Sacramento's Fix I-5 Project:
Impact on Bus Transit Ridership

By

RACHEL A. CARPENTER
B.S. (California Polytechnic State University, San Luis Obispo) 2008

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Civil and Environmental Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____

Chair H. Michael Zhang

____

Patricia L. Mokhtarian

____

Alexander Aue

Committee in Charge

2010

ACKNOWLEDGEMENTS

I would like to thank my committee chair and advisor, Professor Michael Zhang, for his guidance, suggestions and financial support.

I would like to thank my committee members, Professor Patricia Mokhtarian and Professor Alexander Aue, for their advice and constructive comments on my thesis.

I would like to show gratitude to my student colleagues Zhen (Sean) Qian, Yi-Ru Chen, and Wei Shen for their ever-present willingness to help, and also for their friendships.

Finally, I would like to thank my parents, Linda and Dave, and sister, Molly, for their continued support and patience during my studies at UC Davis.

This research was funded by the Cal EPA.

CONTENTS

TABLE OF FIGURES AND TABLES

Figures

Tables

ABSTRACT

The Fix I-5 project was an engineering project that rehabilitated drainage and pavement on Interstate 5 in downtown Sacramento, from May 30, 2008 to July 28, 2008. In order to alleviate congestion, media outreach alerted commuters about projected traffic conditions as well as advised alternative modes or routes of travel. The construction schedule included complete closures of north or southbound portions of Interstate 5. This study analyzed the impact of the Fix I-5 project closures on peak period bus transit ridership of five transit agencies serving the downtown Sacramento core.

The results indicated that gasoline prices and unemployment rates were statistically significant predictors of transit ridership, with increased gasoline prices and unemployment related to increased bus transit ridership. All agencies had overall increases in mean ridership during the study period, but there were also seasonal variations in mean ridership. Removal of trend and seasonal components in the bus transit ridership data sets was accomplished using multiple regression and sinusoidal decomposition. Time series intervention analysis then estimated that the Fix I-5 project had little impact on mean number of bus riders for all five transit agencies. Bus transit agencies with main service areas closest to the Fix I-5 project were most affected, with ridership increases of about three percent or less attributable to Fix I-5. This study did not analyze the impact of Fix I-5 on other modes of transportation, which may have been more affected than bus transit ridership.

CHAPTER 1 INTRODUCTION

Interstate 5 (I-5) is a major interstate that runs north-south, connecting Mexico to Canada through California, and was started in 1947 by the Federal Highway Administration. The downtown Sacramento portion of I-5 was completed in the 1960's and is nicknamed the "Boat Section" because it was constructed below the water level of the Sacramento River, which runs adjacent to the freeway (Caltrans, 2008). In order to construct the boat section of the freeway, Caltrans had to initially drain this section, and engineer a drainage system of pipes and pumps. The boat section was manually monitored during each winter season to ensure pumps were working properly.

After over 40 years and without major renovation, pavement cracking and sediment accumulation required the boat section to undergo repair, and an opportunity was provided for drainage system upgrades. The California Department of Transportation (Caltrans) Engineers' Estimate projected that the rehabilitation of drainage and pavement of Interstate 5 in downtown Sacramento, dubbed "Fix I-5," would take 305 working days at a cost of more than $44 million (C.C. Myers, Inc., 2009). On February 2, 2008, a Rancho Cordova-based engineering firm, C.C. Myers, Inc., won the Fix I-5 project bid with a proposed 85 working days and 29 night and weekend schedule at a substantially lower cost of $36.5 million, with financial incentives for earlier completion (Caltrans, 2009). Aggressive and compressed construction schedules are not novel for C.C. Myers. Their resume includes more than 17 emergency projects for the State of California, including emergency work on the San Francisco Bay Area's 2007 MacArthur Maze meltdown (C.C. Myers, Inc., 2009). Although not emergency work, the Fix I-5 project specifically included a reconstructed six-inch pavement slab, an upgraded drainage system, new de-watering wells, and installation of electronic monitoring equipment (Solak, 2008). The project was completed in a shorter period than predicted, from May 30, 2008 to July 28, 2008 in 35 days and 3 weekends using full unidirectional closures.

The Fix I-5 construction schedule periodically closed entire northbound or southbound portions of Interstate 5 through Sacramento, a relatively new technique for non-emergency construction. Approximately 200,000 vehicles travel on Interstate 5 in Sacramento each day (Schwarzenegger, 2008). Reports projected that during closure periods, traffic congestion could increase nineteen times (Schwarzenegger, 2008). During closure periods, traffic was detoured to arterial streets and other freeways. In order to alleviate congestion, media outreach alerted commuters about projected traffic conditions as well as advised alternative modes of travel. Employers, including California state government which is one of the largest employers in the area with 75,000 commuters, encouraged employees to use alternative modes of travel (Schwarzenegger, 2008).

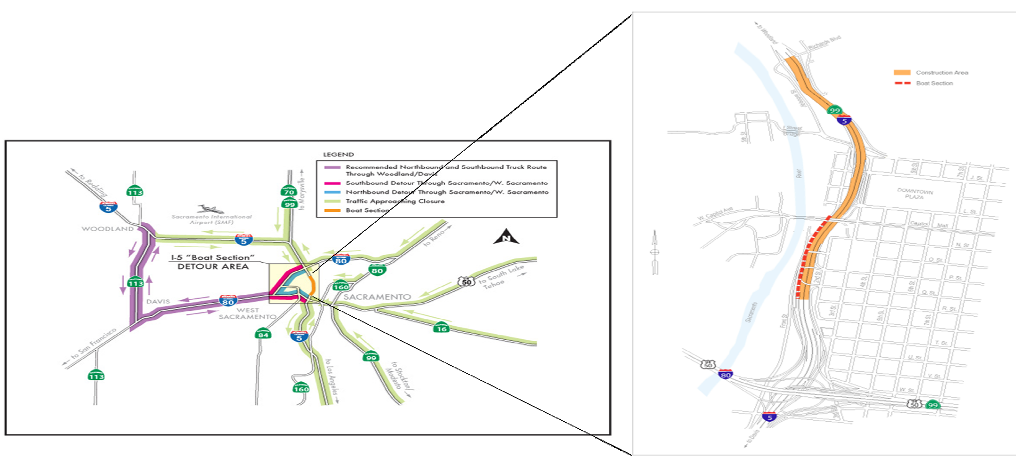


Figure 1.1: The Fix I-5 Construction Area

1.1 Purpose

The main objective of this analysis is to examine the effect that the Fix I-5 project had on commuters' mode choices, more specifically bus transit ridership (supplementary studies are examining the impact of Fix I-5 on other modes of travel). This objective includes the determination of whether the Fix I-5 project caused changes in mean bus transit ridership levels, whether this effect on ridership was permanent or temporary, and the magnitude of the effect. This research includes not only those statistics, but also provides information for service changes for bus transit agencies that need to prepare for future planned construction work, which includes freeway closures such as Fix I-5, and also for unplanned events which force closures.

1.2 Analysis Scope

The primary focus of media outreach was to suggest alternate transportation for those who commute on I-5. State governments and other employers with a large number of employees in the downtown Sacramento core urged employees to use alternate transportation during the Fix I-5 period. Consequently, this study analyzed bus transit agencies' data from the morning (AM) and evening (PM) peak periods. The boundaries of the downtown core were defined as follows: the south boundary defined by the 50/80 freeway, the north boundary defined by Richards Blvd, the west boundary defined by the Sacramento River and the east boundary defined by the Business 80/99 freeway. Bus stops directly below freeway boundaries were considered part of the downtown core. This corresponds to other transit agencies' definitions of downtown Sacramento. Since this analysis focused on commute behavior, only inbound ridership was considered for the AM peak period, while outbound ridership was considered for the PM peak period.

Inbound trips are defined as those trips with a final destination within the downtown core, while outbound trips originate within the downtown core but have a final destination outside it. The AM peak period is the primary morning commute period, but specific hours varied by transit agency. The PM peak period is the primary afternoon commute period and also varied by transit agency. In general, the peak periods occurred between the hours of 5:00AM to 9:00AM, and 3:00PM to 7:30PM. In order to accurately assess bus transit ridership in the downtown Sacramento area, this analysis employed bus transit ridership counts for five transit agencies which provide commute service to the Sacramento downtown core, including: Yuba-Sutter Transit, Yolobus, Roseville Transit, North Natomas Transportation Management Association (TMA) and Sacramento Regional Transit.

As state workers comprise 75,000 commuters in Sacramento, and many state agencies have headquarters in the downtown core, the commute choices made by that group likely had a sizable impact on this study's data sets.

1.3 Gap in Knowledge

In general, many studies have examined transportation-related data using time series methods, although not many have examined bus transit ridership. Few time series studies have analyzed bus transit ridership affected by an outside event (an intervention) using intervention analysis. To date, there are no known studies that examine the intervention of construction work on bus transit ridership.

1.4 Response to the Event

Many public and private agencies united to publicize, prepare and provide for public safety for the Fix I-5 project. These measures included public outreach, intercity and interagency partnerships including the City of Sacramento, City of West Sacramento, Sacramento Area Council of Governments, Downtown Sacramento Partnership, and the Old Sacramento's Merchant's Association. Other efforts included announcements via changeable message signs and highway advisory radio, and California Highway Patrol enforcement in the construction area. Much media outreach was done to warn commuters about traffic conditions and suggest alternative modes of travel. Additionally, various media sources made information about up-to-date information regarding the Fix I-5 project's progress easily available to the general public. The Governor's Executive Order (S-04-08) cited Assembly Bill 32, the California Global Warming Solutions Act of 2006, and advised alternatives to widely used single-occupant vehicle commuting including telecommuting and public transit. Some of the private entities that provided information included News 10, the Sacramento Bee, Sacramento Region 511, and Capital Public Radio, as well as some private business websites. Transit agencies responded to the Fix I-5 project by media outreach that advertised the convenience and availability of transit.

1.4.1 City of Sacramento Traffic Operations Center

An operational tactic for traffic management is the use of traffic operations centers (TOC). The City of Sacramento's single jurisdiction, single agency TOC is operated by the City of Sacramento Traffic Engineering Services Department and funded by Measure A, the gas tax. The goal of their TOC is twofold; first, they must make Sacramento City's transportation network efficient for all transportation modes, and second, they must make the system reliable. Many steps were taken by the TOC in order to ensure their responsibilities were fulfilled during the Fix I-5 project. Planning steps included (City of Sacramento, 2008):

- Identification of potential problem corridors
- Signal maintenance
- Construction of Synchro (transportation modeling software) Model
- Modified signal timing plan
- Coning & striping plan

The TOC makes use of many tools for network monitoring and operation, especially useful during the Fix I-5 project, including (City of Sacramento, 2008):

Figure 1.2: City of Sacramento T.O.C.

- Closed-circuit television (CCTV)
- Advance signal control systems
- Sacramento Police Department Helicopter
- Sacramento Police Officers
- Signal and signage crews
- Traffic cameras (8 Cameras in 2 streams)
- Multi-agency Construction Advisory Team (CAT)
- Traffic Alerts
- Media Contacts

A more detailed summary of the City of Sacramento TOC, based on a field visit on August 21, 2008, is provided in Appendix A.

1.4.2 Government Media Outreach

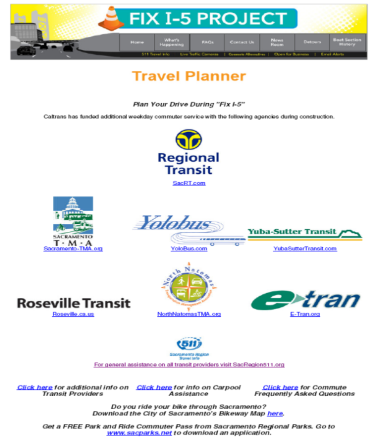


Figure 1.3: The Fix I-5 Website Encouraged Transit

Although all of the sources provided useful information, the official Fix I-5 website, supported by Caltrans, was the most comprehensive and accessible (although no longer active circa August 2008). This website included daily updates ranging from construction updates to detours. It included sections on current work and a history of the portion of I-5 to be repaired (the Boat Section). It also included useful links such as 511 Travel Info, Live Traffic Cameras, and Commute Alternatives. It also provided links to many downtown area businesses, some offering specials to entice people to stay downtown and avoid peak period travel.

Caltrans also hosted three public meetings regarding Fix I-5 in Downtown Sacramento, Natomas and South Sacramento. They gave numerous presentations to audiences including state and local government agencies, residential organizations, private businesses and public officials, and reached an estimated 10,000 people.

In addition to the Fix I-5 website and public meetings Caltrans provided public informational documents. They sent out an email to all Sacramento Personnel Departments which included recommended alternatives to normal work days, including revised work schedules, telecommuting and public transportation. Caltrans provided paycheck stuffers to Sacramento Area employers through Public Outreach Contractors. This document advised departments to reschedule or postpone meetings and events that draw people to downtown. It also informs about a Cal EPA hotline set-up for state workers who needed commute assistance during the Fix I-5 project. Caltrans outreach contractors made information cards available to Sacramento businesses located in the downtown area. These cards provided basic facts about the Fix I-5 project, as well as provided the Fix I-5 website address.

Figure 1.4: Informational Documents Regarding Fix I-5 Closures

Additionally, Assembly member Dave Jones' office sent out a letter to his constituents warning them about the Fix I-5 project, and traffic delays they might encounter. He also encouraged alternate forms of transportation during construction, as well as encouraging shopping or dining with downtown merchants during peak hours.

Although not as comprehensive as the official Fix I-5 website, the City of Sacramento website provided information about the Fix I-5 project. The City of Sacramento also provided parking promotions for six of their parking garages for most of the duration of the Fix I-5 project.

1.4.3 Private Media Outreach

Many private agencies also provided information regarding Fix I-5. In general, these postings included general and up-to-date information about the Fix I-5 Project, but some businesses provided unique information. The News 10 website allowed people to "comment, blog and share photos;" an option not available on the Fix I-5 website. This feature allowed users to share alternate routes through blogs. It also provided Sacramento travel times, as well as easy-to-read color-coded maps that showed lane closure information. The Sacramento Bee provided coverage regarding the Fix I-5 project, through their newspaper publication and website, which provided mobile alerts, a blog jam, and a complete listing of the Fix I-5 stories which were published in the Sacramento Bee newspaper. The Sacramento Region 511 website permanently provides information about traffic, transit, ridesharing and bicycling. They provided minimal coverage regarding the Fix I-5 project, but links to information on transit providers, finding carpools and vanpools, and a guide to bicycle commuting may have been particularly useful to downtown commuters. Capital Public Radio's website provided information about Fix I-5, including a clever 'Jam Factor' scale on their website showing congestion levels on Sacramento area freeways including both north and south bound I-5. Additional Sacramento area businesses posted information about the Fix I-5 project including the NBA Sacramento Monarch's Basketball team, Natomas Racquet Club, California State University Sacramento, Talk Radio 1530 KFBK, and YouTube.

1.4.4 Transit Agency Outreach and Preparation

To prepare riders for the Fix I-5 construction, Regional Transit (RT) posted a press release on their website encouraging people to take transit during the construction period. With additional funding from Caltrans, RT was able to provide supplemental bus and light rail services that increased both capacity and reliability during their peak commuting hours. RT kept ten buses on standby during the construction period and advised passengers to take earlier buses when possible. RT also reminded the public of the 18 park-and-ride lots available throughout Sacramento.

To prepare for the I-5 construction, Yolobus provided an I-5 Construction Options guide in their newsletter. The guide warned passengers of delays and advised them to take earlier morning buses to avoid these delays. Yolobus also took several measures to alleviate overcrowding and delays during the construction period. They had up to two supplemental buses on standby in case other buses were running behind. Yolobus added two morning and two afternoon express trips to both route 45 (service between Sacramento and Woodland) and to route 43 (service between Sacramento and Davis). In addition, Yolobus sold discounted Capitol Corridor train tickets in order to encourage drivers to take transit during the construction period.

To accommodate for the Fix I-5 construction, Roseville Transit posted information on their website regarding the Governor’s Executive Order urging government employees to take transit during the construction. Roseville Transit encouraged new commuter passengers and listed on their website the AM and PM commuter routes with available seating.

In preparation for the Fix I-5 construction, North Natomas T.M.A. posted information in a specific Fix I-5 email newsletter about service changes for the construction period, including loop and route changes that went into effect on June 2, 2008. Additionally, supplemental shuttles and drivers were provided to ease the impact of the anticipated higher ridership during the construction period. The T.M.A. was able to provide additional shuttles with extra funds provided by Caltrans for the construction period, but were required to provide daily counts of AM and PM shuttle riders for each loop. North Natomas T.M.A. also created a special shuttle hotline for passengers to call for up-to-date information about route changes and delays during this period.

In addition to the supplemental schedules, Yuba-Sutter took several other measures to accommodate for the I-5 construction. Route or schedule changes were not made with the exception of minor detours during northbound I-5 closures. Second, Yuba-Sutter had additional buses on call in the event that any early morning buses became overcrowded. Third, Yuba-Sutter used all buses during the construction period, whereas they normally keep three buses non-operational. And finally, Yuba-Sutter closely monitored traffic conditions, which was made possible by improved connections with Caltrans, the City of Sacramento, and Regional Transit.

1.5 Organization of Analysis

The organization of the analysis is as follows. Chapter 2 provides an overview of important concepts in time series which is used in this analysis. It also describes past studies analyzing bus transit ridership, and more specifically those few that used intervention analysis to analyze the impact of an intervention on a time series data set.

Chapter 3 describes the transit agency data, including details of each agency's samples and collection methods, as well as data quality considerations. It also includes information about data cleaning, which was needed to adjust for holidays and limited service days. Finally, data exploration is presented in two sections: measures of centrality and measures of spread for the transit agencies' data sets. Both sections begin by briefly defining the statistics included in that section. Chapter 4 describes the methodology for eliminating trends and cyclic components, and the intervention analysis which examined the impact of the Fix I-5 construction on bus transit ridership. Chapter 5 presents the results of the intervention analysis for each agency, in addition to implications for bus transit agencies for future freeway closures. To conclude, Chapter 6 summarizes the analysis methods and results, and gives recommendations for future work.

CHAPTER 2 LITERATURE REVIEW

Many studies have been conducted analyzing variables that impact transit ridership, primarily using two statistical methods of analysis; time series, and multiple regression. Some, categorized as econometric studies, use those two statistical methods with a focus on economic theory.

Time series is used to analyze a series of data points, to understand the underlying order or context of the data. A review of the literature (Cryer, 1986; Shumway and Stoffer, 2006; Brockwell and Davis, 2002; Anderson, 1976; Kendall, 1973; Kyte et al., 1988) identified a host of different methods used to model time series data, including but not limited to univariate and multiple time series models and transfer function models.

Simple regression is used to analyze the change in a dependent variable as an independent variable changes or is manipulated, while multiple regression uses multiple independent variables (Mann, 2004). However, all regression models assume that the error terms, and therefore response variable observations, are uncorrelated (Kutner et al., 2005). In contrast, time series data often contains observations which are serially dependent (Box and Tiao, 1975). Additional regression methods have been developed that are used for autocorrelated time series data. They employ typical regression techniques, but model the error term using time series models (Tsay, 1984).

Econometrics uses statistical methods to study economic principles (Tinbergen, 1951). The primary focus is the evaluation of economic theory using statistical methods. Discussions of strict and weak stationarity, autoregressive models, and lag structures are found in both econometric time series literature and statistics time series literature.

However, a standard tool in econometrics is to use the structural econometric time series approach (SEMTSA), which uses Box-Jenkins methods but imposes a priori restrictions on the equations based on economic theories (Christ, 1983). Further, this approach is commonly simplified to vector autoregression models (VAR) which omit the moving average polynomial of the ARIMA model (Zellner and Franz, 2004).

Time series was the primary method of analysis used in this study, as autocorrelation was likely to be present in the transit ridership data. Time series analysis encompasses a wide range of models which can handle multiple scenarios within data sets. Time series intervention analysis was used, which provided a methodology to determine the effects of one event on a series. This study used the ARIMA class of time series models, which specify only causality and invertibility as restrictions on the parameters, a feature that was an advantage over models which place additional assumptions on the parameters. Regression was also used to analyze the relationship between multiple independent variables and transit ridership, and for eliminating trends related to independent variables in the transit ridership data sets.

2.1 Time Series

Because time series is a method less commonly used in the field of transportation engineering, a brief overview is given in the following sections.

2.1.1 Background

A time series (x_t) is a sequence of observations collected over time for one variable. Time series can be either continuous or discrete depending on how the observations have been collected. A time series is continuous if observations are taken continuously over time whereas the series is said to be discrete if observations are taken at specific times (Chatfield, 1975). Time series is concerned with chronologically ordered observations of time. Data that is observed over time, both discrete and continuous, is common across many disciplines. In the field of engineering, some examples include series observed over time such as traffic counts and water quality measures. There are many examples in economics, including profits, interest rates, as well as overall economic indicators such as gross domestic product and unemployment rates. In meteorology, a common observation that constitutes a time series is temperature.

Because future observations could be hard to predict, a time series (x_t) is more technically a realization (sample function) of a stochastic process (X_t), which is a family of random variables (Brockwell and Davis, 1987). Time series analysis focuses on studying a time series realization (x_t of X_t) in order to gain insight into the stochastic process (X_t) (Aue, 2009). In practical time series analyses, much of the work is devoted to transforming a nonstationary time series into a stationary process (Fuller, 1976). Conceptually, stationarity is similar to equilibrium within a system. A time series is strictly stationary if its probability structure is not affected by time (Anderson, 1971). In other words, the joint probability distribution of $x_t \ldots x_{t+n}$, is equivalent to the joint probability distribution of $x_{t+h} \ldots x_{t+h+n}$ for all $t, \ldots, t + n \in T$ and h such that $t + h, \ldots , t+h+n \in T$. (Montgomery et al., 2008).

A typically less strict definition of stationarity (for cases where the variance is finite) is called weak stationarity, and is often used because distribution functions are commonly unknown. In order for a time series to be weakly stationary there are two conditions (Shumway and Stoffer, 2006; Brockwell and Davis, 2002):

1. The first moment of x_t is independent of time, t, and is constant.

2. The autocovariance function, defined as $\gamma(h) = Cov\{X_t, X_{t+h}\}$, which depends only on lag *h*, and is independent of *t*.

One important example of a stationary process is called white noise. White noise is commonly denoted $\{Z_t\} \sim WN(0, \sigma^2)$, where Z_t is a sequence of uncorrelated random variables with zero mean and finite variance, σ^2 (Shumway and Stoffer, 2006). White noise is an important building block in time series analysis, as it is the foundation for many more complex processes (Cryer, 1986). It is interesting to note that term *white* noise is derived from white light which is composed of a continuous distribution of wavelengths with the implication that white noise is composed equally of oscillations at all frequencies (Shumway and Stoffer, 2006). Furthermore, if the series of shocks generated are not just uncorrelated (a white noise process), but are independent and identically distributed, the sequence is called i.i.d., denoted $\{Z_t\} \sim IID(0, \sigma^2)$ (Anderson, 1976). Further, if the series is normally distributed, it is both white noise and i.i.d.. Often, a time series (X_t) can be well-explained by a trend component (m_t), a seasonal component (s_t), and a zero mean, random error component (Y_t) (Chatfield, 1975). The process can be represented in the form

$$X_t = m_t + s_t + Y_t.$$

The following provides a short description of each component, although it should be noted that a time series model may exhibit any combination of these components:

The trend component (m_t): Encompasses long run changes in mean. Trends can have many underlying causes including, but not limited to, changes in economic conditions, technological changes and changes in social custom (Farnum and LaVerne, 1989).

The seasonal component (s_t): Encompasses cycles at any recurrent period. This component can include obvious seasonal or annual cycles, or less apparent cycles occurring at any fixed period such as a daily, weekly, or quarterly basis.

The noise component (Y_t): A zero mean, random error component.

There are multiple methodological approaches to the analysis of time series data, more specifically to the removal of trend and seasonal components, including the use of both the time and frequency domains. Analysis in the time domain bases inference on the autocorrelation function, while analysis in the frequency domain pertains to inference based on the spectral density function. Both domains can be used to eliminate seasonal components, while trend components can only be eliminated in the time domain. In this study a decomposition method was used which identified and separately removed the trend and seasonal components from the series. The removal of trend components used methods associated with the time domain, and the removal of seasonal components used methods associated with the frequency domain.

2.1.2 Trend Components

Analysis in the time domain includes methods for removal of both the trend and seasonal components including least squares estimation, smoothing with moving averages, differencing, small trend methods, and moving average estimation (Aue, 2009). Additionally, trend components can be removed using regression techniques (Yaffee, 2000). Aue (2009) provides a detailed description of each method. This study used multiple regression to remove trend components. The multiple regression method is discussed below, in addition to differencing which is referred to in later sections:

Multiple Regression: When there are four predictor variables, P_1, P_2, P_3, P_4 as in this analysis, the model is formulated as

$$X_t = \beta_0 + \beta_1 P_{1t} + \beta_2 P_{2t} + \beta_3 P_{3t} + \beta_4 P_{4t} + \varepsilon_t.$$

In this study, the combination of the predictor variables $(\beta_1 P_{1t} + \beta_2 P_{2t} + \beta_3 P_{3t} + \beta_4 P_{4t})$ constitutes the trend component, while the regression error term (ε_t) constitutes both the seasonal and error terms ($s_t + Y_t$). As discussed previously, standard linear regression models assume that the error terms, ε_t, and therefore, response variable observations, X_t, are uncorrelated (Kutner et al., 2005). Time series data, on the other hand, often contains observations which are serially dependent (Box and Tiao, 1975). Therefore, modifications to standard linear regression would be necessary, including modeling the error terms as time-series autoregressive moving average models (Tsay, 1984; Ostrom, 1978).

Differencing: Applies the difference operator to the original series in order to create a new, stationary series. The lag l difference operator (∇) is defined as (Shumway and Stoffer, 2006):

$$\nabla^{\mathrm{l}} \mathrm{x_t} = x_t - x_{t-l}.$$

In practice, it is common to denote the use of the difference operator by using the backshift operator, B. In this case,

$$\nabla^{\mathrm{l}} \mathrm{x_t} = x_t - x_{t-l} = (1 - B)^l x_t$$

2.1.3 Seasonal Components

This study's decomposition method removed the trend components using multiple regression, and removed seasonal components from the series using a frequency domain approach. The frequency domain, also referred to as the spectral domain, pertains to inference based on the spectral density function. A time series can be decomposed into periodic components, each of which contains variation at that period's frequency, whose variations combine together to cause the overall variation in the time series. Therefore, a time series can be well represented as the sum of significant periodic components (Chatfield, 1980):

$$x_t = \sum_{j=1}^{k} \mathrm{A_j} \cos(2\pi\omega_j \mathrm{t}) + \mathrm{B_j} \sin(2\pi\omega_j \mathrm{t}) + Z_t$$

where A_j and B_j are uncorrelated random variables with mean zero and variances both equal to σ^2 and $\omega = \frac{1}{d}$, where d is the period of the cycle. For example, if there is an annual cycle and the data set contains monthly data points, one period, d, could be 12. Exploratory analysis using the periodogram can help to determine genuine periodic (seasonal) components within the time series, X_t. The definition of the periodogram for $\{X_1,\ldots,X_n\}$ is given below (Brockwell and Davis, 2002):

$$I(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} \mathrm{X_t} \mathrm{e}^{\mathrm{it}\omega} \right|^2$$

where ω is the frequency. The periodogram is the graph of ω and $I(\omega)$ and is an estimation of the power spectral density function. Although the periodogram is not a consistent estimator of the spectral density because the variance of $I(\omega)$ does not decrease as the sample size, n, increases, it will be used in this analysis to determine periodicities, which is a common practice (Chatfield, 1975). If the periodogram is constructed for $-\pi \leq \omega \leq \pi$ the area under the periodogram represents the variance of the time series (Brocklebank, 2003). Therefore, peaks in the periodogram generally indicate frequencies that can explain a significant part of the total variance. For example, a periodogram that displays a large peak at frequency $\omega = 0.25$, indicates a period, $d = 4$, which for quarterly data indicates an annual cycle. If a periodogram does not display any obvious peaks, all frequencies are contributing to the series' variance, and the series may even be a white noise process. The variance by cycles can be decomposed as follows (Aue, 2009),

$$I(\omega_j) = \frac{d_c(\omega_j) + d_s(\omega_j)}{2}$$

where $d_c(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \cos(2\pi\omega_j t)$ and $d_s(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \sin(2\pi\omega_j t)$. As discussed, the periodogram can help to determine seasonalities and peaks in the periodogram can signify a genuine periodic component which explains a large portion of the variance in the time series. However, it is possible that peaks may occur because of random fluctuations in the sample (Priestley, 1981).

This study used spectral analysis of variance to determine whether peaks in the periodogram explain a larger portion of the variance than is expected with sequences such as white noise and ARMA processes.

2.2 Goodness-of-fit Tests

Ideally, after trend and seasonality are removed, the remaining series will be a white noise process. There are many goodness-of-fit tests to determine whether the residuals are white. For an extensive review of diagnostic checks, refer to Li (2004). For the purposes of this study, four goodness-of-fit tests will be utilized, including the sample autocorrelation function (ACF), the portmanteau test (Ljung-Box modification), the rank test, and a test of normality including the squared correlation (R^2) based on a qq plot. An explanation of the four goodness-of-fit tests is described below:

1. *The sample autocorrelation function (ACF)*: The autocorrelation function and sample autocorrelation functions at lag *h* are defined as (Anderson, 1976):

$$\rho_h = \frac{\gamma_h}{\gamma_0} \qquad\qquad \hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}$$

For a series, $Y_1,\ldots,Y_n$, with a large sample size, n, the sample autocorrelations are i.i.d. with zero mean and variance $\frac{1}{n}$ (Brockwell and Davis, 2002). Therefore, in order to test for randomness, a plot of the sample autocorrelation function for any amount of lags *h* should show should that 95% of those lags fall within the bounds $\pm\frac{1.96}{\sqrt{n}}$ if the process is i.i.d. (Aue, 2009).

2. *The portmanteau test (Ljung-Box modification)*: In order to test for randomness, originally, Box and Pierce (1970) suggested the portmanteau test, and developed the statistic, *Q*, as

$$Q(\hat{r}) = n\sum_{j=1}^{h}\hat{r}^2(j)$$

where $\hat{r}$ is defined as the autocorrelation function. Ljung and Box (1978, p. 298) suggest that the Box-Pierce methodology produces "suspiciously low values of Q($\hat{r}$)…" and propose a modified version as

$$Q(\hat{r}) = n(n+2)\sum_{j=1}^{h}\frac{1}{n-j}\hat{r}^2(j)$$

where Q can be approximated as a chi-squared distribution with h degrees of freedom. The hypothesis that the residuals are i.i.d can be rejected at the level α, if $Q > \chi^2_{1-\alpha}(h)$ (Brockwell and Davis, 2002).

3. *The rank correlation test*: The rank test is a test of randomness, to establish whether there remains any systematization in the residuals. For a time series, a trend can be determined by the correlations between the rank order of the time series observations and their time values (Kendall, 1955). In total, there are $P + Q = \frac{1}{2}\mathrm{n}(\mathrm{n}-1)$ pairs, where *P* designates the number of positive correlations, and *Q* designates the number of negative correlations. *P* is represented by Kendall's τ, called the coefficient of rank correlation:

$$\tau = \frac{P+Q}{\frac{1}{2}n(n-1)}$$

The coefficient of rank correlation ranges between 1 (perfect positive correlation) and -1 (perfect negative correlation), with $\tau = 0$ representing an independent, white noise process. Refer to Kendall (1955) for further explanation.

4. *R^2 based on a qq plot*: In order to assess the normality of the residuals, the squared correlation (R^2) value can be calculated based on a quantile-quantile plot (qq plot). A qq plot is a graph that compares the quantiles of two distributions. For this study, the first data set is the ordered residuals from the fitted model assuming a mean zero,

variance one process denoted as Y_j. The second data set is ordered statistics from a random normal sample with mean μ, variance σ^2 denoted as n_j. If the model residuals are normally distributed, the pairs (n_j , Y_j) should have a linear relationship (Shumway and Stoffer, 2000). Hence, perfect normally distributed residuals would display an R^2 value equal to one. If the R^2 value is too small (based on the level α), then the assumption of normality must be rejected. More specifically, the R^2 value can be computed as follows, noting that Φ_j represents the normal distribution:

$$R^2 = \frac{\left[\sum_{j=1}^{n}(n_j - e_j)\Phi_j\right]^2}{\sum_{j=1}^{n}(n_j - e_j)^2 \sum_{j=1}^{n} \Phi_j^{\,2}}$$

Refer to (Shapiro and Francia, 1972) for the critical values of R^2.

For residual testing in this study, lag $h = 20$ was used which is commonly used in time series residual testing (Shumway and Stoffer, 2000).

2.3 Multicollinearity

Multicollinearity occurs when independent variables are highly correlated in a multiple regression model (Kutner et al., 2005). This means that the two correlated variables are not providing independent information which helps to predict the dependent variable. Severe cases of multicollinearity must be corrected, because the result can be unstable regression coefficient estimates. Further, multicollinearity is often indicated by very large standard errors, even though the coefficients are still the best linear unbiased estimators (BLUE) (Washington et al., 2003). If two independent variables are highly correlated, it is difficult to determine which variable is explaining more variation in the dependent variable (both variables' standard errors will become large). Another test for the presence of multicollinearity is the comparison of correlation coefficients to regression coefficients. If their signs are different (+/-) then multicollinearity should be further investigated (Kutner et al., 2005). Two methods for detecting multicollinearity are the variance inflation factor and the condition index.

1. The variance inflation factor (VIF) is defined as $(VIF)_k = \frac{\sigma^2\{\beta_k^*\}}{(\sigma^*)^2}$ where β_k^* are the estimated standardized regression coefficients and $(\sigma^*)^2$ is the variance of the error term[1] for the correlation transformed model (also called the standardized regression model). The multiple regression model discussed previously was $X_t = \beta_0 + \beta_1 P_{1t} + \beta_2 P_{2t} + \beta_3 P_{3t} + \beta_4 P_{4t} + \varepsilon_t$, while the standardized regression model is $X_t^* = \beta_1^* P_{1t}^* + \beta_2^* P_{2t}^* + \beta_3^* P_{3t}^* + \beta_4^* P_{4t}^* + \varepsilon_t^*$. If the mean of the VIF values is greater than 1, serious multicollinearity may exist (Kutner et al., 2005).

2. The condition number (κ) is defined as the largest condition index (CI). It is defined as $\kappa = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$ where λ_{max} is the largest eigenvalue, and λ_{min} is the smallest eigenvalue of the $X'X$ matrix. Condition numbers between 5 and 10 indicate some dependence, while CI values of 30 and above signify strong dependencies (Belsley et al., 1980).

2.4 Lagged Variables

In time series regression models, it is often the case that time lags need to be included (Ostrom, 1978). For example, there is a time lag associated with exposure to carcinogenic substances and the development of cancer. If there are time lags between a change in the

---

[1] In this study, the error term (when testing for multicollinearity) includes a seasonal and noise term ($s_t + Y_t$).

independent variable and the effect on the dependent variable, a lag term should be included in the regression model. In this study, the explanatory variables were unleaded regular gas prices, unemployment rates, gross domestic product and transit fare prices, each of which could have a time lag with transit ridership. However, a time series study of Portland, Oregon transit ridership between 1971 and 1982 focusing on factors that affect ridership show that neither gas price (aggregation level unspecified) nor county employment rates show a time lag for bus transit ridership (Kyte et al., 1988). However, Kyte et al. found a time lag between transit fare prices and ridership. The authors stated that the largest response in ridership to the fare increase occurred almost immediately, and then decayed at a measurable rate for three months. Prior studies have not determined a set of independent variables that consistently predict bus transit ridership. The effects of GDP on ridership have not been studied.

2.5 Box-Jenkins (ARMA) Models

In the time domain, linear filters are often used to transform one time series into another, under the assumption of linearity, and can be defined as:

$$Y_t = \sum_{j=-\infty}^{\infty} \Psi_j X_{t-j}$$

where Ψ_j's are weights for each X_t, and X_t and Y_t are the input and output time series, respectively (Chatfield, 1975; Montgomery et al., 2008). There are many types of linear filters which can be applied to white noise to obtain a more complex linear time series. In general, there are three major classes of linear filters, including autoregressive, moving average and autoregressive-moving average filters. They are described below:

1. *Autoregressive Process, AR(p)*: An autoregressive process can be represented as:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t.$$

The equation's conceptual interpretation is that the current time series observation is a linear weighted combination of the *p* most recent past values of the same time series, plus an error term (Montgomery et al., 2008). The autoregressive polynomial is defined as

$$\phi(Z) = 1 - \phi_1 Z - \phi_2 Z^2 - \cdots - \phi_p Z^p.$$

The roots of the polynomial $X_t = \phi(B)Z_t$ must lie outside of the unit circle to ensure that an AR(*p*) process is stationary; a condition commonly referred to in time series literature as causality (Box et al., 2008).

2. *Moving Average, MA(q)*: An moving average process can be represented as:

$$X_t = Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q}.$$

Observably, a moving average model assumes the current value is a linear weighted combination of q lagged white noise terms. Further, a condition called invertibility is imposed on the weights, θ_j, to ensure a unique MA process for an autocorrelation function (Chatfield, 1980). The moving average polynomial is defined as

$$\theta(Z) = 1 + \theta_1 Z + \phi_2 Z^2 + \cdots + \phi_q Z^q.$$

An MA(q) process is invertible if the roots of $X_t = \theta(B)Z_t$ are outside the unit circle (Box et al., 2008). Invertibility and stationary are two separate conditions; an MA(q) process will always be stationary.

3. *Autoregressive-Moving Average, ARMA(p,q)*: An autoregressive-moving average (ARMA) model assumes that the current observation is a linear weighted

combination of the *p* most recent past observations from the same time series (the AR(*p*) portion), as well as q lagged white noise terms (the MA(q) portion). An autoregressive-moving average process ARMA(*p,q*) can be represented as

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q}.$$

An ARMA (*p,q*) process is causal if the roots of the polynomial $X_t = \phi(B)Z_t$ lie outside of the unit circle, and is only invertible if the roots of $X_t = \theta(B)Z_t$ are outside the unit circle (Box et al, 2008). The coefficients for causality are computed from the expression $\Psi(Z) = \frac{\theta(Z)}{\phi(Z)}$, while the coefficients for invertibility computed from the expression $\Psi(Z) = \frac{\phi(Z)}{\theta(Z)}$.

2.6 Intervention Analysis

Gene Glass (1972, p.463) coined the term *intervention* and described it as follows: "Observation of a variable Z at several equally spaced points in time yields the observations $z_1, z_2, \ldots, z_N$. Suppose that an intervention (T) is made at some point in time before time N into the process presumed to be controlling Z. The time-series is said to be interrupted at a point in time, say n_1 less than N: $z_1, \ldots, z_{n_1}, T,\ z_{n_1+1}, \ldots, z_N$." Box and Tiao (1975) used the term intervention and constructed an analysis method to determine the effect of an intervention, occurring at a known time in a time series. Their intervention model is based on the basic transfer function model,

$$X_{t,2} = \sum_{j=0}^{\infty} \tau_j\, X_{t-j,1} + N_t$$

where $X_{t,1}$ represents the input series, while $X_{t,2}$ represents the output series, which constitutes common notation in transfer function modeling. In intervention analysis, it is more common to replace $X_{t,1}$ with X_t, and also to replace $X_{t,2}$ with Y_t. The basic intervention model can be described:

$$Y_t = \sum_{j=0}^{\infty} \tau_j X_{t-j} + N_t$$

where X_t and Y_t are the input (pulse/step) and output (ridership, after removal of trend and seasonal components) series of the model respectively, τ_j is a linear filter and N_t represents a noise sequence. τ_j is defined as

$$\tau_j = \rho_{YX}(j)\sigma_Y/\sigma_X$$

where $\rho_{YX}$is the cross-correlation between X_t and Y_t, and σ^2 is the variance of each series. In the case of intervention "$T(B) = \sum_{j=0}^{\infty} \tau_j B^j$ is simplified with a rational operator of the form T(B)= $\frac{B^b W(B)}{V(B)}$" where b is the delay parameter, and W and V help to provide coefficients to represent more complicated indicator series built upon a step or pulse function (Brockwell and Davis, 2002, pp. 340-341). The intervention term is then $T(B)X_t$. For a series that might be best represented as an intervention causing a temporary change in the response variable, a pulse indicator variable would be most appropriate:

$$X_t = \begin{Bmatrix} 1 \; if \; t = T \\ 0 \; if \; t \neq T \end{Bmatrix}$$

where t is time, and T is the period of the intervention. For a series that might be best represented as an intervention causing a permanent change in the response variable, a step indicator variable would be most appropriate:

$$X_t = \begin{Bmatrix} 1 \; if \; t \geq T \\ 0 \; if \; t < T \end{Bmatrix}.$$

In general, the Box and Tiao intervention analysis methodology follows a five-step process (Box and Tiao, 1975):

1. Eliminate trend and seasonal components from the original time series (X_t). This study eliminated trend components from the series through a multiple regression, and eliminated seasonal components using sinusoidal decomposition with cycles determined by the periodogram and cycle significance based on spectral ANOVA.

2. Use ordinary least squares (OLS) regression to obtain a initial estimate of τ_j, which represents the transfer model.

3. Model the residuals from the OLS regression as an ARMA(*p,q*) process, which will represent the noise model. For model diagnostics, analyze the residuals using goodness-of-fit tests.

4. Minimize the sum of squares, $\sum_{t=m*+1}^{n}\left(\frac{\hat{\phi}^{N}(B)}{\hat{\theta}^{N}(B)}\right)^{2}(W, V, \phi^{N}, \theta^{N})$, where m*=max ($p_2 + p$, b + $p_2 + q$), in order to obtain final parameter estimates of both the noise and transfer models.

5. Analyze the final model residuals using goodness-of-fit tests. This study used the Sample ACF, qq plot, Ljung-Box test and rank test.

2.7 Review of Relevant Past Studies

Previous studies involving multiple regression and time series analysis are discussed. Many studies have examined transportation-related time series data and used time series methods to analyze the data. Kyte et al. (1988) reviews previous work in transportation related time series, other than bus transit ridership. For example, Atkins (1979) analyzes the effect of speed limit changes on traffic accidents in British Columbia in the 1970's using intervention analysis. Additionally, studies that use vehicle miles travelled (VMT) forecasting models show that VMT can be predicted by independent variables that are similar to those used to predict transit ridership. Common predictors include, but are not limited to, gasoline price (Schimek, 1996 (1521 and 1558); Goodwin et al., 2004; Gately, 1990) and income (Schimek, 1996 (1521 and 1558); Goodwin et al., 2004; Gately 1990). Dahl (1986) summaries previous research on gasoline consumption demand, VMT and miles per gallon (not just VMT), finding negative elasticities for price, and positive elasticities for income. Elasticities measure the responsiveness of one variable to change in another variable. Mokhtarian et al. (2002) analyzed induced demand with respect to highway capacity expansion, and listed predictors of induced vehicle travel as changes in population, demographics, the economy, mode and land use, but not highway capacity expansion. Rose (1982, 1986) examines rail transit ridership using time series and multiple regression techniques. Rose (1986) studies Chicago Transit Authority rail ridership, more specifically, 11 years of monthly average weekday data. He used fares, weekday service miles, cost of car trips (including gas prices), and weather changes and found that gas prices and service levels were significant predictors of rail ridership. But, there are few studies that analyze bus transit ridership with time series models, a fact that was confirmed by librarians at the Physical Science and Engineering Library at UC Davis, and the Institute of Transportation Studies Library at UC Berkeley. Those pertaining to transit ridership (defined as both bus and rail, or just bus) are discussed below.

2.7.1 Predicting Transit Ridership Using Multiple Regression

A number of studies use multiple regression techniques to determine factors that affect bus transit ridership. Those studies take into account autocorrelation in the residuals to ensure valid model results. The following presents studies which use multiple regression as the primary analysis method. Agrawal (1981) analyzed Southeastern Pennsylvania Transportation Authority's City Transit Division's annual full-fare adult ridership between 1964 and 1974. Using multiple regression, he found that three factors were statistically significant in affecting ridership and produced a multiple correlation coefficient of 0.9985. The three significant predictors included average fare (adult riders), jobs in Philadelphia, and bus miles of service, while number of vehicles owned was not a significant predictor. Lane (2009) applied regression techniques to monthly bus and rail transit ridership data from nine US cities between January 2002 and April 2008, and found that gasoline prices were a statistically significant predictor of changes in transit ridership, while service characteristics and seasonality were not significant predictors. Wang and Skinner (1984) analyzed fares, gas prices and monthly ridership data from seven transit authorities across the United States, and using regression techniques, found that as real gasoline prices increased, transit ridership increased, although by a small amount. Also, they found that as real fares increased, ridership decreased. Taylor et al. (2009) analyzed transit ridership from 265 urban areas using 22 independent variables to and show that the majority of transit ridership variation can be explained by variables within the categories of regional geography, metropolitan economy, population characteristics and auto/highway system characteristics. They found a positive correlation between ridership and gas prices, and a negative correlation between ridership and unemployment levels. Gomez-Ibanez (1996) reported an increase in Massachusetts Bay Transportation Authority (MBTA) bus transit ridership in Boston, in part due to service improvements such as phased station modernization and bus replacement, and transit fares which increased less than the inflation rate. Their study also included income, Boston employment, fares, and vehicle miles. Kitamura (1989) showed a causal relationship between car ownership and transit use, more specifically that an increase in car ownership leads to a decrease in transit use, using Dutch National Mobility Panel weekly travel diary data. Cervero (1990) provides a broad overview, and summarizes multiple empirical studies which show that there are many factors affecting transit trips, including characteristics of the traveler such as age, income, auto access, trip purpose, trip length, and also characteristics of the operating environments, such as land use and location settings. Although each study used a different set of independent variables to predict transit ridership, most of the studies that used multiple regression included gas prices, fares, and economic indicators such as unemployment rates.

2.7.2 Transit Ridership and Intervention Analysis

In terms of transit ridership and intervention analysis, there is a scarcity of previous studies. Kyte et al. (1988) use Tri-County Metropolitan Transportation District of Oregon bus transit ridership on various aggregation levels (system, sector and route levels) between 1971 and 1982 to show that service level, transit fares, gasoline price, and employment are statistically significant predictors of bus transit ridership. They also note that to fully explain ridership demand, many more independent variables should be considered. Kyte et al. (1988) used intervention analysis to model changes in bus transit ridership resulting from eleven separate cases of changes in their predictor variables including increased fares, system-wide service changes, and route-level changes, and they observed that the occurrence of multiple events at one time makes it difficult to isolate the impact of any single event on ridership. Their results showed that for the four cases of fare increases, the result in terms of ridership is varied. The separate interventions of system-wide service changes and gasoline supply shortages combine to produce an intervention output of an additional 8,400 bus transit riders. Kyte et al. use elasticities greater than one to determine significance of intervention results. Narayan and Considine (1989) use intervention analysis to analyze two cases of fare increases, in April 1980 and April 1984, and their effects on monthly upstate New York transit ridership, assuming that ridership could be decomposed into a trend, seasonal, intervention and noise term. They assume that the intervention term is best represented as a step function; an "abrupt and permanent change" in ridership (Narayan and Considine, 1989, p. 248). Their methodology differs from the original Box and Tiao intervention analysis methodology, as their model isn't based on the transfer function model, but on regression with correlated error terms, and eleven indicator variables for seasonality, indicator variables for the two fare price increases, and an error term which they claim "correct[s] for autocorrelated errors" (Narayan and Considine, 1989, p. 249). However they didn't use ARMA models for the noise, and it is unclear how they corrected for correlation, because they used *t* tests which require the removal of serial dependence, nonstationarity and seasonality. Both fare interventions produced significant ridership decreases. Considine and Narayan (1988) use data from Chattanooga, Tennessee and intervention analysis to examine the affect of market changes on total ridership, total operating revenues, the ratio of total operating revenues to total revenue miles, and the ratio of total passenger trips to total revenue miles. They slightly modify the Box-Tiao methodology by first using the entire sample to model the noise term, then separately estimating the intervention term, and then minimizing all parameters. They use *t* statistics to test for significance. They show that marketing does significantly affect transit ridership.

2.8 Summary of Literature Review

An extensive literature review identified a number of past studies using transportation-related data. Fewer studies used both multiple regression (taking into account autocorrelation) and time-series methods for the analysis of predictors for transit ridership. There were still fewer time series studies that analyzed transit ridership affected by an intervention, using intervention analysis. To date, there are no known studies that examine the impact of the intervention of construction work on bus transit ridership.

CHAPTER 3 DATA DESCRIPTION

This chapter describes the data used in this analysis. A brief description of methods of ridership data collection is given. Each bus transit agency is described, with their ridership data, and the methods they use to collect ridership data. Data filtering that was required to construct a data set for this analysis is described, with information regarding data imputation for missing data. An analysis was performed on each of the ridership data sets to determine if any independent factors played a significant role in ridership changes during the period of analysis. Data quality with relation to methods of data collection is discussed.

3.1 Methods of Data Collection

Four methods of ridership data collection were used among the five transit agencies that provided service to the downtown core. Those methods included automatic passenger counters (APC), electronic fareboxes, manual counts by route checkers, and manual counts by bus drivers. A description of each method is provided below:

1. *Automatic Passenger Counters (APC)*: APC devices are often door-mounted and use infrared beam technology to automatically count boarding and alighting riders. Many use GPS technology to associate collected data with a time and location.
2. *Electronic Fareboxes*: Electronic fareboxes are devices that collect ridership information. Typically, a bus driver enters a number corresponding to rider type into a key pad on the electronic farebox which stores the data until it is uploaded to a network. Usually, electronic fareboxes do not provide location information.

3. *Manual Counts by Route Checkers*: Route Checkers take manual counts of passengers boarding and alighting at each stop, and the arrival and departure times of these stops.

4. *Manual Counts by Bus Drivers*: Bus Drivers take manual counts of passengers boarding and alighting at each stop, and the arrival and departure times of these stops.

3.2 Data Sample

This section describes the data samples provided by each of the five bus transit agencies. The section is divided into five sub-sections, one for each agency. Each sub-section includes a brief background of each bus transit agency, ridership data collection methods employed by each agency, and the data provided by each agency. Unless otherwise stated, all information regarding each transit agency was obtained through personal correspondence as listed in Table 3.1:

Table 3.1: Data Collection Details

Transit Agency	Contact	Contact's Official Position Title	Type of Personal Correspondence
Regional Transit	James Drake	Assistant Planner	e-mail, phone, mail, in-person
Yolobus	Erik Reitz	Transit Planner	e-mail, phone, mail, in-person
Roseville Transit	Teri Sheets	Alternative Transportation Analyst	e-mail
	Elizabeth Haydu	Administrative Technician	e-mail, phone, in-person
North Natomas TMA	Sarah Janus	Program Coordinator	e-mail
Yuba-Sutter Transit	Dawna Dutra	Analyst	e-mail, phone

3.2.1 Regional Transit

The Sacramento Regional Transit District (RT) operates bus and light rail transit serving 418 square miles of the greater Sacramento metropolitan area (Regional Transit, 2009). They are the largest provider of public transportation within the City of Sacramento, operating 256 buses servicing 97 bus routes with more than 3,600 bus stops which operate from 5 A.M. to 11:30 PM, 365 days per year (Regional Transit, 2009).

RT uses all four data collection methods described in Section 3.1. Regional Transit is the only transit agency within our study that collects ridership data using APC devices, which have been installed in half of the RT bus fleet. Electronic farebox devices are installed on all RT buses, and is the method that RT uses for annual reporting. But because electronic fareboxes don't keep track of location or alighting passengers, this data was not suitable for this study. The ridership data for RT consisted of APC data even though it is not used for official reporting. It records boarding and alighting riders, as well as time and location stamps for each record, which was necessary for filtering purposes. APC devices are still in testing stages. The FTA's National Transit Database requires that two random bus trips must be sampled per day by route checkers, which is why RT also employs this ridership collection method (Drake, 2007). Finally, RT makes use of manual counts by driver for its specialized Community Bus Service (now called Neighborhood Ride) which offers intraneighborhood service within certain communities while also servicing seniors and the disabled.

Because of the expansiveness of RT, total ridership counts are nearly impossible to obtain. On a normal weekday, the RT bus system makes 3,000 trips. As mentioned, approximately half of its bus fleet is equipped with APC devices, which results in the collection of data for 1,500 trips per day. The data is wirelessly uploaded to the RT network, where it undergoes filtering which uses by a relaxed set of rules to remove faulty data. The core set of rules that determine the filtering include:

- The difference between total riders on and total riders off for a bus must be 10% or less,
- The difference between total riders on and total riders off for a block must be 10% or less (a block is a schedule for one physical bus each day),
- The difference between total riders on and total riders off for a trip must be 10% or less,
- The number of stops counted must be “pretty close” to the actual number of stops on the route,
- Records showing obvious technology malfunctions.

As filtering occurs, records are deleted from the database. After the filtering process is complete, about 400 records per day remain. The original raw data only remains as the output from the APC device in the form of a text file. It is difficult to accurately assess RT’s total bus ridership because the data is not a random sample, which is a result of the filtering process and the fact that bus lines are not randomly chosen. Because total ridership data by route is not easily obtained with filtered APC data, the raw APC data was also provided but was not used due to data quality concerns. In addition, data from General Farebox Inc. (counts from the fareboxes on the buses), Parking Lot, Cash, and fare vending machine count data was provided. That additional data was not useful as it provided system-wide information, and was not specific to only those bus lines serving the downtown area.

The APC daily data, including weekdays and weekends, spans from January 7, 2008 to December 30, 2008. Initially, the APC data was a count of boarding and alighting riders of a randomly chosen number of stops within the entire RT service area, allowing for the filtering of bus lines serving the downtown area. The data also contained the time and date of the observation as well as the bus stop identifier and route schedule identifier. RT defined the AM peak period to be 6:30-9:00AM, and the PM peak period to be 3:00-6:00PM. Although the RT sample only contains 49 weekly ridership counts, Cherwony and Polin (1977) used daily bus transit ridership data from Albany, New York to show that only 30 days of transit ridership data is needed to develop a valid travel-forecasting model.

3.2.2 Yolobus

Yolobus is operated by the Yolo County Transportation District and serves Yolo County and surrounding areas including Davis, Sacramento, Winters, and Woodland among others. Unlike RT, Yolobus also provides service to the Sacramento International Airport. Yolobus uses electronic farebox devices as well as manual counts by bus driver to collect ridership information. Yolobus separates its services into three types: regular, commute, and express. Regular services run every day of the week whereas commute and express services only run Monday through Friday. The agency operates 365 days of the year. Since Yolobus offers different types of services to the downtown core, the operating times of those services vary. Regular bus routes (40, 41, 42A/B, 240) collectively run all day from 4:37 AM to 11:48 PM during the week. The commute and express services, which only run Monday through Friday, only run during peak commuting periods. The commuter routes (39, 241) collectively run from 5:35 AM to 8:30 AM and 3:35 PM to 6:34 PM. Similarly, the express routes (43, 44, 45, 230, 231, 232) collectively run from 5:55 AM to 8:32 AM and from 4:03 PM to 7:17 PM. Yolobus restricts passenger travel in downtown Sacramento by not allowing passengers to both board and alight in downtown Sacramento. Instead, passengers are requested to utilize Sacramento RT for local services within downtown Sacramento.

The Yolobus ridership data set contains the total daily ridership counts that span the three-year period from January 2006 to December 2008. The data set is missing two days, July 30, 2006 and July 31, 2006.

3.2.3 Roseville Transit

Roseville Transit is operated by the City of Roseville and mainly serves the City of Roseville, but additionally serves Sacramento commuters. Roseville Transit runs specific commuter routes that serve the Sacramento downtown core, including AM Routes 1-8 and PM Routes 1-8. The commuter routes only run Monday through Friday between the morning peak commute hours of 5 – 9 AM and the afternoon peak commute hours of 3:30 – 7:30 PM. Roseville Transit uses manual counts by bus driver to collect ridership information. Its daily, peak-period, ridership data was provided for the entirety of 2006, 2007, and 2008, including separation by commuter route.

3.2.4 North Natomas TMA

The North Natomas "Flyer" is operated by the North Natomas TMA, serving Natomas as well as downtown Sacramento commuters. The Flyer runs between 20 and 28 passenger buses through several North Natomas neighborhoods, and although there are no timed stops, time points are listed on the schedule (the bus will stop wherever there are passengers waiting). In downtown Sacramento, there are timed stops at set locations. The Flyer includes three routes that serve the downtown core: the Eastside Route, the Westside Route, and the Central Route. In September of 2008, North Natomas TMA began running a Square Route; however, those ridership counts were excluded from the daily totals because that route was added after the construction period had ended and there was no pre-construction or construction data to use for comparison. The Flyer operates Monday through Friday except on certain holidays. North Natomas TMA runs peak period scheduled routes between Natomas and downtown Sacramento. The Eastside Route has three morning and three afternoon loops that run from 5:54 AM – 9:04 AM and from 3:35 PM – 6:54 PM, respectively. The Westside Route has two morning and two afternoon loops that run from 6:00 AM to 7:44 AM and from 4:30 PM to 6:30 PM, respectively. The Central Route also has three morning and three afternoon loops that run from 6:03 AM – 9:04 AM and from 4:07 PM – 7:06 PM, respectively.

North Natomas TMA uses manual counts by driver as well as farebox counts to collect ridership information. Manual counts were also provided by volunteer riders during the duration of the Fix I-5 project. Their daily peak-period ridership data, collected using both manual and automated collection methods, spans the complete 2008 year and is separated by route.

3.2.5 Yuba-Sutter Transit

Yuba-Sutter Transit is operated by Sutter and Yuba Counties and the Cities of Marysville and Yuba City and provides service to Yuba City, Marysville, Linda, Olivehurst, East Nicolaus and Sacramento. Only the Sacramento Commuter Express provides Sacramento downtown commuter service (via Highways 70 and 99). The commuter service runs on weekdays, but not on certain holidays. Yuba-Sutter currently provides nine commuter schedules for each of the peak periods that operate from 5:20 AM to 8:00 AM and from 3:45 PM to 6:50 PM.

Yuba-Sutter Transit uses manual counts by the drivers to collect all ridership information. Their daily, peak-period ridership data is for the Sacramento Commuter Service for the years 2005, 2006, 2007, and 2008. This data was broken down by day and further separated by route. 2008 data was provided in the same format but in an electronic version.

3.3 Data Filtering

In order to modify ridership data provided by each transit agency a four-step procedure was followed for each agency.

3.3.1 General Procedure

Step 1: Filter data to include ridership only for lines which provide service to the Sacramento downtown core, as previously defined by cordon. Table 3.2 provides a list of each transit agency and their bus transit lines that provide service to the downtown:

Table 3.2: Transit Lines Servicing the Downtown Core

Transit Agency	Downtown-Servicing Lines
Regional Transit	2,3,6,7,11,15,29,30,31,33,34,36,38,50E,51,62,63,67,68,86,88,89,109
Yolobus	39,40,41,42A,42B,43,44,45,230,231,232,240,241
North Natomas	Eastside Route, Westside Route, and Central Route
Roseville Transit	AM Routes 1-8, PM Routes 1-8
Yuba-Sutter Transit	Sacramento Commuter Express

Step 2: Filter data according to Table 3.3 to include inbound ridership for the AM peak period, as defined by agency.

Table 3.3 AM Peak Period Definitions of Each Data Set

Transit Agency	AM Peak Period Definition
Regional Transit	6:30-9:00
Yolobus	Daily Data
North Natomas	*Route*: Eastside: 5:54-9:04, Westside: 6:00-7:44, Central: 6:03-9:04
Roseville Transit	5:00-9:00
Yuba-Sutter Transit	5:20-8:00

Step 3: Filter data according to Table 3.4 to include outbound ridership for the PM peak period, as defined by agency.

Table 3.4: PM Peak Period Definitions of Each Data Set

Transit Agency	PM Peak Period Definition
Regional Transit	3:00-6:00
Yolobus	Daily Data
North Natomas	Route: Eastside: 3:35-6:54, Westside: 4:30-6:30, Central: 4:07-7:06
Roseville Transit	3:30-7:30
Yuba-Sutter Transit	3:45-6:50

Step 4: Filter data to include only Tuesday, Wednesday and Thursday ridership. Because modified work schedules are widely used, Monday and Friday are not representative of typical ridership.

The final model ridership is a combination of ridership across bus lines for a given agency, so that a single data point represents total ridership on all lines serving the downtown core. The sample size for the final data sets for this analysis is described in Table 3.5:

Table 3.5: Sample Size of Each Data Set

Transit Agency	Time Period	Aggregation	Sample Size
Regional Transit	2008	Weekly, Peak Period	49
Yolobus	2006-2008	Daily	441
North Natomas	2008	Daily, Peak Period	147
Roseville Transit	2006-2008	Daily, Peak Period	441
Yuba-Sutter Transit	2006-2008	Daily, Peak Period	441

3.3.2 Special Modifications to General Procedure for Regional Transit

RT data required more manipulation in order to perform the necessary filtering. The details are described. The main objective was to use the APC data to obtain the total weekly demand within the downtown core for all 52 weeks in 2008. More specifically, the goal was to obtain this weekly demand data for buses entering the downtown during the AM peak (6:30 AM – 9:00 AM) and for buses leaving the downtown during the PM peak (3:00 PM – 6:00 PM). Because RT ridership data is not collected for every passenger or trip, more complex methods were needed for RT. All of the bus stops within the downtown core were identified using RT generated identifying numbers. RT uses 325 bus stops in the downtown. Then the number of boarding and alighting riders associated with those bus stops during each peak period was obtained. Although the daily APC data is incomplete, it covers almost half of the stops within the downtown area every day. We can assume that the total data collection for one week (Monday through Friday) covers all of the stops within the downtown area and that the sample size for each bus stop is sufficient (Drake, 2007).

Next, using the APC data, the daily average of alighting riders during the AM peak period and the daily average of boarding riders during the PM peak period was calculated. Using the RT bus schedule, the frequency of stops at each bus stop during the peak hours was determined. This frequency is a fixed number every day for a certain schedule. RT had four different schedules throughout 2008; however, comparison between schedules shows that the frequency of the downtown bus stops did not change for the downtown core for 2008. Therefore, this study used the frequencies from the first schedule, Schedule 20, valid between January 6, 2008 through April 5, 2008, for both the AM and PM peak periods. The following equations were used to calculate total ridership:

$$total\ ridership_{inbound} = \sum_{i=1}^{325} average\ "offs"\ per\ stop * frequency\ of\ stops\ at\ bus\ stop\ i\ during\ AM\ peak$$

$$total\ ridership_{outbound} = \sum_{i=1}^{325} average\ "ons"\ per\ stop * frequency\ of\ stops\ at\ bus\ stop\ i\ during\ PM\ peak$$

3.4 Independent Variables

A multiple regression analysis was performed on each of the nine ridership data sets to determine if any independent factors played a significant role in ridership changes during the period of analysis. The regression was performed for all agencies (and all peak periods) using four independent variables: GDP, unemployment rates, gas prices, and fare prices. The smallest period of data aggregation available was used for each independent variable. Table 3.6 describes the final independent variable data:

Table 3.6: Independent Variable Details

Independent Variable	Source	Aggregation	Location	Contact	Contact's Official Position Title
Gross Domestic Product	Bureau of Economic Analysis	Quarterly	National	Lisa Mataloni	Economist
Gasoline Prices	AAA	Monthly	Sacramento City	Michael Geeser	Media and Government Relations Representative
Unemployment Rates	Bureau of Labor Statistics	Monthly	Sacramento/Arden-Arcade/Roseville	Website	
Fares	Yuba-Sutter Transit	Daily	Agency	Dawna Dutra	Analyst
	Roseville Transit			Elizabeth Haydu	Administrative Technician

In terms of GDP data, seasonally unadjusted data was used because the adjustment of GDP data is outsourced, and the Bureau of Economic Analysis doesn't provide or have access to unadjusted GDP data. Additionally, state and metropolitan area GDP is only available on an annual basis and the lowest level of aggregation is national GDP provided on a quarterly basis. GDP was included as a measure of overall economic health. Hoel (1971), in his discussion of linear regression, gave the example of a 0.98 correlation coefficient between teacher's salaries and liquor consumption, noting that in general the economy was doing well and upward trends were common. He warned about spurious correlations which must be considered in correlational studies.

Gas price data was the unleaded gasoline price per gallon averaged for the city of Sacramento between 2006 and 2008. Finally, fares were used for two transit agencies who were affected by changes in basic fare rates namely Yuba-Sutter Transit and

Roseville Transit, with one increase for the three year period (2006-2008) for both agencies. Table 3.7 describes the fare pricing for each agency:

Table 3.7: Fare Pricing Details

Transit Agency	Single Ride, Adult Fare
Regional Transit	$2.00
Yolobus	$1.50
North Natomas	$1.00
Roseville Transit	11/1/2003 – 6/30/2007: $2.75, 7/1/2007 –12/31/2008: $3.25
Yuba-Sutter Transit	8/1/2002 – 6/30/2007: $3.00, 7/1/2007 – 12/31/2008: $3.50

For data aggregated on levels other than daily, the monthly or quarterly average value of the independent variable was repeated for all Tuesdays through Thursdays that existed for that month or quarter (based on the information from the data manipulation section). Consequently, there is a single value for every day of ridership data that represents that month's average of the independent variable. For weeks that straddled two months, the two monthly averages were averaged. For example, Week 17 of 2008 includes April 29, April 30 and May 1, and the unemployment rate for April 2008 is 5.9% while May's unemployment rate is 6.3%. The unemployment rate for this week is calculated as [(2*5.9) + 6.3] / 3 = 6.03333%.

3.5 Data Quality[2]

This section describes data quality considerations, including sub-sections describing quality concerns related to the four ridership data collection methods employed by the five transit agencies, as well as additional data quality concerns. The four collection methods consist of automatic passenger counting (APC) devices, electronic registering

---

[2] Sections 3.5.1, 3.5.2, 3.5.3, and 3.5.4 use information from a report prepared by Jessica Seifert, under the author's direction.

fareboxes (ERFs), manual counts by route checkers, and manual counts by bus drivers. Each section will briefly discuss the collection method, concerns related to data quality and which agencies use that method. All agencies ridership data sets were shortened to Tuesday, Wednesday, and Thursday data sets which did not contain any missing data. The two missing values for Yolobus, discussed previously, July 30, 2006 and July 31, 2006, fell on Sunday and Monday.

3.5.1 Automatic Passenger Counting Devices

APC devices automate ridership data collection by tracking boarding and alighting riders, in addition to including a time and location stamp for each count. RT is the only transit agency within the study that utilizes APC devices, supplied by Clever Devices, Inc. (Drake, 2009). This technology uses an infrared beam to count boarding and alighting riders, and is mounted above the bus doors (Poggioli, 2009). The Clever Devices APC correlates the ridership data to GPS coordinates and scheduled routes so that the data may be viewed on a per-bus, per-door level (Clever Devices, 2009).

Clever Devices, Inc. claims that their APC system demonstrates over 95% accuracy, though they do not provide information on their website that would account for the 5% error (Clever Devices, 2009). Boyle (1998), referring to all APC systems, stated that typically the most common problems are related to software, as transit agencies often have to upgrade their analytical programs, and secondarily hardware problems (device failure and durability). But for the Clever Devices APC system, in large part, the 5% error can be attributed to mechanical malfunctions as well as door bunching, carrying a child, carrying large bags, drivers getting on and off the bus, non-riders making inquiries to bus drivers, and misalignment of sensors (Poggioli, 2009). In addition to technical problems, Boyle (2008, p. 18) defines a "debugging" period in which employees must familiarize themselves with the new technology. From the survey Boyle conducted in 1998, the average debugging period for APC devices was 17 months (Boyle, 2008). The accuracy of APC systems can be evaluated by comparing its ridership data to manual counts, although manual counts may also have data quality problems (see Sections 3.5.3 and 3.5.4) (Boyle, 2008).

As mentioned, RT is the only agency that uses APC devices to collect ridership information. RT was unable to provide APC ridership data for all of the buses that serve the downtown Sacramento region, because APC devices were not installed on the entire bus fleet and because the data was heavily filtered to remove data with obvious errors (see Section 3.2.1 for filtering rules). Also, RT's APC system is still in testing phases which could indicate that the devices are also within the debugging period (Drake, 2009).

3.5.2 Electronic Registering Fareboxes

Electronic registering fareboxes (ERFs) are devices in which bus drivers enter a number corresponding to rider type into a key pad that connects to an electronic farebox (Boyle, 1998). The drivers are also required to enter a value to indicate the route and run number at the beginning of each trip (Boyle, 1998). ERFs do not collect location information, so ridership data is only available at the trip level (Drake, 2009). As is done with APC devices, the data collected from electronic registering fareboxes can be "validated" by a comparison with manual counts or by comparison with the revenue collected from fares (Boyle, 1998).

There are four problems that may be encountered when using electronic registering fareboxes: mechanical problems, operator compliance, software problems, and accuracy of data (Boyle, 1998). The bus operators must enter the correct codes at the beginning of each route and trip and the correct code for the type of passenger (Boyle, 1998). Boyle's survey (1998) indicates that some transit agencies experienced difficulties when adding these additional responsibilities to the bus drivers' duties, although the most successful agencies were the ones that provided continuous ERF training to their drivers. Ultimately, the quality of the data collected from ERFs is affected both by human and software errors.

The transit agencies within this study that used ERF are RT, Yolobus, and North Natomas TMA. RT has electronic fareboxes installed on all of their buses except for the community buses.

3.5.3 Manual Counts by Route Checkers

Most transit agencies utilize manual counts either as their primary method of data collection, or for comparison against electronic methods (Boyle, 1998). Route checkers ride the transit vehicle and take manual counts of passengers boarding and alighting at each stop (Boyle, 1998). They typically have preprinted forms or handheld units that contain all of the stops on that route, with the sole responsibility to count passenger and record bus stop arrival and departure times (Boyle, 1998). Manual counts are the most well-established method of ridership data collection (Boyle, 1998).

The following problems are associated with manual counting by route checkers: accuracy of data, consistency of data, labor intensiveness, reliability of route checkers, and cost of manual counting (Boyle, 1998). Problems with accuracy and consistency of the data are a result of the training and reliability of the route checker, as well as transcription of the handwritten record to an electronic version (Boyle, 1998).

RT and Natomas TMA were the only transit agencies within the study that used manual counts by route checkers to collect ridership data. It should be noted that Natomas TMA used untrained volunteer riders to provide manual counts.

3.5.4 Manual Counts by Bus Drivers

Manual counts by bus drivers are another method of ridership collection. Manual counting by bus drivers is concerned with many of the same problems as manual counting by route checker, including the labor intensiveness and reliability of the counter. But because bus drivers also have many other responsibilities such as driving the bus, monitoring passengers and collecting fares, they may be less focused on counting passengers than route checkers.

All five of the transit agencies within the study use manual counts by bus drivers as either their primary method of data collection or in combination with another technique. Roseville Transit and Yuba-Sutter Transit exclusively use manual counts by bus drivers to collect ridership information, while North Natomas TMA and Yolobus use manual counts by bus driver in addition to electronic registering fareboxes. RT uses manual counts by bus driver in addition to the other three techniques.

3.5.5 Additional Data Quality Considerations

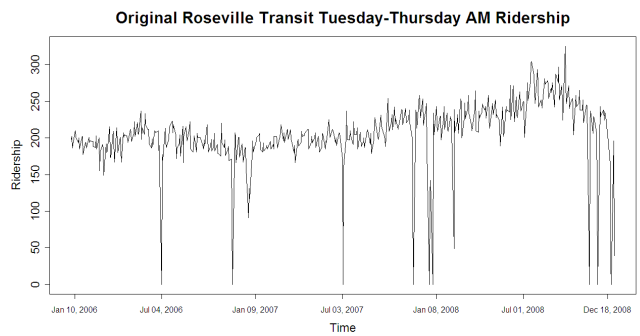
As discussed above, the RT data was heavily filtered and manipulated prior to being obtained by this study. Although quantification of data quality is not possible, RT data is probably the least reliable of the five agencies analyzed. RT ridership data was only a sample of total ridership (unlike all other agencies), and further the collected data was not a random sample. Additionally the system was still in a testing phase. Furthermore, this analysis did not separate riders who board and alight within the downtown core. According to RT, ridership that fell within these categories was less than 5% of total ridership for commute periods, but other ridership data was not available to verify this.

Although the Yolobus data was probably more reliable than RT, their daily ridership totals included regular bus routes (40, 41, 42A/B, 240) which operated all day during weekdays, in addition to commute and express services which only run Monday through Friday during peak commuting periods. The Yolobus ridership sample therefore included some non-commute data.

Since the data from RT and Yolobus was received in an electronic format, there is a possibility of transcription errors on the part of the transit agency. The data from Yuba-Sutter Transit, Roseville Transit, and North Natomas T.M.A. was received in a hardcopy format. There was also a possibility of transcription errors in entering that hardcopy data into data sets used by this study, although all data entry was verified for accuracy by a second person.

3.6 Data Cleaning

The data sets provided by each of the five agencies contained only two cases of actual missing data, both for Yolobus (July 30, 2006 and July 31, 2006). Although cases of missing data were rare, plots of the data that was provided by each agency indicated that some data manipulation would be necessary to account for holidays and limited service days. As an example, Figure 3.1 below displays the original data for Roseville Transit:



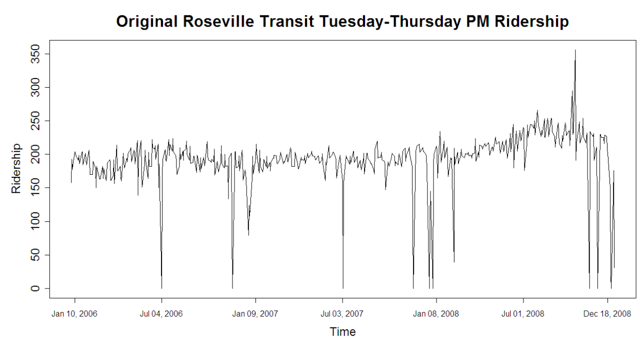


Figure 3.1: Plots of Original Roseville Peak-Period Ridership Data

Plots of each agency's original data set and imputed data set including Tuesday, Wednesday and Thursday ridership can be found in Appendix B. The drops in the plots represent transit holidays and limited service days as well as state holidays. Although not technically missing data, because agencies had provided data for all observations, the buses and the riders (assumed to be workers in downtown Sacramento) were "missing" and therefore ridership was zero (or very low) for transit and state holidays, and unusually low for limited service days. Those occurrences were treated as missing data.

For some missing observations, the missing data could be considered missing completely at random (MCAR). MCAR occurs if missing observations are distributed randomly over all observations, including that variable and any others, and can therefore be considered a simple, random subsample (Allison, 2002). The missing data in this analysis are MCAR, although not ignorable. As discussed in the Literature Review, discrete time series data assumes that the time series is observed at equal intervals. More complex methods are necessary if the observations are not equally spaced, and therefore the missing data in this analysis had to be imputed. There are multiple methods to deal with missing data. Some conventional methods, excluding listwise and pairwise deletion, include dummy variable adjustment, and imputation. A basic dummy variable regression was first used, but an ad hoc imputation method was ultimately used because of the detailed information about the missing value cases and their likely "true" values.

Prior to any data imputation, it was necessary to identify days with no transit service, limited transit service days, and full transit service days that coincide with state holidays for each of the five transit agencies for 2006, 2007 and 2008. Those dates are considered missing observations, and are identified in Appendix D.

From Yolobus data exploration, in general, there was low variation from the mean for Tuesday, Wednesday and Thursday ridership for any given week. However, it also appeared that there was low variation from the mean for the same weekday ridership for three consecutive weeks. For example, the second Tuesday in a month showed similar ridership to the first and third Tuesdays in that month. Therefore, two methods for imputing data for "missing" observations were compared. The methods were tested using Tuesday, Wednesday, and Thursday Yolobus ridership data. The same set of holidays was used to test both methods. The two methods are described below, and the detailed calculations are given in Appendix E:

- Method 1 used the same week that the holiday falls in but different days. T1 is defined as the ridership of the first non-holiday day in the holiday week, and T2 is defined as the ridership of the second non-holiday day in the holiday week. For example, if the holiday fell on a Wednesday, T1 was the ridership on Tuesday and T2 was the ridership on Thursday, whereas if it fell on Tuesday, then T1 applied to Wednesday and T2 to Thursday of the same week. Then the absolute value of the difference, |T1-T2|, was calculated. The differences for all of the holidays were summed and divided by the total number of holidays, giving the average difference. This difference was found to be equal to 152.33.
- Method 2 uses the weeks prior to and after the holiday week but the same day. T1' was defined as the ridership on the same day of the week before the holiday, and T2' was defined as the ridership on the same day of the week after the holiday. For example, if the holiday fell on a Tuesday, T1' was the ridership of the previous Tuesday and T2' was the ridership of the following Tuesday. Then the absolute value of the difference, |T1'-T2'|, was calculated. The differences for all of the holidays were summed and divided by the total number of holidays, giving the average difference. This difference was found to be equal to 154.17.

Since the average difference of Method 1 was smaller than the average difference of Method 2, Method 1 was used for this study. More specifically, the average of T1 and T2, which lie in the same week as the day with the holiday, was used to impute the missing day's ridership. In addition, there were no problems with Method 1 when holidays occurred in consecutive weeks, for example, Christmas Day and New Year's Eve. Data imputation was done using Method 1 for the days that each transit agency ran limited or no services as well as state holidays when they ran full services. Finally, Thanksgiving, Christmas, and New Year's Eve weeks were eliminated from the data as those entire weeks showed extremely low ridership.

The formula used for percent data imputed is % Imputed = (Number of Days Imputed/Total Number of Days) x 100. There were no differences between holidays, or limited service days, so the percent of data imputed for agencies with separate AM and PM peak data sets was constant. Table 3.8 shows that the amount of imputed data for any given agency is at most 2%, and usually much less. This is considered an acceptable level of imputation.

Table 3.8: Percent Imputed Data

Transit Agency	**Percent**
Roseville Transit	0.91%
Yuba-Sutter Transit	0.91%
Yolobus	0.68%
North Natomas T.M.A.	1.36%
Regional Transit	2.04%

3.7 Descriptive Statistics for Transit Ridership[3]

The statistical methods discussed in the literature review are considered parametric statistical methods. Parametric methods make assumptions about the population parameters, more specifically probability distributions are usually assumed to be normal (Mann, 2004). The statistical tests used in this study, including the regression and time series analyses presented later, use parametric methods. The following discussion provides a general statistical overview of each transit agency's ridership data, based on the cleaned data sets, prior to in-depth time series analysis. Both measures of center tendency and measures of dispersion will help to describe the data and its distribution. This section will present statistics, but leaves the interpretation of the statistics to Chapter 5.

3.7.1 Measures of Central Tendency

Measures of center value describe the center of the distribution of a variable. The mean is an arithmetic average which is commonly used to describe distributions. However, the mean statistic is sensitive to extreme values, also known as outliers (Ross, 2005). The median is also a measure of center value, and describes the middle value of the data without being as affected by outliers (Ross, 2005). In order to describe the center values of each data set, Table 3.9 lists the mean and median ridership for each transit agency's data sets.

[3] Section 3.7 makes use of calculations and tables created by Jessica Seifert.

Table 3.9: Measures of Central Tendency: Mean and Median

Transit Agency	Data Aggregation	Mean Ridership	Median Ridership
Roseville Transit	AM	213.9	208.0
	PM	200.4	198.0
Yuba-Sutter Transit	AM	239.2	226.0
	PM	237.5	222.0
Yolobus	Daily	3499.5	3383.0
North Natomas TMA	AM	116.8	127.0
	PM	96.36	96.0
RT	AM	15498.4	15314.0
	PM	13639.6	13641.0

A comparison of the median and the mean provides insight into the shape of the data sets distributions. If the median and mean have similar values, then the distribution is probably symmetric; otherwise, the data may be to some degree skewed (Ross, 2005). The data for this analysis shows that the medians for each agency are similar to their means. This indicates that the distributions of the ridership data are fairly symmetric. More specifically, the medians for Roseville Transit, Yuba-Sutter Transit, Yolobus, and Regional Transit are slightly less than their means, indicating that the distributions may be skewed to the right. Part of the skew in the histograms of Roseville Transit (AM peak period), Yuba-Sutter Transit, and Yolobus can be attributed to slightly higher ridership on Tuesdays compared to Wednesdays and Thursdays. The Roseville and Yuba-Sutter histograms are shown in Figure 3.2, confirming that expectation. The histograms of all data sets are given in Appendix F.

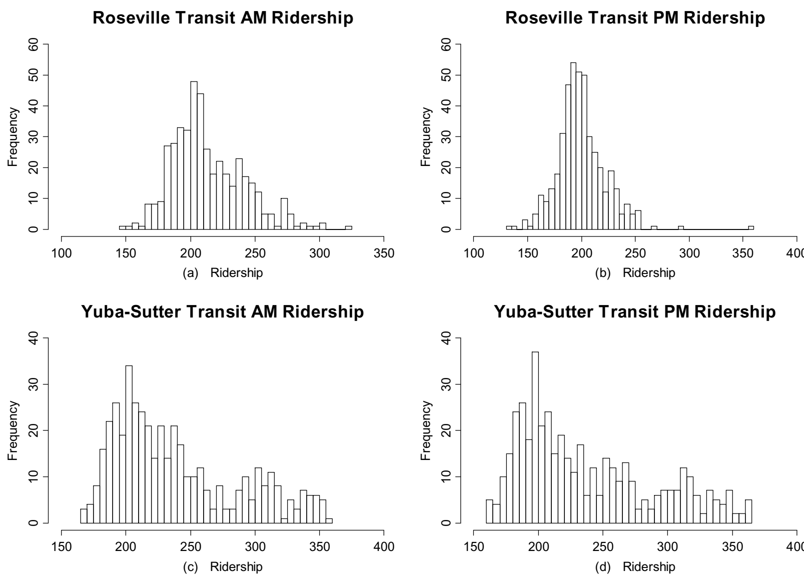


Figure 3.2: Roseville and Yuba-Sutter Transit Histograms

Roseville Transit, Yuba-Sutter Transit, and Yolobus that have data sets spanning the period of 2006 to 2008. All three agencies experienced increased ridership in both 2007 and 2008. In particular, Yuba-Sutter Transit experienced high ridership increases; from 2006 to 2007, average AM ridership increased by 15.2% and from 2007 to 2008, it increased by 31.5%. Similar changes were seen in Yuba-Sutter Transit's PM ridership during those years. The medians in Table 3.10 are much closer to their means. In fact, all agencies display this tendency, indicating that the yearly distributions are much more symmetric than the distributions of the entire data sets.

Table 3.10: Yearly Means and Medians for Transit Agencies with Data Spanning 2006-2008

Transit Agency	**Data Aggregation**	**Mean**			**Median**		
		2006	2007	2008	2006	2007	2008
Roseville Transit	AM	194.9	205.6	241.2	196.0	203.0	240.0
	PM	189.1	192.5	219.6	190.0	194.0	217.0
Yuba-Sutter Transit	AM	195.6	225.4	296.5	195.0	226.0	300.0
	PM	189.9	223.0	299.5	189.0	222.0	302.0
Yolobus	Daily	3175.1	3346.1	3977.2	3188.0	3360.0	3932.0

3.7.1.1 Ridership Means by Period

Since not all data sets span multiple years, the means of the data were also calculated seasonally for the year 2008. The calculations were made based on the following seasons:

- 1st quarter: January – March
- 2nd quarter: April – June
- 3rd quarter: July – September
- 4th quarter: October – December

RT was excluded as its data is observed weekly. All of the agencies experienced increased ridership between the first and second quarters of 2008. North Natomas TMA ridership increased the most during this period, with a 41.6% increase in AM ridership and a 24.9% increase in PM ridership. Similarly, all agencies saw an increase in ridership between the second and third quarters of 2008, with North Natomas TMA again showing the largest increase. However, opposite changes occurred between the third and fourth quarters of 2008. Almost all of the agencies experienced a decrease in ridership during this period; Yolobus was the only agency that saw an increase in ridership (1.7%). The means for each quarter of 2006, 2007 and 2008 are displayed in Table 3.11. In general, it appears that transit ridership decreased in the first and fourth quarters.

Table 3.11: Means by Season for 2006, 2007 and 2008

Transit Agency		Roseville Transit		Yuba-Sutter Transit		Yolobus	North Natomas TMA	
Data Aggregation		AM	PM	AM	PM	Daily	AM	PM
2006 Mean Ridership	1st	190.3	182.4	189.7	182.9	3270.0		
	2nd	203.2	190.7	190.7	184.9	3185.0		
	3rd	197.2	196.2	201.9	196.2	3090.6		
	4th	187.3	186.0	200.6	196.0	3161.8		
2007 Mean Ridership	1st	195.1	191.4	213.4	206.8	3333.8		
	2nd	200.3	192.6	219.9	214.1	3269.8		
	3rd	201.0	188.6	229.0	225.0	3385.3		
	4th	228.6	198.1	240.8	248.8	3403.5		
2008 Mean Ridership	1st	226.3	199.6	253.9	257.2	3423.9	77.9	74.8
	2nd	235.1	213.3	288.8	292.3	3782.8	110.3	93.4
	3rd	266.0	234.6	333.8	335.7	4326.6	146.8	114.7
	4th	234.5	231.3	307.2	310.8	4399.9	130.8	101.3

Means and medians were also calculated based on the Fix I-5 construction period. The three periods in Table 3.12 represent the time before the construction (January 1, 2008 – May 30, 2008), the time during the construction (May 31, 2008 – July 27, 2008), and the time after the construction (July 28, 2008 – December 31, 2008). All of the agencies experienced increases in mean ridership between the pre-construction and construction periods, but the changes in the ridership from the construction to post-construction periods varied by agency and by peak period within agencies. However, these differences are confounded with seasonal differences, as the previous table had shown.

Table 3.12: Means by Construction Period for 2008

Transit Agency	Data Aggre-gation	Mean Ridership			Median Ridership		
		Pre	During	Post	Pre	During	Post
Roseville Transit	AM	228.2	253.8	249.8	229.0	250.5	246.0
	PM	204.3	226.0	233.2	204.0	230.5	230.0
Yuba-Sutter Transit	AM	265.8	314.0	321.8	262.0	311.5	316.5
	PM	268.8	317.2	324.8	268.0	318.5	320.5
Yolobus	Daily	3563.1	4023.5	4393.5	3589.0	3973.0	4447
North Natomas TMA	AM	85.2	146.2	138.1	82.0	148.5	138.5
	PM	76.6	124.3	106.0	75.0	127.0	104.5
RT	AM*	14785.9	15525.8	16235.7	14990.4	15132.5	16423.0
	PM*	13361.6	13907.5	13824.4	13221.9	14188.6	13966.6

* RT AM and PM peak period ridership represents weekly ridership counts.

3.7.2 Measures of Dispersion

But measures of central tendency do not give a complete picture of the data's distribution; measures of dispersion are also included as descriptive statistics and include the standard deviation. The sample variance, s^2, is the average of the squared deviations from the sample mean, $\bar{X}$, while the sample standard deviation, s, is the square root of the variance (Ross, 2004). The relative size of the standard deviation can provide information about how tightly clustered the data are about the mean. Smaller standard deviations indicate that the data are tightly clustered whereas larger standard deviations indicate that the data are relatively more dispersed (Mann, 2004). The standard deviation, together with the mean, can be used to calculate a range in which a certain percentage of the data can be expected to lie: the confidence interval (Ross, 2005). This range provides values in terms of the original data's units that indicate how much of the data are "normally" contained in that range. Standard deviations (s) by construction period are given in Table 3.13.

Table 3.13: Variance and Standard Deviation for Each Transit Agency

Transit Agency	Data Aggregation	Standard Deviation (s)			$\bar{X} \pm s$	$\bar{X} \pm 2s$	$\bar{X} \pm 3s$
		Pre	During	Post			
Roseville Transit	AM	20.54	24.91	24.46	(186.0, 241.8)	(158.1, 269.7)	(130.2, 297.6)
	PM	15.49	20.17	23.49	(177.7, 223.1)	(155.0, 245.8)	(132.3, 268.5)
Yuba-Sutter Transit	AM	29.14	16.33	21.38	(191.4, 287.0)	(143.6, 334.8)	(95.8, 382.6)
	PM	33	19.7	23.02	(185.7, 289.3)	(133.9, 341.1)	(82.1, 392.9)
Yolobus	Daily	240.38	209.77	229.1	(3043.5, 3955.5)	(2587.5, 4411.5)	(2131.5, 4867.5)
North Natomas TMA	AM	12.14	11.65	12.48	(86.7, 146.9)	(56.6, 177.0)	(26.5, 207.1)
	PM	8.75	12.37	12.42	(75.2, 117.6)	(54.0, 138.8)	(32.8, 160.0)
RT	AM	992.62	1265.84	1119.38	(14238, 16759)	(12977, 18019)	(11717, 19280)
	PM	687.33	1043.77	793.33	(12824, 14455)	(12009, 15270)	(11193, 16086)

According to the empirical rule, the following percentages of approximately normal data lie in these respective ranges: 68% in the range $\bar{X} \pm s$, 95% in the range $\bar{X} \pm 2s$, and 99.7% in the range $\bar{X} \pm 3s$ (Ross, 2005). These ranges were calculated for the data sets and are displayed in Table 3.13. The $\bar{X} \pm 3s$ range covers 100% of the data in all but two cases (with Roseville Transit AM and PM data sets containing points that lie outside of the $\bar{X} \pm 3s$, 99.7% range), indicating that the data sets are approximately normal.

3.7.3 Discussion

All agencies had overall increases in mean ridership during the study period, but there were also seasonal variations in mean ridership. An informal analysis of data dispersion indicated that the data sets were approximately normal, with minor skews. Although this study's data failed usual tests of normality, slight departures from normality do not cause serious issues (Kutner et al., 2005) With the possible exception of RT, this study's data sets are random samples with a sufficiently large number of observations. Their populations were considered approximately normally distributed, and parametric methods were justified.

The next Chapter, which uses multiple regression and time series analyses, studies the transit agency data sets to identify independent variables that correlate with increased ridership and which can be used in predictive models to explain the change in ridership means during the Fix I-5 project.

CHAPTER 4 MODEL BUILDING

This chapter describes the methodology that was used to create the time series intervention models for each agency's transit ridership data. The first two sections present the steps taken to transform each of the nine data sets into stationary processes, including detrending using multiple regression analysis, and eliminating seasonal components using sinusoidal decomposition. The last section explains the intervention analysis methodology.

4.1 Multiple Regression

The nine time series plots shown in Appendix B show an overall increasing trend in ridership. As discussed in the Literature Review Section, there are multiple methods of removing trend components in the time domain including least squares estimation, smoothing with moving averages and differencing, as well as regression techniques (Aue, 2009; Yaffee, 2000). Regression techniques allow the modeler to eliminate trends using independent variables. As discussed earlier, each previous study used a different set of independent variables to predict transit ridership, but most of the studies that used multiple regression included gas prices, fares, and economic indicators such as unemployment rates. The following sections describe the relationships between bus transit ridership and each independent variable used in this multiple regression. Plots of each independent variable can be found in Appendix C.

4.1.1 Bus Transit Ridership and Gas Prices

Many studies have shown that gas prices significantly affect ridership, and that the ridership-gas price correlation is positive. Lane (2009) showed that gasoline prices are a statistically significant predictor of positive changes in transit ridership, with positive ridership-gas correlations. Wang and Skinner (1984) used data from seven transit authorities in the U.S. and showed that as real gasoline prices increase, transit ridership increases significantly. In a cross-sectional study, Taylor et al. (2009) analyze transit ridership from 265 urban areas using regional fuel prices provided by the Bureau of Labor Statistics as an explanatory variable and hypothesized a positive correlation. They found that fuel prices were a significant external factor positively influencing aggregate transit ridership. Kyte et al. (1988) shows that gasoline price is a statistically significant predictor of bus transit ridership, explaining that increasing the cost of automobile travel (i.e. gas prices) would motivate a mode change to transit. They find that gasoline prices show a negligible lag in their influence on ridership. For the previous work presented above, those studies that used gas prices in their analysis found them to be significant independent variables.

This study's data found strongly significant and positive Pearson's correlations between bus transit ridership and gas prices ranging between 0.18 and 0.6 for eight of the data sets. The Pearson's correlation coefficient, r, is defined for pairs (x_i, y_i) as (Ross, 2005):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \overline{y})}{(n-1)s_x s_y}.$$

The data sets with the highest correlations (Roseville Transit, Yuba-Sutter Transit and Yolobus) are those having the longest time series, suggesting that perhaps the impact of higher gas prices was beginning to level off by 2008 (i.e. those who were susceptible to the effect of higher prices had already changed earlier than 2008), or possibly that the intervention of the Fix I-5 project and other anomalies of 2008 (economic conditions, serious regional fires during the summer) disrupted the previously regular relationship between gas prices and ridership. Also, the highest correlations between ridership and gas price were for bus transit agencies farthest from the Sacramento downtown core (Yuba-Sutter Transit, Roseville Transit and Yolobus) indicating that commuters with longer commute distances may have been more sensitive to rising gas prices, and therefore, more inclined to use bus transit. The RT AM peak has a counterintuitive negative ridership-gas correlation of -0.2. The ridership-gas Pearson's correlations are shown in Table 4.1.

Table 4.1: Ridership-Gas Price Correlation Coefficients

Transit Agency	**Peak Period**	**Pearson's Ridership-Gas Price Correlation Coefficient**
Regional Transit	AM	-0.20***
Regional Transit	PM	0.18***
Yolobus	Daily	0.39***
North Natomas	AM	0.25***
North Natomas	PM	0.28***
Roseville Transit	AM	0.60***
Roseville Transit	PM	0.43***
Yuba-Sutter Transit	AM	0.59***
Yuba-Sutter Transit	PM	0.59***

***: $p < 0.001$

4.1.2 Bus Transit Ridership and Unemployment Rates

A small number of previous studies have examined the effects of labor statistics, specifically employment, on transit ridership. Agrawal (1981) showed that jobs in

Philadelphia were highly significant positive predictors of transit ridership. Surprisingly, he showed that 250 transit trips per year were created for every nonagricultural job in Philadelphia- about one trip per job per workday! He showed a positive correlation between jobs and transit trips. Kyte et al. (1988) used county employment rates but found it was not a statistically significant predictor of bus transit ridership. They theorized that a cause was the high percentage of his sample that was student riders. Employment showed a negligible lag in terms of its impact on ridership. Interestingly, Agrawal (1981) noted that he initially used both the number of jobs in Philadelphia as well as the unemployment rate, but found that they had the same effect, and therefore dropped the unemployment rate variable as it was only available for a larger area. None of the other studies discussed in the Literature Review included labor statistics in their analysis.

This study's data found positive Pearson's correlations between bus transit ridership and unemployment, with rates ranging between 0.51 and 0.88 across eight of the data sets. The ninth data set, the RT PM peak period, has a 0.19 ridership-unemployment correlation coefficient possibly indicating that the unemployed tend to not use transit during the PM peak period as much as the AM peak period. The ridership-unemployment Pearson's correlations are shown in Table 4.2.

Table 4.2: Ridership-Unemployment Correlation Coefficients

Transit Agency	Peak Period	Pearson's Ridership-Unemployment Correlation Coefficient
Regional Transit	AM	0.51***
Regional Transit	PM	0.19***
Yolobus	Daily	0.85***
North Natomas	AM	0.64***
North Natomas	PM	0.52***
Roseville Transit	AM	0.69***
Roseville Transit	PM	0.64***
Yuba-Sutter Transit	AM	0.88***
Yuba-Sutter Transit	PM	0.88***

***: $p < 0.001$

4.1.3 Bus Transit Ridership and Gross Domestic Product

Although no known studies concerned with transit ridership used GDP in their analysis, GDP is a common variable for predicting VMT, so there is ample precedent for using it as a predictor of travel demand (Schafer, 1998; Schafer and Victor, 2000). GDP is an overall indicator of the economic well-being of the USA. It could be expected that as GDP followed an upward trend, and general purchasing power increased, transit ridership may also increase. But conversely, a substitution phenomenon may result. As GDP decreases, transit ridership may still increase. Commuters may decide to abandon their single-occupant vehicles for a less expensive mode to commute, by substituting transit as their main commute choice. Plausible explanations exist for both positive and negative correlations between ridership and GDP, and there are no previous studies that clarify that relationship. Because there isn't a comprehensive list of independent variables which have been identified that predict bus transit ridership, GDP was included as a potential independent variable.

4.1.4 Bus Transit Ridership and Transit Fares

Many studies have examined the effects of fare pricing on transit ridership, with many studies born in Economics and based on fare elasticities. As Cervero (1990) states, the primary focus of transit pricing research has been fare elasticity estimations. Transit planners commonly use the Simpson & Curtin rule which states there is a fare elasticity of -0.33, or in other words for a 10% increase in fares, transit ridership will decrease by 3.33% (Curtin, 1968). Agrawal (1981) analyzed full-fare adult ridership and found that average fare (adult riders) was statistically significant in affecting ridership. He found that for every 1% increase in fares, ridership decreased by 0.385%, very similar to the Simpson & Curtin rule. Wang and Skinner (1984) used data from seven U.S. transit authorities and found that as real fare increased, ridership decreased significantly. He noted that a more accurate measure of fare would be to create an index that includes different fare types and includes passes; however, he used adult cash fares. Taylor et al. (2009) analyze transit ridership from 265 urban areas and found a significant negative relationship between ridership and fares. Kyte et al. (1988) show that transit fares are statistically significant negative predictors of bus transit ridership, noting that there is a lag structure of three months. Gomez-Ibanez (1996) describes bus transit ridership as a function of transit fares. Narayan and Considine (1989) used a modified intervention analysis approach to model the effects of two cases of fare increases on an upstate New

York transit system and found that both fare increases were significant factors in the decrease of transit ridership.

With respect to this study's data, only Roseville Transit and Yuba-Sutter Transit had fare changes during the study period, and those series exhibited positive Pearson's correlations between bus transit ridership and fares. These results are not consistent with the literature which found negative correlations, and could be the result of a spurious correlation of the fare increase with the rising gas prices that were occurring during the same period (and leading to increased ridership). Fare increases for both agencies occurred at the end of 2007. The correlation between Roseville Transit fares and gas prices is 0.46, while the correlation between Yuba-Sutter Transit fares and gas prices is also 0.46, which are both significant at beyond the α=0.01 level. The ridership-fare Pearson's correlations are shown in Table 4.3.

Table 4.3: Ridership-Fare Correlation Coefficients

Transit Agency	Peak Period	Pearson's Ridership-Fare Correlation Coefficient
Roseville Transit	AM	0.65***
Roseville Transit	PM	0.45***
Yuba-Sutter Transit	AM	0.76***
Yuba-Sutter Transit	PM	0.79***

***: $p < 0.001$

Given the availability of data and hypothesis regarding affects, this study conducted multiple regression analysis using these four independent variables: monthly Sacramento unleaded regular gas price averages, national current dollar adjusted GDP data, monthly unemployment rates for the Sacramento-Arden-Arcade-Roseville, CA metropolitan area, and basic one-rider fare rates for those agencies whose fare pricing changed during the analysis period (Yuba-Sutter Transit and Roseville Transit). During exploratory model building, for each agency, those predictor variables significant at the α=0.1 level were included in this analysis, but all final coefficients are significant at the α=0.01 level, as shown in Appendix G.

4.2 Sinusoidal Decomposition

As previously discussed, Box and Tiao (1975) constructed an analysis method called intervention analysis to determine the effect on a time series of an external event occurring at a known time. The first step in their methodology is to transform the time series into a stationary process. To remove autocorrelation in the data, this study used a sinusoidal decomposition as opposed to the commonly used differencing techniques. Applying the lag l difference operator to the original series yields a loss of l observations in the sample. Sinusoidal decomposition, on the other hand, decomposes a time series into the sine and cosine functions based on the periodic components of the series. The periodogram was examined to identify the periodic components, and next those possibly genuine cycles were tested for significance using spectral ANOVA. For each data set, those cyclic components significant at the α=0.1 level were included in this analysis.

4.3 Intervention

Once the data sets were detrended and seasonal components were eliminated, the first step of the intervention analysis was complete. The basic intervention model can be described as:

$$Y_t = \sum_{j=0}^{\infty} \tau_j \, M_{t-j} + N_t$$

where M_t and Y_t are the input (pulse/step) and output (ridership, after removal of trend and seasonal components)) series of the model respectively, τ_j is a linear filter and N_t represents a noise sequence. For this analysis, a simplified rational operator was used which assumed no delay parameter and allowed only a step or pulse indicator series. The intervention model becomes:

$$Y_t = \omega M_t + N_t$$

where ω represents the change in mean ridership due to the intervention and M_t represents the indicator (0-1) variable. The type of indicator variable was based on data exploration, however after stationarity was achieved, no agency's plot displayed a visual change in mean near the construction period. To be flexible about the nature of the possible impact of the intervention, all series were modeled with both an intervention causing a temporary change in the response variable (a pulse indicator variable), and separately an intervention causing a permanent change in the response variable (a step indicator variable):

$$M_t = \begin{Bmatrix} 1\ if\ t = T \\ 0\ if\ t \neq T \end{Bmatrix} \qquad M_t = \begin{Bmatrix} 1\ if\ t \geq T \\ 0\ if\ t < T \end{Bmatrix}$$

where t is time, and T is the period of the intervention. The pulse indicator variable consisted of one pulse for the duration of the Fix I-5 construction period. The remaining four steps of the Box and Tiao intervention analysis methodology were followed for each data set. After the series was stationary, OLS regression was used to obtain a initial estimate of ω, which represents the transfer model. The initial estimate of ω provides a beginning point for the minimization discussed in Section 2.6. Next, the residuals from the OLS regression were modeled as an ARMA(p,q) process, which represents the noise model. For model diagnostics, the residuals were analyzed using goodness-of-fit tests. For all goodness-of-fit testing this study used the Sample ACF, qq plot, Ljung-Box test and rank test. Next, the sum of squares, $\sum_{t=m*+1}^{n}\left(\frac{\hat{\phi}^{N}(B)}{\hat{\theta}^{N}(B)}\right)^{2}(W, V, \phi^{N}, \theta^{N})$, was minimized in order to obtain final parameter estimates of both the noise and transfer model which were then combined into one intervention model. Then, the final model residuals were analyzed using goodness-of-fit tests. The measure of the change in ridership due to the intervention is ω.

CHAPTER 5 RESULTS

The main objective of this analysis was to determine the impact that the Fix I-5 project had on Sacramento-area bus transit ridership. Interventional analysis was used in order to assess the magnitude of the effect and the projected ridership change of the Fix I-5 project. First, the trend and seasonal components were eliminated using multiple regression and sinusoidal decomposition. Next, intervention techniques were used to determine the impact that the Fix I-5 project had on Sacramento-area transit ridership. The results of this analysis are discussed in the following sections.

5.1 Eliminating Trends: Details of Multiple Regression

Multiple regression was used to detrend each of the nine data sets. Using this technique, systematic trend components were eliminated using independent variables. Those independent variables were identified by examining past models of transit ridership, and identifying other significantly correlated variables. Multiple studies have shown that a variety of independent variables affect transit ridership. This analysis used monthly Sacramento unleaded regular gas price averages, national current dollar adjusted GDP data, monthly unemployment rates for the Sacramento— Arden-Arcade—Roseville, CA metropolitan area, and basic one-ride adult fare prices for Yuba-Sutter Transit and Roseville Transit whose fare pricing changed during the period of analysis. Initially, all models used the first three independent variables, with the exception of Yuba-Sutter Transit and Roseville Transit which also used the fourth (the fare pricing) independent variable. During exploratory model building, for each agency, those predictor variables significant at the α=0.1 level were included in this analysis, but all final coefficients are significant at the $\alpha=0.01$ level. Nine detrended models were created through the four-step iterative process expressed below:

1. Conduct multiple regression using all independent variables,
2. Remove independent variables that were significant at less than the $\alpha=0.1$ level,
3. Test for collinearity using variance inflation factors and condition index,
4. Remove independent variables that displayed variance inflation factors and condition indexes over 10 and 30 respectively.

As a result of the iterative process described above, GDP was removed as an independent variable. The reasons were as follows:

1. Multicollinearity can be formally detected using two separate statistical tests including the variance inflation factor (VIF) and the condition index. When testing for multicollinearity, the VIF for the GDP independent variable for all data sets was much higher than the VIF for all other independent variables, and often the reason for mean VIF values above 1 which can be an indication of multicollinearity. The GDP condition index for all data sets was above 30. The high values of both the variance inflation factors and condition indices suggest serious multicollinearity (Kutner et al., 2004).

2. "Informally, it has been observed that multicollinearity is sometimes manifested by having coefficients that are large in magnitude but opposite in sign, with correspondingly large standard errors, indicating that the impacts of two correlated variables on the dependent variable are largely counteracting each other" (Mokhtarian, 2009). In this analysis, the GDP independent variable frequently had regression

coefficients and pairwise correlation coefficients that exhibited opposite signs and large standard errors. Pearson correlations were compared to the regression coefficients for each of the independent variables for each data set. The signs between the regression and correlation coefficients for 4 of the 9 possible data sets were opposite for the GDP independent variable.[4] Furthermore, opposite coefficient signs were present for only 3 of 22 possible cases for the remaining independent variables (9 cases for gas prices, 9 cases for unemployment rates, 4 cases for fare prices).[5]

3. The adjusted R^2 value for the regression fit changed very little when including the GDP independent variable, indicating that it added little to the explanatory power of the model beyond the other variables included. In fact, for the Roseville Transit AM peak, the North Natomas AM peak and the Regional Transit AM peak, the elimination of GDP as an independent variable increased the adjusted R^2 value, indicating that the increase in variance explained with the inclusion of GDP was so small that it did not compensate for the penalty incurred due to the reduction in parsimony of the model. The results for the adjusted R^2 value are compared in Table 5.1. From Table 5.1, the RT adjusted R^2 values for both peak periods are smaller than the adjusted R^2 values for all other agencies, possibly due to data quality issues.

[4] These data sets were North Natomas AM peak, North Natomas PM peak, Roseville Transit PM peak, and Yolobus (daily ridership total).
[5] The three cases were fare for the Roseville Transit PM peak, gas price for the Regional Transit AM peak, and unemployment for the Regional Transit PM peak.

Table 5.1: Adjusted R^2 With and Without the GDP Independent Variable

Transit Agency	Data Aggregation	Adjusted R^2 (including GDP variable)	Final Adjusted R^2 (without GDP variable)
Yuba-Sutter Transit	AM	0.8805	0.8762
Yuba-Sutter Transit	PM	0.8788	0.8705
Yolobus	Daily	0.7510	0.7387
Roseville Transit	AM	0.6286	0.6287
Roseville Transit	PM	0.4769	0.4585
North Natomas TMA	AM	0.8583	0.8586
North Natomas TMA	PM	0.6880	0.6728
Regional Transit	AM	0.2294	0.2328
Regional Transit	PM	0.2108	0.1086

4. GDP is an overall indicator of the economic well-being of the USA. National GDP data is a broad measure of the country's economic activity and might be very different than the economic conditions in the area served by any transit district. No other transit studies reported results using GDP as an independent variable which may be due to the imprecision of GDP for a selected locality. As mentioned earlier, it is common to use it as an explanatory variable for models of VMT, but those models tend to be at the nation or state level, with annual observations, which is more closely matched to the spatial and temporal granularity at which GDP is available. As discussed previously, it could be expected that as GDP followed an upward trend, and as general purchasing power increased, transit ridership may also increase. But conversely, a substitution phenomenon may result. As GDP decreases, transit ridership may still increase. Commuters may decide to abandon their single-occupant vehicles for a less expensive mode to commute, by substituting transit as their main commute choice. Plausible

explanations exist for both positive and negative correlations between transit ridership and GDP, and there are no previous studies that clarify that relationship.

Once GDP was removed as an independent variable, problems with multicollinearity and opposite coefficient signs disappeared. Appendix G displays multiple regression model selection details for all agencies. Gasoline prices and unemployment rates were significant predictors of transit ridership for all agencies except Regional Transit. The significance of gasoline price and unemployment rate is consistent with the literature (e.g. Lane, 2009; Wang and Skinner, 1984; Taylor et al., 2009; Kyte et al., 1988; Agrawal, 1981). Rosevile Transit and Yuba-Sutter Transit fare increases were not statistically significant predictors of transit ridership, which was inconsistent with the literature (e.g. Agrawal, 1981; Wang and Skinner, 1984; Taylor et al., 2009; Kyte et al., 1988; Narayan and Considine, 1989). Table 5.2 summarizes the final regression models. The magnitudes of the parameter estimates of change in ridership for both the gasoline price and unemployment rate independent variables have a large range, as does the "average" ridership across the five agencies. For example, every $1 increase in gasoline price results in a 9.482 increase in riders for Roseville Transit PM peak period, whereas the same $1 increase in gasoline price results in a 30.108 increase in riders for Yuba-Sutter Transit PM peak period. This can be partly be explained because Roseville Transit's mean ridership is 200.4, while Yuba-Sutter's mean ridership is 237.5 (see Section 3.7.1).

Table 5.2: Statistically Significant Predictors of Bus Transit Ridership

Transit Agency	Peak Period	Final Model Variables	Parameter Estimate (t statistic)
Regional Transit	AM	none	No independent variables
Regional Transit	PM	none	No independent variables
Yolobus	Daily	Gasoline Price	97.37 (4.796)
		Unemployment Rate	335.99 (31.436)
North Natomas	AM	Gasoline Price	39.191 (28.73)
		Unemployment Rate	33.021 (21.63)
North Natomas	PM	Gasoline Price	22.100 (13.48)
		Unemployment Rate	23.967 (16.36)
Roseville Transit	AM	Gasoline Price	20.296 (13.71)
		Unemployment Rate	13.869 (17.80)
Roseville Transit	PM	Gasoline Price	9.482 (6.526)
		Unemployment Rate	11.512 (15.050)
Yuba-Sutter Transit	AM	Gasoline Price	27.611 (18.836)
		Unemployment Rate	33.486 (43.390)
Yuba-Sutter Transit	PM	Gasoline Price	30.108 (18.55)
		Unemployment Rate	36.038 (42.17)

Table 5.3 summarizes the means and medians of each transit agency after trend components were eliminated. As compared to Table 3.12, the means and medians of the periods before, during and after construction contain reduced trend components.

Table 5.3: Means by Construction Period for 2008 After Detrending Data

Transit Agency	Data Aggregation	Mean Ridership			Median Ridership		
		Pre	During	Post	Pre	During	Post
Roseville Transit	AM	230.84	259.04	245.25	230.72	259.04	247.10
	PM	211.21	228.26	225.83	209.84	228.26	226.70
Yuba-Sutter Transit	AM	270.58	320.20	313.11	266.60	320.20	315.64
	PM	271.44	325.18	317.07	267.03	325.18	319.83
Yolobus	Daily	3734.56	4077.30	4226.25	3757.55	4077.30	4234.43
North Natomas TMA	AM	86.98	144.00	136.50	82.08	139.80	139.53
	PM	78.15	114.64	107.75	75.10	112.22	109.77
RT	AM	14785.90	15525.80	16235.70	14990.40	15132.50	16423.00
	PM	13361.60	13907.50	13824.40	13221.90	14188.60	13966.60

5.2 Eliminating Seasonal Components in the Data: Details of Sinusoidal Decomposition

Sinusoidal decomposition was used to eliminate seasonal components in each of the nine data sets, based on the residuals of the multiple regression analyses. A seasonal component encompasses cycles at any recurrent period. As previously discussed, a time series can be represented as the sum of periodic components:

$$x_t = \sum_{j=1}^{k} A_j \cos(2\pi\omega_j t) + B_j \sin(2\pi\omega_j t) + Z_t$$

where x_t is the residual of the multiple regression described in Section 5.1, A_j and B_j are uncorrelated random variables with mean zero and variances equal to σ^2 and $\omega = \frac{1}{d}$, where d is the period of the cycle. The periodogram displayspeaks at the frequency where cyclic behavior occurs. An example of the Yuba-Sutter Transit AM peak period periodogram is displayed in Figure 5.1. The dashed lines are not typically included as part of the periodogram, but are shown here to indicate possible periodic components.

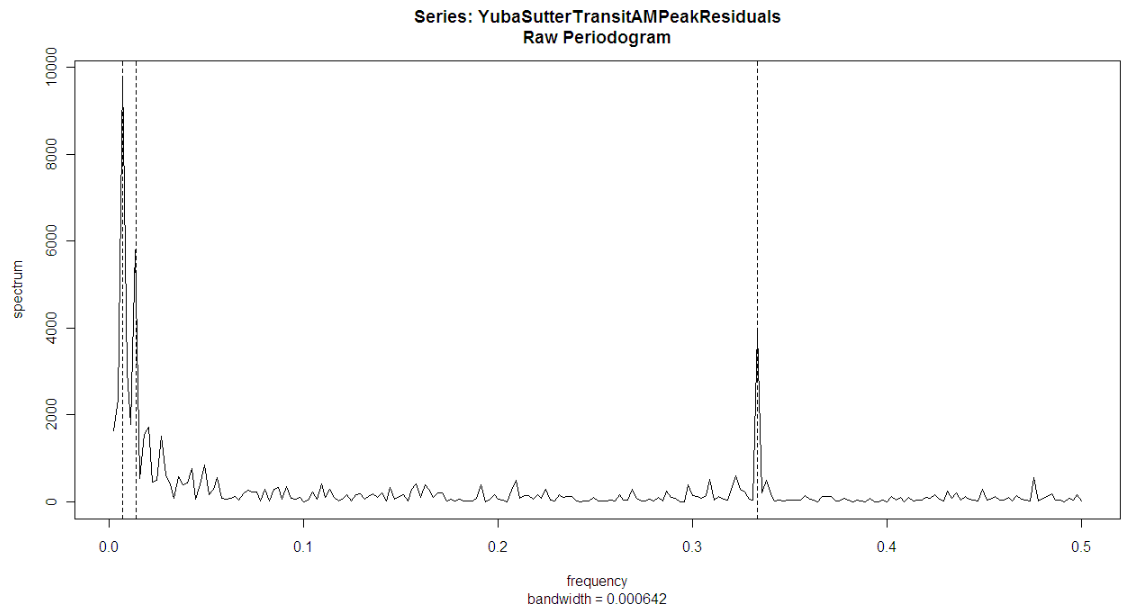


Figure 5.1: Yuba-Sutter Transit AM Peak Periodogram

Given that $\omega = \frac{1}{d}$, the potentially significant periods of the cycles in the Yuba-Sutter Transit AM peak period are 147, 74 and 3 days. Since the data set includes only Tuesday, Wednesday and Thursday data, the 3 day cycle is a surrogate for the weekly cycle. The 74 day cycle is a six month cycle, and the 147 day cycle is a yearly cycle. From the height of each peak, it is clear that the yearly cycle explains the most variance, followed by the half-yearly, then the weekly cycles. However, spectral analysis of variance (ANOVA) was used in order to test each cycle period for statistical significance. The percent of variance explained by each cycle can also be calculated from the results of the spectral ANOVA:

$$I(\omega_j) = d_c^2(\omega_j) + d_s^2(\omega_j)$$

where

$$d_c(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \cos(2\pi\omega_j t)$$

$$d_s(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \sin(2\pi\omega_j t)$$

where $I(\omega_j)$ is the definition of the periodogram, $d_c(\omega_j) = Re(d(\omega_j))$, and $d_s(\omega_j) = Im(d(\omega_j))$. For example, the Yuba-Sutter AM peak period, the results of the spectral ANOVA show that at least one of the components of each cycle, the sine or the cosine component, is significant at the α=0.001 level. The F statistic, which describes the explained variance over the unexplained variance, was used to assess the strength of each cycle. Table 5.4 shows the significant cycles within each data set as well as the variance explained due to each frequency.

Table 5.4: Statistically Significant Periodic Components for Each Agency

Transit Agency	**Data Aggregation**	**Significant Cycles (percent variance explained)**			
Yuba-Sutter Transit	AM (daily)	3(6.2%)	74(11.0%)	147(21.33%)	
Yuba-Sutter Transit	PM (daily)	3(4.32%)	73(15.04%)	147(23.88%)	
Yolobus	Daily	28(4.33%)	38(3.48%)	74(16.83%)	220(7.98%)
Roseville Transit	AM (daily)	3 (12.75%)	58(3.05%)	147(7.29%)	
Roseville Transit	PM (daily)	3(6.85%)	42(2.82%)	147(5.43%)	
North Natomas TMA	AM (daily)	3(3.76%)	15(4.10%)	19(6.63%)	30(13.39%)
North Natomas TMA	PM (daily)	3(6.98%)	8(4.10%)	25(12.86%)	74(16.99%)
Regional Transit	AM (weekly)	4.5(7.18%)	7.1(8.79%)		
Regional Transit	PM (weekly)	3.6(8.81%)	4.2(6.78%)		

The results show that all agencies with daily peak period data show significant weekly cycles, indicated as a three-day cycle. It is expected that a weekly cycle would be present in the bus transit data sets, because within each week there are recurring commute patterns. Although the data sets only include Tuesday, Wednesday, Thursday ridership, in general, Wednesday ridership tends to be higher than Tuesday and Thursday ridership. Pre-cleansed data showed strong weekly cycles with lower ridership on Monday and Friday and higher ridership on Tuesday, Wednesday and Thursday, possibly attributable to 9/80 work schedules (in which an employee works 80 hours in nine days instead of ten) or 4/40 work schedules (in which an employee works 40 hours in four days instead of five). Yuba-Sutter Transit's AM and PM peak periods show almost identical cycles, including weekly, half-yearly and annual cycles, which is an expected result. Because of additional riders in the mid-day period (between peak periods) for Yolobus ridership data, perhaps weekly commuting patterns are hidden. However, there are significant 9-week, three-month, half-yearly, and one-and-a-half year cycles. It should be noted that harmonics, which are multiples of the fundamental frequency, ω (such as 2ω, 3ω, 4ω etc.) were considered, but the one-and-a-half year cycle is not a harmonic of the yearly cycle, nor is the yearly cycle a harmonic of the half yearly cycle. Both of Roseville Transit's peak periods show weekly and yearly cycles. The AM peak period displays a 5-month cycle, while the PM peak shows a 3.5-month cycle. The results also show that North Natomas T.M.A.'s peak periods display very different cycles, with the exception of a weekly cycle displayed by both peak periods. The AM peak period shows a 5, 6, and 10 week cycle while the PM peak shows a 2, 8 week, and half-yearly cycle. Finally the results indicate both of the RT data sets show approximately monthly cycles, while the AM peak also shows a 7-week cycle. The periodogram of each data set is given in Appendix H.

5.3 Intervention Analysis: Details of the Fix I-5 Impact

Intervention analysis is a statistical method to determine the change in the mean level associated with the intervention event, assuming that the intervention event occurs at a known time. In this analysis, the intervention event is the Fix I-5 construction project. This analysis will determine if a temporary or permanent change in ridership better explains each data set, and the magnitude of the mean change in ridership.

Intervention analysis requires a stationary input time series process. In order to achieve stationarity, trend and seasonal components were removed as described above. Then a time series package, Interactive Time series Modeling (ITSM), was used for the intervention analysis. Intervention analysis takes the form:

$$Y_t = \omega M_t + N_t$$

where ω is the change in mean ridership (transfer model), M_t is the indicator function, and N_t is the noise (noise model). The analysis is initialized by assuming the noise is white noise. OLS regression provided a preliminary estimate of the change in mean ridership, ω. Next, the noise model, consisting of the OLS regression residuals, was modeled using a ARMA(p,q) model. Given a maximum and minimum p and q value, ITSM will process all possible combinations and provide the initial noise model with the lowest AICC value for the ARMA(p,q) model, represented as

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q}$$

The AICC criterion chooses p,q, ϕ_p, and θ_q while minimizing the AICC statistic (not the maximum likelihood estimators), and does not necessarily identify the best model fit, but suggests an initial model. Four goodness-of-fit tests, including the Sample ACF, the Ljung-Box modification of the portmanteau test, the rank test, and the qq plot, were used to assess the initial model fit, for example including residual whiteness. In the nine data sets, the AICC criterion produced models that passed the four goodness-of-fit tests at the $\alpha = 0.05$ level. An advantage of using ITSM is that it automatically places causality restrictions on model parameters (see discussion in Chapter 2). Another advantage of ITSM is that it is computationally efficient. The last step of the intervention analysis is a re-estimation, through minimization, of the parameters of the noise and transfer models. The final output, ω, is a number which explains the change in mean of the ridership time series. Both temporary and permanent changes were hypothesized, with goodness-of-fit tests determining which indicator better explained the change in ridership. The results of the intervention analysis for each data set are provided in Appendix I. The results of the comparison of goodness-of-fit tests for the final models, including both permanent and temporary indicator models, for each data set are provided in Appendix J. The three steps below summarize the methodology for the complete analysis of the Yuba-Sutter AM peak period. The steps are consistent for all agencies, but the specific equations will vary based on the specifics of each agency's data set.

1. Eliminate Trend Components:

$$X_t = \beta_0 + \beta_1 gas + \beta_2 unemployment + \varepsilon_t$$

2. Eliminate Seasonal Components:

$$\varepsilon_t = \mathrm{A}_1\cos\left(\frac{2\pi t}{3}\right) + \mathrm{B}_1\sin\left(\frac{2\pi t}{3}\right) + \mathrm{A}_2\cos\left(\frac{2\pi t}{74}\right) + \mathrm{B}_2\sin\left(\frac{2\pi t}{74}\right) + \mathrm{A}_3\cos\left(\frac{2\pi t}{147}\right) + \mathrm{B}_3\sin\left(\frac{2\pi t}{147}\right) + Y_t$$

3. Intervention Analysis:

$$Y_t = \omega M_t + N_t$$

where X_t is the cleaned ridership series, and M_t is the indicator function. Table 5.5 displays the final values for the change in mean ridership of each agency based on the results of the intervention analysis.

Table 5.5: Intervention Analysis Final Model Results

Transit Agency	**Data Aggregation**	**Indicator Function**	**Final Model Estimate ω (mean change in riders)**	**Distance to Sacramento Downtown (miles)***
Yuba-Sutter Transit	AM	Pulse	-7.4	
Yuba-Sutter Transit	PM	Pulse	-6.25	49.8
Yolobus	Daily	Pulse	-13.65	18.7
Roseville Transit	AM	Pulse	-1.15	
Roseville Transit	PM	Pulse	-3.75	19.1
North Natomas TMA	AM	Step	1.2	
North Natomas TMA	PM	Pulse	2.6	6.7
Regional Transit	AM	Step	552.6	
Regional Transit	PM	Step	351.1	3.7

* Distances were calculated from 9th and H Streets in Sacramento to downtown Sutter, the Yolobus offices in Woodland, Roseville downtown, and the North Natomas TMA office. The RT distance was calculated as an average of distances from McKinley Park, William Land Park and CSU Sacramento.

5.4 Significance of Results

Table 5.6 shows the intermediate adjusted R^2 values for the models of detrending, eliminating seasonal components (deseasoned), and intervention analysis, including both cumulative and incremental adjusted R^2 values. The value of R^2 increased with each step of the analysis, indicating that increasing amounts of variance in the original ridership data set were cumulatively explained at each step.

Table 5.6: Final Model Significance

Transit Agency	Data Aggregation	Cumulative Detrended Adjusted R^2	Cumulative Deseasoned Adjusted R^2 (Incremental Addition)	Cumulative Intervention Analysis Adjusted R^2 (Incremental Addition)	Final Model Adjusted R^2 and Significance
Yuba-Sutter Transit	AM	0.8762	0.9228 (0.0466)	0.9997 (0.0769)	0.9997***
Yuba-Sutter Transit	PM	0.8705	0.9255 (0.055)	0.9987 (0.0732)	0.9987***
Yolobus	Daily	0.7387	0.8206 (0.0819)	0.9951 (0.1745)	0.9951***
Roseville Transit	AM	0.6287	0.7105 (0.0818)	0.996 (0.2855)	0.996***
Roseville Transit	PM	0.4585	0.5339 (0.0754)	0.9635 (0.4296)	0.9635***
North Natomas TMA	AM	0.8586	0.8921 (0.0335)	0.999 (0.1069)	0.999***
North Natomas TMA	PM	0.6728	0.7955 (0.1227)	0.9961 (0.2006)	0.9961***
Regional Transit	AM	No Trend	0.0834 (0.0834)	0.991 (0.9076)	0.991***
Regional Transit	PM	No Trend	0.0792 (0.0792)	0.985 (0.9058)	0.985***

***: $p < 0.001$

The final adjusted R^2 value describes the amount of total variance that is explained by the intervention model (including the intermediate steps of eliminating trend and seasonal components). Based on percentage points for the distribution of R^2 (Shapiro and Francia, 1972), the results of all agency's models are statistically significant at the α=0.001 level. The highly significant final R^2 values from the final model indicate that almost all of the variance in ridership has been explained by eliminating trend components, eliminating seasonal components and the effects of the intervention. In Section 3.7, descriptive statistics were given and the conclusions were drawn that that all agencies saw major increases in ridership for the majority of the data's time span. The results of the incremental change in the adjusted R^2 values at each step show that the increases in ridership can be mostly attributed to increasing gas prices and unemployment rates, and to a much smaller degree, the Fix I-5 project. The exception is RT, where the Fix I-5 project accounts for most of the explained variance in ridership. This leads to the conclusion that the RT AM peak ridership is very well explained by the combination of an ARMA(3,3) model and a step function, while the RT PM peak ridership is very well explained by the combination of an ARMA(0,5) model and a step function.

Table 5.7 presents the model estimate of change in number of riders. Intervention analysis does not provide a formal test of statistical significance for the projected increase or decrease in number of riders. The projected change in number of riders (ω) is difficult to interpret in the absence of a benchmark indicating typical ridership levels for a given agency. Table 5.7 explains the effects of the Fix I-5 project on bus transit ridership taking total ridership into account. The column "$\omega/(\omega+CFR)$" describes the change in the mean number of riders divided by the observed ridership, where the observed ridership is viewed as the sum of the ridership due to the Fix I-5 project (ω) and the ridership that would have occurred without the Fix I-5 project (CFR, for "counterfactual ridership" which describes mean ridership levels in the absence of the Fix I-5 project, but during the same period). This percentage represents the proportion of total riders who were added or lost due to the Fix I-5 project. The last column "ω/CFR" represents the increased or decreased proportion of riders as compared to the situation if the Fix I-5 project did not happen. In all cases, the mean change in ridership is less than 4%, signifying a small change.

Table 5.7: Interpretation of Model Significance

Transit Agency	**Data Aggregation**	**Indicator Function**	**Final Model Estimate ω (mean change in riders)**	**Observed Mean Ridership during Fix I-5 Project (ω +CFR)**	**$\omega/(\omega$+CFR) x 100%**	**ω/CFR x 100%**
Yuba-Sutter Transit	AM	Pulse	-7.4	314	-2.36%	-2.30%
Yuba-Sutter Transit	PM	Pulse	-6.25	317.2	-1.97%	-1.93%
Yolobus	Daily	Pulse	-13.65	4023.5	-0.34%	-0.34%
Roseville Transit	AM	Pulse	-1.15	253.8	-0.45%	-0.45%
Roseville Transit	PM	Pulse	-3.75	226	-1.66%	-1.63%
North Natomas TMA	AM	Step	1.2	146.2	0.82%	0.83%
North Natomas TMA	PM	Pulse	2.6	124.3	2.09%	2.14%
Regional Transit	AM	Step	552.6	15525.8	3.56%	3.69%
Regional Transit	PM	Step	351.1	13907.5	2.52%	2.59%

5.5 Discussion

In the multiple regression analyses, gasoline prices and unemployment rates were significant predictors of transit ridership for all agencies except Regional Transit. Both gasoline price and unemployment rate had statistically significant correlations with bus transit ridership. As gasoline prices increased, ridership increased as people switched to bus transit from other modes of transit. As unemployment rates rose, ridership also increased. Possibly other economic hardships related to unemployment, including pay cuts and reduced work hours, resulted in increased bus transit ridership. The relationship between gasoline price and transit ridership is consistent with the literature (Lane, 2009;

Wang and Skinner, 1984; Taylor et al., 2009; Kyte et al., 1988). The relationship between unemployment rates and transit ridership is consistent with Agrawal (1981), as he states that jobs in Philadelphia and the unemployment rate had the same effect on ridership.

The results of the sinsusoidal decomposition show that all but two transit agencies had weekly cycles. Yolobus's data set is daily ridership, but shows no weekly cycle, possibly because of the presence of more non-commuters within the Yolobus dataset. Also, RT doesn't show a weekly cycle, because the RT data observation unit is one week. Possibly bus transit riders had various weekly work schedules that caused weekly cycles. Multiple agencies also had yearly cyclic behavior. Within each year there are understandable recurring patterns. For example, each year there are weather differences which could lead to low ridership in winter months with higher ridership in the warmer months. That might be especially apparent in ridership data, as riders are required to wait in the weather for the bus. Differences in cyclic behavior between the AM and PM peak periods in agencies beside Yuba-Sutter go mostly unexplained; it seems reasonable that commuters who take buses to work would, for the most part, also ride buses home. Possibly the peak periods, as defined by each transit agency, do not capture both commute periods accurately, or many bus transit riders work late or stay downtown to shop and dine after work.

The results of the intervention analysis show that the three agencies farthest from the Fix I-5 construction, namely Yuba-Sutter Transit (49.8 miles away), Roseville Transit (19.1 miles away), and Yolobus (18.7 miles away) saw small decreases in mean transit ridership attributable to the Fix I-5 project (indicated by the (-) sign in Table 5.7). However, the estimated change in number of riders from the intervention analysis model, taken as a proportion of total ridership, was under 3%. One possible hypothesis for the decrease in ridership is a mode shift from transit to driving to avoid the confinement of transit. Specifically, driving allows for more freedom in terms of route change and departure time. An alternative hypothesis is that commuters disproportionally worked at home, compressed their work schedules, or took vacations. The two transit agencies closest to the Fix I-5 project were North Natomas TMA (6.7 miles away) and Sacramento RT (3.7 miles away), and both had slight increases in mean ridership attributable to the Fix I-5 project. Regional Transit, whose service area is centered on the downtown core, had the largest estimated change in number of riders, taken as a proportion of total ridership. However, all results showed very minor impacts on bus transit from the Fix I-5 project.

A possible explanation for the smaller proportionate changes in mean transit ridership for the three more distant transit agencies is that those with longer commutes are more committed to their commute mode. It may be more likely they have monthly bus passes, and there are fewer commute options for those with longer commutes. Commuters served by RT have more commute options, including driving, taking RT, walking, bike riding or ride sharing to downtown Sacramento. However, walking and bike riding may be too strenuous for many, and ride sharing requires organization and commitment. For many in the RT service area, finding a bus may have been relatively easy when faced with the inconvenience of the Fix I-5 closure.

Additionally, those agencies that saw decreased ridership (Yuba-Sutter Transits, Yolobus, and Roseville Transit) were best modeled with a pulse indicator function. This signifies that the decreased ridership was a temporary change. Whereas, for those agencies that saw increased ridership (North Natomas TMA and RT) three of the four peak periods

(North Natomas TMA AM peak, and RT AM and PM peaks) were best modeled as a step indicator function. Although small, a 2%-4% permanent increase in transit ridership may be seen as a victory for most transit agencies. It should be noted that this study used only the simple pulse and step indicator functions, while some peak period ridership may be better represented as more complex linear or exponential indicator functions.

5.6 Implications for Transit Agencies for Future Road Closure Work

The results from this analysis show that the mean change in ridership attributable to Fix I-5 was small. As shown in Chapter 1, transit agencies made great attempts to increase ridership during the Fix I-5 project through media outreach as well as providing increased service. However, only North Natomas TMA and RT, the transit agencies closest to the Fix I-5 project, had increased ridership. Although for RT the increase numbered hundreds of riders, they comprised only a small proportion of RT's total ridership on the lines serving the affected area

For future construction projects, local transit agencies should plan for small proportionate changes in ridership. More distant transit agencies may not be affected or temporarily see decreased ridership for reasons discussed in Section 5.5. This study did not analyze the impact of Fix I-5 on other modes of transportation, which may have been more affected than bus transit ridership. However, Regional Transit light rail ridership counts indicate that light rail also saw minor changes in ridership during the period of the Fix I-5 project, although no statistical analysis was conducted (Kim, 2008).

5.7 Threats to Validity

This analysis was not a controlled experiment, so no statement of causality can be made between bus ridership and the Fix I-5 project. Further, the accuracy of data sets was dependent on each transit agency that supplied the study data. However, errors could be expected to be random, or if systematic (e.g. systematic over counts or undercounts of ridership) controlled for with the detrending and deseasonalizing steps. Validity is a measure of correctness (Maxwell and Delaney, 2004). Internal validity is concerned with possible other explanations for changes in the dependent variable (Maxwell and Delaney, 2004). In this study, internal validity is strengthened by the inclusion of gas prices, unemployment and seasons as covariates. External validity is concerned with the generalizabliliity of the results and how representative the sample is of a population (Maxwell, 2004).

History is not controlled in a time series quasi-experimental design (Campbell and Stanley, 1963). Some other event could have caused the results, producing a threat to internal validity. For example, major wild fires in California during the same time period which caused air quality concerns could have kept people in their homes. The results could have been an artifact of problems with data quality, discussed earlier, or changes in transit agency measurements over time (instrumentation). For example, RT's APC system is relatively new, and still in testing phases. The specific numerical results, even if internally valid, may not project beyond Sacramento (external validity), as the data sample is from a single city with its unique composition of commuters, employment, gas prices and network configuration. However, the methodology is broadly applicable, and it is expected that there would be similar roles of gas prices, unemployment and seasonality elsewhere. Therefore, similar types of models could be expected elsewhere, even if the exact magnitudes of the covariates differed. Although this study's novelty is that it is the first to examine the effects of construction projects on transit ridership, replication studies in the context of other reconstruction events are encouraged to begin a database of knowledge about the likely range of ridership impacts.

CHAPTER 6 CONCLUSIONS

The Fix I-5 project, which encompassed the rehabilitation of drainage and pavement of Interstate 5 in downtown Sacramento, was completed via an aggressive construction schedule of 35 days and 3 weekends between May 30, 2008 and July 28, 2008. The schedule included several complete closures of unidirectional portions of I-5 for 5-10 days at a time. The average daily traffic of approximately 200,000 motorists was detoured from I-5. Media outreach from both the private and public sector aimed to warn about projected traffic conditions while encouraging alternative modes of travel (such as transit). The purpose of this analysis was to determine if the change in mean bus transit ridership levels could be predicted by the Fix I-5 project, and if so, was that change permanent or temporary, and its magnitude and direction.

6.1 Summary

In this analysis, data on bus transit ridership was supplied by five agencies that provide commute service to the Sacramento downtown core, including Yuba-Sutter Transit, Yolobus, Roseville Transit, North Natomas TMA and Regional Transit. Where possible, only commute period travel was considered: more specifically, only inbound travel was considered for the AM peak period, while outbound travel was considered for the PM peak period, with peak period time intervals varying by transit agency. In total nine data sets were analyzed, four for AM travel, four for PM travel and one daily data set for Yolobus.

Although some past studies have used time series models in conjunction with transportation-related data (e.g. Atkins, 1979; Rose, 1982; Rose, 1986), there are few analyses of bus transit ridership. Further, intervention analysis is seldom used for transit data. No known studies have analyzed the impact of construction work on bus transit ridership using time series intervention analysis. No other studies have examined the impact of Sacramento's Fix I-5 project on bus transit ridership.

The five agencies in this study used four data collection methods, some in combination. They included automatic passenger counters, electronic fareboxes, manual counts by route checkers, and manual counts by bus drivers.

In order to modify the original ridership data for this analysis, the data sets were filtered to only include ridership for transit lines which provide service to the Sacramento downtown core, and inbound ridership for the AM peak period and outbound ridership for the PM peak period. Additionally, data imputation was done for days on which the transit agency ran limited or no services as well as state holidays. Total data imputation was under 2.0%.

An analysis of measures of central value and dispersion found that all agencies had overall increases in mean ridership during the study period (3 years for 3 agencies, 1 year for 2 agencies), but there were seasonal variations in mean bus transit ridership. An analysis of data dispersion indicated that the data sets were approximately normal. Data analyses were completed using intervention analysis, with the duration of the Fix I-5 project as the intervention.

Time series intervention analysis requires stationary data input, which entailed the elimination of trend and periodic components. Because the nine time series data sets displayed an overall increasing trend in ridership, trend components were eliminated using multiple regression. Independent, predictor variables were determined through literature review. Many studies have shown that gas prices positively affect ridership (Lane, 2009; Wang and Skinner, 1984; Taylor et al., 2009; Kyte et al., 1988), and the present analysis is no exception (aside from both RT peak periods). This analysis confirm past work as gas prices were significant predictors of ridership, with a positive ridership-gas correlation, with the exception of both RT peak period. Although a small number of past studies use employment as an independent variable (Agrawal, 1981; Kyte et al., 1988), this study found that unemployment is significant for all data sets except both RT peak periods, with positive ridership-unemployment correlations. No past work using bus transit ridership has used GDP as an independent variable. It was found to be highly correlated with the other independent variables (multicollinearity), and was eliminated from the analysis. Multiple studies have examined the effects of fare pricing on transit ridership, and found it to have a negative impact (Cervero, 1990; Curtin, 1968; Agrawal, 1981; Wang and Skinner, 1984; Taylor et al., 2009; Kyte et al., 1988; Narayan and Considine, 1989). This study finds that the fare increases for Roseville Transit and Yuba-Sutter (the only agencies to change fares within the study period) are insignificant predictors of transit ridership. More specifically, the four independent variables included monthly Sacramento unleaded regular gas price averages, national current dollar adjusted GDP data, monthly unemployment rates for the Sacramento— Arden-Arcade—Roseville, CA metropolitan area, and basic one-rider fare rates for those agencies whose fare pricing changed during the analysis period. For each agency, those predictor variables significant at the $\alpha = 0.1$ level were used during exploratory model building, but all final coefficients are significant at the $\alpha=0.01$ level.

The goal of the second stage of data analysis was to eliminate seasonal components using sinusoidal decomposition. Possible cycles were determined based on the periodogram, then tested for significance using spectral ANOVA. For each data set, those cyclic components significant at the 0.1 level were included in this analysis. Table 6.1 shows the significant cycles for each agency.

Table 6.1: Significant Periodic Components of Each Transit Agency

Transit Agency	**Data Aggregation**	**Significant Cycles (days)**
Yuba-Sutter Transit	AM	3, 74, 147
Yuba-Sutter Transit	PM	3, 73, 147
Yolobus	Daily	28, 38, 74, 220
Roseville Transit	AM	3, 58, 147
Roseville Transit	PM	3, 42, 147
North Natomas TMA	AM	3, 15, 19, 30
North Natomas TMA	PM	3, 8, 25, 74
Regional Transit	AM	4.5, 7.1
Regional Transit	PM	3.6, 4.2

The results of the sinsusoidal decomposition show that transit agencies with daily, peak-period ridership data indicate weekly cycles. Additionally, multiple agencies also indicate yearly cyclic behavior. Most agencies show differences in cyclic behavior between the AM and PM peak periods which is unanticipated. Eliminating trend and seasonal components was necessary so that the time series were stationary processes for input into an intervention analysis. Once the data sets were detrended and seasonal trends were eliminated, the first step of the intervention analysis was complete. The basic intervention model can be described as:

$$Y_t = \omega M_t + N_t$$

where ω represents the mean change in riders, and M_t represent the indicator variable. Intervention analysis for both pulse and step indicator variables show that temporary changes in riders was most applicable. Yuba-Sutter Transits, Yolobus, and Roseville Transit saw decreased ridership, and were best modeled with a pulse indicator function, which signified that the decreased ridership was a temporary change. North Natomas TMA and RT saw increased ridership, and three of those four peak periods (North Natomas TMA AM peak, and RT AM and PM peaks) were best modeled as a step indicator function. For all analysis, four goodness-of-fit tests including the Sample ACF, qq plot, Ljung-Box test and rank test were used.

Table 6.2 presents the model's estimates of the increase or decrease in ridership during the construction period, in absolute as well as percentage terms. In all cases, the mean change in ridership is less than 4%, signifying a very small change attributable to Fix I-5.

Table 6.2: Intervention Model Summary

Transit Agency	**Data Aggregation**	**Indicator Function**	**Final Model Estimate ω (mean change in riders)**	**Observed Mean Ridership during Fix I-5 Project (ω +CFR)**	**ω/(ω+CFR) x 100%**	**ω/CFR x 100%**
Yuba-Sutter Transit	AM	Pulse	-7.4	314	-2.36%	-2.30%
Yuba-Sutter Transit	PM	Pulse	-6.25	317.2	-1.97%	-1.93%
Yolobus	Daily	Pulse	-13.65	4023.5	-0.34%	-0.34%
Roseville Transit	AM	Pulse	-1.15	253.8	-0.45%	-0.45%
Roseville Transit	PM	Pulse	-3.75	226	-1.66%	-1.63%
North Natomas TMA	AM	Step	1.2	146.2	0.82%	0.83%
North Natomas TMA	PM	Pulse	2.6	124.3	2.09%	2.14%
Regional Transit	AM	Step	552.6	15525.8	3.56%	3.69%
Regional Transit	PM	Step	351.1	13907.5	2.52%	2.59%

* CFR, for "counterfactual ridership" which describes mean ridership levels in the absence of the Fix I-5 project, but during the same period.

The results of the intervention analysis show that the three agencies farthest from the Fix I-5 construction (Yuba-Sutter Transit, Roseville Transit, and Yolobus) saw small decreases in mean transit ridership (under 3%) during the period of the Fix I-5 project. Hypotheses for the decrease in ridership include a mode shift from transit to driving for more freedom, as well as commuters who worked at home, compressed their work schedules, or took vacations. The agencies that saw decreased ridership were best modeled with a pulse indicator function (a temporary change). The two transit agencies closest to the Fix I-5 project (North Natomas TMA and Sacramento RT) both had increases in mean transit ridership (under 4%) during the period of the Fix I-5 project. For those agencies that saw increased ridership, three of the four peak periods (North Natomas TMA AM peak, and RT AM and PM peaks) were best modeled as a step indicator function (a permanent change). Further, Regional Transit, whose service area is centered on the downtown core, had the largest estimated change in number of riders, taken as a proportion of total ridership. However, all results showed very minor impacts on bus transit from the Fix I-5 project. For future planned or unplanned freeway closures, local transit agencies should plan for small increases in mean riders, while small decreases in riders for more distant transit agencies may not necessitate any changes.

6.2 Future Work

This study provides room for future research, more specifically in the areas of data collection and model building.

First, within the data collection phase, monitoring of collection methods would provide additional insight into and influence on data quality. This study relied on data that was previously collected, some of which had been filtered. Monitoring of collection methods could also ensure sample randomness.

In this study, neither bus miles of service nor headways (the time between consecutive buses) were used as independent variables. Level of service indicators significantly predict transit ridership (e.g. Lane, 2009; Wang and Skinner, 1984; Agrawal, 1978), and it would be desirable to include an independent variable which summaries bus transit comfort levels.

Finally, this study used intervention analysis with step and pulse indicator functions. Future research could utilize more complex indicator functions, which could better explain the effects of an intervention, such as the Fix I-5 project.

REFERENCES

Agrawal, D.C. (1978) The factors affecting mass transportation ridership: An analysis. *Transit* **4(3)**, 55-66.

Allison, Paul D. (2002) *Missing Data.* Thousand Oaks: Sage Publications, Inc.

Anderson, O.D. (1976) *Time Series Analysis and Forecasting: The Box-Jenkins approach.* London, England: Butterworth & Co (Publishers) Ltd.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. New York: John Wiley and Sons, Inc.

Atkins, M. Stella. (1979) A case study on the use of intervention analysis applied to traffic accidents. *The Journal of the Operational Research Society* **30(7)**, 651-659. http://www.jstor.org.

Aue, Alexander. (2009) *STA 137: Applied Time Series Analysis course notes*. Chapters 1-4. Course taken Spring 2009, UC Davis.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Hoboken: John Wiley and Sons, Inc.

Box, G.E.P. and G.M. Jenkins. (1970) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Box, G.E.P., G.M. Jenkins and G.C. Reinsel. (2008) *Time Series Analysis: Forecasting and Control*. 4th edition. Hoboken: John Wiley & Sons, Inc.

Box, G.E. and D.A. Pierce. (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* **65(332)**, 1509-1526. http://www.jstor.org.

Box, G.E.P. and G.C. Tiao. (1975) Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* **70(349)**, 70-79. http://www.jstor.org.

Boyle, D.K, Transit Cooperative Research Program. (1998) *TCRP synthesis 29: passenger counting technologies and procedures.* Washington, D.C.: National Academy Press.

Boyle, D.K., Transit Cooperative Research Program. (2008) *TCRP synthesis 77: passenger counting systems.* Washington, D.C.: Transportation Research Board.

Brocklebank, John C. and David A. Dickey. (2003) *SAS for Forecasting Time Series*. 2nd edition. Cary: SAS Institute, Inc.

Brockwell, Peter J. and Richard A. Davis. (1987) *Time Series: Theory and Methods.* New York: Springer-Verlag New York, Inc.

Brockwell, Peter J. and Richard D. Davis. (2002) *Introduction to Time Series and Forecasting.* 2nd edition. New York: Science+Business Media, LLC.

Brodsky, B.E. and B.S. Darkhovsky. (2000) *Non-Parametric Statistical Diagnosis: Problems and Methods.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

California Department of Transportation. (2009) Awarded Contracts. http://www.dot.ca.gov/hq/esc/oe/planholders/awarded.php?q=&b=&a=2008-02-08&c=&d=&search=Search (accessed October 5, 2009).

California Department of Transportation. (2009) Fix I-5. http://www.fixi5.com No longer active circa August 2008. (accessed August 2008).

Campbell, Donald T., and Julian C. Stanley. (1963) *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally College Publishing Company.

C.C. Myers, Inc. Past Projects. http://www.ccmyers.com/completedprojects.cfm (accessed September 15, 2009).

Cervero, R. (1990) Transit pricing research: a review and synthesis. *Transportation* **17(1)**, 117–139. http://www.springerlink.com.

Chatfield, C. (1975) *The Analysis of Time Series: Theory and Practice.* London, England: Chapman and Hall Ltd.

Chatfield, C. (1980) *The Analysis of Time Series: An Introduction*. 2nd edition. New York: Chapman and Hall.

Cherwony, Walter and Lewis Polin. (1977) Forecasting Patronage on new transit routes. *Traffic Quarterly* **31(1)**, 287-29.

Christ, Carl F. Comment, 1983. (2004) In *The Structural Econometric Time Series Analysis Approach*, ed. Zellner, Arnold and Palm Franz C., 170. Cambridge, England: Cambridge University Press.

City of Sacramento, Department of Transportation. (2008) City of Sacramento Traffic Operations Center I-5 Closure. PowerPoint presentation prepared for the Intelligent Transportation Society of California Annual Meeting 2008.

Clever Devices. (2009) Automatic Passenger Counting. http://www.cleverdevices.com/productdetail.php?ID=70&Hash=a734da96f5e086c6c13fb9bff12c75747d80cbf1 (accessed September 21, 2009).

Considine J. and J. Narayan. (1988) Assessment of the impact of changes in transit systems using intervention analysis. *Transportation Research Part B* **22(1)**, 55-67. http://www.elsevier.com.

Cryer Jonathan D. (1986) *Time Series Analysis*. Boston: PWS Publishers.

Curtin, J.F. (1968) Effect of fare on transit riding. *Highway Research Record* **213(1)**, 8-29.

Dahl, Carol A. (1986) Gasoline Demand Survey. *Energy Journal* **7(1)**, 67-82.

Drake, James. (2007) Sacramento Regional Transit. *Regional Transit, Route Profiles, Fiscal Year 2007.*

Drake, James. Title: Assistant Planner, Regional Transit. Interview by Rachel Carpenter. e-mail, phone, mail, in-person. July 2008-July 2009.

Dutra, Dawna. Title: Analyst, Yuba-Sutter Transit. Interview by Rachel Carpenter. e-mail, phone. August 2008-January 2009.

Farnum, Nicholas R. and LaVerne W. Stanton. (1989) *Quantitative Forecasting Methods*. Boston: PWS-KENT Publishing Co.

Fuller, Wayne A. (1976) *Introduction to Statistical Time Series.* New York: John Wiley & Sons, Inc.

Gallwey, John. Title: Technical Services, UC Berkeley Institute of Transportation Library. Interview by Rachel Carpenter. e-mail. August 2009.

Gately, Dermot. (1990) The U.S. demand for highway travel and motor fuel. *Energy Journal* **11(3)**, 59-73. http://www.ebscohost.com.

Glass, Gene V. (1972) Estimating the effects of intervention into a non-stationary time-series. *American Educational Research Journal* **9(3)**, 463-477. http://www.jstor.org.

Gomez-Ibanez, J. (1996) Big-city transit, ridership, deficits, and politics. *Journal of the American Planning Association* **62(1)**, 30–50. http://www.ebscohost.com.

Goodwin, Phil, Joyce Dargay and Mark Hanly. (2004) Elasticities of road traffic and fuel consumption with respect to price and income: A review. *Transport Reviews* **24(3)**, 275-292. http://www.ebscohost.com.

Haydu, Elizabeth. Title: Administrative Technician, Roseville Transit. Interview by Rachel Carpenter. e-mail, phone, in-person. July 2009-August 2009.

Heyer-Gray, Bob. Title: Librarian, UC Davis Physical Sciences and Engineering Library. Interview by Rachel Carpenter. e-mail. August 2009.

Hoel, Paul G. (1971) *Introduction to Mathematical Statistics*. New York: John Wiley & Sons, Inc.

Janus, Sarah. Title: Program Coordinator, North Natomas Transportation Management Association. Interview by Rachel Carpenter. e-mail. August 2008-July 2009.

Kedem, Benjamin and Konstantinos Fokianos. (2002) *Regression Models for Time Series Analysis.* Hoboken: John Wiley & Sons, Inc.

Kendall, M.G. (1955) *Rank Correlation Methods*. New York: Hafner Publishing Company.

Kendall, M.G. (1973) *Time-Series*. London, England: Charles Griffin & Company Limited.

Kennedy, Peter. (2003) *A Guide to Econometrics*. 5th edition. Cambridge: The MIT Press.

Kim, Changmo. (September 23, 2008) Ridership Count Summary, PowerPoint presentation prepared for the CalEPA Fix I-5 Study.

Kitamura, R. (1989) A causal analysis of car ownership and transit use. *Transportation* **16(2)**, 155–173. http://www.springerlink.com.

Kutner, Michael H., Christopher J. Nachtsheim, John Neter, and William Li. (2005) *Applied Linear Statistical Models*. 5th edition. New York: McGraw-Hill/Irwin.

Kyte, Michael, James Stoner, and Jonathan Cryer. (1988) A time-series analysis of public transit ridership in Portland, Oregon, 1971-1982. *Transportation Research Part A* **22A(5)**, 345-359. http://www.sciencedirect.com.

Lane, B.W. (2009) The relationship between recent gasoline price fluctuations and transit ridership in major US cities. *Journal of Transport Geography* Article in Press. http://www.elsevier.com.

Li, Wai Keung. (2004) *Diagnostic Checks in Time Series*. Boca Raton: Chapman & Hall/CRC.

Ljung G.M. and G. E. P. Box. (1978) On a measure of lack of fit in time series models. *Biometrika* **65(2)**, 297-303. http://www.jstor.org.

Mann, Prem S. (2004) *Introductory Statistics*. 5th edition. Hoboken: John Wiley & Sons, Inc.

Mataloni, Lisa. Title: Economist, Roseville Transit. Interview by Rachel Carpenter. e-mail, August 2009.

Maxwell, Scott E. and Harold D. Delaney. (2004) *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. 2nd edition. London, England: Routledge, a Taylor & Francis Group.

Mokhtarian, Patricia. (2009-2010) *How do commuters react to a temporary freeway closure? An evaluation of the Fix I-5 project in Sacramento, California*. Faculty Research Proposal to the UC Davis Sustainable Transportation Center.

Mokhtarian, Patricia. Title: Professor of Civil and Environmental Engineering, UC Davis. Interview by Rachel Carpenter. e-mail, in-person. February 2009-October 2009.

Mokhtarian, Patricia L., Francisco J. Samaniego, Robert H. Shumway and Neil H. Willits. (2002) Revisiting the notion of induced traffic through a matched-pairs study. *Transportation* **29(2)**, 193–220. http://www.springerlink.com.

Montgomery, Douglas C., Cheryl L. Jennings and Murat Kulahci. (2008) *Introduction to Time Series Analysis and Forecasting*. Hoboken: John Wiley & Sons, Inc.

Narayan, J. and J. Considine. (1989) Assessing the impact of fare increases in a transit system by using intervention analysis. *Journal of Business Research* **19(4)**, 245-254. http://www.sciencedirect.com.

Ostrom, Charles W. Jr. (1978) *Time Series Analysis: Regression Techniques*. Beverly Hills: Sage Publications, Inc.

Poggioli, Susann. Title: Strategic Account Manager, Clever Devices. Interview by Rachel Carpenter. e-mail. September 2009.

Priestley, M.B. (1981) *Spectral Analysis and Time Series. Volume 1: Univariate Series Volume 2: Multivariate Series, Prediction and Control.* San Diego: Academic Press Inc.

Regional Transit. (2009) RT at a glance. http://www.sacrt.com/rtataglance.stm (accessed October 8, 2009).

Reitz, Erik. Title: Transit Planner, Yolobus. Interview by Rachel Carpenter. e-mail, phone, mail, in-person. August 2008-August 2009.

Rose, Geoffrey. (1982) *An aggregate time-series analysis of the effects of fare changes on transit ridership*. M.S. Thesis. Northwestern University, Department of Civil Engineering.

Rose, Geoffrey. (1986) Transit passenger response: Short and long term elasticities using time series analysis. *Transportation* **13(2)**, 131-141. http://springerlink.com.

Ross, Sheldon M. (2005) *Introductory Statistics*. 2nd edition. Burlington: Elsevier Academic Press.

Schafer, Andreas. (1998) The global demand for motorized mobility. *Transportation Research Part A* **32(6)**, 455-477.

Schafer, Andreas and David Victor. (2000) The future mobility of the world population. *Transportation Research Part A* **34(3)**, 171-205.

Shapiro, S.S. and R. S. Francia. (1972) An approximate analysis of variance test for normality. *Journal of the American Statistical Association* **67(337)**, 215-216. http://www.jstor.org.

Schimek, Paul. (1996) Automobile and public transit use in the United States and Canada: Comparison in the postwar trends. *Transportation Research Record* **(1521)**, 3-11. http://www.ebscohost.com.

Schimek, Paul. (1996) Gasoline and travel demand models using time series and cross-section data from United States. *Transportation Research Record* **(1558)**, 83-89. http://www.ebscohost.com.

Schwarzenegger, Arnold. Governor's Executive Order S-04-08. http://gov.ca.gov/executive-order/9629 (accessed October 4, 2009).

Sheets, Teri. Title: Alternative Transportation Analyst, Roseville Transit. Interview by Rachel Carpenter. e-mail. July 2008-August 2008.

Shumway, Robert H. and David S. Stoffer. (2000) *Time Series Analysis and Its Applications*. New York: Springer-Verlag New York Inc.

Shumway, Robert H. and David S. Stoffer. (2006) *Time Series Analysis and Its Applications: With R Examples.* New York: Springer Science+Business Media, LLC.

Solak, Ken. (November 18, 2008) I-5 Rapid Rehabilitation (Fix I-5), Caltrans District 3. PowerPoint presentation prepared for the Self-Help Counties Coalition, Focus on the Future 2008.

Taylor, Brian D., Douglas Miller, Hiroyuki Iseki, and Camille Fink. (2009) Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A* **43(1)**, 60–77. http:///www.elsevier.com.

Tinbergen, Jan. (1951) *Econometrics*. New York: George Allen & Unwind Ltd.

Tsay, R.S. (1984) Regression models with time series errors. *Journal of the American Statistical Association* **79(385)**, 118-124. http://www.jstor.org.

Wang, George H.K., and David Skinner. (1984) The impact of fare and gasoline price changes on monthly transit ridership: empirical evidence from seven U.S. transit authorities. *Transportation Research Part B* **18(1)**, 29-41. http://www.elsevier.com.

Washington, Simon P., Matthew G. Karlaftis and Fred L. Mannering. (2003) *Statistical and Economic Methods for Transportation Data Analysis*. Boca Raton: Chapman & Hall/CRC.

Yaffee, Robert A. and Monnie McGee. (2000) *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS.* San Diego: Academic Press, Inc.

Zellner, Arnold and Palm Franz C. (2004) *The Structural Econometric Time Series Analysis Approach*. Cambridge, England: Cambridge University Press.

APPENDICES

A. City of Sacramento Traffic Operations Center Visit, August 21, 2008

According to Jacques Van Zeventer, PE, TE, California Department of Transportation (Caltrans) District 5 Traffic Manager, Intelligent Transportation Systems (ITS) are important to "get info out (for you) to make better choices about how to get to your destination... You can't build your way out of congestion." An operational tactic for traffic management is through the use of traffic operations centers (TOC). On August 21, 2008, Michael Zhang and his research team consisting of Feng Xiao, Changmo Kim, Zhen Qian, Yi-ru Chen, Wei Shen, Robert Lim and Rachel Carpenter visited Sacramento City Hall, which houses the single jurisdiction, single agency TOC, operated by and serving the City of Sacramento. This TOC is staffed by the city's Traffic Engineering Services Department.

While visiting the TOC we met with the two employees who permanently staff the facility: Telecommunications Technician Shad Bennett and Telecommunications Engineer Ryan Billeci. Mr. Billeci is a Civil Engineer by education, and a California licensed Professional Engineer. According to Mr. Billeci, the Sacramento TOC was born out of three needs. First, the City wanted remote control of its 700 signals, including signal coordination, clock synchronization, the ability to download information and upload timing. They also wanted the ability to remotely observe traffic conditions in their system, and remotely conduct incident management. He said the goal of their TOC is twofold; first, they must make Sacramento City's transportation network efficient for all transportation modes, and second, they must make the system reliable. Today, the City of Sacramento TOC is operational 5 days a week, and consists of a large room with 2 cubicles, in addition to a long table where there are nine computers. There is a system of six large display screens that show live traffic conditions from the CCTV camera system. The City TOC is funded by Measure A, the gas tax.

Mr. Billeci discussed some institutional issues related to the TOC. The TOC participates in interagency collaboration with neighboring organizations including the County of Sacramento and their traffic management center (TMC), the Caltrans District 3 Sacramento TMC, and the cities of Roseville, Citrus Heights, and Elk Grove. As part of an effort to connect these individual agencies, the Sacramento Area Council of Governments (SACOG) funded the Sacramento Transportation Area Network (STARNET). According to the SACOG website, STARNET has four major goals:

- Connect the region's real-time transportation management systems
- Allow sharing or real-time data between systems and between users
- Allow sharing of live video
- Provide real-time information to the public via 511 and other outlets.

The City of Sacramento has the long term goal to establish a traffic and transportation disseminating service similar to the Iowa Department of Transportation website. SACOG contracted with Castle Rock Consultants to implement the STARNET system, which will be in operation by 2009, and the main source of information for the planned all-inclusive website.

During our visit to the City TOC, we were given a presentation about the City of Sacramento TOC during the Fix I-5 project. The following summarizes the presentation:

Traffic signal operation and traffic management is important for four main reasons: its direct economic impact, improvement of air quality, reduction in fuel consumption, and reduction in the number and severity of accidents. Some other functions of the TOC include live traffic surveillance, police, fire and CHP channel monitoring, and Caltrans camera monitoring. In terms of organizational issues, Mr. Billeci said that the City of Sacramento's organizational culture still leans towards constructional improvements, as opposed to operational solutions. Therefore, TOC staff constantly have to prove the importance of the TOC and their jobs. It is hard to recognize a need when a system is running smoothly. Making the difficult task of traffic control look easy results in the TOC appearing to be unnecessary. However, the existence of a well-equipped traffic operation center points to a new direction for the City.

During the Fix I-5 closure, many forces joined to ensure the efficiency of the transportation network. The City TOC was staffed with employees from many agencies and departments including the traffic signal operation group, the Sacramento Police Department, traffic investigation group, right-of-way management, the Sacramento Fire Department, Sacramento Regional Fire Dispatch, and Caltrans.

Within the TOC, preplanning steps were taken to prepare for the Fix I-5 project. Problem corridors were identified, while focusing on downtown Sacramento commuters. Unfortunately, regional commuters, who largely impacted the transportation network, were not taken into consideration. Preventative maintenance measures were taken on 300 signals. Additional ITS elements were identified and employed especially for the Fix I-5 project. These additional elements included wireless closed-circuit television (CCTV) cameras using Dotworkz Systems camera housings. The wireless CCTV cameras were built and maintained by City of Sacramento staff, and included full pan/tilt/zoom (PTZ) control with live motion video. Once impacted corridors were identified, a Synchro traffic model was constructed in order to implement direct routes. The TOC also developed coning and striping modification plans. Some of the anticipated impacted corridors that were identified in preplanning stages by the TOC included:

- 16th Street
- 3rd Street
- 5th Street
- Riverside Boulevard
- Freeport Boulevard
- Broadway
- Truxel Road

In order to prove the importance of the City of Sacramento TOC, cost of delay calculations were done:

Given: $12.02 per vehicular hour of delay (Caltrans Progress Report with gas at $3.17/gallon)

Calculations:

Example 1: Corridor

ADT 10,000 vehicles

1.5 miles long

10 signals

Current average delay per signal is 24 seconds

$300,000 = Monthly delay cost

Example 2

Hourly vehicle volume 1800 vehicles

Current average delay per vehicle is 20 minutes

Duration: 1.5 hours

$570,000 = Monthly delay cost

Solution: By reducing the delay to 10 minutes the public saves $285,000.

The traffic conditions on 16th Street were of major interest to the City TOC. Conditions on 16th street prior to Fix I-5 included an average peak hour of 3000 vehicles with a typical travel time of 6 minutes. During northbound lane closures, the estimated average peak hour increased to 4500 vehicles, and the average travel time during the first two days of closure increased to over 12 minutes. After travel time management strategies were in place, the average travel time was reduced to 7 minutes. Some of the mitigation measures that improved travel times included increased cycle length, the implementation of the preplanned coning plan, and police presence.

Truck traffic posed a serious threat during the I-5 closures, as most City of Sacramento surface streets were not designed for heavy truck loads. The police department dealt with the truck problem by working in collaboration with the TOC. During northbound closures, CCTV video provided TOC operators notice when trucks were exiting I-5 onto downtown surface streets. TOC staff then contacted standby police who cited violators. During southbound closures, trucks were rerouted to avoid the downtown altogether.

After the Fix I-5 presentation, Mr. Bennet and Mr. Billeci recalled actions that they had taken during the I-5 closures. They said that many decisions, including systematic changes, were made spontaneously. Some of these changes included signal timing, and striping and signage. They made changes to cycles that worked so well that they are still being utilized today, even though there will be no further lane closures due to the Fix I-5 project. While the locations of major traffic congestion changed throughout the project, the TOC worked to keep traffic moving smoothly. With the use of remotely accessible field equipment, spontaneous decisions were made about how to best operate the system.

Field equipment is essential to any TOC because the equipment collects data for processing at the TOC. The City of Sacramento TOC uses a variety of data collection devices including inductive loop detectors, CCTV cameras and video-based detection systems. Interestingly, the City of Sacramento uses inductive loop detection (including mid, rear and front detection) for the majority of their detection devices. Mr. Billeci stated that inductive loops, even with their intrusive nature, are still the cheapest form of detection. They are also reliable, as they have approximately a 10-year life. The Sacramento TOC also uses a video-based detection system called Autoscope RackVision, however this system is expensive and maintenance intensive. They use the Autoscope Mini-hub, which provides the interface between the Autoscope RackVision and their Traconex TMP-390 traffic signal controller. The City of Sacramento does not have TS2 capabilities. The City has recently received $4 million for a full upgrade on their transportation system. Part of this funding will be used for a Wavetronix system that has the ability to store traffic data, and provides vehicle speeds.

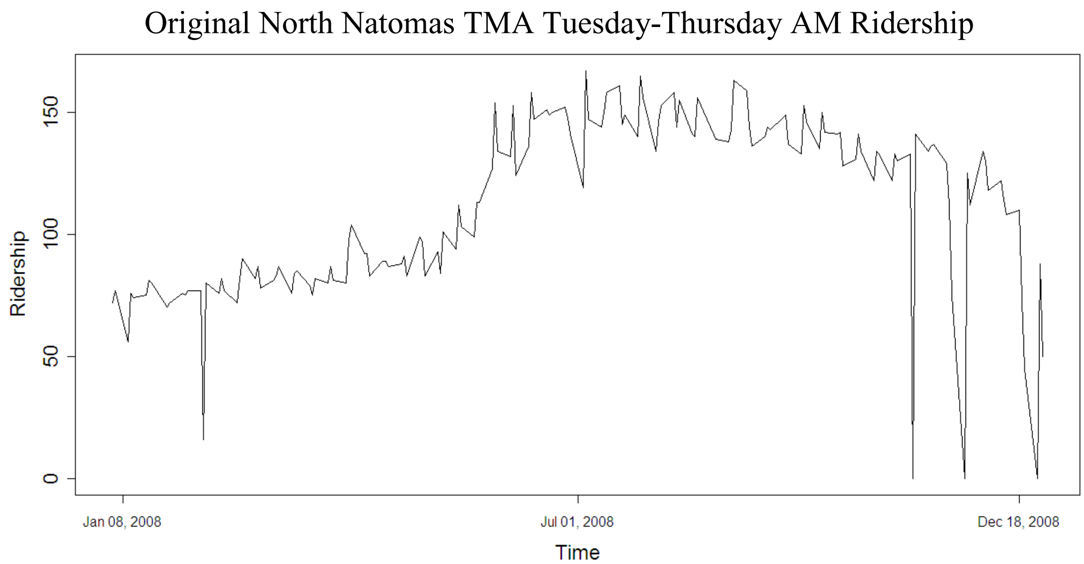
Toward the end of our visit to the TOC, Sacramento City Traffic Engineer**,** Hector Barron, visited the TOC. He said that the City plans to work closely with Caltrans on certain corridors to make them run more smoothly. He also said that the City of Sacramento has debated the idea that all new development would be required to pay a road improvement impact fee. Although Mr. Barron believes it is a good idea, there has been opposition from city officials. Additionally, Sacramento has considered High Occupancy Vehicles (HOV) lanes and High Occupancy Toll (HOT) lanes but Sacramento officials believe it promotes urban sprawl and no lanes have passed the preliminary planning stage.

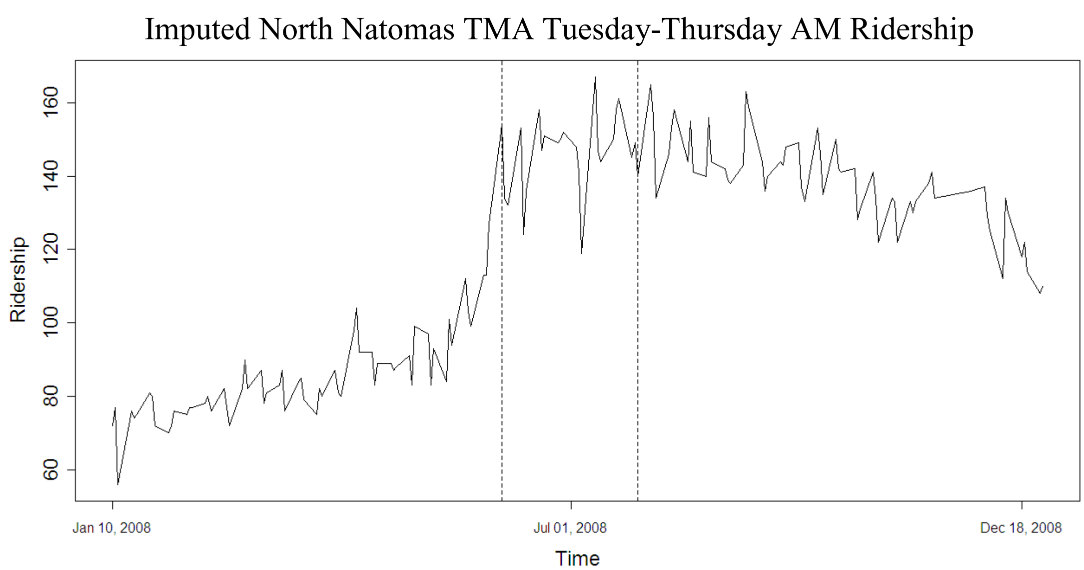
The City of Sacramento has a good start on its TOC. It has partnered with other agencies, and aims to produce an all-inclusive website for transportation related information available to the public. It is staffed by two full time employees who hold operational improvement mindsets- not construction approaches. Operational viewpoints lead to the use of more ITS technologies and greater promotion of what ITS can do for local communities. The various information gathering devices, control room devices, and information dissemination devices all make the traveling public’s trips more effective. The future in transportation continues to look brighter with the use of more ITS equipment for greater operational efficiencies of the roadways.

B. Original and Cleaned Transit Agency Ridership Data Sets

Note: For all imputed data graphs, the first dashed vertical line represents June 3, 2008 (the first imputed day of Fix I-5 project construction), while the second dashed vertical line represents July 24, 2008 (the last imputed day of Fix I-5 project construction).

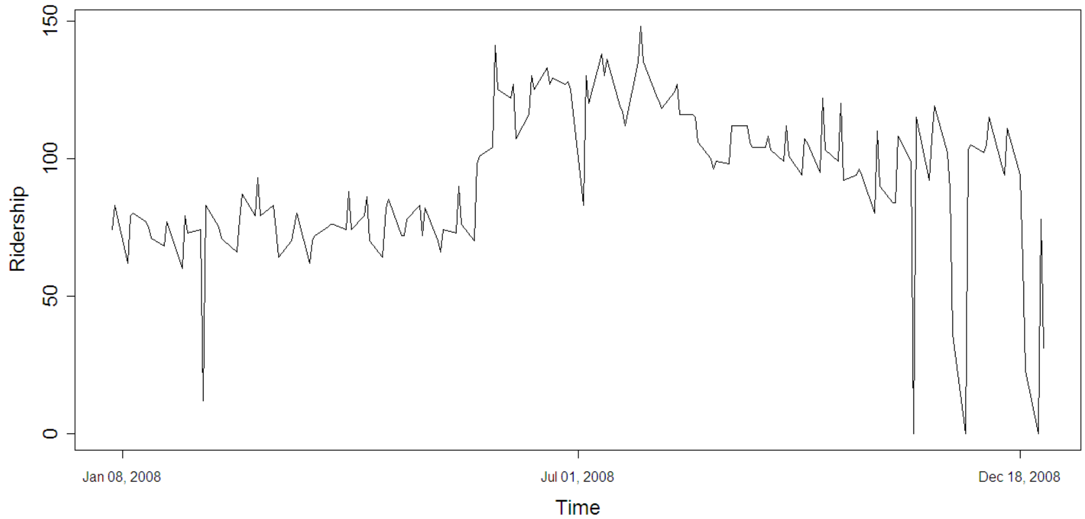
1. North Natomas TMA AM Ridership



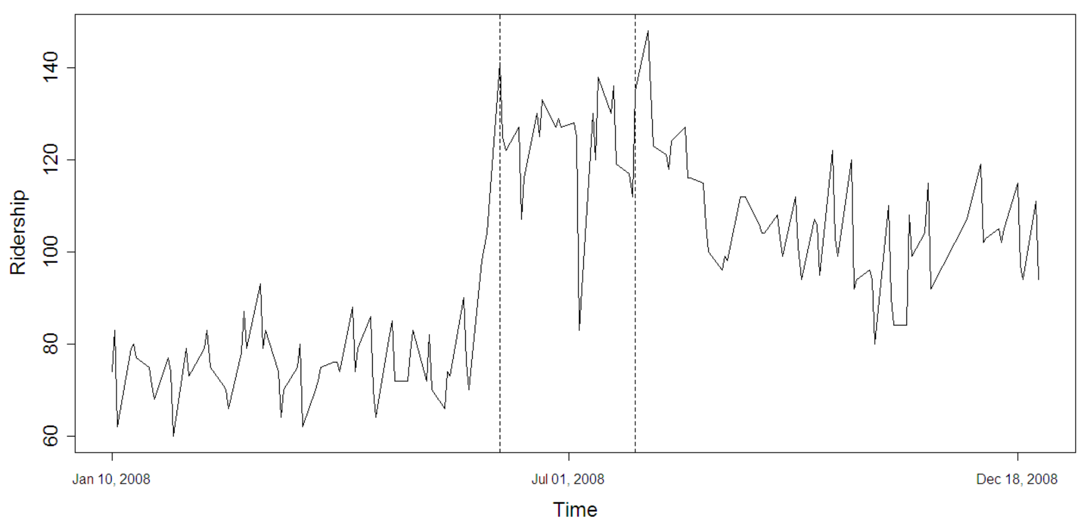


2. North Natomas TMA PM Ridership

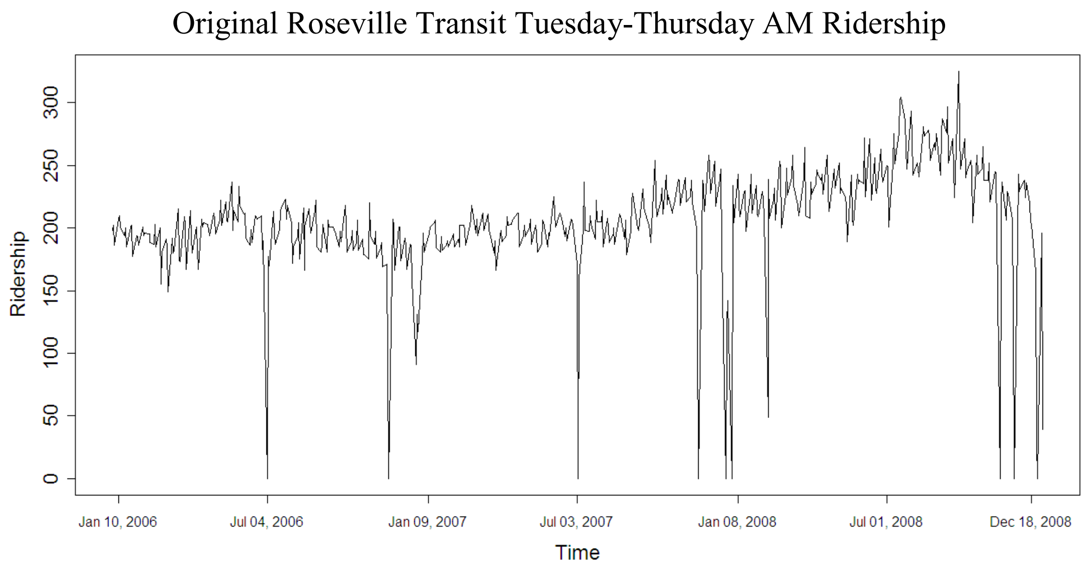
Original North Natomas TMA Tuesday-Thursday PM Ridership

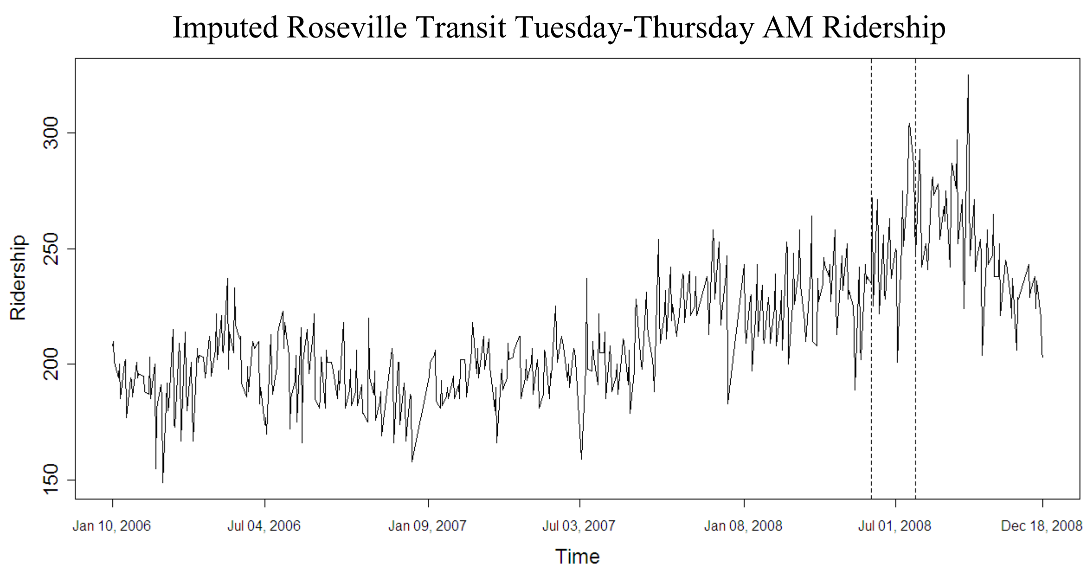


Imputed North Natomas TMA Tuesday-Thursday PM Ridership

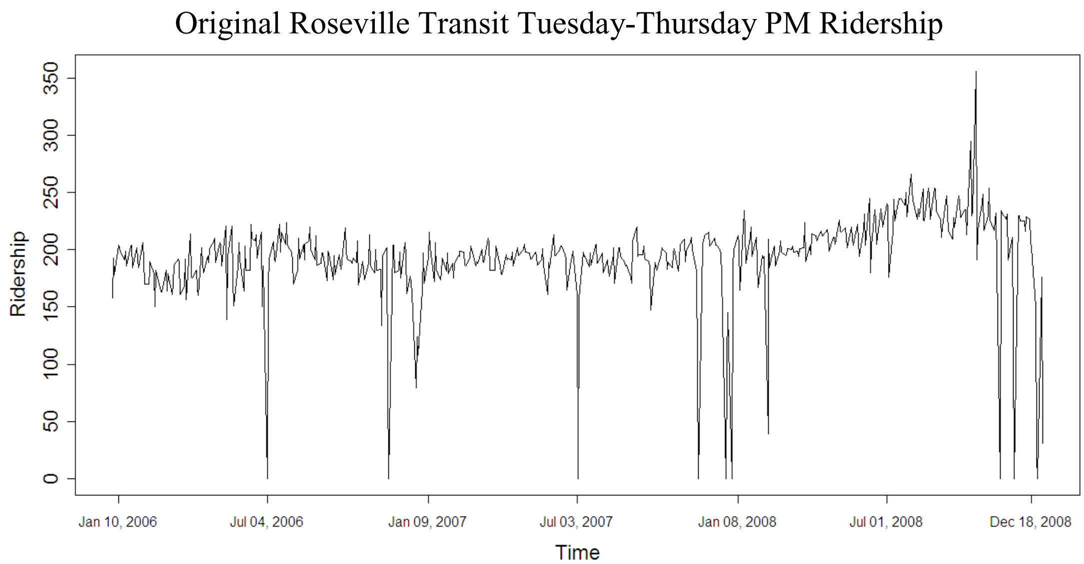


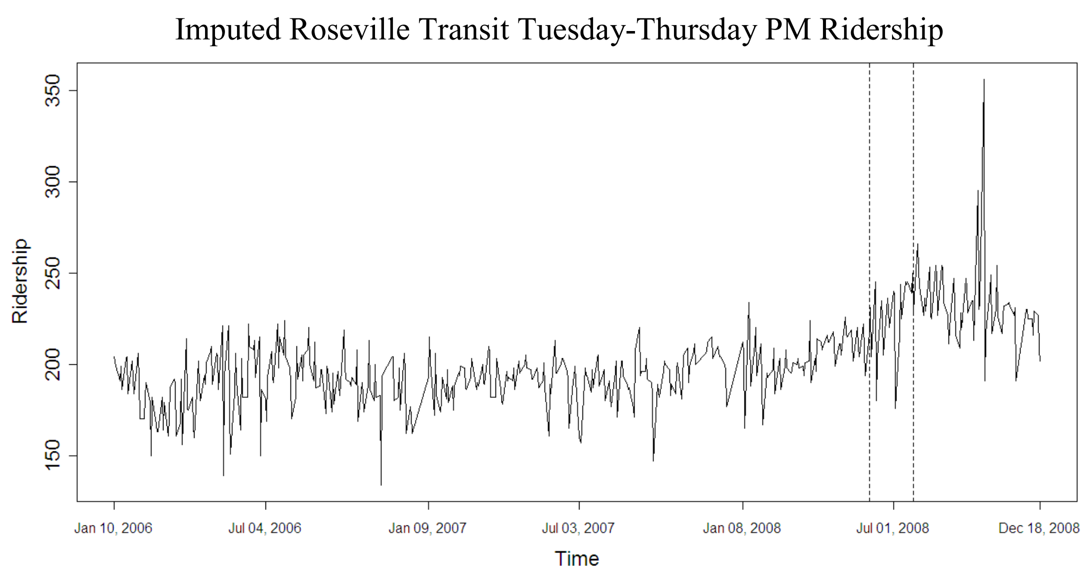
3. Roseville Transit AM Ridership





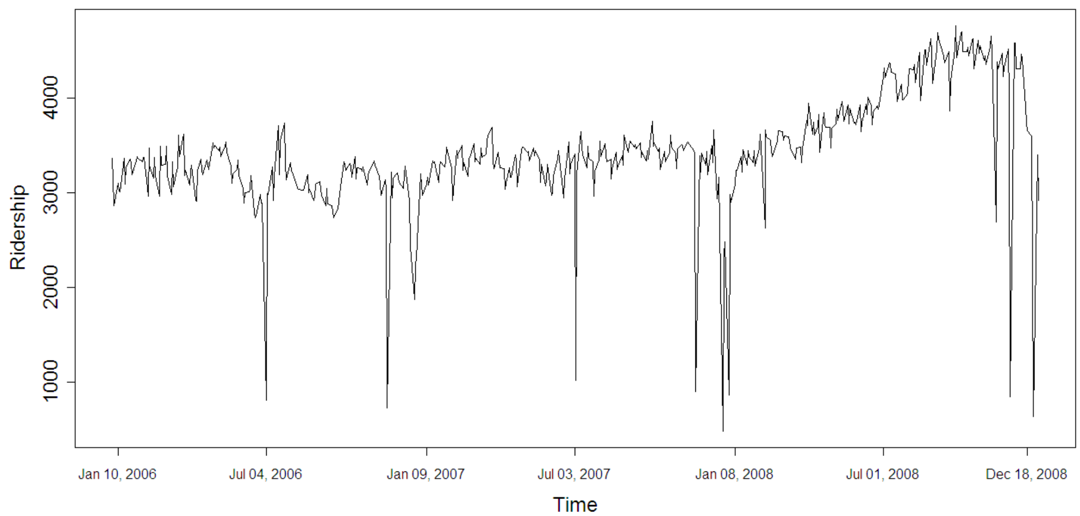
4. Roseville Transit PM Ridership



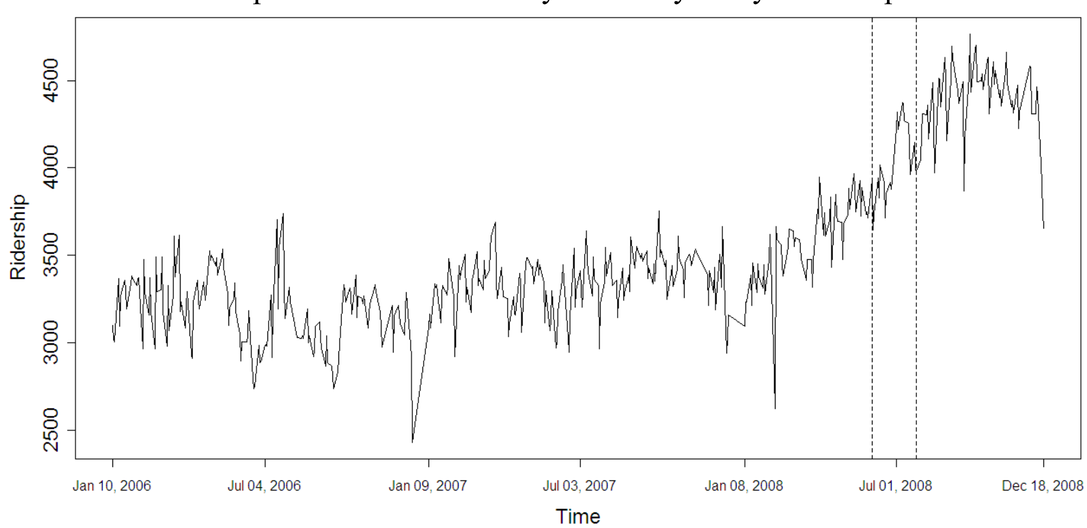


5. Yolobus Ridership

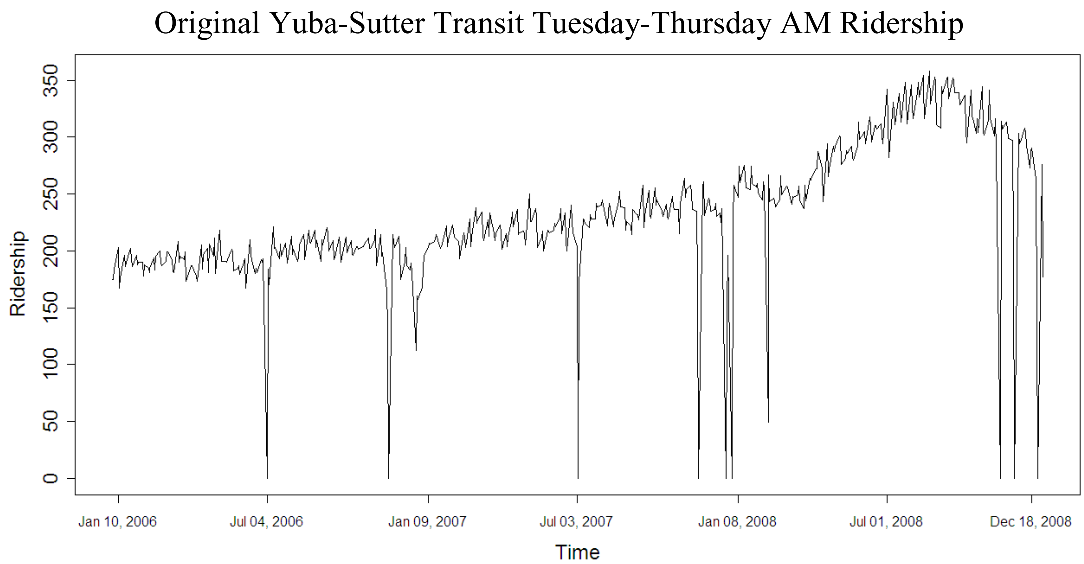
Original Yolobus Tuesday-Thursday Daily Ridership

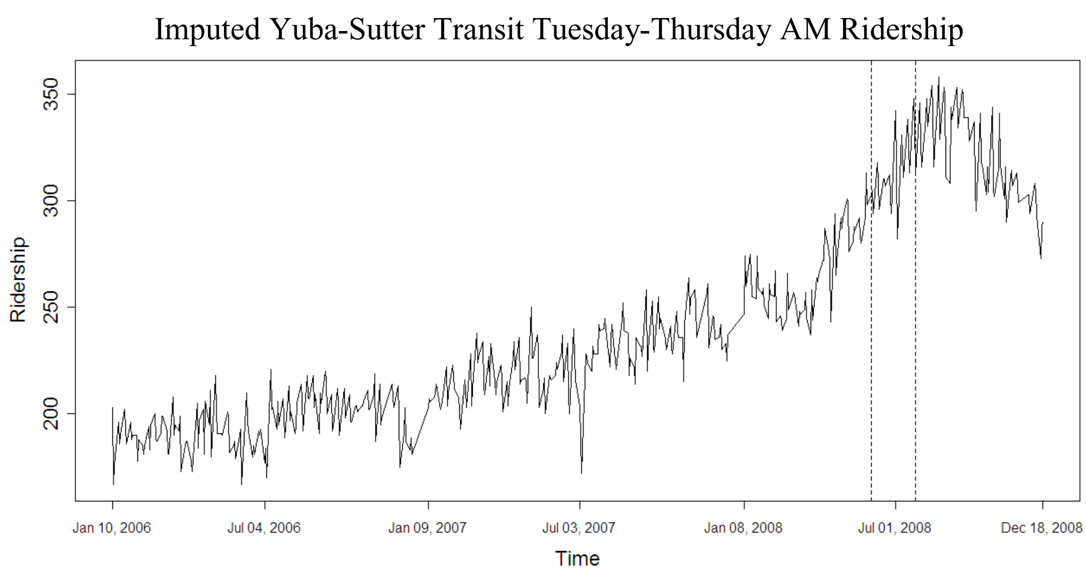


Imputed Yolobus Tuesday-Thursday Daily Ridership

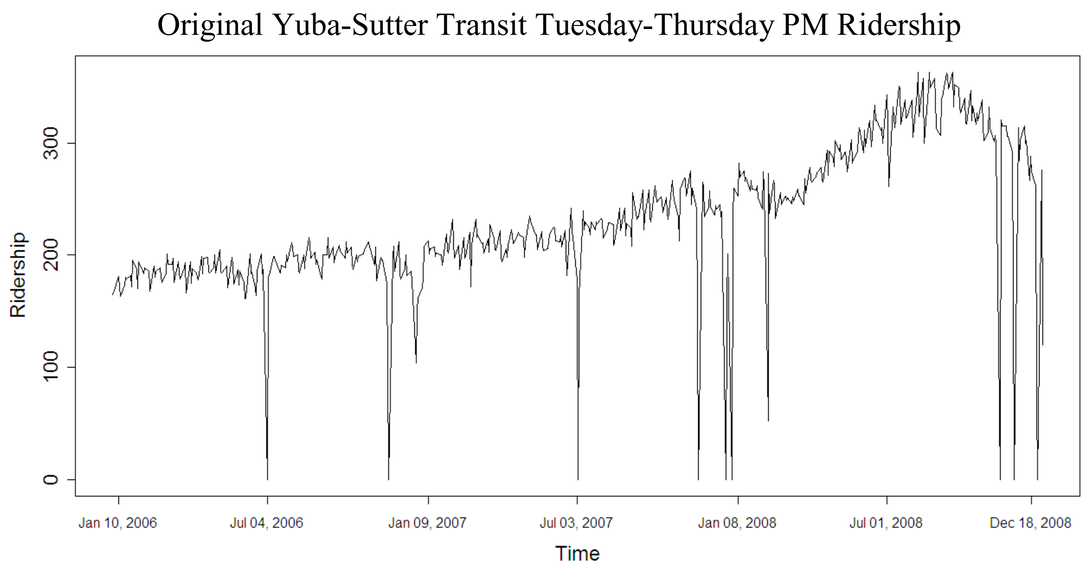


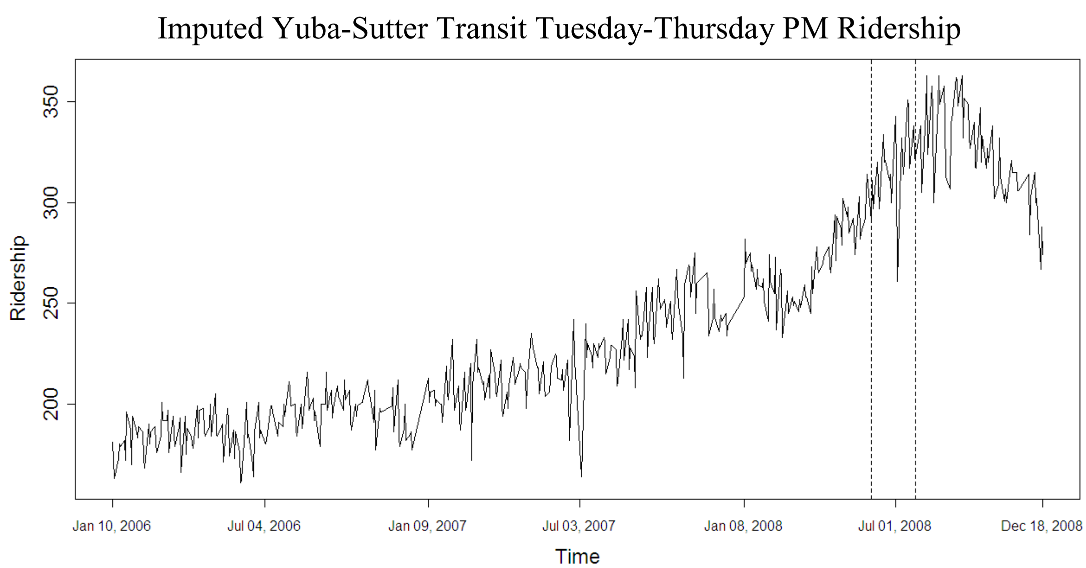
6. Yuba-Sutter AM Ridership



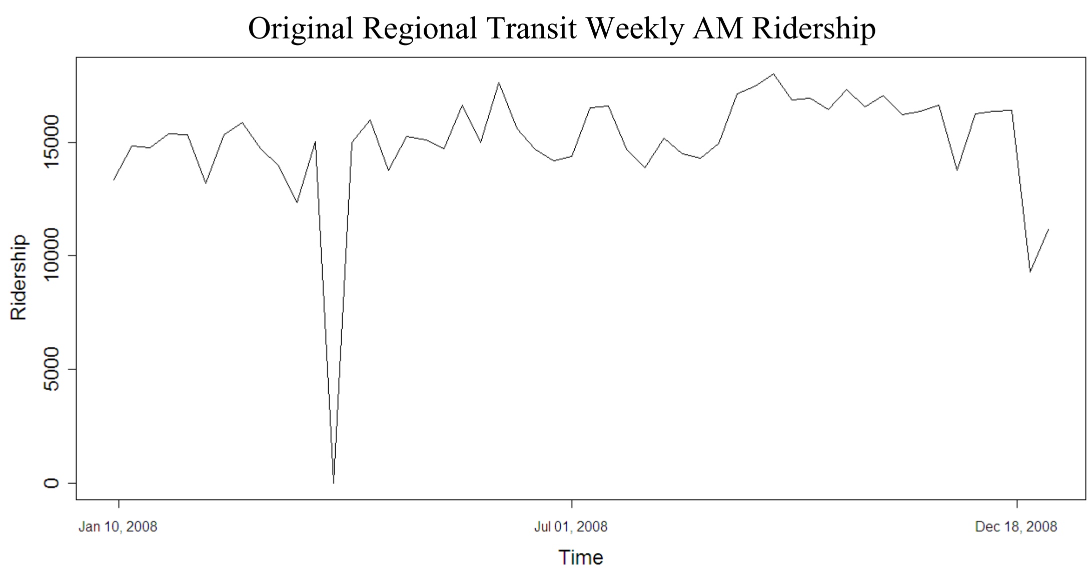


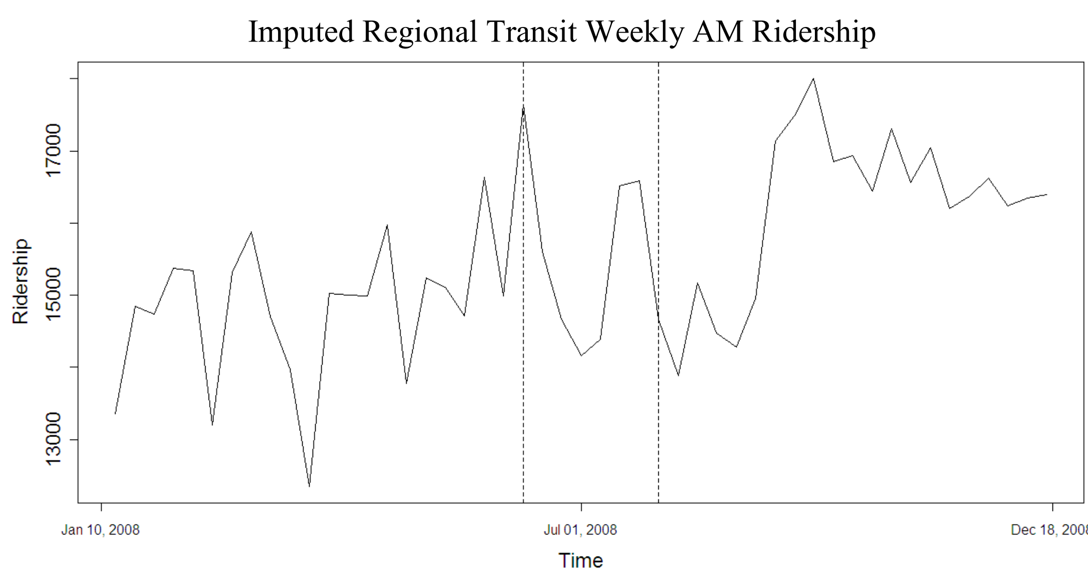
7. Yuba-Sutter PM Ridership





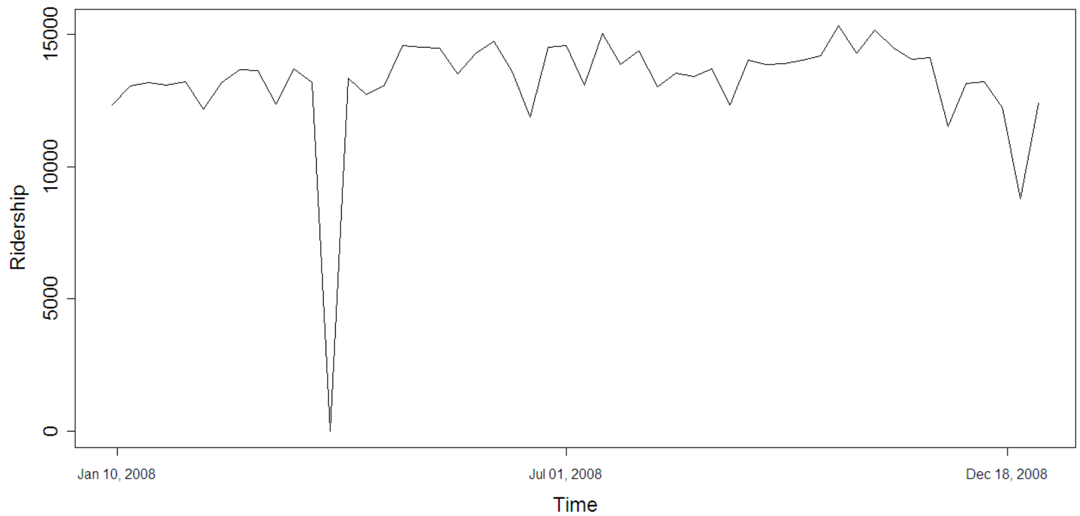
8. Regional Transit AM Ridership



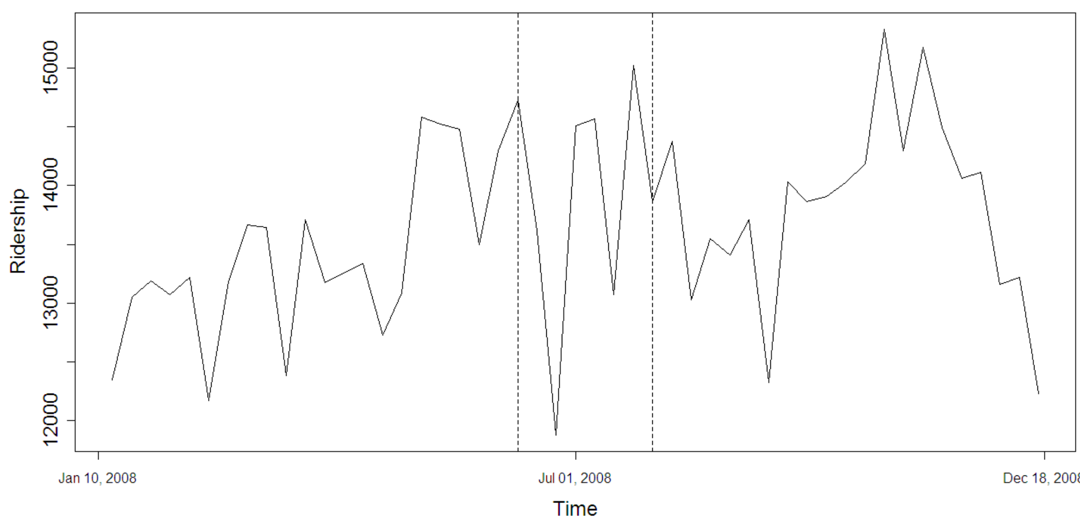


9. Regional Transit PM Ridership

Original Regional Transit Weekly PM Ridership

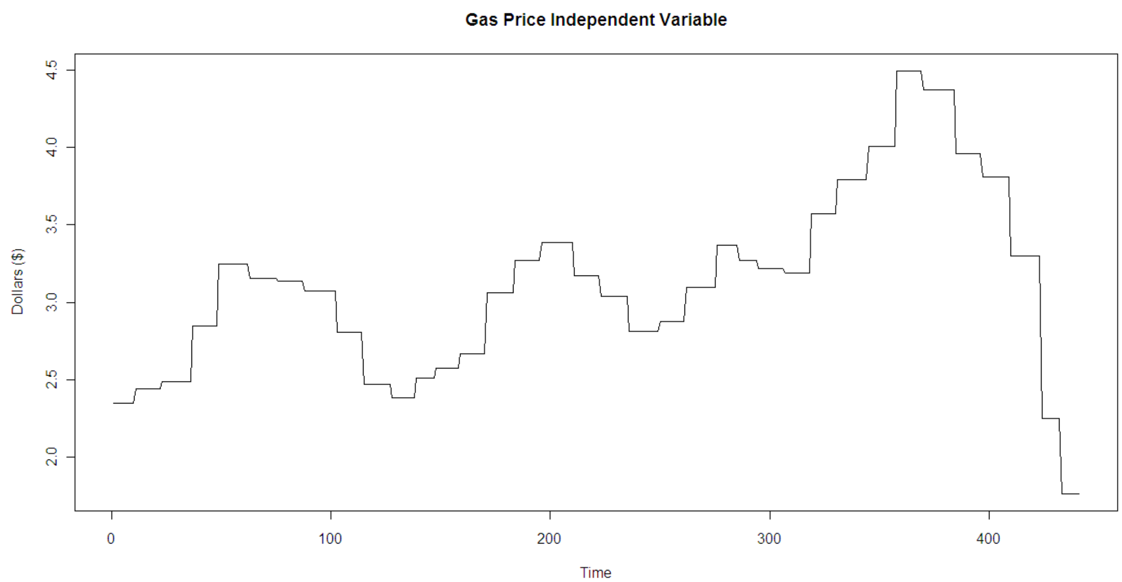


Imputed Regional Transit Weekly PM Ridership

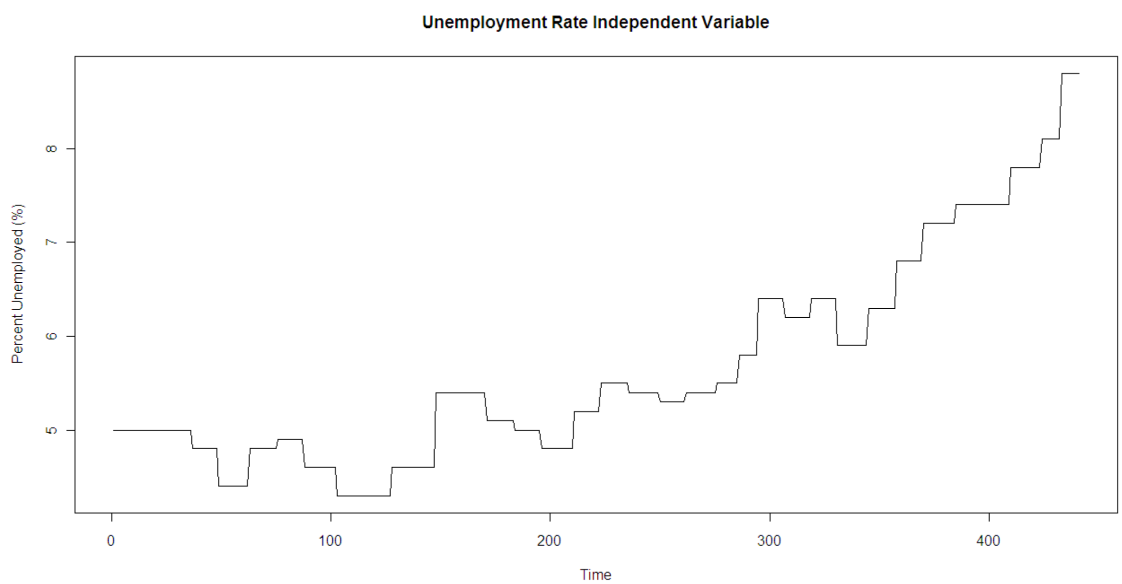


C. Independent Variable Data Sets

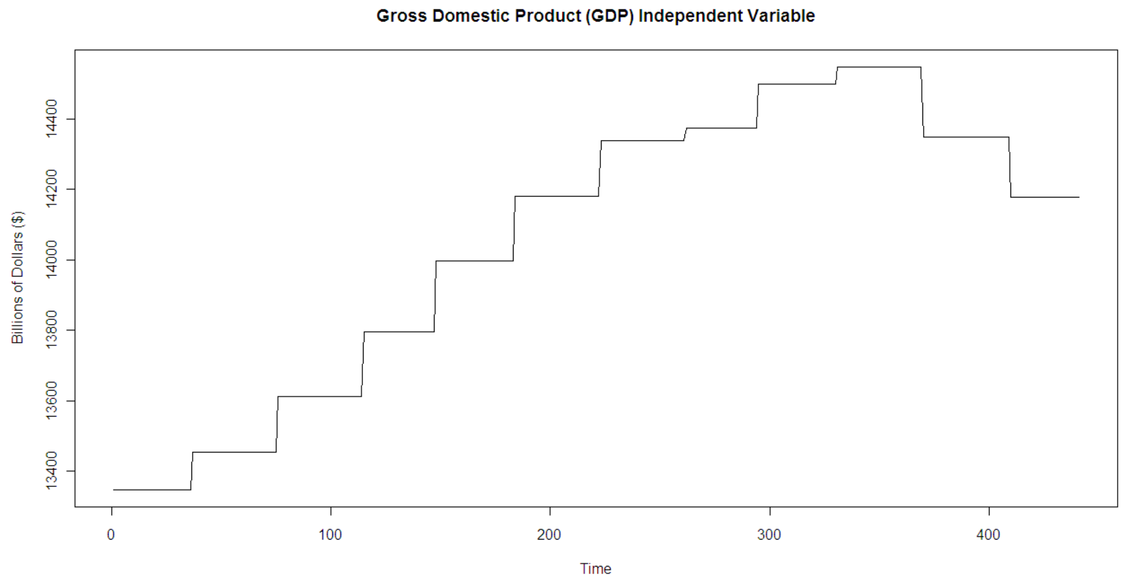
1. Gas Price Independent Variable



2. Unemployment Rate Independent Variable



3. Gross Domestic Product Independent Variable



D. Holiday and Limited Service Imputation Dates

1. Year: 2006

No Service Holidays:

Holiday	Roseville*	Yolobus	Yuba-Sutter*
New Year's Day (Jan 1)		X+	
Day after New Year's (Jan 2)			X
Martin Luther King Jr. Day (Jan 16)	X	Xo	X
President's Day (Feb 20)	X	X+	X
Memorial Day (May 29)	X	X+	X
Independence Day (July 4)	X	X+	X
Labor Day (Sep 4)	X	X+	X
Veteran's Day (Nov 10)	X	Xo	X
Thanksgiving Day (Nov 23)	X	X+	X
Day after Thanksgiving (Nov 24)		Xo	X
Christmas Day (Dec 25)	X	X+	X

* New Year's Day holiday not applicable since it fell on a Sunday and these agencies only operate commuter services during the week

+ Except for routes 40, 41, 42A/B and 240 which operated Sunday schedules

o Except for routes 39, 40, 41, 42A/B and 240 which ran full services

Limited Service Holidays:

Holiday	Roseville	Yolobus	Yuba-Sutter
Day after New Year's (Jan 2)	X		
Cesar Chavez Day (Mar 31)	X		
Columbus Day (Oct 9)	X		
Day after Thanksgiving (Nov 24)	X		
Day after Christmas (Dec 26)	X		

Full service days that coincide with state holidays:

- Yolobus operated full services on Cesar Chavez Day (Mar 31) and Columbus Day (Oct 9)
- Yuba-Sutter operated full services on Cesar Chavez Day (Mar 31) and Columbus Day (Oct 9)

California State Holidays:

Date Observed in 2006	**Holiday**
Monday, January 2	New Year's Day* (observed)
Monday, January 16	Martin Luther King Jr. Day
Monday, February 13	Lincoln's Birthday* (observed)
Monday, February 20	Washington's Birthday
Friday, March 31	Cesar Chavez Day
Monday, May 29	Memorial Day
Tuesday, July 4	Independence Day
Monday, September 4	Labor Day
Monday, October 9	Columbus Day
Friday, November 10	Veteran's** Day (observed)
Thursday, November 23	Thanksgiving Day
Friday, November 24	Day after Thanksgiving
Monday, December 25	Christmas Day

2. Year: 2007
No Service Holidays:

Holiday	Roseville	Yolobus	Yuba-Sutter
New Year's Day (Jan 1)	X	X+	X
Martin Luther King Jr. Day (Jan 15)	X	Xo	X
President's Day (Feb 19)	X	X+	X
Memorial Day (May 28)	X	X+	X
Independence Day (July 4)	X	X+	X
Labor Day (Sep 3)	X	X+	X
Veteran's Day (Nov 12)	X	Xo	X
Thanksgiving Day (Nov 22)	X	X+	X
Day after Thanksgiving (Nov 23)		Xo	X
Christmas Day (Dec 25)	X	X+	X

+ Except for routes 40, 41, 42A/B and 240 which operated Sunday schedules
o Except for routes 39, 40, 41, 42A/B and 240 which ran full services

Limited Service Holidays:

Holiday	Roseville	Yolobus	Yuba-Sutter
Day after New Year's (Jan 2)	X		
Lincoln's Birthday (Feb 12)			
Cesar Chavez Day (Mar 30)	X		
Columbus Day (Oct 8)	X		
Day after Thanksgiving (Nov 23)	X		
Christmas Eve Day (Dec 24)	X		
Day after Christmas (Dec 26)	X		
New Year's Eve Day (Dec 31)	X		

Full service days that coincide with state holidays:

- Roseville Transit operated full service on Lincoln's Birthday (Feb 12)
- Yolobus operated full services on Lincoln's Birthday (Feb 12), Cesar Chavez Day (Mar 30) and Columbus Day (Oct 8)
- Yuba-Sutter operated full services on Lincoln's Birthday (Feb 12), Cesar Chavez Day (Mar 30) and Columbus Day (Oct 8)

California State Holidays:

Date Observed in 2007	Holiday
Monday, January 1	New Year's Day
Monday, January 15	Martin Luther King Jr. Day
Monday, February 12	Lincoln's Birthday
Monday, February 19	Washington's Birthday
Monday, May 28	Memorial Day
Wednesday July 4	Independence Day
Monday, September 3	Labor Day
Monday, October 8	Columbus Day
Monday, November 12	Veteran's Day** (observed)
Thursday, November 22	Thanksgiving Day
Friday, November 23	Day after Thanksgiving
Tuesday, December 25	Christmas Day

3. Year: 2008
No Service Holidays:

Holiday	Natomas	RT*	Roseville	Yolobus	Yuba-Sutter
New Year's Day (Jan 1)	X	X	X	X^{+}	X
Martin Luther King Jr. Day (Jan 21)	X	X	X	X^{o}	X
President's Day (Feb 18)	X		X	X^{+}	X
Cesar Chavez Day (Mar 31)	X				
Memorial Day (May 26)	X	X	X	X^{+}	X
Independence Day (July 4)	X	X	X	X^{+}	X
Labor Day (Sep 1)	X	X	X	X^{+}	X
Veteran's Day (Nov 11)	X		X	X^{o}	X
Thanksgiving Day (Nov 27)	X	X	X	X^{+}	X
Day after Thanksgiving (Nov 28)	X			X^{o}	X
Christmas Day (Dec 25)	X	X	X	X^{+}	X

* Except for routes 15, 30, 34, 38, 51, 67, 68, 86 and 88 which operated a Sunday/Holiday schedule

$^{+}$ Except for routes 40, 41, 42A/B and 240 which operated Sunday schedules

^{o} Except for routes 39, 40, 41, 42A/B and 240 which ran full services

Limited Service Holidays:

Holiday	Natomas	RT*	Roseville	Yolobus	Yuba-Sutter
Day after New Year's (Jan 2)			X		
Lincoln's Birthday (Feb 12)		X			
President's Day (Feb 18)		X			
Cesar Chavez Day (Mar 31)		X	X		
Columbus Day (Oct 13)		X	X		
Veteran's Day (Nov 11)		X			
Day after Thanksgiving (Nov 28)		X	X		
Christmas Eve Day (Dec 24)			X		
Day after Christmas (Dec 26)			X		
New Year's Eve Day (Dec 31)			X		

*Only applicable for routes 3, 7, 29 and 109; all other routes ran full services

Full service days that coincide with state holidays:

- North Natomas TMA operated full services on Lincoln's Birthday (Feb 12) and Columbus Day (Oct 13)
- Roseville Transit operated full services on Lincoln's Birthday (Feb 12)
- Yolobus operated full services on Lincoln's Birthday (Feb 12), Cesar Chavez Day (Mar 31) and Columbus Day (Oct 13)
- Yuba-Sutter operated full services on Lincoln's Birthday (Feb 12), Cesar Chavez Day (Mar 31) and Columbus Day (Oct 13)

California State Holidays:

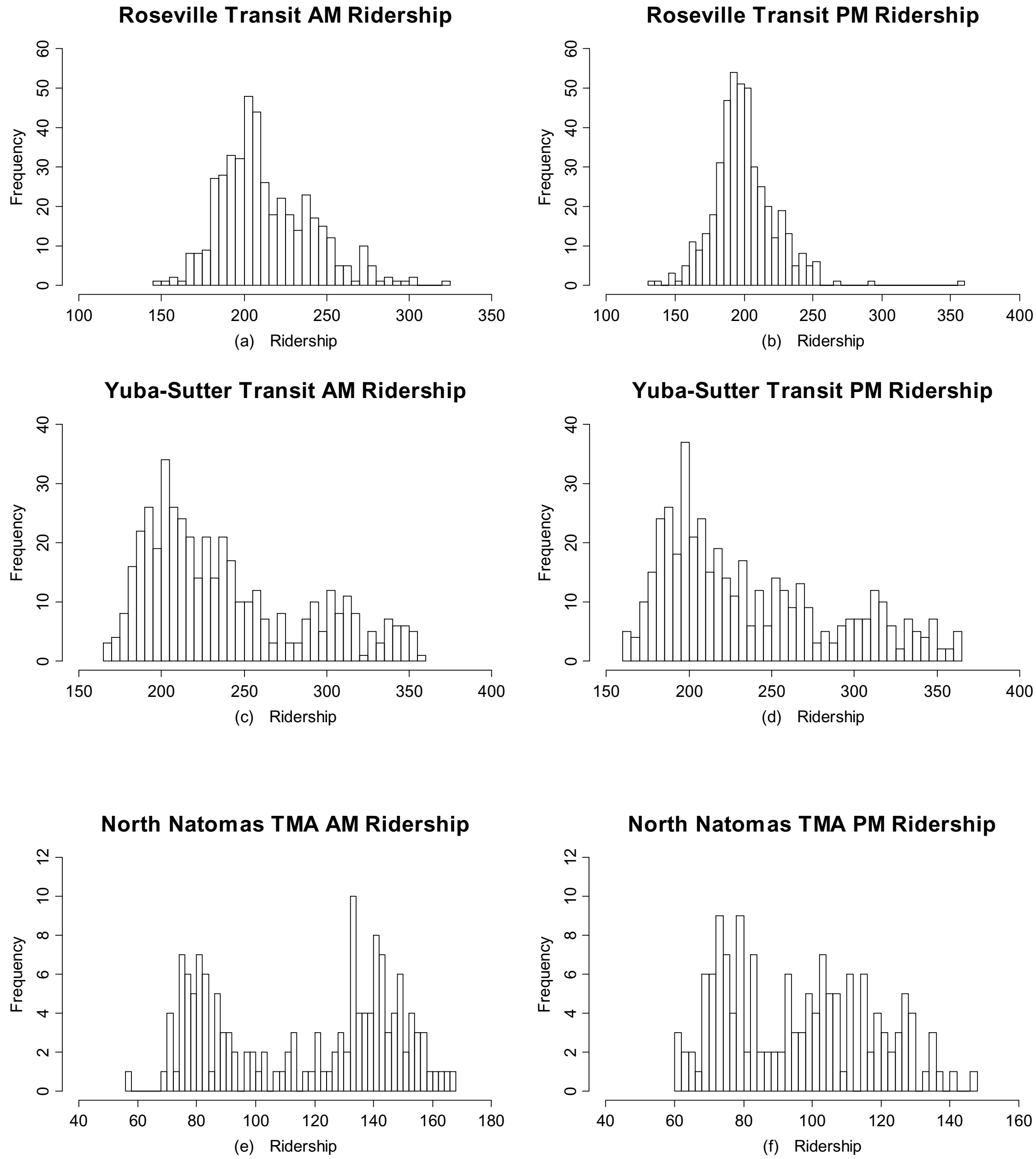
Date Observed in 2008	**Holiday**
Tuesday, January 1	New Year’s Day
Monday, January 21	Birthday of Martin Luther King, Jr.
Tuesday, February 12	Lincoln’s Birthday
Monday, February 18	Washington's Birthday
Monday, March 31	Cesar Chavez Day
Monday, May 26	Memorial Day
Friday, July 4	Independence Day
Monday, September 1	Labor Day
Monday, October 13	Columbus Day
Tuesday, November 11	Veterans Day
Thursday, November 27	Thanksgiving Day
Friday, November 28	Day after Thanksgiving
Thursday, December 25	Christmas Day

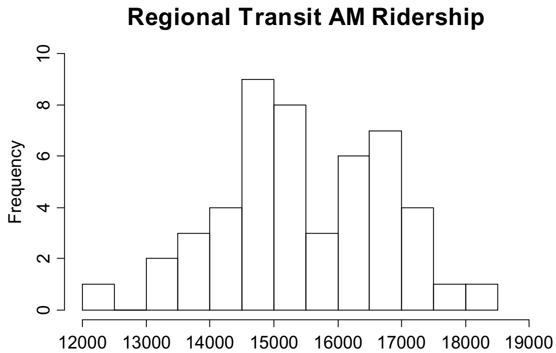
E. Ad Hoc Data Imputation Method Details

Yolobus							
"Missing" Data Cases	"Missing" Data Date	Observation Week	Ridership			Method 1** abs(T1-T2)	Method 2*** abs(T1'-T2')
			Tuesday	Wednesday	Thursday		
1	4-Jul-06	26	2987	2889	2891		
		27	807	2993	2984	9	285
		28	3272	2918	3173		
2	23-Nov-06	46	3177	3097	2975		
		47	3135	3049	730	86	179
		48	3217	2946	3154		
3	4-Jul-07	78	3540	3205	3299		
		79	3413	1014	3203	210	279
		80	3643	3484	3405		
4	22-Nov-07	98	3535	3524	3510		
		99	3450	3415	899	35	95
		100	3389	3216	3415		
5*	25-Dec-07	103	2941	3023	3162		
		104	484	2336	2486	150	2078
		105	863	2987	2903		
6*	1-Jan-08	104	484	2336	2486		
		105	863	2987	2903	84	2611
		106	3095	3235	3235		
7	11-Nov-08	149	4548	4659	4483		
		150	2696	4386	4313	73	77
		151	4471	4225	4318		
8	27-Nov-08	151	4471	4225	4318		
		152	4517	4016	847	501	10
		153	4582	4579	4308		
9*	25-Dec-08	155	3914	3744	3655		
		156	3601	2236	635	1365	
		157	3404	2920			
					Sum	914	925
					Average	152.333333	154.166667

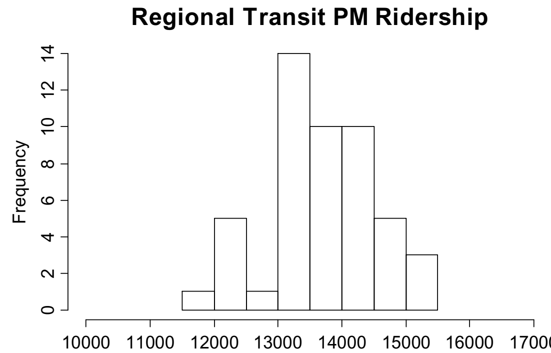
* Week to be removed (Thanksgiving, Christmas and New Year's holidays)
Note: The shaded cells show the original data; these anomalous (holiday or limited service day) cases were treated as missing.
** Method 1 uses adjacent days of the same week for data imputation.
*** Method 2 uses the same days of adjacent weeks for data imputation.

F. Histograms for Each Transit Agency

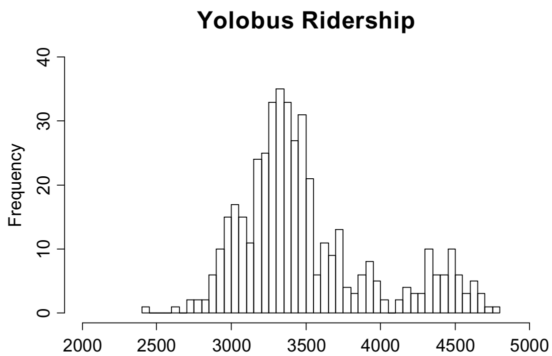




(i) Ridership



(j) Ridership



(k) Ridership

G. Multiple Regression Model Selection

Yuba Sutter Transit Multiple Regression: AM Peak Period

Independent Variable	Model 1		Model 2				Model 3			
	Parameter Estimate (t statistic)	Significance	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	24.42 (14.906)	***	24.43 (14.922)	***	1.450	12.917	27.611 (18.836)	***	1.122	11.236
Unemployment Rate	31.21 (28.913)	***	31.58 (35.454)	***	1.548	13.933	33.486 (43.390)	***	1.122	13.815
GDP	0.009935 (2.630)	**	0.01145 (4.087)	***	2.000	132.547				
Fare Price	4.086 (0.600)									
Intercept	-166.7 (-4.314)	***	-176.9 (-5.096)	***		1.000	-36.599 (-6.979)	***		1.000
Adjusted R-squared	0.8803		0.8805				0.8762			
("X" indicates best model fit)							X			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Yuba Sutter Transit Multiple Regression: PM Peak Period

Independent Variable	Model 1				Model 2			
	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	25.35 (14.292)	***	1.450	14.436	30.108 (18.55)	***	1.122	11.236
Unemployment Rate	31.51 (26.965)	***	2.271	15.304	36.038 (42.17)	***	1.122	13.815
GDP	0.009974 (2.439)	*	3.628	56.549				
Fare Price	18.93 (2.569)	*	4.661	187.683				
Intercept	-221.8 (-5.303)	***		1.000	-60.545 (-10.43)	***		1.000
Adjusted R-squared	0.8804				0.8705			
("X" indicates best model fit)					X			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

		Yolobus Multiple Regression							
		Model 1				Model 2			
		Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Conditio n Index
Independent Variable	Gasoline Price	148.322 (6.584)	***	1.450	12.917	97.37 (4.796)	***	1.122	11.236
	Unemployment Rate	366.568 (29.912)	***	1.548	13.933	335.99 (31.436)	***	1.122	13.815
	GDP	-0.18345 (2.439)	***	2.000	132.547				
	Intercept	3540.017 (7.414)	***		1.000	1293.76 (17.812)	***		1.000
	Adjusted R-squared	0.751				0.7387			
	("X" indicates best model fit)					X			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Independent Variable	Roseville Transit Multiple Regression: AM Peak Period							
	Model 1				Model 2			
	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	20.958 (12.660)	***	1.450	14.419	20.296 (13.71)	***	1.122	11.236
Unemployment Rate	11.877 (10.889)	***	2.271	15.369	13.869 (17.80)	***	1.122	13.815
GDP	-0.0126 (-3.309)	**	3.628	51.752				
Fare Price	27.048 (3.933)	***	4.661	188.803				
Intercept	177.543 (4.458)	***		1.000	71.995 (13.60)	***		1.000
Adjusted R-squared	0.641				0.6287			
("X" indicates best model fit)					x			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Independent Variable	Roseville Transit Multiple Regression: PM Peak Period										
	Model 1		Model 2				Model 3				
	Parameter Estimate (t statistic)	Significance	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Condition Index
Gasoline Price	12.616 (7.773)	***	12.602 (7.763)	***	1.450	12.917	9.482 (6.526)	***	1.122	11.236	11.236
Unemployment Rate	14.040 (13.129)	***	13.385 (15.156)	***	1.548	13.933	11.512 (15.050)	***	1.122	13.815	13.815
GDP	-0.008516 (-2.276)	*	-0.0112 (-4.043)	***	2.000	132.547					
Fare Price	-7.314 (-1.085)										
Intercept	225.031 (5.879)	***	243.225 (7.068)	***		1.000	105.643 (20.323)	***		1.000	1.000
Adjusted R-squared	0.4771		0.4769				0.4585				
("X" indicates best model fit)							x				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	North Natomas T.M.A. Multiple Regression: AM Peak Period									
	Model 1		Model 2				Model 3			
Independent Variable	Parameter Estimate (t statistic)	Significance	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	33.320 (21.225)	***	33.021 (21.63)	***	1.363	8.716	10.404 (3.077)	**	model contains < 2 terms	10.083
Unemployment Rate	36.797 (11.550)	***	39.191 (28.73)	***	1.363	30.587				
GDP	-0.0156 (-0.832)									
Intercept	-35.490 (-0.123)		-275.045 (-20.79)	***		1.000	79.704 (6.489)	***		1.000
Adjusted R-squared	0.8583		0.8586				0.548			
("X" indicates best model fit)			x							

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	North Natomas T.M.A. Multiple Regression: PM Peak Period											
	Model 1				Model 2				Model 3			
Independent Variable	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	21.03 (12.787)	***	1.438	10.006	22.100 (13.48)	***	1.363	8.716	8.268 (3.494)	***	model contains < 2 terms	10.083
Unemployment Rate	32.52 (9.744)	***	7.414	28.087	23.967 (16.36)	***	1.363	30.587				
GDP	0.05557 (2.837)	**	7.825	840.854								
Intercept	-100.6 (-3.331)	***		1.000	-150.043 (-10.56)	***		1.000	66.905 (7.782)	***		1.000
Adjusted R-squared	0.688				0.6728				0.0713			
("X" indicates best model fit)					x							

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Independent Variable	Regional Transit Multiple Regression — AM Peak Period — Model 1: Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	196.475 (0.739)		1.435	10.018
Unemployment Rate	421.563 (0.773)		7.562	28.127
GDP	-1.867 (-0.895)		7.959	851.006
Intercept	3540.017 (7.414)			1.000
Adjusted R-squared	0.2294			
("X" indicates best model fit)				

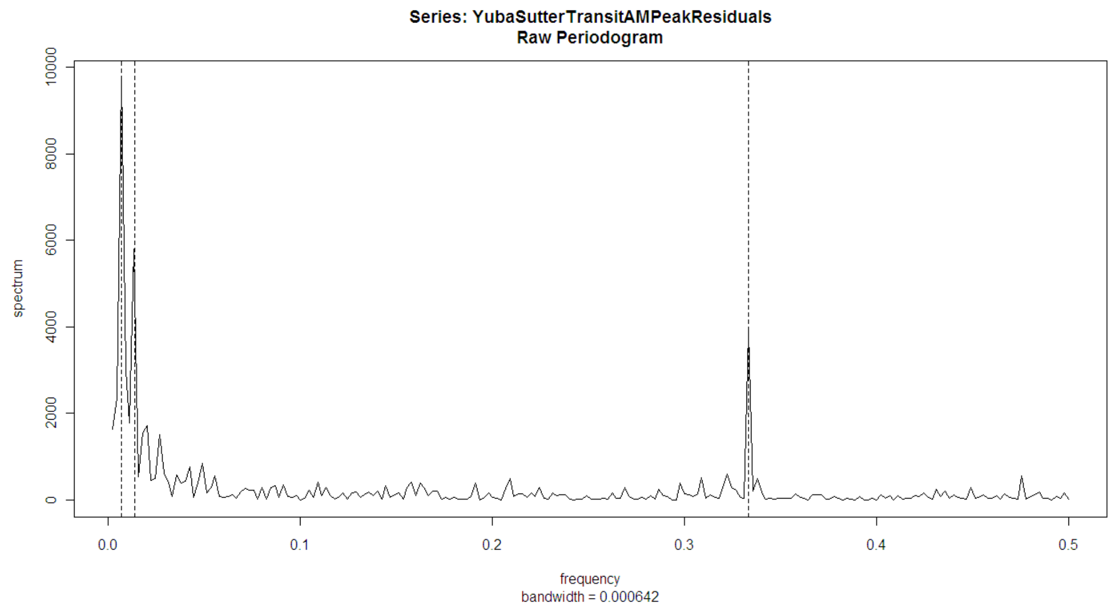
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Regional Transit Multiple Regression									
	PM Peak Period									
	Model 1		Model 2				Model 3			
Independent Variable	Parameter Estimate (t statistic)	Significance	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index	Parameter Estimate (t statistic)	Significance	Variance Inflation Factor	Condition Index
Gasoline Price	539.723 (3.103)	**	542.182 (3.096)	**	1.435	10.863	209.0 (1.287)		model contains < 2 terms	10.083
Unemployment Rate	-457.311 (-1.282)									
GDP	-5.529 (-2.638)	*	-3.096 (-3.455)	**	1.435	298.156				
Intercept	-35.490 (0.123)	**	56292.907 (4.478)	***		1.000	12895.1 (21.863)	***		1.000
Adjusted R-squared	0.2108		0.1998				0.01351			
("X" indicates best model fit)							x			

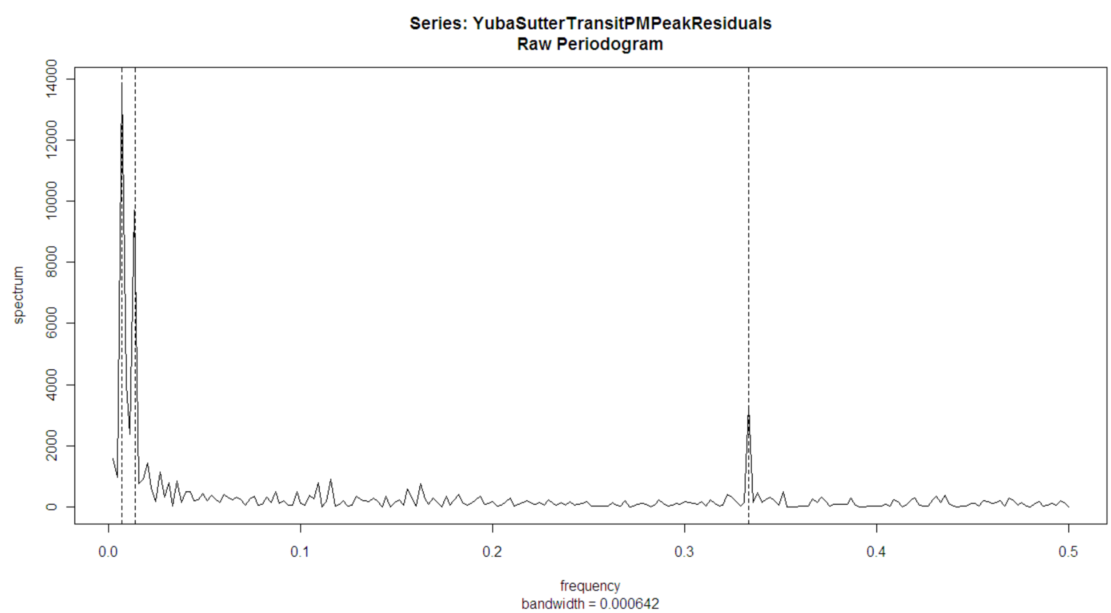
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

H. Transit Agency Periodograms

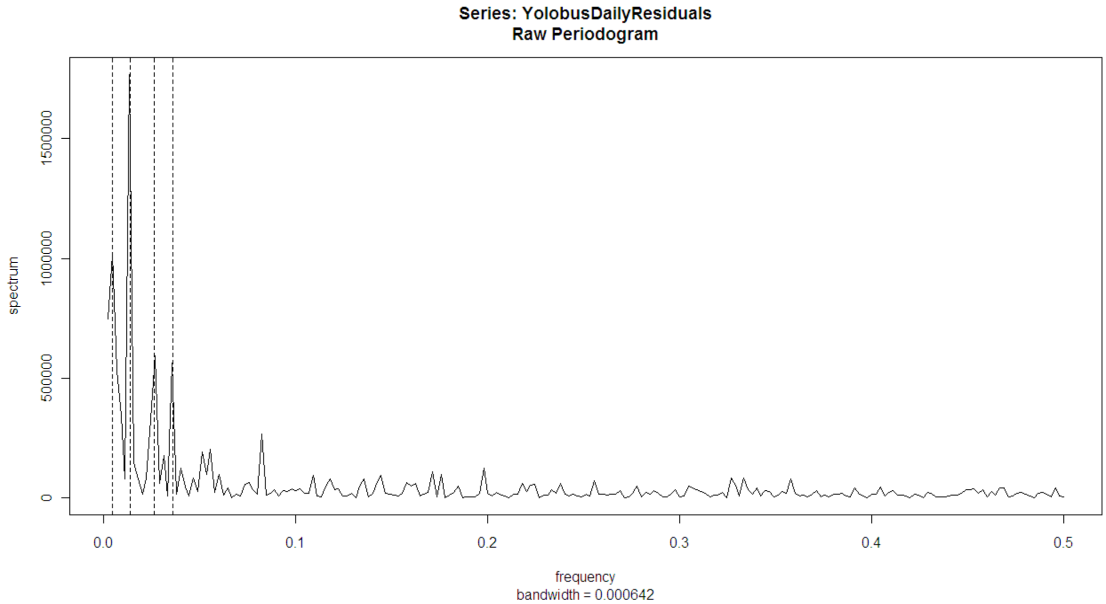
1. Yuba-Sutter AM Periodogram



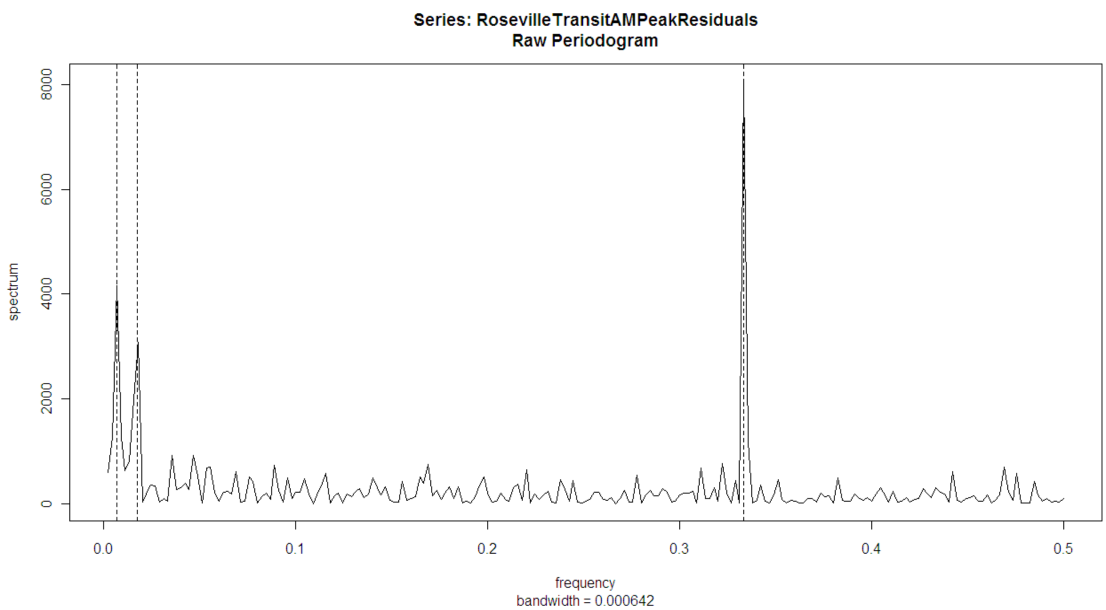
2. Yuba-Sutter PM Periodogram



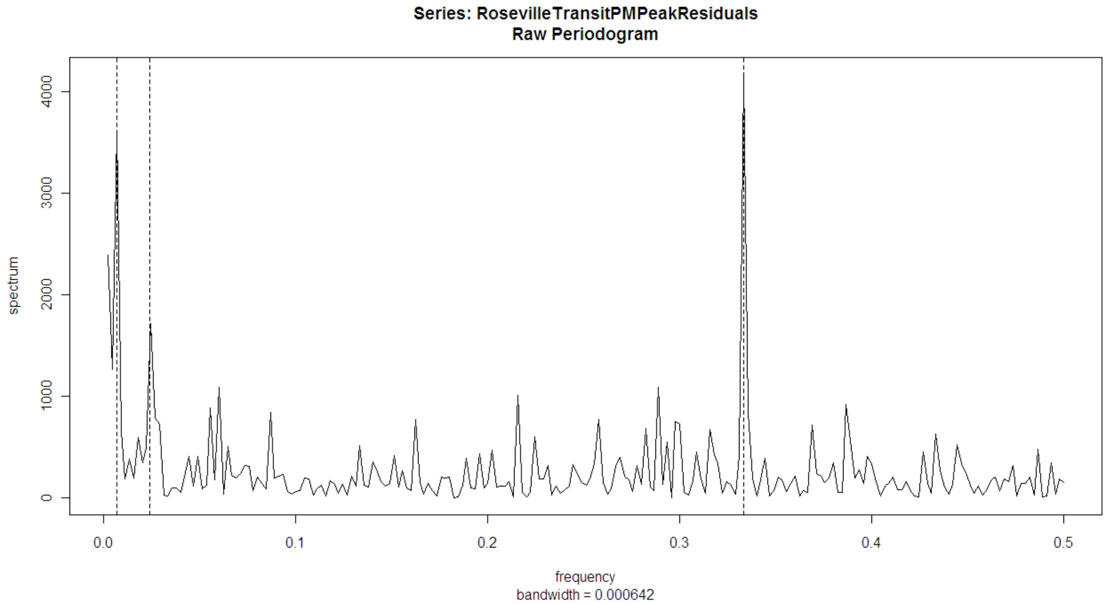
3. Yolobus Daily Periodogram



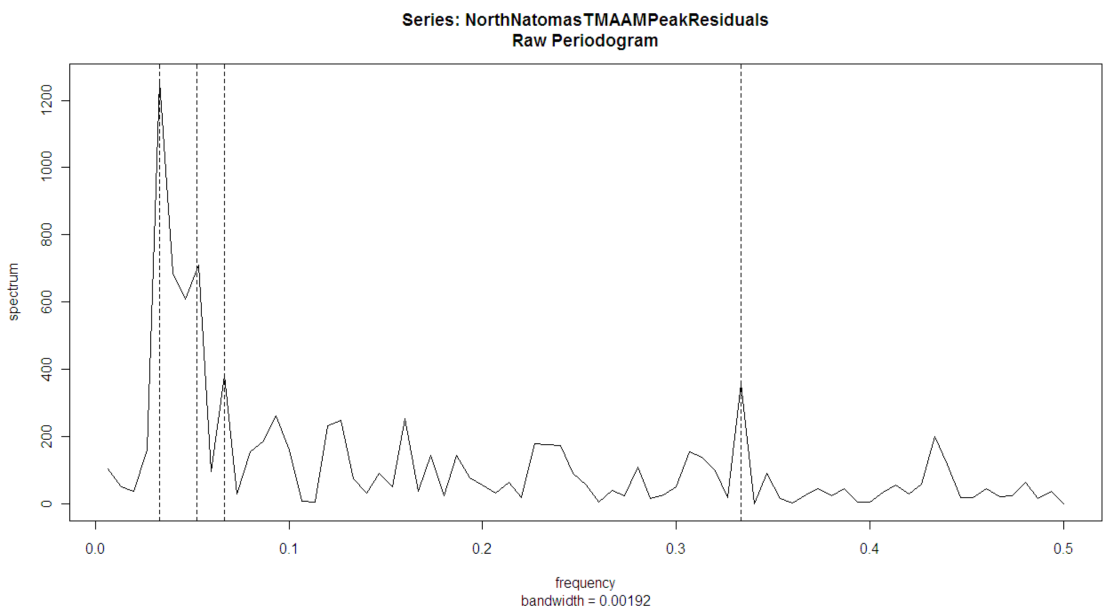
4. Roseville Transit AM Periodogram



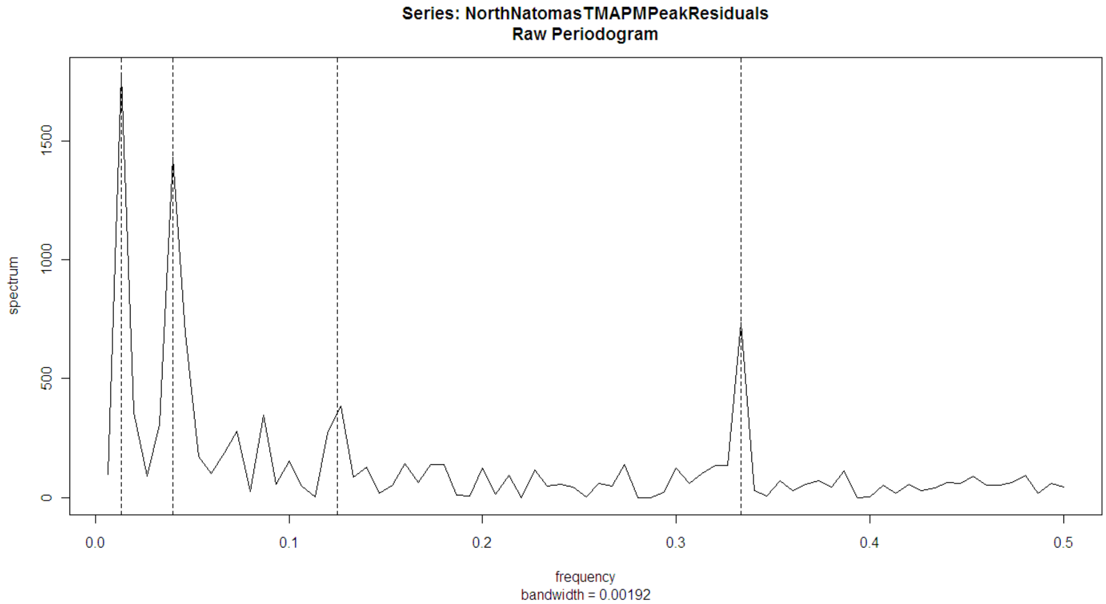
5. Roseville Transit PM Periodogram



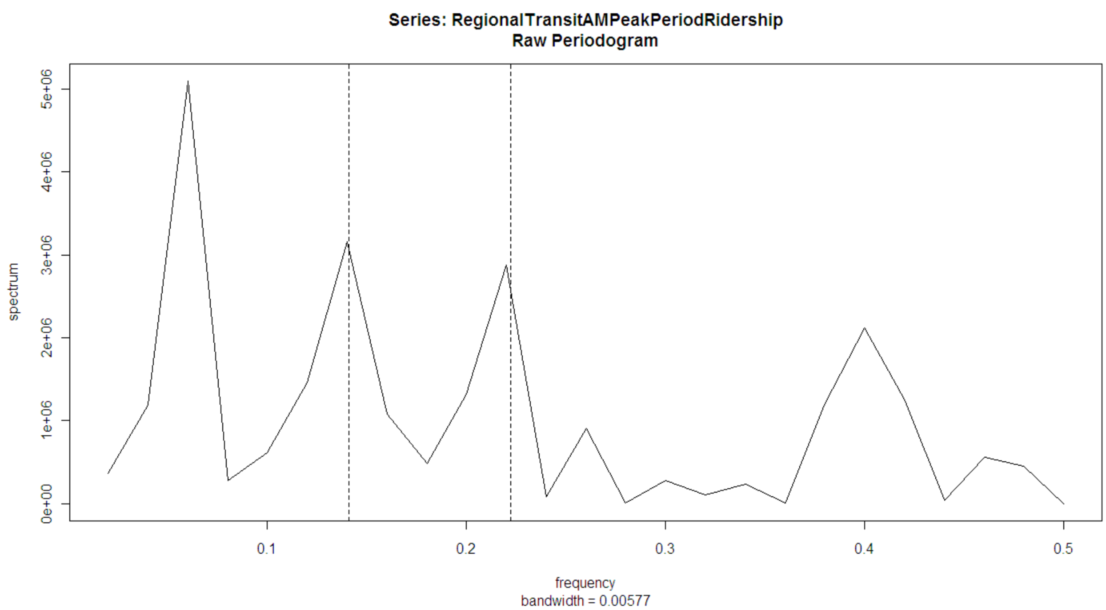
6. North Natomas AM Periodogram



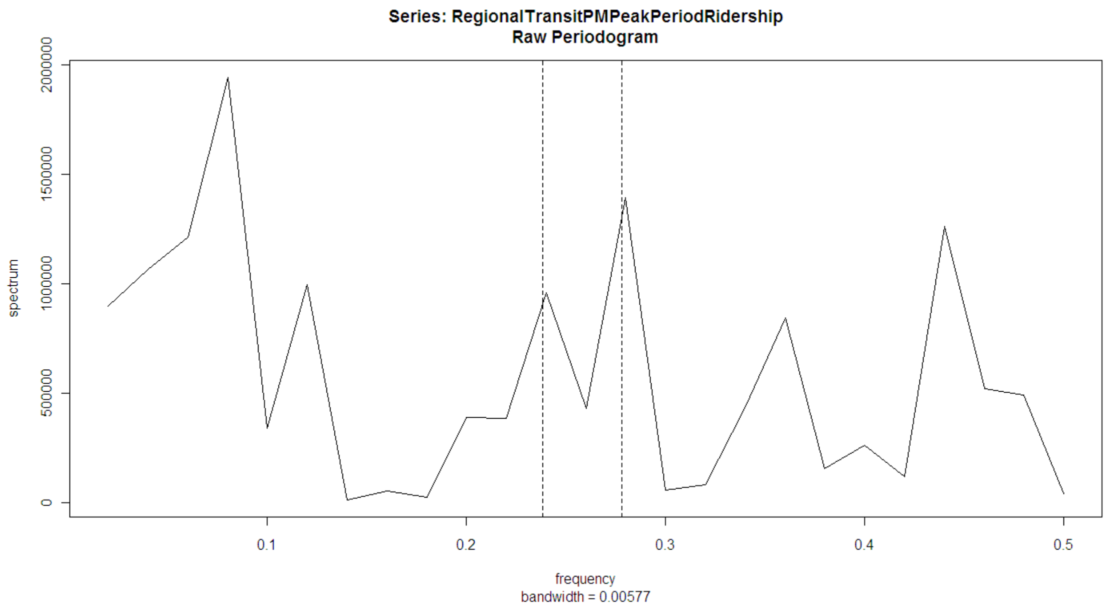
7. North Natomas PM Periodogram



8. Regional Transit AM Periodogram



9. Regional Transit PM Periodogram



I. Intervention Analysis Model Results

Intervention Analysis			Transfer Model		Noise Model			Intervention Model		
Transit Agency	Data Aggregation	Indicator Function	Initial Model Estimate b (AICC)	Model Estimate b (AICC)	ARMA model (p,q) (AICC)	AR Coefficients ϕ_p (Standard Error)	MA Coefficients θ_1 (Standard Error)	Final Model Estimate b (AICC)	AR Coefficients	MA Coefficients
Yuba-Sutter Transit	AM	Pulse	1 (77951.30)	0.4000 (4362.63)	(5,5) (3389.14)	$\phi_1=-0.445003$ (0.088362)	$\theta_1=0.7949$(0.089654)	-7.400 (4243.74)	$\phi_1=0.1000$	$\theta_1=0.2500$
						$\phi_2=-0.553716$ (0.079330)	$\theta_2=0.969499$ (0.115772)		$\phi_2=0.1000$	$\theta_2=0.1000$
						$\phi_3=0.356188$ (0.084764)	$\theta_3=0.137348$ (0.14561)		$\phi_3=0$	$\theta_3=0.1000$
						$\phi_4=0.376708$ (0.067170)	$\theta_4=-0.08698$ (0.11446)		$\phi_4=-0.1000$	$\theta_4=0.3000$
						$\phi_5=0.619758$ (0.062173)	$\theta_5=-0.610824$ (0.065985)		$\phi_5=0.4500$	$\theta_5=-0.3500$
Yuba-Sutter Transit	AM	Step	1 (78200.40)	-0.65000 (4422.45)	(5,5) (3393.98)	$\phi_1=-0.844493$ (0.185336)	$\theta_1=1.185453$(0.182562)	-3.950 (4304.54)	$\phi_1=0.1000$	$\theta_1=0.2500$
						$\phi_2=-0.614862$ (0.190428)	$\theta_2=1.153211$ (0.250482)		$\phi_2=0.1000$	$\theta_2=0.1000$
						$\phi_3=0.506459$ (0.107935)	$\theta_3=0.113551$ (0.220307)		$\phi_3=0$	$\theta_3=0.1000$
						$\phi_4=0.608404$ (0.125308)	$\theta_4=-0.224856$ (0.131760)		$\phi_4=-0.1000$	$\theta_4=-0.3000$
						$\phi_5=0.501889$ (0.117394)	$\theta_5=-0.464088$ (0.087926)		$\phi_5=0.4500$	$\theta_5=-0.3500$
Yuba-Sutter Transit	PM	Pulse	1 (87892.30)	2.800 (4416.29)	(5,5) (3496.89)	$\phi_1=-0.158722$ (0.182004)	$\theta_1=0.483780$(0.175464)	-6.250 (4344.26)	$\phi_1=0.1000$	$\theta_1=0.2000$
						$\phi_2=-0.549409$ (0.190594)	$\theta_2=0.852613$ (0.218845)		$\phi_2=0.0500$	$\theta_2=0.1500$
						$\phi_3=0.642854$ (0.110353)	$\theta_3=-0.333125$ (0.207550)		$\phi_3=.3000$	$\theta_3=-0.2000$
						$\phi_4=-0.071348$ (0.135640)	$\theta_4=0.114305$ (0.141741)		$\phi_4=-0.2000$	$\theta_4=-0.3000$
						$\phi_5=0.548031$ (0.132410)	$\theta_5=-0.532052$ (0.093492)		$\phi_5=0.1000$	$\theta_5=-0.1500$
Yuba-Sutter Transit	PM	Step	1 (88496.10)	-2.100 (4475.39)	(5,5) (3494.86)	$\phi_1=-0.095913$ (0.260545)	$\theta_1=0.421775$(0.252383)	-6.200 (4402.08)	$\phi_1=0.1500$	$\theta_1=0.1500$
						$\phi_2=-0.367560$ (0.265878)	$\theta_2=0.658712$ (0.300076)		$\phi_2=-0.0500$	$\theta_2=0.2500$
						$\phi_3=0.689538$ (0.136619)	$\theta_3=-0.451876$ (0.257691)		$\phi_3=0.3000$	$\theta_3=-0.1500$
						$\phi_4=-0.148921$ (0.191405)	$\theta_4=0.109098$ (0.189802)		$\phi_4=0.3000$	$\theta_4=-0.3500$
						$\phi_5=0.399829$ (0.186089)	$\theta_5=-0.434819$ (0.121521)		$\phi_5=0$	$\theta_5=-0.05000$

Intervention Analysis			Transfer Model		Noise Model			Intervention Model		
Transit Agency	Data Aggregation	Indicator Function	Initial Model Estimate b (AICC)	Model Estimate b (AICC)	ARMA model (p,q) (AICC)	AR Coefficients ϕp (Standard Error)	MA Coefficients $\theta 1$ (Standard Error)	Final Model Estimate b (AICC)	AR Coefficients	MA Coefficients
Yolobus	Daily	Pulse	1 (16037200.00)	38.35 (6720.16)	(3,3) (5762.60)	$\phi 1$=-0.428121 (0.051051)	$\theta 1$=0.772004(0.076086)	-13.650 [illegible](6607.04)	$\phi 1$=0.1000	$\theta 1$=0.2500
						$\phi 2$=0.047334 (0.062840)	$\theta 2$=0.340901 (0.097768)		$\phi 2$=0.1000	$\theta 2$=0.1000
						$\phi 3$=0.812395 (0.049406)	$\theta 3$=-0.512835 (0.076081)		$\phi 3$=0.4000	$\theta 3$=-0.2000
Yolobus	Daily	Step	1 (16031300.00)	46.35 (6776.14)	(1,1) (5765.06)	$\phi 1$=0.813987 (0.056755)	$\theta 1$=-0.475163(0.087644)	31.250 (5666.56)	$\phi 1$=0.6500	$\theta 1$=-0.2500
Roseville Transit	AM	Pulse	1 (98911.10)	0.7000 (4469.70)	(3,3) (3599.40)	$\phi 1$=-0.393116 (0.078331)	$\theta 1$=0.602423 (0.099897)	-1.150 (4449.79)	$\phi 1$=0.1000	$\theta 1$=0.1000
						$\phi 2$=0.083926 (0.092672)	$\theta 2$=0.121255 (0.128440)		$\phi 2$=0	$\theta 2$=0.1000
						$\phi 3$=0.785926 (0.070458)	$\theta 3$=-0.683218 (0.098871)		$\phi 3$=0	$\theta 3$=0.1000
Roseville Transit	AM	Step	1 (99070.20)	-0.5000 (4523.75)	(3,3) (3589.17)	$\phi 1$=-0.456487 (0.096760)	$\theta 1$=0.687583 (0.119019)	-0.450 (4509.82)	$\phi 1$=0.1000	$\theta 1$=0.1000
						$\phi 2$=0.007258 (0.115468)	$\theta 2$=0.229019 (0.153139)		$\phi 2$=0	$\theta 2$=0.1000
						$\phi 3$=0.731301 (0.087855)	$\theta 3$=-0.599024 (0.118995)		$\phi 3$=0	$\theta 3$=0.1000
Roseville Transit	PM	Pulse	1 (105325.00)	-3.350 (4495.93)	(4,1) (3641.89)	$\phi 1$=1.015030 (0.057432)	$\theta 1$=-0.941892 (0.034943)	-3.750 (4483.04)	$\phi 1$=0.1000	$\theta 1$=0
						$\phi 2$=-0.064694 (.067206)			$\phi 2$=0	
						$\phi 3$=0.156347 (.067111)			$\phi 3$=0.2000	
						$\phi 4$=-0.127175 (.050744)			$\phi 4$=-0.5000	
Roseville Transit	PM	Step	1 (106003.00)	1.700(4556.08)	(4,1) (3393.98)	$\phi 1$=1.022552 (0.056744)	$\theta 1$=-0.947341(0.034933)	1.500 (4543.46)	$\phi 1$=0.0500	$\theta 1$=0.0500
						$\phi 2$=-0.064116 (0.067428)			$\phi 2$=0	
						$\phi 3$=0.154332 (0.067371)			$\phi 3$=0.2000	
						$\phi 4$=-0.131703 (0.050795)			$\phi 4$=0.1000	

Intervention Analysis			Transfer Model		Noise Model			Intervention Model		
Transit Agency	Data Aggregation	Indicator Function	Initial Model Estimate b (AICC)	Model Estimate b (AICC)	ARMA model (p,q) (AICC)	AR Coefficients ϕp (Standard Error)	MA Coefficients θ1 (Standard Error)	Final Model Estimate b (AICC)	AR Coefficients	MA Coefficients
North Natomas TMA	AM	Pulse	1 (13831.70)	0.1000 (1377.58)	(5,2) (1060.93)	ϕ1=2.153686 (0.083018)	θ1=-1.906989 (0.011761)	-0.850 (1369.64)	ϕ1=0.1500	θ1=0.2000
						ϕ2=-1.556697 (0.196312)	θ2=0.991083 (0.011727)		ϕ2=-0.1000	θ2=0.1500
						ϕ3=0.390641 (0.232365)			ϕ3=0.5000	
						ϕ4=0.012665 (0.196243)			ϕ4=-0.0500	
						ϕ5=-0.071214 (0.062924)			ϕ5=0	
North Natomas TMA	AM	Step	1 (13775.10)	1.400 (1435.80)	(1,0) (1060.75)	ϕ1=0.366248 (0.077196)		1.200 (1416.96)	ϕ1=0.3500	
North Natomas TMA	PM	Pulse	1 (13014.90)	2.950 (1366.79)	(2,2) (1057.00)	ϕ1=1.672595 (0.118292)	θ1=-1.436799 (0.164590)	2.600 (1358.94)	ϕ1=0.2000	θ1=0.1000
						ϕ2=-0.744316 (0.119998)	θ2=0.482199 (0.157398)		ϕ2=0.1000	θ2=0
North Natomas TMA	PM	Step	1 (13253.80)	0.1000 (1429.27)	(1,0) (1058.87)	ϕ1=0.324071 (0.078070)		0.0500 (1415.17)	ϕ1=0.3000	
Regional Transit	AM	Pulse	1 (54085000.00)	-0.4340 (931.379)	(3,3) (815.387)	ϕ1=-0.723440 (0.325700)	θ1=1.170438 (0.486214)	762.20 (921.889)	ϕ1=0	θ1=0.3500
						ϕ2=0.230768 (0.322303)	θ2=0.652557 (0.580528)		ϕ2=0.6500	θ2=-0.3000
						ϕ3=0.750456 (0.198139)	θ3=-0.169992 (0.506359)		ϕ3=0.0500	θ3=0.3000
Regional Transit	AM	Step	1 (54055500.00)	514.60 (945.367)	(3,3) (813.324)	ϕ1=-0.864145 (0.254531)	θ1=1.367380 (0.294042)	552.60 (941.554)	ϕ1=-0.1500	θ1=0.4500
						ϕ2=0.038451 (0.304989)	θ2=0.931355 (0.346533)		ϕ2=0.6000	θ2=-0.2500
						ϕ3=0.619069 (0.178947)	θ3=0.046151 (0.276103)		ϕ3=0.0500	θ3=0.2500

Intervention Analysis			Transfer Model		Noise Model			Intervention Model		
Transit Agency	Data Aggregation	Indicator Function	Initial Model Estimate b (AICC)	Model Estimate b (AICC)	ARMA model (p,q) (AICC)	AR Coefficients ϕp (Standard Error)	MA Coefficients θ1 (Standard Error)	Final Model Estimate b (AICC)	AR Coefficients	MA Coefficients
Regional Transit	PM	Pulse	1 (26935100.00)	233.90 (888.122)	(0,5) (774.776)		θ1=0.115495 (0.088153)	456.60 (880.193)		θ1=0.0500
							θ2=0.918939 (0.059602)			θ2=0.8000
							θ3=0.468561 (0.127693)			θ3=0.4000
							θ4=0.469089 (0.059602)			θ4=0.6000
							θ5=0.786910 (0.088153)			θ5=0.4000
Regional Transit	PM	Step	1 (26927100.00)	210.10 (906.624)	(0,5) (776.078)		θ1=0.132875 (0.085714)	351.10 (899.168)		θ1=0.1500
							θ2=0.924971 (0.057584)			θ2=0.7500
							θ3=0.502574 (0.124987)			θ3=0.5500
							θ4=0.463867 (0.057584)			θ4=0.5000
							θ5=0.800003 (0.085714)			θ5=0.5500

J. Goodness-of-fit Tests

		Yuba Sutter Transit			
		AM Peak Period		PM Peak Period	
		PULSE INPUT INDICATOR	STEP INPUT INDICATOR	PULSE INPUT INDICATOR	STEP INPUT INDICATOR
Residual Tests	The Sample ACF (# lags outside 95% bounds)	2	3	0	0
	The Portmanteau Test (Ljung-Box) (p-value)	0.08445	0.06974	0.97388	0.96028
	The Rank Test (p-value)	0.48041	0.29893	0.38718	0.10814
	Tests for Normality (R^2)	0.996501	0.996709	0.981917	0.982649
Model Selection	("X" indicates best model fit)	X		X	

		Yolobus	
		Daily Ridership	
		PULSE INPUT INDICATOR	STEP INPUT INDICATOR
Residual Tests	The Sample ACF (# lags outside 95% bounds)	1	2
	The Portmanteau Test (Ljung-Box) (p-value)	0.86634	0.26009
	The Rank Test (p-value)	0.13297	0.04008
	Tests for Normality (R^2)	0.972679	0.974811
Model Selection	("X" indicates best model fit)	X	

		Roseville Transit			
		AM Peak Period		PM Peak Period	
		PULSE INPUT INDICATOR	STEP INPUT INDICATOR	PULSE INPUT INDICATOR	STEP INPUT INDICATOR
Residual Tests	The Sample ACF (# lags outside 95% bounds)	1	1	1	1
	The Portmanteau Test (Ljung-Box) (p-value)	0.59792	0.58891	0.46587	0.4959
	The Rank Test (p-value)	0.34501	0.36623	0.14328	0.04078
	Tests for Normality (R^2)	0.986268	0.98624	0.921784	0.919529
Model Selection	("X" indicates best model fit)		X	X	

		North Natomas T.M.A.			
		AM Peak Period		PM Peak Period	
		PULSE INPUT INDICATOR	STEP INPUT INDICATOR	PULSE INPUT INDICATOR	STEP INPUT INDICATOR
Residual Tests	The Sample ACF (# lags outside 95% bounds)	6	3	0	0
	The Portmanteau Test (Ljung-Box) (p-value)	0.06878	0.07897	0.87743	0.7747
	The Rank Test (p-value)	0.28453	0.3272	0.57247	0.4561
	Tests for Normality (R^2)	0.9895	0.990658	0.98097	0.986188
Model Selection	("X" indicates best model fit)		X	X	

		Regional Transit			
		AM Peak Period		PM Peak Period	
		PULSE INPUT INDICATOR	STEP INPUT INDICATOR	PULSE INPUT INDICATOR	STEP INPUT INDICATOR
Residual Tests	The Sample ACF (# lags outside 95% bounds)	0	1	0	0
	The Portmanteau Test (Ljung-Box) (p-value)	0.54685	0.49637	0.94061	0.98919
	The Rank Test (p-value)	0.27744	0.37928	0.54625	0.95875
	Tests for Normality (R^2)	0.98812	0.990232	0.971005	0.983763
Model Selection	("X" indicates best model fit)		X		X