methodologies, and discussed by congestion dimension and variable importance. Finally, results are related back to variable selection.

5.2 Pearson Correlations

The goal in variable selection is to uncover those variables that might be correlated with the measures of congestion. A review of the simple correlation matrix indicates that of the 52 independent variables, all but 13 have a statistically significant correlation at the 5% level with at least one of the congestion measures, with 19 having such a significant correlation with all three measures. The congestion duration measure (PkHrs) is the dependent variable with the broadest set of significant correlations, with 36 statistically significant correlations, to include a high correlation of 0.696 (UASqMi). The congestion extent measure (PortLMCong) has the poorest matches with potential predictors; only 22 correlations are statistically significant, with a high of -0.559 (FwyArtMi_Cap). The congestion intensity measure (TTI) falls between the two.

Notably, the lowest correlations for all three congestion variables are not statistically significant.

Table 14: Pearson correlations for the three dependent (congestion) variables

	Correlations			Ranks		
Variables	**TTI**	**PortLMCong**	**PkHrs**	**TTI**	**PortLMCong**	**PkHrs**
TTI	**1**	**0.477**	**0.828**	NA	NA	NA
PortLMCong	**0.477**	**1**	**0.476**	NA	NA	NA
PkHrs	**0.828**	**0.476**	**1**	NA	NA	NA
PctPopCh**	-0.153	0.139	-0.167	38	27	38
Rep-Dem	**0.271**	0.043	**0.216**	25	46	34
NetMi_SqMi**	-0.082	-0.136	-0.081	46	28	43
FwyMi_SqMi	-0.051	**-0.236**	-0.051	48	16	47
FwyLM_NetLM	**0.210**	-0.032	**0.220**	34	48	33
FwyLM_KCmtr	**-0.220**	**-0.420**	-0.185	32	3	37
FwyArtMi_Cap*	**-0.439**	**-0.559**	**-0.407**	10	1	13
DecBeforeNow	**0.397**	0.017	**0.487**	12	51	9
Links_Node**	0.087	0.156	0.118	45	26	41
Cmtr_SqMi*	**0.198**	**0.225**	**0.249**	35	19	31
Pers_SqMi*	**0.254**	**0.245**	**0.310**	28	14	26
Veh_HH**	-0.183	0.021	-0.154	36	50	40
Inc_Cap	**0.472**	0.120	**0.552**	8	29	7
Empl_Cap	**0.319**	-0.011	**0.360**	21	52	19
Pers_Rest	**-0.230**	-0.066	**-0.247**	31	41	32
Inflows_Wkr**	-0.006	-0.094	0.015	51	38	50
AvgCmtTime*	**0.626**	**0.476**	**0.650**	1	2	3
PctSOV*	**-0.480**	**-0.253**	**-0.374**	7	12	16
VRM_SqMi*	**0.527**	**0.329**	**0.537**	4	7	8
PopSm*	**-0.421**	**-0.241**	**-0.638**	11	15	5
PopMed*	**-0.280**	**-0.252**	**-0.279**	24	13	28
PopLg	**0.248**	0.173	**0.352**	29	24	20
PopVLg*	**0.527**	**0.384**	**0.640**	5	4	4
GeoNE**	0.140	-0.157	0.053	50	23	44
GeoS**	0.011	0.194	-0.062	41	25	46
GeoMW	-0.151	**-0.219**	0.020	39	20	49
GeoW**	0.003	0.096	0.010	52	36	52
PctTrks**	-0.133	-0.084	-0.094	42	39	42
PctOldYng*	**-0.248**	**-0.257**	**-0.338**	30	11	23
Nodes_UpNetMi*	**0.394**	**0.225**	**0.369**	13	18	17
PctPrPvmt	**0.353**	0.095	**0.343**	17	37	22
Crashes_Kcap**	-0.105	-0.038	-0.047	44	47	48
YrPrecipIn**	0.067	-0.052	0.014	47	44	51
SpTms_Mcap*	**-0.310**	**-0.308**	**-0.271**	22	8	29
GPop_NetMi	**0.214**	0.079	**0.212**	33	40	36
GWkr_UpNetMi*	**0.269**	**0.207**	**0.323**	27	21	24
GVeh_HH	**0.492**	0.100	**0.456**	6	34	11

	Correlations			Ranks		
Variables	**TTI**	**PortLMCong**	**PkHrs**	**TTI**	**PortLMCong**	**PkHrs**
GMedInc_HH	**0.271**	-0.063	**0.362**	26	42	18
GWkr_Cap**	0.118	0.108	0.167	43	32	39
GJobs_SqMi	**0.355**	-0.044	**0.215**	16	45	35
GJobs_HH	**0.303**	0.114	**0.313**	23	30	25
GJobs_Wkr*	**0.390**	**0.206**	**0.474**	14	22	10
PctJobsJRDTcts**	0.012	-0.110	-0.054	49	31	45
PctPopJPTcts*	**0.362**	**0.264**	**0.348**	15	10	21
LeePoly*	**0.337**	**0.276**	**0.553**	19	9	6
Med_Mult*	**0.448**	**0.227**	**0.384**	9	17	14
PctGovtEmp	-0.150	-0.058	**-0.309**	40	43	27
PctRetEmp	**-0.319**	-0.023	**-0.418**	20	49	12
Pat_KWkrs	0.168	0.105	**0.271**	37	33	30
GDP_VMT	**0.352**	0.097	**0.374**	18	35	15
UASqMi*	**0.623**	**0.366**	**0.696**	2	5	1
UAPop-K*	**0.618**	**0.365**	**0.685**	3	6	2
Variables with a Statistically Significant Correlation	**35**	**22**	**36**			
Highest Correlation	**0.626**	**-0.559**	**0.696**			
Lowest Correlation	0.003	-0.011	0.010			
Average Correlation	0.277	0.176	0.299			

Notes: Values in bold are different from 0 with a significance level alpha=0.05

* Variable is statistically correlated with all three dependent variables.

** Variable is not statistically correlated with any dependent variable.

Table 15: Top ten Pearson correlations by congestion variable

TTI	**Rank**	**PortLMCong**	**Rank**	**PkHrs**	**Rank**
AvgCmtTime*	**1**	FwyArtMi_Cap	**1**	UASqMi*	**1**
UASqMi*	**2**	AvgCmtTime*	**2**	UAPop-K*	**2**
UAPop-K*	**3**	FwyLM_KCmtr	**3**	AvgCmtTime*	**3**
VRM_SqMi*	**4**	PopVLg*	**4**	PopVLg*	**4**
PopVLg*	**5**	UASqMi*	**5**	PopSm	**5**
GVeh_HH	**6**	UAPop-K*	**6**	LeePoly	**6**
PctSOV	**7**	VRM_SqMi*	**7**	Inc_Cap	**7**
Inc_Cap	**8**	SpTms_Mcap	**8**	VRM_SqMi*	**8**
Med_Mult	**9**	LeePoly	**9**	DecBeforeNow	**9**
FwyArtMi_Cap	**10**	PctPopJPTcts	**10**	GJobs_Wkr	**10**

* Variable is in the top ten for all three dependent variables.

The top ten Pearson correlations are shown in Table 15. Five variables are in the top ten for each congestion dimension: AvgCmtTime, UASqMi, UAPop-K, VRM_SqMi, and PopVLg. Three of these are urban area size variables, which may not be surprising given the relationships uncovered in Table 4 above. Recall, however, that these reflect binary relationships and do not consider the interactive play these variables are likely to have.

While it is desirable for the independent variables to be highly correlated with the dependent variables, it is not so desirable for them to be correlated with one another. The nature of the variables themselves, however, as characteristics of urbanized areas, suggests that they are not likely to be independent and that there would be some overlap. A review of the variance inflation factors (VIFs) for each of the independent variables indicates that this is indeed the case. The VIF is derived from the R^2 from a multiple regression of each independent variable on all the other independent variables. (The tolerance is $1 - R^2$ and the VIF is the reciprocal of the tolerance.) VIF values above 2.50 are troublesome and indicate high multicollinearity (Allison 1999). While this is less of a problem in a global predictive model, high multicollinearity presents a real challenge in understanding the complex relationships between urban characteristics and identifying the more important predictor variables. Highly correlated independent variables can obscure the impacts of other variables in the same model. As Table 16 shows, this data set is highly correlated – only two variables have VIFs under the 2.50 threshold. This lack of independence between the "independent" variables indicates that models based on multinomial ordinary least squares (OLS) regression would be problematic. Other methods are needed. (It should be noted that two of the dummy variables (PopSm and GeoNE) were dropped from the variable set in the tolerance calculations and do not have a VIF value. These were dropped to avoid the perfect collinearity problem and allow the VIFs to be calculated. If dummy variables other than these two are dropped, the VIFs for all the dummy variables are changed, however, they still remain above the 2.50 threshold and the VIFs for the other independent variables remain unchanged to two decimal places.)

Table 16: Variance Inflation Factors (VIFs) for the independent variables

Variable	VIF	Variable	VIF	Variable	VIF
PctPopCh	3.55	VRM_SqMi	14.54	GWkr_UpNetMi	3.09
Rep-Dem	2.15	PopSm		GVeh_HH	17.93
NetMi_SqMi	28.62	PopMed	6.23	GMedInc_HH	5.22
FwyMi_SqMi	30.15	PopLg	10.83	GWkr_Cap	5.65
FwyLM_NetLM	33.27	PopVLg	16.70	GJobs_SqMi	7.65
FwyLM_KCmtr	23.41	GeoNE		GJobs_HH	3.62
FwyArtMi_Cap	6.86	GeoS	16.23	GJobs_Wkr	7.37
DecBeforeNow	11.22	GeoMW	6.14	PctJobsJRDTcts	3.99
Links_Node	7.28	GeoW	23.88	PctPopJPTcts	8.20
Cmtr_SqMi	167.72	PctTrks	2.57	LeePoly	3.52
Pers_SqMi	181.88	PctOldYng	5.95	Med_Mult	11.31
Veh_HH	5.82	Nodes_UpNetMi	4.59	PctGovtEmp	2.95
Inc_Cap	12.04	PctPrPvmt	5.00	PctRetEmp	5.36
Empl_Cap	9.24	Crashes_Kcap	16.16	Pat_KWkrs	2.67
Pers_Rest	2.77	YrPrecipIn	4.87	GDP_VMT	14.37
Inflows_Wkr	6.55	SpTms_Mcap	2.69	UASqMi	25.98
AvgCmtTime	9.19	GPop_NetMi	2.41	UAPop-K	22.45
PctSOV	14.37				

Note: shaded variables have VIFs below 2.50

5.3 Partial Least Squares (PLS) Regression Results

As noted in Chapter 4, PLS handles highly correlated independent variables by combining them into orthogonal components for analysis (Garson 2014). Moreover, since it estimates relationships between matrices, it also allows the inclusion of the three dependent variables in the same model so that their interactions can be included in the analysis. With this in mind, four PLS regression models are developed: one combined model with the three dependent and the 50 independent variables (two dummy variables are excluded to prevent perfect collinearity), and three additional models with just one of the three dependent variables and the 50 independent variables. These last models are used to assess the effects of the independent variables on each of the dependent variables separately. In model development, a 95% confidence interval is used and the software is keyed to determine automatically the optimal number of components. (XLSTAT uses an iterative algorithm to develop orthogonal components that maximize the explained variance of and the relationship between the dependent and independent variables. Components are developed such that the collective explanatory/predictive power of the components increases until the global quality begins to decline, at which point the number of components is determined to be optimal. See below for a description of the various indices used in this process.) The optimal number of components according to this criterion differs among the models: 3, 1, 1 and 2 for the combined, TTI, PortLMCong and PkHrs models, respectively. Finally, the various results, to include model quality, goodness of fit, variable importance in the projection, standardized coefficients and residuals are analyzed and discussed.

5.3.1 Model Quality and Goodness-of-Fit. Table 17 shows the number of components and the model quality for each model. The Q^2 cumulative index is a measure of the model's global quality, that is, the contribution of all components to its predictive quality. Values range between -1 and 1 (although only very bad models have negative values) with higher values indicating better model quality. The R^2Y cumulative and R^2X cumulative indices are measures of the explanatory/predictive power of the model for the dependent and independent variables, respectively, with values ranging between 0 and 1. The cumulative indices measure the cumulative effect for all the components together. When there is only one component, of course, the cumulative measure is the same as the measure for just the one component. It should be noted that the R^2Y value is the same as the R^2 and can be used to measure goodness-of-fit.

Table 17: Model quality results for the four separate models

Item	Combined	TTI	PortLMCong	PkHrs
Number of Components	3	1	1	2
Q^2 Cumulative Index	0.504	0.515	0.242	0.726
R^2Y Cumulative Index	0.649	0.591	0.370	0.812
R^2X Cumulative Index	0.370	0.186	0.159	0.265

From these measures, the PkHrs model appears to be the model with the highest global quality, followed somewhat closely by the TTI model. The PortLMCong model is substantially weaker than these, while the combined model's Q^2 cumulative index falls just below the TTI model. The combined model, however, has very good R^2Y numbers (better than all but the PkHrs model) and better R^2X scores than any other model. This indicates that the combined model does a better job of representing both the independent and dependent variables in the model without sacrificing too much of the global quality. When considering only the goodness-of-fit for each of the congestion variables (Table 18), the combined model tops each of the separate models except the PkHrs model, where the R^2 values are very comparable (0.802 vs. 0.812). All in all, the combined model appears to be the best of the four and only this model will be used for the rest of this section.

Table 18: Goodness-of-fit results for the three separate models

	Combined			Separate		
Item	TTI	PortLMCong	PkHrs	TTI	PortLMCong	PkHrs
Observations	100.000	100.000	100.000	100.000	100.000	100.000
Sum of weights	100.000	100.000	100.000	100.000	100.000	100.000
DF	96.000	96.000	96.000	98.000	98.000	97.000
R^2	0.665	0.479	0.802	0.591	0.370	0.812
Std. deviation	0.039	0.122	0.655	0.043	0.133	0.636
MSE	0.001	0.014	0.412	0.002	0.017	0.392
RMSE	0.039	0.120	0.642	0.043	0.132	0.626

5.3.2 Variable Importance to the Projection (VIP). One of the outputs of PLS regression is the VIP, which is a measure of the importance of an independent variable in component development. VIPs are calculated for each component in the analysis, with higher VIP values indicating that the variable is more influential. Highly influential variables have values above 1.000, moderately influential variables have values between 0.800 and 1.000, and variables with low or no influence have values below 0.800 (XLSTAT 2014). VIPs for each component in the combined model are shown in Table 19, rank ordered by the average for all three components, weighted by the component's Q^2 quality index. (Weighting the average in this way gives increased influence to the VIPs in the better quality, more meaningful components.) VIPs below 0.800, the threshold noted above for moderate to high influence, are shaded to facilitate understanding. Note that the VIP scores for all but four variables are either all above the 0.800 threshold (27 variables) or all below the threshold (19 variables) on each component. Interestingly, the four variables with the mixed values all have low VIPs for Component 1 and moderate VIPs for Components 2 and 3. A case could be made for dropping the bottom 19 (or even the bottom 23) variables if a predictive model were the goal. As the goal here is the exploring the relationships between urban characteristics and congestion, however, retaining these variables is important for further analysis.

Table 19: Variable Importance in the Projection (VIP)

		Component			Weighted
No.	**Variable**	**1**	**2**	**3**	**Average**
17	AvgCmtTime	2.007	1.895	1.840	1.984
49	UASqMi	1.986	1.801	1.755	1.949
50	UAPop-K	1.965	1.764	1.713	1.924
22	PopVLg	1.803	1.624	1.580	1.766
19	VRM_SqMi	1.624	1.474	1.453	1.593
7	FwyArtMi_Cap	1.504	1.628	1.592	1.528
13	Inc_Cap	1.426	1.310	1.330	1.403

No.	Variable	Component			Weighted
		1	2	3	Average
43	LeePoly	1.370	1.231	1.196	1.342
35	GVeh_HH	1.305	1.329	1.294	1.309
18	PctSOV	1.290	1.183	1.185	1.269
40	GJobs_Wkr	1.278	1.163	1.135	1.254
44	Med_Mult	1.244	1.128	1.099	1.221
8	DecBeforeNow	1.169	1.164	1.188	1.169
28	Nodes_UpNetMi	1.156	1.048	1.034	1.134
42	PctPopJPTcts	1.114	1.044	1.024	1.099
48	GDP_VMT	1.018	1.068	1.061	1.028
46	PctRetEmp	0.982	1.034	1.022	0.992
29	PctPrPvmt	0.974	1.007	0.984	0.980
32	SpTms_Mcap	0.977	0.965	0.961	0.975
27	PctOldYng	0.954	0.982	0.968	0.959
21	PopLg	0.912	1.027	1.157	0.937
6	FwyLM_KCmtr	0.830	1.230	1.206	0.909
34	GWkr_UpNetMi	0.923	0.849	0.831	0.908
11	Pers_SqMi	0.915	0.845	0.897	0.902
20	PopMed	0.910	0.853	0.840	0.898
39	GJobs_HH	0.886	0.880	0.868	0.885
14	Empl_Cap	0.878	0.862	0.984	0.877
36	GMedInc_HH	0.783	0.819	0.872	0.791
10	Cmtr_SqMi	0.748	0.681	0.731	0.735
38	GJobs_SqMi	0.704	0.793	0.782	0.722
15	Pers_Rest	0.669	0.720	0.701	0.679
2	Rep-Dem	0.663	0.597	0.703	0.652
47	Pat_KWkrs	0.653	0.593	0.580	0.641
45	PctGovtEmp	0.648	0.599	0.655	0.640
33	GPop_NetMi	0.612	0.718	0.753	0.634
5	FwyLM_NetLM	0.538	0.743	0.731	0.579
12	Veh_HH	0.423	0.599	0.583	0.457
37	GWkr_Cap	0.452	0.425	0.423	0.447
1	PctPopCh	0.317	0.924	0.897	0.438
4	FwyMi_SqMi	0.304	0.861	0.846	0.415
24	GeoMW	0.321	0.601	0.605	0.377
9	Links_Node	0.381	0.342	0.451	0.375
3	NetMi_SqMi	0.310	0.620	0.606	0.372
26	PctTrks	0.355	0.318	0.319	0.347
30	Crashes_Kcap	0.222	0.422	0.598	0.265
23	GeoS	0.070	0.889	0.872	0.233
41	PctJobsJRDTcts	0.137	0.183	0.244	0.148
16	Inflows_Wkr	0.056	0.357	0.392	0.117
25	GeoW	0.086	0.109	0.435	0.097
31	YrPrecipIn	0.065	0.063	0.274	0.069

5.3.3 Standardized Coefficients. Table 19 provides an insight into the relative importance of the variables, but it is not complete. While variable importance deals with the development of the separate components, it does not address the questions of the size and direction of the effects on the dependent variable. These two questions can be answered by analyzing the standardized coefficients. Standardized coefficients are a measure of the number of standard deviations the dependent variable will change with a one standard deviation change in the independent variable. Regression coefficients are commonly in different units of measurement, which makes comparisons between the independent variables problematic. Coefficients are standardized so that the independent variables can be compared head-to-head to determine which has the greater effect. It should be noted here that the VIP still retains importance. Variables with low VIPs (below 0.8) are problematic, regardless of the size of the standardized coefficient, and "should not be taken in account in the analysis" (Jakobowicz 2014).

Table 20 shows the standardized coefficients for each of the congestion variables in the combined model and the average VIP (weighted by the component's Q^2 quality index). These effects, "revealed" in the PLS regression analysis, are categorized in terms of size and direction based on the standardized coefficients. The size of the revealed effects range from low to high: a high effect is assessed if the standardized coefficient is above 0.100; a low effect is assessed if the standardized coefficient is below 0.050; and a moderate effect assessed otherwise. (These cut-offs are determined after analyzing the standardized coefficient dataset together with the VIPs, and reflect a relative value more than an absolute one.) The direction of the effect is either positive or negative, depending on the sign of the standardized coefficient, with a positive sign indicating a negative effect on congestion (recall that increased congestion is a negative). Since there are three dependent variables and thus three coefficients for each independent variable, there is a chance that the size and effect for all three will not be the same. Indeed, for over half of the variables (28 of 50), this is the case. This clearly suggests that the dimensions of congestion are linked to the characteristics of the urban areas in different ways. The speed of population growth (PctPopCh), for example, seems to have a more negative association with the portion of the network that is congested (PortLMCong) than the TTI or the duration of the peak travel period. As another example, the portion of the network that comprises freeways (FwyLM_NetLM) has a positive relationship with the extent of the congestion problem (PortLMCong), but a negative one with TTI and peak hour duration. In these cases with discrepancies between the coefficients (17 for size and 13 for direction, including two cases for both size and direction), the standardized coefficients are compared and an overall category assessed. These “summarized” effects are indicated in bold. Note that there are variables with high VIPs that have small standardized coefficients (e.g., GVeh_HH with a 1.309 VIP and standardized coefficients of 0.24, -0.17, and 0.24). This suggests that a variable that is important in model development may not always have a large effect on model outcomes (and it is the size and direction of effect that are important in this analysis). Conversely, there are also variables with low VIPs (below 0.800) that have relatively large standardized coefficients. As noted above, these low VIP variables should not be considered in the analysis and are shaded in the table below.

A comment on the dummy variables is warranted here. Two of the dummies (PopSm and GeoNE) serve as the base cases for their groups (Pop and Geo), and hence do not have standardized coefficients. The other six dummies do have standardized coefficients which must be interpreted in relation to the base cases. This leads to some interesting results. When compared to small cities and with all other things being equal, congestion levels are lower in midsize cities, much higher in large cities and higher in very large cities. Moreover, these relationships hold for all dimensions of congestion. When compared to cities of the Northeast and with all other things being equal, congestion levels are higher in cities in the South and lower in cities in the Mid-west and West. Unlike the city size variables, the relationship differs slightly across dimensions. While the direction of effect is the same, the relative improvements in congestion levels in midwestern and western cities vary. The interpretation of dummy variables can be tricky at times, but worth the effort for the nuances they can uncover.

Table 20: Standardized coefficients and revealed effects

		VIP	Standardized Coefficients			Revealed Effects	
No.	**Variable**	**(Wtd Avg)**	**TTI**	**LMCong**	**PkHrs**	**Size**	**Direction**
1	PctPopCh	0.438	0.024	0.070	0.026	**Low**	Negative
2	Rep-Dem	0.652	0.053	0.027	0.066	**Low**	Negative
3	NetMi_SqMi	0.372	-0.041	-0.060	-0.046	**Mod**	Positive
4	FwyMi_SqMi	0.415	-0.040	-0.083	-0.041	**Mod**	Positive
5	FwyLM_NetLM	0.579	0.006	-0.034	0.009	Low	**Positive**
6	FwyLM_KCmtr	0.909	-0.067	-0.114	-0.070	**Mod**	Positive
7	FwyArtMi_Cap	1.528	-0.092	-0.129	-0.097	**High**	Positive
8	DecBeforeNow	1.169	0.050	-0.008	0.061	**Low**	**Negative**
9	Links_Node	0.375	-0.001	0.011	-0.007	Low	**Negative**
10	Cmtr_SqMi	0.735	0.009	0.011	0.004	Low	Negative
11	Pers_SqMi	0.902	0.010	0.009	0.004	Low	Negative
12	Veh_HH	0.457	0.004	0.029	0.006	Low	Negative
13	Inc_Cap	1.403	0.072	0.020	0.086	**Mod**	Negative
14	Empl_Cap	0.877	0.053	-0.002	0.067	**Low**	**Negative**
15	Pers_Rest	0.679	-0.015	0.015	-0.018	Low	**Positive**
16	Inflows_Wkr	0.117	-0.007	-0.033	-0.004	Low	Positive
17	AvgCmtTime	1.984	0.113	0.116	0.125	High	Negative
18	PctSOV	1.269	-0.028	-0.017	-0.025	Low	Positive
19	VRM_SqMi	1.593	0.046	0.029	0.045	Low	Negative
20	PopSm						
21	PopMed	0.898	-0.041	-0.050	-0.042	Low	Positive

No.	Variable	VIP (Wtd Avg)	Standardized Coefficients			Revealed Effects	
			TTI	LMCong	PkHrs	Size	Direction
22	PopLg	0.937	0.099	0.087	0.119	**High**	Negative
23	PopVLg	1.766	0.078	0.069	0.084	Mod	Negative
24	GeoNE						
25	GeoS	0.233	0.048	0.084	0.055	**Mod**	Negative
26	GeoMW	0.377	-0.027	-0.057	-0.026	**Low**	Positive
27	GeoW	0.097	-0.024	-0.006	-0.034	Low	Positive
28	PctTrks	0.347	-0.011	-0.011	-0.010	Low	Positive
29	PctOldYng	0.959	-0.071	-0.074	-0.080	Mod	Positive
30	Nodes_UpNetMi	1.134	0.067	0.051	0.076	Mod	Negative
31	PctPrPvmt	0.980	0.014	-0.016	0.013	Low	**Negative**
32	Crashes_Kcap	0.265	0.031	0.029	0.042	Low	Negative
33	YrPrecipIn	0.069	0.019	0.005	0.026	Low	Negative
34	SpTms_Mcap	0.975	-0.047	-0.066	-0.047	**Mod**	Positive
35	GPop_NetMi	0.634	-0.010	-0.025	-0.016	Low	Positive
36	GWkr_UpNetMi	0.908	0.053	0.046	0.060	**Mod**	Negative
37	GVeh_HH	1.309	0.024	-0.017	0.024	Low	**Negative**
38	GMedInc_HH	0.791	0.036	-0.013	0.047	Low	**Negative**
39	GWkr_Cap	0.447	0.020	0.025	0.020	Low	Negative
40	GJobs_SqMi	0.722	0.018	-0.021	0.022	Low	**Negative**
41	GJobs_HH	0.885	0.013	-0.008	0.011	Low	**Negative**
42	GJobs_Wkr	1.254	0.040	0.021	0.041	Low	Negative
43	PctJobsJRDTcts	0.148	-0.002	-0.016	0.001	Low	**Positive**
44	PctPopJPTcts	1.099	0.052	0.062	0.054	Mod	Negative
45	LeePoly	1.342	0.059	0.048	0.063	**Mod**	Negative
46	Med_Mult	1.221	0.041	0.023	0.043	Low	Negative
47	PctGovtEmp	0.640	-0.051	-0.034	-0.062	**Mod**	Positive
48	PctRetEmp	0.992	-0.030	0.018	-0.037	Low	**Positive**
49	Pat_KWkrs	0.641	0.028	0.012	0.032	Low	Negative
50	GDP_VMT	1.028	0.008	-0.020	0.004	Low	**Positive**
51	UASqMi	1.949	0.105	0.086	0.117	**High**	Negative
52	UAPop-K	1.924	0.086	0.067	0.093	Mod	Negative

Bold text indicates summarized effects.
Shaded cells indicate low VIPs.

5.3.4 Revealed Effects versus Expected Effects. Table 21 shows these revealed effects are compared to the expected effects noted in the variable tables in Chapter 4. The discrepancies between the expected and revealed effects are indicated in bold and variables with low VIPs (below 0.800) are shaded. As noted above, these variables are problematic and any discrepancies in effects are suspect. Note that the two dummy variables that serve as the base cases (PopSm and GeoNE) do not have individual revealed effects, which would be determined in relationship to the other dummies in the group and within the model as a whole. (Recall again that increased congestion is a negative effect.)

Table 21: Expected effects compared to revealed effects

			Expected Effects		Revealed Effects	
No.	**Variable**	**Focus**	**Size**	**Direction**	**Size**	**Direction**
1	PctPopCh	Supply	High	Negative	**Low**	Negative
2	Rep-Dem	Supply	Low	Negative	Low	Negative
3	NetMi_SqMi	Supply	Mod	Positive	Mod	Positive
4	FwyMi_SqMi	Supply	Mod	Positive	Mod	Positive
5	FwyLM_NetLM	Supply	Mod	Positive	**Low**	Positive
6	FwyLM_KCmtr	Supply	High	Positive	**Mod**	Positive
7	FwyArtMi_Cap	Supply	Mod	Positive	**High**	Positive
8	DecBeforeNow	Supply	Low	Negative	Low	Negative
9	Links_Node	Supply	Mod	Positive	**Low**	**Negative**
10	Cmtr_SqMi	Demand	Mod	Negative	**Low**	Negative
11	Pers_SqMi	Demand	Mod	Negative	**Low**	Negative
12	Veh_HH	Demand	Mod	Negative	**Low**	Negative
13	Inc_Cap	Demand	Mod	Negative	Mod	Negative
14	Empl_Cap	Demand	Mod	Negative	**Low**	Negative
15	Pers_Rest	Demand	Mod	Positive	**Low**	Positive
16	Inflows_Wkr	Demand	Low	Negative	Low	**Positive**
17	AvgCmtTime	Demand	Mod	Negative	**High**	Negative
18	PctSOV	Demand	Low	Negative	Low	**Positive**
19	VRM_SqMi	Demand	Low	Positive	Low	**Negative**
20	PopSm	Demand	Mod	Negative		
21	PopMed	Demand	Mod	Negative	**Low**	**Positive**
22	PopLg	Demand	Mod	Negative	**High**	Negative
23	PopVLg	Demand	Mod	Negative	Mod	Negative
24	GeoNE	Demand	Low	Unknown		
25	GeoS	Demand	Low	Unknown	**Mod**	**Negative**
26	GeoMW	Demand	Low	Unknown	Low	**Positive**
27	GeoW	Demand	Low	Unknown	Low	**Positive**
28	PctTrks	Flow	Mod	Negative	**Low**	**Positive**
29	PctOldYng	Flow	Low	Negative	**Mod**	**Positive**
30	Nodes_UpNetMi	Flow	Mod	Negative	Mod	Negative
31	PctPrPvmt	Flow	Low	Negative	Low	Negative
32	Crashes_Kcap	Flow	Low	Negative	Low	Negative
33	YrPrecipIn	Flow	Low	Negative	Low	Negative
34	SpTms_Mcap	Flow	Low	Negative	**Mod**	**Positive**
35	GPop_NetMi	Spread	Low	Negative	Low	**Positive**
36	GWkr_UpNetMi	Spread	Low	Negative	**Mod**	Negative
37	GVeh_HH	Spread	Low	Negative	Low	Negative
38	GMedInc_HH	Spread	Low	Negative	Low	Negative

No.	Variable	Focus	Expected Effects Size	Expected Effects Direction	Revealed Effects Size	Revealed Effects Direction
39	GWkr_Cap	Spread	Low	Negative	Low	Negative
40	GJobs_SqMi	Spread	Mod	Negative	**Low**	Negative
41	GJobs_HH	Spread	Mod	Negative	**Low**	Negative
42	GJobs_Wkr	Spread	Mod	Negative	**Low**	Negative
43	PctJobsJRDTcts	Other	Mod	Negative	**Low**	**Positive**
44	PctPopJPTcts	Other	Mod	Negative	Mod	Negative
45	LeePoly	Other	Mod	Positive	Mod	**Negative**
46	Med_Mult	Other	Mod	Negative	**Low**	Negative
47	PctGovtEmp	Other	Low	Negative	**Mod**	**Positive**
48	PctRetEmp	Other	Low	Positive	Low	Positive
49	Pat_KWkrs	Other	Low	Positive	Low	**Negative**
50	GDP_VMT	Other	Mod	Negative	**Low**	**Positive**
51	UASqMi	Other	Mod	Negative	**High**	Negative
52	UAPop-K	Other	Mod	Negative	Mod	Negative

Bold text indicates discrepancies between expected and revealed effects.
Shaded cells indicate low VIPs.

While the sizes of the effects are often estimated incorrectly (for 26 of the 50 variables), the errors are small in all cases but one, going up or down by one gradation (e.g., low to moderate or high to moderate). This is likely because of the uncertainty of the borders between low, moderate and high, or perhaps simply in the interpretation of what is considered low, moderate and high. Either way, for these cases, it is likely of little concern. In one case, however, the error is two gradations: the change in population is expected to have a large impact, when the impact is revealed to be small. While the direction of the effect is as expected (faster growth linked to higher congestion levels), the size of the effect suggests that the rate of city growth is not a big player in the congestion problem. Moreover, as the VIP is low, then confidence in this variable as a player is of concern.

More problematic are the discrepancies with the direction of the effects. In 14 cases, the direction of the effect is opposite of that expected, and in three other cases, the expected direction of the effect is unknown (i.e., is not estimated in Chapter 4 above). In most of these cases (12 of 17), the revealed effect on congestion is positive, when the expected effect is negative or unknown; in the other five cases, it is the opposite. For example, one might expect that a larger relative transit network (VRM_SqMi) would have a positive impact on congestion (congestion would decrease); instead, the effect is revealed to be negative. This could indicate the effects of an unidentified variable, or an insufficient reaction-type problem (i.e., the larger relative transit systems arose in response to increasing congestion but not quite enough, which may require a time-lag analysis to delineate). Alternatively, the estimates of both the size and the direction of the effects could simply be poor estimates, although I think there is a sound case for the estimates made. (It could also be that the revealed effect is accurate and that the net impact of transit on congestion is negative. Transit may not remove enough cars from the network to offset the contribution to congestion of large, plodding buses with their many stops, a possibility that might be better studied in a micro-analysis.) Regardless, the revealed effects are important in understanding the relationships between the selected urban characteristics and congestion so these conflicts need some discussion. It is important to note that ten of the 17 discrepancies in direction are for variables with low VIPs. These ten are included in the discussion, even though the discrepancy is suspect.

The twelve “negative, but positive” or “unknown, but positive” cases are in order of size of effect, high to low: PctOldYng, SpTms_Mcap, PctSOV, PopMed, GDP_VMT, and seven low VIP variables (PctGovtEmp, Inflows_Wkr, GeoMW, GeoW, PctTrks, GPop_NetMi, and PctJobsJRDTcts).

- PctOldYng – the percentage of the population that is either old or young is an attempt to address the “driving while distracted” set, with the idea that older and younger drivers are more likely to be part of this set. Since drivers' ages are not available for each urban area, population ages are used, with the assumption that the percentages of drivers in each age group are uniform across the nation. If this is true, then reasonably, more distracted drivers are linked with worse congestion. The revealed effect is the opposite, so either the assumption is invalid or other mitigating variables are at play. In urban areas with abundant alternative transportation modes, many old-young people may choose to use these modes in lieu of driving. In such instances, a larger percentage of the old-young could actually translate to fewer drivers. Instead of more distracted drivers having a negative effect on congestion, fewer drivers have a positive effect.

- SpTms_Mcap – this variable is a surrogate for the special events that disrupt traffic flows, with more special events having a negative impact on congestion. The revealed effect, however, is positive; more sports teams per capita are linked to lower congestion. How can this be? A review of the rank-ordered data indicates that the higher values for this measure are for the mid-size and even some of the smaller cities. (For example, Boulder is the smallest city, but has the fifth largest SpTms_Mcap value.) This is because the larger number of sports teams in the larger cities is more than offset by the increased population in these cities. So this variable actually points to urban areas with less of a congestion problem than can perhaps more readily absorb the increased traffic woes associated with special events. Clearly, this surrogate is not up to the task of capturing the degree of disruption in traffic flows from special events; better data are required. It may be, however, that even with better data, the effect would be unclear. The number of special events may be a manifestation of city size and population and their disruptive effects may be a wash.

- PctSOV – in popular culture, the large percentage of drivers commuting in SOVs seems to be the "poster child" for urban area congestion problems. Much has been written about the negative effects of so many SOV drivers and the need to reduce their number. It seems reasonable, then, that a larger percentage of SOV drivers would be linked to worse congestion. Instead, this assessment shows the opposite. This totally counter-intuitive result is likely a chicken-and-egg type problem. People may drive SOVs in large numbers where there are fewer alternative modes and where traffic flows smoothly enough so that there are fewer penalties for using SOVs. Indeed, rank ordering the PctSOV shows that the larger cities are clustered together at the bottom of the list. Perhaps, smaller percentages of SOVs are a reaction to congestion rather than a factor of it.
- PopMed – the UMR population groupings, as a whole, are expected to have a moderate, negative effect on congestion; i.e., as city populations increase, congestion worsens. When the groupings are broken out into four dummy variables, however, the collective effect must also be broken out into expected effects for the four groups. The average data from Table 5 above show that, in general, the smaller cities have less congestion in all dimensions than the larger cities, with congestion in large and very large urban areas being the most problematic. The PLS regression results show that mid-size city congestion is only slightly worse than that in the small cities, which suggests that a better expected effect for small and medium cities would be positive; the expected effect

for large and very large cities would remain negative. This positive effect for mid-size cities is what the results revealed.

- GDP_VMT – real GDP per VMT is expected to have a negative effect on congestion; as it increases, congestion increases as well. This is the effect for two of the three congestion dimensions (intensity and duration), but the effect is very low, with both standardized coefficients falling below 0.01. The effect on congestion extent, however, is positive and this effect, while still low, is an order of magnitude larger at -0.20 (which explains the overall positive effect assessed). Greater GDP per VMT, therefore, is linked to a smaller portion of lane-miles that are congested, a counterintuitive finding. Perhaps, here, the total wealth of the urban area, while being linked with more intense and longer duration travel, can also serve to provide funds to tackle some of the problem areas in the network, thereby reducing the extent of the problem.
- PctGovtEmp – with government employees predicted to work more regular jobs and thus contribute more to the peak hour rush, a negative impact on congestion is expected. Instead, as the percentage of government workers increases, congestion decreases. Perhaps government workers do not work more regular hours. Indeed, at the local level, the sheriff's department and the school system are often the largest government departments and workers in these fields often work outside the "normal" 9-to-5 day. Regardless, the VIP for this variable is low, so these effects are problematic.
- Inflows_Wkr – the net inflows of workers from outside the urban area relative to those already in the urban area would logically have a negative effect on

congestion with a greater relative in-flow being linked with worse congestion. The data reveal the opposite, although with a low VIP, the results are suspect. Since the relative inflow is a ratio variable (net in-commuters to total workers), it may be that it is easier to have a high value in smaller urban areas. In the larger areas; the inflow numbers in the numerators are simply overpowered by the total worker numbers in the denominators.

- GeoMW and GeoW – there is a question of the impact of culture (and in particular, planning culture) on congestion and the dummy variables representing the four Census regions of the country (Northeast, South, Mid-west, and West) try to address it. For each variable the effect is unknown and any results are new information. In the analysis, the Northeast region dummy variable is the base case (i.e., is excluded from the model to prevent perfect collinearity) and the results must be interpreted with this in mind, which can be tricky and less straightforward than the population dummy variables. Here, the results suggest that relative to the Northeast, being in the South is linked with worse congestion in all dimensions, while being in the Mid-west or West is linked with lower congestion in all dimensions. This does not fully jibe with the Census region averages in Table 6, but those averages do not consider the interactions among the variables that are included in the PLS regression results. As the VIP is very low, however, confidence in this new information is also low.
- PctTrks – most people would probably agree that a larger percentage of trucks on the freeways would be associated with higher congestion (a negative effect), but the PLS regression results reveal that the effect is positive. These effects are quite

small, however, (the standardized coefficients are -0.011, -0.011, and -0.010), and the VIP for this variable is low, so these effects are suspect. Still, further research using a larger sample might better tease out the effects of trucks on urban congestion.

- GPop_NetMi – a higher Gini coefficient of population per network mile means a less equitable distribution of network mileage across all census tracts in the urban area, which is expected to lead to worse congestion. So as the Gini coefficient increases, congestion is expected to increase also. Surprisingly, it decreases. While the standardized coefficients range between -0.010 and 0.025 and the effects are quite small, they are nonetheless positive. A review of the rank-ordered values of this variable reveals no clear pattern in the data; cities with known high TTIs are scattered throughout the list. It could be that people have adapted to the network they have, or the network has responded to the demands of the people and the people are not uniformly demanding network access. It could also be that uniform access to the network is not a good measure of congestion for these same reasons. Regardless, the VIP is low, so these effects are somewhat questionable.
- PctJobsJRDTcts – the percentage of jobs in job-dense, job-rich Census tracts is a measure of centrality, with the idea that higher levels of centrality are associated with higher levels of congestion. The PLS regression results show that this is true only for congestion duration and even that effect is extremely small (the standardized coefficient is 0.001) The revealed effects for intensity and extent are positive, although they are also very small (standardized coefficients of -0.002

and -0.016, respectively). It could be that this variable does not fully capture the notion of centrality. It could also be that the low VIP makes the results questionable at best.

The five “positive, but negative” or “unknown, but negative” cases are again in order of size of effect, high to low: LeePoly, VRM_SqMi, and three low VIP variables (GeoS, Links_Node, and Pat_KWkrs).

- LeePoly – a higher level of poly-centricity might reasonably be linked to less congestion, with the idea that more centers of employment in an urban area allow more of the network to be utilized in the commute. This does assume that households have made location decisions based on employment center proximity, which may not always be the case, especially in two-worker households or when school quality does not mesh with network access. This also assumes that the network is uniform in its support of all employment centers, and since the results seem to say that increased poly-centricity is linked to worse congestion, this assumption may not always be valid. It may be that the transportation networks supporting the newer nuclei in the urban region are unable to handle the commuter loads that can often arise quite rapidly. A more micro-analysis would be needed to determine if this were the case.
- VRM_SqMi – the density of transit service has a low, negative effect on congestion – the higher the density the worse the congestion, which is somewhat counterintuitive. One might reasonably think that more widespread transit service would alleviate traffic flows, but if increases in transit service lagged the increases in privately owned vehicles (POVs), then this variable could be a

marker for increased congestion. It might also be that increased transit traffic might interrupt traffic flows to a greater extent than it removes POVs from the network. Another possibility is that any increases in transit coverage are a direct response to increased congestion, so more coverage would be linked to worse congestion. It could also be that different types of transit are correlated with congestion in different ways. For example, rail cars and buses with dedicated guideways may be linked with congestion much more differently than transit that uses the street network. Since the VRM_SqMi includes all types of transit, these impacts may be offsetting and contributing to the unexpected effect.

- GeoS – like the GeoMW and GeoW variables in the "negative, but positive" paragraph above, the expected effect for this variable is unknown, so the results are new information more than a contradiction of an expected effect. Also like the other two dummy variables, the results for GeoS must be interpreted given the GeoNE dummy variable as the base case. Here it seems that relative to the Northeast, being in the South is linked with worse congestion in all dimensions. The VIP is low, however, so the results may not be correct.
- Links_Node – it is expected that increased connectivity is linked with lower congestion levels; the more routes one has to a destination, the more one can avoid traffic. This assumes, as do many of the rational actor models, that information is perfect an all congested areas are known as are all alternative routes around the congested areas. This is clearly not the case – we all have run into unexpected congestion with no idea how to get around it. The effect here is

extremely small (standardized coefficients range from -0.001 to 0.011 and the VIP is low) so the results are problematic.

- Pat_KWkrs – it is theorized that a creative workforce, as measured by patents per thousand workers, might be "good for congestion," but its revealed impact is negative. This might be an indicator that "creative" people are not that different from "normal" people, and the measure of patents is another measure of urban area vitality, with the idea that urban areas with higher levels of vitality or energy have higher congestion levels. Regardless, the VIP is low, so the results are suspect.

5.3.5 Most Important Variables as Determined by PLS Regression Results. The top ten variables in terms of size of effect as measured by the size of the standardized coefficients are shown in Table 22. Seven of these are common to all congestion dimensions (AvgCmtTime, UASqMi, PopVLg, PopLg, FwyArtMi_Cap, FwyLM_KCmtr, and PctOldYng) and three of these deal with urban area size (UASqMi, PopVLg, and PopLg). The presence of the size variables is not that all surprising given the population group averages in Table 5 and the Pearson correlations in Table 14. Their continued presence in a multivariate assessment does, however, suggest that size seems to matter greatly, at least as far as congestion is concerned. It may be that large cities can really do little to eliminate congestion, but instead just have to adapt to it.

Table 22: Top ten standardized coefficients by congestion variable

TTI	**Rank**	**S.C.**	**PortLMCong**	**Rank**	**S.C.**	**PkHrs**	**Rank**	**S.C.**
AvgCmtTime*	1	0.113	FwyArtMi_Cap*	1	-0.129	AvgCmtTime*	1	0.125
UASqMi*	2	0.105	AvgCmtTime*	2	0.116	PopLg*	2	0.119
PopLg*	3	0.099	FwyLM_KCmtr*	3	-0.114	UASqMi*	3	0.117
FwyArtMi_Cap*	4	-0.092	PopLg*	4	0.087	FwyArtMi_Cap*	4	-0.097
UAPop-K	5	0.086	UASqMi*	5	0.086	UAPop-K	5	0.093
PopVLg*	6	0.078	GeoS	6	0.084	Inc_Cap	6	0.086

TTI	Rank	S.C.	PortLMCong	Rank	S.C.	PkHrs	Rank	S.C.
Inc_Cap	7	0.072	FwyMi_SqMi	7	-0.083	PopVLg*	7	0.084
PctOldYng*	8	-0.071	PctOldYng*	8	-0.074	PctOldYng*	8	-0.080
FwyLM_KCmtr*	9	-0.067	PctPopCh	9	0.070	Nodes_UpNetMi	9	0.076
Nodes_UpNetMi	10	0.067	PopVLg*	10	0.069	FwyLM_KCmtr*	10	-0.070

* Variable is in the top ten for all three dependent variables.

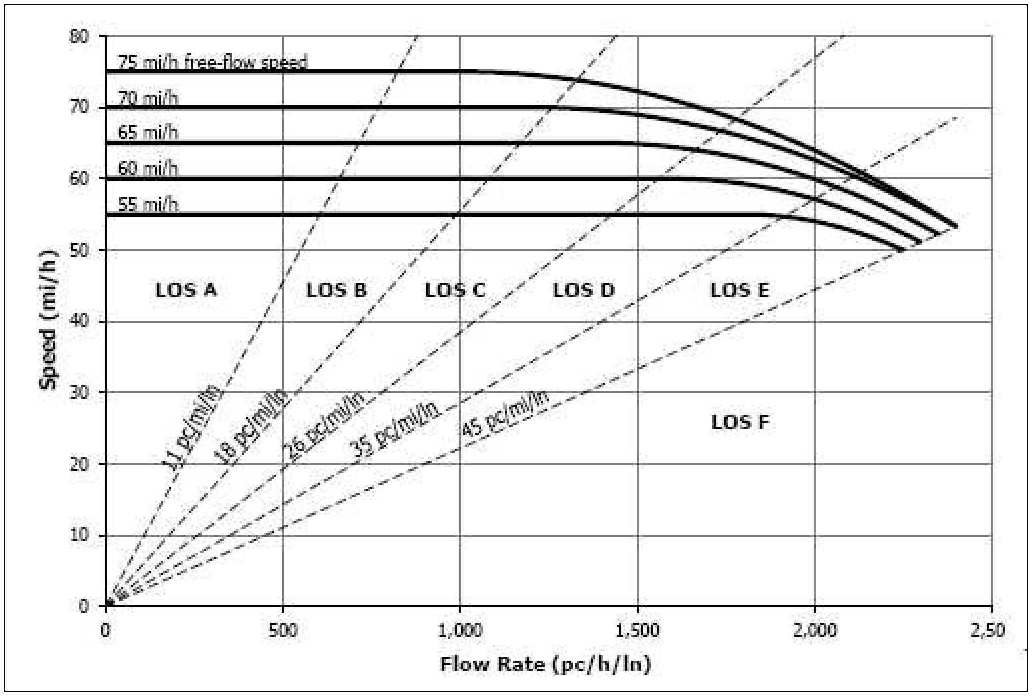
Of the other four in the top ten that are common to all dimensions, three seem reasonably self-evident: a measure of travel time (AvgCmtTime) has a negative effect on congestion (as it increases congestion gets worse), while two measures of network supply (FwyArtMi_Cap and FwyLM_KCmtr) have a positive effect (as they increase, congestion improves). The final member of the "common seven" is unexpected. PctOldYng has an effect on congestion that is surprising both in direction (revealed positive versus expected negative, as discussed above) and in size (revealed moderate versus expected low).

There are three variables in the top ten that are common to two of the dimensions of congestion: UAPop-K, Inc_Cap, and Nodes_UpNetMi are common to both TTI and PkHrs. In other words, the dimensions of intensity and duration have the same variables in the top ten, although not in the same order. The revealed effects for each are as expected. The dimension of extent has three different variables in its top ten: FwyMi_SqMi has a negative association with congestion and GeoS and PctPopCh have a positive one. The size and direction of the effects for the geographic dummy variable, GeoS, are discussed above as is the unexpected size of effect from PctPopCh (revealed to be low but expected to be high).

The remaining variables in the data set, those which have not been discussed in this section, generally have revealed results that were as expected and follow the ideas laid out in variable development. Most were low in importance and/or small in the size of their effect and are not likely to be important players in congestion analysis and remediation, at least not according to the PLS results. These results do, however, assume a linear relationship between the components and the congestion variables. Is this assumption valid?

5.4 The Linearity Assumption

A review of the binary scatter plots shows no obvious non-linear relationships between the independent and the dependent variables, so the relationship is presumed to be linear. An analysis of the residuals in the PLS regression also does not suggest that these relationships are anything but linear. There are, however, two concerns here; the former presumes linearity unless non-linearity is clear and the latter reflects relationships between the dependent variables and the components and not the independent variables themselves. There is other evidence that suggests that at least some variables have a non-linear relationship with congestion. For example, the freeway speed-flow curves (Figure 4) from the 2010 Highway Capacity Manual (TRB 2010), which show the relationship between the levels of service, the speed and the flow rate. These curves indicate that speeds are largely unaffected by traffic volumes until volumes begin to reach facility capacity, at which time speeds begin to fall off with increasing rapidity. This would seem to have implications for other variables as well. It does not seem coincidental that urban area size measures figure so prominently in the most influential variables. Perhaps a non-linear relationship is at play.



Source: 2010 Highway Capacity Manual
Notes: LOS = level of service; pc = passenger car; ln = lane

Figure 4: Freeway speed-flow curves

One way to get around this linearity issue is to use an analytical method that does not assume a linear (or non-linear) relationship between the dependent and independent variables, but instead, considers the variables as they are. Decision tree methodologies do just this.

5.5 Chi-square Automatic Interaction Detection (CHAID) Analysis Results

A decision tree analysis allows the uncovering of the key independent variables by analyzing the splits that occur as the tree is grown. There are at least two ways of assessing the significance of these splits: by a variable's use as a "splitting" variable, together with the portion of the observations that it is used to split, and by the order of a variable's use in a first splitting operation, that is, the order in a stepwise reduction process in which the variables are used to begin tree growth. (This will be explained more fully below.)

In general, the earlier a variable is used in the splitting process, the larger the portion of observations involved and the more important the variable is in the growth of the tree (Shmueli, Patel and Bruce 2007). This is not always true, of course, especially as the tree grows deeper. This "whole tree" approach, which is the primary purpose of a CHAID analysis, allows the development of good predictive models, and also furthers the understanding of which variables are important on the various branches of the tree; it generally considers variable importance in the context of the whole tree.

There is, however, an alternative to the whole tree perspective; a decision tree analysis also allows the assessment of variable importance at the first split. This is important because the first split is the only split that considers all the observations for the dependent variable. In subsequent splitting operations as the tree grows "deeper," ever smaller subsets of the total observations are considered in each split. While it may be that variable x is useful in splitting small urban areas with a high GDP per VMT and located in the Mid-west, it may not be useful at all in splitting all urban areas, regardless of characteristic. To consider just the first split, a stepwise approach to tree growth is used. In this iterative process, the decision tree is grown and the first splitting variable is identified and then eliminated from the variable set. A second decision tree is then grown from the reduced variable set and so on until all variables have been used in a first split. The order in which the variables are used in a first split and then eliminated from the variable set is an indicator of their importance. This process has the added benefit of allowing the relationships between the dependent and independent variables to be graphed and reviewed for indications of non-linearity.

5.5.1 Tree Structure and Variable Importance. A CHAID analysis is completed for each of the three dependent congestion variables, with a focus on identifying the splitting variables. (Unlike PLS regression, CHAID does not allow all three dependent variables to be considered together in one model so that their interrelationships would be included in the model results. Instead, the dependent variables are considered separately.) The XLSTAT default values are used, which include a maximum tree depth of six tiers, a significance level of five percent, and merge and split levels of five percent each. These default values seem reasonable: the tree depth seems to be appropriate given a sample size of 100 cities (more tiers would mean more smaller branches, each with increasingly less meaning) and the significance levels are widely used research parameters. Once completed, the independent variables and the number of cases (i.e., urban areas) involved in each split are identified and the splitting variables are rank ordered to determine their importance in the model. There are several instances where an independent variable is used more than once in the splitting operations. In some, this involved a split along the same branch; in others, the split is on a different branch. In both situations, the numbers of cases are added in the variable importance calculations.

The tree structures for each of the three models are shown in Figures 5-7 below. Each node has a number assigned by the software and includes four data items: the splitting variable, the range of values for the splitting variable, the percent of cases included in the split, and the predicted value of the dependent variable. (Note that since there are 100 cases, the percentage of cases is the same as the number of cases.) Terminal nodes are shaded.

The tree structure for the TTI model includes 39 nodes, 19 of which are terminal. There are 19 splitting operations involving 16 different independent variables. There are 36 nodes in the tree structure for the PortLMCong model with 19 of these being terminal. Sixteen different independent variables are used in 17 splitting operations. The PkHrs model has 21 terminal and 40 total nodes. There are 19 splitting operations involving 14 different independent variables.

The tree structure is relatively easy to understand. As an example, in the TTI model, the variable GDP_VMT allows the best split based on the splitting criteria noted above. The tree (Node 1) is split into two branches, one (Node 2) with 14% of the cases (observations), which have GDP_VMT values between 2.033 and 7.712, and one (Node 3) with 84% of the cases, which have GDP_VMT values between 7.712 and 14.524. The predicted TTI in the former is 1.166 and in the latter, 1.204. Considering only these two branches, as GDP_VMT increases, TTI also increases. Each node can be so analyzed to determine how the splitting variable behaves on the branches off that node. If only two branches are grown from a node, then the "revealed" relationship must be linear; if more than two are grown, then non-linear relationships may be revealed. In the next tier, two splits are performed using two different variables. For urban areas with smaller GDP_VMT, the splitting variable is GMedInc_HH; for urban areas with larger GDP_VMT, the splitting variable is PopMed. For the former, a spread variable yields the best split while in the latter, it is a population dummy variable. At different points in tree development, then, variables that are not useful anywhere else may well come to the fore.

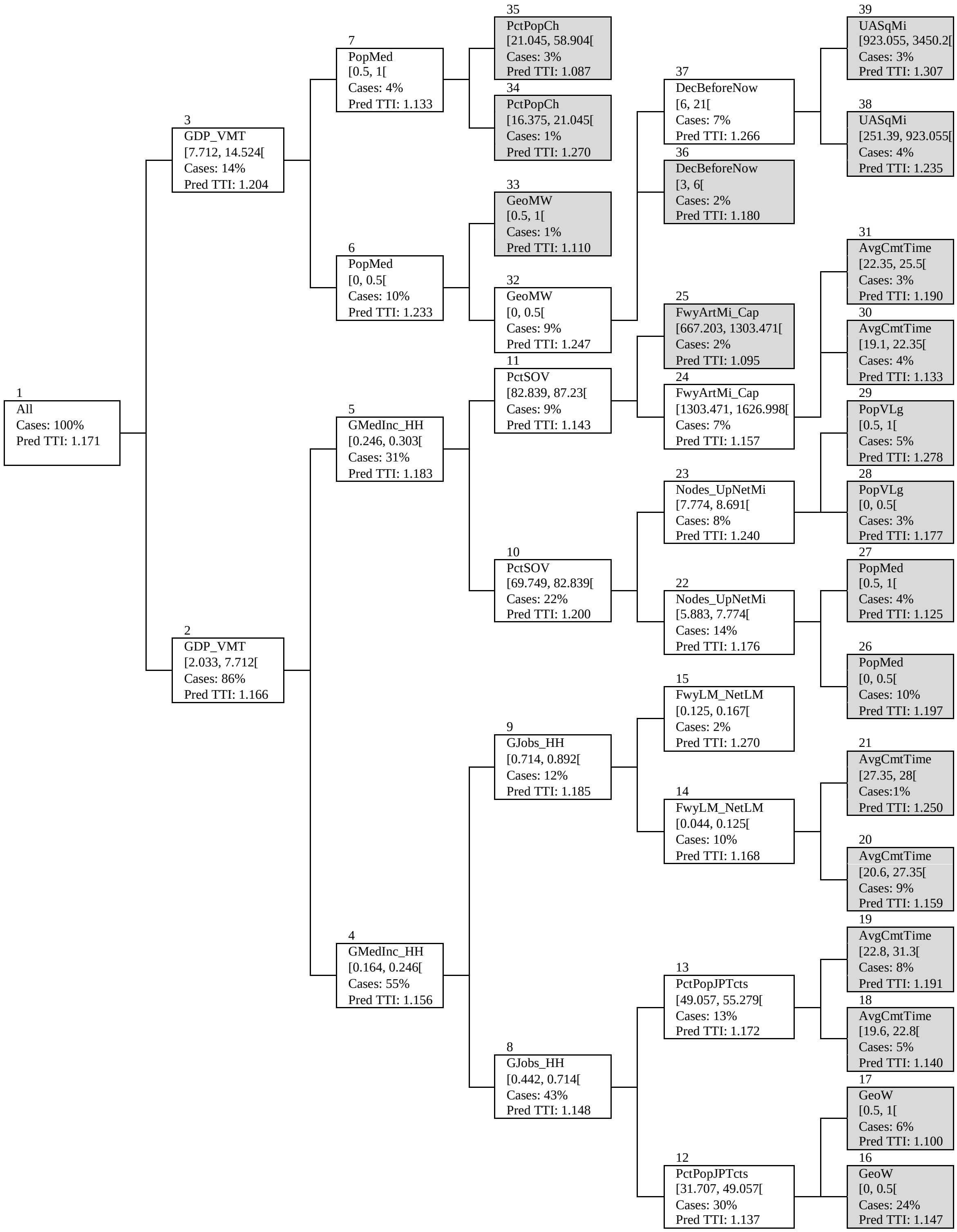


Figure 5: CHAID tree structure for the congestion intensity (TTI) model

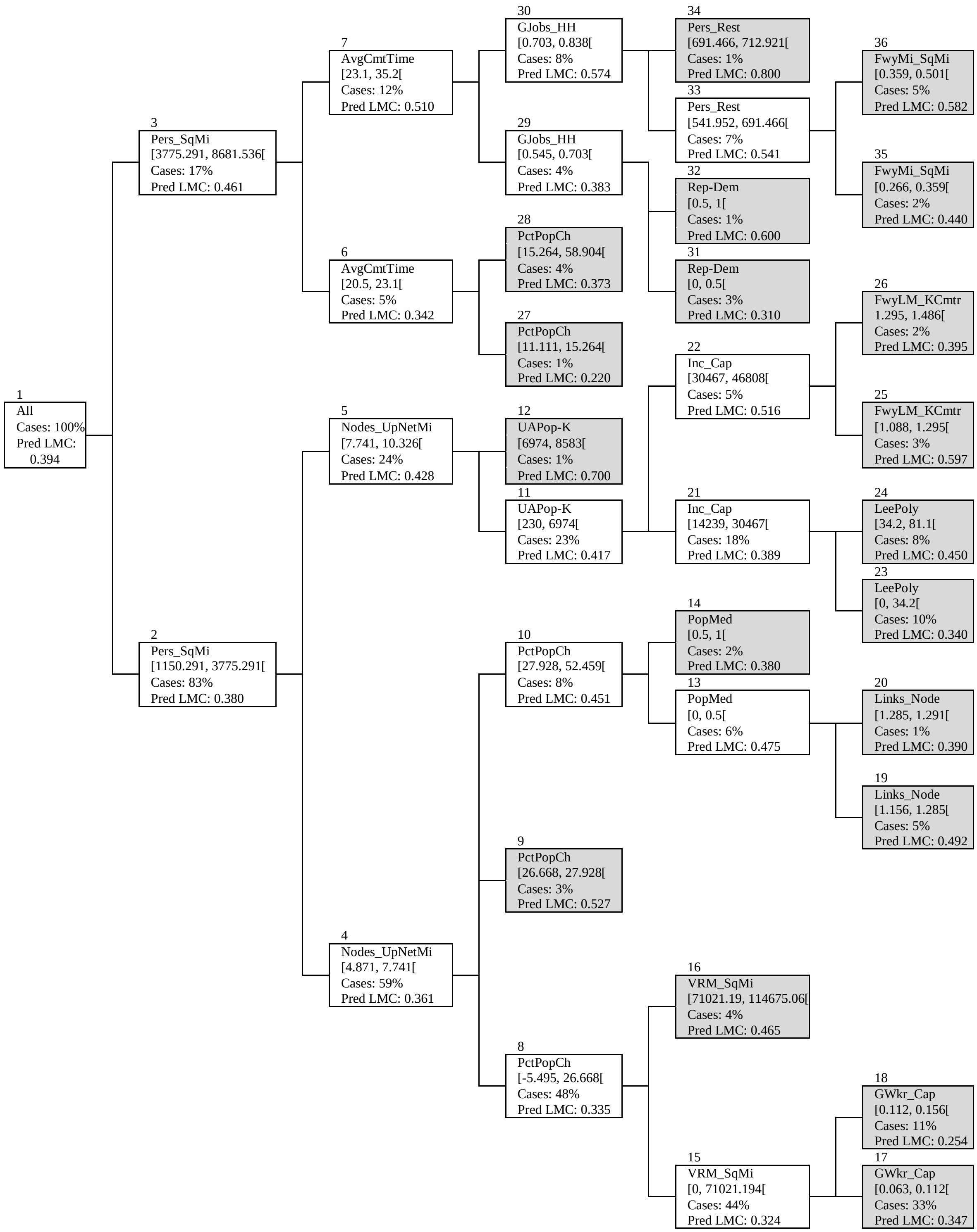


Figure 6: CHAID tree structure for the congestion extent (PortLMCong) model

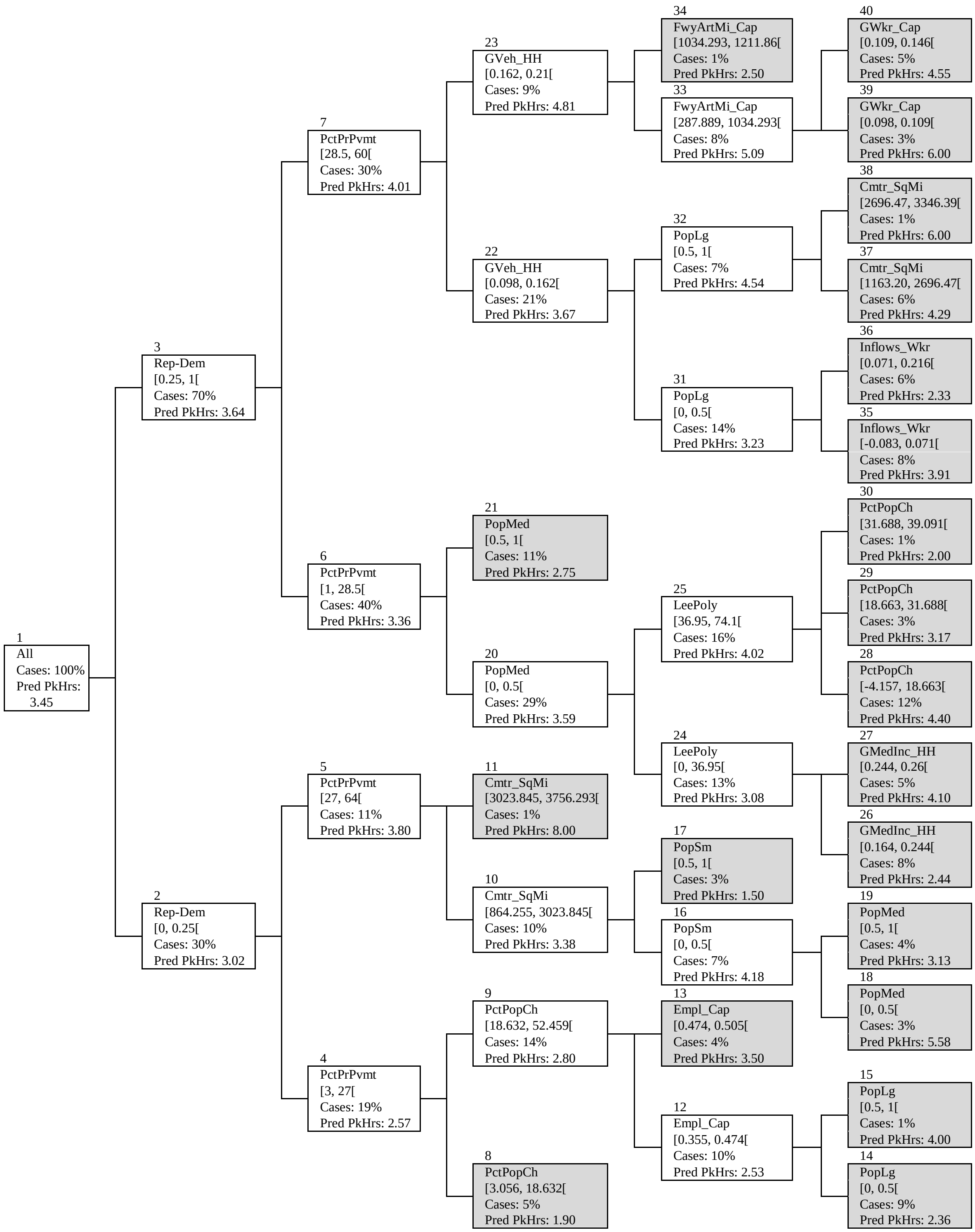


Figure 7: CHAID tree structure for the congestion duration (PkHrs) model

The independent variables involved in the CHAID splitting operations are shown in Table 23 with their relative importance as measured by the number of different cases in the splits in which they were involved. Of these important variables, only two are common among all dimensions. Interestingly, they are two urban size variables. One is the dummy variable for mid-size cities (PopMed) and the other is the change in city size variable (PctPopCh). Again, city size is at the top of the leader board in variable importance.

Table 23: Splitting variables for each CHAID model in order of importance

Intensity (TTI) Model			**Extent (PortLMCong) Model**			**Duration (PkHrs) Model**		
Variable	**Splits**	**Cases**	**Variable**	**Splits**	**Cases**	**Variable**	**Splits**	**Cases**
GDP_VMT	1	100	Pers_SqMi	1	100	PctPrPvmt	2	100
GMedInc_HH	1	86	Nodes_UpNetMi	1	83	Rep-Dem	1	100
GJobs_HH	1	55	PctPopCh*	2	64	PopMed*	2	47
PctPopJPTcts	1	43	VRM_SqMi	1	48	PctPopCh*	2	35
PctSOV	1	31	GWkr_Cap	1	44	PopLg	2	31
AvgCmtTime	3	30	UAPop-K	1	24	GVeh_HH	1	30
GeoW	1	30	Inc_Cap	1	23	LeePoly	1	29
PopMed*	2	28	LeePoly	1	18	Cmtr_SqMi	2	18
Nodes_UpNetMi	1	22	AvgCmtTime	1	17	Empl_Cap	1	14
FwyLM_NetLM	1	12	GJobs_HH	1	12	Inflows_Wkr	1	14
GeoMW	1	10	Pers_Rest	1	8	GMedInc_HH	1	13
DecBeforeNow	1	9	PopMed*	1	8	PopSm	1	10
FwyArtMi_Cap	1	9	FwyMi_SqMi	1	7	FwyArtMi_Cap	1	9
PopVLg	1	8	Links_Node	1	6	GWkr_Cap	1	8
UASqMi	1	7	FwyLM_KCmtr	1	5			
PctPopCh*	1	4	Rep-Dem	1	4			

* Variable is listed as statistically significant for all three dependent variables.

5.5.2 The First Split and Variable Importance. A series of CHAID analyses are completed for each of the three dependent congestion variables, with a focus on identifying the splitting variables involved in the first split only. Again, the XLSTAT default values are used, which include a maximum tree depth of six tiers, a significance level of five percent, and merge and split levels of five percent each. Once completed, the independent variable used in the first split is identified and then deleted from the variable set. (Note: XLSTAT does not have a sub-routine to complete this iterative process, so it must be completed manually.) The CHAID analysis is run again and again with the ever smaller variable set until all independent variables have been involved in a first split. Since an analysis is completed for each of the three dependent variables and the order of independent variable elimination varies for each of the three sets of runs, 156 separate runs (3 x 52) are required. As the process wears on and the variable set grows smaller, there comes a time when there is no variable that will make the first split at the original default significance settings. At this point, the significance levels are loosened so that a first split is attained. The order in which the variables are used in a first split and then eliminated from the variable set reflects their relative importance to the particular dimension of congestion as a whole. (One limitation of the variable importance in the tree structure calculations above is that every variable after the first split is involved only with a branch of the tree. Their importance to the branch may not extend to the whole tree. This approach gets around this limitation.)

Table 24 shows the order of involvement of the independent variables in a first split operation for each of the three congestion dimensions, along with the significance level of the split. The variables that are not able to split with a significance level above 0.05 are shaded; these variables are unlikely to be important to the particular dimension of congestion. (While one may argue for inclusion of variables able to split at a significance level of 0.10, this study uses a cut-off of 0.05.)

Table 24: Order of variable involvement in a first split by congestion dimension

Intensity (TTI) Model		Extent (LMCong) Model		Duration (PkHrs) Model	
Split Variable	**Significance**	**Split Variable**	**Significance**	**Split Variable**	**Significance**
GDP_VMT	5	Pers_SqMi*	5	Rep-Dem	5
FwyArtMi_Cap*	5	GeoMW	5	Links_Node	5
GWkr_UpNetMi*	5	Inc_Cap*	5	Pers_SqMi*	5
GMedInc_HH	5	Inflows_Wkr	5	Empl_Cap	5
VRM_SqMi*	5	PopMed*	5	FwyLM_NetLM	5
GJobs_SqMi	5	PopLg*	5	AvgCmtTime*	5
PctSOV*	5	Med_Mult*	5	DecBeforeNow	5
Empl_Cap	5	GJobs_HH*	5	FwyLM_KCmtr*	5
LeePoly*	5	Nodes_UpNetMi*	5	GWkr_UpNetMi*	5
Rep-Dem	5	UAPop-K*	5	PctGovtEmp	5
Pers_SqMi*	5	GWkr_UpNetMi*	5	Cmtr_SqMi	5
FwyLM_KCmtr*	5	FwyLM_KCmtr*	5	PctOldYng	5
FwyMi_SqMi	5	FwyMi_SqMi	5	FwyArtMi_Cap*	5
DecBeforeNow	5	GJobs_Wkr*	5	PctPopJPTcts*	5
Cmtr_SqMi	5	PctPopJPTcts*	5	GJobs_HH*	5
PopLg*	5	FwyArtMi_Cap*	5	GMedInc_HH	5
PctRetEmp	5	LeePoly*	5	PopMed*	5
GJobs_HH*	5	PctSOV*	5	PctSOV*	5
PopMed*	5	SpTms_Mcap*	5	PctPrPvmt	5
AvgCmtTime*	5	AvgCmtTime*	5	VRM_SqMi*	5
Pers_Rest	5	PopSm*	5	Nodes_UpNetMi*	5
PctPopJPTcts*	5	UASqMi*	5	Inc_Cap*	5
GJobs_Wkr*	5	VRM_SqMi*	5	GVeh_HH	5
PctPrPvmt	5	PopVLg*	5	Pers_Rest	5
Inc_Cap*	5	NetMi_SqMi**	10	SpTms_Mcap*	5
Nodes_UpNetMi*	5	PctPrPvmt	10	PopLg*	5
SpTms_Mcap*	5	Cmtr_SqMi	15	PctRetEmp	5
GVeh_HH	5	GeoNE**	15	UAPop-K*	5
UAPop-K*	5	GeoW**	15	Med_Mult*	5
Med_Mult*	5	GPop_NetMi**	15	GJobs_Wkr*	5
PopSm*	5	GJobs_SqMi	15	LeePoly*	5
UASqMi*	5	FwyLM_NetLM	20	UASqMi*	5
PopVLg*	5	Pers_Rest	20	PopSm*	5
Veh_HH**	10	PctOldYng	20	PopVLg*	5
Inflows_Wkr	10	GeoS**	20	Inflows_Wkr	10
NetMi_SqMi**	10	PctJobsJRDTcts**	20	Pat_KWkrs**	10
Crashes_Kcap**	10	PctPopCh**	20	Crashes_Kcap**	10
PctTrks**	15	YrPrecipIn**	20	GDP_VMT	10
Links_Node	15	Pat_KWkrs**	20	GJobs_SqMi	15
GeoMW	15	PctTrks**	20	GPop_NetMi**	15
FwyLM_NetLM	15	GVeh_HH	20	FwyMi_SqMi	20
PctPopCh**	20	Veh_HH**	25	NetMi_SqMi**	20
GeoNE**	20	PctGovtEmp	50	PctPopCh**	25
PctOldYng	25	Links_Node	50	PctTrks**	50
GPop_NetMi**	50	GDP_VMT	50	PctJobsJRDTcts**	50

Intensity (TTI) Model		Extent (LMCong) Model		Duration (PkHrs) Model	
Split Variable	**Significance**	**Split Variable**	**Significance**	**Split Variable**	**Significance**
Pat_KWkrs**	50	Empl_Cap	50	Veh_HH**	75
GWkr_Cap**	75	Crashes_Kcap**	75	GeoNE**	75
PctGovtEmp	75	GMedInc_HH	75	GeoS**	75
GeoW**	99	DecBeforeNow	99	GeoW**	99
YrPrecipIn**	99	PctRetEmp	99	YrPrecipIn**	99
GeoS**	99	GWkr_Cap**	99	GeoMW	99
PctJobsJRDTcts**	99	Rep-Dem	99	GWkr_Cap**	99

* Important in all three dimensions
** Not important in any dimension

The TTI model has 33 important variables, the LMCong model has 24, and the PkHrs model has 34. There are 21 variables that are important in all three dimensions and there are 13 that are not important in all dimensions; the other 18 are important in one or two dimensions only. The variation between the dimensions in important variables suggests that there are different underlying factors associated with each dimension of congestion and that different strategies would likely be required for remedial actions.

The top ten variables in terms of importance in first splits are shown in Table 25. The top variables are determined by ranking the variables by dimension based on their order of use in first splits, adding the ranks of the three dimensions, and then sorting the sum of the ranks. This moves the variables that are important to congestion collectively to the top of the list; the top ten are marked. Finally, the top ten variables are sorted by rank by dimension so that their relative order within the dimension is readily discernable. Skipped rankings indicate a variable that, while important within one dimension, is not important in all dimensions collectively.

All ten of these variables are common to all congestion dimensions, although the variables rankings within each dimension are different. Two variables deal with urban area size (PopMed and PopLg) which continues to reinforce the notion that with congestion, size matters. Three others are repeat top tens from the PLS regression results: FwyArtMi_Cap, FwyLM_KCmtr and AvgCmtTime. None of these are real surprises; the size and magnitude of the effects for each variable are about as expected. Of the remaining five, two (PctSOV and VRM_SqMi) have some counterintuitive effects that are discussed in the PLS results section above, and two (GWkr_UpNetMi and GJobs_HH) that indicate that the distribution of the workforce across the urban footprint is linked to the levels of congestion. The final variable in the top ten is population density (Pers_SqMi). It seems clear from these results that density has a negative relationship with congestion (as urban areas become more dense, congestion worsens), although many new urbanists believe that increased density is a necessary precondition for renewed urban vitality and the development of alternative modes of transportation that could reduce the prominence of the automobile, thereby leading to a reduction in congestion.

Table 25: Top ten first split variables by congestion variable

TTI	**Sig.**	**Rank**	**PortLMCong**	**Sig.**	**Rank**	**PkHrs**	**Sig.**	**Rank**
FwyArtMi_Cap*	5	2	Pers_SqMi*	5	1	Pers_SqMi*	5	3
GWkr_UpNetMi*	5	3	PopMed*	5	5	AvgCmtTime*	5	6
VRM_SqMi*	5	5	PopLg*	5	6	FwyLM_KCmtr*	5	8
PctSOV*	5	7	GJobs_HH*	5	8	GWkr_UpNetMi*	5	9
Pers_SqMi*	5	11	GWkr_UpNetMi*	5	11	FwyArtMi_Cap*	5	13
FwyLM_KCmtr*	5	12	FwyLM_KCmtr*	5	12	GJobs_HH*	5	15
PopLg*	5	16	FwyArtMi_Cap*	5	16	PopMed*	5	17
GJobs_HH*	5	18	PctSOV*	5	18	PctSOV*	5	18
PopMed*	5	19	AvgCmtTime*	5	20	VRM_SqMi*	5	20
AvgCmtTime*	5	20	VRM_SqMi*	5	23	PopLg*	5	26

* Variable is in the top ten for all three dependent variables.

5.5.3 Linearity and Congestion Dimension. Understanding the nature of the relationships between the dependent variables and the independent variables is important.

As discussed above, these relationships are commonly assumed to be linear. An analysis of the scatter plots for the first split results, however, suggests that this is not always the case. In the first split operations performed in Section 5.5.2, the 100 observations (i.e., urban areas) are divided into two or more groups (tree branches). The number of observations in each group and a predicted value for the dependent variable are computed in tree development. These values are plotted, trend lines are added, and the resulting graphs are assessed with a focus on identifying relationships that are not linear. In most cases (42 of 52 for TTI, 39 of 52 for PortLMCong, and 40 of 52 for PkHrs), there are only two groups in the first split, so the relationship is revealed to be linear, even though it may not be. In 14 of the 35 variables with three or more groups in the first split, the scatter plots show relationships that are linear, increasing (arcs bending upwards for negative or downwards for positive) or decreasing (arcs bending to the right for both negative and positive). In the 21 remaining cases, the scatter plots show evidence of a non-linear relationship, with three U-shaped, twelve inverted U-shaped, and six in an up-and-down, indeterminate pattern. Table 26 shows these relationships, which may vary between the dimensions, along with their relative importance (order of use) in the first split operations. The variables that are not involved in a first split at the 0.05 level of significance are shaded; for these variables the relationships are suspect. Also included in the table is the direction of effect indicated by the trend line. For the linear (L), the increasing, and the decreasing relationships, this direction of effect accurately reflects the scatter plot. For the indeterminate non-linear (NL), the U, and the inverted U relationships, this direction of the trend line is suspect.

Table 26: Relationships between the dependent and independent variables

			TTI			PortLMCong			PkHrs		
No.	**Variable**	**Focus**	**Imp**	**Rel**	**Dir**	**Imp**	**Rel**	**Dir**	**Imp**	**Rel**	**Dir**
1	PctPopCh	Supply	42	L	**Pos**	37	L	**Neg**	43	L	**Pos**
2	Rep-Dem	Supply	10	L	**Neg**	52	U	**Pos**	1	L	**Neg**
3	NetMi_SqMi	Supply	36	InvU	Pos	25	L	Pos	42	InvU	Pos
4	FwyMi_SqMi	Supply	13	L	Pos	13	L	Pos	41	L	Pos
5	FwyLM_NetLM	Supply	41	L	Neg	32	L	Neg	5	L	Neg
6	FwyLM_KCmtr	Supply	12	InvU	**Neg**	12	L	**Pos**	8	InvU	**Pos**
7	FwyArtMi_Cap	Supply	2	L	Pos	16	L	Pos	13	NL	Pos
8	DecBeforeNow	Supply	14	U	Neg	49	InvU	Neg	7	NL	Neg
9	Links_Node	Supply	39	L	**Neg**	44	InvU	**Pos**	2	InvU	**Neg**
10	Cmtr_SqMi	Demand	15	L	Neg	27	L	Neg	11	L	Neg
11	Pers_SqMi	Demand	11	NL	Neg	1	L	Neg	3	Decr	Neg
12	Veh_HH	Demand	34	L	**Pos**	42	L	**Neg**	46	NL	**Pos**
13	Inc_Cap	Demand	25	NL	Neg	3	L	Neg	22	L	Neg
14	Empl_Cap	Demand	8	L	**Neg**	46	InvU	**Pos**	4	Decr	**Neg**
15	Pers_Rest	Demand	21	L	Pos	33	L	Pos	24	L	Pos
16	Inflows_Wkr	Demand	35	L	Pos	4	L	Pos	35	L	Pos
17	AvgCmtTime	Demand	20	Decr	Neg	20	L	Neg	6	Decr	Neg
18	PctSOV	Demand	7	Decr	Pos	18	L	Pos	18	L	Pos
19	VRM_SqMi	Demand	5	Decr	Neg	23	L	Neg	20	L	Neg
20	PopSm	Demand	31	L	Pos	21	L	Pos	33	L	Pos
21	PopMed	Demand	19	L	Pos	5	L	Pos	17	L	Pos
22	PopLg	Demand	16	L	Neg	6	L	Neg	26	L	Neg
23	PopVLg	Demand	33	L	Neg	24	L	Neg	34	L	Neg
24	GeoNE	Demand	43	L	**Neg**	28	L	**Pos**	47	L	**Neg**
25	GeoS	Demand	51	L	**Neg**	35	L	**Neg**	48	L	**Pos**
26	GeoMW	Demand	40	L	**Pos**	2	L	**Pos**	51	L	**Neg**
27	GeoW	Demand	49	L	Neg	29	L	Neg	49	L	Neg
28	PctTrks	Flow	38	L	Pos	40	L	Pos	44	L	Pos
29	PctOldYng	Flow	44	L	Pos	34	L	Pos	12	L	Pos
30	Nodes_UpNetMi	Flow	26	L	Neg	9	L	Neg	21	L	Neg
31	PctPrPvmt	Flow	24	L	Neg	26	L	Neg	19	L	Neg
32	Crashes_Kcap	Flow	37	L	Pos	47	L	Pos	37	L	Pos
33	YrPrecipIn	Flow	50	L	**Neg**	38	L	**Pos**	50	L	**Neg**
34	SpTms_Mcap	Flow	27	L	Pos	19	L	Pos	25	L	Pos
35	GPop_NetMi	Spread	45	L	Neg	30	L	Neg	40	L	Neg
36	GWkr_UpNetMi	Spread	3	L	Neg	11	L	Neg	9	L	Neg
37	GVeh_HH	Spread	28	L	Neg	41	L	Neg	23	L	Neg
38	GMedInc_HH	Spread	4	L	**Neg**	48	InvU	**Pos**	16	L	**Neg**
39	GWkr_Cap	Spread	47	L	**Pos**	51	U	**Neg**	52	NL	**Neg**
40	GJobs_SqMi	Spread	6	L	Neg	31	L	Neg	39	L	Neg
41	GJobs_HH	Spread	18	L	Neg	8	L	Neg	15	L	Neg
42	GJobs_Wkr	Spread	23	L	Neg	14	L	Neg	30	L	Neg
43	PctJobsJRDTcts	Other	52	L	**Neg**	36	InvU	**Pos**	45	L	**Pos**
44	PctPopJPTcts	Other	22	L	Neg	15	L	Neg	14	L	Neg
45	LeePoly	Other	9	L	Neg	17	L	Neg	31	L	Neg
46	Med_Mult	Other	30	L	Neg	7	L	Neg	29	L	Neg
47	PctGovtEmp	Other	48	L	Pos	43	L	Pos	10	L	Pos
48	PctRetEmp	Other	17	L	Pos	50	InvU	Pos	27	L	Pos
49	Pat_KWkrs	Other	46	L	Neg	39	L	Neg	36	L	Neg
50	GDP_VMT	Other	1	L	**Neg**	45	InvU	**Pos**	38	L	**Neg**
51	UASqMi	Other	32	L	Neg	22	L	Neg	32	L	Neg

			TTI			PortLMCong			PkHrs		
No.	**Variable**	**Focus**	**Imp**	**Rel**	**Dir**	**Imp**	**Rel**	**Dir**	**Imp**	**Rel**	**Dir**
52	UAPop-K	Other	29	L	Neg	10	L	Neg	28	L	Neg

Notes: Values are shaded for variables unable to split at a 0.05 significance level.
All variable data are shaded when unable to split at a 0.05 significance level in all areas.
Differences in direction of effect are in bold.

For linear relationships, the direction of effect is as it seems. A positive direction indicates that an increase in the independent variable is associated with a decrease in the dependent variable. (Recall that congestion is a negative situation and that a positive effect is when congestion decreases.) A negative direction results in the opposite situation. For increasing non-linear relationships, the dependent variable increases or decreases increasingly more quickly as the independent variable increases, positively or negatively. The situation is reversed for a decreasing relationship, where the positive or negative effects begin to decrease as the independent variable increases. In the U and inverted U relationships, the effect of a change in the independent variable changes direction as the variable increases. This effect is either higher in the middle than on the ends (inverted U) or lower in the middle (U).

The non-linear relationships are of particular interest to efforts at congestion remediation. For example, the freeway lane-miles per commuter (FwyLM_KCmtr) have an inverted U relationship with congestion intensity and duration, but a linear relationship with extent. Additions to freeway lane-mileage after a point might well have a diminishing effect on TTI and the number of peak hours, while continuing to improve the portion of the network that is congested. (The term “might” is used here since there is no evidence of causality in the relationships, only correlation.)

Note that there are 14 cases where there is disagreement on the direction of effect between the dimensions; five are two positives and one negative and nine are two negatives and one positive. Thirteen of these involve results from a split at a significance level below 0.05, so the disagreement is problematic. The remaining variable (FwyLM_KCmtr) involves only statistically significant splits, but does have two inverted U and one linear relationship. Since the direction of effect is derived from the trend line, the disagreement in this instance is possibly due to the trend line in the inverted U relationship not accurately representing the overall effects. (It could also be, of course, that the relationship is different for this dimension of congestion, though this seems somewhat unlikely since this occurs on one variable only.) In the 38 other cases, there is no disagreement and the effects are the same in all dimensions. A positive effect of one variable on TTI, for example, is also a positive effect on PortLMCong and PkHrs.

5.5.4 Expected vs. Revealed Effects. The expected effects theorized in Chapter 4 include both size and direction. In the PLS analysis, the regression output provided both size and direction data. The CHAID results are not so generous. While the directions of effect can be gleaned from the scatter plots, the sizes of effect are less evident. The coefficients from the trend lines are non-standardized, so comparisons are meaningless. While it is possible to derive standardized coefficients for all 156 scatter plots, the cost-benefit of this effort is questionable, especially since parsing the sizes of effects into three groups (high, moderate and low) has its own issues. Instead, the focus here is on the directions of effect only.

Table 27 shows the effects expected for congestion in general with the revealed effects for each congestion dimension. Also included in the table are the variables' focus and relative importance (as determined by the order in which they are used in the first split operations). Since this order varies by congestion dimension, the variables are listed in the order of discussion in Section 4. (For convenience, Table 24 above lists the variables in order of importance for each dimension.)

There are 26 cases where the expected effects differ from the revealed effects for at least one of the three congestion dimensions. Of these 26, eleven are cases where all the revealed data result from splitting operations at significance levels above the 0.05 threshold and another four are cases where the disagreement between the expected and revealed effect are only in the results of non-statistically significant splits. If the results of these 15 cases are discounted, there are 11 remaining cases where the expected effects differ from the revealed effects and the revealed effects are from statistically significant splits. In five of these 11, the expected effect is negative, while the revealed effect in the statistically significant cases is positive; in five, the expected effect is positive and the revealed effect is negative or mixed; and in one, the expected effect is unknown and the revealed effect is positive.

Table 27: Expected effects compared to revealed effects

			Expected		**TTI**		**LMCong**		**PkHrs**	
No.	**Variable**	**Focus**	**Size**	**Dir**	**Imp**	**Dir**	**Imp**	**Dir**	**Imp**	**Dir**
1	PctPopCh	Supply	High	**Neg**	42	**Pos**	37	**Neg**	43	**Pos**
2	Rep-Dem	Supply	Low	**Neg**	10	**Neg**	52	**Pos**	1	**Neg**
3	NetMi_SqMi	Supply	Mod	Pos	36	Pos	25	Pos	42	Pos
4	FwyMi_SqMi	Supply	Mod	Pos	13	Pos	13	Pos	41	Pos
5	FwyLM_NetLM	Supply	Mod	**Pos**	41	**Neg**	32	**Neg**	5	**Neg**
6	FwyLM_KCmtr	Supply	High	**Pos**	12	**Neg**	12	**Pos**	8	**Pos**
7	FwyArtMi_Cap	Supply	Mod	Pos	2	Pos	16	Pos	13	Pos
8	DecBeforeNow	Supply	Low	Neg	14	Neg	49	Neg	7	Neg
9	Links_Node	Supply	Mod	**Pos**	39	**Neg**	44	**Pos**	2	**Neg**
10	Cmtr_SqMi	Demand	Mod	Neg	15	Neg	27	Neg	11	Neg
11	Pers_SqMi	Demand	Mod	Neg	11	Neg	1	Neg	3	Neg
12	Veh_HH	Demand	Mod	**Neg**	34	**Pos**	42	**Neg**	46	**Pos**
13	Inc_Cap	Demand	Mod	Neg	25	Neg	3	Neg	22	Neg
14	Empl_Cap	Demand	Mod	**Neg**	8	**Neg**	46	**Pos**	4	**Neg**
15	Pers_Rest	Demand	Mod	Pos	21	Pos	33	Pos	24	Pos
16	Inflows_Wkr	Demand	Low	**Neg**	35	**Pos**	4	**Pos**	35	**Pos**
17	AvgCmtTime	Demand	Mod	Neg	20	Neg	20	Neg	6	Neg
18	PctSOV	Demand	Low	**Neg**	7	**Pos**	18	**Pos**	18	**Pos**
19	VRM_SqMi	Demand	Low	**Pos**	5	**Neg**	23	**Neg**	20	**Neg**
20	PopSm	Demand	Mod	Pos	31	Pos	21	Pos	33	Pos

			Expected		TTI		LMCong		PkHrs	
No.	**Variable**	**Focus**	**Size**	**Dir**	**Imp**	**Dir**	**Imp**	**Dir**	**Imp**	**Dir**
21	PopMed	Demand	Low	Pos	19	Pos	5	Pos	17	Pos
22	PopLg	Demand	Low	Neg	16	Neg	6	Neg	26	Neg
23	PopVLg	Demand	Mod	Neg	33	Neg	24	Neg	34	Neg
24	GeoNE	Demand	Low	**Unk**	43	**Neg**	28	**Pos**	47	**Neg**
25	GeoS	Demand	Low	**Unk**	51	**Neg**	35	**Neg**	48	**Pos**
26	GeoMW	Demand	Low	**Unk**	40	**Pos**	2	**Pos**	51	**Neg**
27	GeoW	Demand	Low	**Unk**	49	**Neg**	29	**Neg**	49	**Neg**
28	PctTrks	Flow	Mod	**Neg**	38	**Pos**	40	**Pos**	44	**Pos**
29	PctOldYng	Flow	Low	**Neg**	44	**Pos**	34	**Pos**	12	**Pos**
30	Nodes_UpNetMi	Flow	Mod	Neg	26	Neg	9	Neg	21	Neg
31	PctPrPvmt	Flow	Low	Neg	24	Neg	26	Neg	19	Neg
32	Crashes_Kcap	Flow	Low	**Neg**	37	**Pos**	47	**Pos**	37	**Pos**
33	YrPrecipIn	Flow	Low	**Neg**	50	**Neg**	38	**Pos**	50	**Neg**
34	SpTms_Mcap	Flow	Low	**Neg**	27	**Pos**	19	**Pos**	25	**Pos**
35	GPop_NetMi	Spread	Low	Neg	45	Neg	30	Neg	40	Neg
36	GWkr_UpNetMi	Spread	Low	Neg	3	Neg	11	Neg	9	Neg
37	GVeh_HH	Spread	Low	Neg	28	Neg	41	Neg	23	Neg
38	GMedInc_HH	Spread	Low	**Neg**	4	**Neg**	48	**Pos**	16	**Neg**
39	GWkr_Cap	Spread	Low	**Neg**	47	**Pos**	51	**Neg**	52	**Neg**
40	GJobs_SqMi	Spread	Mod	Neg	6	Neg	31	Neg	39	Neg
41	GJobs_HH	Spread	Mod	Neg	18	Neg	8	Neg	15	Neg
42	GJobs_Wkr	Spread	Mod	Neg	23	Neg	14	Neg	30	Neg
43	PctJobsJRDTcts	Other	Mod	**Neg**	52	**Neg**	36	**Pos**	45	**Pos**
44	PctPopJPTcts	Other	Mod	Neg	22	Neg	15	Neg	14	Neg
45	LeePoly	Other	Mod	**Pos**	9	**Neg**	17	**Neg**	31	**Neg**
46	Med_Mult	Other	Mod	Neg	30	Neg	7	Neg	29	Neg
47	PctGovtEmp	Other	Low	**Neg**	48	**Pos**	43	**Pos**	10	**Pos**
48	PctRetEmp	Other	Low	Pos	17	Pos	50	Pos	27	Pos
49	Pat_KWkrs	Other	Low	**Pos**	46	**Neg**	39	**Neg**	36	**Neg**
50	GDP_VMT	Other	Mod	**Neg**	1	**Neg**	45	**Pos**	38	**Neg**
51	UASqMi	Other	Mod	Neg	32	Neg	22	Neg	32	Neg
52	UAPop-K	Other	Mod	Neg	29	Neg	10	Neg	28	Neg

Notes: Values are shaded for variables unable to split at a 0.05 significance level.
All variable data are shaded when unable to split at a 0.05 significance level in all areas.
Differences in direction of effect are in bold.

The five "negative, but positive" cases are: Inflows_Wkr, PctSOV, PctOldYng, SpTms_Mcap, and PctGovtEmp. All of these are also "negative, but positive" cases in the PLS regression analysis above. Since the statistically significant relationships with the congestion variables for all five variables are either linear or decreasing, the reasoning for the disagreement in effect would be the same. Please refer to Section 5.3.3 for the discussion on these five variables.

The five "positive, but negative" cases: FwyLM_NetLM, FwyLM_KCmtr, Links_Node, VRM_SqMi, and LeePoly. Two of these (VRM_SqMi, and LeePoly) are also "positive, but negative" cases in the PLS regression analysis above and are discussed in that section. Both exhibit linear or decreasing relationships with the congestion variables, so there is little else to add. The remaining three are discussed below:

- FwyLM_NetLM – it seems reasonable that a larger freeway system in relation to the entire network would be associated with lower congestion. According to the trend line data, however, the effect is negative: relatively more freeways are linked with greater congestion. These results are the same for all three dimensions of congestion, although only for the duration dimension are the results significant. Still one wonders why. Ever larger freeway/network ratios are probably associated with the larger urban areas where congestion is already problematic, which suggests that this effect is confounded by other urban characteristics.
- FwyLM_KCmtr – one would expect that more lane miles of freeway for each commuter would translate into lower congestion. Instead, the trend line data show the opposite. Worse congestion accompanies a rise in the freeway lane-mile/commuter ratio. This is likely a result of the inverted U relationship between this ratio and the intensity and duration dimensions of congestion. (The relationship with extent is linear.) After an initial rise in congestion levels, which may be a manifestation of supply lagging demand, there reaches a point where the demand is met and continual increases in supply begins to

improve the congestion situation. If this point is far enough along the curve, the overall trend line would show a negative effect.

- Links_Node – more intra-connectivity would reasonably be associated with lower congestion levels (a positive effect), but the trend line data show a negative effect. The relationship with two of the congestion dimensions (extent and duration), however, is an inverted U; the other one is linear. The inverted U allows an increase in the links per node initially to accompany a worsening of congestion before congestion begins to improve. This could be the effect of the supply of intra-connecting roads lagging their demand early on before demand is satisfied. It could also be that increased intra-connectivity does not affect congestion until a threshold is achieved (although congestion worsens due to other factors), after which increased intra-connectivity begins to have an increasing effect.

The one "unknown, but positive" case is one of the geographic dummy variables, GeoMW, which has a positive revealed effect on intensity and extent but a negative revealed effect on duration. However, only the positive effect on extent is based on a statistically significant split. Apparently, an urban area in the Mid-west census region is less likely to be congested than in other regions, especially with regard to the extent of the system congested. Whether this is due to regional attitudes, city size or some other factor is unclear. This is the only geographic dummy variable to have a statistically significant first split and then only in the one dimension. Of the other geographic dummy variables, only GeoW showed results consistent in each dimension. To be in the West is unfavorable as far as congestion is concerned.

5.6 Analysis Results by Congestion Dimension

The results above show that the results from the various analyses differ depending on the congestion dimension. With this in mind, the results are now displayed by dimension. Variable size effects come from the correlation and PLS regression analyses, while the direction effects come from the correlation, PLS regression and the CHAID 1st split analyses. The relationship effect comes from just the CHAID 1st split analysis. In cases where there is disagreement in the size or direction of the effects, the results from the various methods are compared, the importance of the results considered, and an effect that best represents the variable is estimated. Variables are grouped by importance into four categories: important (those that are important in all four analyses (correlations, PLS regression, CHAID tree and CHAID 1st split)); somewhat important (those important in three analyses); somewhat unimportant (those important in one or two analyses); and, unimportant (those that are not important in any analysis). Data cells for variables that are somewhat unimportant or unimportant are shaded; variable name cells are shaded when data cells in all dimensions are shaded.

Table 28: Effects and importance by congestion dimension

		Intensity (TTI)				**Extent (PortLMCong)**				**Duration (PkHrs)**			
	Variable Code	**Size**	**Dir**	**Rel**	**Imp**	**Size**	**Dir**	**Rel**	**Imp**	**Size**	**Dir**	**Rel**	**Imp**
	Supply Variables												
1	PctPopCh	Low	Pos	L	SU	Low	Neg	L	SU	Low	Pos	L	SU
2	Rep-Dem	Low	Neg	L	SU	Low	Neg	U	SU	Low	Neg	L	SI
3	NetMi_SqMi	Low	Pos	InvU	U	Mod	Pos	L	U	Low	Pos	InvU	U
4	FwyMi_SqMi	Low	Pos	L	SU	Low	Pos	L	SI	Low	Pos	L	U
5	FwyLM_NetLM	Low	Neg	L	SU	Low	Pos	L	U	Low	Neg	L	SU
6	FwyLM_KCmtr	Mod	Pos	InvU	SI	Mod	Pos	L	I	Mod	Pos	InvU	SU
7	FwyArtMi_Cap	Mod	Pos	L	I	High	Pos	L	SI	Mod	Pos	NL	I
8	DecBeforeNow	Low	Neg	U	I	Low	Pos	InvU	SU	Mod	Neg	NL	SI
9	Links_Node	Low	Neg	L	U	Low	Neg	InvU	SU	Low	Pos	InvU	SU
	Demand Variables												
10	Cmtr_SqMi	Low	Neg	L	SU	Low	Neg	L	SU	Low	Neg	L	SI
11	Pers_SqMi	Low	Neg	NL	SI	Low	Neg	L	I	Low	Neg	Incr	SI

		Intensity (TTI)				Extent (PortLMCong)				Duration (PkHrs)			
	Variable Code	**Size**	**Dir**	**Rel**	**Imp**	**Size**	**Dir**	**Rel**	**Imp**	**Size**	**Dir**	**Rel**	**Imp**
12	Veh_HH	Low	Pos	L	U	Low	Neg	L	U	Low	Neg	NL	U
13	Inc_Cap	Mod	Neg	NL	SI	Low	Neg	L	SI	Mod	Neg	L	SI
14	Empl_Cap	Mod	Neg	L	SI	Mod	Pos	InvU	SU	Mod	Neg	Incr	I
15	Pers_Rest	Low	Pos	L	SU	Low	Pos	L	SU	Low	Pos	L	SU
16	Inflows_Wkr	Low	Pos	L	U	Low	Pos	L	SU	Low	Pos	L	SU
17	AvgCmtTime	High	Neg	Decr	I	High	Neg	L	I	High	Neg	Incr	SI
18	PctSOV	Low	Pos	Decr	I	Low	Pos	L	SI	Low	Pos	L	SI
19	VRM_SqMi	Low	Neg	Decr	SI	Low	Neg	L	I	Low	Neg	L	SI
20	PopSm	Low	Pos	L	SI	Low	Neg	L	SI	Mod	Pos	L	I
21	PopMed	Low	Pos	L	I	Low	Pos	L	I	Low	Pos	L	I
22	PopLg	Mod	Neg	L	SI	Mod	Neg	L	SU	Mod	Neg	L	I
23	PopVLg	Mod	Neg	L	I	Mod	Neg	L	SI	Mod	Neg	L	SI
24	GeoNE	Low	Neg	L	U	Low	Pos	L	U	Low	Neg	L	U
25	GeoS	Low	Neg	L	U	Low	Neg	L	U	Mod	Pos	L	U
26	GeoMW	Low	Pos	L	SU	Low	Pos	L	SU	Low	Neg	L	U
27	GeoW	Low	Neg	L	SU	Low	Neg	L	U	Low	Neg	L	U
Flow Variables													
28	PctTrks	Low	Pos	L	U	Low	Pos	L	U	Low	Pos	L	U
29	PctOldYng	Mod	Pos	L	SU	Mod	Pos	L	SU	Mod	Pos	L	SI
30	Nodes_UpNetMi	Mod	Neg	L	I	Mod	Neg	L	I	Mod	Neg	L	SI
31	PctPrPvmt	Low	Neg	L	SI	Low	Neg	L	SU	Low	Neg	L	I
32	Crashes_Kcap	Low	Pos	L	U	Low	Pos	L	U	Low	Pos	L	U
33	YrPrecipIn	Low	Neg	L	U	Low	Pos	L	U	Low	Neg	L	U
34	SpTms_Mcap	Low	Pos	L	SI	Mod	Pos	L	SI	Low	Pos	L	SI
Spread Variables													
35	GPop_NetMi	Low	Neg	L	SU	Low	Neg	L	U	Low	Neg	L	SU
36	GWkr_UpNetMi	Mod	Neg	L	SI	Low	Neg	L	SI	Mod	Neg	L	SI
37	GVeh_HH	Low	Neg	L	SI	Low	Neg	L	SU	Low	Neg	L	I
38	GMedInc_HH	Low	Neg	L	SI	Low	Pos	InvU	U	Low	Neg	L	SI
39	GWkr_Cap	Low	Neg	L	U	Low	Neg	U	SU	Low	Neg	NL	SU
40	GJobs_SqMi	Low	Neg	L	SU	Low	Pos	L	U	Low	Neg	L	SU
41	GJobs_HH	Low	Neg	L	I	Low	Neg	L	SI	Low	Neg	L	SI
42	GJobs_Wkr	Low	Neg	L	SI	Low	Neg	L	SI	Low	Neg	L	SI
Other Variables													
43	PctJobsJRDTcts	Low	Neg	L	U	Low	Pos	InvU	U	Low	Pos	L	U
44	PctPopJPTcts	Mod	Neg	L	I	Mod	Neg	L	SI	Mod	Neg	L	SI
45	LeePoly	Mod	Neg	L	SI	Low	Neg	L	I	Mod	Neg	L	I
46	Med_Mult	Low	Neg	L	SI	Low	Neg	L	SI	Low	Neg	L	SI
47	PctGovtEmp	Low	Pos	L	U	Low	Pos	L	U	Low	Pos	L	SU
48	PctRetEmp	Low	Pos	L	SI	Low	Neg	InvU	SU	Low	Pos	L	SI
49	Pat_KWkrs	Low	Neg	L	U	Low	Neg	L	U	Low	Neg	L	SU
50	GDP_VMT	Low	Neg	L	I	Low	Pos	InvU	SU	Low	Neg	L	SU
51	UASqMi	High	Neg	L	I	Mod	Neg	L	SI	High	Neg	L	SI
52	UAPop-K	Mod	Neg	L	SI	Mod	Neg	L	I	Mod	Neg	L	SI

There are 11 variables that are important in their relationship with congestion intensity (TTI) and 17 more that are somewhat important. This leaves 24 that are not important. Not surprisingly, the important variables include three urban area size variables (UASqMi, PopVLg and PopMed). Note that there are four demand variables among those rated as important and another four variables (three other and one spread) that may well be linked with demand. In intensity considerations, demand-related variables seem perhaps more influential. There are also eight cases where the relationships between the variables are not linear, all for supply or demand variables.

There are eight variables that are important and another 13 that are somewhat important in their relationship with congestion extent (PortLMCong). Over half the variables considered (31 of 52) are either somewhat unimportant or unimportant. Again size variables are well-represented in the top of the table, and again four demand variables are in the top category, along with two other variables that could be linked with demand. Nine variables, in all focus categories except flow, have relationships with the PortLMCong that are not linear.

Finally, there are eight variables that are important in their relationship with congestion duration (PkHrs) (three of which are urban size related), 21 that are somewhat important, and 23 that are somewhat unimportant or unimportant. In duration, as in intensity and extent above, it seems that demand variables dominate the top category – four of the top eight are demand and one other and one spread variable are also demand-related. Note that there are 16 cases where there is disagreement in effect size and 11 cases in effect direction. There are ten cases where the relationships between the variables are not linear; nine of these are for supply or demand variables.

5.7 Variable Importance

As the main goal for this research is to identify those urban characteristics that have the most important relationships with congestion, it is worth summarizing variable importance from Table 28 above. Table 29 shows that there are 19 variables that are important or somewhat important in all three dimensions of congestion, 20 that are not important in all three dimensions, and 13 that are of varying importance among dimensions. These will be discussed further in the section below. Note that there are 24 cases where the size of the effect is not as expected (shaded in gray) and 14 cases where the direction of the effect is not as expected (in bold font). In all cases but one, the size difference was just one gradation up or down; for PctPopCh, the difference was two gradations.

Table 29: Variable importance summarized

Focus	**Important (Size/Direction of Effect)**	**Of Varying Importance (Size/Direction of Effect)**	**Not Important (Size/Direction of Effect)**
Supply	FwyArtMi_Cap (High/Pos)	Rep-Dem (Low/Neg)	PctPopCh (Low/Neg)
		FwyMi_SqMi (Mod/Pos)	NetMi_SqMi (Mod/Pos)
		FwyLM_KCmtr (Mod/Pos)	FwyLM_NetLM (Low/Pos)
		DecBeforeNow (Low/Neg)	**Links_Node (Low/Neg)**
Demand	Pers_SqMi (Low/Neg)	Cmtr_SqMi (Low/Neg)	Veh_HH (Low/Neg)
	Inc_Cap (Mod/Neg)	Empl_Cap (Mod/Neg)	Pers_Rest (Low/Pos)
	AvgCmtTime (High/Neg)	PopLg (High/Neg)	**Inflows_Wkr (Low/Pos)**
	PctSOV (Low/Pos)		GeoNE (Base variable)
	VRM_SqMi (Low/Neg)		GeoS (Mod/Neg)
	PopSm (Base variable)		GeoMW (Low/Pos)
	PopMed (Low/Pos)		GeoW (Low/Pos)
	PopVLg (Mod/Neg)		
Flow	Nodes_UpNetMi (Mod/Neg)	**PctOldYng (Mod/Pos)**	**PctTrks (Low/Pos)**
	SpTms_Mcap (Mod/Pos)	PctPrPvmt (Low/Neg)	Crashes_Kcap (Low/Neg)
			YrPrecipIn (Low/Neg)
Spread	GWkr_UpNetMi (Mod/Neg)	GVeh_HH (Low/Neg)	**GPop_NetMi (Low/Pos)**
	GJobs_HH (Low/Neg)	GMedInc_HH (Low/Neg)	GWkr_Cap (Low/Neg)
	GJobs_Wkr (Low/Neg)		GJobs_SqMi (Low/Neg)
Other	PctPopJPTcts (Mod/Neg)	PctRetEmp (Low/Pos)	**PctJobsJRDTcts (Low/Pos)**
	LeePoly (Mod/Neg)	**GDP_VMT (Low/Pos)**	**PctGovtEmp (Low/Pos)**
	Med_Mult (Low/Neg)		**Pat_KWkrs (Low/Neg)**
	UASqMi (High/Neg)		
	UAPop-K (Mod/Neg)		

5.8 Analysis Results vs. Variable Selection

The overall results from above are now superimposed on the variable selection tables in Chapter 4 and discussed to "close the loop". Variable importance is shown by shading (no shading for important in all dimensions, dark gray for not important in any dimension, and light gray for important in some dimensions but not in others). Variable codes are added for ease of reference, revealed effects are noted under expected effects, and variables with possible linearity problems are indicated with an asterisk. As noted above, there are a number of cases where the expected effects and the revealed effects are different. Most of these are effect size-related and only one of which, PctPopCh, is more than one gradation in difference (expected – high, revealed – low); and this variable is unimportant in all congestion dimension. The differences in directions of effect are of more concern and, although discussed above in the specific analysis results sections above, they may merit additional comment in the discussion below.

Table 30: Overall results for variables impacting supply

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	**OPERATIONAL**			
Network Size	Percent change in population 2000 to 2010 (PctPopCh)	High/Negative Revealed **Low**/Negative	UMR	Measure of inadequacy of network size -Structural functionalism -Lag-time concept
	Political party control in 2000 (political affiliation of mayor)* (Rep-Dem)	Low/Negative Revealed Low/Negative	City Records World-statesmen website	Measure of transportation investments -Structural functionalism -Political party trends
Network Density	Network miles per square mile* (NetMi_SqMi)	Mod/Positive Revealed Mod/Positive	FHWA Census	Measure of network ability to accommodate demand -Structural functionalism
	Freeway miles per square mile (FwyMi_SqMi)	Mod/Positive Revealed Mod/Positive	FHWA Census	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
Network Structure	Freeway lane miles per network lane mile (FwyLM_NetLM)	Mod/Positive Revealed **Low**/Positive	UMR FHWA	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
Network Robustness	Freeway lane miles per thousand commuters* (FwyLM_KCmtr)	High/Positive Revealed **Mod**/Positive	UMR	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
	Freeway miles + arterial miles per capita* (FwyArtMi_Cap)	Mod/ Positive Revealed **High**/Positive	FHWA UMR	Measure of network ability to accommodate demand -Structural functionalism -Land rent theory
	City Age (Census urban area reached 50k in population (decades before 2010))* (DecBeforeNow)	Low/Negative Revealed Low/Negative	Census Wikipedia city pages	Measure of network ability to accommodate demand -Structural functionalism -Changing urban needs over time
Network intra-connectivity	Network nodes / Network links* (Links_Node)	Mod/Positive Revealed **Low**/**Negative**	TransCAD Census	Measure of available alternate routes -Structural functionalism -Packet-switching network theory

Note: Shading is based on variable importance: none - important in all dimensions, light gray - important in some dimensions but not in others, and dark gray - not important in any dimension.
* Possible non-linear relationships with one or more dependent variables.

Of the nine supply variables, only one is important in all congestion dimensions; four are not important in any dimension and four are mixed in their importance. The important variable, FwyArtMi_Cap, is a measure of network robustness, which seems to be the more influential category of supply variables. The variables used to assess the size, density, structure and intra-connectivity of the network are either not important in all dimensions or of mixed importance. One variable has a counter-intuitive direction of effects, Links_Node, but that variable is unimportant in all dimensions and this result is somewhat suspect. There are six variables with possible non-linear relationships with one or more of the dependent variables and one of these is the important variable, FwyArtMi_Cap. Care must be taken when interpreting the revealed effects for these variables. All in all, supply variables seem relatively less important in their links with congestion than variables in other categories.

Table 31: Overall results for variables impacting demand

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	**OPERATIONAL**			
Internal Productions	Commuters per square mile (Cmtr_SqMi)	Mod/Negative Revealed **Low**/Negative	UMR Census	Measure of commuter productions -Social exchange theory -4-Step Urban Travel Demand Model
	Persons per square mile* (Pers_SqMi)	Mod/Negative Revealed **Low**/Negative	UMR Census	Measure of total productions -Social exchange theory -4-Step Urban Travel Demand Model
	Cars per household* (Veh_HH)	Mod/Negative Revealed **Low**/Negative	Census ACS12-1	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
	Income per capita* (Inc_Cap)	Mod/Negative Revealed Mod/Negative	Census ACS12-1	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
Internal Attractions	Employment per capita* (Empl_Cap)	Mod/Negative Revealed Mod/Negative	Census ACS12-1	Measure of commuter attractions -Transportation demand is derived -Rational choice theory -Various sociological theories
	Persons per restaurant (Pers_Rest)	Mod/Positive Revealed **Low**/Positive	US Census Economic Census 2007	Measure of other attractions -Transportation demand is derived -Rational choice theory -Various sociological theories
External Productions	In-commuting flows per worker (Jobs in UA tracts - Workers in UA tracts) (Inflows_Wkr)	Low/Negative Revealed Low/**Positive**	CTPP 5-Year ACS 2006-2010	Measure of external productions -Social exchange theory -Land rent theory -Transportation demand is derived
Trip Distribution	Average commuting time in minutes* (AvgCmtTime)	Mod/Negative Revealed **High**/Negative	Census ACS12-1	Measure of time on network -Rational choice theory -Land rent theory
Mode Split	Percent of commuters in single occupant vehicles (SOV)* (PctSOV)	Low/Negative Revealed Low/**Positive**	Census ACS12-1	Effects of decreasing highway demand -Rational choice theory -Various sociological theories
	Transit vehicle revenue miles per square mile* (VRM_SqMi)	Low/Positive Revealed Low/**Negative**	NTD Census	Effects of decreasing highway demand -Rational choice theory -Various sociological theories

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
Variations in Demand between Urban Areas	Dummy variables based on city size (population) (PopSm) (PopMed) (PopLg) (PopVLg)	Mod/Negative Revealed (Base variable) **Low**/**Positive** **High**/Negative Mod/Negative	UMR	Measure of variations in demand -Rational choice theory -Various sociological theories
	Dummy variables based on geography (GeoNE) (GeoS) (GeoMW) (GeoW)	Low/Unknown Revealed (Base variable) **Mod**/**Negative** Low/**Positive** Low/**Positive**	Census	Measure of variations in demand -Rational choice theory -Various sociological theories

Note: Shading is based on variable importance: none - important in all dimensions, light gray - important in some dimensions but not in others, and dark gray - not important in any dimension.
* Possible non-linear relationships with one or more dependent variables.

There are 18 demand variables, eight of which are dummies associated with either population or geography. Of these 18, eight are of importance in all dimensions, three of mixed importance and seven of no importance. The important variables (Pers_SqMi, Inc_Cap, AvgCmtTime, PctSOV, VRM_SqMi, and all population dummies except PopLg) are largely focused on the trip production part of the four step travel modeling process, assuming that the population variables are drivers of productions more so than attractions. The mode split variables also figure prominently, but the revealed effects are in the opposite direction as expected, so there may be some confounding interactive effects at play in these two cases. Interestingly, the unimportant variables include Veh_HH, which although a key driver in trip productions in the four-step model, appears less useful in differentiating between urban areas based on congestion levels. Also unimportant are the geographical locations of the cities, the net traffic inflows and the number of attractions (as measured by Pers_Rest). It seems likely that these too, are similar enough cross the cities as to be less useful in differentiation. There are seven variables with unexpected directions of effect, although three of these are the geographic dummy variables, where the expected direction is unknown, so the difference is less an unexpected effect than just new knowledge. Another unexpected direction is for one of the population dummies (PopMed) and this results from the interpretation of this variable in relation to the base variable more so than in relation to congestion itself. The other discrepancies in direction seem valid and are discussed in detail above. Seven variables show evidence of non-linear relationships with at least one dimension of congestion, and care should be exercised in interpreting the revealed effects. Demand variables, as a whole, seem to be very important in their relationships with congestion.

Table 32: Overall results for variables impacting flow

VARIABLE		**EXPECTED EFFECT**	**SOURCE**	**JUSTIFICATION**
CONCEPTUAL	**OPERATIONAL**			
Trucks	Percent of trucks on freeways (PctTrks)	Mod/ Negative Revealed **Low/Positive**	FHWA	Measure of truck impact on flow -Differences in truck-car nimbleness
Distracted Driving	Percent of population 16-24 plus percent of population 65 and over (PctOldYng)	Low/ Negative Revealed **Mod/Positive**	Census ACS12-1	Measure of flow interruptions due to distracted drivers -Consequences of human interaction -Loss aversion
Intersections with traffic signals and stop-controlled signage	Nodes per network mile (upper level system only) (Nodes_UpNetMi)	Mod/ Negative Revealed Mod/Negative	TransCAD Census	Measure of flow interruptions due to signals/signage -Queuing theory
Road Condition	Pavement condition (percent in poor condition) (PctPrPvmt)	Low/ Negative Revealed Low/Negative	TRIP Urban Roads Report	Measure of decreases in flow caused by lower speeds due to poor pavement -Human nature and driving skills
Traffic Incidents	Accident rate x VMT per capita (Crashes_Kcap)	Low/ Negative Revealed Low/Negative	NHTSA FHWA UMR	Measure of flow interruptions due to traffic accidents -Consequences of human interaction -Loss aversion
Weather	Annual precipitation (YrPrecipIn)	Low/ Negative Revealed Low/Negative	NCDC	Measure of flow interruptions due to bad weather -Mother nature and geography

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
Special Events	Number of upper level sports teams (SpTms_Mcap)	Low/ Negative Revealed **Mod/Positive**	Various Internet Websites	Measure of flow interruptions due to special events -Structural functionalism -Social exchange theory

Note: Shading is based on variable importance: none - important in all dimensions, light gray - important in some dimensions but not in others, and dark gray - not important in any dimension.

On the other hand, flow variables appear to be less important in their relationships with congestion. Just two of the seven are important in all dimensions and one of these has an unexpected direction of effect, which may indicate the presence of a confounding variable. The other important variable, Nodes_UpNetMi, seems to be a clean and expected link with congestion, although it may have limited use in remediation efforts (assuming that the variable is causal); many of the existing intersections are needed to allow network access and alternatives to intersections can often be prohibitive in cost. Three flow variables are unimportant and the remaining two are of mixed importance. The two other variables with unexpected directions of effects are of limited or no importance and have been discussed already. Interestingly, all flow variables appear to have linear relationships with all three congestion variables.

Table 33: Overall results for measures of spread across urban area census tracts

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
Network Layout	Gini coefficient of population per network mile (GPop_NetMi)	Low/Negative Revealed Low/**Positive**	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of network layout efficiency -Structural functionalism -Changing urban needs over time
	Gini coefficient of workers per upper network mile (GWkr_UpNetMi)	Low/Negative Revealed **Mod**/Negative	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of network layout efficiency -Structural functionalism -Changing urban needs over time
Internal productions	Gini coefficient of car ownership (aggregate vehicles per HH) (GVeh_HH)	Low/Negative Revealed Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	OPERATIONAL			
	Gini coefficient for median income per HH* (GMedInc_HH)	Low/Negative Revealed Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of trip productions -Social exchange theory -4-Step Urban Travel Demand Model
	Gini coefficient of workers per capita* (GWkr_Cap)	Low/Negative Revealed Low/Negative	CTPP 5-Year ACS 2006-2010	Measure of time on network -Level of mixed land use -Land rent theory
Urban Spatial Structure	Gini coefficient of employment (jobs) density (GJobs_SqMi)	Mod/Negative Revealed **Low**/Negative	CTPP 5-Year ACS 2006-2010	Measure of degree of monocentricity -Central place theory -Land rent theory
	Gini coefficient for jobs/HH balance (GJobs_HH)	Mod/Negative Revealed **Low**/Negative	CTPP 5-Year ACS 2006-2010	Measure of degree of monocentricity -Central place theory -Land rent theory
	Gini coefficient of jobs/worker balance (GJobs_Wkr)	Mod/Negative Revealed **Low**/Negative	CTPP 5-Year ACS 2006-2010	Measure of degree of monocentricity -Central place theory -Land rent theory

Note: Shading is based on variable importance: none - important in all dimensions, light gray - important in some dimensions but not in others, and dark gray - not important in any dimension.
* Possible non-linear relationships with one or more dependent variables.

Three spread variables are important in all dimensions of congestion: GWkr_UpNetMi, GJobs_HH, and GJobs_Wkr. The first involves network layout, while the last two involve urban spatial structure. All three, however, involve the degree of equal distribution of the workforce within the urban footprint, and all three have the expected effect – an increase in the inequality of the distribution is linked with worse congestion. The other five spread variables are either of mixed importance (two) or not important (three). Only one has an unexpected direction of effect (and that variable is an unimportant one, so it is unlikely to be of consequence) and two have potential non-linear relationships with one or more of the dependent variables (and these are of mixed or no importance). On the whole, most spread variables are not so useful in this assessment and the time and effort involved in their calculation may make them “not worth the effort.”

Table 34: Overall results for other variables potentially impacting congestion

VARIABLE		EXPECTED EFFECT	SOURCE	JUSTIFICATION
CONCEPTUAL	**OPERATIONAL**			
Centrality	Percent of Employment in Job-Rich, Job-Dense Tracts* (PctJobsJRDTcts)	Mod/Negative Revealed **Low**/**Positive**	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of the spread of employment -Central place theory -Land rent theory
Sprawl	Percent of Population in Job-Poor Tracts (PctPopJPTcts)	Mod/Negative Revealed Mod/Negative	CTPP 5-Year ACS 2006-2010 TransCAD	Measure of the spread of population -Central place theory -Land rent theory
Urban Spatial Structure	Degree of poly-centricity (higher more poly-centric) (LeePoly)	Mod/Positive Revealed Mod/**Negative**	Lee and Gordon (2007)	Measure of degree of monocentricity -Central place theory -Land rent theory
Land Costs	Housing affordability (Med_Mult)	Mod/Negative Revealed **Low**/Negative	Int'l Housing Afford. Survey	Measure of the bid-rent function -Land rent theory
Government Employment	Percent of employees working for government (PctGovtEmp)	Low/Negative Revealed Low/**Positive**	Census ACS12-1	Measure of private-public employment split -Degree of peak hour participation
8-hour work day	Percent of employment in retail* (PctRetEmp)	Low/Positive Revealed Low/Positive	Census ACS12-1	Measure of employees not working a standard 8-hour day -Rational choice theory -Degree of peak hour participation
Creativity	Patents per 1000 workers (Pat_KWkrs)	Low/Positive Revealed Low/**Negative**	Brookings Institute	Measure of participation in the status quo -Transportation demand is derived -Rational choice theory -Various sociological theories
Activity Density	Real GDP per VMT* (GDP_VMT)	Mod/Negative Revealed **Low**/**Positive**	BEA UMR FHWA	Measure of city density of activity -Structural functionalism -Social exchange theory
Size	Urban area square miles (UASqMi)	Mod/Negative Revealed **High**/Negative	Census	Measure of city size -Rational choice theory -Various sociological theories
Size	Urban area population (UAPop-K)	Mod/Negative Revealed Mod/Negative	Census	Measure of city size -Rational choice theory -Various sociological theories

Note: Shading is based on variable importance: none - important in all dimensions, light gray - important in some dimensions but not in others, and dark gray - not important in any dimension.

* Possible non-linear relationships with one or more dependent variables.

There are ten other variables that are considered to “round out the field”. Of these, five are important in all dimensions of congestion, three are unimportant, and two are of mixed importance. Two of the five important variables are size-related (UASqMi and UAPop-K) and three generally deal with the overall urban layout: sprawl (PctPopJPTcts), poly-centricity (LeePoly), and land costs (Med_Mult). While LeePoly has an unexpected direction of effect (which could indicate the interactive effects of other confounding variables), all other important variables behave about as expected. The issue of centrality, the level of government employment and the creative nature of the workforce are not important, at least as far as the variables used as metrics are concerned. Of some interest, but of less importance are the level of retail employment and the amount of money “available” for each vehicle mile of travel. There are three variables with possible non-linear relationships with the congestion variables, but these are unimportant or of mixed importance. All in all, these other “round-out” variables were a valuable addition to the analysis.

CHAPTER 6: CONCLUSIONS

6.1 Summary

This exploratory research sought to identify the set of urban characteristics that are correlated with traffic congestion. After a review of the literature concerning congestion and urban areas, with a focus on theories and concepts, models, and urban structure, three dependent congestion variables representing the three dimensions of congestion (intensity, extent and duration) and 52 potential predictor variables were identified for 100 urban areas in the United States, using 2010 data predominantly. Variables were analyzed using multiple methods. Simple correlation, PLS regression, CHAID decision trees, and CHAID first split analysis results were all considered in determining the relationships between the predictor and response variables. Each method was used to uncover the influential variables for each dimension and then these method-based results were compared with one another to determine the variables' influence across all four methods. Of the 52 predictor variables, 19 were determined to be important (i.e. well-correlated) in all three dimensions of congestion, 20 were not important in any of the three dimensions, and 13 were important in some dimensions, but not in others. While in most cases the direction of effect was as expected, there were 17 instances where the effect was in the opposite direction, most likely due to the presence of interaction effects from confounding variables. In most cases, the revealed effects suggested a linear relationship with the dependent variable; however, there were 18 cases where the relationship was possibly non-linear.

Of the 19 variables that were most correlated with congestion, one was supply-focused (FwyArtMi_Cap), eight were demand-focused (Pers_SqMi, Inc_Cap, AvgCmtTime, PctSOV, VRM_SqMi, PopSm, PopMed, and PopVLg), two were flow-focused (Nodes_UpNetMi and SpTms_Mcap), three were spread-focused (GWkr_UpNetMi, GJobs_HH, and GJobs_Wkr), and five were in the "other" category where the focus was unclear or overlapping (PctPopJPTcts, LeePoly, Med_Mult, UASqMi, and UAPop-K). The revealed effects of congestion were in the expected direction for 15 of these 19; four, however, had counterintuitive revealed effects (PctSOV, VRM_SqMi, SpTms_Mcap and LeePoly).

6.2 Research Corroborated

Most of the congestion research heretofore has focused on congestion impacts, mechanics and remediation. While the first is relatively independent of causality concerns, the last two are built around cause-and-effect relationships. This research did not establish any causal connections, nor can they be inferred. Nonetheless, there are some links between this analysis and the previous literature that might be noted, especially in the congestion remediation area. Efforts to improve congestion levels through adding supply are reasonable, particularly if the focus is on both arterial and freeway capacity. Efforts to improve congestion levels by decreasing demand are also reasonable, although the results of some strategies (e.g. decreasing PctSOV or increasing VRM_SqMi) might not have the expected results and results may diminish as cities become large and very large. Flow-focused efforts to reduce congestion are reasonable, but perhaps to a lesser degree and might be better targeted to the upper level system. There are some congestion remediation proposals that are spread-focused, although land-use solutions are long-term propositions. Still, the data show that a more uniform distribution of jobs and housing is linked with lower congestion levels. Finally, there is support for the idea that lower levels of sprawl are correlated with lower levels of congestion.

6.3 Lessons Learned

Congestion is a most complex issue. Non-linear relationships, the presence of confounding, and perhaps unknown, variables, and the varying degrees of importance among reasonably formulated variables all serve to muddy the waters of understanding. Nonetheless, some broad lessons are learned, in addition to the specific findings noted in the summary. Size variables seem to be overly represented among those deemed important and population density seems to be influential. Certain measures of spread are also important. All of which point to the idea that size matters, at least as far as congestion goes. Another key lesson learned is that some variables just do not seem to matter and can be safely excluded from the congestion discussion. One final word of caution in the lessons learned: this assessment looked exclusively at correlative relationships. No causal inferences were determined or implied.

6.4 Future Research

As noted, this research is exploratory in purpose. There is no intent to explain congestion or identify any causal relationships. It would be extremely difficult, if not impossible, to “get at” the causality issue in a macro analysis such as this one. Human behavior, and traffic congestion certainly falls under this heading, is eminently adaptable and notoriously difficult to predict. While it is true that Sherlock Holmes noted in *The Sign of Four:*

> "While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will be up to, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician."
> ― Sir Arthur Conan Doyle

percentages may well change, especially over the long periods of time needed to implement congestion remedies. Hence, the percentages are likely to become increasingly difficult to predict. Nonetheless, efforts to understand congestion are not misplaced. Additional research using only those variables identified here as important, and especially those with likely linear relationships, is a reasonable "next step" forward. As causality issues are better studied at the micro-level, such research should be targeted at one or at a small group of cities, similar in size and geography. Also of potential benefit is research within a single congestion dimension. Most discussion of congestion in the literature seems to consider congestion as a whole, and possible approaches to more effective congestion remediation could be uncovered with a single dimension focus. Finally, there may be some merit to additional exploratory research conducted on a more micro level. It could be that some important relationships are lost in higher level analyses. It could also be that some of the counter-intuitive relationships found in this study no longer hold in a more localized assessment.

6.5 Final Thoughts

Larger cities, in terms of square mileage and population, tend to be accompanied by increased poly-centricity, a more variegated landscape, and a more sprawling footprint as people try to balance housing and jobs and employers try to adapt to the changing worker pool. A quick look at the urban areas in the study shows that it is the larger ones that have the most congestion issues. It may turn out that congestion is simply another one of the “amenities” (although negative in nature) offered by urban regions that the citizenry will need to embrace, or at least accept. As this amenity does not seem to be a major deterrent to city growth, it could be that Anthony Downs is right when he said: “Traffic congestion is not essentially a problem. It’s the solution to our basic mobility problem.” (Downs 2004, p. 20) Individuals can and do adapt.

REFERENCES

Aftabuzzaman, M. (2011) "Understanding the Road Traffic Congestion Relief Impacts of Public Transport." A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD), Institute of Transport Studies, Department of Civil Engineering, Monash University, Melbourne, Australia.

Aftabuzzaman, M., Currie, G., Sarvi, M. (2010) "Evaluating congestion relief impacts of public transport in monetary terms." *Journal of Public Transportation*, Vol. 13, No. 1, pp. 1-24.

Allison, P.D. (1999) *Multiple Regression – A Primer.* Thousand Oaks, CA: Pine Forge Press.

Aziz, H.M.A., Ukkusuri, S.V. (2012) "Unified Framework for Dynamic Traffic Assignment and Signal Control with Cell Transmission Model." *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2311, pp. 73–84.

Bairoch, P. (1988) *Cities and Economic Development: From the Dawn of History to the Present*. Chicago, IL: University of Chicago Press.

Baum-Snow, N. (2007) "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, Vol. 122, No. 2.

Balaker, T., Staley, S. (2006) *The Road More Traveled: Why the Congestion Crisis Matters More Than You Think, and What We Can Do About It.* Lanham, MD: Roman and Littlefield Publishers, Inc.

Beckx, C., Arentze, T., Int Panis, L., Janssens, D., Vankerkom, J., Wets, G. (2009) "An integrated activity-based modelling framework to assess vehicle emissions: approach and application". *Environment and Planning B: Planning and Design,* Vol. 36, No. 6, pp. 1086–1102.

Bertini, R.L. (2005) "Congestion and Its Extent." In: *Access to Destinations: Rethinking the Transportation Future of our Region*, Ed: D. Levinson and K. Krizek, Elsevier.

Bigss, D., Ville, B., Suen, E. (1991) "A method of choosing multiway partitions for classification and decision trees." *Journal of Applied Statistics*, Vol. 18, No. 1, pp. 49-62.

Boarnet, M.G. (2008) "Transportation Infrastructure and Sustainable Development: New Planning Approaches for Urban Growth." *Access*, Vol. 33 (Fall), pp. 27-33.

Boarnet, M.G., Crane, R. (2001) *Travel by Design: The Influence of Urban Form on Travel.* New York, NY: Oxford University Press.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1998) *Classification and Regression Trees*. Boca Raton, FL: Chapman and Hall/CRC.

Buliung, R.N., Kanaroglou, P.S. (2006) "Urban Form and Household Activity-Travel Behavior." *Growth and Change*, Vol. 37, No. 2, pp. 172-199.

Cambridge Systematics, Inc. (2005) *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. A report prepared for the Federal Highway Administration in September.

Census Bureau. (2010) "2010 Census Urban and Rural Classification and Urban Area Criteria." Accessed February 10, 2013 at: http://www.census.gov/geo/reference/urban-rural-2010.html.

Cervero, R. (2003a) "Are Induced-Travel Studies Inducing Bad Investments?" *Access*, Vol. 22 (Spring), pp. 22-27.

Cervero, R. (2003b) "Road Expansion, Urban Growth, and Induced Travel: A Path Analysis," *Journal of the American Planning Association*, Vol. 69, No. 1, pp. 145–163.

Cervero, R. (2001) "Efficient urbanization: economic performances and the shape of the metropolis." *Urban Studies*, Vol. 38, No. 10, pp. 1651-1671.

Chin, S.M., Franzese, O., Greene, D.L., Hwang, H.L., Gibson, R.C. (2002a) *Temporary Losses of Highway Capacity and Impacts on Performance*. Oak Ridge National Laboratory Report ORNL/TM-2002/3.

Chin, S.M., Franzese, O., Greene, D.L., Hwang, H.L., Gibson, R.C. (2002b) *Temporary Losses of Highway Capacity and Impacts on Performance: Phase 2*. Oak Ridge National Laboratory Report ORNL/TM-2002/3.

Christaller, W. (1966) *Central Places in Southern Germany*. Translated by Carlisle Bascom in 1966. Englewood Cliffs, NJ: Prentice-Hall.

Chowdury, M., Derov, N., Tan, P., Sadek, A. (2005) "Prohibiting Left-Turn Movements at Mid-Block Unsignalized Driveways: Simulation Analysis." *Journal of Transportation Engineering*, Vol. 131, No. 4, pp. 279-285.

Cook, K. S., Rice, E. R. W. (2001) "Exchange and power: Issues of structure and agency." In Jonathan H. Turner (Ed.) *Handbook of Sociological Theory*. New York: Kluwer Academic/ Plenum Publishers.

Cox, W. (2013) "9th Annual Demographia International Housing Affordability Survey (2012: 3rd Quarter)." Accessed May 23, 2013 at: http://www.demographia.com/dhi.pdf.

Cox, W. (2011) “7th Annual Demographia International Housing Affordability Survey (2010: 3rd Quarter).” Accessed May 23, 2013 at: http://www.demographia.com/dhi2011.pdf.

Cox, W., Pisarski, A.E. (2004) *Blueprint 2030: Affordable Mobility and Access for all of Atlanta and Georgia*. Report to the Governor’s Mitigation Task Force. Atlanta, GA. Accessed May 23, 2013 at: http://ciprg.com/ul/gbt/atl-report-20040621.pdf.

Crampton, G.R. (2000) “Urban economic structure and the optimal rail system.” *Urban Studies*, Vol. 37, No. 3, pp. 623-632.

Crane, R., Chatman, D.G. (2003) “As Jobs Sprawl, Whither the Commute?” *Access*, Vol. 23 (Fall), pp. 14-19.

Dahl, J., Lee, C. (2012) “Empirical Estimation of Capacity for Roundabouts Using Adjusted Gap-Acceptance Parameters for Trucks.” *Transportation Research Record: Journal of the Transportation Research Board*, Issue 2312, pp. 34–45.

Demographia. (2013) “Definition of Urban Terms.” Accessed February 10, 2013 at: http://www.demographia.com/db-define.pdf.

Demographia. (2006) “New York Employment and Commuting.” Accessed February 6, 2014 at: http://www.demographia.com/db-nyc-employ.pdf.

Downs, A. (2004) “Why Traffic Congestion Is Here to Stay. . . and Will Get Worse.” *Access*, Vol. 25 (Fall), pp. 19-25.

Driedger, L. (1991) *The Urban Factor: Sociology of Canadian Cities*. Don Mills, Ontario: Oxford University Press.

Dumbaugh, E. (2012) “Rethinking the Economics of Traffic Congestion.” *Atlantic Cities*, 1 June.

Duranton, G., Turner, M.A. (2011) “The Fundamental Law of Road Congestion: Evidence from US Cities.” *American Economic Review*, Vol. 101 (October), pp. 2616–2652.

Federal Highway Administration (FHWA). (2013a) “Urban Congestion Trends.” Accessed April 24, 2013 at: http://www.ops.fhwa.dot.gov/publications/fhwahop13016/fhwahop13016.pdf.

Federal Highway Administration (FHWA). (2013b) “Traffic Congestion and Reliability: Linking Solutions to Problems.” Accessed April 4, 2013 at: http://www.ops.fhwa.dot.gov/congestion_report_04/chapter4.htm.

Federal Highway Administration (FHWA). (2012a) “Operations Story: 21st Century Operations Using 21st Century Technologies.” Accessed April 5, 2012 at: http://ops.fhwa.dot.gov/aboutus/one_pagers/opstory.htm.

Federal Highway Administration (FHWA). (2012b) “Focus on Congestion Relief.” Accessed April 24, 2013 at: http://www.fhwa.dot.gov/congestion/describing_problem.htm.

Federal Highway Administration (FHWA). (2009) “Road Pricing: Resources.” Accessed April 24, 2013 at: http://www.fhwa.dot.gov/ipd/revenue/road_pricing/resources/webinars/webinar_090710.htm.

Federal Highway Administration (FHWA). (2008) “Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation.” Accessed April 24, 2013 at: http://ops.fhwa.dot.gov/congestion_report/chapter3.htm.

Fleming, S. (2007) “In Transportation and in Technology, Packets Beat Circuits.” Georgia Public Policy Foundation. Accessed April 24, 2013 at: http://www.georgiapolicy.org/in-transportation-and-in-technology-packets-beat-circuits/

Forkenbrock, D., Mathur, S., Schweitzer, L. (2001) *Transportation Investment Policy and Urban Land Use Patterns.* Public Policy Center, Iowa City: University of Iowa.

Gärling, T., Gillholm, R., Gärling, A. (1998) “Reintroducing attitude theory in travel behavior research: The validity of an interactive interview procedure to predict car use.” *Transportation*, Vol. 25, pp. 129–146.

Garson, G.D. (2014) *Partial Least Squares: Regression and Structural Equation Models.* Asheboro, NC: Statistical Publishing Associates.

Gayah, V.V. (2012) “Two-Way Street Networks: More Efficient than Previously Thought?” *Access*, Vol. 41, pp. 10-15.

Gill, J., Gain, J. (2002) “Why Does Voting Get So Complicated? A Review of Theories for Analyzing Democratic Participation.” *Statistical Science*, Vol. 17, No. 4, pp. 383-404.

Glaister, S., Graham, D.J. (2006) “Proper Pricing for Transport Infrastructure and the Case of Urban Road Congestion.” *Urban Studies*, Vol. 43, No. 8, pp. 1395-1418.

Gunay, B. (2009) “Using Wavelet Transforms for Better Interpretation of Traffic Simulation.” *Traffic Engineering & Control*, Vol. 50, No. 10, pp. 450-453.

Had2Know.com (2014) Gini Coefficient Calculator. Accessed January 6, 2014 at: http://www.had2know.com/academics/gini-coefficient-calculator.html.

Hansen, M. (1995) “Do New Highways Generate Traffic?” *Access,* No. 7 (Fall).

Hardin, G. (1968) “The Tragedy of the Commons.” *Science*, Vol. 162, pp. 1243-8.

Hartgen, D.T., Fields, M.G. (2009) *Gridlock and Growth: The Effect of Traffic Congestion on Regional Economic Performance*. Reason Foundation Policy Study 371. Los Angeles, CA: The Reason Foundation, available at: http://reason.org/news/show/gridlock-and-growth-the-effect.

Hartgen, D.T., Fields, M.G. (2006) *Building Roads to Reduce Traffic Congestion in America's Cities: How Much and at What Cost?* A Report for the Reason Foundation. Los Angeles, CA: The Reason Foundation, available at: http://reason.org/news/show/building-roads-to-reduce-traff-1.

Hartgen, D.T., Fields, M.G., Moore, A. (2011) *Impacts of Transportation Policies on Greenhouse Gas Emissions in U.S. Regions.* A Report for the Reason Foundation. Los Angeles, CA: The Reason Foundation, available at: http://reason.org/news/show/greenhouse-gas-policies-cost-transp.

Hatzopoulou, M., Miller, E. (2010) "Linking an activity-based travel demand model with traffic emission and dispersion models: Transport's contribution to air pollution in Toronto". *Transportation Research Part D,* Vol. 15, pp. 315–325.

Hensher, D.A. (2003) “Formulating an urban passenger transport policy: a reappraisal.” *Australian Economic Papers*, Vol. 18, No. 32, pp. 119-130.

Holden, C. (2008) “Hang Up and Get Moving.” *Science*, Vol. 319, No. 5860, pp. 138-139.

Horner, M.W. (1998) “Direct Household Assignment: An Exploratory Study of Simplifying Urban Travel Forecasts using GIS.” Master’s Thesis. Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC.

Jacobs, J. (1961) *The Death and Life of Great American Cities*. New York: Random House.

Jakobowicz, E. (2014) E-mail correspondence with Partner of Addinsoft (maker of XLSTAT) reference the meaning of VIP, June 6.

Kass, G.V. (1980) “An exploratory technique for investigating large quantities of categorical data.” *Applied Statistics*, Vol. 20, No. 2, pp. 119-127.

Kelly, B. (2012) “A ‘Green Wave’ Reprieve.” Traffic Engineering & Control, Vol. 53, No. 2, pp. 55-58.

Koch, J.W. (1995) *Social Reference Groups and Political Life*. Lanham, MD: University Press of America, Inc.

Kotkin, J. (2013) "Houston Rising – Why the Next Great American Cities Aren't What You Think." *The Daily Beast*, April 8.

Lee, B., Gordon, P. (2007) Urban Spatial Structure and Economic Growth in Us Metropolitan Areas. A paper presented at the 46th Annual Meeting of the Western Regional Science Association in Newport Beach, California, February. Accessed December 5, 2013, at: http://www-bcf.usc.edu/~pgordon/pdf/structure_growth_200701.pdf

Littman, T. (2013) "Congestion Costing Critique: Critical Evaluation of the 'Urban Mobility Report.'" Accessed June 25, 2013, at: http://www.vtpi.org/UMR_critique.pdf.

Littman, T. (2012) "Congestion Reduction Strategies: Identifying and Evaluating Strategies to Reduce Traffic Congestion." From the Victoria Transport Policy Institute's TDM Encyclopedia, updated 5 January 2012. Accessed April 8, 2013 at: http://www.vtpi.org/tdm/tdm96.htm.

Litman, T. (2001) "Generated Traffic; Implications for Transport Planning," *ITE Journal*, Vol. 71, No. 4, pp. 38-47.

Llewelyn-Davies (Banister, D., Hall, P.). (2004) Transport and City Competitiveness – Literature Review. A Report for the U.K. Department for Transportation.

Maitra, B., Sikdar, P.K., Dhingra, S.L. (1999) "Modeling Congestion on Urban Roads and Assessing Level of Service." *Journal of Transportation Engineering*, Vol. 125, No. 6 (November/December).

Maitra, S., Yan, J. (2008) "Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression." Casualty Actuarial Society, 2008 Discussion. Accessed May 22, 2014 at: http://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf

Mallinckrodt, J. (2010) "VCI, a Regional Volume/Capacity Index Model of Urban Congestion." *Journal of Transportation Engineering*, Vol. 136, No. 2 (February 1).

Martin, W., McGuckin, N. (1998) *Travel Estimation Techniques for Urban Planning*. NCHRP Report 365, Transportation Research Board. Washington, DC: National Academy Press.

McGroarty, J. (2010) "Recurring and Non-Recurring Congestion: Causes, Impacts, and Solutions." University of Cincinnati, Neihoff Urban Studio – W10 January 29. Accessed May 23, 2013 at:
http://www.uc.edu/cdc/niehoff_studio/programs/great_streets/w10/reports/recurring_non-recurring.pdf

Meijers, E.J., Burger, M.J. (2010) "Spatial structure and productivity in US metropolitan areas." *Environment and Planning A*, Vol. 42, pp. 1383-1402.

Merriam-Webster Dictionary. (1997) Springfield, MA: Merriam-Webster.

Neumann, J. (2005) "Kurt Lewin at the Tavistock Institute." *Educational Action Research*, Vol. 13, No. 1, pp.119-136

Newman, P., Kenworthy, J. (1999) Sustainability and Cities: Overcoming Automobile Dependence. Washington, DC: Island Press.

Ogunbodede, E.F. (2007) "Assessment of Traffic Congestions in Akure (Nigeria) Using GIS Approach: Lessons and Challenges for Urban Sustenance." Paper presented at the International Conference on Whole Life Urban Sustainability and its Assessment at Glasgow Caledonian University, Scotland in June.

Organization for Economic Cooperation and Development (OECD); European Conference of Ministers of Transport (ECMT) (2007) *Managing Urban Traffic Congestion*. Paris, France: OECD Publishing.

Ostrom, E. (1990) *Governing the Commons*. New York, NY: Cambridge University Press.

Prud'homme, R., Lee, C-W. (1999) "Size, sprawl, speed and the efficiency of cities." *Urban Studies*, Vol. 36, No. 11, pp. 1849-1858.

Ritzer, G. (1996) *Sociological Theory*, 4th Edition. New York, NY: McGraw Hill Companies, Inc.

Ratner, B. (2012) *Statistical and Machine Learning Data Mining*, 2d Edition. Boca Raton, FL: CRC Press.

Rothwell, J., Lobo, J., Strumsky, D., Muro, M. (2013) "Patenting Prosperity: Invention and Economic Performance in the United States and its Metropolitan Areas." Washington, DC: The Brookings Institution. Accessed January 16, 2014 at:
http://www.brookings.edu/research/reports/2013/02/patenting-prosperity-rothwell

Schrank, D., Eisele, B., Lomax, T. (2012) *TTI's 2012 Urban Mobility Report; Powered by INRIX Traffic Data*. Accessed May 22, 2014 at: http://mobility.tamu.edu/ums/report/.

Schrank, D., Eisele, B., Lomax, T. (2011) *TTI's 2011 Urban Mobility Report; Powered by INRIX Traffic Data*. Accessed May 22, 2014 at: http://mobility.tamu.edu/ums/report/.

Sanchez, G. (2013) "PLS Path Modeling with R." Trowchez Editions. Berkeley, 2013. Accessed May 2, 2014, at: http://www.gastonsanchez.com/PLS Path Modeling with R.pdf

Shen, W., Zhang, H.M. (2010) "Pareto-improving ramp metering strategies for reducing congestion in the morning commute." *Transportation Research Part A*, Vol. 44, pp. 676–696.

Shmueli, G. (2014) "The use of dummy variables in predictive algorithms." BzST, Business Analytics, Statistics, Technology. Accessed June 25, 2014 at: http://www.bzst.com/2014/03/the-use-of-dummy-variables-in.html.

Shmueli, G. (2007) "Classification Trees: CART vs. CHAID." BzST, Business Analytics, Statistics, Technology. Accessed July 6, 2014 at: http://www.bzst.com/2006/10/classification-trees-cart-vs-chaid.html.

Shmueli, G., Patel, N.R., Bruce, P.C. (2007) *Data Mining for Business Intelligence*. Hoboken, NJ: John Wiley & Sons, Inc.

Sivak, M. (2013) *Distance Driven and Economic Activity in the Individual U.S. States: 1997-2011*. A Report from the University of Michigan Transportation Research Institute, Ann Arbor, Michigan, available at: http://www.umich.edu/~umtriswt.

Skabardonis, A., Varaiya, P.P., Petty, K.F. (2002) "Measuring Recurrent and Non-recurrent Traffic Congestion." Paper presented at the 82nd Annual Meeting of the Transportation Research Board in Washington, DC in January 2003.

Smart Growth America and the State Smart Transportation Initiative. (2012) *The Innovative DOT: A handbook of policy and practice*. Accessed April 25, 2013 at: http://www.smartgrowthamerica.org/the-innovative-dot.

Sorensen, P. (2009) "Moving Los Angeles." *Access*, Vol. 35 (Fall), pp 16-24.

Spagnoli, F. (2008) "Income Inequality (1): Gini." Accessed May 22, 2014 at: http://filipspagnoli.wordpress.com/2008/05/05/human-rights-facts-4/

Strayer, D.L., Drews, F.A., Crouch, D.J. (2006) "A Comparison of the Cell Phone Driver and the Drunk Driver." *Human Factors*, Vol. 48, No. 2, pp. 381–391.

Strickland, S.G., Berman, W. (2005) "Congestion Control and Demand Management." *Public Roads*, Vol. 59, No. 2 (Winter).

Tversky, A., Kahneman, D. (1981) "The Framing of Decisions and the Psychology of Choice."*Science*, New Series, Vol. 211, No. 4481, pp. 453-458.

Transportation Research Board (TRB). (2010) *Highway Capacity Manual*. National Research Council: Washington, DC.

Uddin, W. (2011) *Performance Evaluation of Roundabouts for Traffic Delay and Crash Reductions in Oxford, MS*. Center for Advanced Infrastructure Technology Report UM-CAIT/2011-01. University, MS: The University of Mississippi, June 30.

Varaiya, P. (2008) "Congestion, ramp metering and tolls." *Philosophical Transactions of the Royal Society A*, Vol. 366, pp. 1921–1930.

Viegas, J.M. **(**2001) "Making Urban Road Pricing Acceptable and Effective: Searching for Quality and Equity in Urban Mobility." *Transport Policy*, Vol. 8, No. 4, pp. 289-294.

Wang, Q., Campbell, E., Parsons, R. (2010) *Pupil Transportation: Travel Behavior, Traffic Impacts, and Potentials for Improvement*. University of Vermont Transportation Research Center Report # 10-020.

Watkins, K.E., Hallenbeck, M.E. (2010) "Impact of Weather on Freeway Travel Times in the Rain City." Paper presented at the 89th Annual Meeting of the Transportation Research Board in Washington, DC in January 2010.

Weiner, E. (1997) *Urban Transportation Planning in the United States: An Historical Overview* (Fifth Edition). Publication DOT-T-97-24. Washington, DC: USDOT.

Wessa, P. (2014), Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7. Accessed January 6, 2014 at: http://www.wessa.net.

Wirth, L. (1938) "Urbanism as a Way of Life." *American Journal of Sociology*, Vol. 44, pp. 1-24.

Wu, X., Liu, H.X. (2011) "A shockwave profile model for traffic flow on congested urban arterials." *Transportation Research Part B: Methodological*, Vol. 45, No. 10, pp. 1768-1786.

Yager, C. (2013) *An Evaluation of the Effectiveness of Voice-to-Text Programs at Reducing Incidences of Distracted Driving*. Report SWUTC/13/600451-00011-1. College Station, TX: Texas A&M Transportation Institute.

Yu, X.A., Prevedouros, P.D. (2013) "Left-Turn Prohibition and Partial Grade Separation for Signalized Intersections: Planning-Level Assessment." *Journal of Transportation Engineering*, Vol. 139, No. 4, pp. 399-406.

Zafirovski, M. (2001) “From Market Catallactics To Universal Social Theory: Another Look At The Paradigm Of Rational Choice.” *Theory & Science*, Vol. 2, No.1.

Zhao, P, Luë, B, de Roo, G. (2010) “Urban expansion and transportation: the impact of urban form on commuting patterns on the city fringe of Beijing.” *Environment and Planning A*, Vol. 42, pp. 2467-2486.

Zimmermann, E.W., Ybarra, S., Staley, S.R., Donohue, N. (2011) *The Most for Our Money: Taxpayer Friendly Solutions for the Nation’s Transportation Challenges*. Washington, DC: Transportation for America.