

TIME-SERIES ANALYSIS OF BICYCLE COUNT DATA

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yutaka Motoaki

January 2015

© 2015 Yutaka Motoaki

Abstract

This research is to improve the analysis of cycling demand subject to weather conditions by analyzing time series of automatic cycling counts. I examined the performance of several ridership prediction models, including the Negative Binomial regression and time-series models such as SARIMA and SARIMAX. Using cycling counts for Portland, I show that the SARIMAX model that includes weather conditions (temperature and precipitation) as explanatory variables performs best in out-of-sample prediction. Future research in State Space models is needed for overcome the problems of SARIMAX when predicting ridership in periods with poor weather.

Biographical Sketch

Yutaka Motoaki was born and grew up in Tokyo, Japan. He is an MS/PhD student at Civil & Environmental Engineering at Cornell University. He received a BS in Environmental Economics from Oregon State University and MA in Economics from Binghamton University. His academic interests include non-motorized transportation, behavioral science, and disaster evacuation.

Acknowledgements

Thanks to my advisor, Ricardo Daziano for his guidance.

Table of Contents

ABSTRACT	II
BIOGRAPHICAL SKETCH.....	III
ACKNOWLEDGEMENTS	IV
LIST OF TABLES.....	VI
LIST OF FIGURES.....	VII
CHAPTER 1: INTRODUCTION	8
CHAPTER 2: METHODOLOGY	10
CHAPTER 3: CYCLIST COUNT DATA ANALYSIS	15
CHAPTER 4: RESULTS	21
CHAPTER 5: DISCUSSION/CONCLUSION.....	31
REFERENCES	33

List of Tables

Table 1. Coefficients Estimates.	25
---------------------------------------	----

List of Figures

Figure 1. Hawthorne Bridge in Portland, OR.	16
Figure 2. Average cyclists per hour.	16
Figure 3. Average cyclists per hour per day, per day of the week (left) and month (right).	16
Figure 4. Average monthly precipitation (left) and temperature (right).	16
Figure 5. Average ridership per day, during the month of November.	17
Figure 6. Average temperature per day, during the month of November.	17
Figure 7. Time series of daily ridership.	18
Figure 8. Autocorrelation function (ACF).	18
Figure 9. Average ridership and ACF after differencing.	20
Figure 10. Observed and fitted daily ridership between April and September, SARIMA.	21
Figure 11. Predicted ridership for the month of October, SARIMA.	22
Figure 12. Observed and fitted daily ridership between April and September, NB.	23
Figure 13. Observed and fitted daily ridership between April and September, SARIMA.	23
Figure 14. Observed and fitted daily ridership between April and September, SARIMAX.	24
Figure 15. Predicted ridership for the month of October, SARIMAX.	24
Figure 16. Summary of predictions (SARIMA, SARIMAX, NB, observed values).	26
Figure 17. Prediction of the SARIMAX model for the month of November.	17
Figure 18. MCMC plots.	27
Figure 19. Observed and fitted daily ridership between April and September, State Space Model.	29
Figure 20. Predicted ridership for the month of October, State Space Model.	29

Chapter 1: Introduction

Background

There has been considerable growth in bicycle ridership over the past few decades – the total number of bike trips in the U.S. more than tripled between 1977 and 2009, while the bike share of total trips almost doubled, rising from 0.6% to 1.0% (Pucher, et al., 2011). The rising bicycle ridership suggests that careful planning and appropriate investment in bicycle infrastructure are necessary in order to accommodate increasing bicycle travel. Since accurate ridership prediction will be indispensable for those planning and investment decisions, it is important to understand various ridership prediction models which are capable of identifying significant factors related to the motivation for people to bicycle. This study examines the forecasting capability of various econometric models for bicycle counts.

In past studies, bicycle counts data have been primarily treated as cross-sectional. As is often the case with traffic count analysis, Poisson and Negative Binomial (NB) models have been used to analyze cross-sectional and time-series count data of bicycle riders. For example, Nasal and Miranda-Moreno (2011) estimated both a Poisson regression model and a Negative Binomial regression model using hourly ridership count and found that recreational facilities are more affected by weather conditions than utilitarian facilities in general, and non-recreational facilities are more sensitive to weather conditions on weekends. However, in their analysis, an independent and identically distributed error term was assumed – an assumption unlikely to hold in time series data. In addition, the data was treated as cross-sectional without checking for stationarity, ignoring thus the dynamics of cycling ridership, trends, seasonality, or within-week cycles. Modeling time series count data using Poisson regression or NB regression may result in inefficient parameter estimates as the time series data are often serially correlated. An appropriate prediction model is necessary in order to account for the serial correlation that exists in bicycle ridership. The key objective of this study therefore is to examine the effects of weather on bicycle ridership using cross-sectional count models and time-series models, and to examine their relative

performances. I will discuss the strengths and weaknesses of these models and then examine the performance of alternative models, namely the so-called “state space” models. The rest of this chapter is organized as follows. In the next section, I will describe the econometric models I used for the analysis. I will then explore in detail the bicycle count data collected using an inductive loop bicycle counter in Portland, Oregon. Finally, I will present our results and discuss the limitations of this study and the directions for the future research.

Chapter 2: Methodology

Models for Count Data

The models for integer event count data are well developed and applied in various fields. Cross-sectional count data are often modeled using a Poisson regression model or a Negative Binomial regression if over-dispersion is present. When a Poisson model is appropriate for an outcome Y , the probabilities of observing any specific count, y , are given by the formula:

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where λ is known as the population rate parameter (which usually needs to be estimated). A Poisson random variable Y has expectation $E(Y) = \lambda$, and variance $\text{var}(Y) = \lambda$. The fact that the expectation and variance coincide provides a quick check on whether a Poisson model might be appropriate for a sample of observations. The Poisson regression is simply an extension of the Poisson model, where parameters for covariates are estimated to describe the relationship between covariates (e.g., weather variables) and responses covariates (e.g., ridership counts). This relationship can be parameterized by a log-linear model,

$$\log(\lambda(X_1, X_2, \dots, X_n)) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

$$E(\lambda) = \text{var}(\lambda) = \mu$$

The Poisson regression has the severe limitation that the variance of the counts as well as that of explanatory variables is equal to the mean. If this fails to be true (i.e., count data shows “over-dispersion”), the estimates of the coefficients can still be consistent, but the standard errors can

be biased. The Negative Binomial regression (NB) is an extension of the Poisson regression that can account for over-dispersion among variables. More specifically, the NB model will estimate an additional parameter, θ (dispersion parameter), for the variance:

$$var(\lambda) = \mu + \theta\mu^2$$

The Negative Binomial regression is often more appropriate than the Poisson regression when the data is observational as I would not expect that every variable that contributes to the rates of events is measured, and so there will always be residual variation. I will therefore use a Negative Binomial regression for our analysis, which is specified as follows:

$$\log(\mu_{m,d}) = \alpha + \beta X_{m,d} + \gamma + \epsilon_{m,d},$$

where

m, d = indexes representing the month and day of the week, respectively;

$\mu_{m,d}$ = mean number of bicycle counts during a specific month m , and day of the week d ;

$X_{m,d}$ = weather conditions (precipitation, temperature, and dew point depression)

γ = a dummy variable for weekend days

$\epsilon_{m,d}$ = independent error term

Models for Time Series Data

The models for continuous autoregressive time series data were introduced by Box and Jenkins (1976), and have been applied in different fields such as finance and economics. The Box and Jenkins model such as the seasonal autoregressive integrated moving average (SARIMA) model is capable of taking into account the trend and seasonality (and hence the serial correlation)

normally present in time series data. An extension to this model was proposed by Box and Tiao (1975) by adding the ability to examine the effects of various regressors (or/and interventions) as explanatory variables along with the trend and seasonal components. This model, called Auto Regressive Integrated Moving Average with Exogenous Input (ARIMAX), can be expressed as follows:

$$Y_t = \beta X_t + E_t$$

where Y_t = the dependent variable for a particular time t ;
 X_t = the deterministic effects of independent variables;
 E_t = the stochastic variation;

The stochastic variation can be represented by an ARIMA model as ARIMA (p,d,q) (for a non-seasonal time series), or a SARIMA model (for a seasonal time series) denoted as SARIMA (p,d,q)×(P,D,Q)_s. In these models, p is the order of the non-seasonal autoregressive (AR) process, P is the order of the seasonal AR process, d is the order of the non-seasonal difference, D is the order of the seasonal difference, q is the order of the non-seasonal moving average (MA) process, Q is the order of the seasonal MA process and the subscript s is the length of seasonality (Box et al., 2008).

State Space Models

Recently there has been an increasing interest in the application of state space models in time series analysis. One major drawback of ARIMA models is the requirement of stationarity. The analysis on nonstationary time series requires a preliminary transformation of the data to get stationarity. The stationarity requirement becomes problematic in two ways: (1) stationarity may

be difficult to attain through differencing; and (2) the coefficient estimates of regressors, if introduced in the model, are difficult to interpret. Multiple differencing is sometimes required for the series to exhibit stationarity, and additionally, the criterion for the series to be stationary is arbitrarily based on statistical tests such as Dickey-Fuller test. Inclusion of explanatory variables can provide a greater explanatory power to the model; however, since most series require differencing, the interpretation of the physical meanings of coefficient estimates of differenced variables can be arbitrary. The state space models solve all of those problems.

State space models allow a direct analysis on data that exhibits non-stationarity. In the state space model, the development of the system, y_t , is determined by unobserved series of states, θ_t , whose relation with y_t is specified by model. In general, a dynamic linear state space model is written as:

$$\text{Observation Equation : } Y_t = F_t \theta_t + v_t, \quad v_t \sim N_m(0, V_t)$$

$$\text{State Equation: } \theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N_p(0, W_t)$$

The state vector, θ_t is specified by a prior distribution. For example, a Normal prior distribution for the k-dimensional state vector at $t = 0$ is $\theta_t \sim N_m(0, C_t)$. In the classical approach, the estimation of a vector of unknown parameters is done by maximum likelihood. However, I will apply the Bayesian approach as it offers a more consistent formulation of the problem (Petrís and Campagnoli, 2009). Estimation of unknown parameters is solved by computing conditional distributions of the quantities of interest given the most recent data using filtering. Filtering is the recursive steps needed to compute the densities $p(\theta_t|Y_t)$ in the state space model. In the filtering problem, it is assumed that the data arrives sequentially in time and the object of filtering is to update our knowledge of the system each time new data arrives. Filtering involves the following

steps:

- One step ahead predictive distribution for θ_t given Y_{t-1} , based on the filtering density $p(\theta_{t-1}|Y_{t-1})$:

$$p(\theta_t|Y_{t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|Y_{t-1}) d\theta_{t-1}$$

- One step ahead predictive distribution for the next observation:

$$p(y_t|Y_{t-1}) = \int p(y_t|\theta_t)p(\theta_t|Y_{t-1}) d\theta_t$$

- The posterior distribution $\pi(\theta_t|y_{1:t})$ using the prior distribution $p(\theta_t|Y_{t-1})$ and the likelihood $p(y_t|\theta_t)$:

$$p(\theta_t|Y_t) = \frac{p(y_t|\theta_t)p(\theta_t|Y_{t-1})}{p(y_t|Y_{t-1})}$$

The Kalman filter allows us to compute the predictive and filtering distributions recursively, using $\theta_0 \sim N(u_0, v_0)$ to compute $p(\theta_1|y_1)$, and proceeding recursively as new data become available.

In this study, I will apply the random walk plus noise model:

$$\begin{aligned} Y_t &= \alpha_t + v_t, & v_t &\sim N_m(0, \sigma_v^2) \\ \mu_t &= \alpha_{t-1} + w_t, & w_t &\sim N_p(0, \sigma_w^2) \end{aligned}$$

where v_t and w_t are all mutually independent and independent of α_t . The prior distribution of α_t is assumed to be d-inverse Gamma.

Chapter 3: Cyclist Count Data Analysis

Introduction

The bicycle count data was collected in the Hawthorne Bridge in Portland, Oregon using an inductive loop bicycle counter. The counter detects bicycles by monitoring changes in an electric current in sub-pavement loops of cable and is capable of distinguishing cyclists from other traffic (Nasal and Miranda-Moreno, 2011). The data was acquired between 5 am and 7 pm each day from April to November in 2010. As shown in Figure 4-1, the Hawthorne Bridge is equipped with a cycle track and carries commuters from the east side of the city into the downtown area. The City of Portland has a population of about 1.6 million with a relatively high percentage of bicycle commuting (2.6 %) compared to other large US cities (Dill& Carr, 2003).

Data Description

Figure 2 shows the average hourly counts of bicycle ridership observed at the Hawthorne Bridge. The ridership peaks twice a day, in the morning (8am) and in the evening (5pm), which indicates that the facility is used primarily for utilitarian purposes. Figure 3 shows the average daily counts and the average monthly counts of bicycle ridership. The ridership stays relatively flat on weekdays reaching its peak on Wednesday and dramatically descends on weekend, which further indicates that the bridge is used as a utilitarian facility. Figure 3 also shows the average daily counts over months. The ridership gradually increases toward warmer months, achieving its peak in August and gradually decreases afterward. A drastic drop in the ridership is observed between October and November. To illustrate the cause of this drop in the ridership, Figure 4 shows the average temperature and precipitation in Portland between April and November of 2010. As the figure shows, the temperature and precipitation change dramatically in November. In particular, the temperature decreases rapidly from October to November.

Based on the data described above, I will estimate the following four models: SARIMA, Negative Binomial, SARIMAX, and the State Space model with a random walk plus noise to compare their fits and forecasting accuracy. As described above, the month of November in Portland is characterized by much colder weather with a high level of precipitation and much lower ridership (see Figures 5 and 6); therefore I will first use data from April to September to make out-of-sample predictions for October. The issue associated with the presence of inclement weather in November will be further discussed later on.

Figure 1 Hawthorne bridge in Portland, OR



Figure 2 Average cyclists per hour

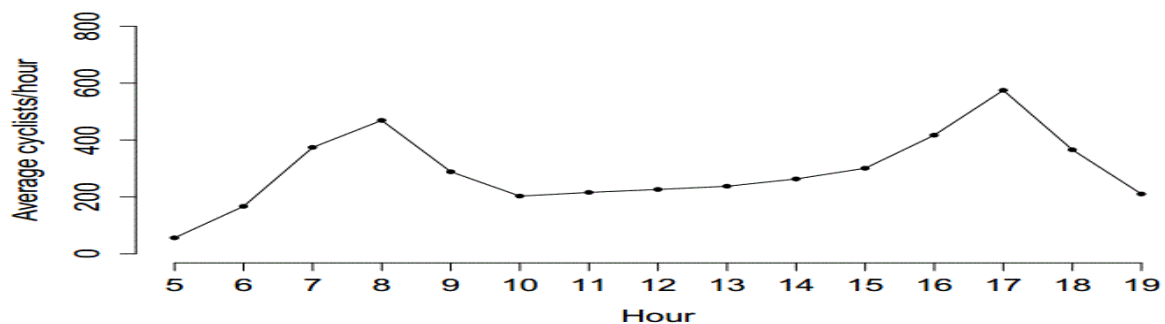


Figure 3 Average cyclists per hour per day, per day of the week (left) and month (right)

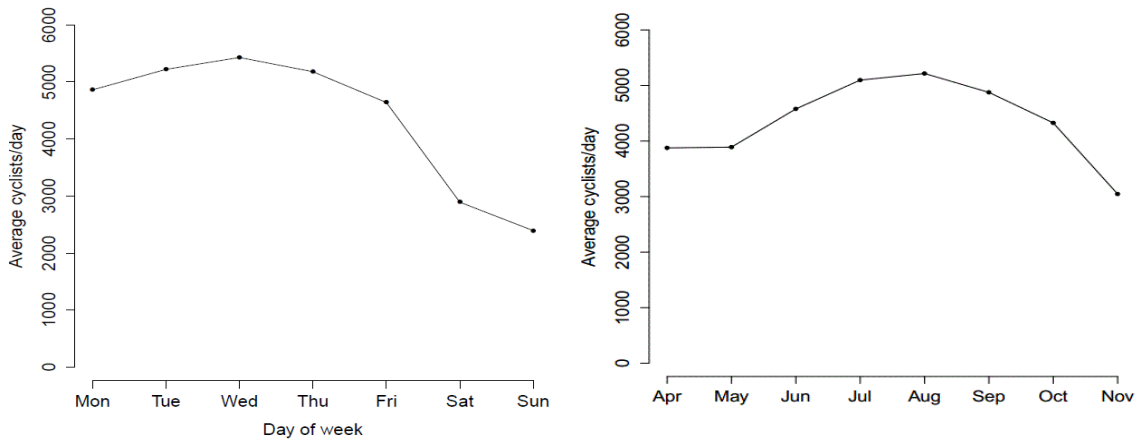


Figure 4 Average monthly precipitation (left) and temperature (right)

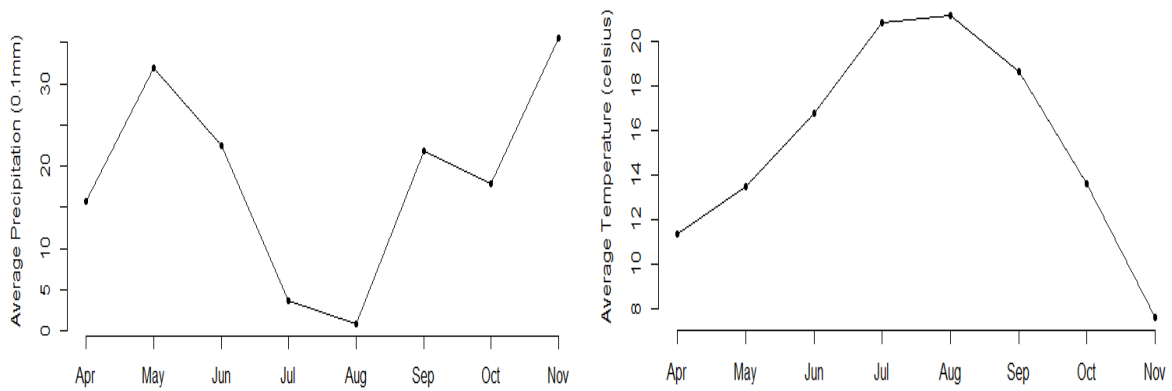


Figure 5 Average ridership per day, during the month of November

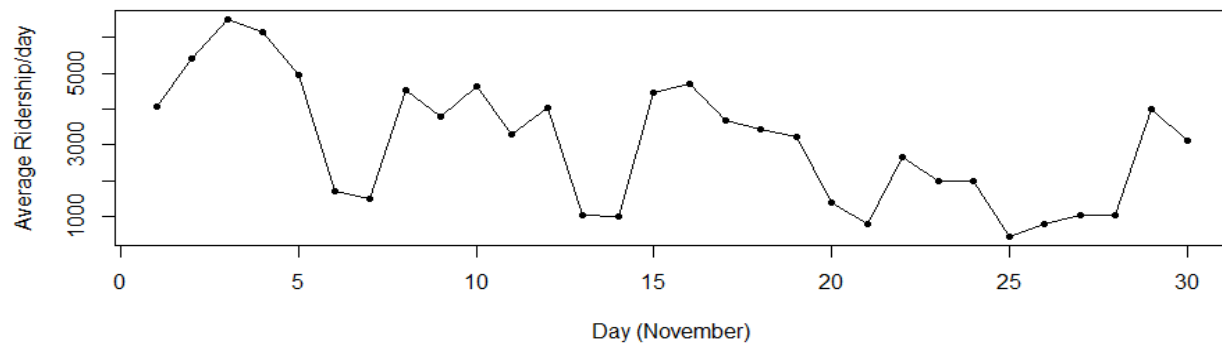
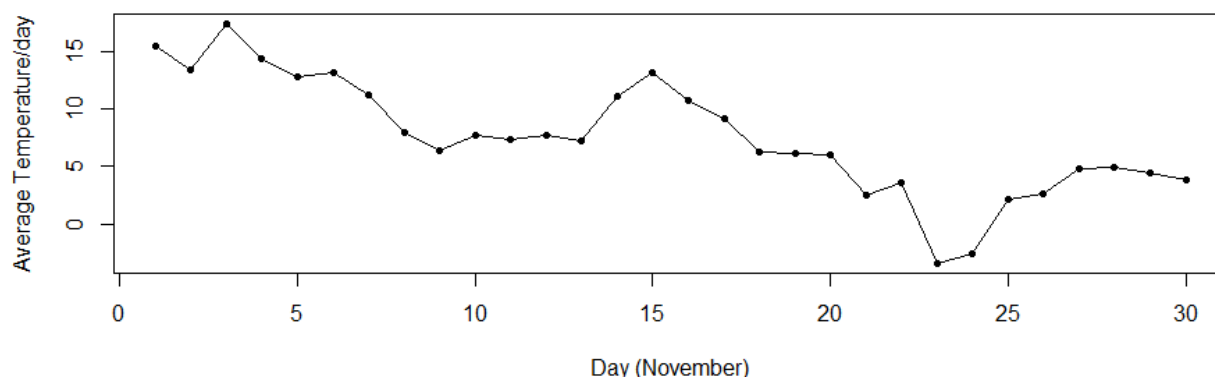


Figure 6 Average temperature per day, during the month of November



Stationarity

I next examined the ridership as a time-series object. Figure 7 shows a plot of the average cyclist counts per day and Figure 8 shows a plot of autocorrelation function (ACF). The ACF plot exhibits a significant auto-correlation at lag 1 and a seasonal auto-correlation with neighboring effects at lag 7, i.e., a group of autocorrelations at the lag 6, 7, and 8 appear repeatedly every 7 periods. An augmented Dickey-Fuller Test of lag order 10 shows the presence of a unit root, i.e. the series is non-stationary. Figure 4-9 show plots of the series and ACF of the average ridership after differencing. After differencing, the Augmented Dickey-Fuller Test rejected the presence of a unit-root at the 1% level; meaning that the series is now stationary. There still appears a strong seasonal autocorrelation at period 7 as shown in the ACF plot in Figure 9.

Figure 7 Time series of daily ridership

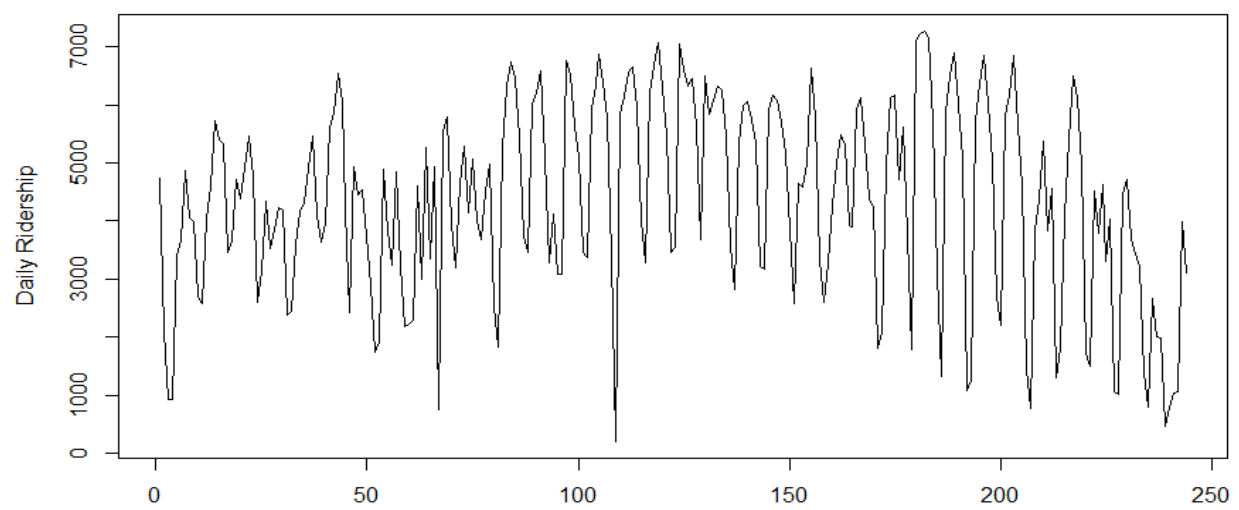


Figure 8 Autocorrelation function (ACF)

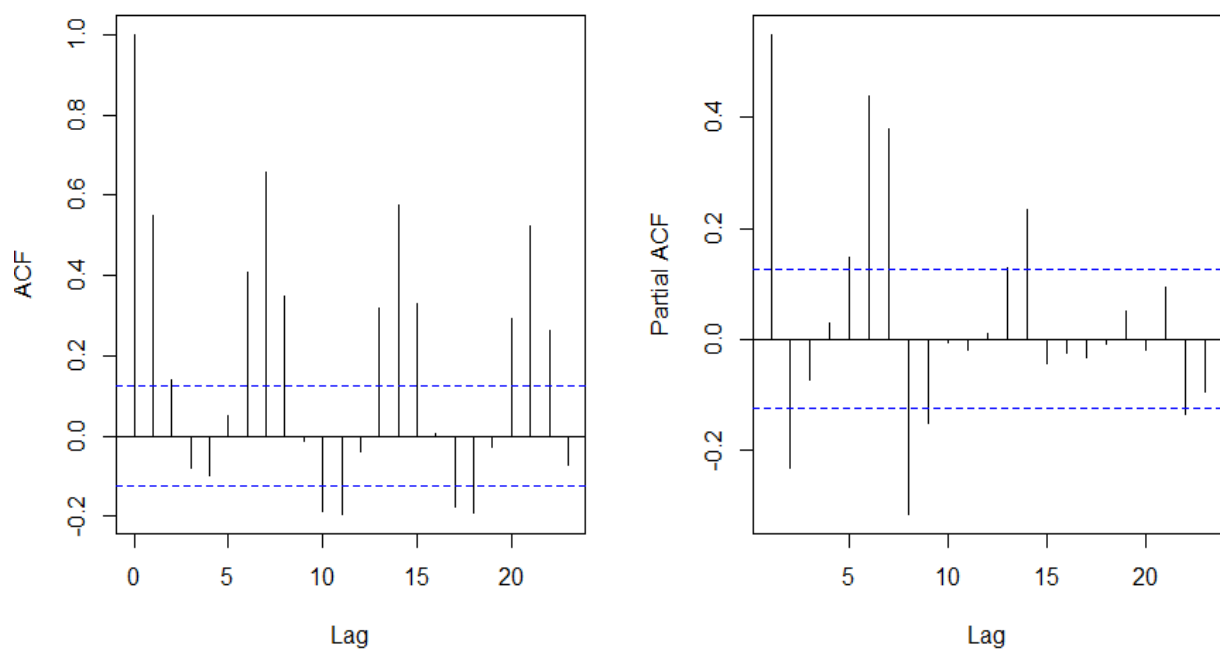
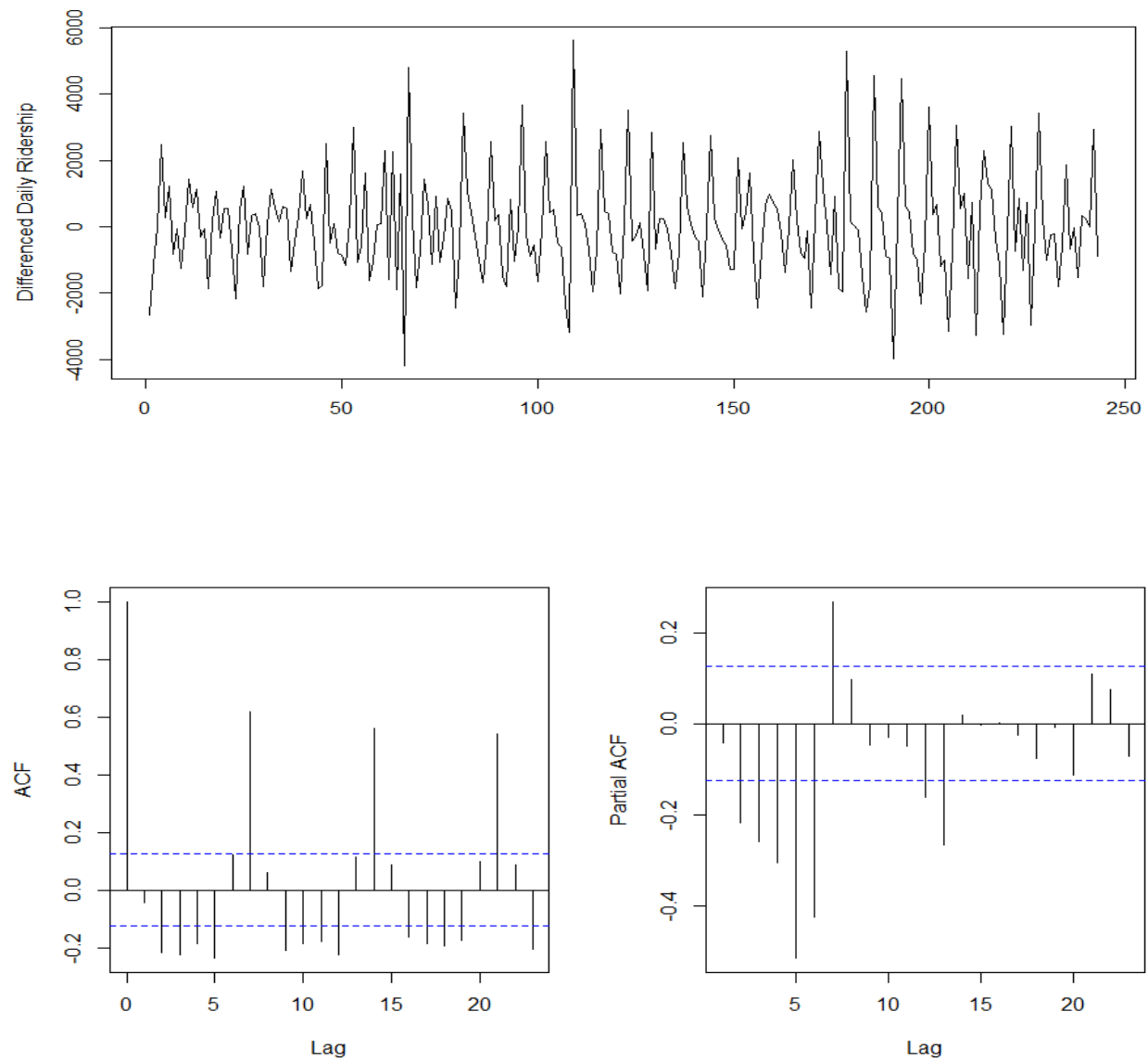


Figure 9 Average ridership and ACF after differencing



Chapter 4: Results

SARIMA

I first estimated SARIMA model and produced predicted values for the ridership in October. After study of ACF and PACF of the series and with the AIC criterion, I identify the appropriate model as ARIMA (5,1,2)(0,0,1)₇. The model was successful in removing auto-correlation as the Ljung-Box test gives a chi-square value of 10.3 with a p-value of 0.42. Figure 10 shows the observed daily ridership between April and September and the fitted values from the model. Figure 11 shows the observed ridership in October and the forecast from the SARIMA model. As can be seen in Figure 11, the forecast from SARIMA tend to overpredict the ridership, especially for weekends. The root mean square error is 1048.56, and the root mean forecast error is 2543.5.

Figure 10 Observed and fitted daily ridership between April and September, SARIMA

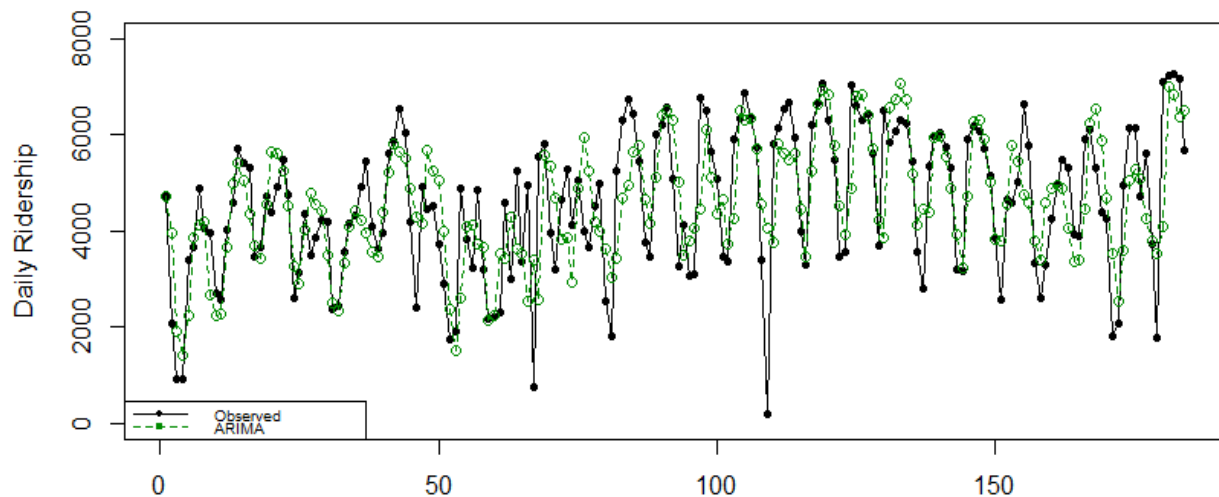
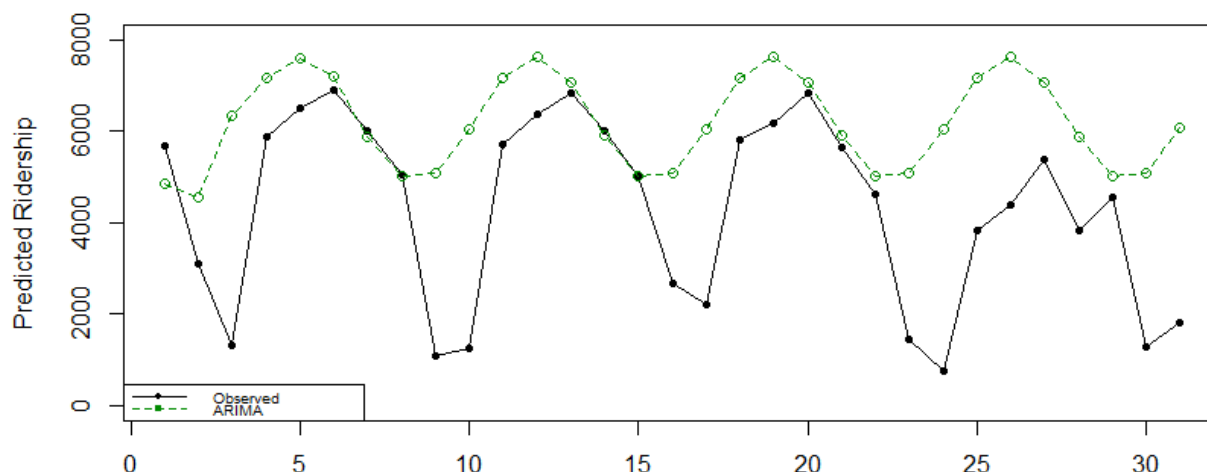


Figure 11 Predicted ridership for the month of October, SARIMA



Negative Binomial

I next estimated a NB model with temperature, squared temperature, precipitation, weekend dummy, and dew point depression as explanatory variables. Figure 12 shows the fitted values, and Figure 13 shows the forecast from the NB model. The weather variables seem to explain the ridership very well. The prediction from NB model is much better than that from the ARIMA model, though the forecast tends to underpredict the ridership for weekdays and overpredict for weekends. The root mean square error is 805.21, and the root mean forecast error is 993.44.

SARIMAX

After the study of ACF and PACF, I specified the model as a SARIMAX (0,1,1)(0,0,1)₇. The model was successful in removing auto-correlation as the Ljung-Box test gives a chi-square value of 7.22 with a p-value of 0.70. Figure 14 shows the fitted values, and Figure 15 shows predicted ridership in October. As is obvious from Figure 15 and Table 1, SARIMAX outperforms both the SARIMA model and the NB model in terms of both model fit and forecasting accuracy. The root mean square error is 760.2, and the root mean forecast error is

842.29.

Figure 12 Observed and fitted daily ridership between April and September, NB

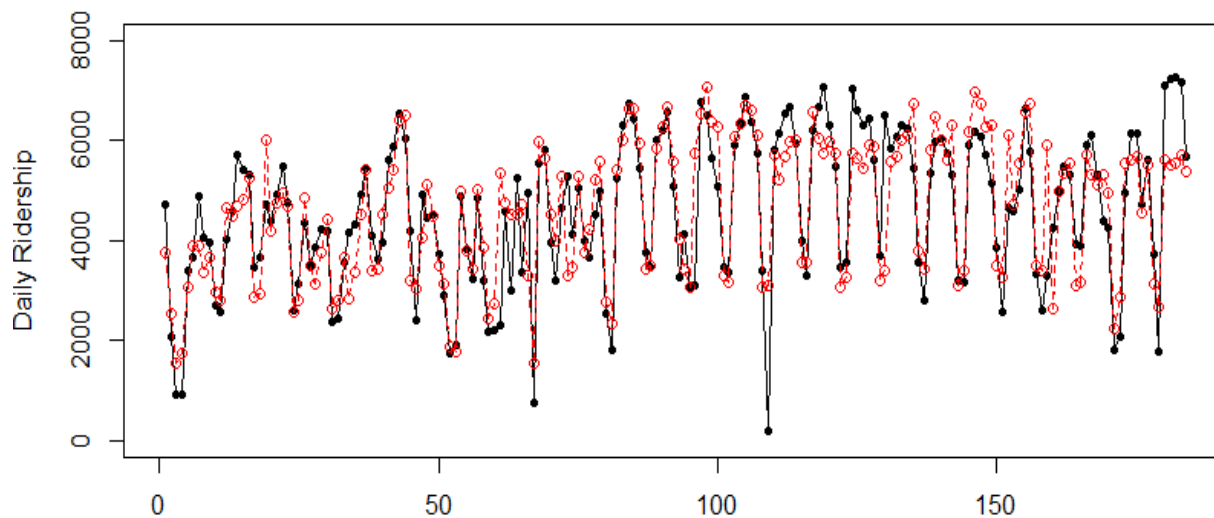


Figure 13 Predicted ridership for the month of October, NB

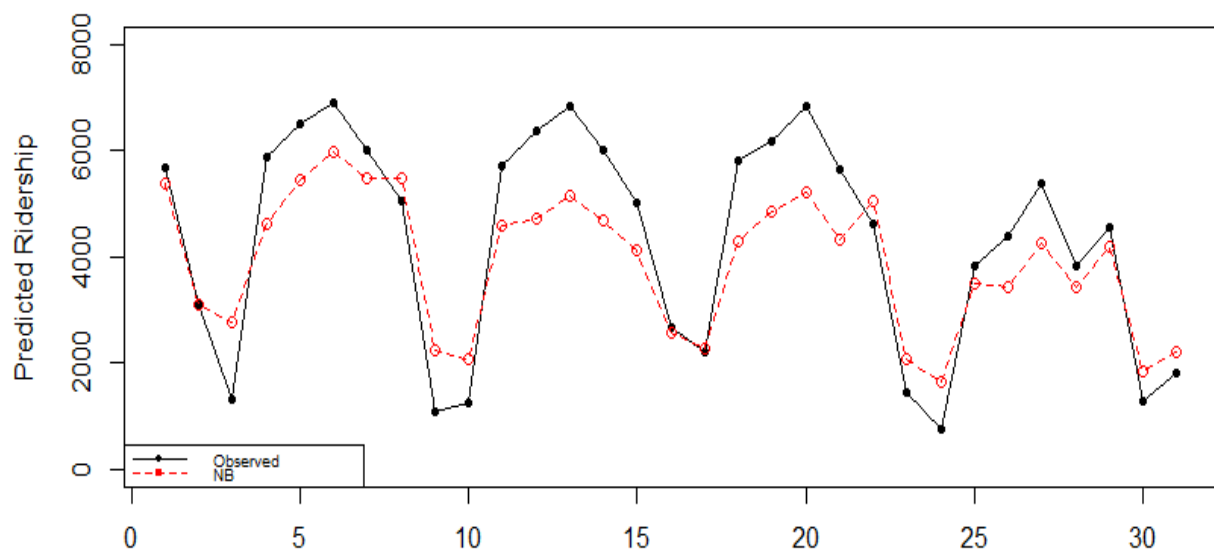


Figure 14 Observed and fitted daily ridership between April and September, SARIMAX

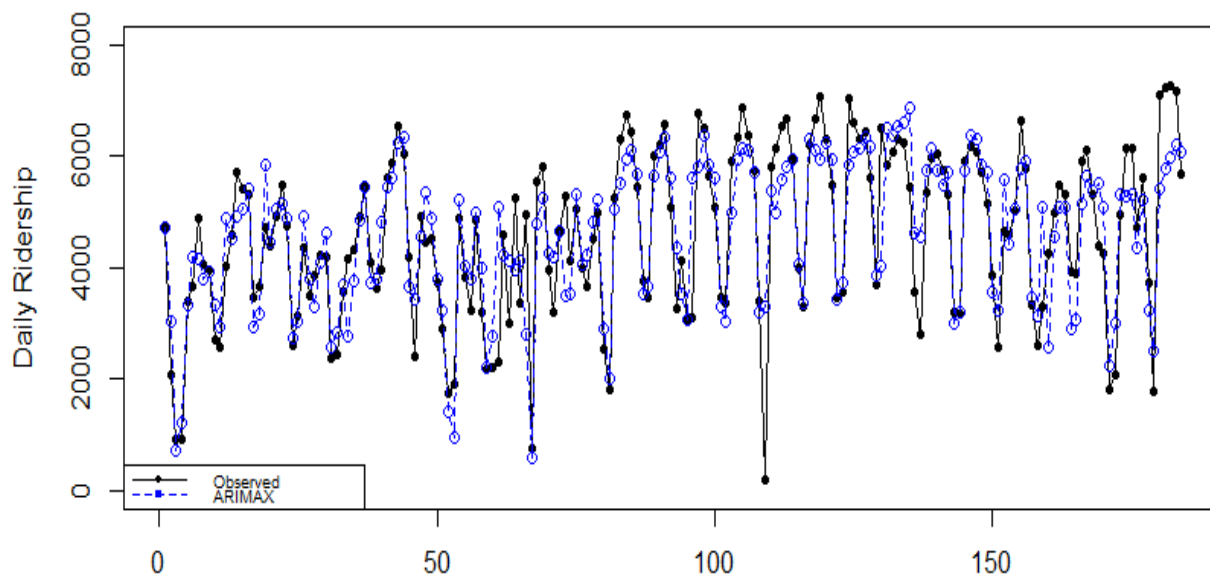


Figure 15 Predicted ridership for the month of October, SARIMAX

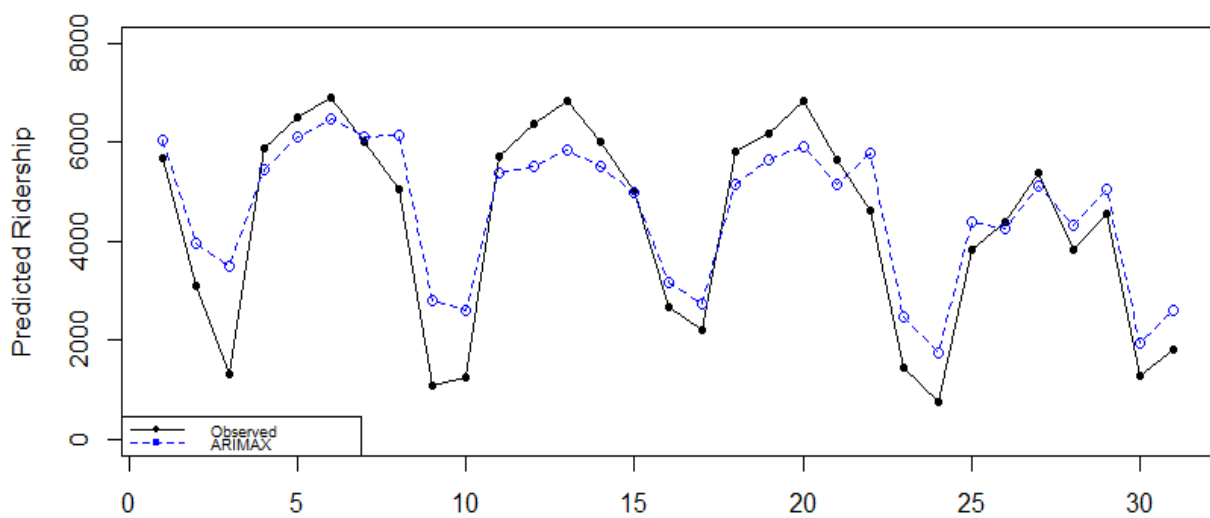
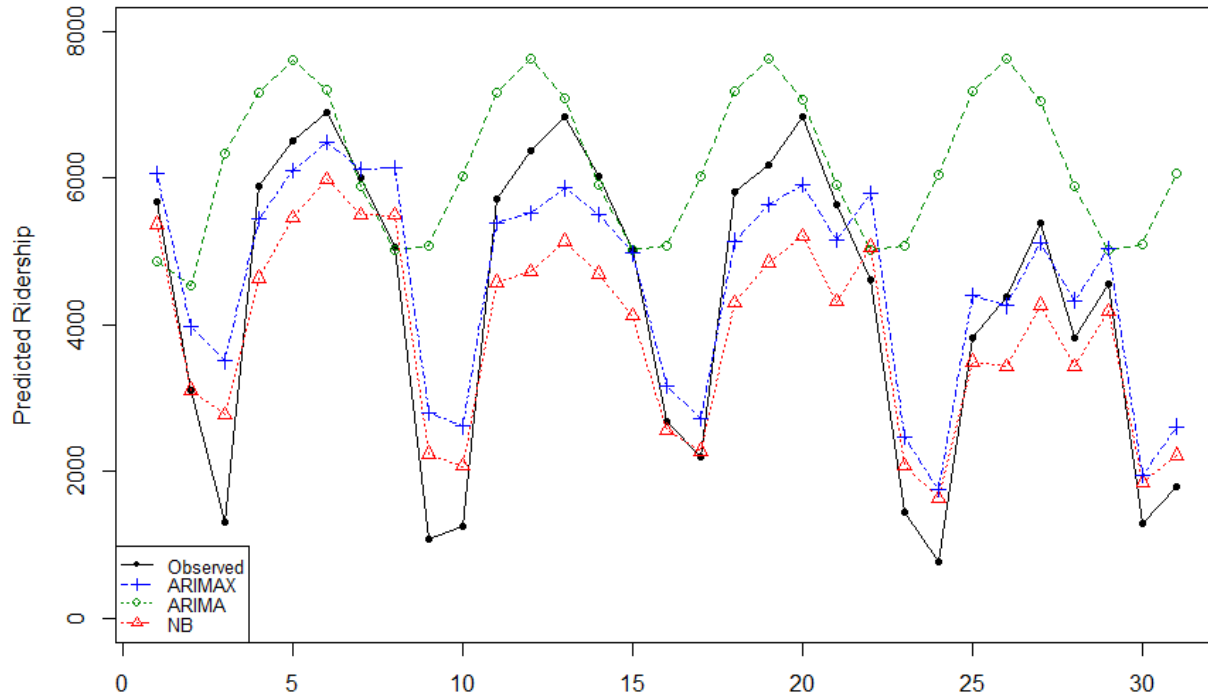


Table 1 Coefficients Estimates

	ARIMA(5,1,2)(0,0,1) ₇		ARIMAX(0,1,1)(0,0,1) ₇		Negative Binomial	
	Coef	t - stat	Coef	t - stat	Coef	t - stat
				-		
MA1	-1.219	-44.663	-0.869	14.480	-	-
MA2	0.978	24.037	-	-	-	-
Seasonal MA1	0.206	2.772	0.208	2.893	-	-
AR1	0.648	8.503	-	-	-	-
AR2	-0.739	-8.530	-	-	-	-
AR3	-0.231	-2.285	-	-	-	-
AR4	-0.191	-2.176	-	-	-	-
AR5	-0.233	-3.140	-	-	-	-
Precipitation	-	-	-4.19	-3.051	-0.002	-3.54
Temperature	-	-	477.31	6.101	0.12	6.37
Temperature square	-	-	-10.81	-5.482	-0.003	-5.15
Weekend	-	-	-2285.77	-16.09	-0.60	-15.50
dpd	-	-	121.57	3.826	0.032	3.83
Log-likelihood	-1536.41	-	-1474.90	-	-1534.55	-
AIC	3090.82	-	2965.80	-	3083.10	-
Root mean square error	1048.56	-	760.20	-	805.21	-
Root mean square forecast error	2543.5	-	842.29	-	993.44	-

Figure 16 Summary of predictions (green: SARIMA; blue: SARIMAX; red: NB; black: observed values)

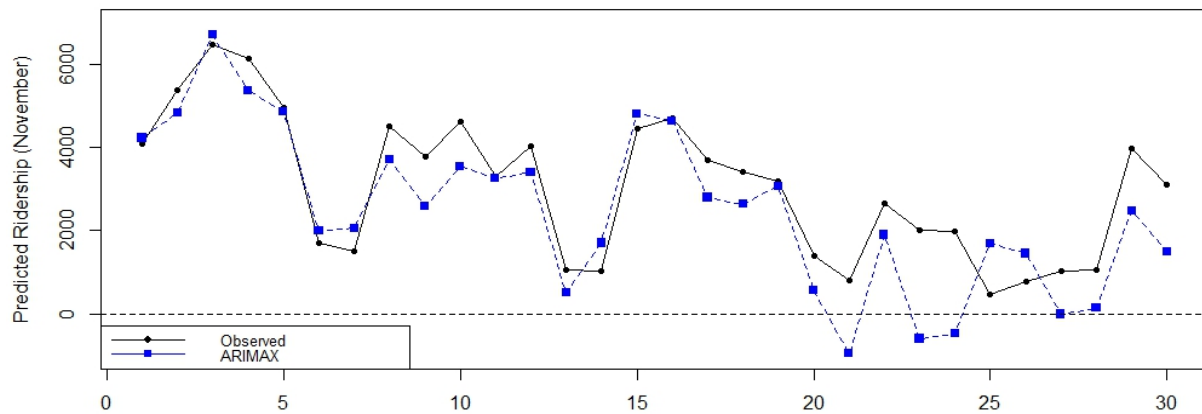


The performance of the fitted models was investigated using the relative forecast error. The result of this study suggests that the SARIMAX model has the best prediction power. This is due to the fact that SARIMAX is able to take into account both serial correlation and exogenous factors influencing the ridership. However, my finding has some limitations – when modeling nonnegative integer-valued count data such as traffic count, Box and Jenkins models may be inappropriate because of the normality assumption on which the ARIMA model is based.

To illustrate this problem, I forecasted the ridership in November using the same models. As discussed above, the weather in the latter part of November is characterized by much colder temperatures and higher precipitation than any other month included in the data. As a result, some of the predicted ridership for November from the SARIMAX model becomes negative as

shown in Figure 17. This suggests that although SARIMAX model has the highest forecasting power as well as the best fit, it is less appropriate as ridership count becomes smaller. In those cases, the NB model performs reasonably well without the problem of negative forecast values.

Figure 17 Prediction of the SARIMAX model for the month of November



My conclusion above suggests a doubt on the conventional statistical analysis on time-series count data. Poisson regression and Negative Binomial regression guarantees the integer forecast values, but ignores autocorrelation in the series. Although ARIMA-type models have been popular method to analyze time-series data, I found that using ARIMA-type models has at least three major problems when applied on count data:

1. Stationality requirement (pre-transformation of data)
2. Difficulty in making inferences on explanatory variables when data is transformed
3. Negativity of forecast values resulting from the normality assumption

For example, in this study I needed to take a first difference in order to get stationarity in the data. As a result, the coefficient estimates of weather variables became difficult to interpret. The negative sign of precipitation in our estimated model means that a positive difference of

precipitation between days has a negative impact on the ridership; however, the positive difference can mean either increase in precipitation in the current period or a decrease in the last period. State Space models (or more specifically, dynamic regressions) solve all of the problems listed above as they do not require stationarity in data.

State Space Model

The estimation of State Space model involved estimation of the two unknown variances, dV and dW . Figure 18 displays the MCMC output. The ergodic means seem to be stable in the last part of the plot. Figure 19 shows the observed daily ridership between April and September and the fitted values from the State Space model, and Figure 20 shows the predicted ridership in October. Since the model follows a random walk with noise, the forecast values are simply the filtered estimate from the last day in September.

Figure 18 MCMC plots

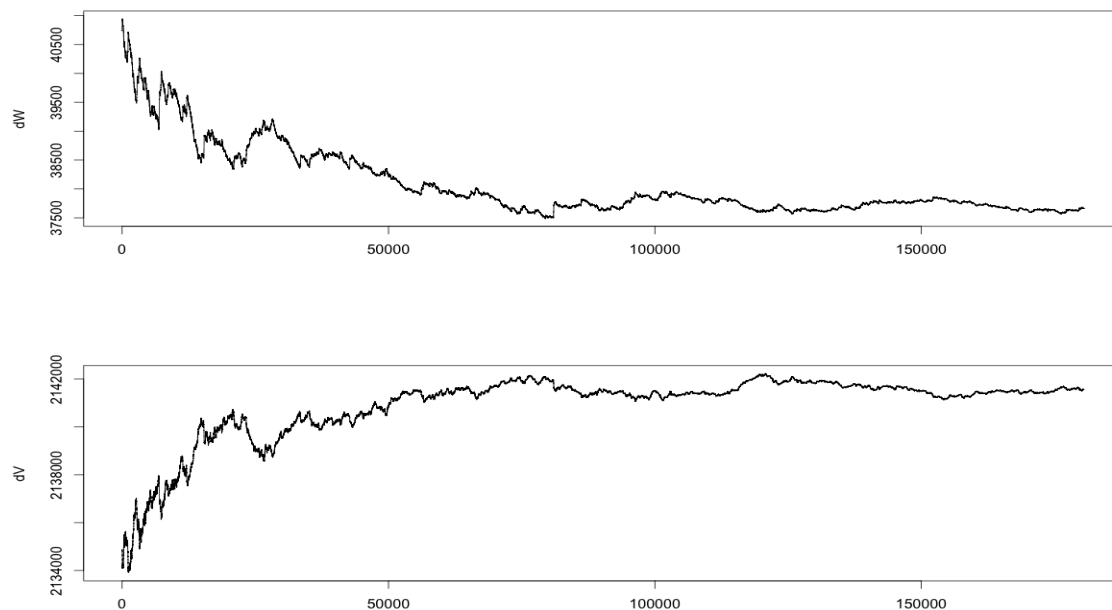


Figure 19 Observed and fitted daily ridership between April and September, State Space Model

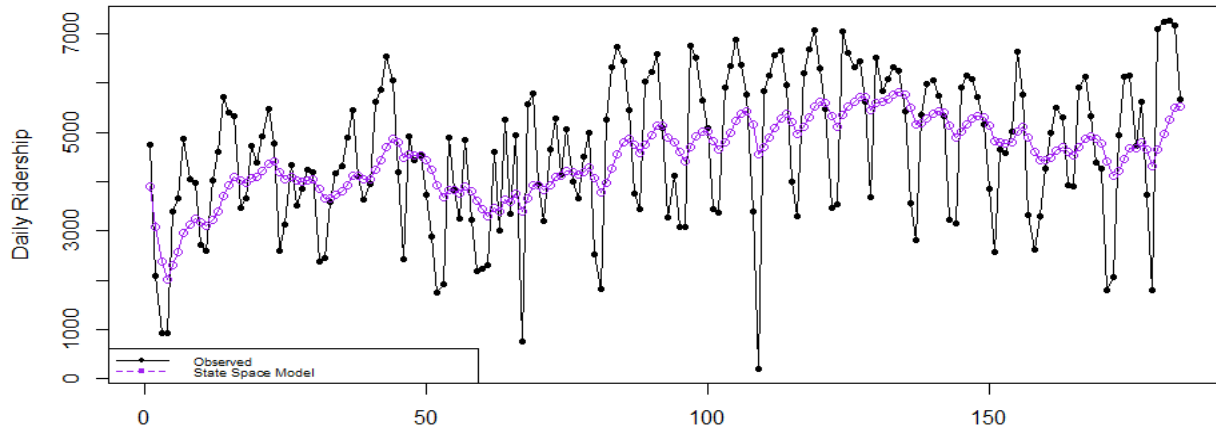
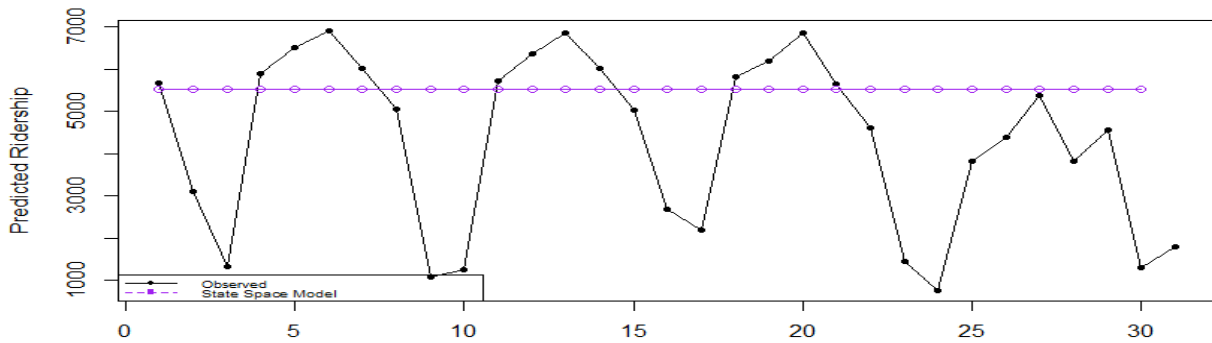


Figure 20 Predicted ridership for the month of October, State Space Model



State space models may include explanatory variables as in the following specification.

$$Y_t = X\theta_t + v_t, v_t \sim \tau(0, \sigma_v^2)$$

$$\theta_t = Z\theta_{t-1} + w_t, w_t \sim \tau(0, \sigma_w^2)$$

$$\text{where } Y_t = (Y_{1,t}, \dots, Y_{n,t})', \theta_t = (\beta_1, \dots, \beta_k)' \sim \varphi(l, m),$$

$$X = \begin{bmatrix} 1 & f_2(x_1) & \dots & f_k(x_1) \\ \vdots & \vdots & & \vdots \\ 1 & f_2(x_n) & \dots & f_k(x_n) \end{bmatrix}$$

$$Z = \text{diag}(\delta_1, \dots, \delta_k)$$

$$\sigma_v^2 = \text{diag}(\phi_1^{-1}, \dots, \phi_n^{-1})$$

$$\sigma_w^2 = \text{diag}(\omega_1^{-1}, \dots, \omega_n^{-1})$$

The model can be extended in order to accommodate count data. With alternative distributional assumptions of v_t , w_t , and θ_t , I will be able to estimate a dynamic Poisson regression and dynamic Negative Binomial regression, which (1) does not require stationarity, (2) produces integer-forecast values, and (3) produces coefficient estimates that are much more intuitive to interpret.

Chapter 4: Discussion/Conclusion

In this project I have analyzed different econometric methods for making statistical inference on the demand for nonmotorized transportation, with a focus on cycling. I investigated the factors that influence people's decision to use bicycles as a means of transportation. In particular, the methodological framework I applied in this study looked at seasonality and time-dependency of bicycle ridership as well as effects of weather conditions on ridership. Since the time series data was only one-year long, I did not (could not) investigate possible trends in ridership over years. Given the availability of data, an examination of long term ridership growth would be an interesting research area to explore as it is important for planning for growth in traffic demand and consequentially for infrastructure investments. In this research, I focused on short-run prediction accuracy of several ridership prediction models, including the Negative Binomial regression and time-series models such as ARIMA and ARIMAX (e.g., Nihan & Holmesland, 1980 and Houston & Richardson, 2002). The result of this study suggests that the SARIMAX model has the best prediction as that model is able to take into account both serial correlation and exogenous factors influencing the ridership. However, I also found that Box and Jenkins models may be inappropriate for count data because of the normality assumption on which the general ARIMA models are based, especially when ridership counts are small. This problem is especially relevant for months with weather conditions that discourage the use of cycling. The special family of State Space models has the potential to address the non-normality issues. Although estimation of State Space models is computationally intensive, software implementations have increased with the progress in computing capabilities and the modernization of computer software. Many well-known statistical and econometric software packages currently have options for the use of state space methods (Durbin & Koopman, 2012). However, most of them

are currently only capable of estimating the basic state space model that predicts averages – identical to the model I used in this study. The estimation of more generalized state space models requires much more intensive computer programming, and therefore I leave this task to future research.

References

- Akar, G. and K. Clifton. 2009. The influence of individual perceptions and bicycle infrastructure on the decision to bike. *Transportation Research Record* 2140:165–172.
- Balsas, C. J. L. 2002. New Directions for bicycle and pedestrian planning and education in the U.S. *Planning Practice and Research* 17(1):91-105.
- Balsas, C. J. L. 2003. Sustainable transportation planning on college campuses. *Transport Policy* 10:35-40.
- Bekkum, J. E. and J. M. Williams. 2011. Cycle commuting and perceptions of barriers: stages of change, gender and occupation. *Health Education* 111 (6):476-497.
- Bishop, M. 1962. *A History of Cornell*. Ithaca, NY: Cornell University Press.
- Blanco, H., M. Alberti, A. Forsyth, K.J. Krizek, D.A. Rodríguez, E. Talen, and C. Ellis. 2009. Hot, congested, crowded and diverse: Emerging research agendas in planning. *Progress in Planning* 71 (4):153-205.
- Bonham, J. and B. Koth. 2010. Universities and the cycling culture. *Transportation Research Part D* 15:94-102.
- Brownson, R. C., T. K. Boehmer, and D. A. Luke. 2005. Declining Rates of Physical Activity in the United States: What Are the Contributors? *Annu.Rev. Public Health* 2005. 26:421–43.
- Campus Planning Office. 2008. *Cornell Master Plan*. <http://www.masterplan.cornell.edu> (accessed March 15, 2012).
- Chaplin, L. E. 1998. *Campus Biking: Challenges and Strategies*. http://www.bike.cornell.edu/pdfs/Report/Campus_Biking.pdf (accessed March 15, 2012).
- Chaplin, L. E., M. J. Adleman, P. C. T. deBoer, M. Kratzschmar, B. Lane, M. E. Moran, Y. M. Nelson, D. M. Nutter, B. Rausch, S. L. Tallant, M. Thompson, and J. Ullberg. 1991. *Bicycling on the Cornell University Campus: Final Report from the Ad Hoc Bicycle Committee*.

- http://www.bike.cornell.edu/pdfs/Bicycle%20Ad%20Hoc%20Committee%20Report_May-91.pdf
(accessed March 14, 2012).
- Cornell University 1993. Cornell Cycles: A New Call for Transportation Alternatives.
<http://www.bike.cornell.edu/pdfs/Cornell%20Cycles%202.pdf> (accessed February 20, 2012).
- Cornell University. 2007. Transportation-Focused Generic Environmental Impact Statement (t-GEIS)
Travel Survey Results: Executive Summary.
http://www.tgeisproject.org/TGEIS_Documents/TravelSurvey_ExecSummary_FINAL_19Feb07.pdf
(accessed March 13, 2012).
- Cornell University. 2008. Transportation Impact Mitigation Strategies: Draft Report.
http://www.tgeisproject.org/TGEIS_Documents/Final-Draft-TIMS-060308.pdf (accessed January 29, 2012).
- Cornell University. 2009. *Cornell Climate Action Plan*.
http://www.sustainablecampus.cornell.edu/climate/Climate_Action_Plan_Archive_2009/CA_P2009.pdf (accessed March 13, 2012).
- Cornell University. 2011a. Transportation Demand Management.
<http://www.sustainablecampus.cornell.edu/transportation/demand.cfm> (accessed January 29, 2012).
- Cornell University. 2011b. *Transportation Team*.
<http://www.sustainablecampus.cornell.edu/about/index.cfm> (accessed March 12, 2012).
- Cornell University. 2012. *Cornell Facts 2011-2012*. <http://www.cornell.edu/about/facts/cornell-facts-2011-12.pdf> (accessed March 12, 2012).
- Cornell University Survey Research Institute. 2005. Tompkins County/Cornell Employee Commuter Survey Phase I: Cornell Employees. June 2005. http://www.tompkins-co.org/itctc/parknride/Phase%20I_Summary%20Report-final.pdf (accessed March 8, 2012).

- City of Ithaca. 2012. Maps and Demographics. <http://www.ci.ithaca.ny.us/maps/index.cfm> (accessed April 2, 2012).
- Dang, M. 2007. University Changes Free Bus Pass Policy. *The Cornell Daily Sun*. April 25, 2007. <http://cornellsun.com/node/23149> (accessed January 29, 2012).
- Forsyth, A. and K. J. Krizek. 2010. Promoting Walking and Bicycling: Assessing the Evidence to Assist Planners. *Built Environment* 36 (4):429-446.
- Forsyth, A. and K. J. Krizek. 2011. Urban Design: Is there a distinctive view from the bicycle? *Journal of Urban Design* 16 (4):531-549.
- Gatersleben, B. and K. Appleton. 2007. Contemplating cycling to work: attitudes and perceptions in different stages of change. *Transportation Research Part A* 41(4):302–312.
- Heinen, E., B. Van Wee, and K. Maat. 2010. Commuting by Bicycle: An Overview of the Literature. *Transport Reviews* 30 (1):59-96.
- Husick, Tanya. 2011. Interview concerning Cornell bicycle policy and design standards. Personal correspondence.
- Jones, T., A. Chisholm, H. Harwatt, D. Horton, A. Jopson, C. Pooley, G. Schelderman, and M. Tight. 2009. Understanding Walking and Cycling: A Multi-Method Approach to Investigating Household Decision Making in Relation to Short Journeys in Urban Areas. Paper presented at the Cycling and Society Research Group Symposium, University of Bolton, September 7. http://www.lec.lancs.ac.uk/research/society_and_environment/cycling/CSRG_Symposium_paper_Tim_Jones_Oxford_Brookes.pdf (accessed January 5, 2012).
- Krizek, K. J., A. Forsyth, and L. Baum. 2009. Walking and Cycling International Literature Review. <http://www.planethealthcymru.org/sitesplus/documents/886/Walking-and-cycling-international-literature-review%20%282009%29.pdf> (accessed December 30, 2011).
- Krizek, K., A. Forsyth, and A. W. Agrawal. 2010. *PABS Users Guide*. San José, CA: Mineta Transportation Institute. <http://transweb.sjsu.edu/project/2907.html> (accessed October 24, 2011).

- League of American Bicyclists. 2010. *Highlights of the 2009 National Household Travel Survey*.
<http://www.bikeleague.org/resources/reports/pdfs/nhts09.pdf> (accessed March 15, 2012).
- Litman, T. 2003. Integrating Public Health Objectives in Transportation Decision-Making. *American Journal of Health Promotion* 18 (1):103–108.
- Pope, N. B. 2010. An Examination of the Socio-Demographic Characteristics and Perceptions of Cycling among Students at Georgia State University. *Public Health Theses*. Paper 89.
- Pucher J., J. Dill and S. Handy. 2010. Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive Medicine* 50:S106-S25.
- Pucher, J., R. Buehler and M. Seinen. 2011. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A* 45:451–475.
- Rabl, A., and A. de Nazelle. 2012. Benefits of shift from car to active transport, *Transport Policy* 19 (1):121-131.
- Shannon, T., B. Giles-Corti, T. Pikora, M. Bulsara, T. Shilton, and F. Bull. 2006. Active commuting in a university setting: assessing commuting habits and potential for modal change. *Transport Policy* 13, 242-253.
- Stinson, M. A. and C. R. Bhat. 2004. Frequency of Bicycle Commuting: Internet-Based Survey Analysis. *Transportation Research Record* 1878:122–130.
- The Ithaca Journal. 1991. Turn-of-the-century Ithacans shared present-day vision: Bike paths. The Ithaca Journal. November 2, 1991.
- Toor, W. and S. Havlick. 2004. *Transportation and Sustainable Campus Communities: Issues, Examples, Solutions*. Washington, DC: Island Press.
- Town of Ithaca. 2008. *Transportation-focused DRAFT Generic Environmental Impact Statement (t-DGEIS) Volume I*. http://www.tgeisproject.org/TGEIS_Documents/t-DGEIS%20Vol%20I_July%202008.pdf (accessed March 12, 2012).