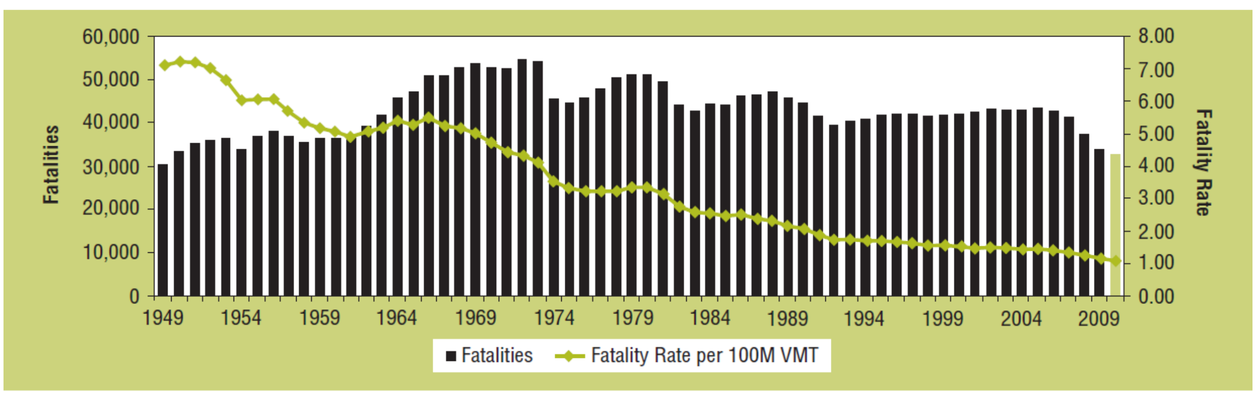
Pius Wong
Time Series Analysis, Spring 2012; Prof. Dragan Djurdjanovic

"Modeling and Prediction of Motor Vehicle Crash Counts over Time in the U.S."

1. Introduction

1. Motivation

Motor vehicle crashes on US roadways have been a serious problem for decades, consistently causing death, injury, and financial losses each year. For example, crash-related fatalities per year have fluctuated between 30-55 thousand since 1950, even though the normalized fatality rate per mile traveled has decreased (Fig. 1). Mileage has increased, working against the reduced fatality rate.



1949–1974: National Center for Health Statistics, HEW, and State Accident Summaries (Adjusted to 30-Day Traffic Deaths by NHTSA)
FARS 1975–2009 (Final) 2010 Annual Report File (ARF); Vehicle Miles Traveled (VMT): Federal Highway Administration.

Figure 1. Crash-related fatalities and fatality rates per 100 million vehicle miles traveled (VMT), 1948-2010. Reproduced from US Department of Transportation report [1].

More locally in Texas alone, in 2010, crashes caused the deaths of 3023 people and an estimated $20 billion dollars in economic losses [2]. The economic losses include property damage and productivity losses due to lost work.

Because of the harmful consequences of crashes, government and industry are interested in better predicting, preventing, and responding to them. Models of crash occurrence can help them do that. With good crash prediction, local police and emergency services can be preemptively dispatched. At the larger regional level, governments, insurance, and shipping companies can better budget for expected crash numbers. Finally, models of crash counts may help identify causes of crashes, as well as help flag major changes in crash patterns.

The purpose of this project is to examine the use of time-series models for crash counts. Numerous factors contribute to individual crashes, including alcohol use, driver age, speed, traffic, location, lighting, road characteristics, date, time, and vehicle characteristics. At a broader level, regional crash counts similarly have many influences, including alcohol laws, driver age limits, speed limits, climate, time period, population, population density, car ownership rates, road infrastructure, and police presence. Instead of incorporating all these factors into one model, a simpler ARMA model could be an alternative. This project attempts to find when ARMA modeling could be useful.

2. Data Set

Government agencies are required to record crash counts due to federal law. By legal definition, a “crash” involves death, injury, or property damage of $1000 or more, and only these severe crashes are reported. From these databases, I obtained daily crash count data sets for the states of Texas [3] and Illinois [4] from 2004-2010 and the city of Austin [5] from 2007-2010 (Fig. 2).

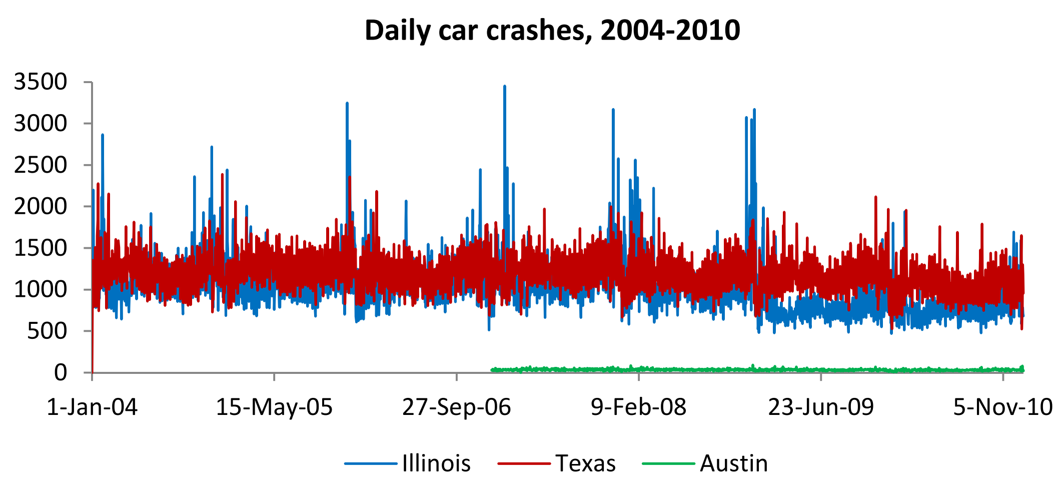


Figure 2. Raw daily crash counts for Texas, Illinois, and Austin.

These data sets were chosen because of personal and local interest, availability and accessibility, and to allow comparison across different regions. As shown, in this time period Texas and Illinois both have similar average crash counts of about 1100 crashes/day, while the city of Austin has two orders of magnitude less at about 33 crashes/day. Texas and

Illinois differ greatly in geography and climate. While Texas has the greater population, Illinois has the greater population density, which may contribute to why their average crash counts are similar. Spikes are more apparent for Illinois, and a related question is whether ARMA models can predict these, or if they can explain any seasonal occurrence of crashes.

2. Crash Count Dynamics

1. Analysis Method

ARMA models were fit to the data to analyze general dynamics of crash counts. The goal was to identify any common orders, trends, and seasonalities among the three regions. Furthermore, because the daily data is visibly so noisy (Fig. 1), the data were summed by week and by month for further analysis of the data at different sampling intervals. The effects of region and sampling interval on the model were sought.

As an example of the analysis method, the Texas state daily data are analyzed in detail here. First, the data set was detrended by fitting a least-squares linear trend through it (Fig. 2) and then subtracting the trend from the data. A linear trend was assumed due to slow change in mean value in this time period for all three data sets. The resulting data set had a constant mean of zero.

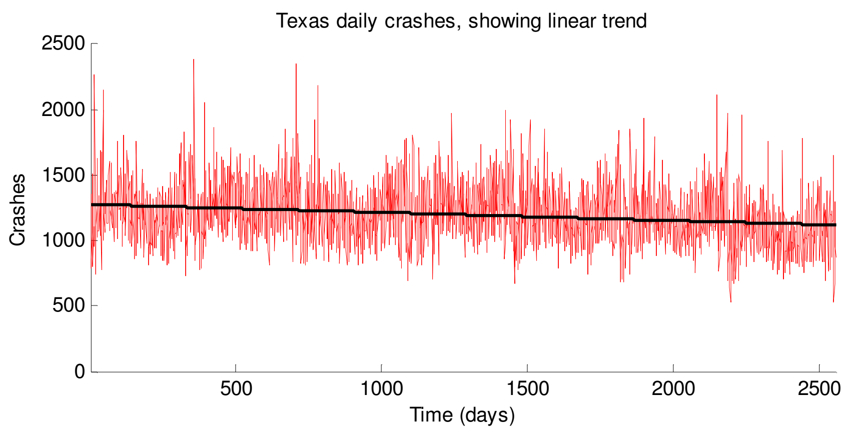


Figure 2. Fitting and subtracting a linear trend from raw data.

Then ARMA models of increasing order were fit to the detrended data until F-testing showed that the model root-sum-square (RSS) values no longer significantly decreased. Standard Matlab code was used for this analysis. The autocorrelation function of the resulting ARMA model noise or residuals was analyzed to ensure white noise (Fig. 3). Using the Matlab "resid" function, 95% of autocorrelation values had to lie within the critical 2/sqrt(N) threshold to indicate white noise.

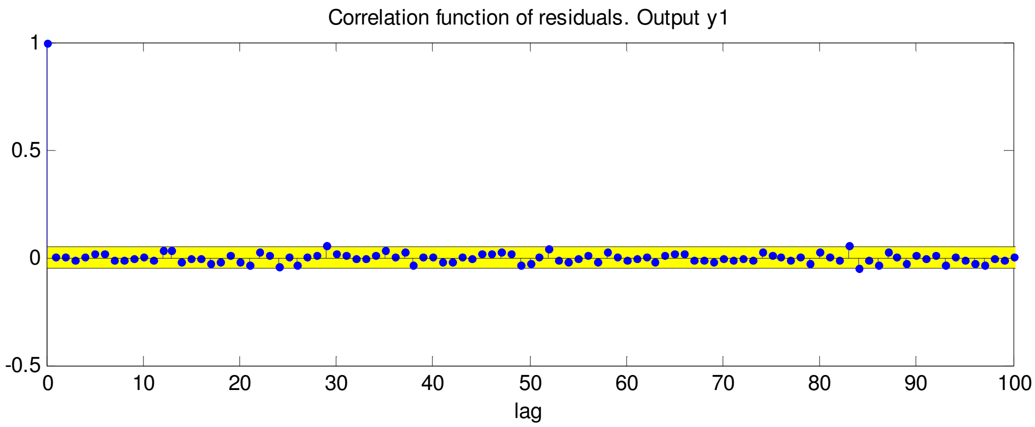


Figure 3. Autocorrelation function of daily Texas state ARMA(10,9) model.

This Texas case, and most all other cases, gave white noise residuals using the standard F-test method. In one case only (monthly Texas data), the standard method gave a model with possibly non-white noise residuals, and its model was increased further until white noise was observed. This is detailed below and in the Appendix.

After confirming white noise residuals, the resulting model was plotted to visually analyze its fit. The daily Texas model here shows some of the erratic behavior of the original data. To help characterize and compare models across different regions and sampling intervals, a kind of "signal-to-noise" ratio was obtained by finding the ratio of *model* variance to *model noise* variance.

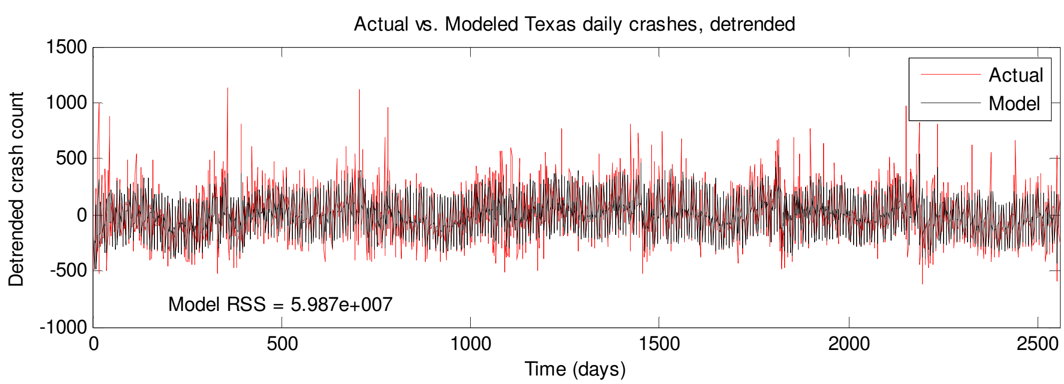


Figure 4. Actual and modeled detrended daily Texas crash count.

The roots of the autoregressive part of the ARMA model were plotted in the complex plane, in order to identify possible trends and seasonalities in the crash data where root magnitudes approached 1. The daily Texas model roots shown here indicate most strongly a number of patterns, compared to the other models obtained (Appendix 6.1). As shown (Fig. 5), the single real root $\lambda=0.98$ weakly suggested a constant linear trend, while the three pairs of complex roots nearly on the unit circle ($|\lambda|>0.9999$) indicated seasonalities of 7, 7/2 and 7/3 days.

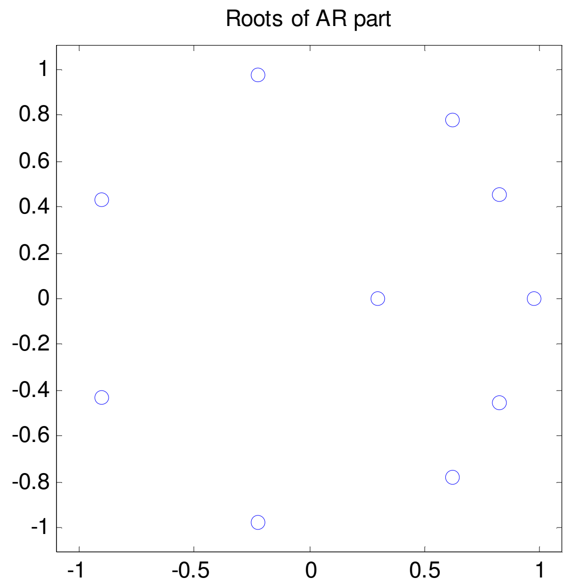


Figure 5. Roots of the AR part of the detrended Texas daily crash model.

The presence of the linear trend was tested by finding the associated parsimonious model and checking for non-significant increases in model RSS. In this case, the parsimonious model $P_t = (1-B)X_t$ was modeled with a one-order-lower ARMA(9,9). The F-test compared the unrestricted (10,9) model RSS with the restricted (9,9) model RSS according to the standard equation:

$$F = [(A_1-A_0)/s] / [A_0/(N-r)]$$

In this case of testing the constant trend parsimonious model for the daily Texas data, the F-value exceeded the critical threshold $F^{95\%}_{(s,N-r)}$, meaning that the parsimonious model had a significantly larger RSS and was not adequate. In other words, the constant trend could not be confirmed, which may be expected given that this root was not closer in magnitude to 1.

Details of the F-tests for checking parsimonious for the daily data are shown in Appendix 6.2. To summarize, none of the parsimonious models associated with trends and seasonalities indicated by the plots of roots were found to be adequate, for not only the Texas daily data but also the Illinois and Austin data. F-statistics consistently showed that the unrestricted, non-parsimonious models significantly improved the model RSS values. This method was applied to the Texas, Illinois, and Austin data, for daily, weekly, and monthly sampling intervals.

2. Effect of Sampling Interval

One goal of checking larger than daily sampling intervals was to see if different sampling intervals revealed new information about crash dynamics. Returning back to the specific case of Texas, this was true. The model of the weekly data gave an ARMA(12,11) model with roots indicating seasonalities at approximately 25.9, 4.33, and 2.36 weeks, corresponding to about 181, 30.3, and 16.5 days, or about 6, 1, and ½ months (Fig. 6). These roots had magnitudes within <0.01 distance from 1 (Appendix 6.3), but parsimonious models were not checked for the weekly sampling interval data. This is because weekly data only indicated seasonalities at periods corresponding approximately to whole or fractional months, and no exact integer relationship exists between weeks and months, which would make any parsimonious model somewhat artificial. Additionally, no roots suggested seasonality for weekly Illinois and Austin data.

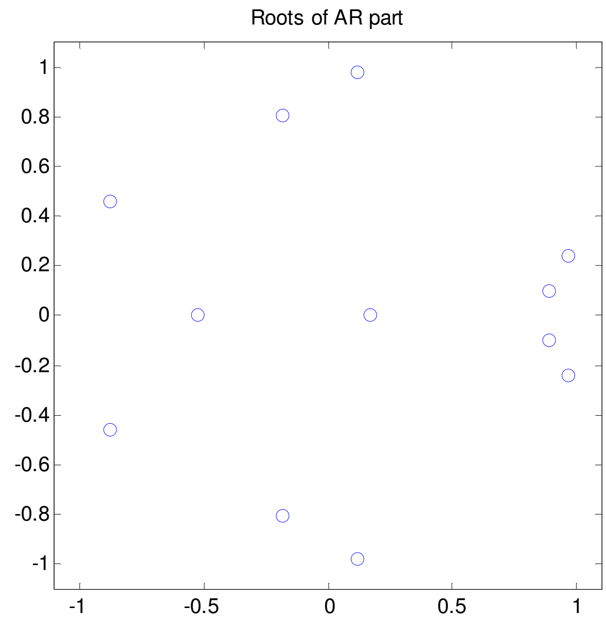


Figure 6. Roots of the AR part of the detrended Texas weekly crash model.

Monthly data was similarly analyzed for Texas, Illinois, and Austin. The initial model obtained for the Texas data was an AR(1) model whose residual autocorrelation function showed almost exactly 5% of its values exceeding the critical threshold of whiteness, which was not ideal. Higher order models were fit until both the RSS decreased and a good autocorrelation function was found, leading to an ARMA(6,5) model. This model had roots weakly suggesting seasonality of period 5.96 months (λ=0.989), and strongly at 2.4 months (λ=0.998), although parsimonious models were inadequate. This monthly Texas data was the only case where the standard F-testing procedure was overridden to find a higher order model.

3. Summary of Results

Appendix 6.4 gives other remaining plots of the relevant data, models, residual autocorrelation functions, and autoregressive roots for the Texas, Illinois, and Austin data at daily, weekly, and monthly sampling intervals. The table below summarizes the modeling results for the data sets. Each region and sampling interval yielded different models of inconsistent order between (1,0) and (12,11).

However, one consistent pattern among all the regions was that daily data exhibited 7-day seasonality, which makes sense if crashes follow a weekly pattern according to traffic rates. Seasonalities of other periods may be found among specific regions and at certain sampling intervals.

Sampling	Modeling Feature	Region: Texas	Illinois	Austin
Daily	ARMA order	**(10,9)**	**(8,7)**	**(6,5)**
	Samples (N)	2557	2557	1461
	ARMA noise variance	2.38E+04	4.82E+04	74.1
	ARMA model variance	2.17E+04	3.44E+04	15.5
	Ratio (model variance)/ (noise variance)	91.2%	71.4%	21%
	Possible seasonalities	7, 7/2, 7/3 days	7, 7/2, 7/3 days	7 days
Weekly	ARMA order	**(12,11)**	**(2,1)**	**(2,1)**
	Samples (N)	364	364	207
	ARMA noise variance	2.46E+05	7.98E+05	593
	ARMA model variance	1.79E+05	5.84E+05	94.0
	Ratio (model variance)/ (noise variance)	72.7%	73.2%	16%
	Possible seasonalities	6, 1, 1/2 months	none	none
Monthly	ARMA order (initial F-tests)	**(1,0)**	**(4,3)**	**(8,7)**
	Samples (N)	83	83	47
	ARMA noise variance	4.54E+06	1.35E+07	3.70E+03
	ARMA model variance	8.27E+05	9.58E+06	5.17E+03
	Ratio (model variance)/ (noise variance)	18.2%	70.9%	140%
	Possible seasonalities	none	none	none
	ARMA order (extra model)	**(6,5)**		
	ARMA noise variance	2.43E+06		
	ARMA model variance	3.45E+06		
	Ratio (model variance)/ (noise variance)	142.1%		
	Possible seasonalities	2.39 months	6 months	N/A
Adequate parsimonious models		none	none	none

Furthermore, the above data suggests that the signal-to-noise ratio for ARMA models of crashes tends to be less than 1, meaning that the noise is very large relative to the model. This emphasizes the difficulty of modeling crashes and separating out real trends from noise.

3. Prediction of Crashes

1. Analysis Method

A subset of the above data was analyzed to see how well ARMA models could predict future crashes. Daily data was deemed less useful for prediction, since even if the number of crashes could be predicted for a particular day, a large region like a state probably could not respond practically to that. Cities may find it useful, but only if they can rapidly mobilize and change service levels for each day. The monthly data was not analyzed because the sample sizes became less than 100, which could reduce model robustness. The weekly state data were the focus of this analysis.

The prediction step was chosen to be more than one month, or 5 weeks ahead for the weekly data. This was chosen because, at least at the federal level, 90% of crashes take up to 30 days to be reported and recorded [6]. Assuming that crash data is only accurate 4 weeks ago, a 5-step-ahead prediction would be needed to predict next week's crash count. i.e. calculating the prediction of $x_{t-4}(5)$.

Initially to examine the prediction method, arbitrary time periods were chosen to fit the model, with the remaining data used to check the prediction. Plots of continuously updated predictions were examined for their accuracy. For example, when predicting 2009-2010 Texas data based on 2004-2008, a 95%CI of the prediction could be calculated based on the first 5 terms of the Green's function and the standard deviation of the model residuals (Fig. 3.1.1).

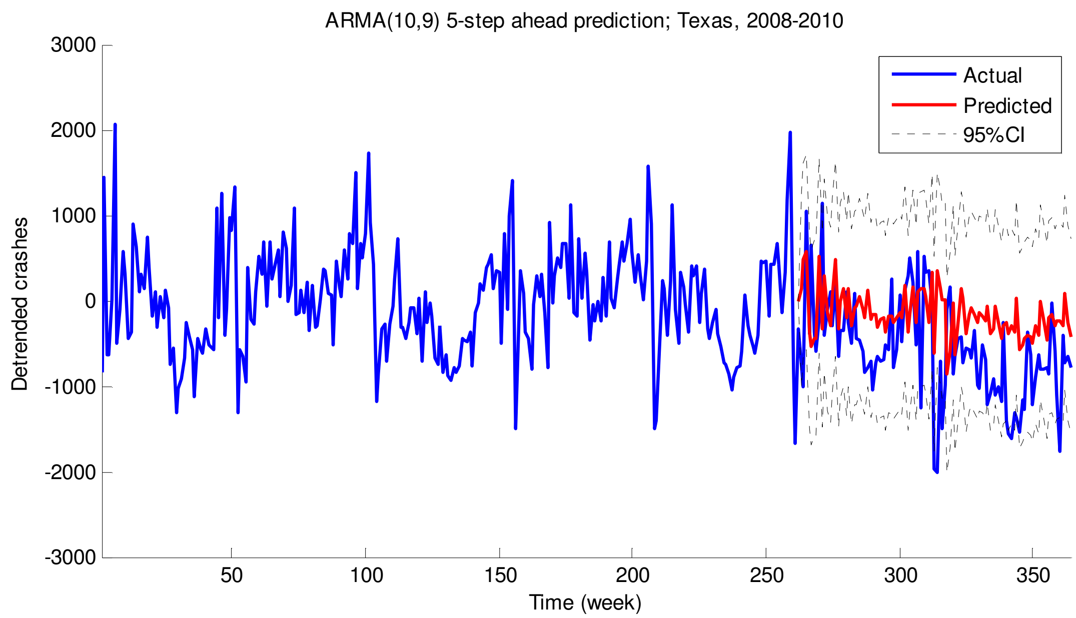


Figure 3.1.1. 5-step-ahead prediction of 2009-2010 Texas weekly detrended crashes.

A quantitative measure of the goodness of the prediction was the actual variance of the prediction error overall the whole predicted period. In this case, for example, the prediction error variance was 3.79e5. In contrast, a prediction of 2006-2010 data based on 2004-2005 is shown below (Fig. 3.1.2), and it has a larger prediction error variance of 2.93e6, even if the 95% CI intervals in that case were narrower. The prediction error improved when a shorter time period was being predicted, or when more data was used to form the prediction model.

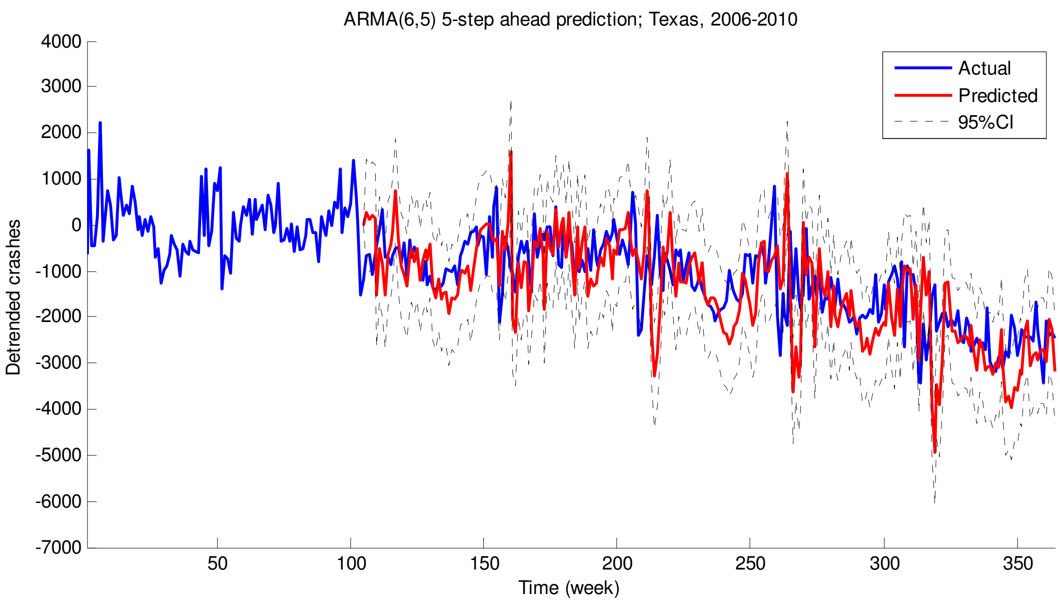


Figure 3.1.2. 5-step-ahead prediction of 2006-2010 Texas weekly detrended crashes.

Illinois data was more difficult to predict, as shown in Fig. 3.1.3., for example. The spikes seen here are not modeled well. More significantly, a drop in the mean value of the data occurs beginning around 2008, which consistently disrupted prediction models no matter what time period was the model basis, as several were tested. Essentially the time series was not stationary, rendering the ARMA models inadequate for prediction. However, this deviation from the model beginning in 2008 can usefully indicate when a new model must be made.

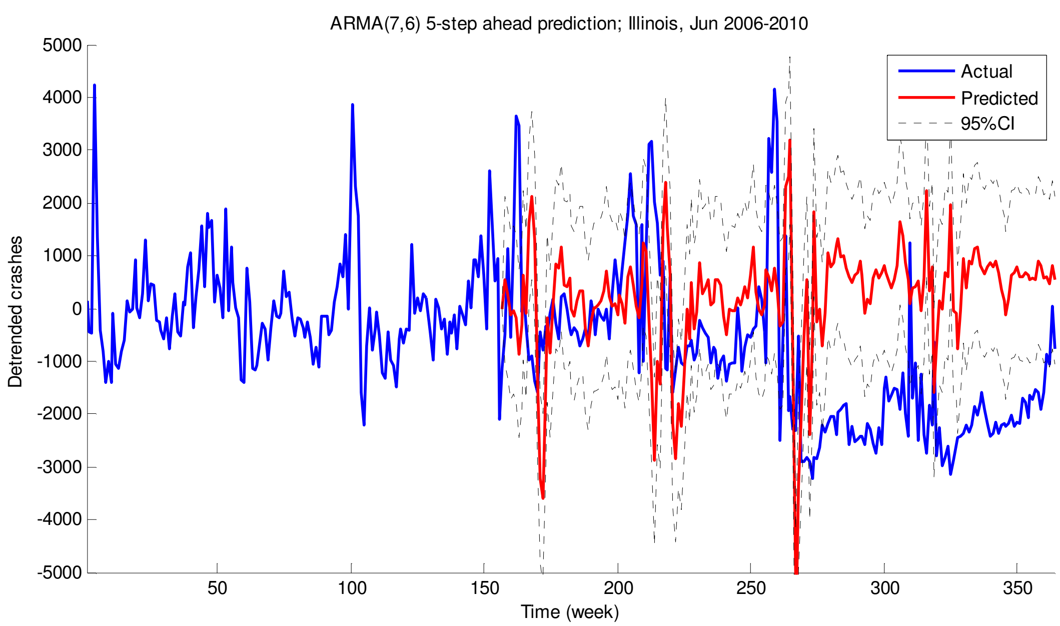


Figure 3.1.3. 5-step-ahead prediction of 2007-2010 Illinois weekly detrended crashes.

2. **Effect of Model Time Period**

To examine the effect of the model time period on predictions, code was written to model 2-year windows in the Texas weekly crash data, and then predict the following 2 years of crashes. The windows were rolling, where it moved up by one week. 152 ARMA models and corresponding 5-step-ahead predictions could be made in this way.

For the Texas data, a plot of the ARMA model orders as the window moved showed that the orders were not consistent and changed frequently (Fig. 3.2.1). This suggests that the start and end of the time period for modeling is very influential on the ARMA model, and likely can change the resulting detected seasonalities and predictions.

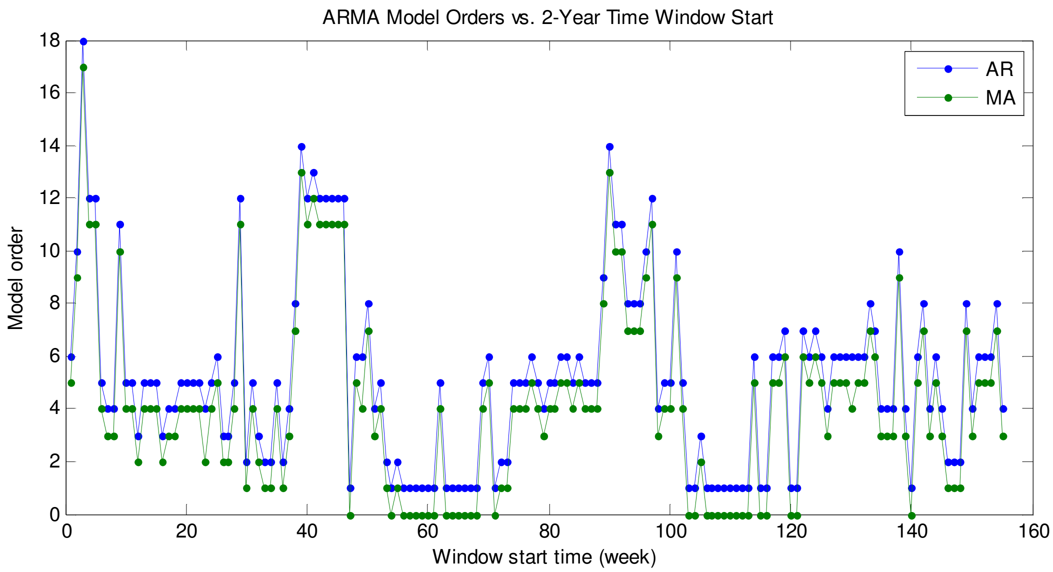


Figure 3.2.1. ARMA model orders for Texas weekly data, in rolling 2-year time windows.

A distribution plot (Fig. 3.2.2) of the AR orders of these models showed that AR orders 5 and 1 appeared most frequently, and that orders greater than 7 were less common. This suggests that model order still tends toward certain values for this data and is not random.

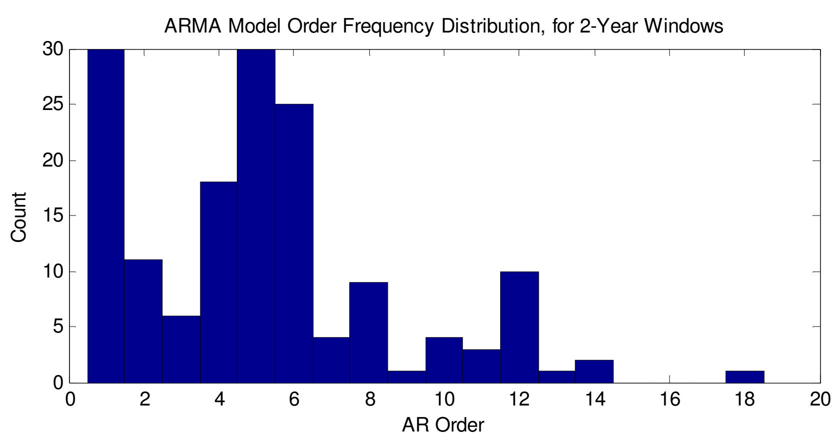


Figure 3.2.2. ARMA model AR order histogram for Texas weekly data, in rolling 2-year time windows.

Finally the prediction error variance of the next 2 years also was inconsistent, as the model time window rolled onward. A plot of the prediction error variance is shown in Fig. 3.2.3, suggesting that prediction errors can be extremely large if the wrong time window is chosen.

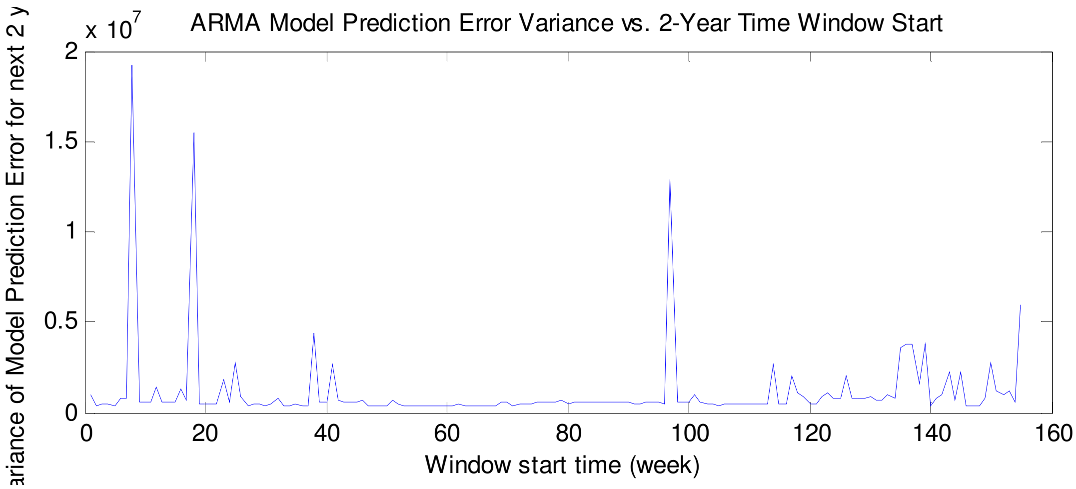


Figure 3.2.3. 2-year prediction error variance of ARMA models, based on rolling 2-year time windows.

The Illinois state data suggested similar findings. However, in addition, it had another frequency distribution of models for the rolling 2-year windows, as shown in Fig. 3.2.4. ARMA models with AR order 1 were most frequent, suggesting that this is associated with highly noisy data series that are difficult to predict.

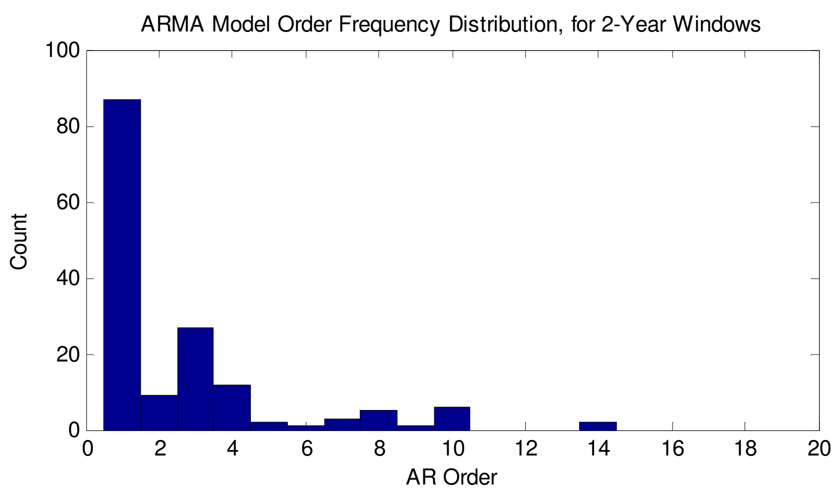


Figure 3.2.2. ARMA model AR order histogram for Illinois weekly data, in rolling 2-year time windows.

3. Effect of Gas Prices

A vectorial ARMA model was investigated for predicting weekly crash data, to see if it could improve predictions. It was hypothesized that gas prices may affect crash counts, since higher prices might cause fewer people to drive, leading to fewer crashes. The weekly gas prices for Texas are shown in Fig. 3.3.1.

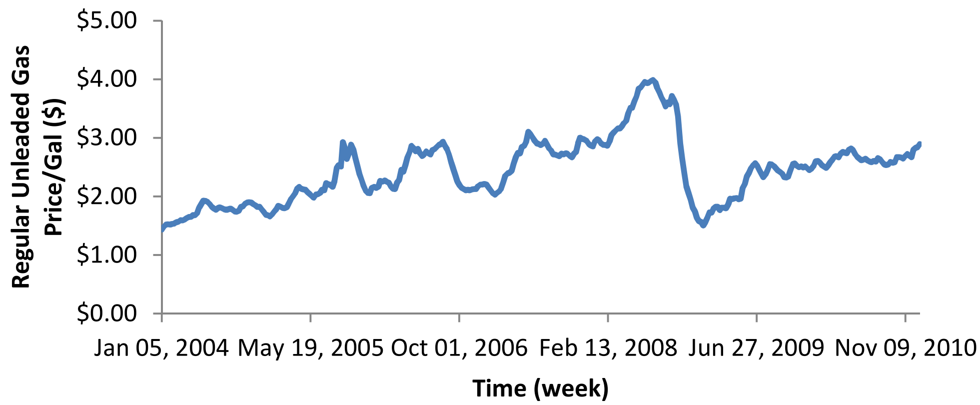


Figure 3.3.1. Weekly Texas prices for regular gasoline, 2004-2010.

Data was examined again using 2004-2008 data to predict 2009-2010 data. The standard Matlab code for one-input-one-output ARMA models was used to systematically fit a vectorial model to the gas price and detrended crash data. F-testing and the corresponding model parameters are given in Appendix 6.5. An overall model was found where the crash data (x_2) could be represented by an (8,7) model and the gas price data (x_1) could be represented by a (2,1) model. The resulting prediction was plotted (Fig 3.3.2).

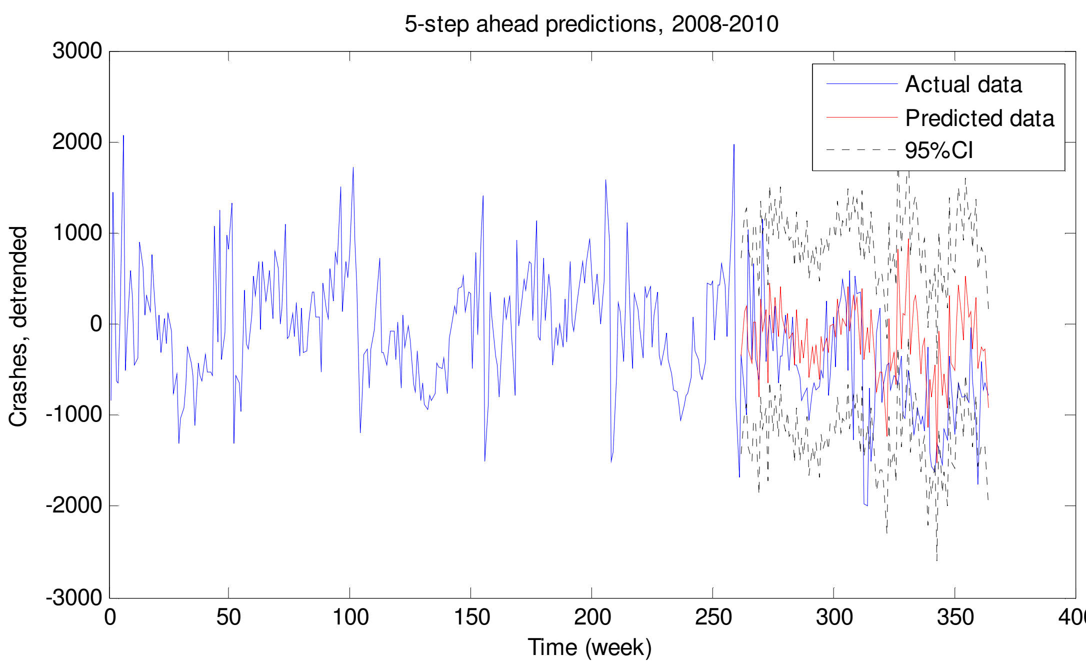


Figure 3.3.2. 5-step-ahead prediction of 2009-2010 Texas weekly detrended crashes, using vectorial ARMA involving gas prices.

This prediction is similar to the one previously found using the single time series (Fig. 3.1.1). In fact, the vectorial model prediction error variance is 4.15e5, which is about 10% larger than that of the single time series model. In this case the vectorial model did not help improve prediction accuracy (as measured by error variance) nor precision (as shown by the 95%CI). This could be partly due to no strong link between the two time series, and also because the gas data may not be stationary either, especially from 2008-2009. The sharp decrease in crashes after 2008 maybe suggests that the economic downturn caused fewer people to drive, thereby decreasing crashes. To examine this, other time series could be tested in vectorial models, such as economic indicators like GDP or unemployment rates.

4. Discussion

The modeling analysis performed here revealed several characteristics of crash count dynamics. First, and most significantly, model orders and their dependent trends and seasonalities vary, and they are greatly determined by the region, time period, and sampling interval from which they are based.

Nevertheless, the daily crash count data here consistently showed a 7-period seasonality, likely corresponding with the weekly cycles of higher traffic throughout the workweek and less traffic during the weekend. Other seasonalities could be detected for specific places, times, and intervals. The 6-month seasonality detected for Texas and Illinois could be weather-related or school-related, as crashes may follow seasonal storms or biannual migrations of students.

Regarding prediction, care must be taken to avoid predicting nonstationary time series. However, if that cannot be avoided, as is the case with the Illinois crashes starting in 2008, these models become tools to see when crash patterns are changing, or when a new model is needed. If a prediction is showing very large errors, a new model should be made with updated data. Despite that warning, one-month old crash data may be good enough to predict next week's crashes in some places, including Texas state.

The results here have some limitations to their implications. First, the data has only examined two states and one city, and clearly other regions can be examined. Second, the data is more useful to insurance and trucking companies than government, since government is much more interested in crash fatalities rather than crashes themselves. Then this data only covers severe crashes, which discounts much financial losses from less serious accidents. Finally these models cannot directly explain crash causes.

Despite these limitations, this exercise has still provided useful information about crash patterns and how to employ ARMA models with them. Industry and government could still use such models to help budget for impending crashes. Future work should address the limitations described above by expanding the regions studied, analyzing fatalities, and using models that may account for non-stationary time series.

5. References

[1] U.S. Department of Transportation, National Highway Traffic Safety Administration. "2010 Motor Vehicle Crashes: Overview." Research Note. Revised February 2012.

[2] Texas Department of Transportation. "Comparison of motor vehicle traffic deaths, vehicle miles, death rates, and economic loss, 2003-2010." Research Report. 2010.

[3] Texas Department of Transportation. "Texas Motor Vehicle Crash Statistics." http://www.txdot.gov/txdot_library/drivers_vehicles/publications/crash_statistics/default.htm. Accessed March 2012.

[4] Illinois Department of Transportation, via Lori Midden, Traffic Statistics Unit Manager. Contacted March 2012.

[5] Texas Department of Transportation, via Carol Thomas, TxDOT Traffic Operation. Contacted April 2012.

[6] U.S. Department of Transportation, National Highway Traffic Safety Administration. "Statistical methodology to make early estimates of motor vehicle traffic fatalities." November 2010.

[7] U.S. Energy Information Administration. "US Regular Gasoline Prices". US Department of Energy. www.eia.gov April 2012.

[8] U.S. Department of Transportation, National Highway Traffic Safety Administration. "An analysis of the significant decline in motor vehicle traffic fatalities in 2008". June 2010.

6. Appendices

1. ROOTS OF DAILY ARMA MODELS

root	Texas	ARMA(10,9)				
root	real	imag	magnitude	angle (rad)	angle(deg)	period
1	-0.90091	0.433716	0.9999	2.6929	154.3	2.333216
2	-0.90091	-0.43372	0.9999	3.5903	205.7	
3	-0.22257	0.974904	1	1.7953	102.9	3.499837
4	-0.22257	-0.9749	1	4.4879	257.1	
5	0.623333	0.781913	1	0.8978	51.4	6.99851
6	0.623333	-0.78191	1	5.3854	308.6	
7	0.980019	0	0.98	0	0.0	random walk
8	0.831595	0.456602	0.9487	0.5021	28.8	12.51324
9	0.831595	-0.4566	0.9487	5.7811	331.2	
10	0.298797	0	0.2988	0	0.0	N/A

	Illinois	ARMA(8,7)				
root	real	imag	magnitude	angle (rad)	angle(deg)	period
1	-0.90096	0.433799	0.999953	2.692865	154.3	2.333216
2	-0.90096	-0.4338	0.999953	3.590321	205.7	
3	-0.22255	0.974523	0.999612	1.795318	102.9	3.499837
4	-0.22255	-0.97452	0.999612	4.487868	257.1	
5	0.622504	0.780877	0.998639	0.897774	51.4	6.99851
6	0.622504	-0.78088	0.998639	5.385411	308.6	
7	0.968672	0	0.968672	0	0.0	random walk
8	0.165634	0	0.165634	0	0.0	

	Austin	ARMA(6,5)				
root	real	imag	magnitude	angle (rad)	angle(deg)	period
1	-0.2221	0.974961	0.999938	1.794778	102.8	3.498
2	-0.2221	-0.97496	0.999938	4.488407	257.2	
3	0.621884	0.782134	0.999236	0.899043	51.5	6.999
4	0.621884	-0.78213	0.999236	5.384142	308.5	
5	0.956505	0	0.956505	0	0.0	
6	0.031001	0	0.031001	0	0.0	

2. PARSIMONIOUS MODEL F-TESTS FOR DAILY CRASH DATA

Texas	**testing**	**constant trend**	**7th roots**	**period 7**	**period 7/2**	**period 7/3**
	test	ARMA(9,9) vs. (10,9)	ARMA(3,9) vs. ARMA(10,9)	ARMA(8,9) vs. ARMA(10,9)	ARMA(8,9) vs. ARMA(10,9)	ARMA(8,9) vs. ARMA(10,9)
restricted	A1	6.05E+07	6.15E+07	6.19E+07	6.17E+07	6.29E+07
unrestricted	A0	5.99E+07	5.99E+07	5.99E+07	5.99E+07	5.99E+07
no restricted	s	1	7	2	2	2
samples	N	2557	2557	2557	2557	2557
no params	r	20	20	20	20	20
	F	27.5858	9.8732	42.7134	39.4718	65.0024
	Fcrit95	3.845127	2.013185	2.999272	2.999272	2.999272

Illinois	**testing**	**constant trend**	**7th roots**	**period 7**	**period 7/2**	**period 7/3**
	test	(7,7) vs (8,7)	(1,7) vs (8,7)	(6,7) vs (8,7)	(6,7) vs (8,7)	(6,7) vs (8,7)
restricted	A1	1.24E+08	1.32E+08	1.27E+08	1.26E+08	1.32E+08
unrestricted	A0	1.22E+08	1.22E+08	1.22E+08	1.22E+08	1.22E+08
no restricted	s	1	7	2	2	2
samples	N	2557	2557	2557	2557	2557
no params	r	16	16	16	16	16
	F	43.2804	31.6613	53.7346	44.4304	109.9783
	Fcrit95	3.845121	2.01318	2.999267	2.999267	2.999267

Austin	**testing**	**constant trend**	**period 7**	**period 7/2**
	test	(6,5) vs (5,5)	(6,5) vs (5,5)	(6,5) vs (5,5)
restricted	A1	1.08E+05	1.07E+05	1.13E+05
unrestricted	A0	1.06E+05	1.06E+05	1.06E+05
no restricted	s	1	1	1
samples	N	1461	1461	1461
no params	r	12	12	12
	F	16.4907	15.1278	84.4977
	Fcrit95	3.847884	3.847884	3.847884

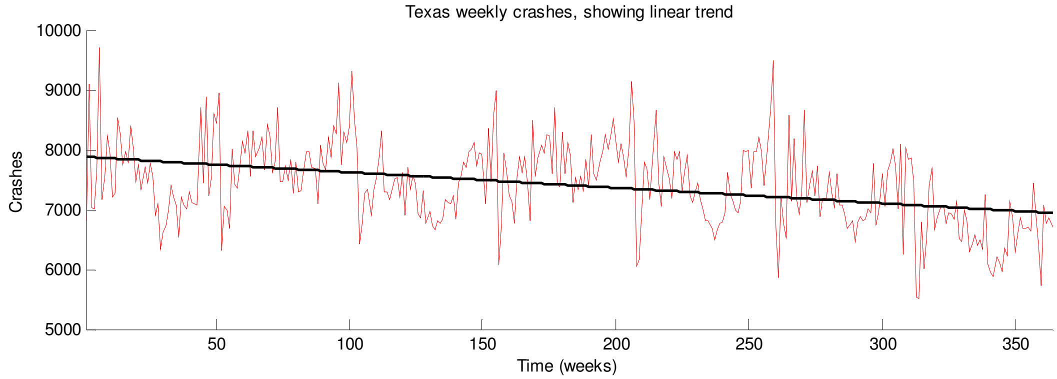
3. WEEKLY MODEL ROOTS

	Texas	ARMA(12,11)					
root	real	imag	magnitude	angle (rad)	angle(deg)	period	period est
1	-0.87702	0.45976	0.990225	2.658751	152.3	2.36321	2.333333
2	-0.87702	-0.45976	0.990225	3.624435	207.7		
3	0.118486	0.984012	0.991119	1.450962	83.1	4.33035	3.5
4	0.118486	-0.98401	0.991119	4.832224	276.9		
5	-0.18405	0.805096	0.825865	1.795539	102.9	N/A	7
6	-0.18405	-0.8051	0.825865	4.487646	257.1		
7	-0.52406	6.42E-17	0.524057	3.141593	180.0	constant	
8	0.970148	0.240149	0.999429	0.242661	13.9	25.892	
9	0.970148	-0.24015	0.999429	6.040525	346.1		
10	0.890959	0.099417	0.896488	0.111125	6.4	N/A	
11	0.890959	-0.09942	0.896488	6.172061	353.6		
12	0.170984	0	0.170984	0	0.0		

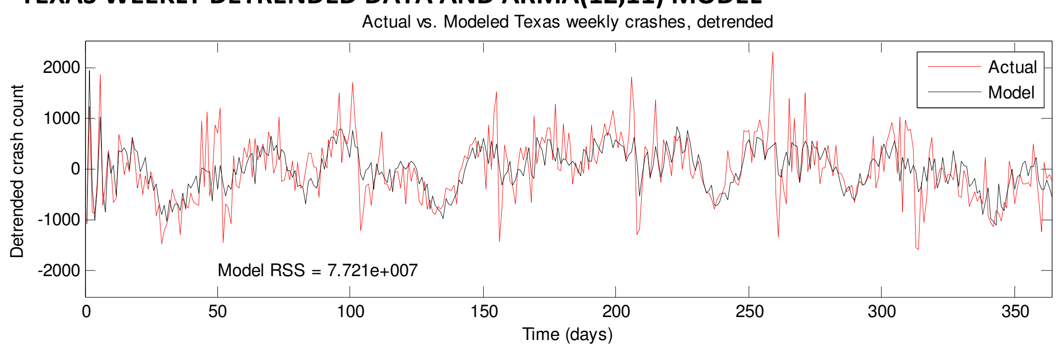
4. RELEVANT PLOTS OF MODELS

TEXAS DAILY PLOTS (see main body of report)

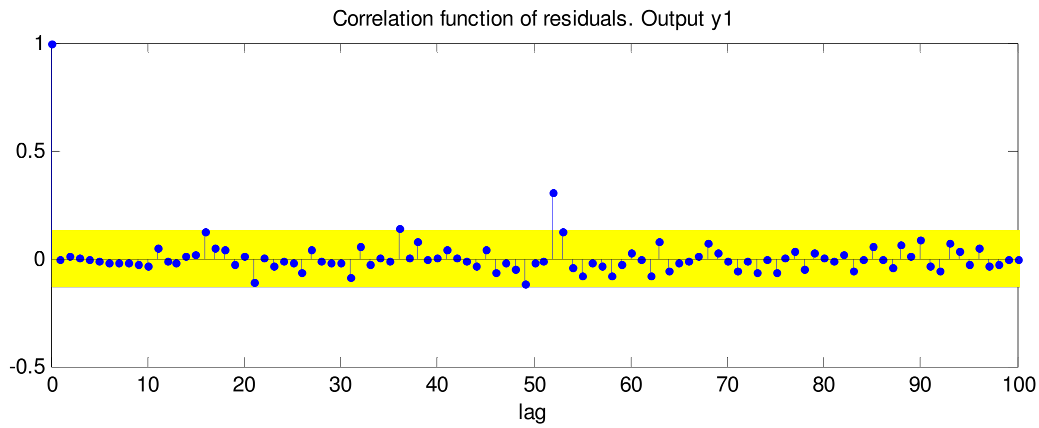
TEXAS WEEKLY RAW DATA AND TREND



TEXAS WEEKLY DETRENDED DATA AND ARMA(12,11) MODEL

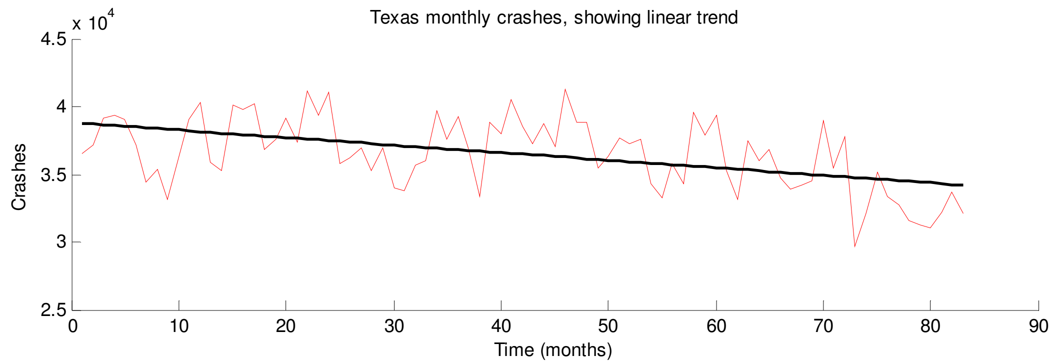


TEXAS WEEKLY MODEL RESIDUAL AUTOCORRELATION FUNCTION

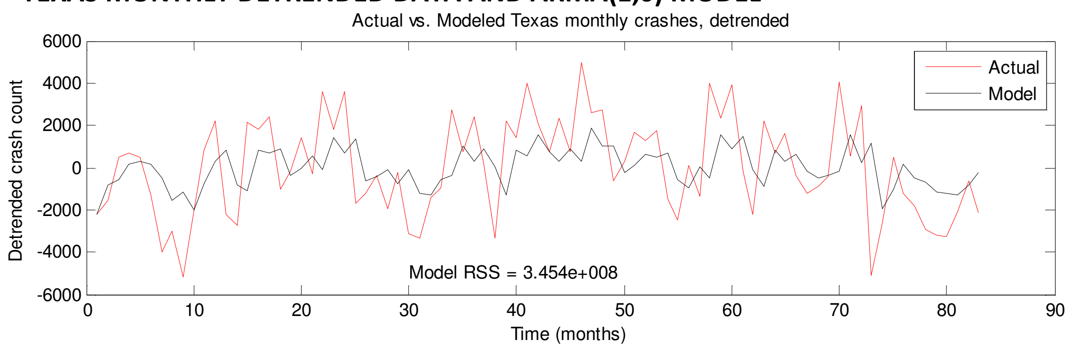


TEXAS WEEKLY MODEL AUTOREGRESSIVE ROOTS (see main body of report)

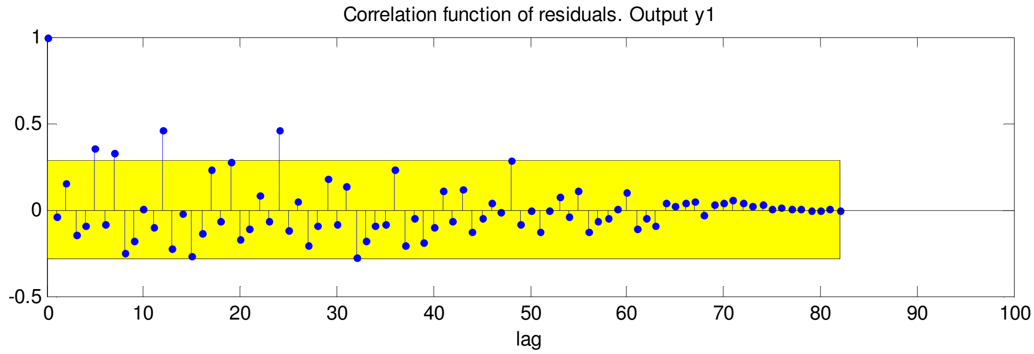
TEXAS MONTHLY RAW DATA AND TREND



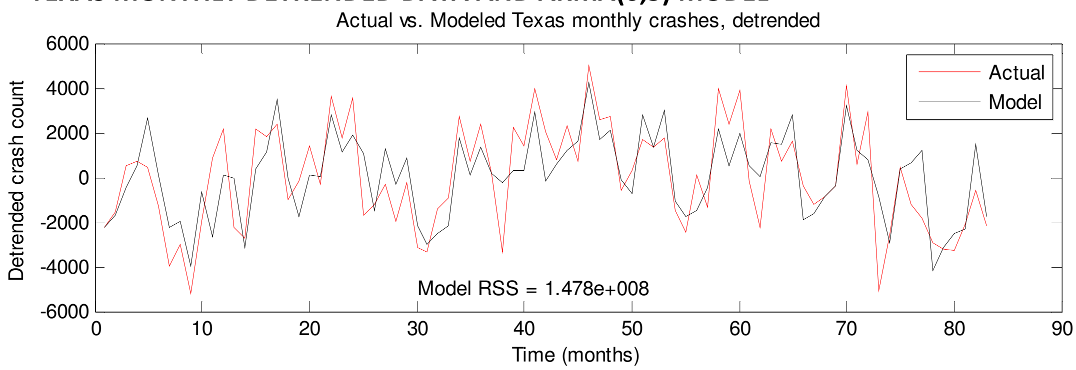
TEXAS MONTHLY DETRENDED DATA AND ARMA(1,0) MODEL



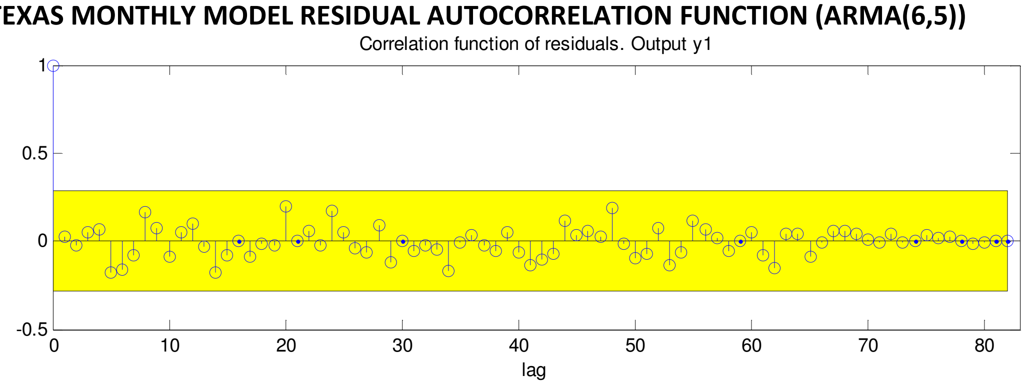
TEXAS MONTHLY MODEL RESIDUAL AUTOCORRELATION FUNCTION (ARMA(1,0))



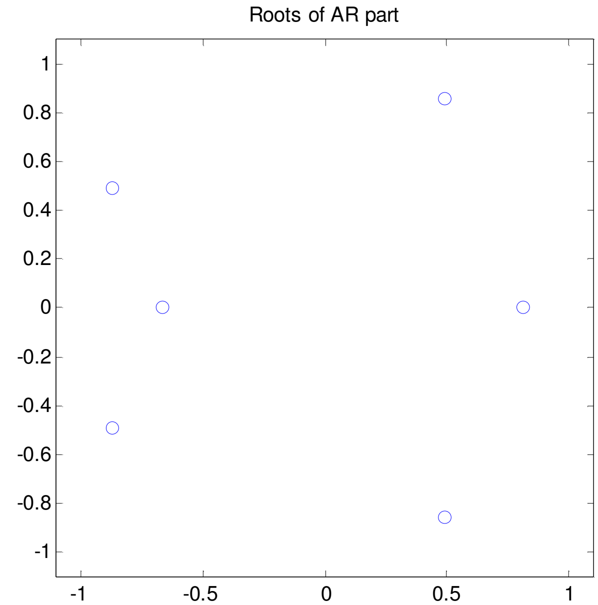
TEXAS MONTHLY DETRENDED DATA AND ARMA(6,5) MODEL



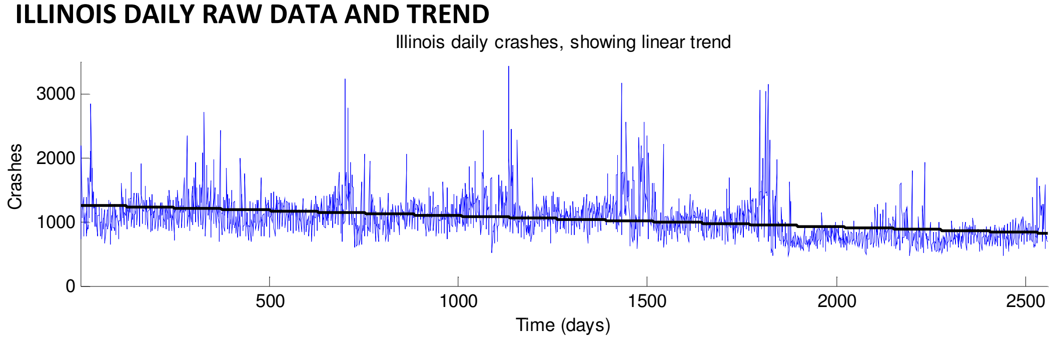
TEXAS MONTHLY MODEL RESIDUAL AUTOCORRELATION FUNCTION (ARMA(6,5))



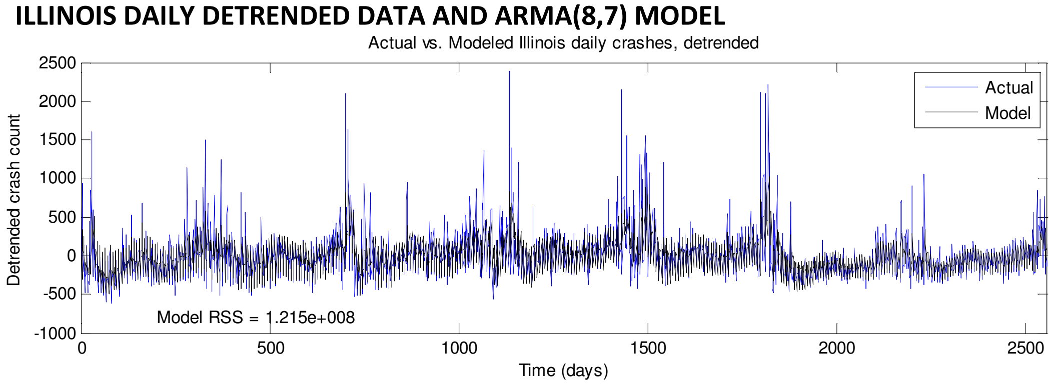
TEXAS MONTHLY MODEL AUTOREGRESSIVE ROOTS (ARMA(6,5))



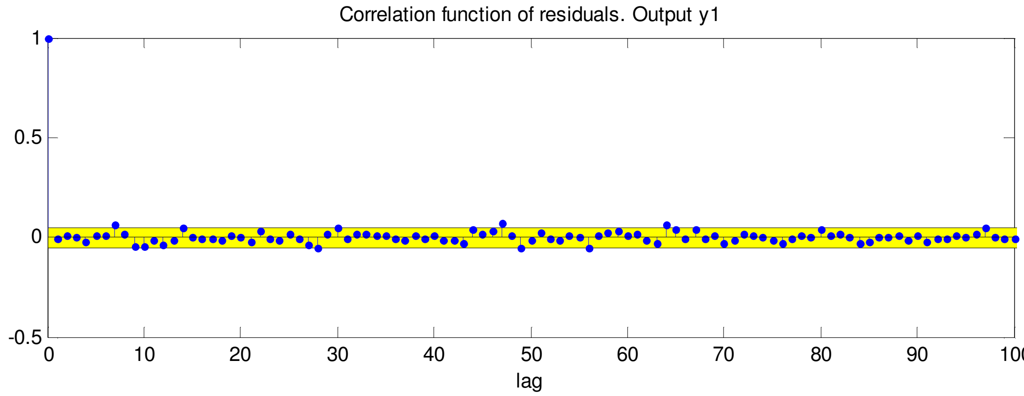
ILLINOIS DAILY RAW DATA AND TREND



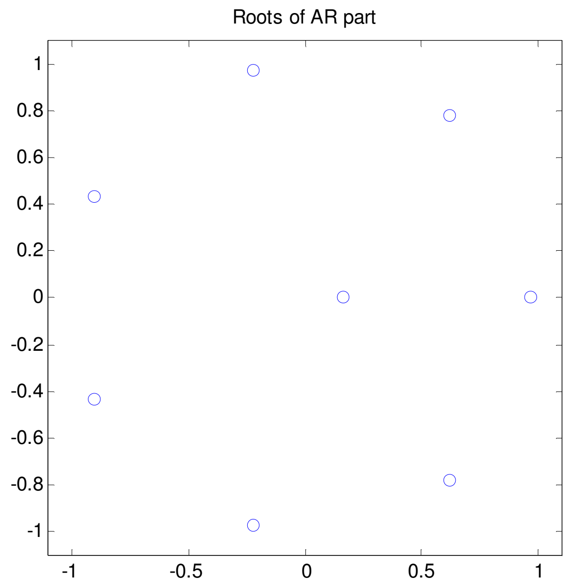
ILLINOIS DAILY DETRENDED DATA AND ARMA(8,7) MODEL



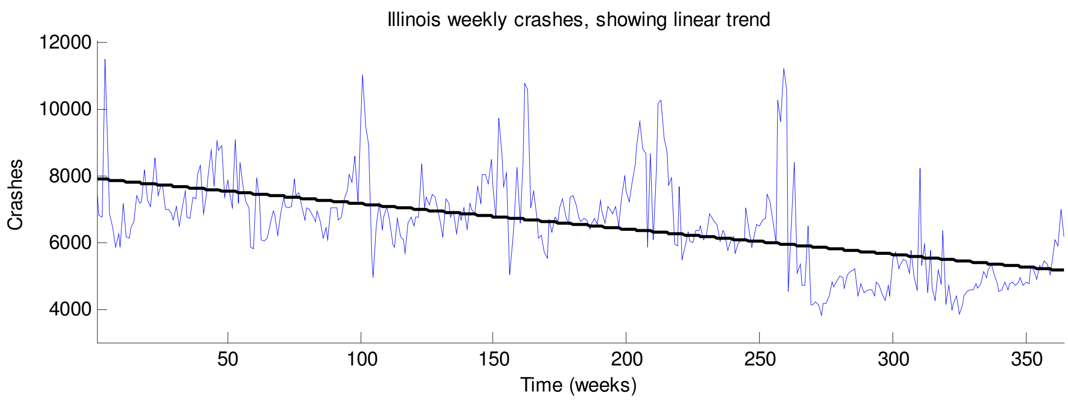
ILLINOIS DAILY MODEL RESIDUAL AUTOCORRELATION FUNCTION



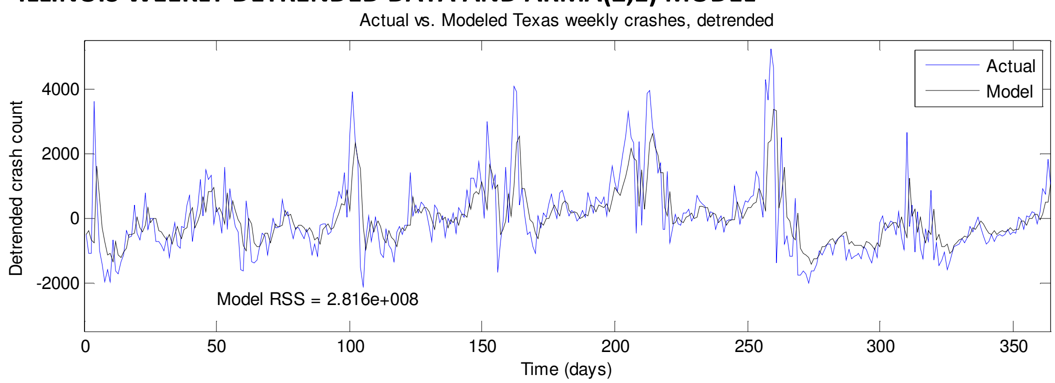
ILLINOIS DAILY MODEL AUTOREGRESSIVE ROOTS



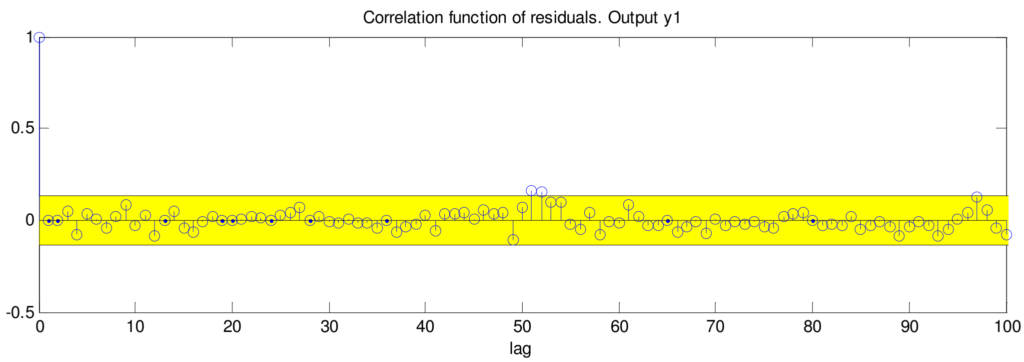
ILLINOIS WEEKLY RAW DATA AND TREND



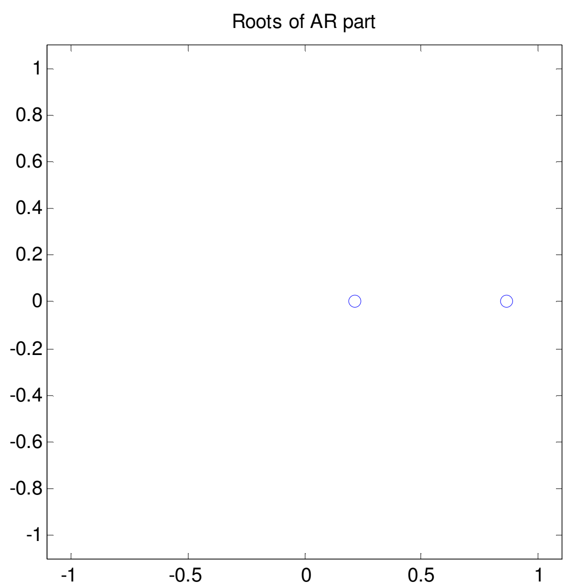
ILLINOIS WEEKLY DETRENDED DATA AND ARMA(2,1) MODEL



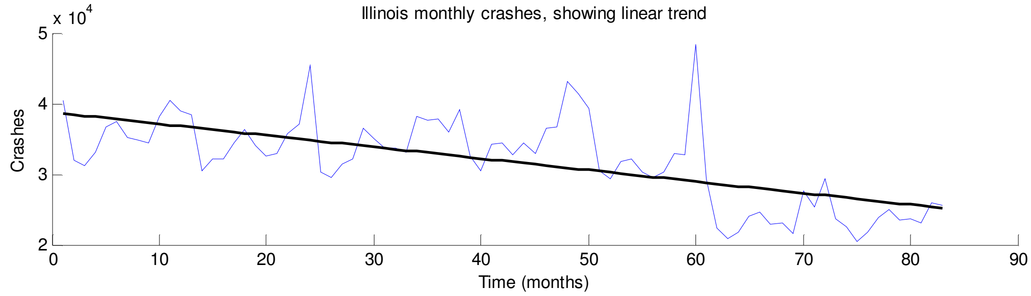
ILLINOIS WEEKLY MODEL RESIDUAL AUTOCORRELATION FUNCTION



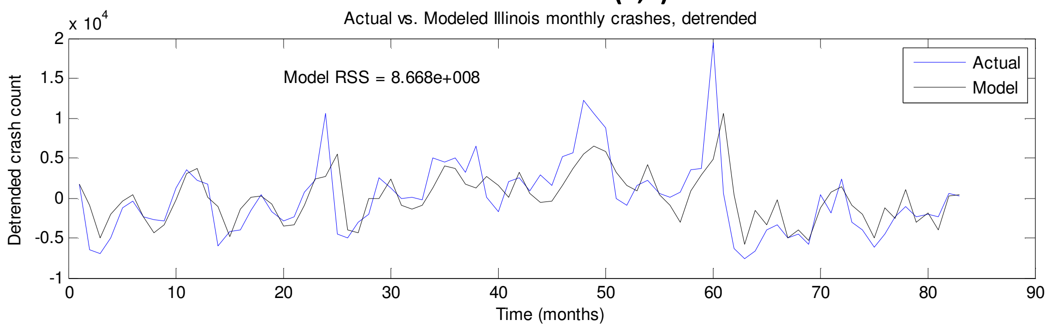
ILLINOIS WEEKLY MODEL AUTOREGRESSIVE ROOTS



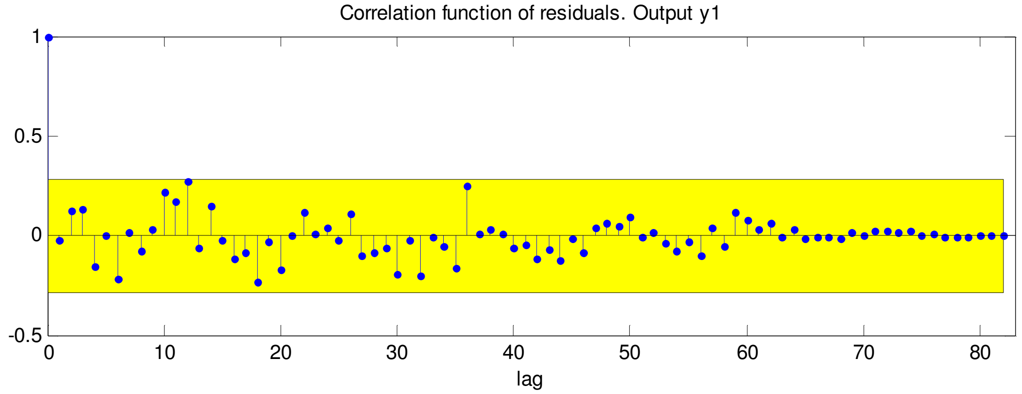
ILLINOIS MONTHLY RAW DATA AND TREND



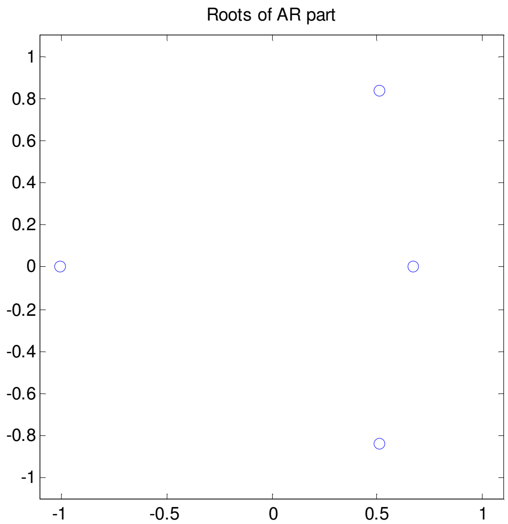
ILLINOIS MONTHLY DETRENDED DATA AND ARMA(4,3) MODEL



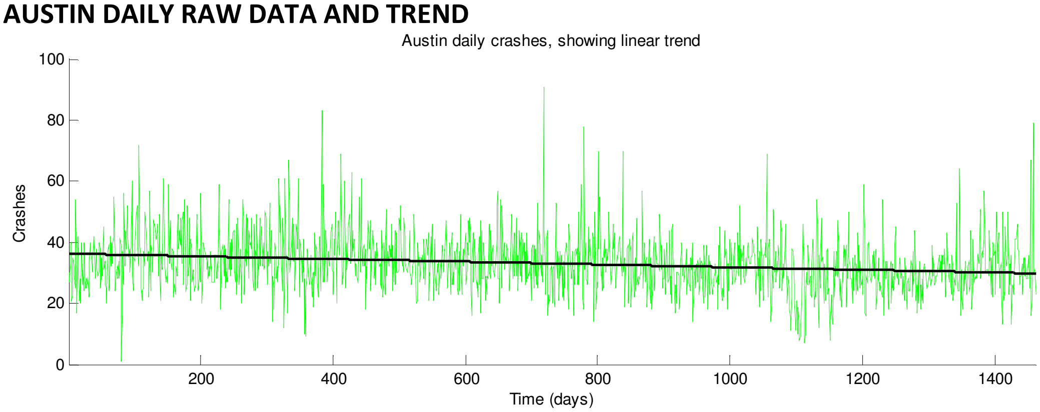
ILLINOIS MONTHLY MODEL RESIDUAL AUTOCORRELATION FUNCTION



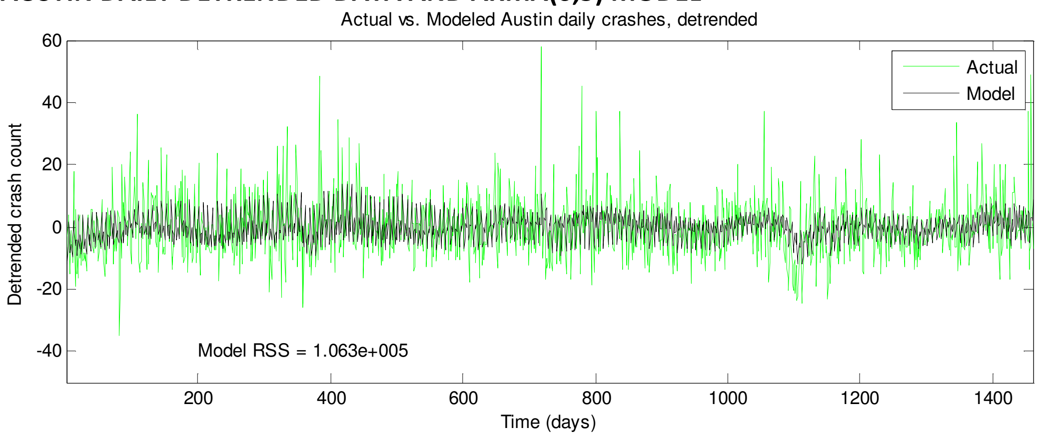
ILLINOIS MONTHLY MODEL AUTOREGRESSIVE ROOTS



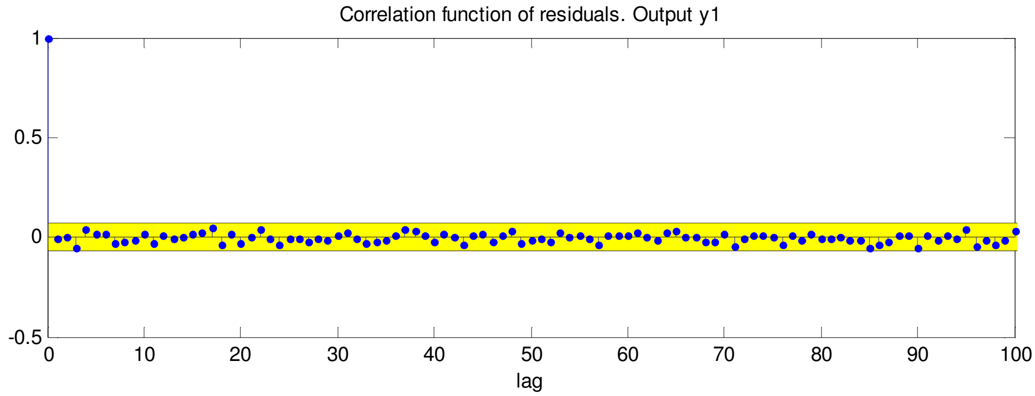
AUSTIN DAILY RAW DATA AND TREND



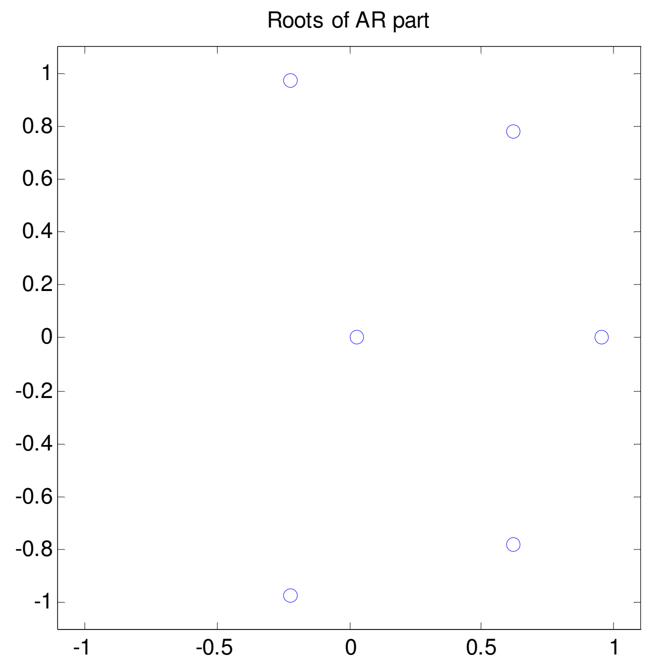
AUSTIN DAILY DETRENDED DATA AND ARMA(6,5) MODEL



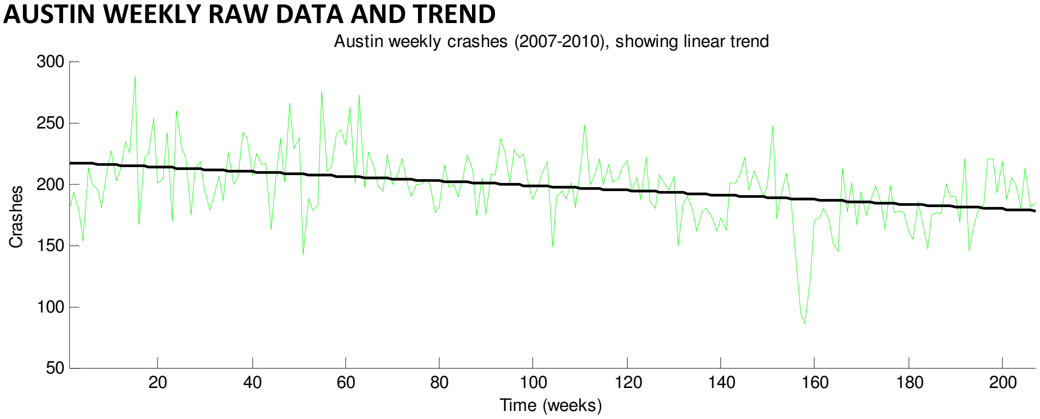
AUSTIN DAILY MODEL RESIDUAL AUTOCORRELATION FUNCTION



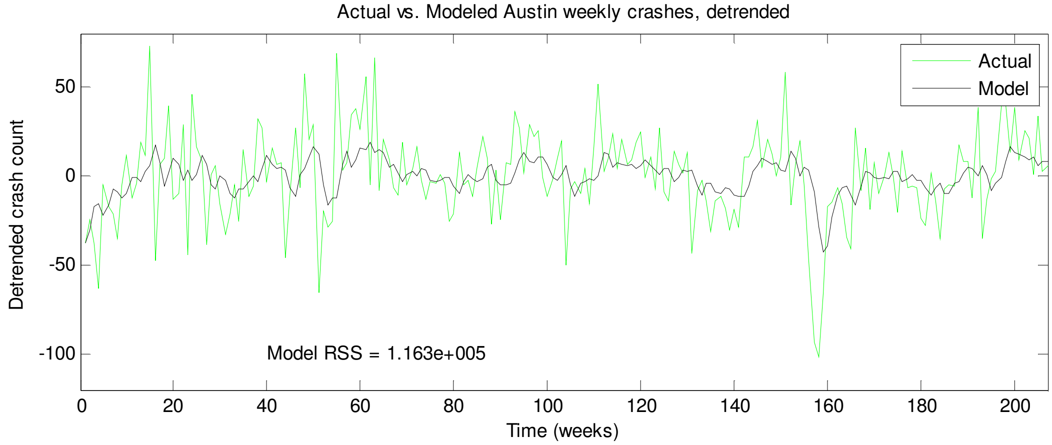
AUSTIN DAILY MODEL AUTOREGRESSIVE ROOTS



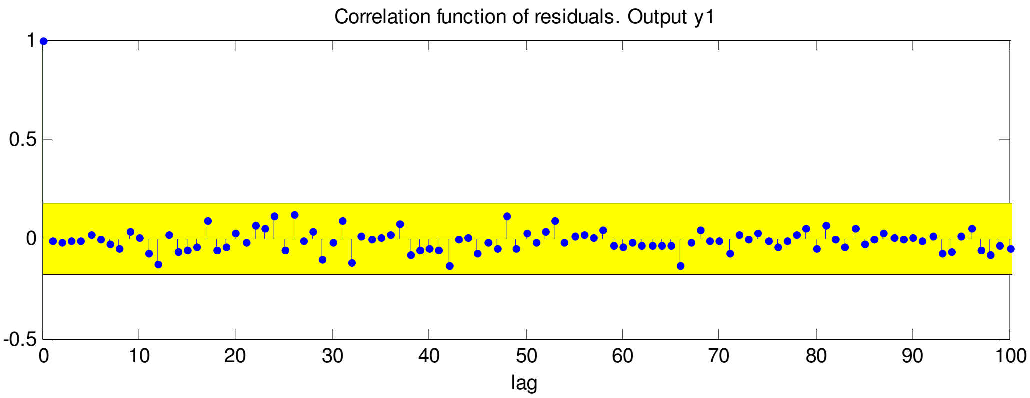
AUSTIN WEEKLY RAW DATA AND TREND



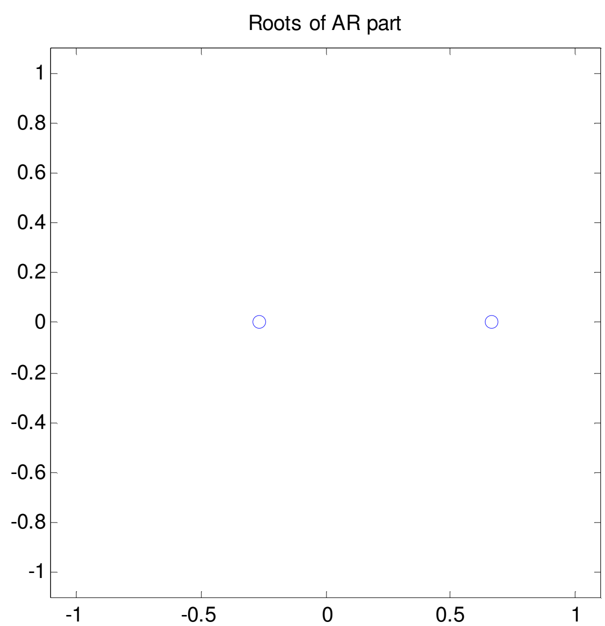
AUSTIN WEEKLY DETRENDED DATA AND ARMA(2,1) MODEL



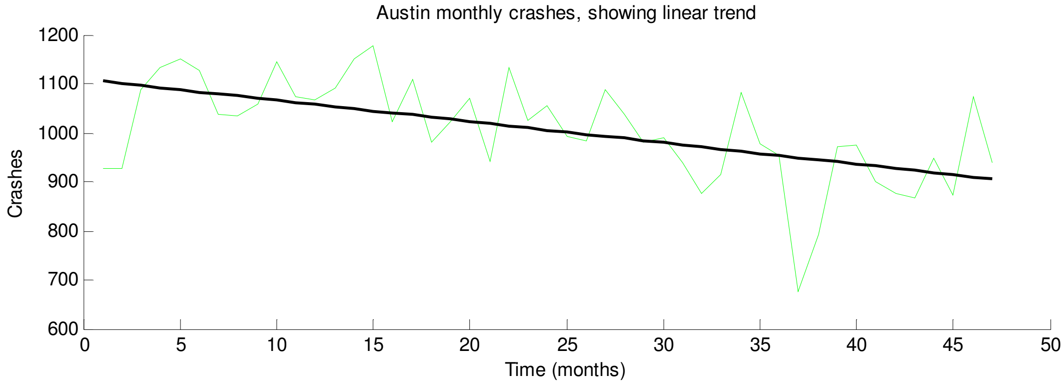
AUSTIN WEEKLY MODEL RESIDUAL AUTOCORRELATION FUNCTION



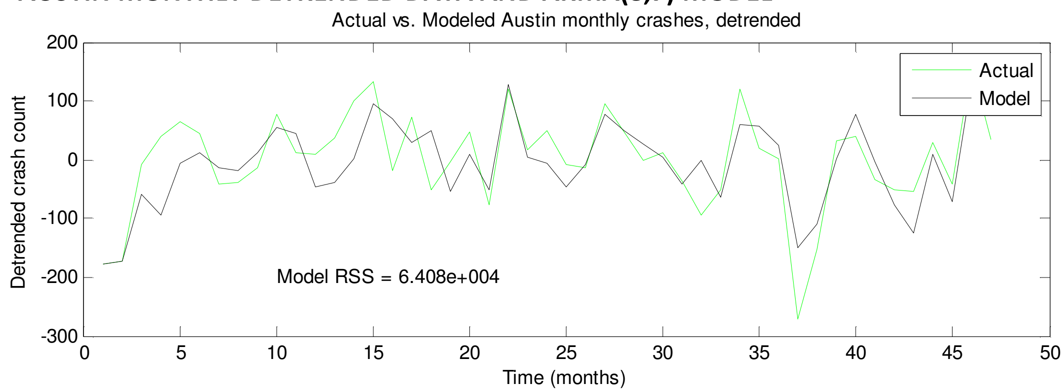
AUSTIN WEEKLY MODEL AUTOREGRESSIVE ROOTS



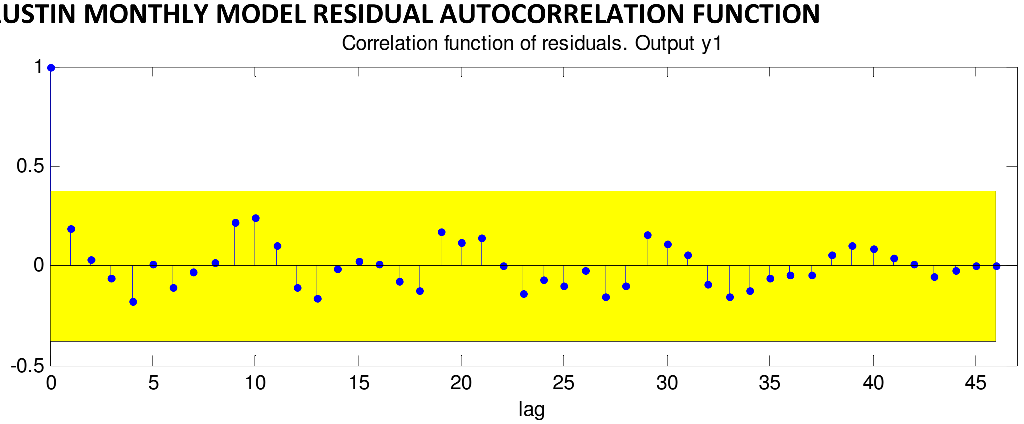
AUSTIN MONTHLY RAW DATA AND TREND



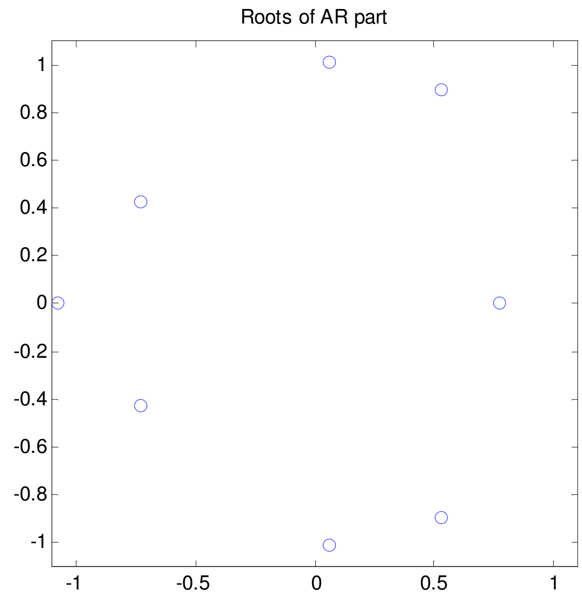
AUSTIN MONTHLY DETRENDED DATA AND ARMA(8,7) MODEL



AUSTIN MONTHLY MODEL RESIDUAL AUTOCORRELATION FUNCTION



AUSTIN MONTHLY MODEL AUTOREGRESSIVE ROOTS



5. VECTORIAL ARMA F-TESTING: TEXAS

x2 = output (crashes) Texas, 2008-10

x1 = input (gas price)

subscripts = row, col, delay

	Optimizing x2	**Model order**						
	Parameters	**(2,1)**	**(4,3)**	**(6,5)**	**(8,7)**	**(10,9)**	**(7,6)**	**(8,6)**
x2 terms	phi221	-0.94582	0.794306	-0.5168	-1.85962	-0.1004	-1.26044	0.038332
	phi222	0.108941	0.690734	0.179841	1.964175	-0.2434	0.854597	-0.83176
	phi223		0.29434	-0.25733	-1.48439	0.840343	-0.31618	-0.29682
	phi224		-0.157968	-0.27196	0.407134	0.182324	0.030882	-0.4576
	phi225			0.271373	0.275864	0.205997	0.381715	0.41562
	phi226			-0.02679	-0.69767	0.453547	-0.28129	0.609613
	phi227				0.678202	-0.45689	-0.13766	-0.13326
	phi228				-0.07588	0.368067		0.055383
	phi229					0.224833		
	phi2210					-0.23268		
x1 terms	phi211	-386.055	-826.6637	-1061.46	-681.735	-868.166	-610.785	-832.034
	phi212	385.2658	259.5809	1535.442	1922.197	1445.67	2015.11	1431.075
	phi213		-295.4835	-1077.93	-3287.52	-353.688	-2377.99	29.6125
	phi214		842.2111	1389.947	4568.519	-573.935	2402.299	170.7565
	phi215			-1082.88	-4626.55	-1.54737	-2373.01	-691.518
	phi216			292.7548	4117.882	19.78648	1600.975	-835.805
	phi217				-3270.35	532.9491	-418.33	817.4558
	phi218				1255.649	218.394		41.28626
	phi219					-1228.92		
	phi2110					786.0997		
a2 terms	theta21	-0.6337	1.287158	-0.12337	-1.54914	0.342347	-0.87148	0.446525
	theta22		1.272468	0.115442	1.431731	-0.12859	0.551073	-0.73819
	theta23		0.820431	-0.13151	-0.90371	0.926939	-0.03327	-0.58897
	theta24			-0.35683	-0.09359	0.65144	-0.09943	-0.71779
	theta25			0.498727	0.778478	0.736477	0.746832	0.430388
	theta26				-1.13543	0.869333	-0.4739	0.941946
	theta27				0.85071	-0.24749		
	theta28					0.50896		
	theta29					0.797141		
	RSS	8.59E+07	6.94E+07	6.27E+07	5.31E+07	5.17E+07	6.01E+07	5.50E+07
	gamma(a22)	3.44E+05	2.91E+05	2.77E+05	2.47E+05	2.53E+05	2.72E+05	2.55E+05
	F-test		(2,1) vs (4,3)	(4,3) vs (6,5)	(6,5) to (8,7)	(8,7) to (10,9	(7,6) to (8,7)	(8,6) to (8,7)
	s		6	6	6	6	3	1
	N		261	261	261	261	261	261
	r		12	18	24	30	24	24

F	9.8872	4.3021	7.1541	1.0836	12.9202	8.5403
F95%	2.135099	2.136008	2.136963	2.137968	2.64269	3.880995

	Optimizing x1	**Model order**		
	Parameters	**(2,1)**	**(4,3)**	**(2,0)**
x2 terms	phi120	-6.10E-06	-1.02E-06	-4.95E-06
	phi121	9.90E-07	1.36E-06	-2.78E-06
	phi122	1.34E-05	1.15E-05	1.34E-05
	phi123		-2.19E-05	
	phi124		1.31E-05	
x1 terms	phi111	-1.7234	-3.312719	-1.40644
	phi112	0.724073	4.501663	0.407319
	phi113		-2.965068	
	phi114		0.776393	
a2 terms	theta11	-0.39914	-2.018281	
	theta12		1.751858	
	theta13		-0.500605	
	theta14			
	theta15			
	theta16			
	theta17			
	theta18			
	theta19			
	RSS	1.7026	1.6215	1.7541
	gamma(a22)	0.0069	0.0069	0.007
	F-test		(2,1) vs (4,3)	(2,0) vs (2,1)
	s		6	1
	N		261	261
	r		12	6
	F		2.0756	7.7132
	F95%		2.135099	3.878184