A Report to the Arizona Department of Transportation

Forecast and Capacity Planning for Nogales' Ports of Entry

J. René Villalobos, Arnold Maltz*, Liangjie Xue, Octavio Sánchez, Laura Vázquez

Department of Industrial Engineering

Ira A. Fulton School of Engineering

***Department of Supply Chain**

WP Carey School of Business

Table of Contents

Table of Figures

Table of Tables

1 Executive Summary

This document provides the final report of the activities performed under the project Nogales POEs Traffic Study: Forecast and Capacity Planning for Nogales' Ports of Entry sponsored by the Arizona Department of Transportation (ADOT) under Grant JPA 08-024T. Some of the main activities of this study include:

- A baseline analysis of the Nogales Ports of Entry (POEs), Mariposa POE and DeConcini POE. Including analysis of historical data for these POEs, a visit to the Mariposa POE and conclusions gathered from any relevant previous studies.
  - Different types of traffic were investigated, including commercial traffic (mainly truck), POV (Privately Owned Vehicle), pedestrian, bus and train.
  - Through our analysis, we discovered that the truck traffic contained a very strong seasonality pattern while other modes of traffic did not.
  - Previous to our study, there were not many studies dedicated to forecasting border crossing traffic.
  - None of the studies we reviewed had dealt with the seasonality pattern we observed here.
  - Economic indices were usually incorporated in the models; however, one should take caution when choosing the proper indices to incorporate into the model.
  - The Mariposa POE was the only one of the Nogales POEs that processed truck crossings, thus a traffic spilt between POEs only occurred with POV, buses and pedestrian traffic.
  - The traffic split between the two POEs was stable throughout the years. For pedestrians, the DeConcini POE consistently accounted for nearly 95% of the pedestrian traffic. POV traffic had a ratio of roughly 60:40 (DeConcini: Mariposa) before 2007, and then 70:30 (DeConcini: Mariposa) from late 2007 onwards. Bus traffic had a ratio of roughly 25:75 (Deconcini: Mariposa) over the years analyzed.
- Testing of various model alternatives on the historical data for the different modes of traffic to find the best methods for creating our forecasts.
  - We built different types of models on the historical data, including different types of regression models and time series models. The performance of the models was compared, and the best performing models were chosen to produce the forecasts.
  - For the POV traffic, we built the model based on the number of vehicles, since the POVs were processed by vehicle.
  - Models for bus traffic were built on the number of bus passengers, since bus vehicle capacity might not be fully utilized.
  - Generally, the time series models were better for short term forecasts.

    - We found that the exchange rate between the Mexican Peso and US dollar was the most influential economic variable for truck traffic.
    - We tested the external variables on other traffic types, but none of them was statistically significant. However, we found that including Arizona employment data improved the quality of the models for pedestrian traffic.
    - No model was built for rail traffic because only the Union Pacific operates through Nogales, and company-specific decisions seemed to drive the history.
- Using the chosen models to provide forecasts of border crossings for the next 5, 10 and 15 years into the future
    - Time series models were used to produce all the short term (5 year) forecasts for all the traffic modes.
    - Regression models were used to produce the long term (10-year and 15-year) forecasts for POV, pedestrian and bus.
    - Time series models were used to produce the long term (10-year and 15-year) forecasts for trucks.
    - Long term forecasts for the economics indices were not available, so we defined plausible scenarios and used these scenarios in our models for crossing traffic.
    - According to our forecasts we found that the number of Commercial Vehicle (mainly truck) crossings might increase up to 50% in 15 years when compared to the number of crossings recorded in 2008.
    - The POV traffic and pedestrians were more sensitive to the changes in the economic climate and therefore their forecasts are less reliable than those obtained for commercial vehicles.
    - Our forecasts suggest that, in the near future, POV, pedestrian and bus traffic will decrease slightly. We do not believe they will continually decrease; however, we could not be sure when the declines will reverse.
    - These near term trends are probably driven by the economic downturn that began in late 2007.
- Creating a simulation model to test the capacity of the Nogales POE given our forecasted future traffic demands. Some of the results produced through this simulation include the following:
    - If our forecasts are correct, the maximum queue length based on our capacity estimates will be approximately 2300 trucks (over one day's backlog). The bottleneck location is the super-booth area for most of our scenarios.
    - Given existing infrastructure and time constraints (i.e. 11 hour workday), the current Mariposa POE does not have capacity to service our predicted maximum levels of traffic.

2 Introduction

This report documents the findings and the activities performed under ADOT grant JPA 08-024T. The overall purpose of this study was to forecast the number of border crossings by mode of traffic at the Nogales-Mariposa and DeConcini Ports of Entry (POEs), and to assess the interaction between the Mariposa and DeConcini Ports of Entry. Significant population growth and economic development in the Ambos Nogales area requires new comprehensive planning to address growing demands placed on the two land POEs. In addition, this growth and development calls for an examination of port of entry needs and opportunities.

In order to meet the expected increase in traffic at the international POEs in Nogales, the federal government and the State of Arizona plan to expand the capacity of the Nogales POEs in the near future. Sizing this new capacity requires forecasting the demand for each of the POEs as a foundation for developing appropriate expansion plans.

This report covers the activities completed from the start date of the project: 06/01/2008 to the end date of the project: 12/31/2009. The major product of this study will be a final report which contains projections of the number of border crossings by mode of transportation over five, ten and fifteen year periods and a description of the interaction between the Mariposa and DeConcini Ports of Entry, which is primarily in the area of passenger vehicles and, to a lesser extent, pedestrians and buses.

The general steps to be completed as part of this project include the following:

1. Identify, assess and classify previous studies dealing with traffic forecasts of the targeted Ports of entry
2. Analyze and document current conditions of the POEs
3. Develop preliminary assessment of forecast models and refine scope of work
4. Present preliminary findings and proposed model to ADOT
5. Develop the accepted forecast models
6. Collect data and validate forecast models
7. Determine infrastructure capacity
8. Interim report preparation
9. Final report preparation

In the rest of this report we provide a brief summary of the activities performed to accomplish these tasks.

3 Refinement of Tasks

The proposed initial tasks to conduct *Nogales POEs Traffic Study: Forecast and Capacity Planning for Nogales' Ports of Entry* were presented to the Technical Advisory Committee (TAC) in the inter-plenary kickoff meeting in August 2008.

After the objectives were approved by the TAC we developed a set of detailed activities required to complete the scope of work for the project. These activities are as listed above.

These tasks were then updated during a meeting with ADOT on June 20, 2009 when it was also decided that in addition to the three modes of traffic that had already been analyzed as a part of this study (Commercial Vehicle, Pedestrian and POV), bus and train traffic should be added to the scope of this study.

The remainder of this report is organized according to the following activities:

- Documentation of previous studies related to the scope of this project
- Analysis of the historical data of border crossings for various modes of traffic
- Analysis of the current state and traffic split of both Nogales POEs
- Summary of data collected during a visit to the Mariposa POE
- Testing of different model types on the historical data for each mode of traffic, with a focus on regression based and time series based models
- Application of the models to truck, POV, pedestrian and bus traffic
- Forecasts of five, ten and fifteen year time spans for commercial vehicles, POV, pedestrians, and bus passengers
- Description of Mariposa POE simulation model and results
- Conclusions drawn from this study
- Suggestions for future research topics

4 Documentation of Previous Studies

The main purpose of this activity was to identify the previous studies having a direct or indirect relation with either of the Nogales Ports of Entry so that redundant work would be avoided. In order to document and analyze the previous studies dealing with the Nogales ports we did the following:

- Identify relevant previous projects
- Complete a literature review on similar studies related to border crossing traffic.
- Read through each study and develop a matrix which includes the documents researched and any relevant contributions they may have to the current project
- Develop a brief summary of the findings from the past projects
- Identify those areas that have not been covered by previous projects
- Incorporate, if feasible, the identified knowledge gaps into the current project

As an initial activity of the *Nogales POEs Traffic Study: Forecast and Capacity Planning for Nogales' Ports of Entry*, previous studies were identified, gathered, and summarized. The studies were identified through a literature search using citation indices, internet tools and from citations from the studies themselves. The following is a list of the studies reviewed:

1. Currency Movements and International Border Crossings (2000)
2. Unified Nogales/Santa Cruz County Transportation 2000 plan (2000)
3. Estimating Texas-Mexico North American Free Trade Agreement Truck Volumes (2001)
4. Specification of a Borderplex Econometric Forecasting Model (2001)
5. Cross Border Cargo Vehicle Flows (2002)
6. Assessment of Automated Data Collection Technologies for Calculation of Commercial Motor Vehicle Border Crossing Travel Time Delay (2002)
7. El Paso Customs District Cross-Border Trade Flows (2003)
8. Borderplex Bridge and Air Econometric Forecast Accuracy (2004)
9. Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study: Strategic & Geographic Area Overview (2004)
10. Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study: Existing and Future Travel Demand (2004)
11. Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility Study : Travel Demand Analysis Process (2004)

12. Canada-United States-Ontario-Michigan Border Transportation Partnership Planning/Need and Feasibility : Study Partnership of Transportation Problems and Opportunities Report (2004)
13. Traffic Forecast Based on Real Data (2004)
14. An Error Correction Analysis of US-Mexico Trade Flows (2005)
15. Analyzing highway flow patterns using cluster analysis (2005)
16. Tradeoffs between security and Inspection Capacity: Policy Options for Land Border Ports of Entry (2006)
17. Socioeconomic determinants of Mexican Circular and Permanent Migration (2006)
18. AZ Multimodal Freight TM1:Analysis of Freight Dependent Industries (2007)
19. AZ Multimodal Freight TM2:Assessment of Arizona's Existing Freight Infrastructure (2007)
20. AZ Multimodal Freight TM3:Strategic Directions for Freight Planning (2007)
21. Use of Box and Jenkins Time Series Technique in Traffic Volume Forecasting (2007)
22. Nogales Railroad Small Area Transportation Study (2007)
23. Bottleneck Study of Mariposa POE (2008)
24. Mariposa/I-19 Connector Route Study (2008)

One of the tasks included in this project was to compile a summary of previous studies. The project team elected to summarize previous findings using two instruments:

- An Excel Matrix
- A written summary of the previous studies

The two instruments are described next:

An Excel matrix was prepared with the various studies in the leftmost column and the other relevant information such as the year of the study, a brief summary, main methods used and author(s) of the study described in the columns to the right. The Excel Matrix consists of five separate sheets – Factors Considered, Procedure, Scope, Detail and Data Source.

A summary of each study was prepared that describes the main elements of the document and indicates the findings that seem to be relevant to this project. It is suggested that the reader first look at the matrix to see which studies may contain relevant data, and then go to either the study summaries or the studies themselves to find the information they are seeking. The written summary is included as a literature summary appendix to this report, and the Excel matrix is also included in that appendix. Complete bibliographical data can be found in the literature appendix. Conclusions drawn from Previous Studies:

After reviewing previous studies related to border crossing traffic we discovered that for the majority of the studies concerning the southern US border, Texas or California are the areas of focus. To the best of our knowledge, there is no study dedicated to forecasting the border crossing traffic for the POEs in Arizona.

We also reviewed a number of studies regarding AZ highways and POEs. From this review we found one thing to note, as pointed out by a previous, the Mariposa POE Bottleneck Study; Border Wizard, the tool used by GSA for planning, actually does not forecast the volume of traffic, but rather takes the forecasted volume as an input (Study 23 in Documentation of Previous Studies).

From the review of methodology, we found there was no widely accepted systematic way of building models for border crossing traffic, because the infrastructure, types of traffic and other conditions varies widely between ports. The mainstream methods used in the literature are regression based models and time series analysis based models. However, previous studies did not have to account for the significant seasonal variation associated with Arizona POEs. Our methodology is thus somewhat different from that used in other studies of border flows.

5 Baseline Analysis of Current Conditions

A preliminary phase of this study was to assess the existing conditions of each of the ports of entry in the Nogales area. This baseline analysis consisted of several processes which are listed below:

- Finding and analyzing the historical data of the border crossings for each mode of traffic. Then identifying the characteristics of this data to propose potentially suitable methods for further analysis.
- Determining the current traffic split among each of the ports of entries (Mariposa and DeConcini).
- Visiting the Mariposa port of entry to measure the time it currently takes a commercial vehicle to cross the border. These data were later used to create a simulation model to test the capacity of the Mariposa POE.
- Reviewing and gathering conclusions from previous studies related to either border crossing traffic or to Arizona highways and ports of entry.

5.1 Introduction

The two international points of entry (POEs) connecting the cities of Nogales, Arizona with Nogales, Sonora in Mexico are vital for the economy of these two cities as well as the surrounding region. These two POEs, the Mariposa POE and the DeConcini POE, (see M and D in Figure 5-1) are also extremely important for trade between the United States and Mexico. For instance, one of the main economic drivers of Santa Cruz County and Nogales is the fresh produce industry which relies heavily on these POEs as they are the principal import points for winter fresh vegetables from Mexico to the United States.

Additionally, Nogales, Sonora is one of the Mexican border cities with a high level of industrial (maquiladoras) development. Consequently, the increased presence of American (and foreign in general) companies on the Mexican side of the border generates the need for daily transportation of materials across international boundaries. The shipping of goods proves to be a challenging task for the Logistics and Traffic departments of these businesses because the greater the congestion at the POEs in Nogales, the less competitive these companies become and alternatives such as moving to locations with more efficient POEs may then be considered.

Figure 5-1 Mariposa and DeConcini POEs at Nogales

In our analysis of current conditions, we assessed the existing conditions of the Commercial Traffic, POVs (Privately Owned Vehicles) and pedestrian traffic crossing the POEs in Nogales. First, we provide an overview of the historical data and then proceed to a more specific assessment of the commercial traffic which showed a cyclic pattern that, to the best of our knowledge, had not been addressed by any previous study. Next, we analyzed the traffic split between the Mariposa and DeConcini POEs. Last, we provided a brief description of the conclusions we drew from our research of relevant previous studies.

5.2 Historical Data

There are three principal modes of traffic which we explored: commercial traffic (trucks), POV, and pedestrians. In the original scope of the project, we had not planned on taking into account rail freight or bus traffic since they account for a very small percentage of traffic crossings. However it was later determined that they should also be considered. The data for these two modes of traffic is also presented in this section.

The historical monthly data (from January, 1994 - current) used was gathered from the Bureau of Transportation Statistics website (BTS). The daily truck crossing data of 2008 and the data regarding the traffic split between different POEs was obtained from US Customers and Border Protection Tucson Field Office (Donahue 2009).

The historical monthly crossing data for the commercial traffic, POVs and pedestrians are depicted in Figure 5-2. Note the vertical line marks the date of the event "9/11". We believe that the event "9/11" brought significant changes to border crossing traffic. Note that we plotted the number of Privately Owned Vehicles (POVs) crossing the border but not the number of persons crossing the border by POV because the POV crossings are processed vehicle by vehicle. However, the change in the number of POVs should be highly correlated with the number of persons crossing the border by POV.

As Figure 5-2 indicates, the truck data has very strong cyclic properties and subsequent statistical analysis quantified this behavior. As noted above, both POV and pedestrian traffic showed significant changes right after "9/11", while truck and bus crossings appear to be relatively unchanged. The correlation between these different modes of traffic is displayed in Table 5-1. Three approaches were used to calculate correlations: 1) using the entire range of data, 2) including only the data before "9/11" (until 2009/08); 3) including only the data after "9/11". From both the graph and our correlation data we can see that after "9/11" the changes in POV traffic and pedestrian traffic are negatively correlated with each other.

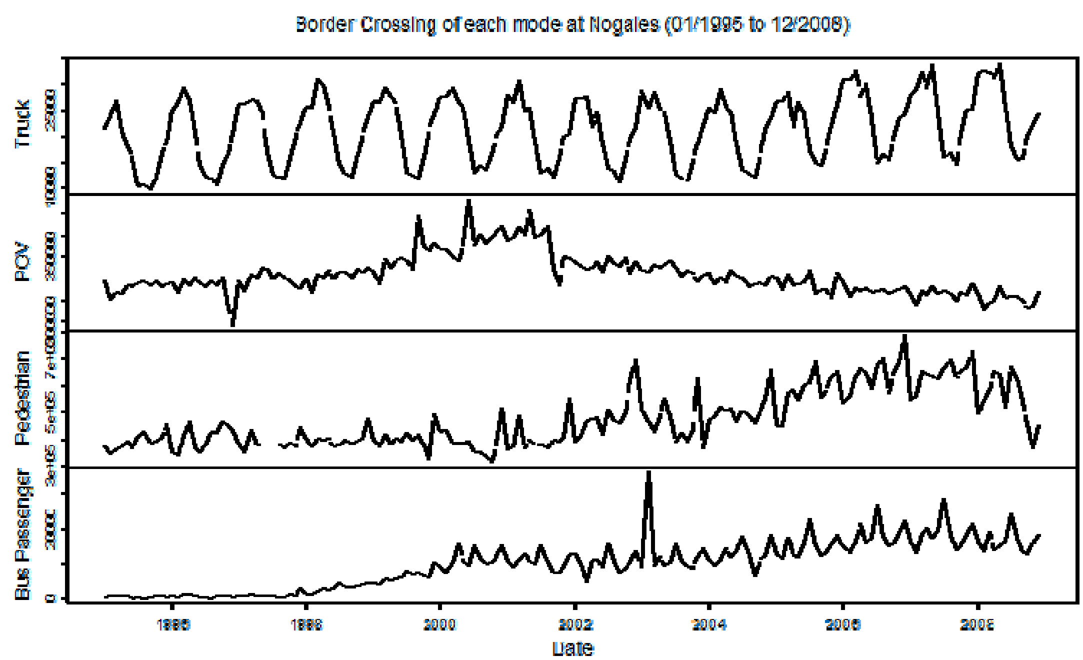

Figure 5-2 Crossings by Mode

Table 5-1 Correlation between different traffic modes

		Truck	POV	Pedestrian	Bus Passenger
All data	**Truck**	1.000	-0.102	0.216	0.202
	POV	-0.102	1.000	-0.395	-0.057
	Pedestrian	0.216	-0.395	1.000	0.707
	Bus Passenger	0.202	-0.057	0.707	1.000
Before 9/11	**Truck**	1.000	0.045	0.113	0.116
	POV	0.045	1.000	0.023	0.845
	Pedestrian	0.113	0.023	1.000	0.022
	Bus Passenger	0.116	0.845	0.022	1.000
After 9/11	**Truck**	1.000	-0.208	0.193	0.156
	POV	-0.208	1.000	-0.422	-0.367
	Pedestrian	0.193	-0.422	1.000	0.548
	Bus Passenger	0.156	-0.367	0.548	1.000

From the first four rows of Table 5-1 we can also tell that out of any two modes of traffic the strongest correlation was between Pedestrian and Bus traffic followed by the correlation between POV and Pedestrian traffic. By separating the data into "before and after 9/11", we observed that the POV and pedestrian traffic had little correlation "before 9/11"; however, they showed a strong negative correlation "after 9/11". Also, we observed that the POV and bus traffic had strong positive correlation beforehand, but they showed a strong negative correlation after "9/11". The pedestrian and bus traffic are positively correlated. However, by separating the data, we can tell that this correlation mainly happened after "9/11". Thus, It appears that the preference for personal border crossing shifted from vehicle to foot and bus after 9/11.

Among the four modes of traffic, the pedestrian traffic contained the most variation, and the commercial vehicle traffic was the most stable. Note that for 2008, the pedestrian data exhibited a significant drop while the other two modes remained relatively stable. This could be interpreted as the pedestrian traffic being more sensitive to changes in the economic climate, considering the current recession.

Table 5-2 below lists the yearly number of crossings for each type of traffic. One interesting fact is that the number of POV crossings has been decreasing since 2001, and that 2007 and 2008 were both lower than POV crossings in 1995. In contrast, truck and pedestrian crossings have trended upward since 2001, with the exception of a decrease in pedestrian crossings in 2008.

Table 5-2 Yearly number of crossing of each mode

	Truck	POV	Pedestrian	Bus Passenger
1995	206,032	3,368,337	4,698,049	7,608
1996	229,337	3,316,799	4,864,717	8,637
1997	242,830	3,587,985	4,643,538	11,477
1998	258,828	3,698,273	4,796,884	34,470
1999	256,426	4,186,962	4,806,076	75,976
2000	254,694	4,681,567	4,677,819	136,471
2001	249,237	4,590,933	4,874,738	126,530
2002	242,237	3,978,640	5,911,866	125,264
2003	243,365	3,836,372	5,583,533	156,406
2004	247,553	3,571,230	6,131,407	150,073
2005	266,233	3,445,984	6,930,198	178,306
2006	289,590	3,282,781	7,726,045	217,093
2007	295,267	3,180,548	7,722,877	221,410
2008	303,757	3,026,767	6,568,207	195,741

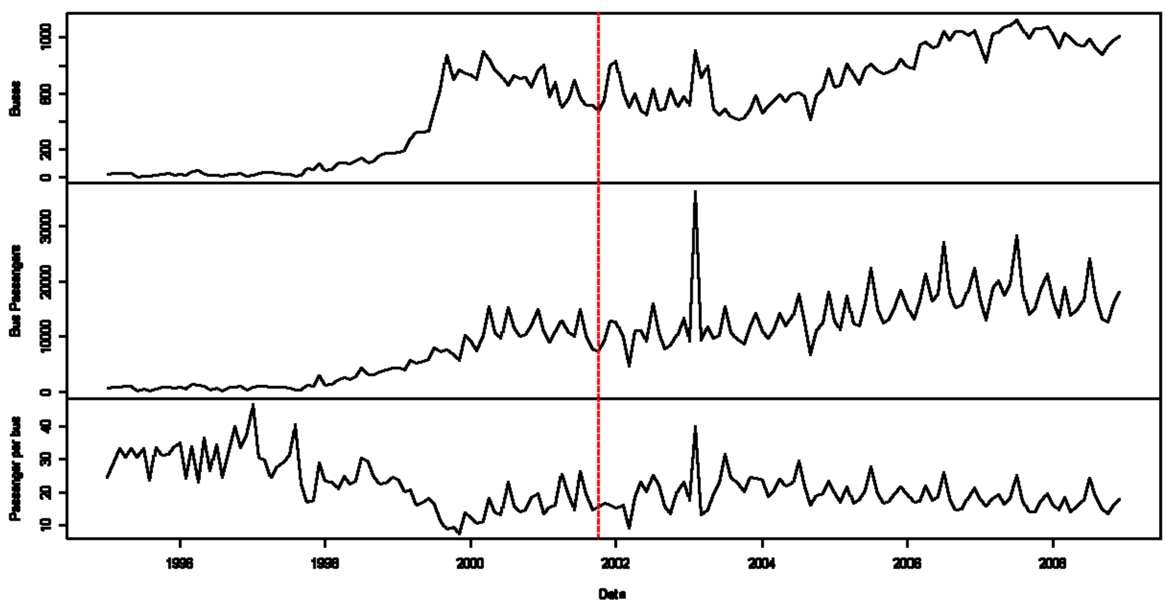

Figure 5-3 Historical data of bus crossings and bus passengers

Figure 5-3 shows the historical data of the bus crossings and the number of passengers crossing by bus. The number of crossings by bus started to increase in the middle of

1997, and had a sharp jump in 1999. After that it was relatively stable with a slight decreasing trend until 2005. In 2005, there was another significant increase which lasted until 2007, when the number of bus crossings once again stabilized. Similarly, the number of passengers crossing by bus started to increase at the end of 1997 and stabilized during the year 2000. After that, the number of bus passengers remained relatively stable with a slight increasing trend. The only exception occurred during 2003, when an abnormally steep spike occurred. The bottom panel of Figure 5-3 is the average number of passengers per bus, which shows that the average number of passengers per bus started to decrease in 2005, although the downward trend is slight.

The number of bus passengers is much smaller than the number of passengers crossing by other modes. We found that although the number of bus passengers has increased very quickly during the last few years, it still only comprises a small fraction of the total number of passenger crossings. In 2008 for example, the average number of monthly pedestrian crossings was 547,351, the average monthly POV vehicles was 706,023, but the monthly bus passengers was 16,312, which was roughly 2.9% of the pedestrian crossings and 2.3% of the POV vehicles.

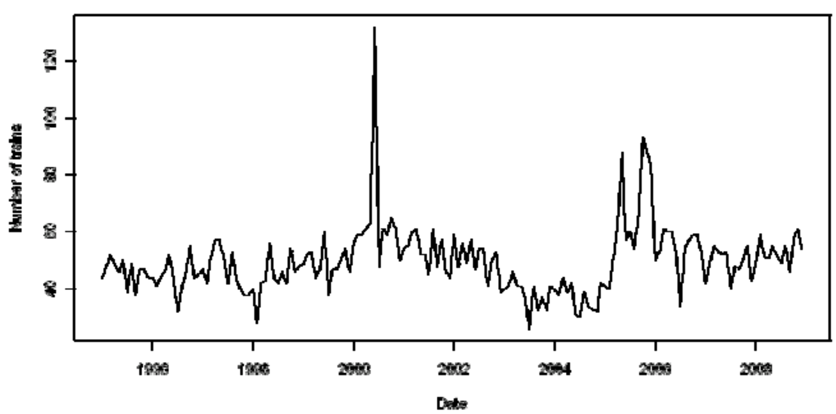

Figure 5-4 The number of trains crossing the border

Figure 5-4 shows the number of trains crossing the border from January 1995 to December 2008. We did not have reported train crossings for February 1995 and April 1995. We used the average of the preceding and the following month to represent these missing values. Before 2000, the number of trains was relatively stable with a slight increasing trend. In the middle of 2000, there was a large spike and after this occurrence the number of trains followed a decreasing trend which continued until early 2005. 2005 saw another sudden increase, and since then train crossings have been relatively stable. Note that train crossings are partly dependent on the number of schedules Union Pacific chooses to run, and that the actual amount of freight crossing the border depends on the length and consists of the trains Union Pacific chooses to run.

5.3 Traffic split between the Nogales POEs

Commercial vehicles cross only at the Mariposa POV; therefore we did not have any data for the traffic split of the trucks. However, POV, pedestrians and bus crossings occurred at both of the POEs. We had a limited amount of data, starting from October 2004, for the traffic recorded by mode and by POE. Figure 5-5, Figure 5-6 and Figure 5-7 depict the split of the POV traffic, pedestrian traffic and bus traffic (number of buses) respectively. From Figure 5-5 we observe that the POV traffic has a ratio of roughly 60:40 (DeConcini: Mariposa) from 2004 to 2007, and then 70:30 (DeConcini: Mariposa) from late 2007 onward. Figure 5-6 shows that the majority of pedestrian traffic passes through the DeConcini POE, and this split has been relatively stable throughout the years. Figure 5-7 shows that the bus traffic has a ratio of roughly 25:75 (DEconcini: Mariposa) all over the years, except from April 2007 to September 2007. For the recent months (including the whole year of 2008), the ratio tended to be quite stable. We believe there are several causes for this stable trend in pedestrian traffic:

- The Mariposa POE is not adapted for handling pedestrian traffic.
- The DeConcini POE is closer to most of the Nogales population and business compared to Mariposa
- The DeConcini POE has more booths for pedestrian traffic than the Mariposa POE.
- The route via Mariposa POE is the preferred route to the bus traffics, which is opposite to that of the POV and pedestrian traffics.

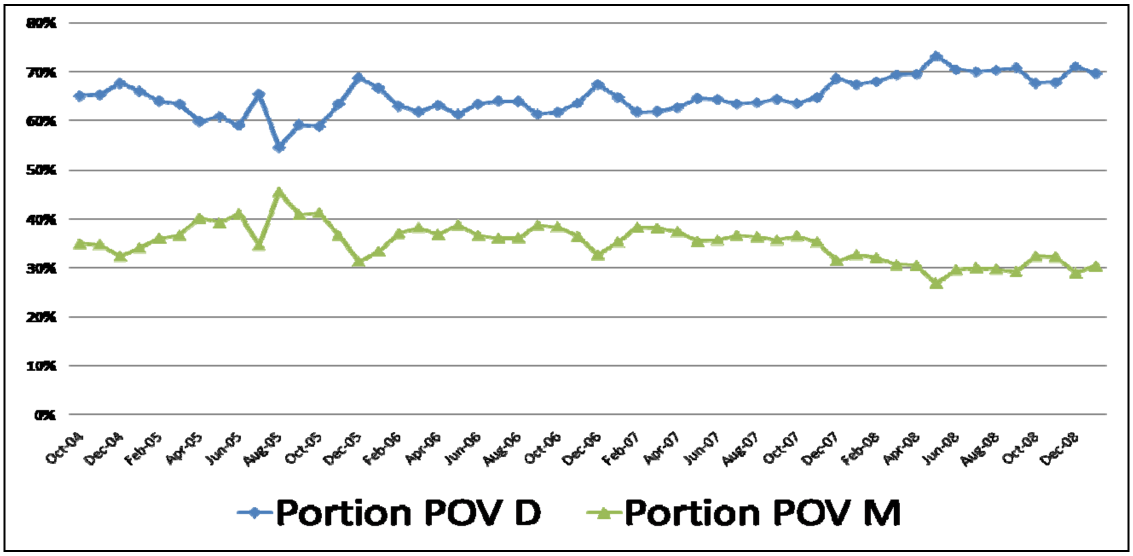

Figure 5-5 POV traffic split

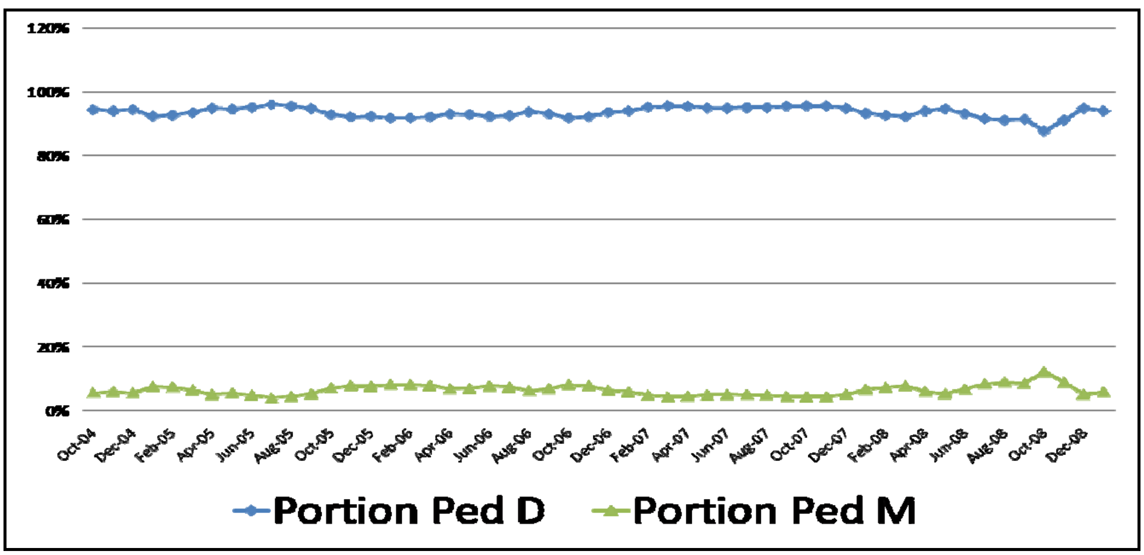

Figure 5-6 Pedestrian traffic split

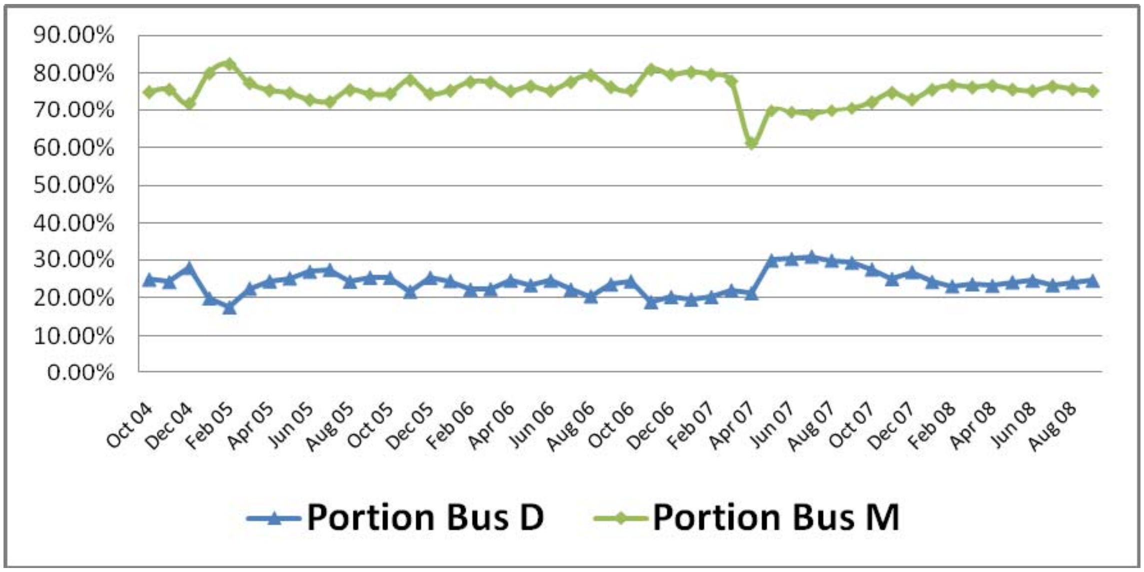

Figure 5-7 Bus traffic split

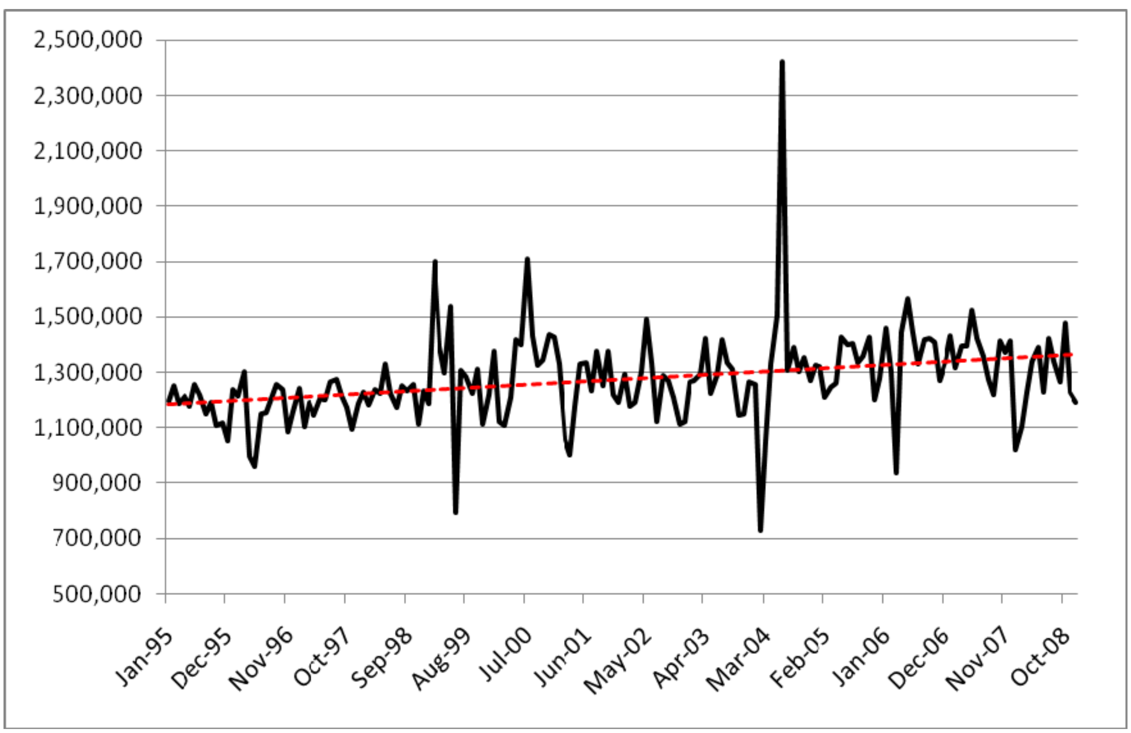

Figure 5-8 Total number of persons crossing (POV+Pedestrian+Bus Passengers) at the POEs in Nogales

Figure 5-8 depicts the total number of persons crossing (POV passenger, bus passengers and pedestrian) crossing at both POEs in Nogales by month, where the red straight line is a fitted trend line The change in total number of persons crossings the border from 1995 to 2008 was relatively small however the fluctuations from month to month were at times very significant. The greatest change occurred between the months of July and September 2004 with a decrease of about 1.7 million crossings, which was preceded by a very large increase. We do not have any concrete explanation for these fluctuations however we hypothesize that it may have to do with changes in the measurement process.

5.4 Mariposa POE Site visit

Our visit to the Mariposa POE was conducted on Tuesday May 25, 2009. The main purpose of this visit was to measure the time for a commercial vehicle (mainly referring to trucks) to cross the border. To gather our measurements we had 4 observation points, which were Weigh-in-Motion (WIM), SBs (Super Booths=primary inspection), ADOT inspection and the exit to the highway. The four observation points are marked in Figure 5-9.

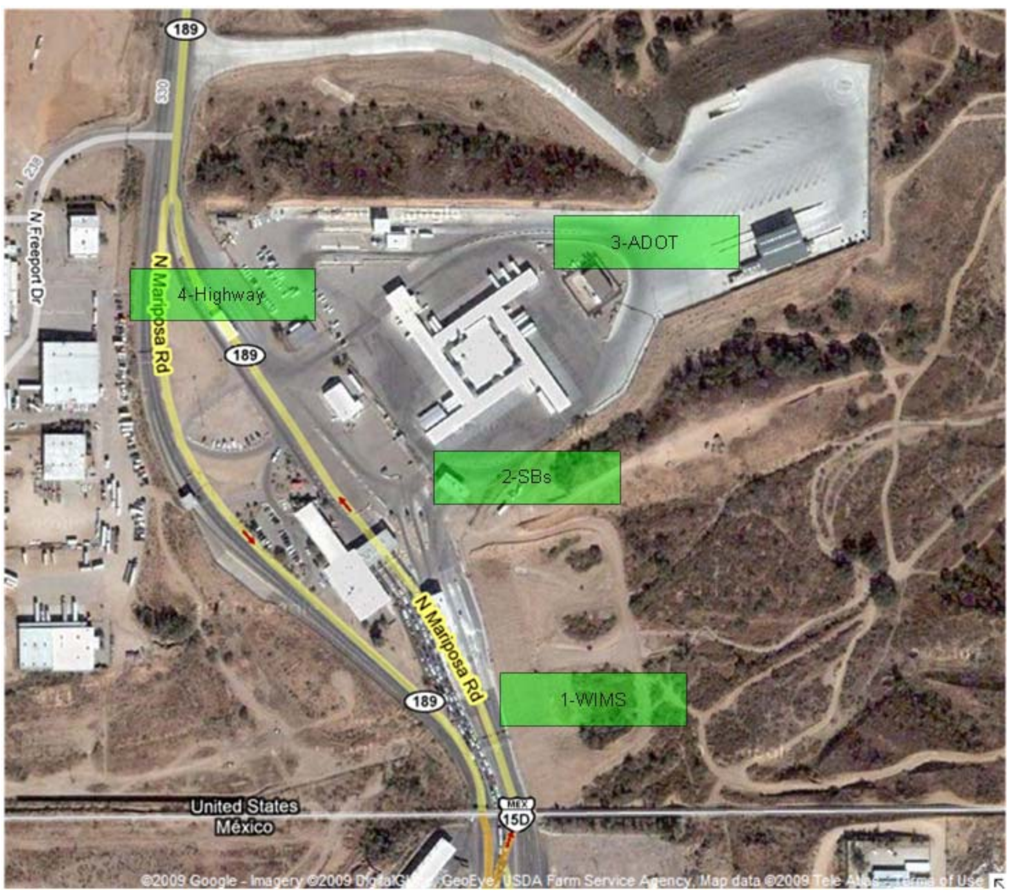

Figure 5-9 Measurement Points for Mariposa Crossing Times

We recorded the plate number of the vehicles passing every observation point and the time of passing. We also wrote a brief description of the vehicles in case we misread the plate or recorded different license plate numbers since it was very common for the border crossing vehicles to have multiple license plate numbers. We started our observations at 10:30 am and finished at 4:30 pm. These observations were only taken for commercial vehicles, since we were not granted access to observe other types of crossings. Also due to clearance issues, we were only able to gain a general idea of the amount of time spent at each location: WIM, SB, and ADOT (i.e. the time we recorded is a combination of the waiting time and processing time at these locations).

We observed approximately 600 trucks during the six hour time period. We summarize the time of passing of each observation point in Table 5-3. The histogram of number of trucks by hourly interval is provided in Figure 5-10. The bar marked as "14:17:27" is shorter, since the border was closed for half an hour during that time slot.

Table 5-3 Current result summary

	WIM	ADOT	CBP
Average	0:04:10	0:57:21	0:27:07
Standard Deviation	0:04:12	0:37:37	0:48:16

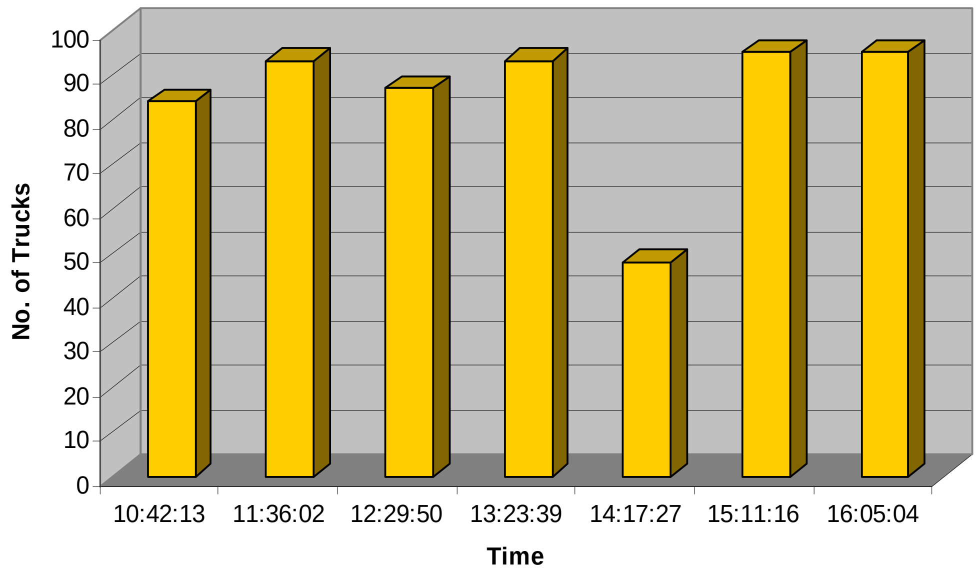

Figure 5-10 Number of Trucks Crossing by Time Period

After suitable processing we used this data to build a simulation model to assess the current capacity of the Mariposa POE.

5.5 Summary of Baseline Analysis

Our literature review also provided some useful insights for our model building. Economic indices like the Index of Industrial Production and the exchange rate were used in many previous studies. This motivated us to incorporate some of these indices in our model. Some preliminary analysis was also conducted to gain a more thorough understanding of the traffic characteristics. Through this process we identified the different modes of traffic to study and examined the historical data for each mode of traffic as well as the traffic split among the ports of entry. We noticed that the cyclic pattern shown in the truck crossing data was not addressed in other related work, although it has long been "taken for granted" in the Nogales import community. Valuable information was also obtained for our later capacity assessment simulation work through our visit to the Mariposa POE.

6 Model Alternatives

In this section, we test different types of models on the historical data to find the best alternative for forecasting. Generally, the models can be categorized into two types, regression based and time series based models. We begin with a brief introduction of each type of model, and then we use the commercial vehicle model as an example to describe the way we selected the models. Following this section, we present the models we built and the resulting forecasts. As with the baseline analysis section, an appendix to this section provides a more detailed review of the related technical issues.

6.1 Regression models

Univariate regression model

The univariate linear regression model, which is the simplest type of regression model, only takes time as a regressor (regression variable). Its basic equation is shown in equation (6.1.1) (Montgomery, Peck, and Vining 2006a). In this equation y is the target traffic, t is the time and ϵ is the irregular fluctuation around the trend, which is usually assumed to follow a normal distribution.

$$y = \beta_0 + \beta_1 t + \epsilon \tag{6.1.1}$$

The βs are the coefficients we need to estimate. This is the first type of model we applied, however, we will explain later in this report why this was not the best choice for our forecasts.

Multivariate regression model

The second model we tested was the multivariate model because from the available literature, we found that border crossing traffic may be influenced by several exogenous variables, such as the GDP of the countries that share a common border. Unlike the univariate regression model, this type of model takes exogenous variables into consideration. The model has the form shown in equation (6.1.2) (Montgomery, Peck, and Vining 2006b), where y represents the target traffic and $x_i, i = 1, 2, \cdots, k$ are the exogenous variables. In our study, each economic index will be an exogenous variable.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{6.1.2}$$

Based on previous research results and the conditions of the Nogales POEs, we identified a list of candidate exogenous variables as shown in Table 6-1. However, since there were only 14 years of available data, a limited number of variables could be used in the regression model. Thus, a variable selection procedure was used to identify the "best" variables to include in the model.

Table 6-1 List of candidate external variables

Data Name	Time range	Frequency
US national GDP	from 1949 to 2008 Q4	quarterly
Mexican national GDP	from 1993 to 2008 Q4	quarterly
Exchange rate (1USD in MNX)	Since Jan 1994	daily, monthly
Arizona GDP	1997 -2007	yearly
US fuel price (Gasoline and Diesel)	1994 Jan to 2008 Dec	monthly
Arizona Population	1990 to 2008	yearly
Sonora Population	1995 to 2008	yearly
US Index of Industrial Production(IIP)	Since 1919	monthly
MX Index of Industrial Production(IIP)	Since 1990	monthly
US Consumer Price index (CPI)	Since 1990	monthly
MX Consumer Price index (CPI)	Since 1990	monthly
Real exchange rate	Calculated from exchange rate and CPIs Since Jan 1995	monthly

Two tier regression model for the truck traffic

We noticed from our baseline analysis that the truck traffic has a stable cyclic pattern. The existence of this cyclic pattern prevents us from using the multiple regression models directly, however, since this pattern is stable, we can build a two tier regression model. In the two tier model, we first built a regression model on the yearly data, and then split it into months according to monthly percentages. In contrast, for POV and pedestrian traffic, we built the model directly on the original monthly data as they had no obvious seasonality. Furthermore, it should be noted that the regression models used are all linear models.

Figure 6-1 is the box plot of the truck crossings of each month. This plot reveals some useful information about the truck data:

1. The box plot is a confirmation of cyclic pattern as we observed from Figure 5-2.
2. For all our data the percentage of total crossings for the year for each month stays relatively stable. For example if the number of crossings in January 1995 was 10% of the total number of crossings for that year, the percentage of total crossings for 2008 occurring in January 2008 will also be roughly 10%.

The month of May is the month with the most outliers in the number of crossings, while April is the month with most variation.

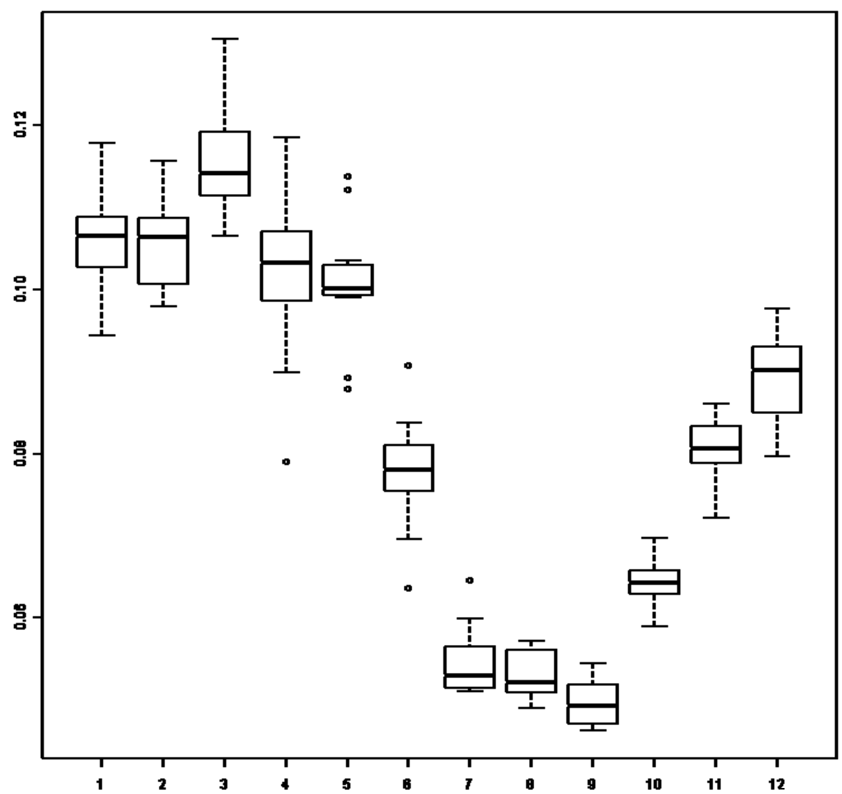

Figure 6-1 Box plot of truck crossings by month of the year

Mathematically, we explain the two tier model as described subsequently. Suppose we have N years of monthly data points available. Let y_{ij} be the data for month i in year j. Let $T = \sum_{j=1}^{N}\sum_{i=1}^{12} y_{ij}$ be the total number of crossings in the data set. Then the portion corresponding to month i can be calculated as $p_i = \left(\sum_{j=1}^{N} y_{ij}\right)/T$. Therefore, when the number of crossings for year j is calculated, namely y^j, the estimated number of crossings of month i in year j can be calculated as $\hat{y}_{ij} = y^j \times p_i$. Note that all the p_i's are calculated from the data in the training data set, the data set we used to build the model When applying the method to new data, we still use the p_i's calculated from the training data set values on which we built the model.

Variable selection

Given the small size of the variable pool and the limited number of data points we used an exhaustive method for variable selection. Using this method we enumerated all the possible combinations of up to 5 variables, and then built the corresponding regression models. The resulting models were then evaluated using several criteria:

- R-square[1] : The R square value can be interpreted as the proportion of variation explained by the model. When using linear models, the R-square value will be between 0 and 1, however, when using other types of models, this cannot be guaranteed.
- VIF (Variance Inflation Factor)[2] : measures the multicollinearity between models' variables. Multicollinearity exists when at least one variable can be represented by the linear combination of other variables. In other words, this measures whether several of our independent variables are highly correlated and thus can be replaced by only one or two variables.
- Expert knowledge about the relationship among the variables selected and the target variable to be estimated.

For our variable selection, we applied the above criteria to both the training data set and the validation data set. When there was a tie, we chose the model with fewer variables.

6.2 Time series model

Another type of model commonly used in previous studies was the time series model. Particularly, we considered the ARIMA (Autoregressive-integrated-moving average) model[3] (Farnum and Staton 1989; Shumway and Stoffer 2006a). In order to build a credible time series model, we needed to further explore the characteristics of the

[1] Refer to R square section of appendix of statistical details for further explanation
[2] Refer to VIF section of appendix of statistical details for further explanation
[3] Refer to the ARIMA model section of appendix of statistical detail for further explanation

data. For example, the first question we needed to address was whether to use a regular model or a seasonal model.

The ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function)[4] act as tools for determining the appropriate type of time series model as well as the structure of the model. Figure 6-2 depicts the ACF and PACF of the truck data. These functions allow us to determine seasonal and other patterns of the data. Note the unit of the lag is year, so 0.5 means 6 months. The ACF at lag 0.5 has a negative value near -1 while the value at lag 1.0 is near 1, which confirms the need to use a seasonal ARIMA model to forecast border crossings.

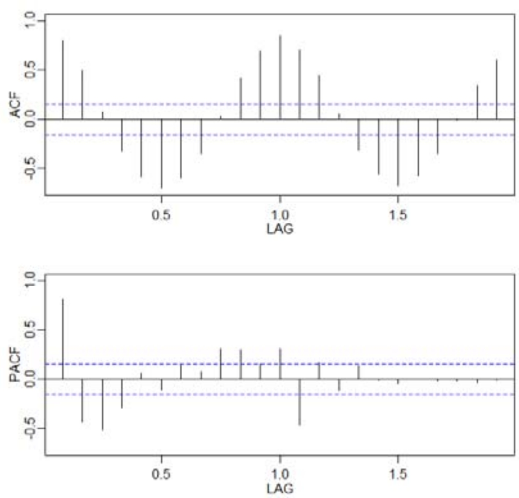

Figure 6-2 ACF and PACF plot of the Truck data

Univariate time series model

We mainly considered the ARIMA model and Holt-Winter's model for the univariate time series models. We have mentioned the ARIMA model in the previous paragraph, which is a type of time series model. The Holt-Winter's model is a more specific time series model, which is capable of handling both trend and seasonality in the data simultaneously. Due to the strong presence of seasonality in the truck traffic, we first tried the additive Holt-Winter's model on the truck traffic data. For the POV and the pedestrian flows, we used the non seasonal ARIMA model. Note that the Holt-Winter's model can be converted to a corresponding ARIMA model. The details of these two models are explained in the appendix of statistical details.

[4] Refer to ACF and PACF section of appendix of statistical detail for further explanation

The Holt-Winter's model decomposes the target data into three parts: *level*, which is the non seasonal mean of the data; *Trend*, which is the slope of the likely line through data points; and an index of *seasonality*. Mathematically, it can be written as:

$$y_t = a_t + b_t t + S(t) + \varepsilon_t \tag{6.2.3}$$

where a_t is the unseasoned level of time series at time t, b_t is the slope of the trend at time t, s_t is index of season $i, i = 1, 2, \cdots, L$. i corresponds to the season of current time t. The parameter estimating methods and the updating forecast methods are explained in the Appendix of Statistical Details.

An ARIMA model is usually written as $ARIMA(p, d, q)$, where p is the AR (Autocorrelation) order, d is the degree of differencing, and q is the MA (Moving Average) order. When applying the ARIMA model, it is important to first decide the structure of the model. PACF and ACF act as tools for determining the structure of an ARIMA model. Since it is possible to have potential models that work equally well, it is preferable to come up with a list of reasonable ARIMA models and then select from this candidate list. Therefore, instead of deciding the (p, d, q) directly from ACF and PACF, we defined ranges for (p, d, q), and tested all the possible combinations of the parameters within the established ranges. We used Theil's U statistic[5], which is a measure of the similarity between two time series, as a criterion for model selection. R square was not used because when a data set contains nonlinearities, a large R square does not necessarily imply a good model. We use the same method to find the structural parameters in our multivariate time series models.

Multivariate time series models were another type of model chosen to forecast border crossings. To build this kind of model, we introduced exogenous variables into the model rather than only taking the data itself into consideration. We referred to the previous studies we reviewed to decide what exogenous variables should be incorporated in the model. We also referred to the variables selected in the multivariate regression model, and field knowledge.

A seasonal ARIMA model has seven structural parameters to determine (Shumway and Stoffer 2006b), which are shown in Table 6-2. A model with those parameters is usually reported as $ARIMA(p, d, q)(P, D, Q)_L$.

We used the same method as we used in the univariate ARIMA model building to get a list of good model candidates, and then selected models from this candidate list.

[5] The definition of Theil's U statistic is stated in the Appendix of Statistical Details

Table 6-2 list of ARIMA structural parameters

p	AR (Autocorrelation) order
d	The degree of differencing
q	MA (Moving Average) order
P	Seasonal AR order
D	The degree of seasonal differencing
Q	Seasonal MA order
L	Seasonal period

6.3 Comparison of the models

Before applying the models to generate forecasts, we first tested the performance of the models on our data. We split the historical data into two subsets, a training set and a validation set. As described in the Historical Data section, we have data available from January 1995 to December 2008. We designated the last three years' data as the validation set, and used the rest as the training set.

We use the truck data to illustrate the procedure we used to compare the models:

1. Prepare the variables to use in the multivariate regression model and the multivariate time series model
2. Build model of each type of model on the training data set
3. Use the models to forecast the traffic of the last three years (validation data set)
4. Compare the forecasted values to the validation set values

We defined some criteria for model selection; however, we may not strictly select the model with the best criteria. There are many reasons for doing this:

- The criteria may not be able to fully reflect the performance of the model
- A "too good" performance on the training set may lead to over fitting and thus the model would perform badly on the forecast task
- Other issues that are not incorporated in the model, but need to be considered, for example the stability of the model.

Variable selection

We use the variable selection procedure described in 6.1. Table 6-3 shows the best R square values we obtained using different numbers of regressors (independent regression variables). These values were obtained by applying the resulting forecast models to the validation data. As we observed from the table, there was a significant increase in the R square value when using two regressors as opposed to just one.

However, when the number of regressors was greater than 2 there was not much benefit in terms of the gain in R square value. In addition, some of the variables were highly correlated, which could cause multicollinearity issues, resulting in a forecast model that was unstable. In order to minimize multicollinearity issues we used the Variance Inflation Factor (VIF) metric to choose variables that were not highly correlated.

Table 6-3 Best R square VS number of regressors

Number of Regressors	Best R Square on Validation data
1	0.6388321
2	0.8365505
3	0.8611884
4	0.8616332
5	0.8649311

Table 6-4 Some of the variable selection results

Model	Training R Square	Validation R Square	VIF	VIF	VIF
Truck~USIIP+Xrate	0.9675	0.6710	2.8646	2.8646	
Truck~MXIIP+Xrate	0.9671	0.6558	2.2812	2.2812	
Truck~AZemp+sonpop	0.9711	0.6524	8.2115	8.2115	
Truck~USIIP	0.9667	0.6388			
Truck~RXrate+USIIP	0.9668	0.6342	1.3426	1.3426	
Truck~MXIIP	0.9667	0.6331			
Truck~MXIIP+RXrate	0.9668	0.6279	1.2049	1.2049	
Truck~RXrate	0.9668	0.6201			
Truck~Xrate	0.9668	0.6043			
Truck~AZpop+MXIIP	0.9711	0.5786	3.0764	3.0764	
Truck~MXIIP+sonpop	0.9709	0.5681	2.2072	2.2072	
Truck~AZemp+sonpop+USDiesel	0.9714	0.5636	8.6778	9.0878	3.5406

Note: AZpop: Arizona Population; AZemp: Arizona Employment; Xrate: Exchange rate; RXrate: real Exchange Rate; sonpop: Sonora Population; IIP: Index of Industrial Production;MX: Mexico

Table 6-4 shows part of the variable selection process for the trucks. Column 1 is the model we used, the variable to the left side of "~" is the target mode of traffic, and the variables to the right side of"~" are the variables in the model. Column 2 is the R square value on the training data, and Column 3 is the R square value on the validation data. All the columns after Column 3 are VIF values. The items were sorted according to the Validation R Square values in descending order.

Those models having VIF values greater than 10 were excluded from this table, as this indicates multicollinearity issues (Montgomery, Peck, and Vining 2006c). If there was only one regressor no VIF value was provided, for two regressors the VIF for these two regressors will be identical and for greater than three regressors each regressor has its own VIF. This table shows that of the models analyzed, a regression model using the Index of Industrial Production for the US (US IIP) and the exchange rate between US Dollar and to Mexican Peso would render the best results for forecasting the truck traffic border crossings.

We first decided the structural parameters of the multivariate ARIMA model. Table 6-5 lists the results of some of the ARIMA models tested, sorted by the Theil's U statistic (obtained by using the time series model on the training data set. A lower U indicates a better fit). In Table 6-5, we filtered out the models whose residual violates normality assumptions as these would potentially create misleading forecasts. One needs to be careful when choosing the parameters. All of the parameters listed in Table 6-5 were generally good candidates. When selecting among the list of parameters, the experience of the modeler, the plot of the fitted values as well as the residuals and reasonableness of the model all play important roles in the selection process. Here we chose the model with $(p,d,q)(P,D,Q)_L = (1,1,4)(2,1,2,)_{12}$ to compare with other type of models, which is NO. 4 in Table 6-5. The subscript 12 means that we use a seasonal ARIMA model with seasonal period of 12 months.

Table 6-5 List of ARIMA models for truck

ID	p	d	q	P	D	Q	Validation R square	Theil's U (Training)	Theil's U (Validation)
1	5	0	4	2	1	0	0.900769	0.028775	0.039771
2	5	0	4	2	1	1	0.891993	0.028457	0.040779
3	3	1	4	2	1	2	0.890986	0.029925	0.041243
4	1	1	4	2	1	2	0.887973	0.030065	0.04177
5	1	1	5	2	1	2	0.885679	0.029985	0.042002
6	1	1	5	2	1	1	0.884808	0.03024	0.0422
7	1	1	4	2	1	1	0.885581	0.030341	0.042243
8	1	1	4	1	1	1	0.88376	0.030368	0.042512
9	0	1	4	1	1	2	0.881525	0.030079	0.042681
10	6	0	3	2	0	2	0.884909	0.035388	0.042749

The comparison result

We were mostly concerned about the ability of the models to forecast future traffic crossings. Therefore, we used the models built on the training data set to forecast three years ahead and compared the forecasted values with the real data in the validation data set. Table 6-6 shows the comparison among the multivariate regression model, the Holt-Winter's method and the multivariate time series model. We could see that the multivariate time series model (ARIMA) outperforms the other two methods in terms of R square and Theil's U statistic.

Table 6-6 Compare of different models

Method	R square (The higher the better)	Theil's U statistic (The lower the better)
Multivariate Regression	0.765	0.06315865
Holt-Winter's	0.760	0.05936151
Multivariate time series	**0.889**	**0.04156882**

Figure 6-3 is a graph of the three model forecasts. From the graph, we can tell that all the models fit well to the real data at the beginning. However, the Regression model tended to underestimate and the Holt Winter's method tended to overestimate later. From this example, we preferred to use the multivariate ARIMA model in our forecast.

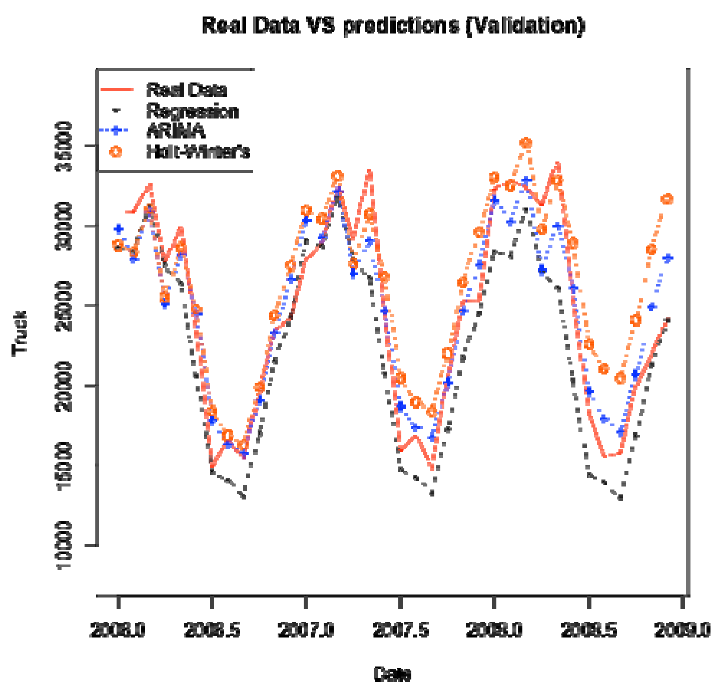

Figure 6-3 Forecasts vs. Actual

6.4 Model alternatives for other modes

In section 6.3 above, we used the truck data as an example to show the model alternatives, and the results show that the ARIMA model outperforms other models. Therefore, our first choice was to use the ARIMA model on the other modes of traffic. Since the Holt Winter's method could be converted into an ARIMA model, we only considered ARIMA models on the other data sets. Before we made this choice we applied the same variable selection procedure on the other modes of traffic. The results of this testing did not show any variables which improved the quality of the models, such as the US IIP and the exchange rate which we found for the truck data. Thus, we mainly relied on the ARIMA model with no exogenous variables to forecast the other traffic modes.

In particular we note that the bus and train modes of traffic run on a relatively stable schedule, which did not typically change in response to economic variations as do the other modes of traffic.

POV

We first generated the ACF and PACF of the POV data as in Figure 6-4. The ACF tails off and the PACF cuts off after 3. Although we did not see any spike after 1, we chose to use a seasonal ARIMA model. We believed the POV traffic also had some patterns that repeated from year to year. Table 6-7 lists some ARIMA models we tested on the POV historical data. We left out the last three years' data for validation as we did for the truck data. The last three columns are the R square value on the validation set, the Theil's U statistic on the Training set and the Theil's U statistic on the Validation set respectively. The table was ordered according to the Theil's U statistic on the Validation set in ascending order. As with the truck data, choice of model was not solely based on the order of the parameters listed in this table. Some other factors were also considered, such as the validation plots of the models.

We picked the model with structural parameters $(p,d,q)(P,D,Q)_L = (6,2,6)(2,0,1)_{12}$, and plotted the forecasted result against the real data in Figure 6-5. The fitted data seemed to overestimate the traffic. However, we also noticed that the real data had several fluctuations and the model was only able to capture the main trend excluding these fluctuations. If we tune the parameters to follow this fluctuating pattern, we may end up over fitting the model, thus generating an extremely implausible forecast. Instead, it may be appropriate to estimate a fixed correction, depending on discussions with subject experts.

Table 6-7 List of ARIMA models for POV

p	d	q	P	D	Q	Validation R Square	Theil's U(T)	Theil's U(V)
5	0	3	2	0	3	0.2712246	0.03882831	0.02571694
2	1	6	2	0	3	0.2568555	0.03870172	0.02588999
6	2	6	2	0	1	0.2476303	0.03940341	0.02618603
6	2	6	2	0	0	0.2508674	0.03869041	0.02628177
5	2	3	2	0	3	0.2257206	0.03882943	0.0263639
6	0	3	2	0	1	0.2199377	0.03977416	0.02655877
5	2	3	1	0	3	0.09908995	0.03884228	0.02833369
6	2	6	0	0	1	0.1044766	0.0393614	0.02846729
4	2	5	0	0	3	0.08501245	0.0391253	0.02860754
3	2	6	0	0	3	0.05758358	0.03913235	0.02901835
6	2	5	1	0	0	0.05689941	0.0395814	0.02913005

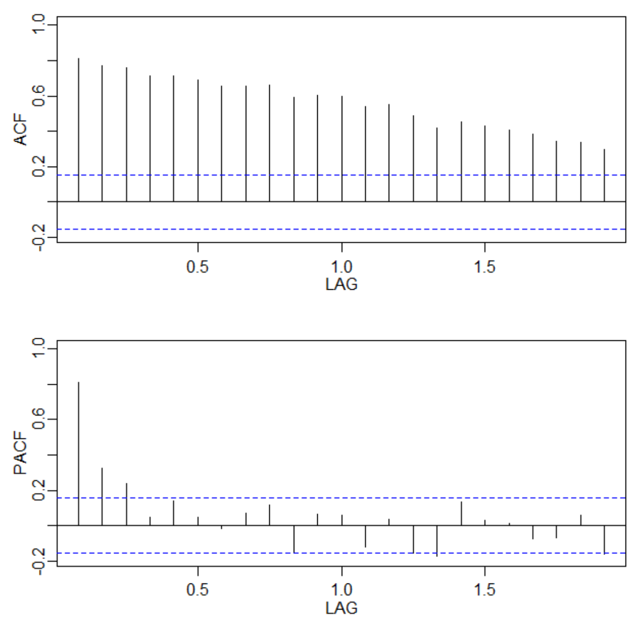

Figure 6-4 ACF and PACF of the POV data

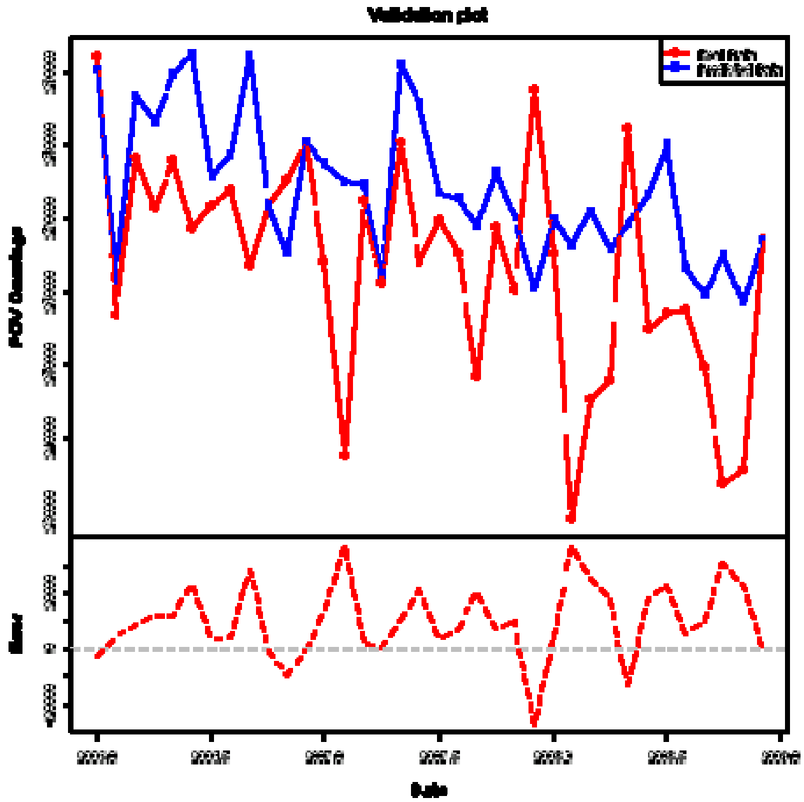

Figure 6-5 Plot of the fitted data to the real data on validation set (POV)

Pedestrian

As we stated at the beginning of the section, we could not find an exogenous variable that would allow us to build a reasonable regression model for the pedestrian data. However, since we believe the majority of the people crossing the border by foot are locals, we thought that employment in Arizona might influence this crossing. Therefore, we incorporated Arizona employment into our time series model. We show the ACF and PACF of the data as in Figure 6-6. The ACF tails off, while the PACF dies off after 4 steps.

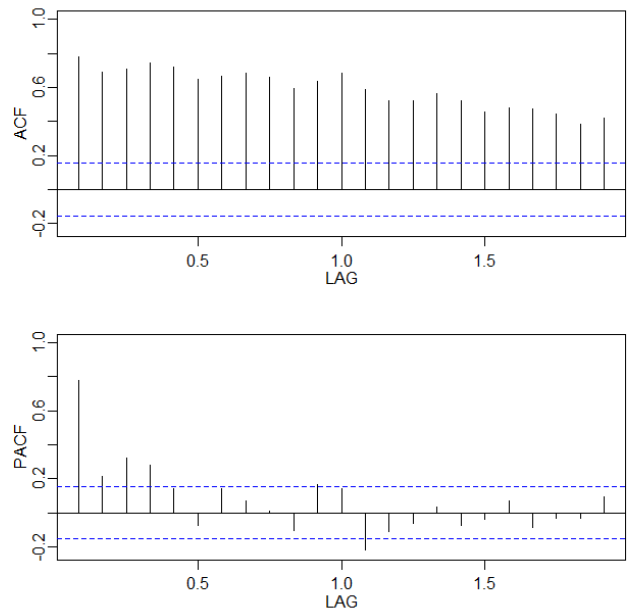

Figure 6-6 ACF and PACF of the Pedestrian data

Similarly, we have a list of relatively good models, which are listed in Table 6-8. We chose the model with structural parameters $(p,d,q)(P,D,Q)_L = (3,2,1)(2,0,2)_{12}$ to see how it performed on the forecast. Figure 6-7 plots the fitted data against the real data in the validation set. We can see before the middle of 2008, the fitted values follow the real data relatively well. However, a big drop occurred in late 2008, which was not captured by the model.

Table 6-8 List of ARIMA models for Pedestrian

p	d	q	P	D	Q	Validation R Square	Theil's U(T)	Theil's U(V)
3	2	1	1	0	1	0.34989	0.057766	0.054331
3	2	1	2	0	2	0.348045	0.057426	0.054253
3	2	1	1	0	0	0.347597	0.0586	0.054569
3	2	1	0	0	1	0.337429	0.058797	0.055034
3	2	1	1	0	2	0.32706	0.057466	0.054939
3	2	1	2	0	1	0.312486	0.057518	0.055427
6	2	6	1	0	0	0.311573	0.053171	0.055578
2	2	5	1	0	0	0.297323	0.056882	0.056197
6	2	5	1	0	0	0.293468	0.055045	0.056611
3	2	4	2	0	2	0.291586	0.055999	0.056388

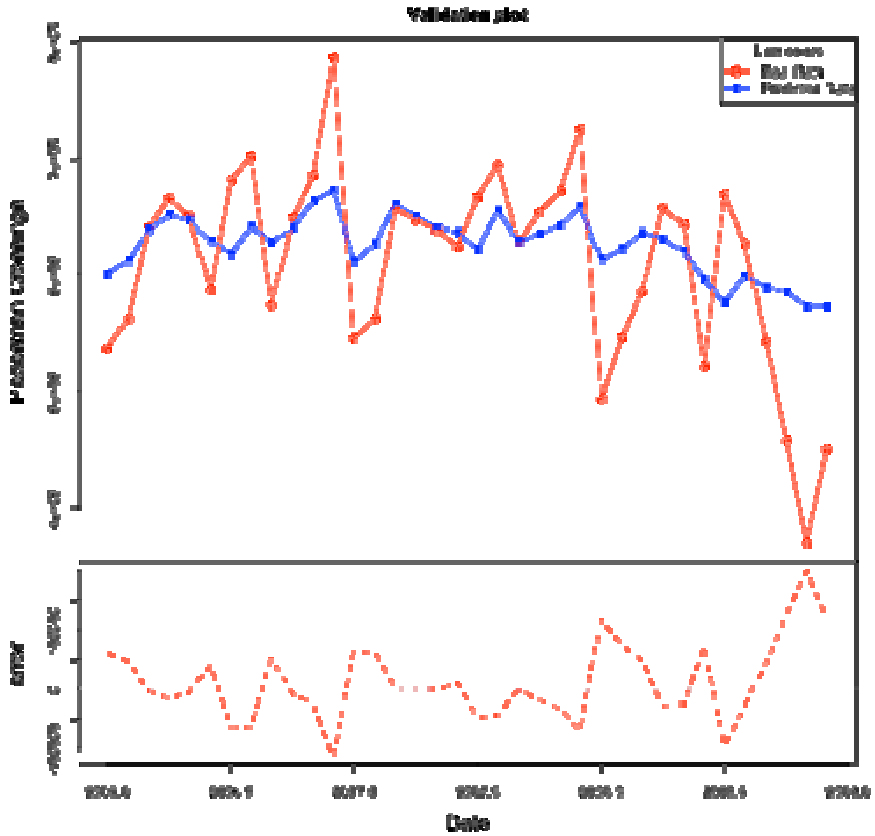

Figure 6-7 Plot of the fitted data to the real data on validation set (Pedestrian)

Bus

We depicted the historical data of the bus crossings in section 5.2 Historical Data. We show the graph of historical bus traffic and bus passengers crossing the border as Figure 6-8 here for review. Note that the bus traffic started to increase by the end of 1997 and then began increasing faster in 1998. The amount of bus traffic jumped up significantly in the middle of 1999. According to a fact sheet from USDOT (U.S. DOT 2002), "the NAFTA timetable also called for the United States and Mexico to lift all restrictions on regular route, scheduled cross-border bus service by January 1, 1997." We believe this jump was associated with the implementation of the NAFTA. Therefore, we decided to use the data after NAFTA had been implemented, and the impact of this implementation had stabilized. For convenience, we used crossings since January 2000.

Comparing the bus traffic and the bus passenger data, we found that there was a slight difference between the patterns of these two data sets. For example, the bus passenger data did not show any decrease in its general trend between 2000 and 2008, while the bus traffic started to decrease after 2000, and then began increasing in 2005. We decided to build the model based on bus passenger data rather than the number of buses crossing the border. First, there are many companies involved in the bus operations. There are always new companies joining in and other companies leaving this business. This makes the number of bus crossings more difficult to predict. Secondly, bus capacities may not be fully utilized. If this is the case, predicting the number of buses will not reflect the number of passengers crossing the border.

Figure 6-9 depicts the ACF and PACF function of the bus passenger data. Note that there was a stem at lag 1, which is 1 year. This spike indicated that there was some autocorrelation with an interval of 12 months. However, when examining the bus passenger data, we could not find a stable seasonality effect such as we found in the truck data. Thus the two tier regression model we used for the truck data was not viable here. Instead we decided to use the time series model. However, similar to the POV and pedestrian data, we did not think the time series model was capable of giving a good extended forecast; therefore, a regression model based on the yearly bus passengers was also built to produce the extended forecast.

We tested different models to find a relatively good time series model for the bus data. We used the ARIMA model with $(p,d,q)(P,D,Q)_L = (9,0,7)(1,0,0)_{12}$. In this case, the training data was from January 2000 to December 2005, and the validation data set was from January 2006 to December 2008. We found that the data for February 2003 was abnormally high, which prevented us from finding a good model, thus we used the average of January 2003 and March 2003 data to replace the original data point. Figure 6-10 shows the fitted value against the real value on the validation data. Due to the variety in the data, the model was unable to follow each fluctuation in the real data, but the general trend does not deviate. The Theil's U statistic is 0.091 on the validation set, which was high compared to those from other modes. However, this was a relatively good result among the models we tested.

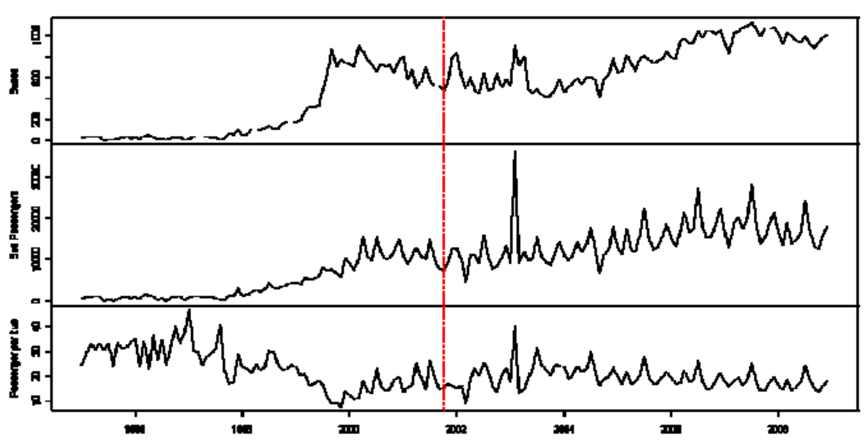

Figure 6-8 Historical data of bus crossings and bus passengers

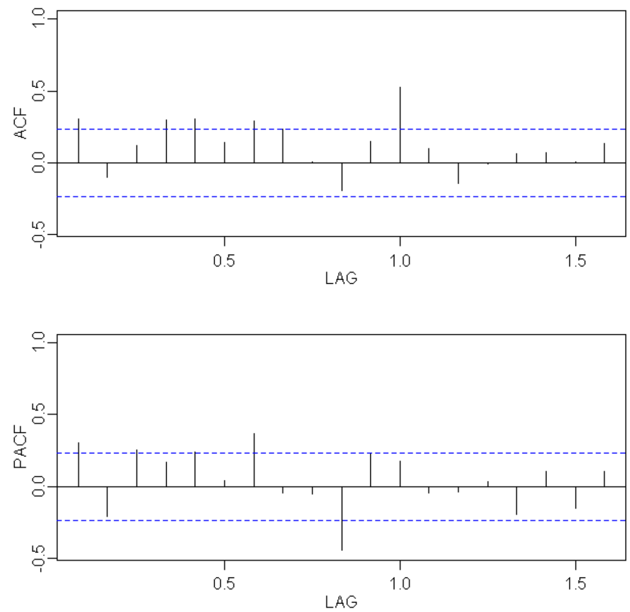

Figure 6-9 ACF and PACF of the bus passengers

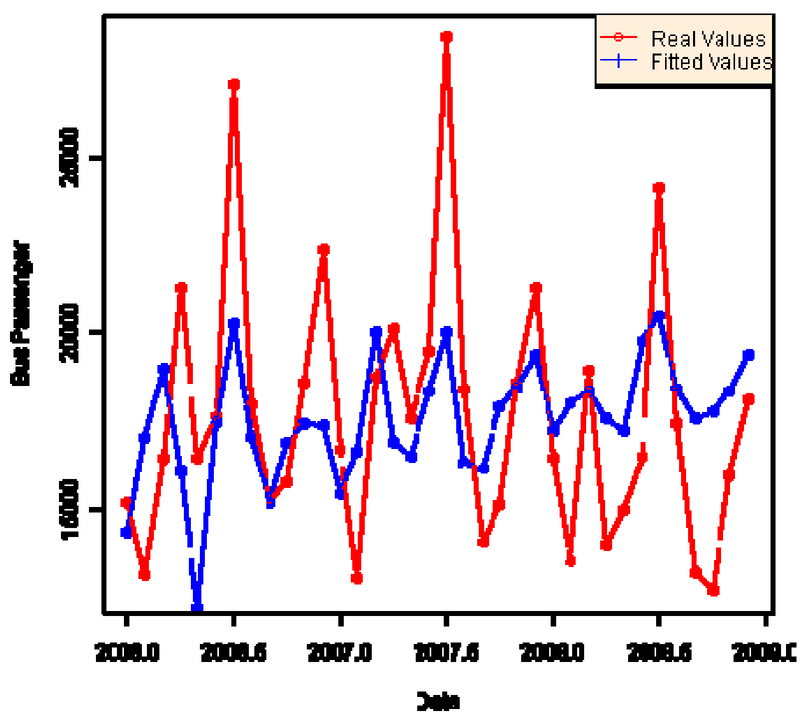

Figure 6-10 Plot of the fitted data to the real data on validation set (Bus Passenger)

Rail

Besides the relatively stable schedule of the trains, train traffic was also highly dependent on the availability of equipment and underlying customer demand. Recall the historical data of the train crossing we listed in section 5.2 Historical Data. We plot the graph here again in Figure 6-11. There were three huge spikes during the last 14 years. During the years 2003 and 2004, traffic was significantly lower than other years. Aside from these two instances, the railway traffic was relatively stable, though some fluctuations existed. Realistic projections of rail traffic will depend critically on Union Pacific's assessment of customer demand and other external factors such as the success of Punta Colonet, rerouting away from the center of Nogales, and expansion of the port of Guaymas.

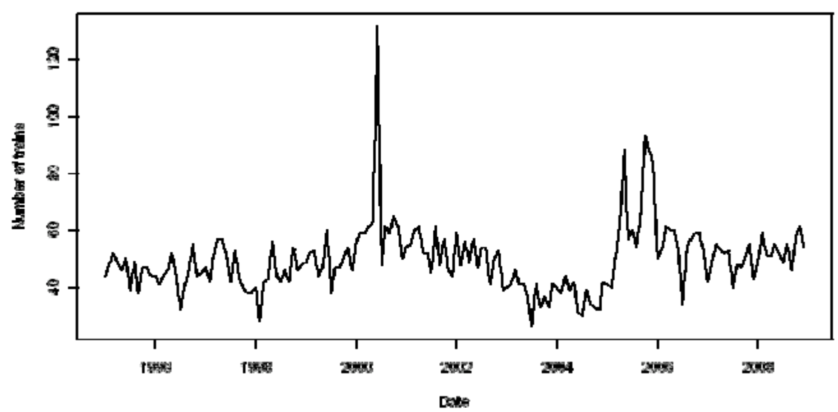

Figure 6-11 Historical data of the number of trains crossing the border

7 Models

After a preliminary evaluation of each of the possible model types the following framework was adopted to generate forecasts:

- Time series models were used to conduct all the short term (5 year) forecasts for all the traffic modes
- Regression models were used to conduct the long term (10-year and 15-year) forecasts for POV, pedestrian and bus.
- Time series models were used to conduct the long term (10-year and 15-year) forecasts for truck.

After our initial analysis of the train data, we decided not to build any forecast model for the train traffic for the following reasons: 1) The trains were running on a relatively stable schedule. 2) Only one company was involved in the railway transportation business in this area. Any changes in schedule were highly dependent on this particular company.

In this section, we show the models we actually used for the forecasts. Note the modes we mentioned in the previous section, Model Alternatives, were built on training data, which left out the last three years' data for validation purpose. The models shown in this section were built upon the full data set, and aimed to forecast the future traffic. For each model we built, we show the model and the diagnostic methods for the corresponding model. We believed the best way to explain the methods is with an example, so the methods will be explained when first used in the following part of this section. An overall summary of the techniques used are summarized in the appendix called "*Statistical Detail*".

7.1 Models for Truck Traffic

We used the ARIMA model with $(p,d,q)(P,D,Q)_L = (1,1,4)(2,1,2)_{12}$, and then we estimated the coefficient for each parameter based on the full dataset. The computer package used to do this estimation was R(R Development Core Team 2009), a freely available statistical software package. The computer outputs are shown as in Table 7-1[6] . Table 7-1 lists the coefficients of the parameters and their standard deviation

[6] Refer to the ARIMA model part in the statistical detail appendix for the detailed meaning of this result report

estimates below the corresponding coefficient (The lines start with "s.e."). Figure 7-1 shows the fitted values and the real values, which shows the fitted values are very close to the real values for historical data.

One important criterion used to check the adequacy of the model was to analyze the residuals, which were the differences between the fitted values and the real values in the training set. In our models, we assumed the residuals are normally distributed with mean 0. So we needed to check the validity of this assumption. In the time series model, we also wanted to make sure there was no trend existing in the residuals. A good way to check this was with the ACF and PACF function as we used before. However, instead of applying these functions on the original data, we applied them on the residuals. If there was no significant autocorrelation between the residuals, we believed there was no trend in the residuals.

Figure 7-2 shows a series of diagnostic plots for the model. The upper panel shows the standardized residuals. The middle left panel is the ACF of the residuals. From the plot we conclude that there was no significant autocorrelation among the residuals, and thus we claim that there is no trend contained in the residuals. The middle right panel is the Normal plot of the residuals. All the points are tightly clustered around the straight line in this plot, which is a good indication that the residuals are normally distributed. The lower panel is the plot of Ljung-Box test statistics. The X axis in the plot is the lag, while the y axis is the p-value, and the blue dash line is the limit. If the p-value is within limits at lag i, then the residuals have no autocorrelation at lags of i or less. From this graph, we can say residuals have no autocorrelation at lags up to 30. Thus we believe that the model and parameters satisfy the requirements for good model fit.

Table 7-1 Computer output: The coefficients for the parameters, Truck model

```
          ar1     ma1      ma2      ma3     ma4     sar1     sar2
sma1
      -0.9318  0.4383  -0.9395  -0.3056  0.1567  -0.6268  -0.0325
0.0643
s.e.   0.2324  0.2388   0.1265   0.1256  0.0837   0.2505   0.1700
0.2434
          sma2  constant     USIIP      Xrate
      -0.4692    3.7757  164.8073  -273.2759
s.e.   0.1989       NaN   75.8264   266.4386
```

Xrate stands for exchange rate

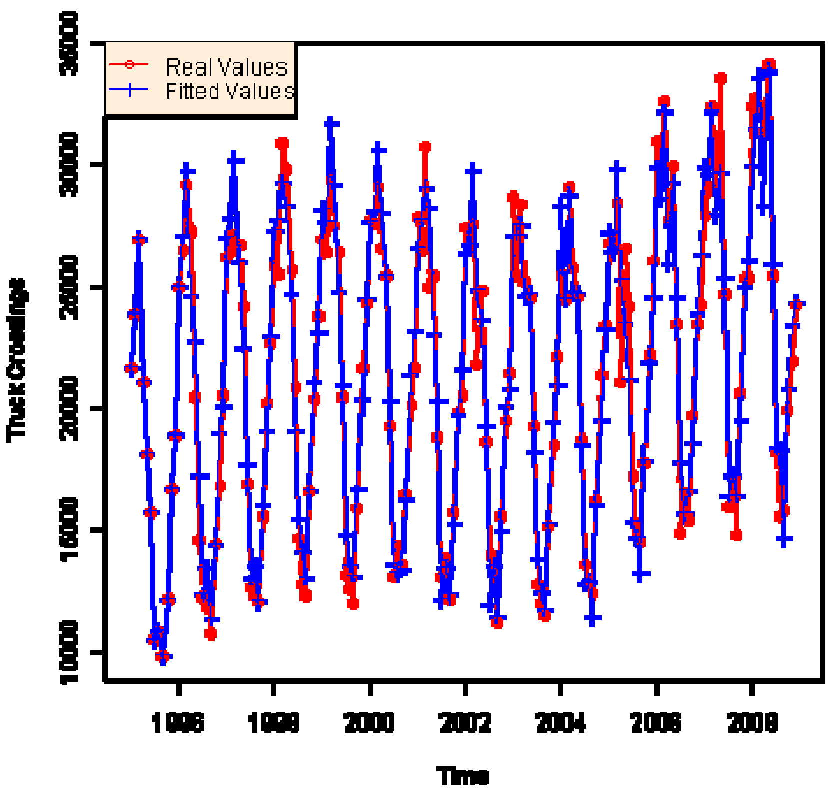

Figure 7-1 Fitted values vs Real values, Truck

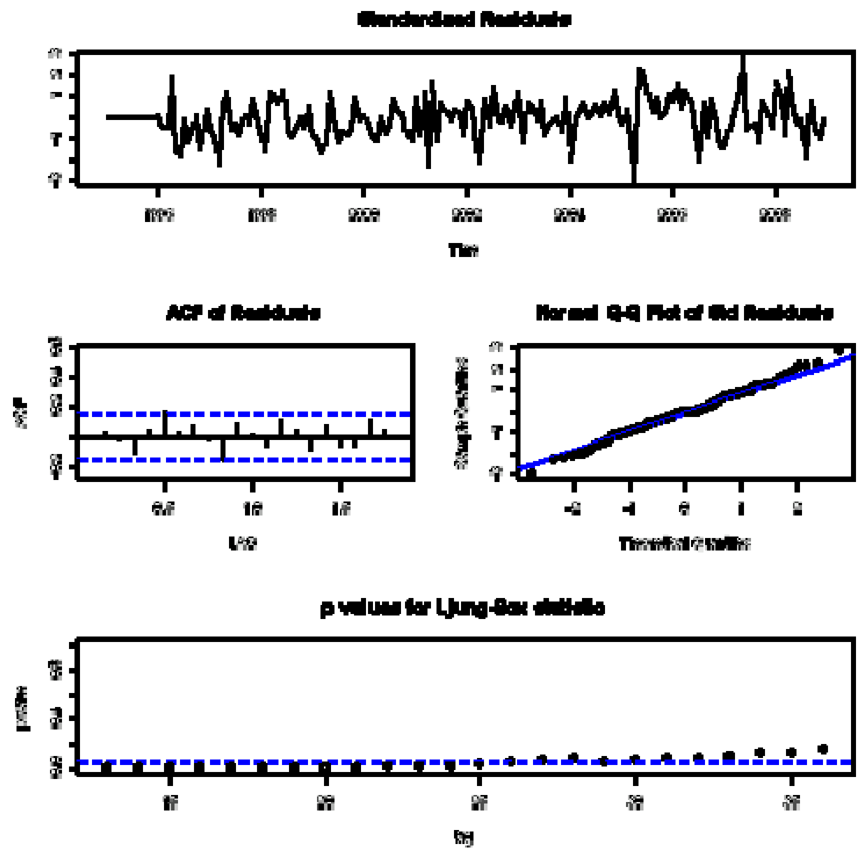

Figure 7-2 Diagnosis plots of the Truck model

7.2 Models for POV Traffic

We used both a time series model and a regression model for the POV traffic. The time series model was used to produce the short term forecast (5 year), and the regression models were used to produce the extended forecasts.

The time series model used was a seasonal ARIMA model with structural parameters $(p,d,q)(P,D,Q)_L = (2,1,6)(2,0,3)_{12}$. Table 7-2 shows the computer output of the ARIMA model, which contains the coefficients of the parameters. Figure 7-3 plots the real values and fitted values together. We can see the fitted values follow the trend of the real values generally well except for those points whose value increase or decrease suddenly. Figure 7-4 shows the diagnostic plots of the POV time series model. The upper panel shows the standardized residuals. We can see that there are some places the residuals are abnormally high or low. Comparing the position of their corresponding time stamp to the time stamps of the sudden increase and decrease in historical POV data as shown Figure 7-3, we can see that the time stamps match each other. We intend not to incorporate these sudden increases or decreases in our model. Otherwise, the model might be over fitted. From the ACF plot of the residual and Ljung-Box test statistics, we believe that the residuals are normally distributed with mean 0, except for a few exceptional points.

Table 7-2 Computer output: The coefficients for the parameters, POV model

```
        ar1      ar2     ma1     ma2      ma3      ma4     ma5     ma6
    -1.1039  -0.9917  0.5475  0.2996  -0.6448  -0.0728  0.0675  0.0576
s.e.  0.0161   0.0144  0.0843  0.0993   0.0990   0.0895  0.0893  0.0821
       sar1    sar2     sma1     sma2    sma3       XmatT
     0.8037  0.0094  -0.7022  -0.0747  0.1956    153.8429
s.e. 0.3838  0.3860   0.3801   0.3554  0.1005  11102.8878
```

XmatT contains the time index

As we can tell from the Figure 7-5, the real values cannot be fitted by a single linear model. Thus, we decided to use piecewise linear regression, and used different pieces as different scenarios in our later forecasts. One thing we needed to decide was how many break points to have and where to place these break points. We used the method introduced in (Achim Zeileis et al. 2002) and (Achim Zeileis et al. 2003) to finish these two tasks simultaneously. The breakpoints we located were 56 and 80, which were corresponding to August, 1999 and August 2001. Figure 7-5 shows the breakdown of the historical data. An interesting finding here is that the POV traffic actually started to drop in August 2001, which just one month prior to

"9/11". After "9/11", the decrease was magnified, and it continued to late 2008.

For each segment, we labeled the data from 1 to *n* according to the sequence of their original time stamp, where *n* was the length of data in that segment. We fitted a simple linear regression model for each segment. The fitted line and the real values are plotted in Figure 7-5. The computer outputs of these models are shown in Table 7-3, which contains the coefficients of each segment's model and corresponding tests. Segment 1, 2 and 3 are corresponding to the segments labeled 1, 2 and 3 in Figure 7-5. The slopes of the three segments are 1004.6, 1777.5 and -1057.45 respectively. The model for segment 2 is not as good in terms of the p-value (the smaller the better) as in segments 1 and 3. Segment 2 contained some big spikes, and our model was a linear model, which was not capable of capturing all these spikes. However, according to the Figure 7-5, this fit was still acceptable since the general increasing trend was captured. The whole set of historical data may have different trends in different time frames, thus we use the piecewise regression to find the possible segments and the slope of each segment. Each segment provides a possible trend within a time span. When we have difficulty in conducting a forecast, we may refer to these segments to estimate the future trend. In our study, we used each segment as a scenario.

Table 7-3 Computer output: models of different segments, POV data

```
Segment 1:
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 267733.4     5874.3  45.577  < 2e-16 ***
ind           1004.6      179.3   5.603 7.32e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21690 on 54 degrees of freedom
Multiple R-squared: 0.3677,   Adjusted R-squared: 0.3559
F-statistic:  31.4 on 1 and 54 DF,  p-value: 7.316e-07

Segment 2:
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 374845.5    13180.1  28.440   <2e-16 ***
ind           1777.5      922.4   1.927    0.067 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31280 on 22 degrees of freedom
Multiple R-squared: 0.1444,   Adjusted R-squared: 0.1055
F-statistic: 3.713 on 1 and 22 DF,  p-value: 0.06699

Segment 3:
Coefficients:
```

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept) 338274.70    3211.91  105.32   <2e-16 ***
ind          -1057.45      62.68  -16.87   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14940 on 86 degrees of freedom
Multiple R-squared: 0.7679,    Adjusted R-squared: 0.7652
F-statistic: 284.6 on 1 and 86 DF,  p-value: < 2.2e-16
```

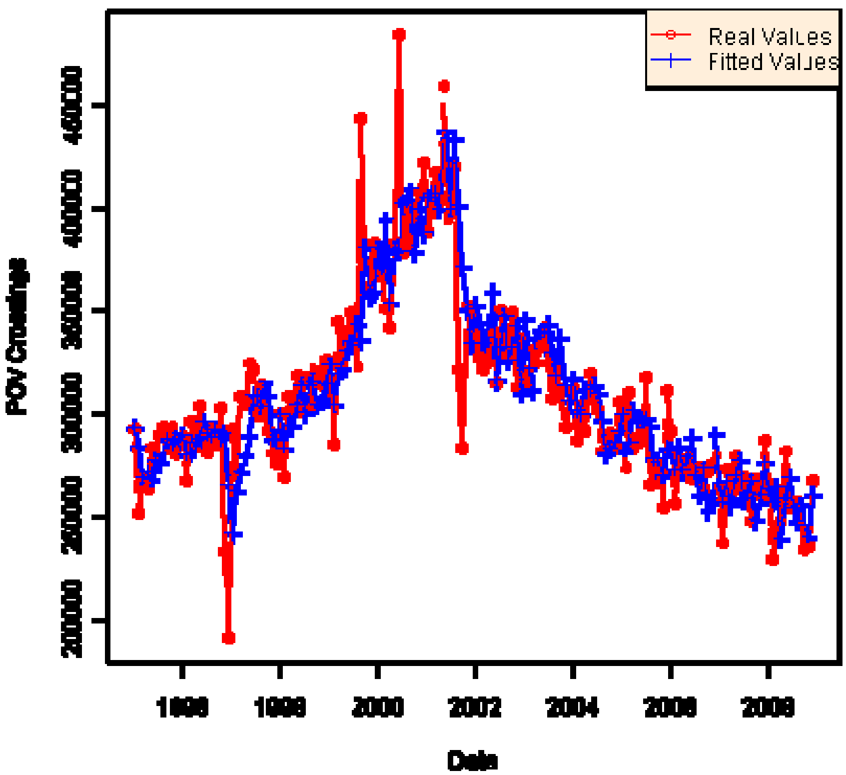

Figure 7-3 Fitted values vs Real values, POV

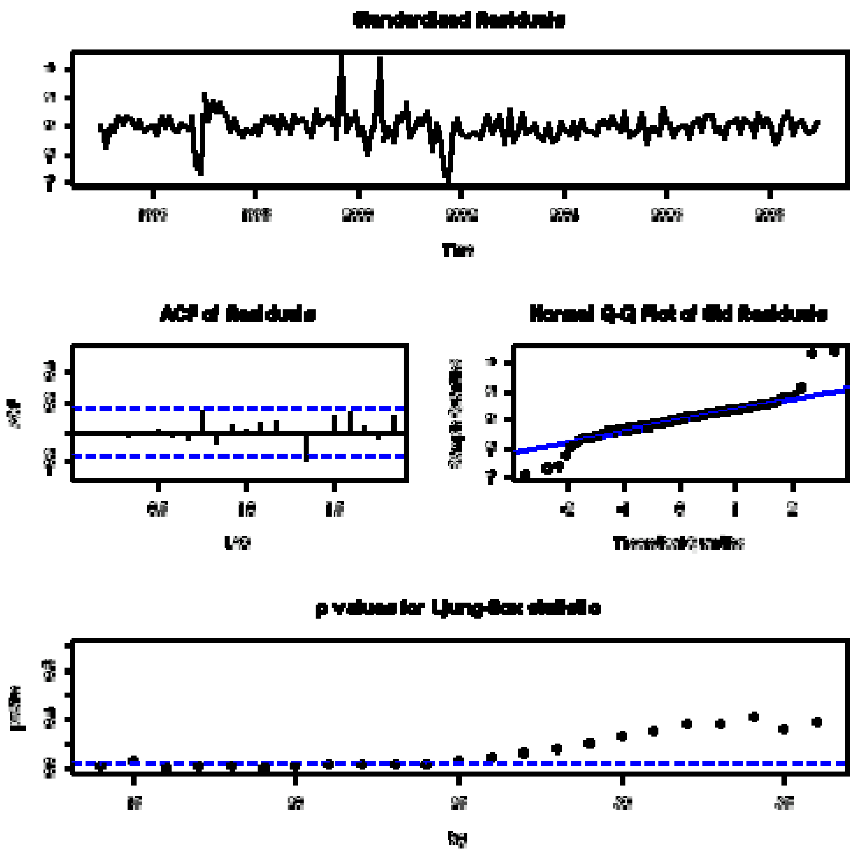

Figure 7-4 Diagnosis plots of the POV

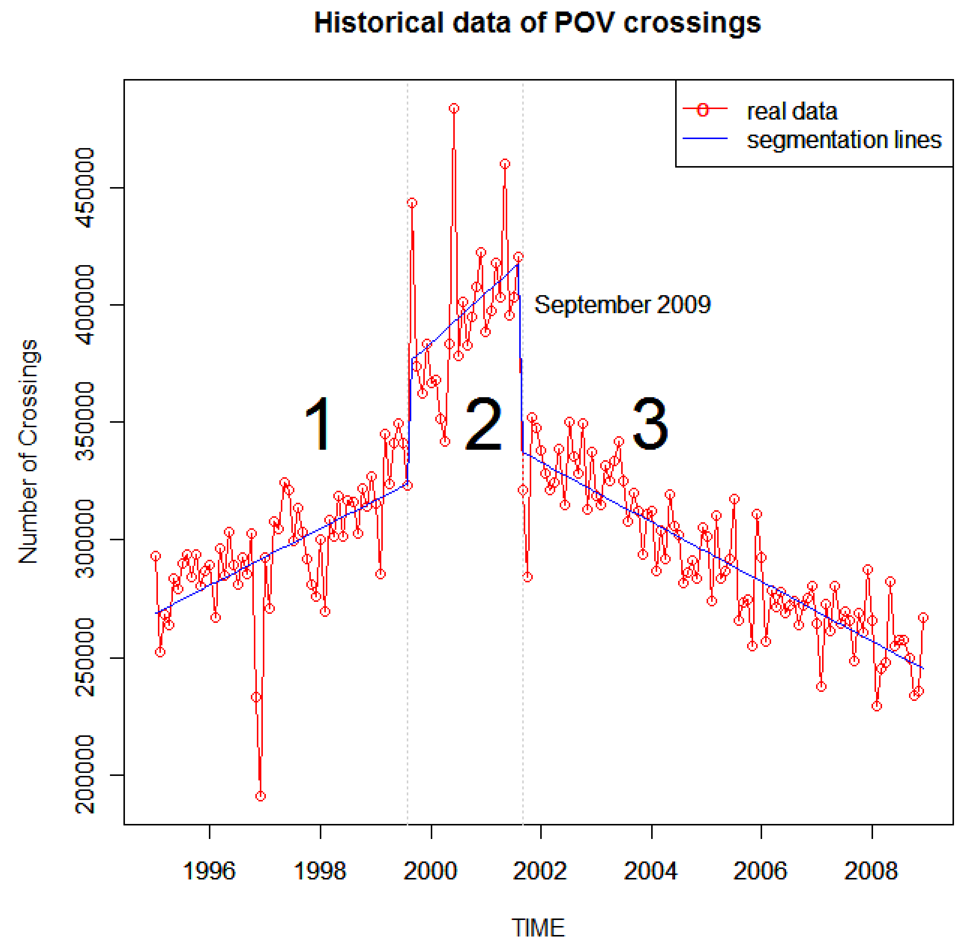

Figure 7-5 Piecewise regression on POV data

7.3 Models for Pedestrian Traffic

According to our data as well as interviews with people who had work experience in the Nogales area, we found that the pedestrian traffic was the most sensitive of the three modes of traffic under study, i.e. the pedestrian traffic contained the most variability. Similar to the POV traffic, we built a time series model for the pedestrian traffic for short term forecasting and simple regression models for extended forecasting. As we stated in the Model Alternative section, we incorporated "Arizona Employment" as an exogenous variable into the time series model. However, given the high level of variability in the pedestrian crossing history, the long term forecast should be used with caution, and multiple scenarios might be considered.

The ARIMA model we used had structural parameters $(p,d,q)(P,D,Q)_S = (3,2,1)(2,0,2)_{12}$. Table 7-4 shows the computer output of the coefficients of the ARIMA models, which contains the coefficients of the parameters.

Table 7-4 Computer Output: The coefficients for the parameters, Pedestrian model

```
          ar1      ar2      ar3      ma1     sar1    sar2    sma1
sma2
      -0.5697  -0.4191  -0.2584  -1.0000  0.2254  0.2998  0.0090  -
0.0237
s.e.   0.0756   0.0851   0.0783   0.0155  0.4671  0.4512  0.4775
0.3786
      constant     AZemp
      -557.051  700.5331
s.e.       NaN  211.8245
```

AZemp: "Arizona Employment"

Since the pedestrian traffic data contained so much variability, it was not appropriate to produce the long term forecast using a time series model. Thus we used piecewise linear regression to build regression models on historical data of the pedestrian traffic, and used them for different scenarios in the forecast. We used the same method as for the POV traffic to locate the number and locations of the break points, and then fitted simple regression models for each segment. Note that the "Arizona Employment" factor was not involved in these simple regression models. The break points were 66, 97 and 142, which were corresponding to June 2000, January 2003 and August 2006 respectively. Figure 7-8 depicts the breakdown of the historical data and the corresponding fitted model of each segment. We labeled the segments as 1, 2, 3 and 4 according to time stamp of each segment as shown in Figure 7-8. The slopes of the lines corresponding to segment 1, 2, 3 and 4 are 236.4, 6745, 5053.0 and

-7149. We can see the slopes differ not only in the absolute values, but also the sign. The slopes of the first three segments are positive, while the slope for segment 4 is negative, which indicates the trend changed from increasing to decreasing.

Table 7-5 Computer Output: Models of different segments, Pedestrian data

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 389459.6     8206.6   47.46   <2e-16 ***
ind            236.4      212.9    1.11    0.271
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32960 on 64 degrees of freedom
Multiple R-squared: 0.0189,   Adjusted R-squared: 0.003569
F-statistic: 1.233 on 1 and 64 DF,  p-value: 0.271


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   329561      22393  14.717 5.49e-15 ***
ind             6745       1222   5.521 5.97e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60840 on 29 degrees of freedom
Multiple R-squared: 0.5125,   Adjusted R-squared: 0.4957
F-statistic: 30.48 on 1 and 29 DF,  p-value: 5.966e-06


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 425631.8    17990.6  23.659  < 2e-16 ***
ind           5053.0      681.1   7.419  3.2e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59340 on 43 degrees of freedom
Multiple R-squared: 0.5614,   Adjusted R-squared: 0.5512
F-statistic: 55.04 on 1 and 43 DF,  p-value: 3.201e-09


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   702805      31521  22.296  < 2e-16 ***
ind            -7149       2041  -3.503  0.00183 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 78060 on 24 degrees of freedom
Multiple R-squared: 0.3383,   Adjusted R-squared: 0.3107
F-statistic: 12.27 on 1 and 24 DF,  p-value: 0.001830
```

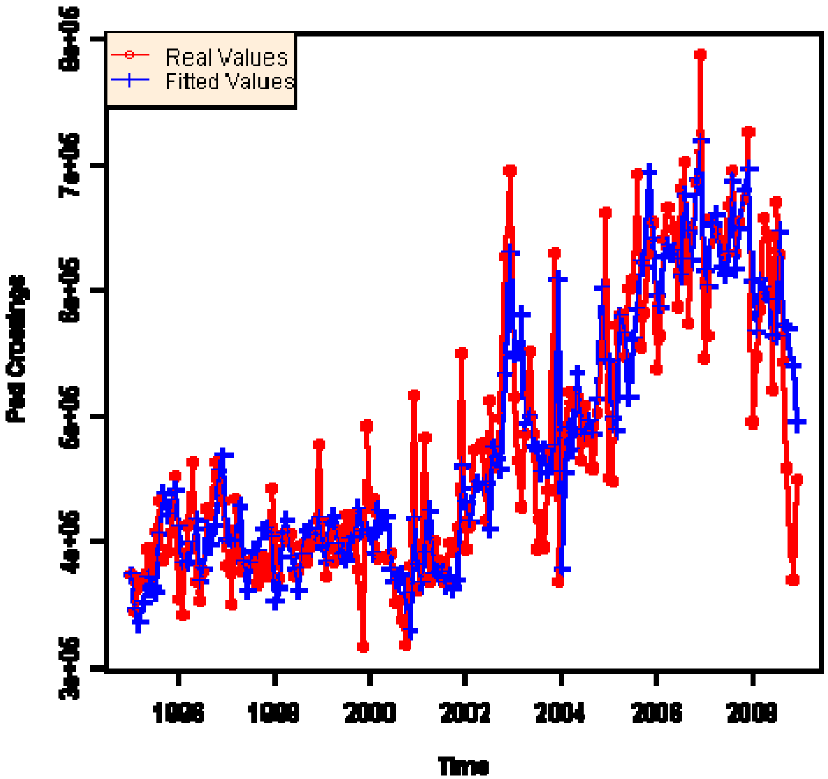

Figure 7-6 Fitted values vs real values

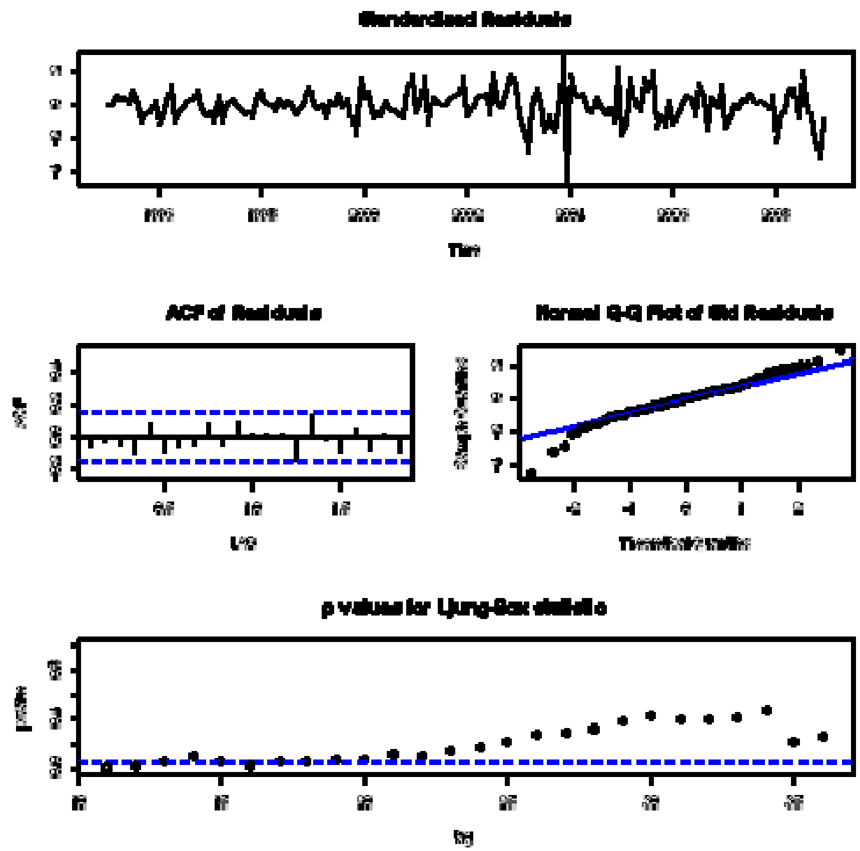

Figure 7-7 Diagnosis plots of the Pedestrian

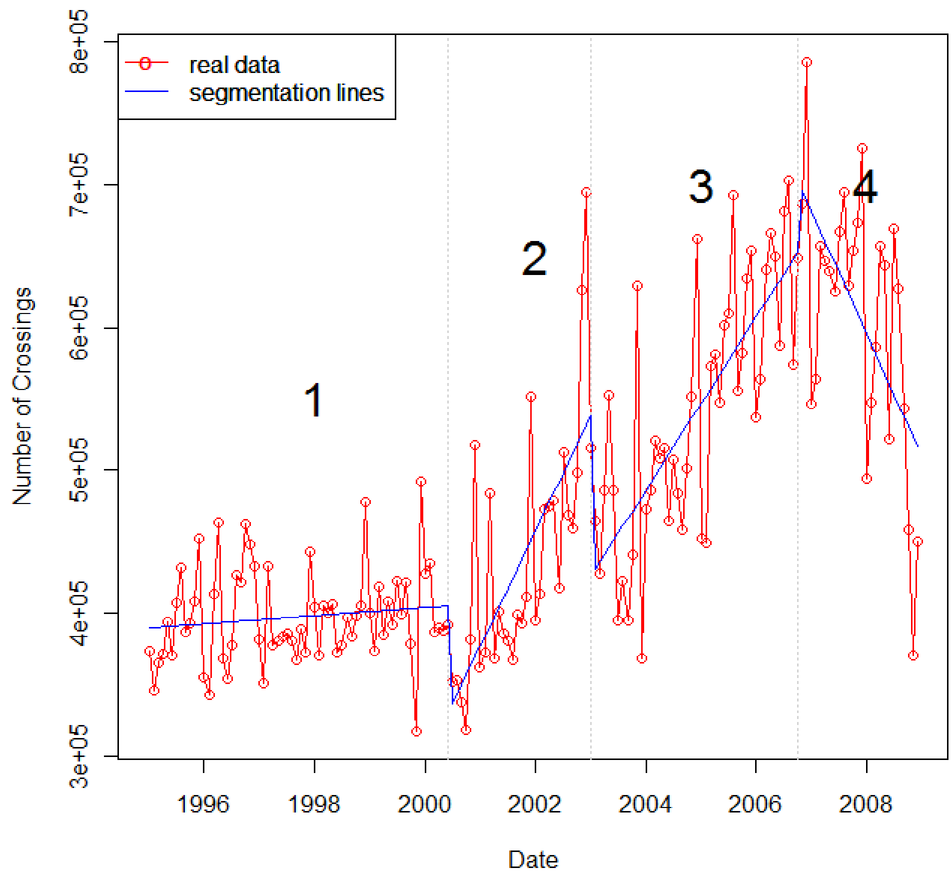

Figure 7-8 Piecewise regression on Pedestrian data

7.4 Models for Bus Traffic

We decided to build the model of bus traffic based on the number of bus passengers instead of the number of buses in section 6.4. When building the models for bus traffic, we had some issues that we did not face in other data. When we tested the model alternatives, we left out the last three years' data for validation purpose. We used the full set of data (note here the full data set indicates the data from January 2000 to December 2008) to build the forecast model. However, when we put all the data into the model, we found the ACF and PACF completely changed. Figure 7-9 compares the ACF and PACF of the data in training data set and the whole data set, where the left panels are ACF and PACF of the training data set, and the right panels are the ACF and PACF of the whole data set. To explain the cause of this difference, let's revisit the historical data since 2000, which are plotted in Figure 7-10 with red dots. We can see that the traffic before 2006 was increasing and the 2006 traffic increased compared to that of 2005, while 2007 was almost the same as 2006. The traffic started to decrease in 2008. The ARIMA model we used in the model alternative section actually did not work here, since singular numbers are produced and thus the model will not converge. We built new models according to the new ACF and PACF functions of the full data. The model we were using had the structural parameters $(p,d,q)(P,D,Q)_L = (9,1,7)(0,0,1)_{12}$. The coefficients of the model are shown in Table 7-6. The diagnostic plots are shown in Figure 7-11. From the plots, we can tell that the residuals conform to the normal assumption.

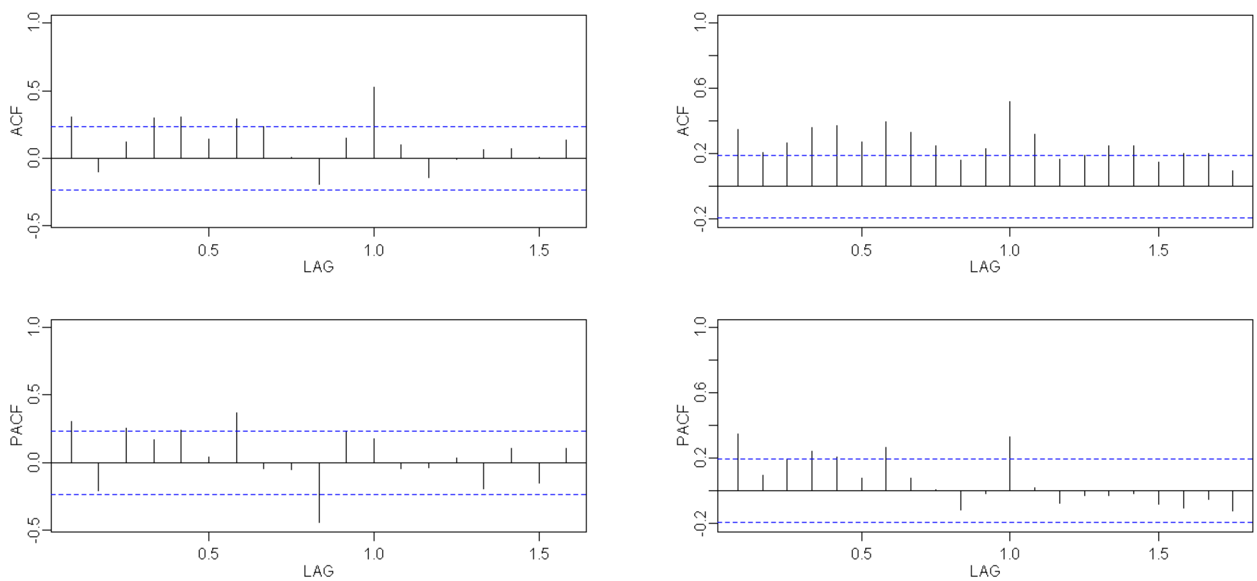

Figure 7-9 Comparison of ACF and PACF.

Left side are from the training data, right side are from the whole data set

Table 7-6 Computer Output: ARIMA model coefficients, Bus Passenger

	ar1	ar2	ar3	ar4	ar5	ar6	ar7	
	-1.0606	-1.5195	-1.4992	-1.0064	-0.7799	-0.5072	-0.1904	
s.e.	0.6870	1.1418	1.7192	2.0331	1.9004	1.6589	1.2629	
	ar8	ar9	ma1	ma2	ma3	ma4	sma1	XmatT1
	0.0074	0.1390	0.4316	0.4197	0.3989	-0.5887	0.3502	6.3708
s.e.	0.7695	0.3901	0.6211	0.6400	0.6447	0.6289	0.1009	1118.6583

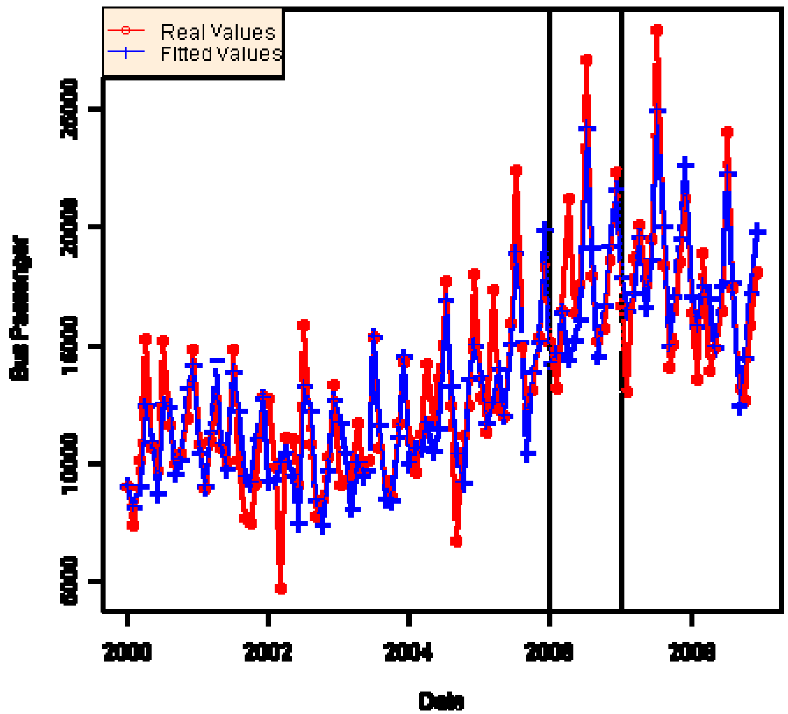

Figure 7-10 Historical data for bus passengers since 2000 vs. fitted data
(the February, 2003 data is replaced by average of January 2003 and March, 2003)

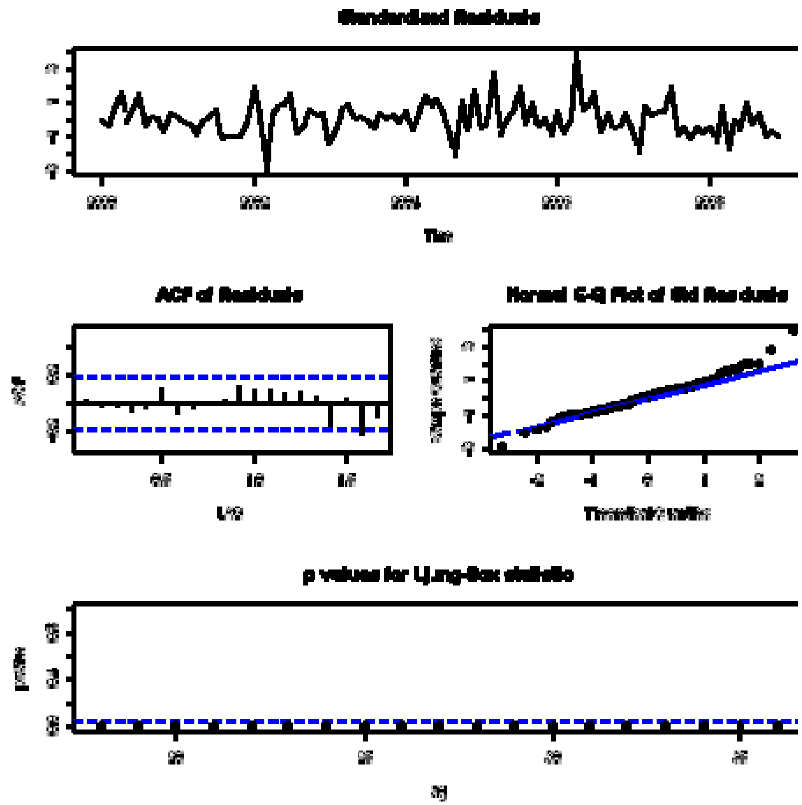

Figure 7-11 Diagnostic plots of the Bus Model

8 Forecasts

Once the forecasting models were built and validated using the available data, the next step was to use these models to provide forecasts of border crossings for the next 5, 10 and 15 years. However, given the unstable economic conditions at the time the models were built and the data was collected, instead of giving a single estimate for each of these time periods into the future it was decided to prepare multiple forecasts based on different scenarios.

For instance, during the model building stage, it was found that the Mexican Peso to US Dollar Exchange rate and the US Index of Industrial Production (IIP) significantly influenced border crossing traffic, especially commercial vehicle crossings. Thus, we first analyzed the trend scenarios of these two indices. Based on these scenarios we developed forecasts for border crossings for the different modes of traffic. In the following sections we provide forecasts for four modes of traffic: commercial vehicles, Privately Owned Vehicle (POV) pedestrians and Bus. For each mode, we provide a 5-year, 10-year and 15-year traffic forecast.

8.1 Forecasts for Commercial Vehicles

Overview of Exchange Rate and Index of Industrial Production(IIP)

The historical data for the exchange rate (If we don't indicate specifically, the exchange rate means the exchange rate between US Dollar and Mexican Peso, represented by the value of 1 US Dollar in Pesos) and US IIP(Index of Industrial Production) are available from the Federal Reserve Board. Various companies provide forecasts of these two indices, but most of these forecasts are only for a horizon of 36 months. We obtained the 36-month forecasts from the organization forecasts.org, and we call these forecasts the external forecasts. Since our intention was to give 5-year, 10-year and 15-year forecasts, we used forecast models in combination with these external forecasts. In cases where there was no external forecast available, we used only our own forecast. Due to the complexity of the forecast, we did not have a perfect forecast of these indices. What our forecasts did was to capture the general trend of the indices. We used simple regression for the 5 and 10 year forecast and used piecewise linear regression for the 15-year forecast. In the development of our forecasts, we considered different scenarios by assuming different trends of the underlying forecasting regressors (USIIP and Exchange rate). To begin we first reviewed the historical trends of the exchange rate and US IIP data. Figure 8-1 and Figure 8-2 plot the historical data (beginning March 2003) and the 36 month forecast (beginning May 2009) of the Exchange Rate and US IIP respectively, both of which were obtained from the organization forecasts.org. Figure 8-3 and Figure 8-4 show the data for a longer time span, where all available historical data was included and the forecasted data was excluded.

Considering that there was a devaluation of the Mexican Peso in late 1994 (Joseph A. Whitt 1996), we used the data starting from January 1995 to estimate the long term trend for the Exchange Rate. Also, because another significant devaluation of the Mexican Peso occurred during late 2008/early 2009, the trend of the Exchange Rate beginning in 1995 provided insightful information for the trend of the Exchange Rate after 2008/2009. The US IIP data, with its history dating back to 1919, provided relatively better historical records for estimating the future trends, especially for those trends occurring after recessions.

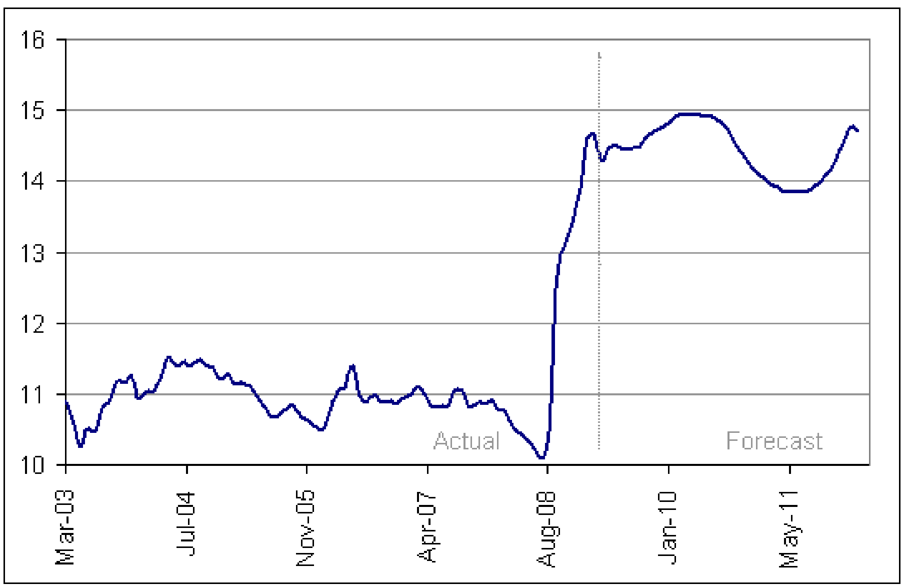

Figure 8-1 Mexican Peso to US Dollar Exchange Rate Forecast:
Past Trend & Future Projection (forecasts.org)

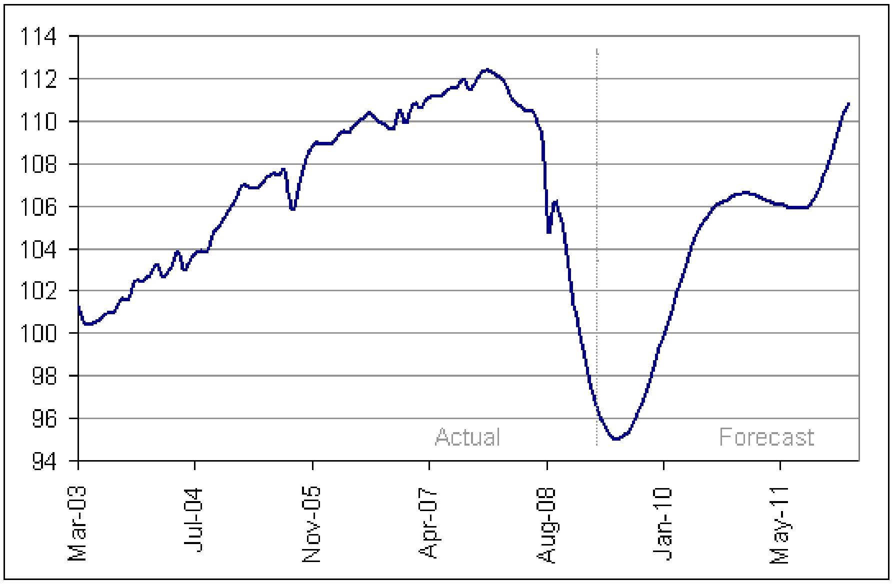

Figure 8-2 U.S. Industrial Production Index Forecast: Past Trend & Future Projection (forecasts.org)

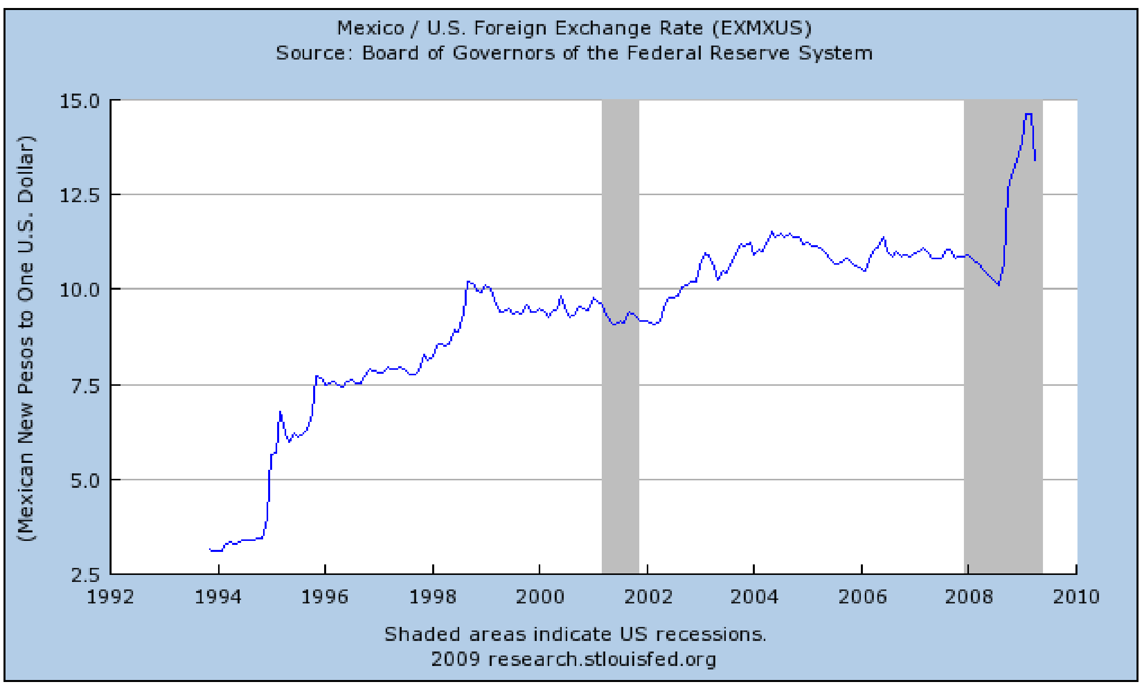

Figure 8-3 Mexican Peso to US Dollar Exchange Rate Historical Trend (Federal Reserve)

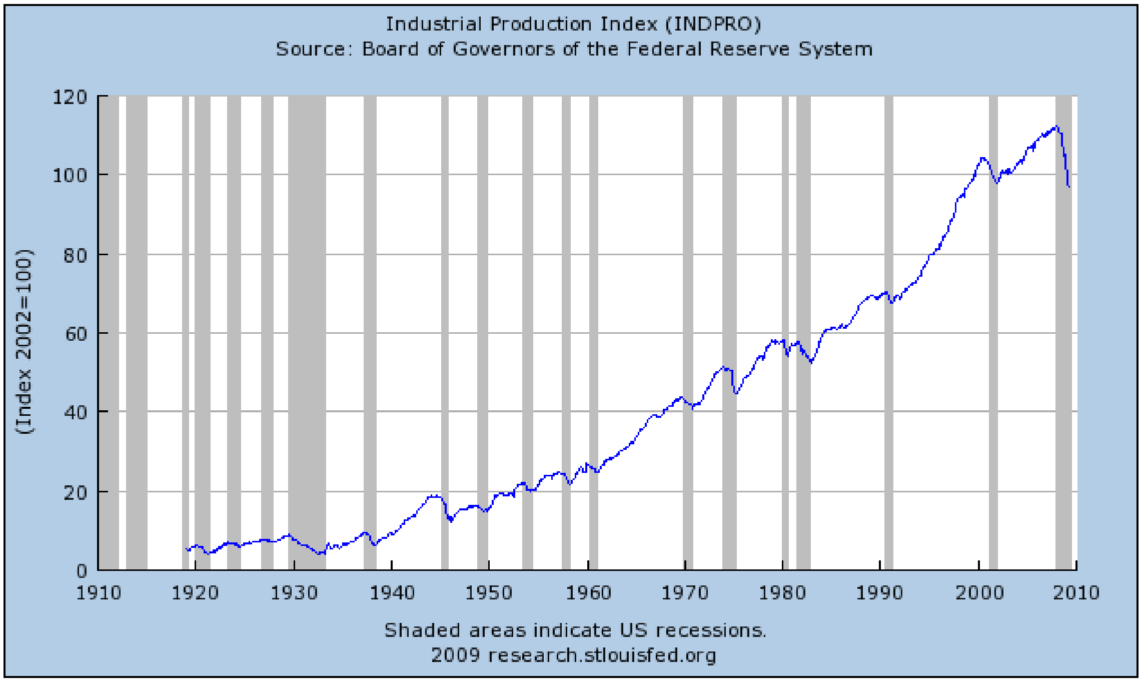

Figure 8-4 Historical data of U.S. Industrial Production Index (Federal Reserve)

Five-Year forecast

In order to estimate the 5 year trend of the exchange rate, we divided the historical data into groups of five-year length segments, and fitted those segments separately. We plotted the segmented data as shown in Figure 8-5. From this figure, we can tell that there are different possible trends for a 5 year time span, as shown by line segments 1, 2 and 3 which have very different slopes. For instance, segment 3 is almost a horizontal line, suggesting a very stable exchange rate. A similar situation occurred with the IIP data, which is shown in Figure 8-6, i.e., different 5-year segments resulted in significantly different trends.

In order to develop different forecast scenarios we chose different combinations of the exchange rate and the IIP trends as input to the models to obtain different forecasts of traffic. We defined different trends for the two indices (or variables) as shown in Table 8-1. There were 9 different combinations in total. For notation simplicity, we assigned each possible trend combination a two digit code, where the first digit represented the trend of the exchange rate and the second one represented the trend of US IIP. For example, the code 31 meant forecasts using "Staying relatively stable" for exchange rate and "Growing fast" for the US IIP. Figure 8-7 shows all the 9 different combinations of the future Exchange Rate and US IIP graphically. Figure 8-8 shows the forecasts of all the different scenarios graphically. Here the X axis represents the time and the Y axis represents the number of trucks crossing the border. Figure 8-9 aggregates the results to yearly level. The solid blue lines in Figure 8-8 are the forecasted values while the red dash lines represent one-standard deviation intervals. Note that all the scenarios have the same result for the first three years because they all used the same 36 months forecasts of the two indices from forecasts.org.

Table 8-1 Possible trend types for exchange rate and IIP within 5-year span

Exchange rate	US IIP
Growing fast (1)	Growing fast (1)
Growing mildly (2)	Growing slowly (2)
Staying relatively stable (3)	Staying relatively stable (3)

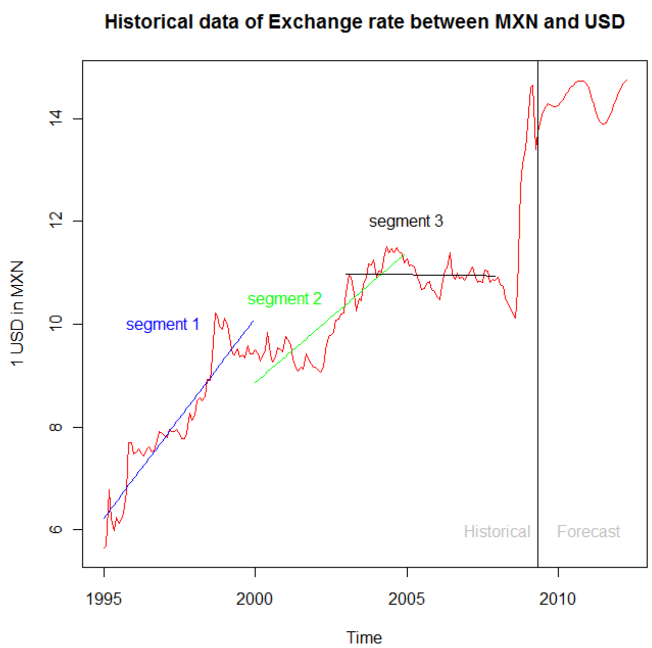

Figure 8-5 Historical Exchange Rate data with external forecast (5 year segments)

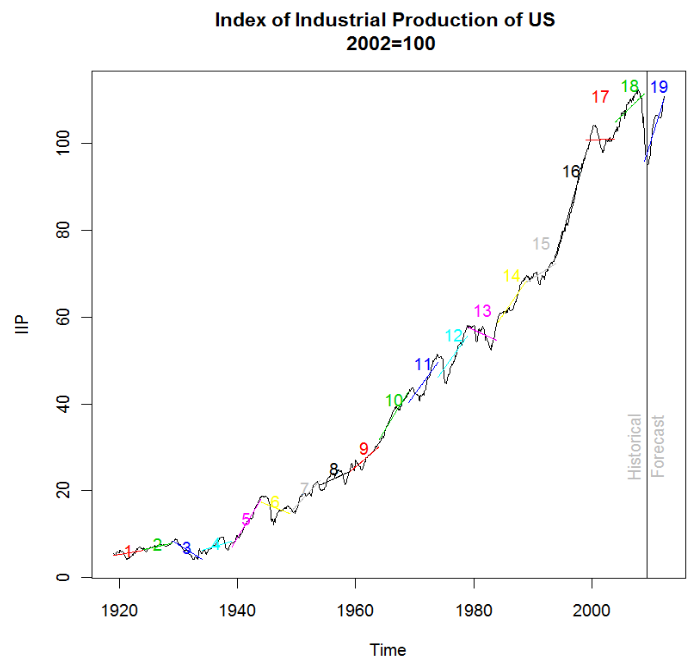

Figure 8-6 Historical data of US IIP with forecast from forecasts.org (5-year segments)

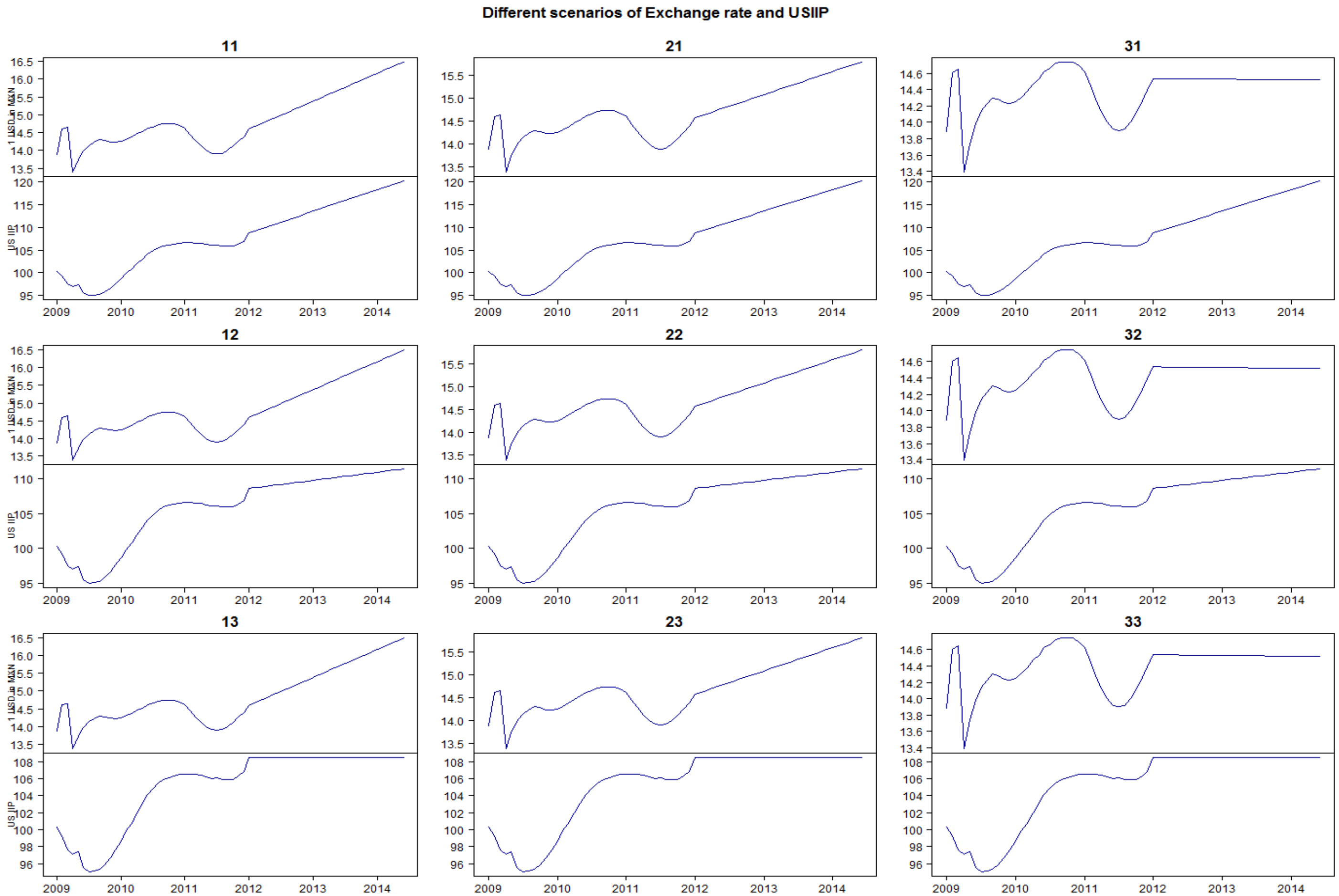

Figure 8-7 Different scenarios of Exchange Rate and US IIP

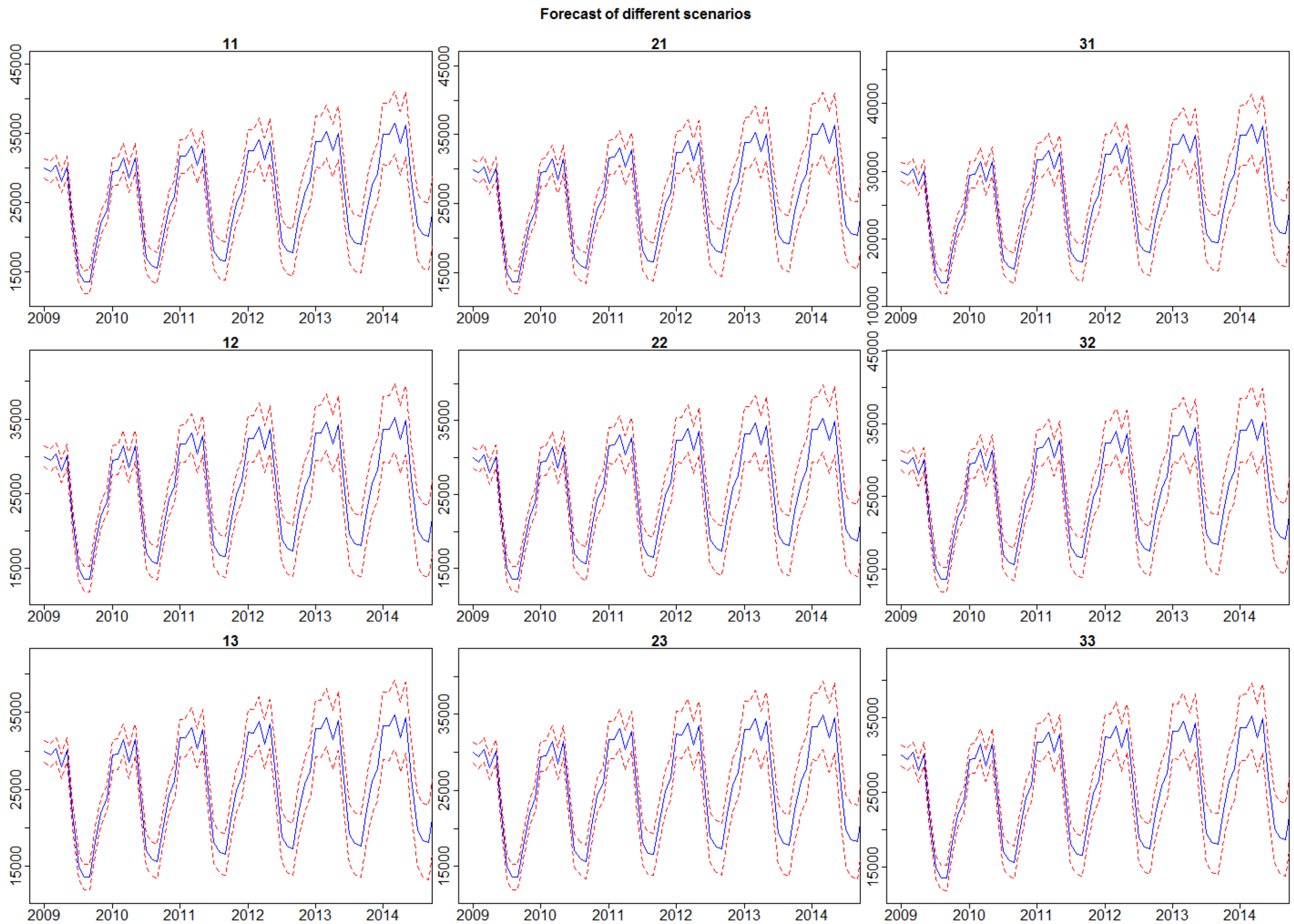

Figure 8-8 Forecasts under different scenarios;

solid blue line: the forecast, dashed red lines: one time standard deviation interval

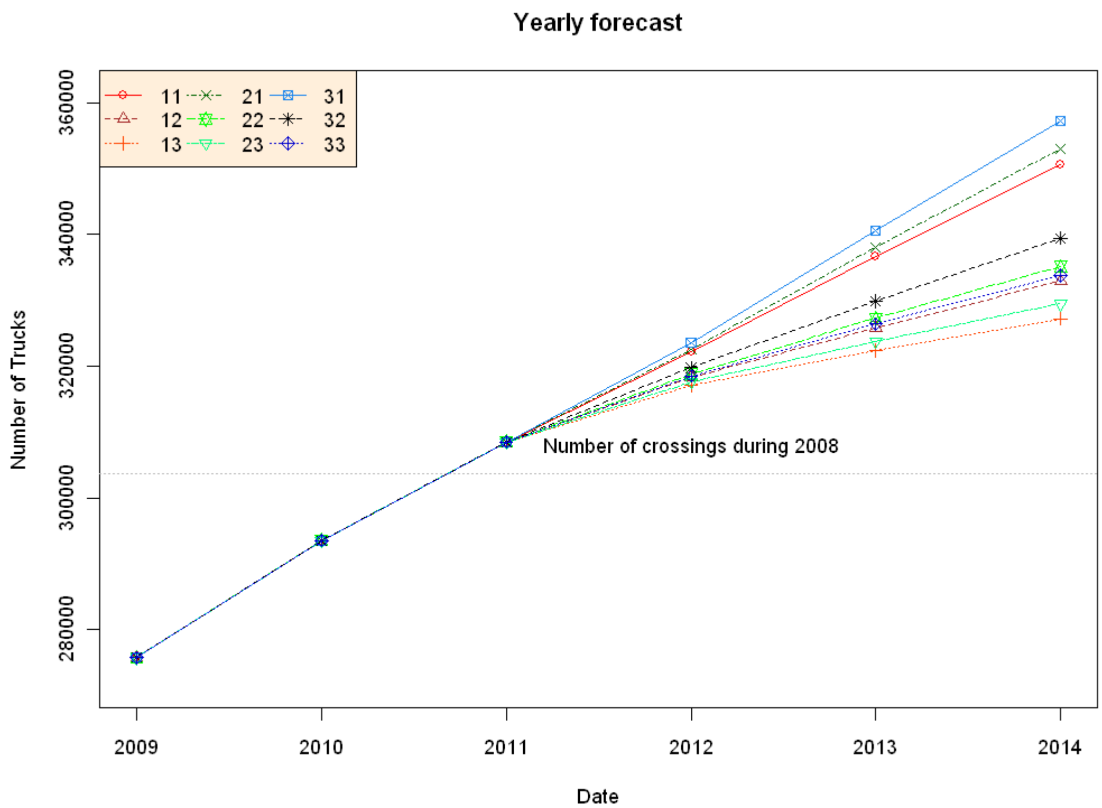

Figure 8-9 Yearly aggregation of the 5-year truck crossings forecast

Table 8-2 Five year forecast of different scenarios, compared to 2008

Increment of 2014 (%) 2008=100			
US IIP - Growth Speed (+ to -)	+ Exchange Rate Growth speed		-
	11	**21**	**31**
	15.4	16.2	17.6
	12	**22**	**32**
	9.6	10.3	11.7
	13	**23**	**33**
	7.7	8.5	9.9

Table 8-2 shows the percentage of increase of commercial vehicle crossings compared to the number of crossings in 2008. We can see the increment will be between 7.7% and 17.6%, based on the different trends of the exchange rate and the change of US IIP. The biggest increase will happen if the Exchange rate stays relatively stable and the US IIP grows fast. Comparing Table 8-2 by columns, we can tell that for the same trend of Exchange Rate, a "growing fast" trend of US IIP renders the largest increase of truck crossings. We can also conclude that for the same US IIP trend, a stable trend of Exchange Rate results in the biggest increase of truck crossings. Ten-Year forecast

For the ten-year forecast, we applied a similar procedure. When examining the trend of the Exchange Rate over a ten-year time span, we found that it was unlikely to be stable, as can be seen in Figure 8-10. Therefore, we only prepared two scenarios for the Exchange Rate, "growing fast" and "growing mildly". Figure 8-11 shows all the 10-year segments of the historical US IIP data. For the US IIP, historical data leads us to believe that all three possible trends could still occur during a 10-year time span, so we kept the same three US IIP scenarios as we did for the 5-year forecast. We used a similar coding method to that used in the five-year forecast to indicate the different scenario combinations. Table 8-3 lists all the scenarios we considered for the Exchange Rate and US IIP. Due to the long term uncertainty, we only give yearly forecasts as opposed to the monthly forecasts that were given in the 5-year forecast.

Table 8-3 Possible trend types for exchange rate and IIP within 10-year span

Exchange rate	US IIP
Growing fast (1)	Growing fast (1)
Growing mildly (2)	Growing slowly (2)
	Keeping relatively stable (3)

Figure 8-13 below shows the forecast of yearly commercial vehicle crossings under different scenarios and Table 8-4 shows the increase in number of crossings forecasted in 2019 when compared to those in 2008. From this table we can see that the ten year increase will be in the range of **18.8%** and **32.9%**. When comparing across columns, we can see that a "Growing mildly" trend of Exchange Rate renders a larger increase in the crossing of commercial vehicles. When comparing across rows, we can see that a "Growing fast" trend of US IIP renders a faster increase of the commercial vehicle crossings. Figure 8-13 depicts these differences graphically; here we can see the difference of increase is more significant among the scenarios with different US IIP trends. For the maximum growth of truck traffic, the US IIP should increase fast and the exchange rate kept relatively stable. For the minimum growth of truck traffic, the US IIP should stay relatively stable and the exchange rate grows fast.

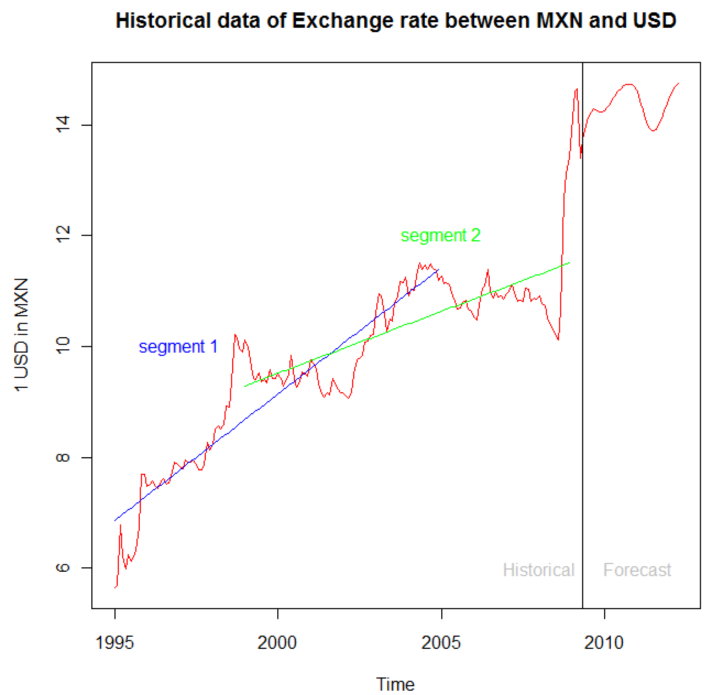

Figure 8-10 Historical Exchange Rate data with external forecast (10-year segments)

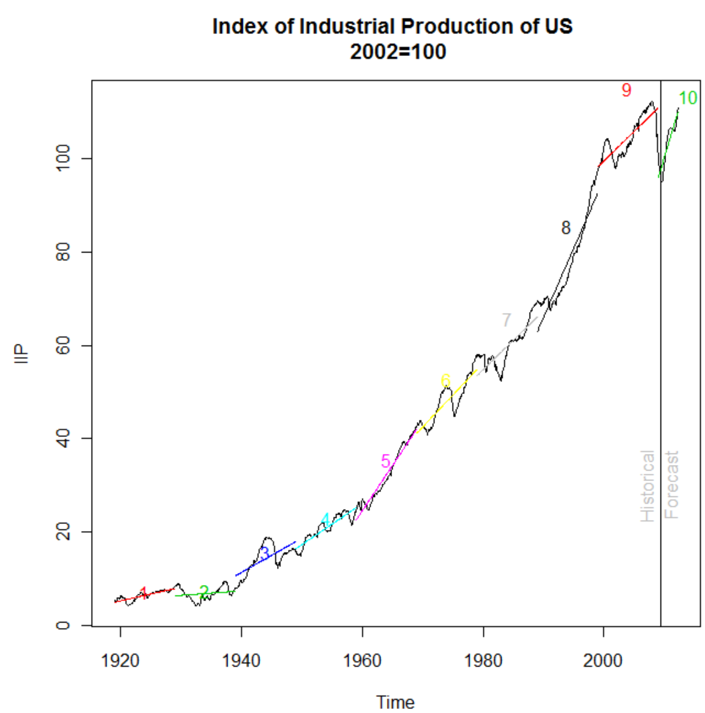

Figure 8-11 Historical data of US IIP with forecast from forecasts.org (10-year segments)

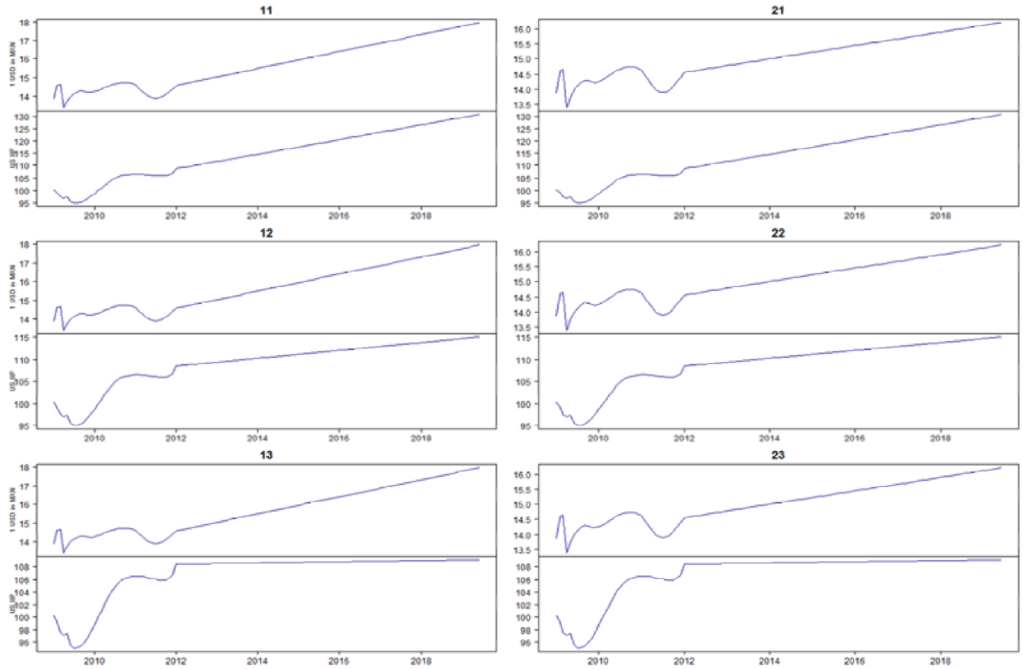

Figure 8-12 Different scenarios of exchange rate and US IIP (10-year segments)

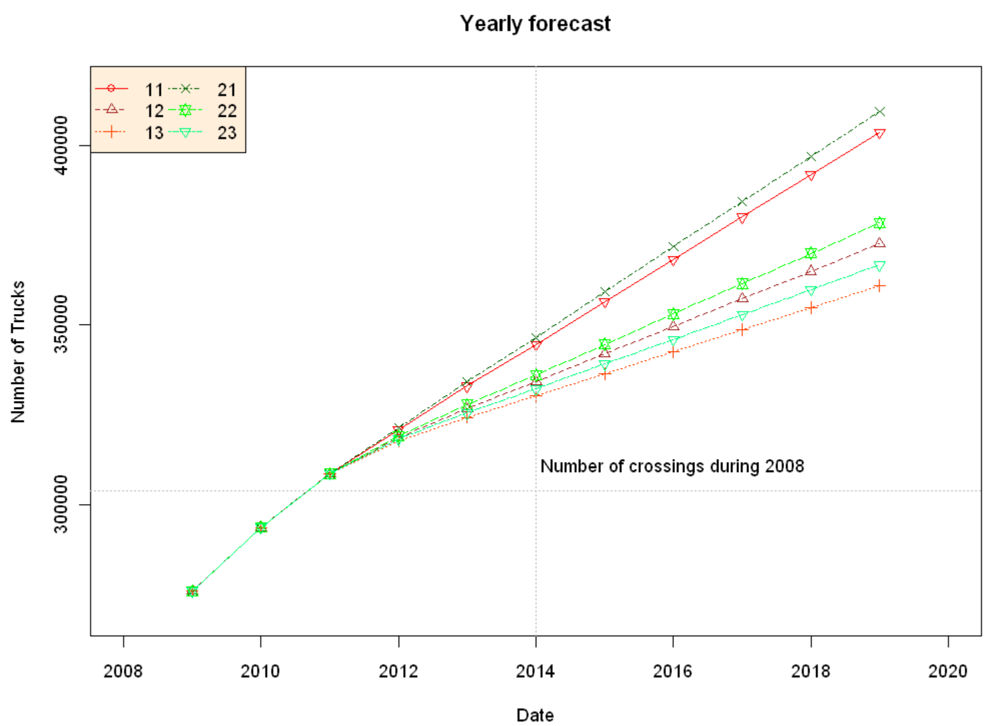

Figure 8-13 Yearly aggregation of the 10-year truck crossings forecast

Table 8-4 Ten-year forecast of different scenarios, compared to 2008

	Increment of 2019 (%) 2008=100	
	+ Exchange Rate Growth speed -	
US IIP -Growth Speed+	**11**	**21**
	32.9	34.8
	12	**22**
	22.7	24.6
	13	**23**
	18.8	20.8

Fifteen-Year forecast

For the fifteen year forecast we had to use a different approach to handle each of the scenarios for exchange rate changes because we only had 14 years of historical data available. Instead of separating the data into different segments and determining the speed of growth (stable, mild or fast), we used different forms of piecewise linear regression methods to build the scenarios. We used a package named "segmented" (Muggeo 2008) in the R system (R Development Core Team 2009) to locate the breakpoints. The two scenarios for Exchange rate are shown in Figure 8-14, where the blue lines indicates scenario 1 and the green lines indicates scenario 2. For the US IIP, we used a similar approach as the one used in 5-year and 10-year forecasts. Figure 8-15 shows the 15-year segments of the historical US IIP data. We categorized the trends into three different types as listed in Table 8-5.

Figure 8-16 shows all the possible combined scenarios of exchange rate and US IIP. Figure 8-17 shows the forecasted yearly truck crossings within a 15-year time span. The two vertical dash lines in Figure 8-17 mark the years 2014 and 2019. For this forecast, we focused on the data points after 2019. Table 8-6 shows the increase the yearly truck crossings for the year of 2024 compared to that of 2008. From this table we can see the increase will be between **29.1%** and **47.2%** in accordance with our various scenarios. When examining Figure 8-17, one can see that for forecasts with the same US IIP trend, the predicted forecasts over the 15-year time span will be very close. In long term, the US IIP may play a more important role than the exchange rate in influencing border crossing traffic. A fast growing US IIP trend Is the scenario associated with the fastest growth in truck traffic.

Table 8-5 Possible trend types for exchange rate and IIP within 15-year span

Exchange rate	US IIP
Blue Scenario (1)	Growing fast (1)
Green Scenario (2)	Growing slowly (2)
	Keeping relatively stable (3)

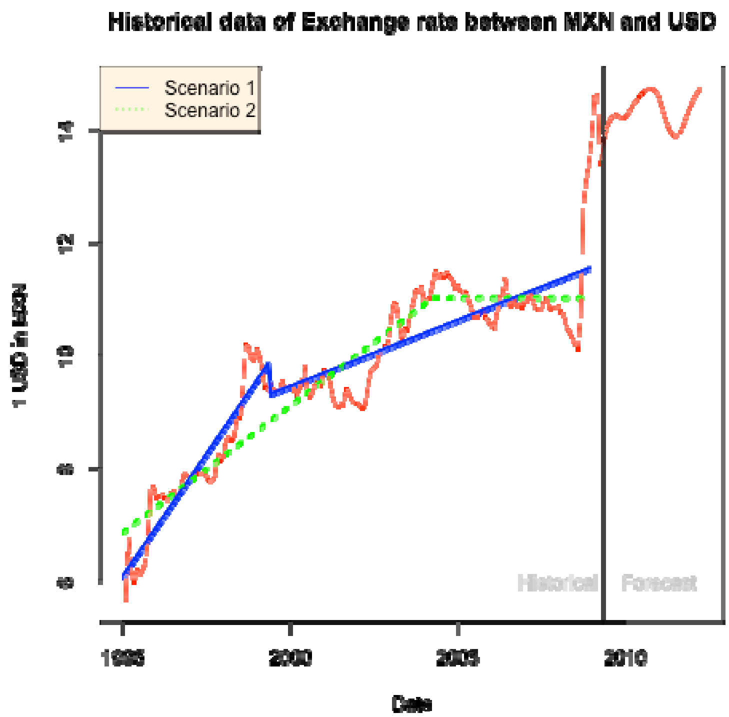

Figure 8-14 Different segment methods for Exchange Rate

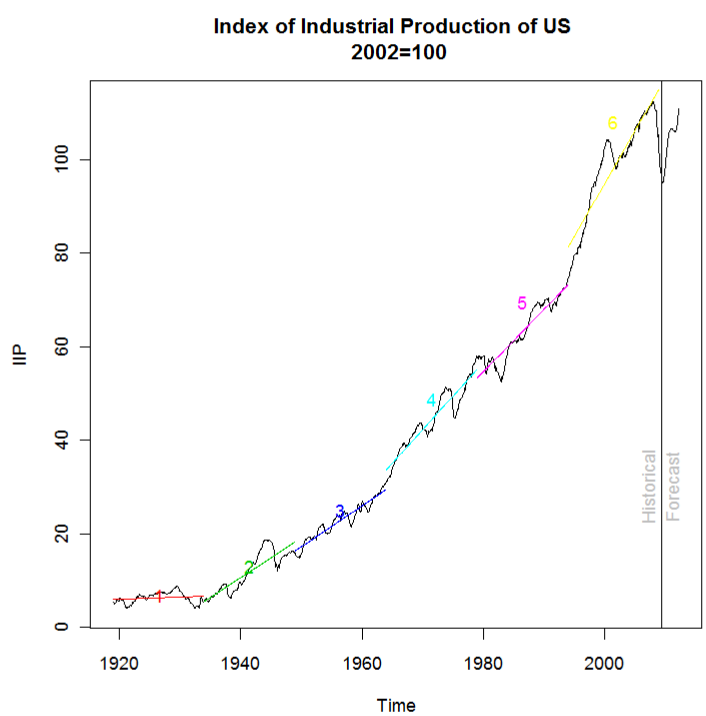

Figure 8-15 Historical data of US IIP with forecast from forecasts.org (15-year segments)

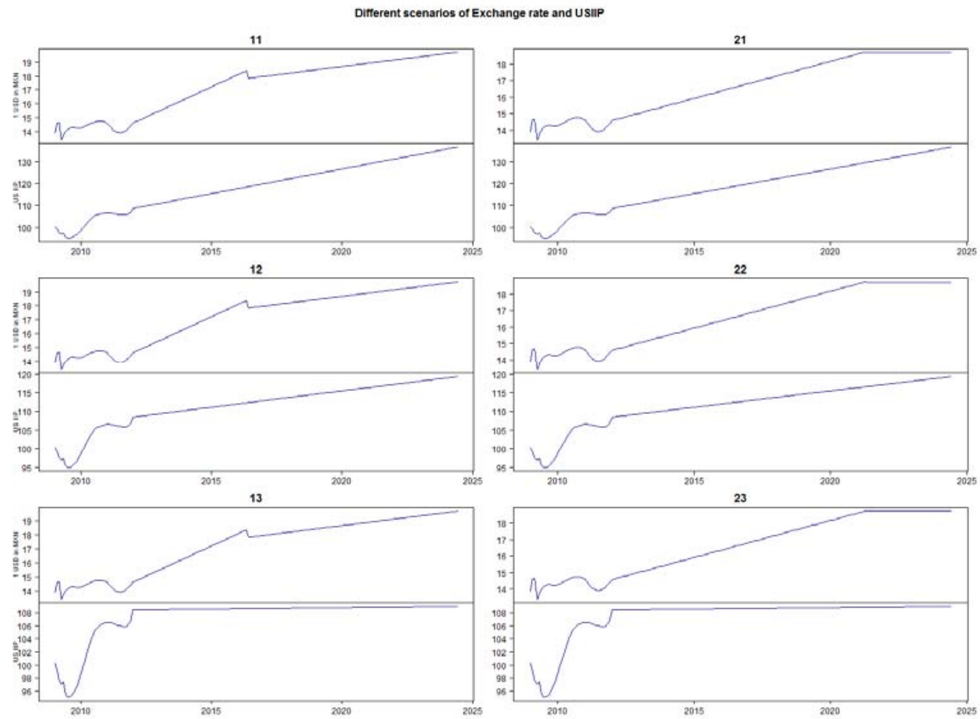

Figure 8-16 Different scenarios of exchange rate and US IIP

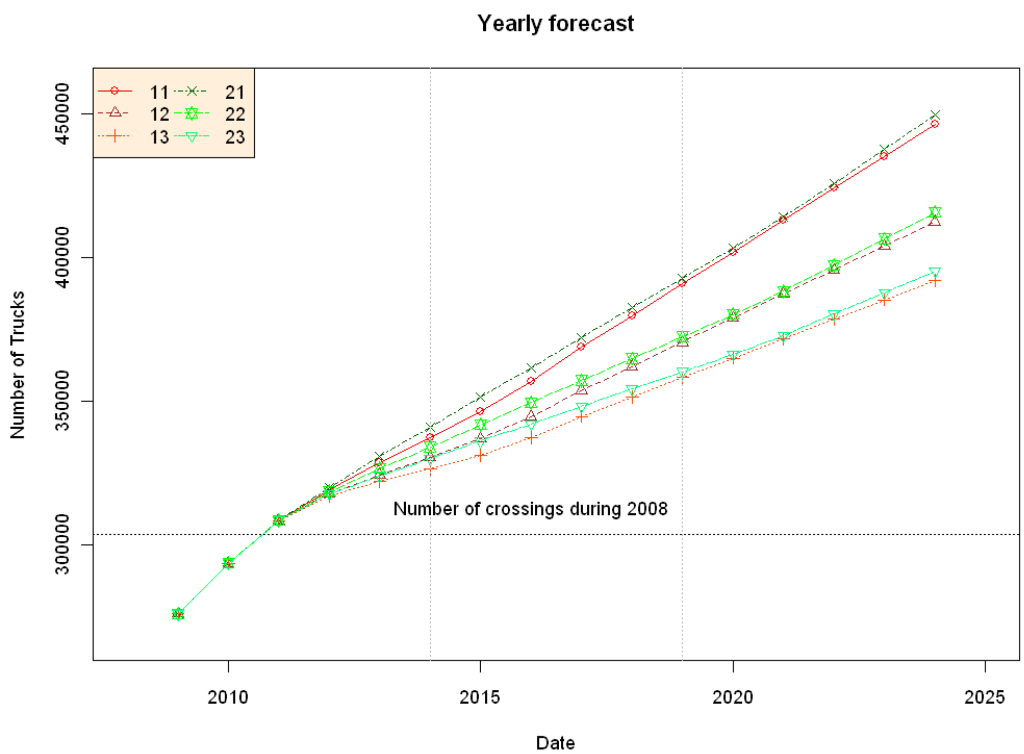

Figure 8-17 Yearly crossing forecasts of different scenarios

Table 8-6 Fifteen-year forecast of different scenarios, compared to 2008

	Increment of 2024 (%) 2008=100	
US IIP -Growth Speed+	**11**	**21**
	47.2	42.3
	12	**22**
	35.9	37.0
	13	**23**
	29.1	30.2

8.2 Forecast for the POV

As we stated previously, we used the time series model to produce the five-year forecast and used the regression model to produce the extended forecasts.

Figure 8-18 depicts the five-year forecast of POV crossing, which mainly is an extension of the decreasing trend of segment 3 in Figure 7-5. Considering the recession started in late 2007, this forecast seemed reasonable. However, we were not sure what the trend would be after the economy recovers from the recession. Segment 1 in Figure 7-5 shows the trend of POV crossings after the 1994 Mexican Peso crisis, which was increasing until "9/11" happened.

Figure 8-19 depicts the forecast for 10 years and 15 years, where we assumed the POV traffic would start to recover after the current recession is over. Extra attention should be paid to the turning point marked by the red dashed circle around 2014. Although it was drawn around 2014, it was meant to suggest that the turning point will occur when the economy recovers from the recession, which will happen at an unknown point of time into the future. The two scenarios in Figure 8-19 were based on the trends of segment 1 and segment 2 in Figure 7-5 respectively. They showed a significant difference in long run. Table 8-7 shows the forecasted POV crossing under these two scenarios. When comparing the highest previous crossing level, which was 2000, scenario 1 was equal to the previous high, while scenario 2 slightly exceeded it.

Table 8-7 Forecasted POV crossing

	Historical Highest(2000)	Bench mark 2008	2019	2024
Scenario 1	4682 K	3027 K	3264 K	3988 K
Scenario 2	4682 K	3027 K	3770 K	5050 K
Difference between scenarios			506 K	1062 K

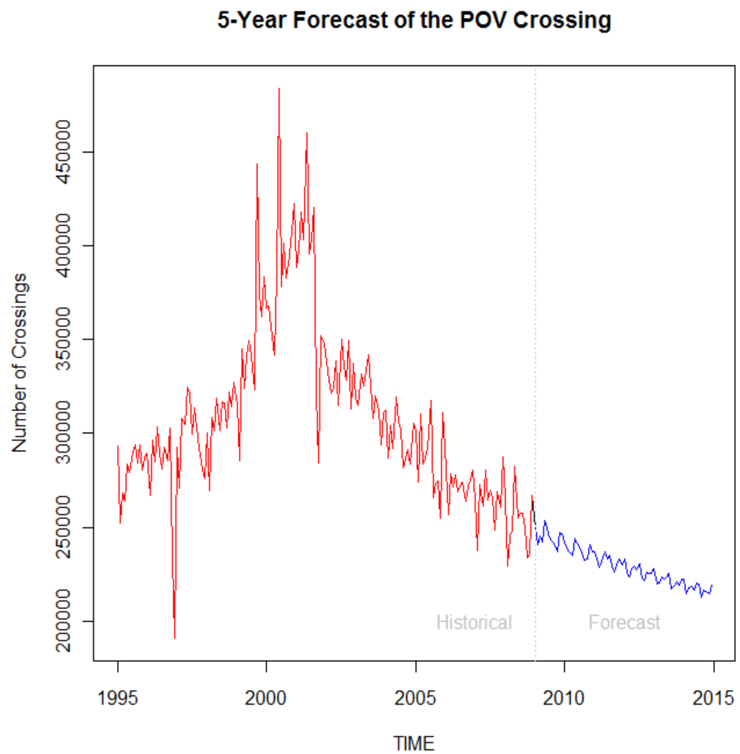

Figure 8-18 Five Year forecast of the POV crossing

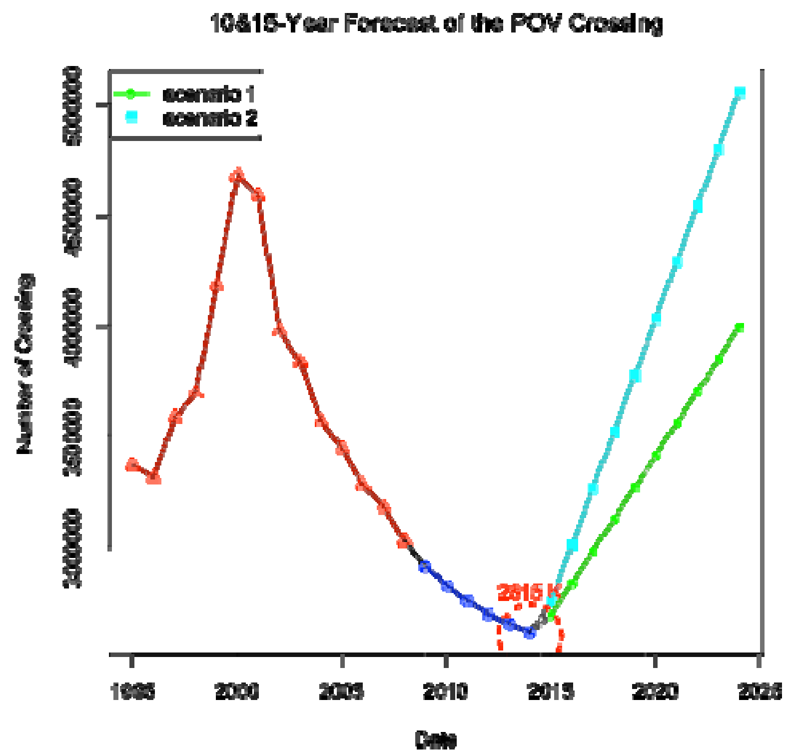

Figure 8-19 10 & 15-Year forecast of the POV crossing

8.3 Forecast for pedestrian crossings

As we did for the POV data, we produced the 5-year forecast of the pedestrian traffic by ARIMA model and the extended forecast by the regression method. We used "Arizona Employment" as an external variable in the ARIMA model, thus we first produced a forecast of "Arizona Employment". We used a 2nd order polynomial function to fit the "Arizona Employment", which is shown in Figure 8-20. Again, the forecast for "Arizona Employment" was not meant as an accurate forecast, but as an attempt to capture the main trend.

Figure 8-21 depicts our 5-year forecast of the pedestrian traffic, which was a monthly forecast. The overall trend was going down, which continued the trend of segment 4 in Figure 7-8. As we mentioned in previous section, we were unsure when the current recession would be over, thus we were not sure when the descending trend of the pedestrian crossings would end, since pedestrian crossings are very sensitive to economic climate changes.

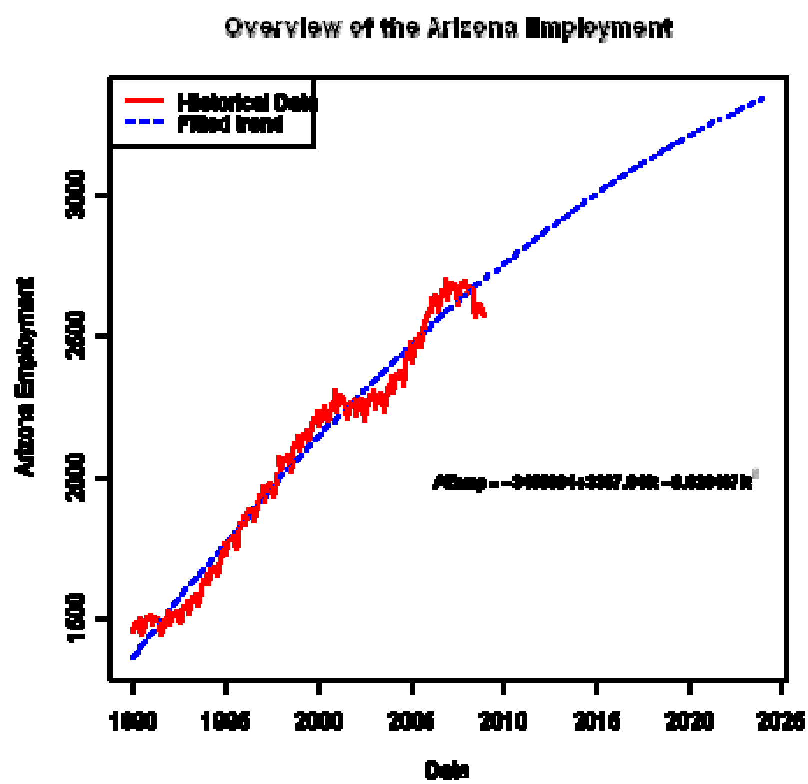

Figure 8-20 Historical data and a 2-order polynomial fit

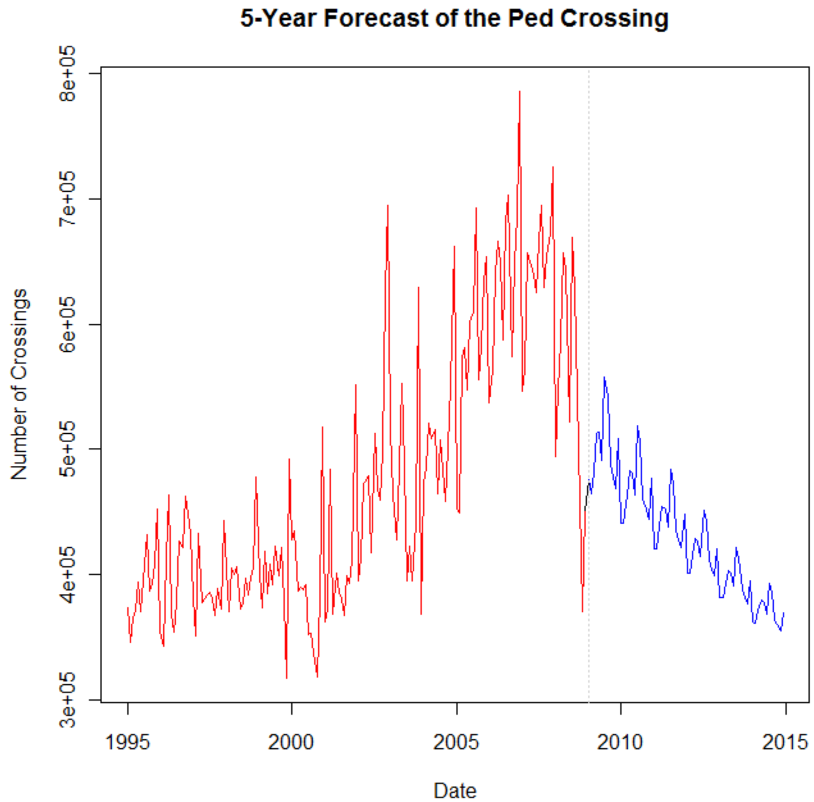

Figure 8-21 5-Year forecast of the pedestrian crossings

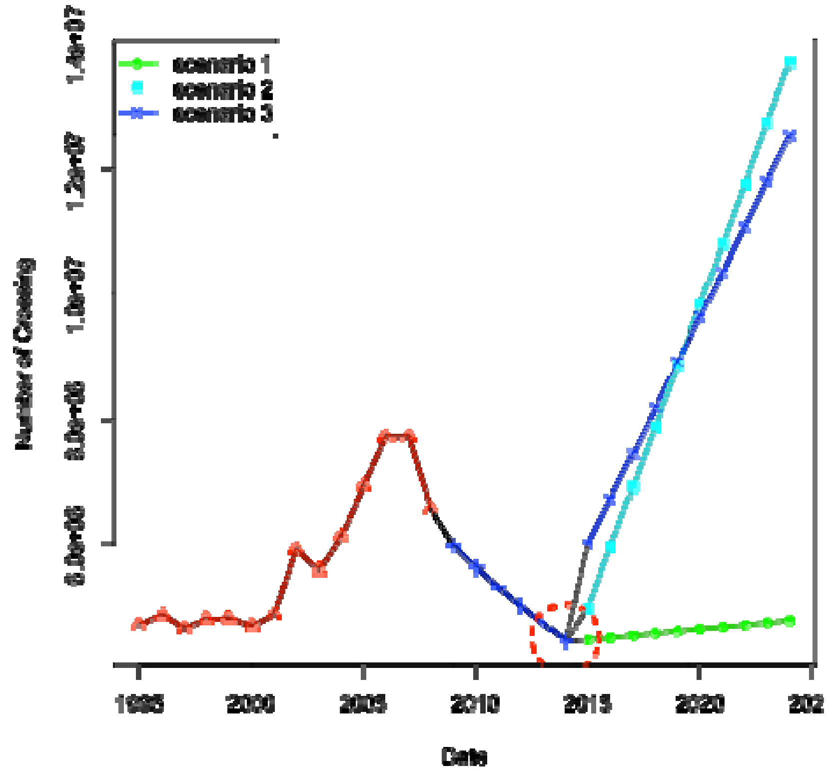

Figure 8-22 10 & 15-year forecast of pedestrian crossings

Table 8-8 Forecasted pedestrian crossing

	Historical Highest(2006)	Bench mark 2008	2019	2019/2008 (%)	2024	2024/2008 (%)
Scenario 1	7726 K 7726 K	6568 K 6568 K	4602 K	70.07%	4772 K	72.66%
Scenario 2			8858 K	134.87%	13715 K	208.82%
Scenario 3			8905 K	135.58%	12543 K	190.97%

Figure 8-22 shows the 10 & 15-year forecasts of pedestrian crossings. Scenarios 1 to 3 correspond to the trend of segments 1 to 3 in Figure 7-8. The dashed red circle in Figure 8-22 indicated the end of the economic recession, which would occur at an undetermined point in time in the future. Table 8-8 shows the predicted yearly crossings of pedestrians in 10 & 15 years and how these compared to the number in 2008. In scenario 1, the 2019 crossing of pedestrian will be around 70% of 2008, while the other two scenarios will be about 135%. For 15 years, scenario 1 will be about 73% of 2008, scenario 2 will be 209% of 2008 and scenario 3 will be 191% of 2008. Scenarios 2 and 3 are very similar in long run, while both of them had a significant difference when compared to scenario 1. Both scenarios 2 and three predicted the increasing rate to be much faster than that of scenario 1.

8. 4 Forecast for Bus Passengers

We used the time series model we built in the model section to produce the five-year forecast, and used simple regression models to produce the extended forecast. The number of passengers between 2002 and 2007 increased much faster than other years, so we used the data from 2000 to 2007 to build one regression model, and used all the data to build another one. Thus, we have two scenarios for forecasts.

1. Scenario 1: Used all the data and the regression model had a slope of 88.54.
2. Scenario 2: Used data between 2002 and 2007, and the regression model had a slope of 155.6.

Note that when building the models, we numbered the time periods consecutively. For example, for the data between 2002 and 2007, we marked January 2002 as 1, February 2002 as 2, and so on.

All the forecasts were given at a yearly level. As we observed from the data there is a great deal of variability in the data so we think a monthly forecast is not likely to be useful.

Figure 8-23 shows the five year forecast of the bus passengers. According to the ARIMA model, bus traffic will stay relatively stable over the next few years if the current condition does not change. Figure 8-24 shows the yearly forecasts of the bus passengers. Similarly, the turning point circled by the dashed red circle was an imaginary point, which indicated the ending of the current recession. Table 8-9 shows the forecasted bus passengers of 2014, 2019 and 2024 respectively. Also, we compared the predicted number of crossings to the crossings of 2008. In both of the scenarios, the number of crossings will increase. However, the scenarios differ in terms of the increasing rate. For the 2019 and 2024 forecasts, we had two scenarios, which were based on the different regression models we described previously. The future increases will be higher if the factors influencing bus passenger traffic are similar to those between 2002 and 2007. However, the factors driving bus passenger traffic still should be subject to further study.

Table 8-9 Forecasted bus passengers

	Benchmark 2008	2014	2019	2019/2008(%)	2024	2024/2008(%)
Scenario 1	195741 (196 K)	179706 (180 K)	243 K	135.00%	307 K	170.56%
Scenario 2			292 K	162.22%	404 K	224.44%

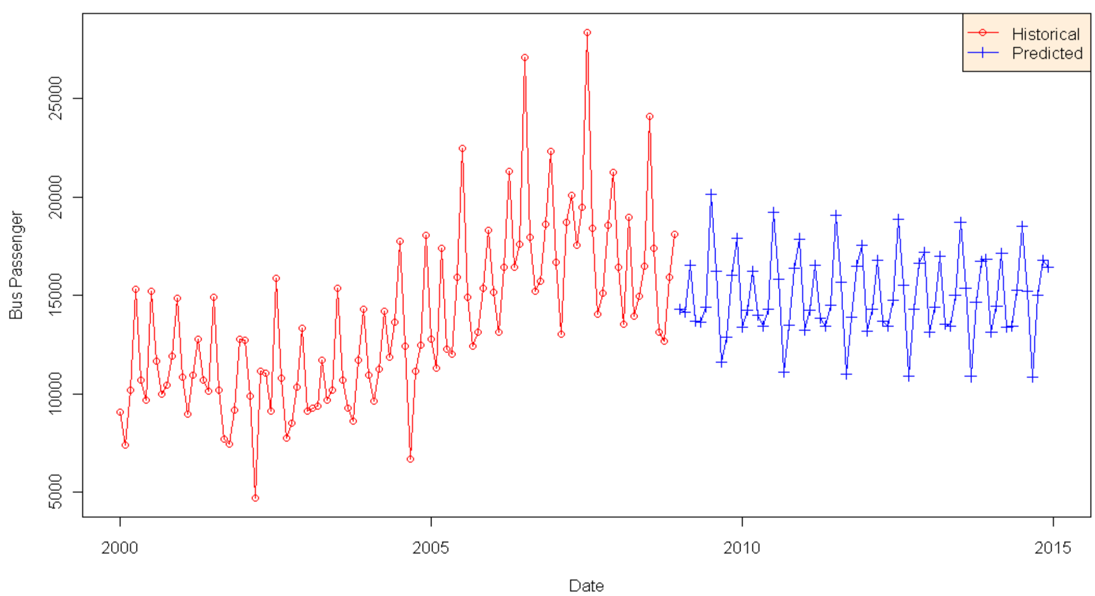

Figure 8-23 5-Year Forecast for the Bus passengers

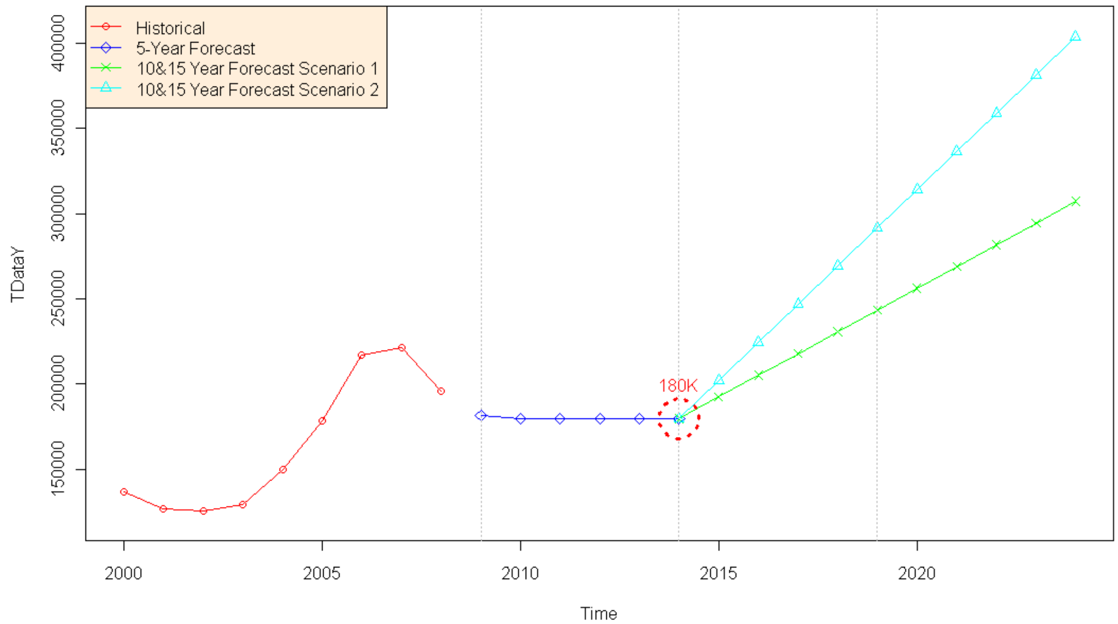

Figure 8-24 10 & 15 Year forecasts for Bus passengers

9 The Simulation Model

The simulation model we used for this study was an updated version of the model used in the ADOT project entitled *Logistics Capacity Study of the Guaymas-Tucson Corridor* (Villalobos et al.). This model was updated using data our team observed on a visit to the Mariposa POE conducted on Tuesday May 29, 2009. To begin our updates we made the following modifications to the physical infrastructure of the original simulation:

- Increased the number of highways/lanes all the way through the primary inspection area from two to four
  - The routing process assigned before a truck visibly enters the system was then changed accordingly
- Increased the number of inspection stations at each stop in the primary inspection area from two to four
- One highway/lane is designated as Free and Secure Trade (FAST) and assigned a different inspection time than that of the other three normal lanes
  - Added FAST as an attribute to trucks to determine which vehicles will be allowed to enter the FAST lane

In the previous version of the simulation trucks could cross over between FAST and normal lanes prior to entering the super-booths but in our updated version trucks may not make this crossover. Furthermore, the trucks in the topmost lane (FAST lane) are routed directly to the highway exit after completing primary inspections. Based on the data we collected in our visit we found that there was an insignificant amount of variation between the primary inspection times of a truck using the normal lanes as opposed to the FAST lanes therefore primary inspection times are the same for all vehicles. However, the simulation is set up so that the FAST versus normal lane inspection times can be easily changed later if variation between the inspection times is observed. The final determined inspection times of each area of the POE are displayed in the Table 9-1:

Table 9-1: Inspection times in Simulation Model

Inspection	Time Distribution (min)
Pre-Screening	ERLANG (0.72, 3)
Primary Inspection	ERLANG (1.33, 3)
*** 20% including ADOT in Super-Booth**	ERLANG (2, 3)
Document Revision	ERLANG (30.745, 3)
Full Inspection	ERLANG (82.2, 3)
Hazardous and weapons Enforcement	ERLANG (82.2, 3)
X-Ray	ERLANG (8.27, 3)
ADOT	TRIANGULAR (25, 30, 35)

Another change we made based on the observations from our visit was the percentage of trucks that are required to pass through additional inspections after completing primary inspections. This includes all CBP area inspections (Document, X-ray, Full, and Weapons Enforcement Inspections) and ADOT inspections. We determined that trucks that have a total time in the system of less than 4 min and 45 seconds left the system directly after primary inspections. We used this cutoff time and the data times we measured to determine the final percentage of trucks routed through only primary inspections. All other percentages of trucks requiring each type of inspection were kept consistent with those used in the Guaymas study (Villalobos et al.). These percentages are all shown in Table 9-2.

Table 9-2: Percentage of trucks requiring each type of inspection

Percentage	Description
100 %	Pre-Screening
100 %	Primary Inspection
30.74 %	Released to enter the US from Primary inspection (FAST lane)
69.26 %	Required further inspections and enter the compound (normal lanes)
***Out of the 69.26 % that require more inspection:**	
33 %	Required X-Ray
17 %	Required Full Inspection or Hazardous and Weapons Inspection
83 %	Required Documentation Review
20 %	Required to enter the ADOT yard for Inspection

After we updated and validated our simulation model we then tested the ability of the current infrastructure of the Mariposa POE to handle the daily demands of truck traffic we had predicted with our forecasts. In order to do this we ran the simulation model 5 times with a total runtime for 24 hours for each scenario. The scenarios are differentiated by the total amount of trucks tested for each scenario. Their respective arrival times (distributed according to an Erlang distribution) are summarized in Table 9-3.

Table 9-3 Daily Traffic Demand and Arrival times for each Scenario

Scenario	Daily Traffic Demand	Time between arrivals (Erlang distribution)
1-1	1928	0.00571
1-2	1800	0.00611
1-3	1759	0.00625
1-4	1945	0.00566
1-5	1816	0.00606
1-6	1775	0.00620
1-7	1976	0.00557
1-8	1847	0.00596
1-9	1806	0.00609
2-1	2131	0.00516
2-2	1969	0.00559
2-3	1909	0.00576
2-4	2161	0.00509
2-5	2000	0.00550
2-6	1939	0.00567
3-1	2302	0.00478
3-2	2139	0.00514
3-3	2042	0.00539
3-4	2325	0.00473
3-5	2159	0.00509
3-6	2062	0.00533

The logical flow of entities (trucks) in the simulation is explained in further detail in the diagram Figure 1 in Appendix H of the Guaymas study (Villalobos et al.). In summation, the logical process flow is as follows: when a truck enters the system it must pass through all primary inspections then depending on what attributes it has already been assigned it will either be routed straight to the highway exit or it will go through additional inspections and then be routed to the highway and exit the system.

The whole system can be divided into four different sections:

1. Pre-Screening and Primary Inspections: These are the first two steps in the process and all trucks are required to go through them.
2. Secondary Inspection: Different tasks can be done in this section such as: normal secondary inspection, Full (100%) inspection, weapons and enforcement inspection and others.
3. X-ray: three stations for x-ray inspection.
4. ADOT compound: ADOT's Motor Vehicle Division safety inspection and Federal Motor Carrier Safety Administration (FMCSA) inspections are conducted here.

The physical movement of the trucks can be observed in the animation of the simulation shown in Figure 9-1. Currently the trucks cross the border in four lanes, one of them being a FAST lane assigned to trucks with pre-cleared operators and CTPAT certified ownership, and the other three being regular lanes. All trucks will then enter a pre-screening station, follow to one of the four primary inspection super-booths, and then proceed to either Nogales, Arizona (if they were in the FAST lane) or else go on for further inspection in a counter clockwise (CCW) motion around the compound.

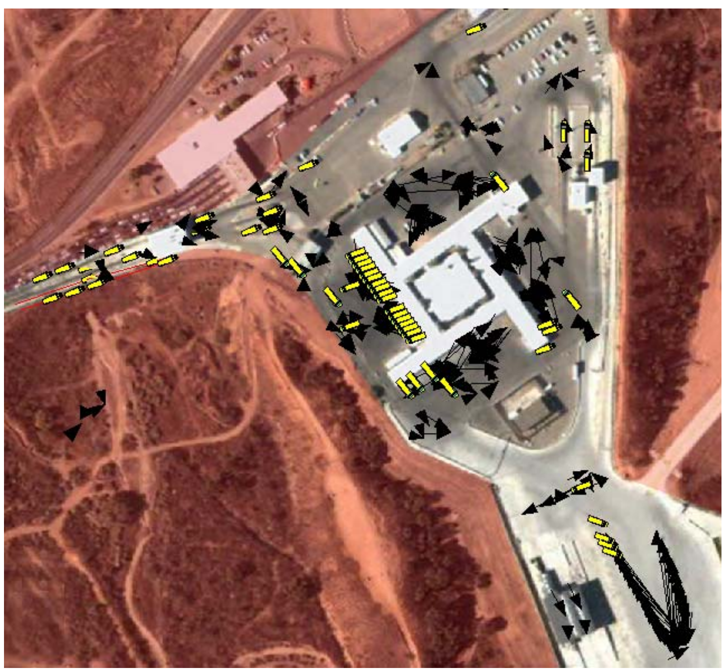

Figure 9-1 Image of Simulation

The results after running our simulation under the previously described scenarios are displayed in

Table 9-4 below. In this table, the first two columns show the scenario number and the number of trucks used as a daily demand input for each scenario. The third and fourth columns represent the total number of hours required to process all trucks and how many of those are additional hours over the current 11 hour workday that the port is open. The fifth column shows the average amount of time (in minutes) that a truck will spend in the system. The sixth and seventh columns show the 95% low and high confidence intervals for the maximum number of trucks that will wait in queue on the highway. The last two columns on the right show the bottleneck locations and their approximate utilizations for each scenario.

Table 9-4 Results of running the simulation

Scenario	# Trucks	Required Process time	Extra hours required	Avg. time in system (min)	Max in Queue (low 95%)	Max in Queue (high 95%)	Bottleneck	Approx. Utilization
1-1	1928	15.50	4.50	389.710	1888.26	1893.74	Super-booths	87.70%
1-2	1800	14.66	3.66	368.655	1759.58	1767.42	Super-booths	80.75%
1-3	1759	14.66	3.66	361.327	1719.37	1727.03	X-ray	81.50%
1-4	1945	16.22	5.22	395.704	1904.04	1908.96	Super-booths	81.30%
1-5	1816	15.14	4.14	367.047	1773.80	1781.40	Super-booths	78.84%
1-6	1775	14.64	3.64	362.902	1735.65	1740.15	Super-booths	79.42%
1-7	1976	17.04	6.04	401.391	1934.06	1940.34	Super-booths	84.87%
1-8	1847	15.62	4.62	370.460	1807.50	1813.50	Super-booths	82.69%
1-9	1806	15.15	4.15	363.639	1764.67	1772.33	Super-booths	81.27%
2-1	2131	17.51	6.51	424.579	2091.24	2096.76	Super-booths	84.12%
2-2	1969	16.92	5.92	399.541	1928.55	1936.45	Super-booths	78.84%
2-3	1909	15.60	4.60	387.990	1868.91	1875.69	Super-booths	89.04%
2-4	2161	17.91	6.91	432.178	2121.45	2128.35	Super-booths	84.56%
2-5	2000	16.51	5.51	407.981	1960.05	1964.35	Super-booths	81.25%
2-6	1939	15.89	4.89	388.628	1896.77	1904.83	Super-booths	86.60%
3-1	2302	18.39	7.39	458.475	2262.94	2270.06	Super-booths	87.69%
3-2	2139	17.21	6.21	426.991	2098.73	2107.67	Super-booths	81.43%
3-3	2042	16.65	5.65	412.149	2000.89	2008.31	Super-booths	81.44%
3-4	2325	18.82	7.82	471.270	2285.52	2291.08	Super-booths	87.71%
3-5	2159	17.28	6.28	433.375	2119.19	2127.61	Super-booths	83.49%
3-6	2062	16.70	5.70	416.790	2020.59	2030.21	Super-booths	87.21%

From this table we can observe the following:

- The maximum number of trucks that will wait in a queue on the highway according to our 95% confidence intervals is within the range of 2119 and 2127 trucks.
- For almost all scenarios the bottleneck location is the super-booths (Insp_PrePri_Norm1, Insp_PrePri_Norm2, Insp_PrePri_Norm3), with the exception of Scenario 1-3 where the bottleneck location is X-ray inspection.
- Based on our forecasts for daily truck traffic we can see that the current system is already at capacity due to the fact that in every scenario additional hours over the typical 11 hour workday are required for all trucks to be processed.

In order to validate our simulation we used the data times collected on our Mariposa POE visit and updated the simulation created in the Guaymas study (Villalobos et al.) so that the simulations output times match those we observed on our visit. We found that the most accurate manner in which to compare the times recorded in our visit with those in the simulation was by syncing the times a truck spent in the CBP area. In other words the average time a truck spends in the CBP area that we observed in our visit matches the times spent in the CBP area of trucks in our simulation model. Furthermore we changed the percentage of trucks that did not pass through any secondary inspections to match those observed in our visit. For further details regarding these changes refer to simulation appendix.

In conclusion, the results of running the simulation model compared to the actual inspection times measured in our visit to the Mariposa POE give us confidence in the validity of the results produced by our simulation model. Put another way, we found that if we tuned the simulation to our observed results for CBP times and percentage of trucks diverted, the overall average time spent in the system was relatively consistent with our observations. From our results we can see that given the forecasted future demands of traffic the system is already at capacity and would be unable to handle these traffic demands with the current infrastructure of the POE and length of workday (11 hours). We also found that for all but one of our forecasted scenarios the bottleneck of the system occurred at the same location, which we found to be the super-booths.

10 Conclusions and Recommendations

10.1 Findings

After completing the designated activities we agreed upon with ADOT we have drawn the following conclusions from our study:

1. The traffic characteristics at the POEs at Nogales are very different from that of other POEs, specifically in the seasonality pattern shown in the truck traffic. This is because of the high volume of fresh produce crossing this POE, which varies drastically in different seasons.
2. Economic indices are likely to be correlated with the level of border crossing traffic, especially for commercial traffic.
3. We provided various scenarios for each forecast due to the uncertainty of the future. The truck crossing traffic may increase up to 50% compared to the crossings in 2008, while the most conservative forecast showed the increase will be 30%. The POE should be prepared to handle the increase of the traffic with infrastructure and human resources.
4. The POV flows and pedestrian traffic flows are more sensitive to the economy than the truck traffic, thus the forecasts for these two types of traffic are likely to be less accurate than the forecasts for truck traffic.
5. As an example, we show our predicted truck crossing of 2009 against the recorded values in Table 10-1 below. The two columns of data are very close to each other.

Table 10-1 Pedicted truck traffic vs actual record of the year 2009

	Predicted Value	Actual Record
Jan	29,968	29,667
Feb	29,458	27,926
Mar	30,329	28,952
Apr	27,974	29,773
May	30,104	26,213
Jun	21,819	22,779
Jul	14,935	14,712

6. As with any long term prediction, one should be cautious when using the extended forecasts.

7. The simulation model suggests that the current setup of the Mariposa POE would be unable to handle future traffic demands. This lends credibility to the approved project to expand and reconfigure the port.

10.2 Future Research Opportunities

In our models, we used some exogenous variables such as the exchange rate. However, it was very difficult to forecast the future exchange rate and we were unable to find any single person or institute willing to give this kind of long term forecast. As a remedy to this problem we propose to use Delphi techniques to collect opinions from experts about the future trends of exogenous factors, and build our scenarios based on these opinions.

Another problem we encountered was that collecting all the necessary data for this project was difficult to accomplish during the model building process. For example, some data we found was incomplete. Therefore we suggest having a professional, technical and independent "data clearing house" to serve as the repository of border data and research results. We believe that such a "data clearing house" would not only be beneficial for this study, but also other similar studies involving data collection in border areas.

After all our analysis, we found many questions we could not answer based solely on the historical data. For instance, why after at least a 6-year upward trend has the POV traffic kept shrinking since 9/11? Furthermore, if the POV traffic does not start to increase even when the current recession is over, at what level will it become stabilized (i.e. it can’t keep shrinking down to zero)? Is there any relationship between the trends of POV and pedestrian crossings after "9/11" since they appear to be opposite to each other? What is the economic impact of not having the proper infrastructure or procedures for border crossings which either prevents people from crossing the border or make walk rather than drive across? Also, how is it that the traffic split between the Mariposa and DeConcini ports relates directly to the capacity of each one? How do people make the decision of what port to use that makes the overall system "efficient"?

11 References

Achim Zeileis, Christian Kleiber, Walter Kraemer, and Kurt Hornik. 2003. Testing and Dating of Structural Changes in Practice. *Computational Statistics & Data Analysis* 44: 109-123.

Achim Zeileis, Friedrich Leisch, Kurt Hornik, and Christian Kleiber. 2002. strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software* 7, no. 2: 1-38.

BTS. RITA | Bureau of Transportation Statistics (BTS). *Bureau of Transportation Statistics*. http://www.bts.gov/.

Donahue, ChrisLetter. 2009. Nogales POEs daily truck crossing data of 2008. February.

Farnum, Nicholas R., and Laverne W. Staton. 1989. *Quantitative Forecasting Methods*. Wadsworth Publishing Company, January.

Federal Reserve. St. Louis Fed: Series: EXMXUS, Mexico / U.S. Foreign Exchange Rate. http://research.stlouisfed.org/fred2/series/EXMXUS?cid=95.

———. St. Louis Fed: Series: INDPRO, Industrial Production Index. http://research.stlouisfed.org/fred2/series/INDPRO.

forecasts.org. Mexican Peso to U.S. Dollar Currency Exchange Rate Extended Forecast. http://www.forecasts.org/extended/peso.htm.

———. Industrial Production Index Extended Forecast. http://www.forecasts.org/extended/iip.htm.

Joseph A. Whitt, Jr. 1996. The Mexican peso crisis. *Economic Review*, no. Jan: 1-20.

Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2006a. Simple Linear Regression. In *Introduction to Linear Regression Analysis*, 12. 4th ed. WILEY, July.

———. 2006b. Multiple Linear Regression. In *Introduction to Linear Regression Analysis*, 64. 4th ed. WILEY, July.

———. 2006c. Variance Inflation Factors. In *Introduction to Linear Regression Analysis*, 111,335-340. 4th ed. WILEY, July.

Muggeo, Vito M.R. 2008. *segmented: Segmented relationships in regression models*.

R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. http://www.R-project.org.

Shumway, Robert H., and David S. Stoffer. 2006a. *Time Series Analysis and Its Applications: With R Examples*. 2nd ed. Springer, May 25.

———. 2006b. Building ARIMA Models. In *Time Series Analysis and Its Applications: With R Examples*, 143-164. 2nd ed. Springer, May 25.

U.S. DOT. 2002. *North American Free Trade Agreement – U.S. Department of Transportation Regulations*. Washington, D.C.: U.S. DOT, March 14. http://www.fmcsa.dot.gov/documents/nafta/Fact-Sheet.pdf.

Villalobos, J. René, Arnold Maltz, Omar Ahumada, Gerardo Treviño, Octavio Sánchez, and C. García Hugo. *Logistics Capacity Study of the Guaymas-Tucson Corridor*. ADOT & Arizona State University. http://www.canamex.org/PDF/FinalReport_LogisticsCapacity_Guaymas-TucsonCorridor.pdf.