Bike Sharing
Time Series Analysis

Group Members:

Abby(Jingyi) Liu

Qianying Diao

Qiang Wang

Xinrong Chen

Jamie(Xiaojie) Pan

Agenda

Introduction and Overview

Univariate Models

Multivariate Models

TF Model with regressors

Vector Model

Part 1

Introduction & Overview

Introduction

Background Overview

- **Bike Sharing System** is an innovation of traditional vehicles rental services.
- **Daily counts** of rental bikes from **2011.1.1 to 2012.12.31** in Capital Bikeshare System in Washington D.C..
- Also includes daily weather, humility, wind speed, precipitation.

Variable Description

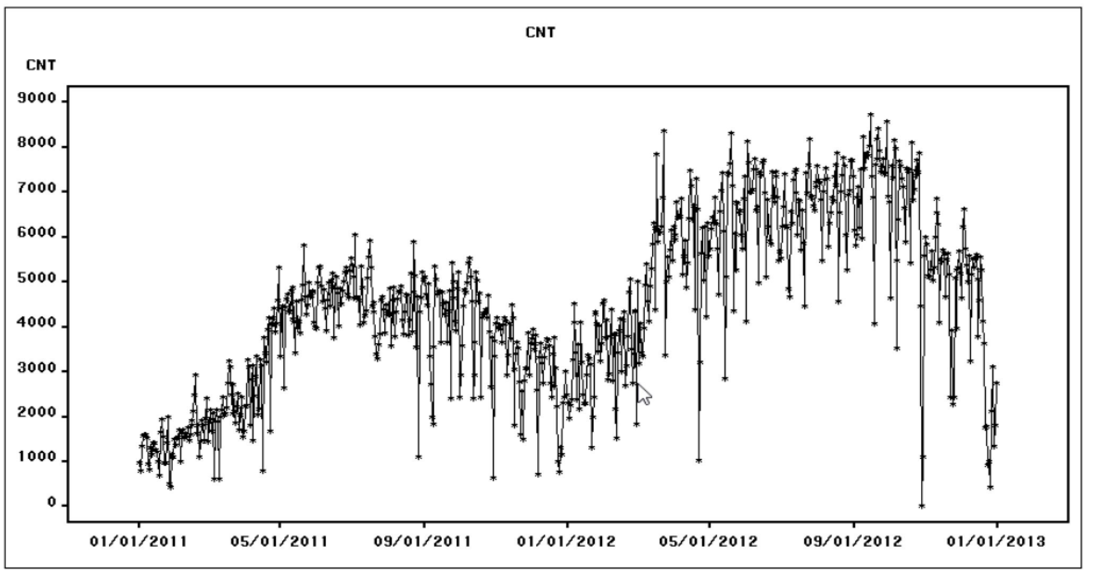
Dependent Variable

- **cnt:** Count of total rental bikes trips including both casual and registered

Independent Variables

- **atemp:** Normalized feeling temperature in Celsius.
- **hum:** Normalized humidity.
- **windspeed:** Normalized wind speed.
- **prcp:** historical daily precipitation observations.
- **start_holiday:** dummies variable, 1 for previous day is working day current day is holiday.
- **end_holiday :** dummies variable, 1 for previous day is holiday current day is work day.

Introduction



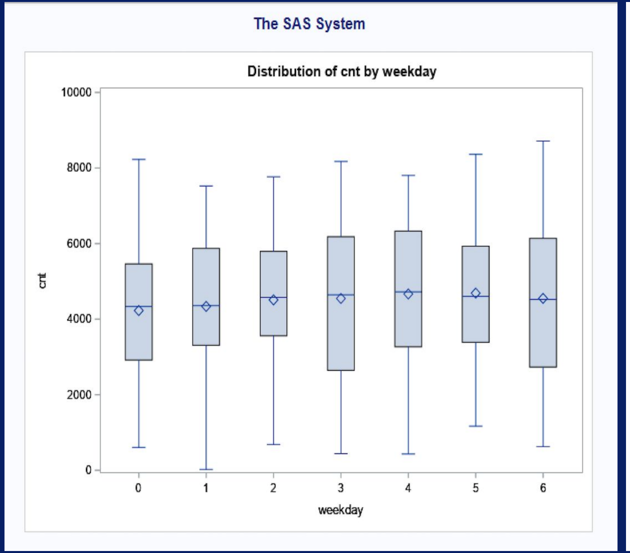
Features:

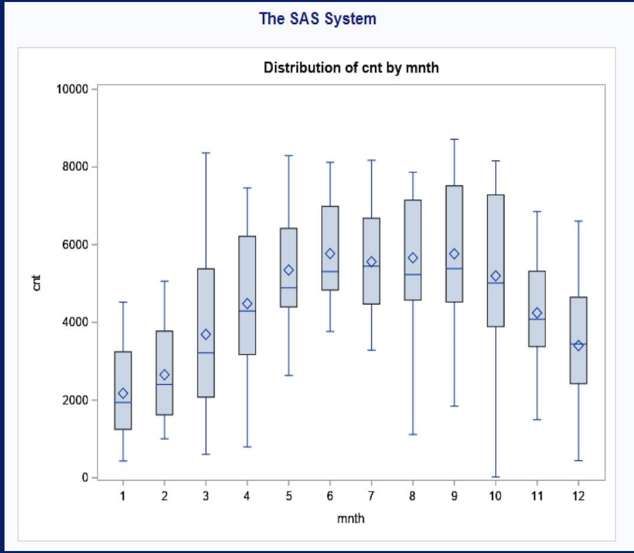
- Nonstationary series
    - Need first difference
- Increasing change in the variability of the series over time
    - a log transformation can be applied
- Clear seasonality

Part 2

Univariate Models

Univariate Models - Seasonal dummies

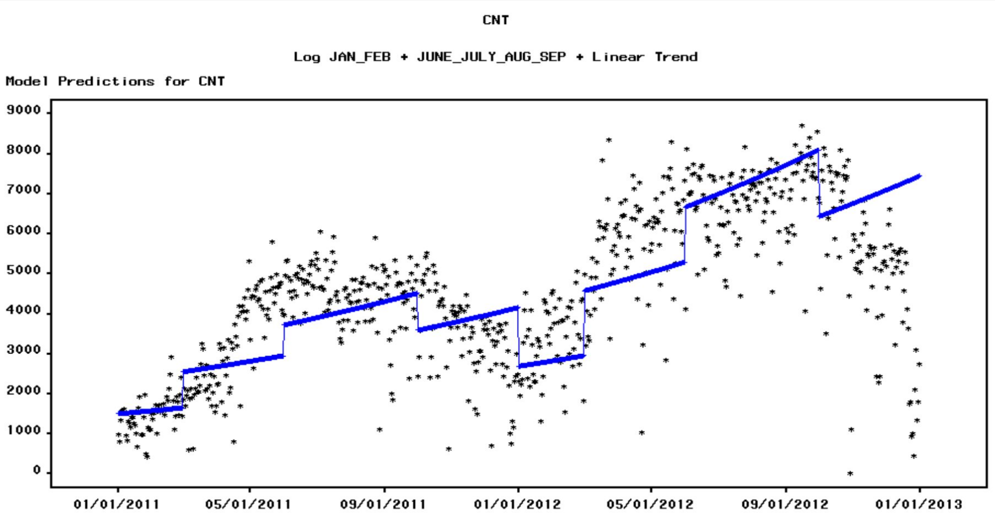




- Create two dummies for peak and bottom months
- ‘JAN_FEB’ dummies distinguish Jan. and Feb with other months
- ‘JUNE_JULY_AUG_SEP’ dummies distinguish June, July, August, and September with other months.

- Weekdays: No significant changes --- No model should be built
- Months: Peak: June to September. Bottom: January and February -- Create dummies

Univariate Models - Seasonal dummies



Model Prediction

The model generally grasp most of the ups and downs.

Parameter Estimate

All parameters are significant

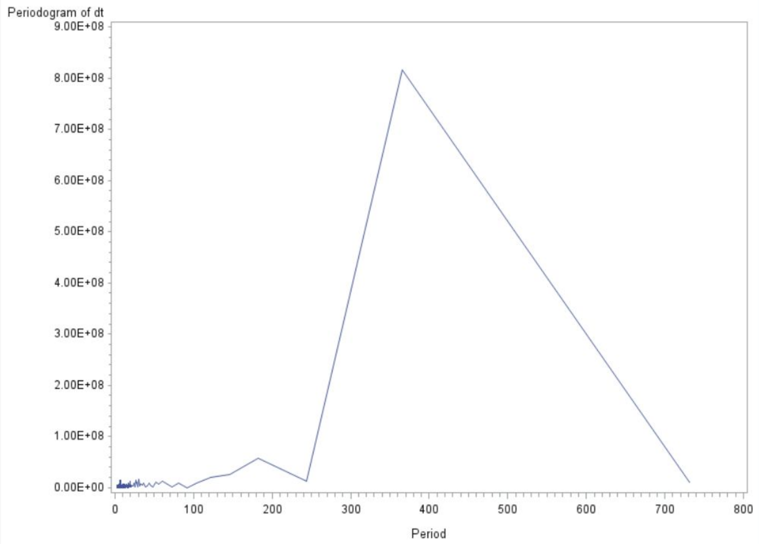
Parameter Estimates

CNT

Log JAN_FEB + JUNE_JULY_AUG_SEP + Linear Trend

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	7.75107	0.0311	249.4030	<.0001
JAN_FEB	-0.43567	0.0376	-11.5796	<.0001
JUNE_JULY_AUG_SEP	0.22996	0.0307	7.4814	<.0001
Linear Trend	0.00159	0.000079	20.2492	<.0001
Model Variance (sigma squared)	0.11403	.	.	.

Univariate Models - Cyclical

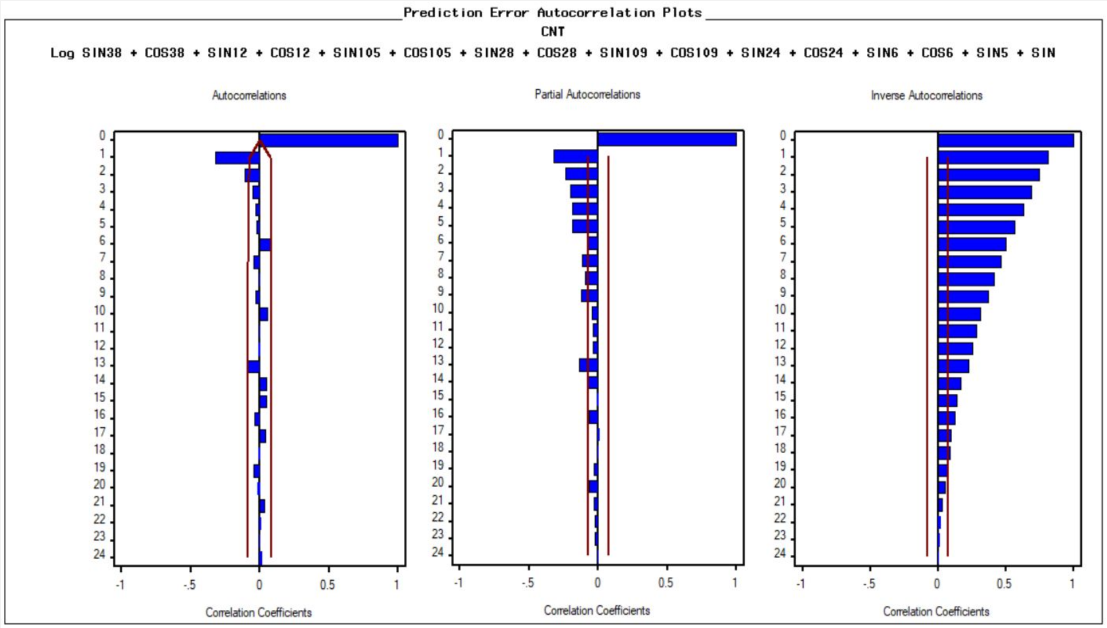


1	Obs	FREQ	PERIOD	P_01
2	**3**	0.01719	365.5	816186123
3	**5**	0.03438	182.75	58007683.1
4	**6**	0.04298	146.2	25420661.9
5	**7**	0.05157	121.833	20569437.7
6	**25**	0.20629	30.458	15529566
7	**110**	0.93689	6.706	15052975.3
8	**29**	0.24067	26.107	14403743.3
9	**106**	0.90251	6.962	14288650.9
10	**13**	0.10314	60.917	13474732.8
11	**39**	0.32662	19.237	12266775.6

Periodogram → Highest 10 amplitudes →

- Non-stationary
- Not White Noise
- **RMSE = 2615.7**

Univariate Models - Cyclical + First Difference



Take First Difference

- Stationary
- Not WN
- **RMSE = 1310.7**

- ACF decays quickly
- PACF decays quickly
- ARMA(1,1) for Error Model

Univariate Models - Cyclical + First Difference + Error Model

Statistic of Fit

Statistic of Fit	Value
Mean Square Error	1556499.9
Root Mean Square Error	1247.6
Mean Absolute Percent Error	245.70072
Mean Absolute Error	992.69333
R-Square	0.595

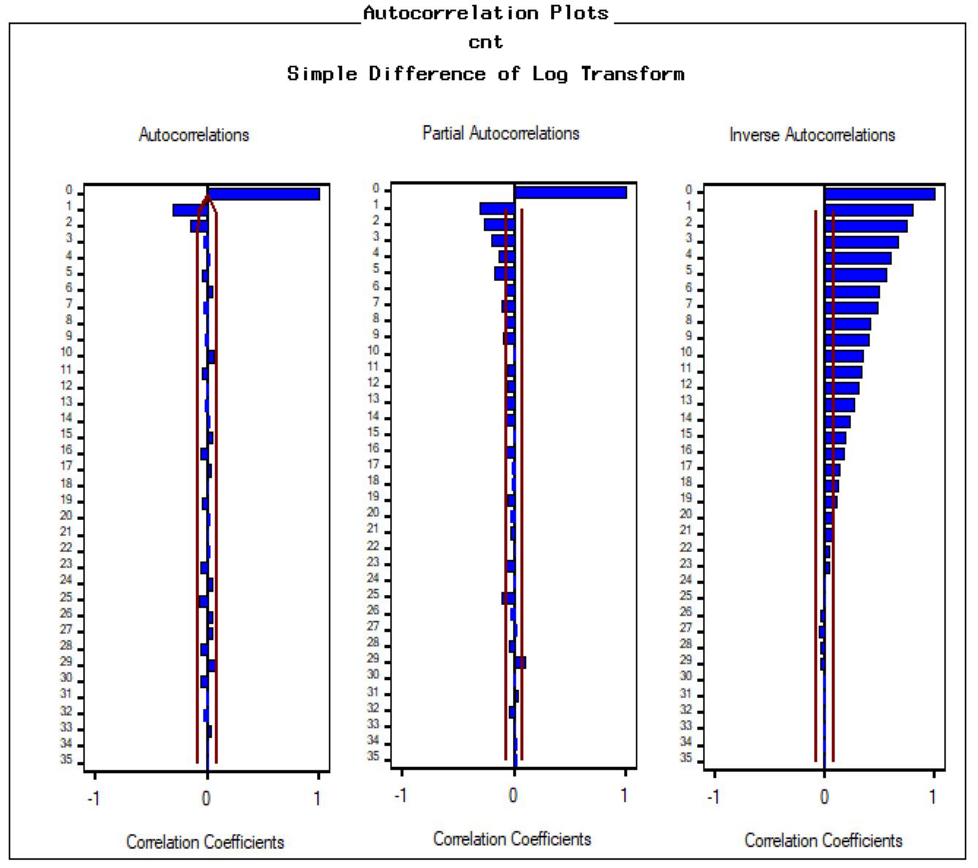
Parameter Estimate

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Intercept	0.00188	0.000512	3.6719	0.0004
Moving Average, Lag 1	0.96732	0.0113	85.3323	<.0001
Autoregressive, Lag 1	0.24225	0.0415	5.8306	<.0001
SIN38	0.03204	0.0194	1.6549	0.1020
COS38	-0.00183	0.0194	-0.0946	0.9249
SIN12	0.00992	0.0208	0.4777	0.6342
COS12	-0.02415	0.0208	-1.1625	0.2486
SIN105	0.02663	0.0172	1.5482	0.1257
COS105	-0.00443	0.0172	-0.2575	0.7975
SIN28	0.01024	0.0198	0.5179	0.6060
COS28	0.02786	0.0197	1.4126	0.1618
SIN109	0.03947	0.0171	2.3142	0.0233
COS109	0.02318	0.0171	1.3597	0.1779
SIN24	0.03396	0.0199	1.7088	0.0915
COS24	-0.02746	0.0199	-1.3816	0.1711
SIN6	0.03006	0.0251	1.1974	0.2348
COS6	-0.01871	0.0251	-0.7447	0.4587
SIN5	0.02608	0.0277	0.9434	0.3484
SIN4	-0.02891	0.0321	-0.9011	0.3703
COS5	0.05556	0.0270	2.0605	0.0427
COS4	-0.08702	0.0290	-2.9966	0.0037
SIN2	0.02967	0.0480	0.6177	0.5386
COS2	-0.40753	0.0483	-8.4422	<.0001
Model Variance (sigma squared)	0.07276	.	.	.

- **Lowest RMSE = 1247.6**
- Low MAPE = 245.7
- Pass WN test
- Significant p-values
  - COS 2 - 1 year
  - COS 4 - 1/2 year
  - COS 109 - 1 week

ARMA(1,1) fits well !

Univariate Models - ARIMA



Original series (CNT) is nonstationary

Take first difference

CNT(1) is stationary

Univariate Models - ARIMA

ARIMA(1,1,1)

Statistic of Fit	Value
Mean Square Error	1736875.0
Root Mean Square Error	1317.9
Mean Absolute Percent Error	302.29158
Mean Absolute Error	940.46107
R-Square	0.548

ARIMA(0,1,2)

Statistic of Fit	Value
Mean Square Error	1756347.5
Root Mean Square Error	1325.3
Mean Absolute Percent Error	302.30539
Mean Absolute Error	953.34912
R-Square	0.543

- Out-out samples - 100 obs
- Log Transformation
- Without intercept
- No obvious seasonality

ARIMA(1,1,1) is better!

Univariate Models - Comparison

Model	Model Variance	Root Mean Square Error	Mean Absolute Percent Error
Seasonal Dummy	0.11403	2651.9	388.38455
Cyclical + ARIMA(1,1,1)	0.07276	1247.6	245.70072
ARIMA(0,1,2)	0.07633	1325.3	302.30539
ARIMA(1,1,1)	0.07666	1317.9	302.29158

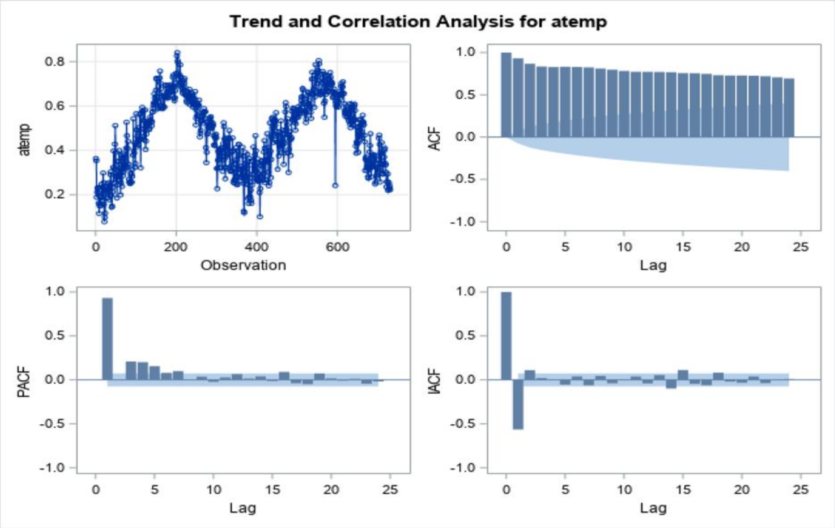
*All models here take the log-transformation.

Part 3

Multivariate Models

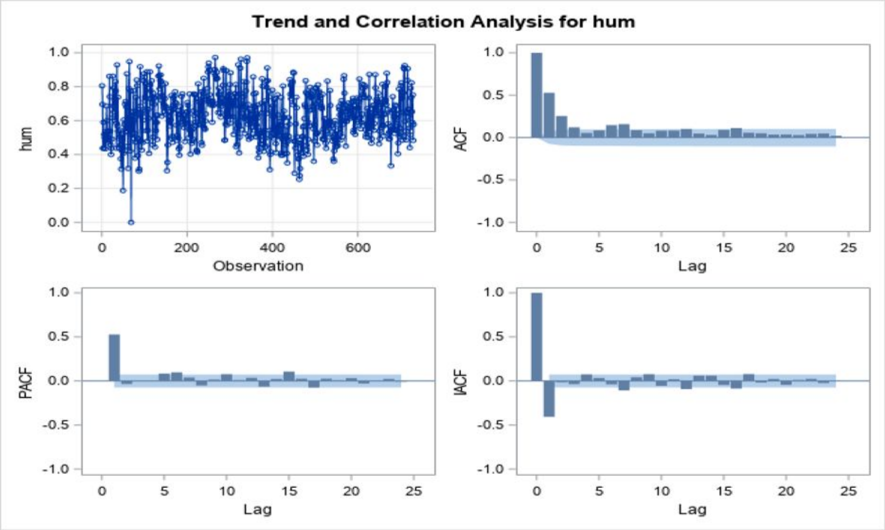
Multivariate Models - Preview Inputs

Average Temperature



Nonstationary

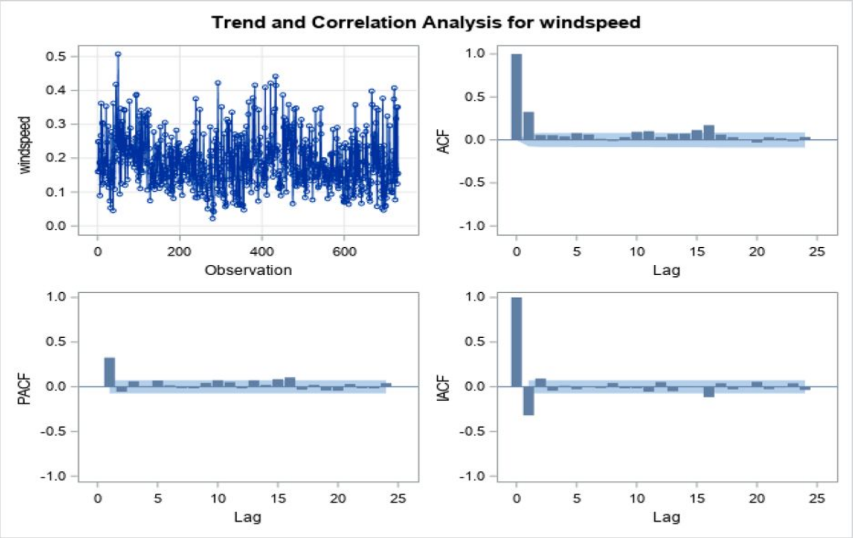
Humidity



Stationary, NOT White Noise

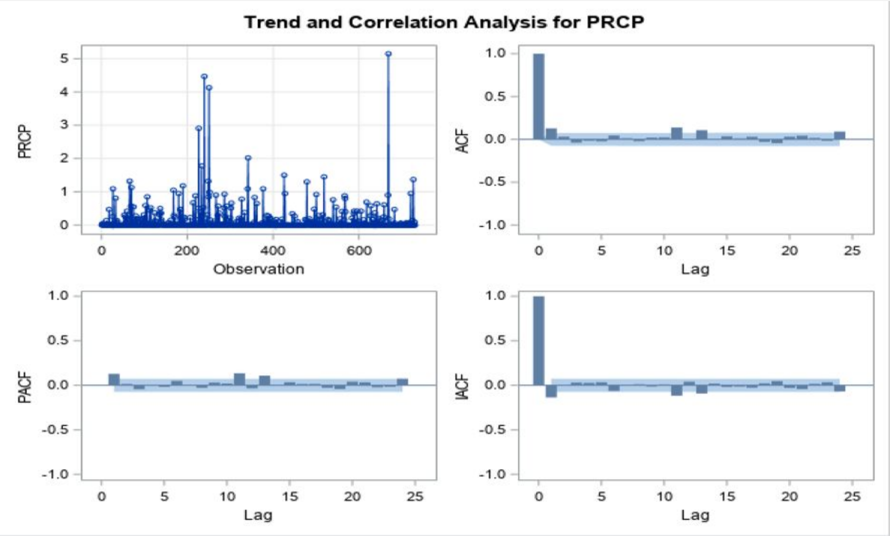
Multivariate Models - Preview Inputs

Windspeed



Stationary, NOT White Noise

Precipitation



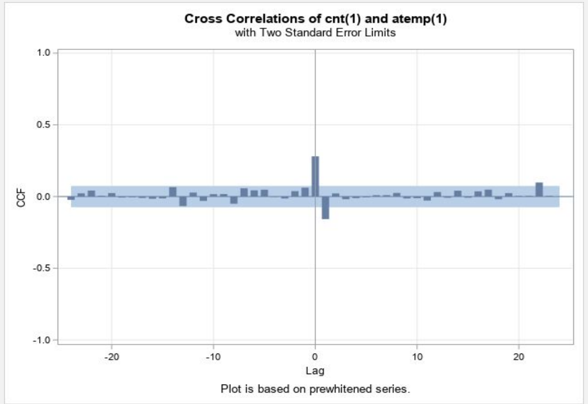
Stationary, NOT White Noise

Multivariate Models - Prewhitening Inputs

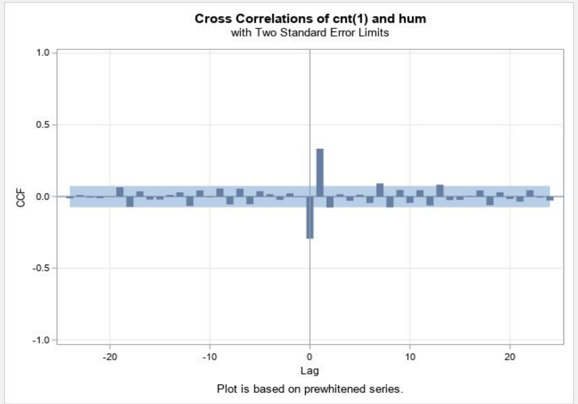
Input Variables	Prewhitening Process
ATEMP	ARIMA(1,1,2)
HUM	ARIMA(1,0,1)(0,0,1)s
WINDSPEED	Factor Model, Q = (1,16)
PRCP	Factor Model, Q = (1,11)

Multivariate Models - Cross-correlation analysis

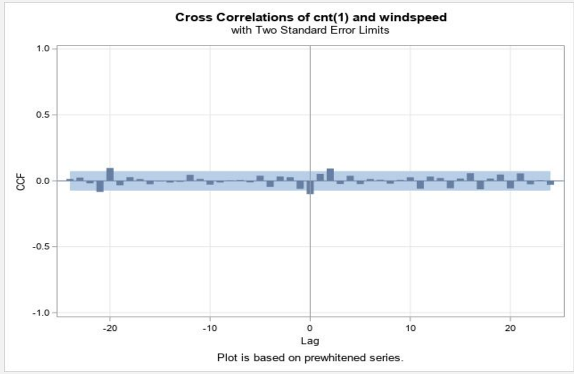
Average Temperature



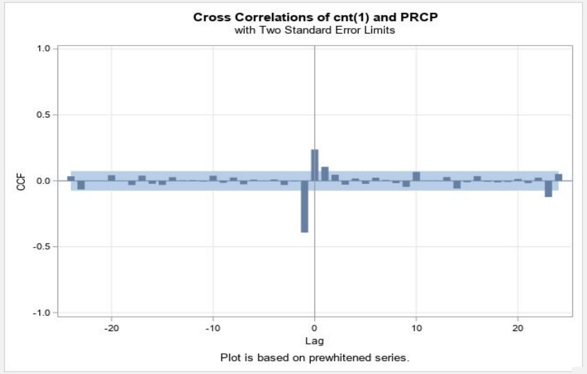
Humidity



Windspeed



Precipitation



Multivariate Models - Estimate TF model

Prewhitened Variables	(b, r, s)
ATEMP	b = 0, r = 0, s = 1
HUM	b = 0, r = 0, s = 2
WINDSPEED	b = 0, r = 0, s = 1
PRCP	A significant negative lag

Exclude from TF model

Multivariate Models - Check for adequacy(univariate)

Crosscorrelation Check of Residuals with Input atemp									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	1.27	5	0.9377	-0.005	-0.011	0.000	-0.008	-0.029	-0.026
11	3.08	11	0.9896	0.019	-0.002	0.035	-0.024	-0.012	-0.012
17	4.38	17	0.9991	0.020	0.000	0.020	0.013	0.002	0.028
23	10.45	23	0.9881	-0.008	0.003	-0.014	-0.011	0.089	-0.008
29	20.26	29	0.8846	-0.024	-0.020	0.047	-0.091	0.044	-0.006
35	33.03	35	0.5634	0.001	-0.066	0.100	-0.019	-0.053	0.012
41	36.76	41	0.6595	0.011	-0.038	0.023	0.042	-0.033	-0.016
47	40.29	47	0.7448	-0.001	0.038	0.029	-0.032	0.031	-0.022

Crosscorrelation Check of Residuals with Input windspeed									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	9.34	5	0.0963	0.000	-0.026	0.101	0.010	0.043	-0.002
11	12.81	11	0.3059	0.005	0.008	-0.028	0.005	0.043	-0.045
17	22.63	17	0.1616	-0.000	0.047	-0.049	-0.004	0.073	-0.059
23	29.60	23	0.1612	0.016	0.064	-0.050	0.049	-0.009	-0.010
29	38.64	29	0.1088	-0.031	0.100	-0.027	-0.008	0.019	0.018
35	41.37	35	0.2123	0.007	0.050	-0.018	-0.004	-0.025	-0.014
41	44.09	41	0.3424	0.037	0.013	-0.022	0.009	-0.000	-0.041
47	47.40	47	0.4563	0.015	0.045	-0.038	0.025	0.006	0.014

Crosscorrelation Check of Residuals with Input hum									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	2.28	4	0.6852	-0.009	0.007	-0.001	0.018	-0.050	0.013
11	9.82	10	0.4562	-0.026	0.072	-0.057	0.016	-0.012	0.028
17	12.76	16	0.6903	-0.035	0.046	0.003	-0.006	0.016	0.020
23	18.42	22	0.6810	-0.044	-0.003	-0.004	-0.065	0.035	0.019
29	21.70	28	0.7948	-0.056	0.013	0.002	0.001	0.021	-0.027
35	26.92	34	0.8008	0.004	-0.003	0.027	-0.062	0.047	0.019
41	37.83	40	0.5684	-0.028	-0.019	0.036	-0.049	0.082	-0.059
47	41.98	46	0.6414	0.003	0.005	0.030	0.010	0.020	-0.065

P-value are larger than 0.05

Our target and input series are not cross-correlated

TF model is appropriate

Multivariate Models - Check for adequacy(multivariate)

Crosscorrelation Check of Residuals with Input atemp									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	1.35	5	0.9301	-0.011	-0.008	-0.025	0.000	-0.026	-0.019
11	4.89	11	0.9363	0.038	0.040	-0.001	-0.010	-0.041	-0.002
17	7.82	17	0.9703	0.027	-0.015	0.039	0.011	0.034	0.016
23	15.10	23	0.8910	-0.026	0.011	-0.032	0.025	0.084	0.024
29	30.17	29	0.4054	-0.070	-0.011	0.053	-0.093	0.059	-0.028
35	37.96	35	0.3358	0.015	-0.054	0.076	-0.011	-0.037	0.018
41	39.55	41	0.5352	-0.011	-0.018	0.029	0.018	-0.023	0.000
47	42.54	47	0.6575	-0.007	0.045	0.022	-0.012	0.032	-0.021

Crosscorrelation Check of Residuals with Input hum									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	3.13	4	0.5360	-0.011	0.007	0.003	0.019	-0.060	-0.013
11	13.68	10	0.1880	-0.036	0.092	-0.057	0.029	-0.020	0.019
17	16.55	16	0.4150	-0.051	0.021	-0.001	0.004	-0.011	0.027
23	22.04	22	0.4577	-0.052	-0.010	0.003	-0.053	0.044	-0.001
29	25.09	28	0.6228	-0.047	0.017	-0.022	0.016	0.010	-0.029
35	32.75	34	0.5286	0.018	-0.021	0.034	-0.080	0.045	0.012
41	48.83	40	0.1596	-0.057	-0.010	0.047	-0.066	0.078	-0.078
47	51.79	46	0.2582	0.032	0.020	0.008	-0.031	0.013	-0.037

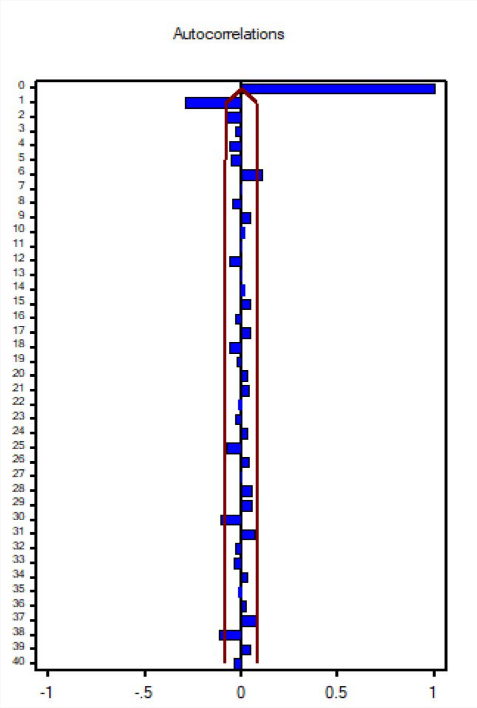
Crosscorrelation Check of Residuals with Input windspeed									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	20.76	5	0.0009	0.010	-0.047	0.146	-0.003	0.069	0.014
11	24.26	11	0.0117	0.000	0.004	-0.027	0.031	0.029	-0.048
17	28.89	17	0.0355	0.045	0.021	-0.031	0.004	0.047	-0.028
23	33.41	23	0.0743	0.028	0.054	-0.032	0.038	-0.004	-0.005
29	45.45	29	0.0266	-0.039	0.110	-0.031	-0.007	-0.020	0.040
35	50.63	35	0.0424	-0.012	0.070	-0.006	-0.011	0.005	-0.043
41	55.11	41	0.0694	0.023	0.041	-0.032	0.010	0.010	-0.052
47	59.33	47	0.1070	0.040	0.017	-0.036	0.034	-0.000	0.038

P-value are larger than 0.05

Our target and input series are not cross-correlated

TF model is appropriate

Multivariate Models - TF model



TF model

Not WN,
Try different
error models

TF + ARIMA(0,1,1)

Statistics of Fit

CNT

Log atemp[Dif(1) N(1)] + windspeed[N(1)] + hum[N(2)] + IMA(1,1) NOINT

Statistic of Fit	Value
Mean Square Error	1511930.6
Root Mean Square Error	1229.6
Mean Absolute Percent Error	210.73422
Mean Absolute Error	921.67957
R-Square	0.607

TF + ARIMA(1,1,1)

Statistics of Fit

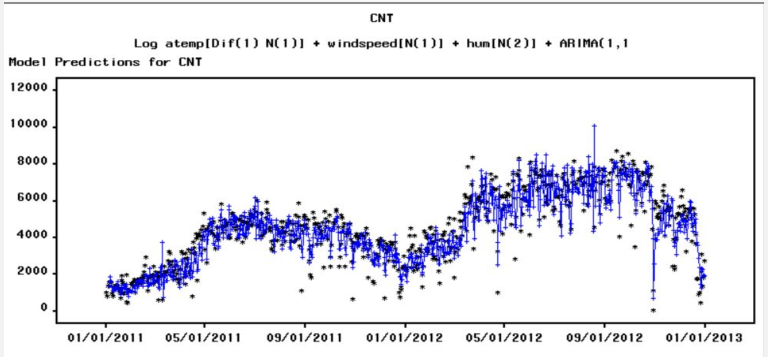
CNT

Log atemp[Dif(1) N(1)] + windspeed[N(1)] + hum[N(2)] + ARIMA(1,1,1) NOINT

Statistic of Fit	Value
Mean Square Error	1242800.8
Root Mean Square Error	1114.8
Mean Absolute Percent Error	202.25135
Mean Absolute Error	857.75356
R-Square	0.677

Comparison of TF-noise model

Multivariate Models - TF-noise model



Parameter Estimates

CNT

Log atemp[Dif(1) N(1)] + windspeed[N(1)] + hum[N(2)] + ARIMA(1,1,1) NOINT

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Moving Average, Lag 1	0.84890	0.0275	30.8383	<.0001
Autoregressive, Lag 1	0.20554	0.0505	4.0682	0.0001
ATEMP[Dif(1) N(1)]	1.08876	0.1513	7.1942	<.0001
ATEMP[Dif(1) N(1)] Num1	-0.39755	0.1473	-2.6984	0.0083
WINDSPEED[N(1)]	-0.86002	0.1338	-6.4267	<.0001
WINDSPEED[N(1)] Num1	0.49965	0.1425	3.5053	0.0007
HUM[N(2)]	-1.00334	0.0794	-12.6372	<.0001
HUM[N(2)] Num1	-0.18655	0.0783	-2.3819	0.0193
HUM[N(2)] Num2	-0.11360	0.0765	-1.4854	0.1409
Model Variance (sigma squared)	0.05477	.	.	.

- Stationary
- Not WN
- **RMSE = 1114.8**
- **MAPE = 202.25135**

Model has been improved

Part 4

TF Model with Regressors

TF Model with regressors - Regressors Detection

Feedback Relationships with PRCP

- **Significant Regressors**:
  PRCP, PRCP of Tomorrow(k = -1)

- **Interpretation:**
  People can check tomorrow's weather. The decision today will depend on precipitation volumes of yesterday, today and tomorrow.

  Before the inclement weather, the level of bike trips count is high.

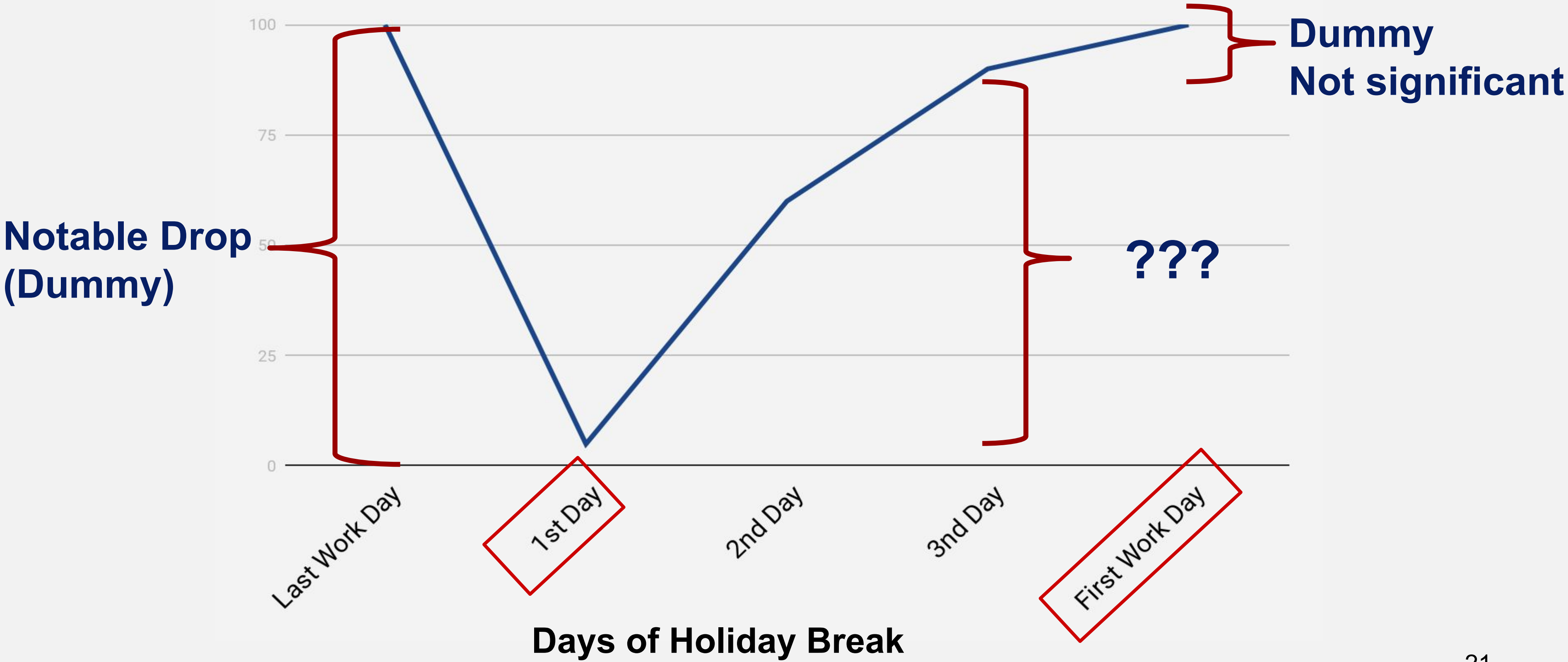
Holiday Effect

Holiday	2011		2012	
Martin Luther King Jr. Day	Jan. 15th	Jan. 18th	Jan. 14th	Jan. 17th
Valentine's Day		Feb. 14th		Feb. 14th
Presidents' Day	Feb. 19th	Feb. 22th	Feb. 18th	Feb. 21th
Labor Day	Sep. 3rd	Sep. 6th	Sep. 1st	Sep. 4th
Thanksgiving Day	Nov. 24th	Nov. 25th	Nov. 22th	Nov. 23th
Black Friday		Nov. 26th		Nov. 23th
Christmas Eve	Dec. 24th		Dec. 24th	

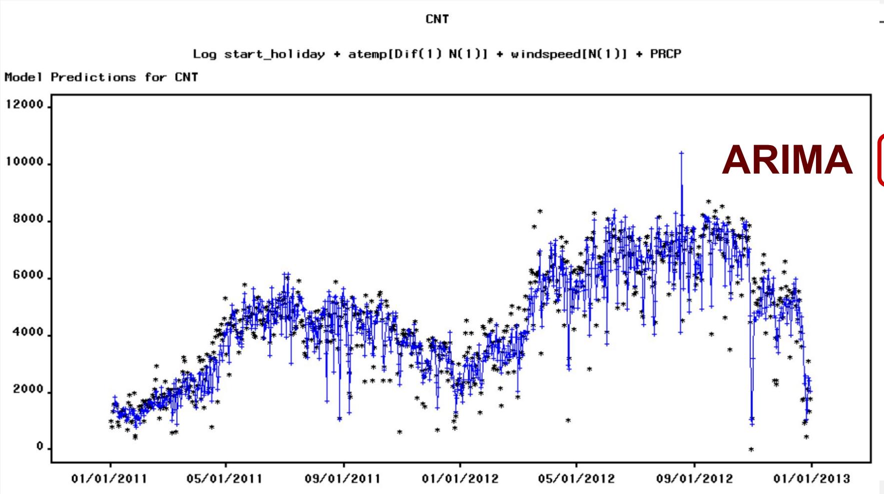
Outliers:
Notable deviation substantially higher or lower than predicted values in former ARIMA or TF models

Dummy Variables:
Break start days (Significantly Negative)
The first working day (Not Significant)

TF Model with regressors - Count of Bike Trips during Holiday



TF Model with regressors - Model Identification and Comparison



Parameter Estimates

CNT

Log start_holiday + atemp[Dif(1) N(1)] + windspeed[N(1)] + PRCP + PRCP_LAG + hum[N(2)] +

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	0.00283	0.0016	1.7485	0.0839
Moving Average, Lag 1	0.85473	0.0274	31.2373	<.0001
Autoregressive, Lag 1	0.24369	0.0508	4.7986	<.0001
start_holiday	-0.21953	0.0688	-3.1915	0.0020
ATEMP[Dif(1) N(1)]	1.01249	0.1352	7.4876	<.0001
ATEMP[Dif(1) N(1)] Num1	-0.40382	0.1318	-3.0632	0.0029
WINDSPEED[N(1)]	-0.61818	0.1222	-5.0581	<.0001
WINDSPEED[N(1)] Num1	0.33574	0.1299	2.5855	0.0114
PRCP	-0.05024	0.0244	-2.0563	0.0428
PRCP_LAG	-0.29191	0.0243	-12.0110	<.0001
HUM[N(2)]	-0.69635	0.0767	-9.0834	<.0001
HUM[N(2)] Num1	-0.18074	0.0730	-2.4753	0.0153
HUM[N(2)] Num2	-0.13229	0.0686	-1.9274	0.0572
Model Variance (sigma squared)	0.04379	.	.	.

TF

Regressors

	TF Model	TF Model with regressor
Root Mean Square Error	1114.8	891.87
Mean Absolute Percent Error	202.25	68.99

Part 5

Vector Model

Vector Model

Cross Correlations

Schematic Representation of Cross Correlations													
Variable/Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
cnt_delta_std	++	--	-.	-.	..	..	+.	..	..	..	..	..	..
PRCP1	++	++	..	..	..	..	..	..	..	..	..	.+	..
+ is > 2*std error, - is < -2*std error, . is between													

Partial Autoregression

Schematic Representation of Partial Autoregression												
Variable/Lag	1	2	3	4	5	6	7	8	9	10	11	12
cnt_delta_std	-+	-+	-.	-+	-+	.+	..	..	..	..	..	..
PRCP1	-+	-+	-.	-.	-.	-+	-.	..	..	..	.+	..
+ is > 2*std error, - is < -2*std error, . is between												

VMA(2) Parameter Estimates

Model Parameter Estimates						
Equation	Parameter	Estimate	Standard Error	t Value	Pr > |t|	Variable
cnt_delta_std	MA1_1_1	0.55096	0.03353	16.43	0.0001	e1(t-1)
	MA1_1_2	-0.05861	0.03596	-1.63	0.1036	e2(t-1)
	MA2_1_1	0.25936	0.03202	8.10	0.0001	e1(t-2)
	MA2_1_2	-0.00144	0.03470	-0.04	0.9669	e2(t-2)
PRCP1	MA1_2_1	0.55967	0.03711	15.08	0.0001	e1(t-1)
	MA1_2_2	-0.09792	0.03719	-2.63	0.0086	e2(t-1)
	MA2_2_1	0.06901	0.04136	1.67	0.0957	e1(t-2)
	MA2_2_2	-0.03966	0.04289	-0.92	0.3555	e2(t-2)

Vector Model

Matrix Function

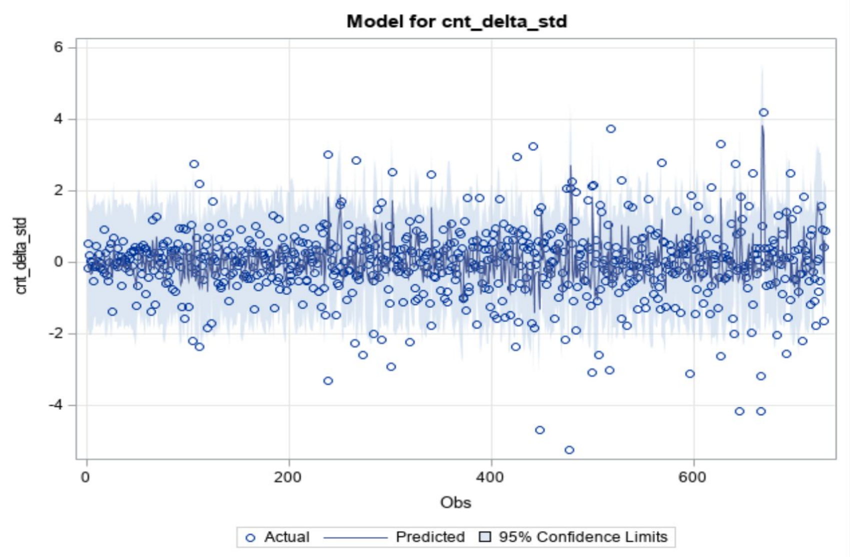
$$\begin{bmatrix} \Delta CNT\_STD \\ PRCP\_STD \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} - \begin{bmatrix} 0.55096 & -0.05861 \\ 0.55967 & -0.09792 \end{bmatrix} \times \begin{bmatrix} e_{1t-1} \\ e_{2t-1} \end{bmatrix} - \begin{bmatrix} 0.25936 & -0.00144 \\ 0.06901 & -0.03966 \end{bmatrix} \times \begin{bmatrix} e_{1t-2} \\ e_{2t-2} \end{bmatrix}$$

Correlations of Residuals

Schematic Representation of Cross Correlations

Variable/Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
cnt_delta_std	+.	..	..	-.	..	..	+.	..	..	..	..	..	.-
PRCP1	.+	..	..	..	..	..	.+	..	..	..	..	.+	..

+ is > 2*std error, - is < -2*std error, . is be



Thank you for listening !