Intervention Analysis In the Utility Industry

Using Different Time-Series Models

Fujiang Wen

City of Dallas



SOUTH CENTRAL SAS® USERS GROUP

Discussion Focus

- Dallas Water Utilities Quick Facts
- Retail Demand and Recent Restrictions
- Evaluation Methods and Analysis
- Summary

Dallas Water Utilities

- **System connections**

  Over 300,000 retail customers (residential, Industrial, commercial and institutional)
  Wholesale cities: 28

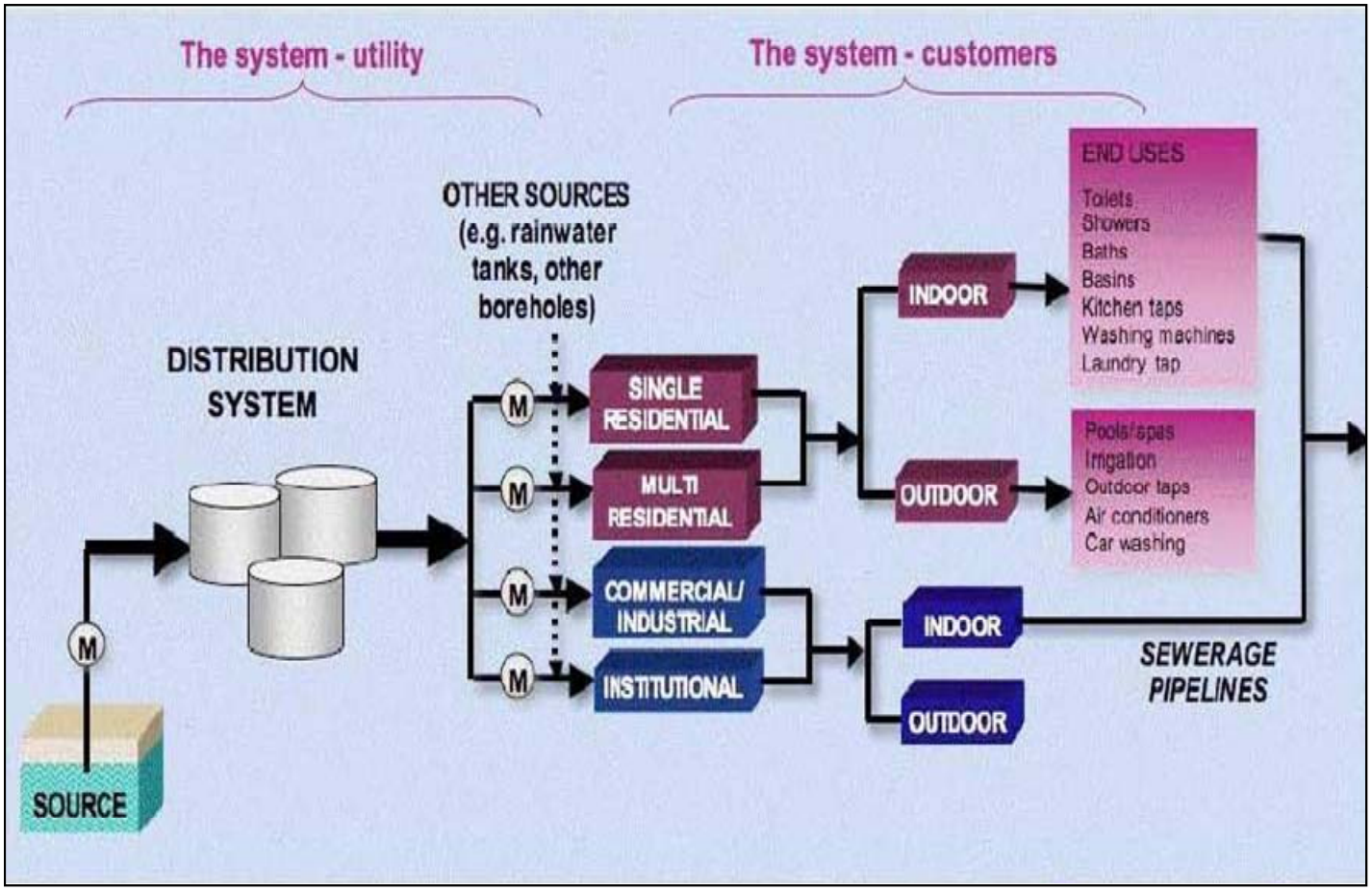
- **Population served**

  1,200,000 in retail
  1,229,000 in wholesales

- **Demand and revenue**

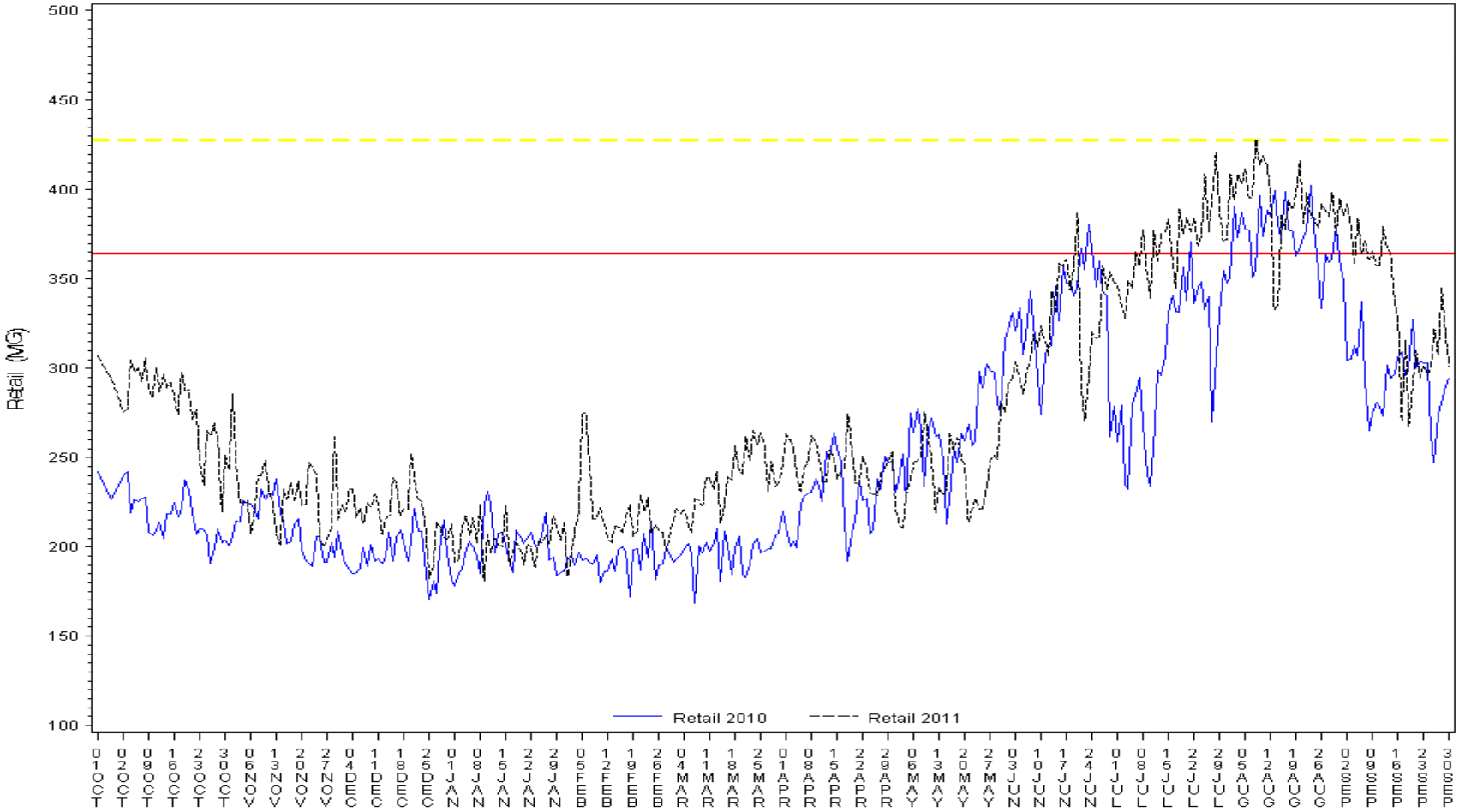
  Average daily demand of 428 MG
  Revenues of $355 million

Utilities System Flowchart



Source: Sam Kayaga, WEDC, End-use Analysis for Demand forecasting & Options Analysis

Historical Water Demand and 2012 System Restriction



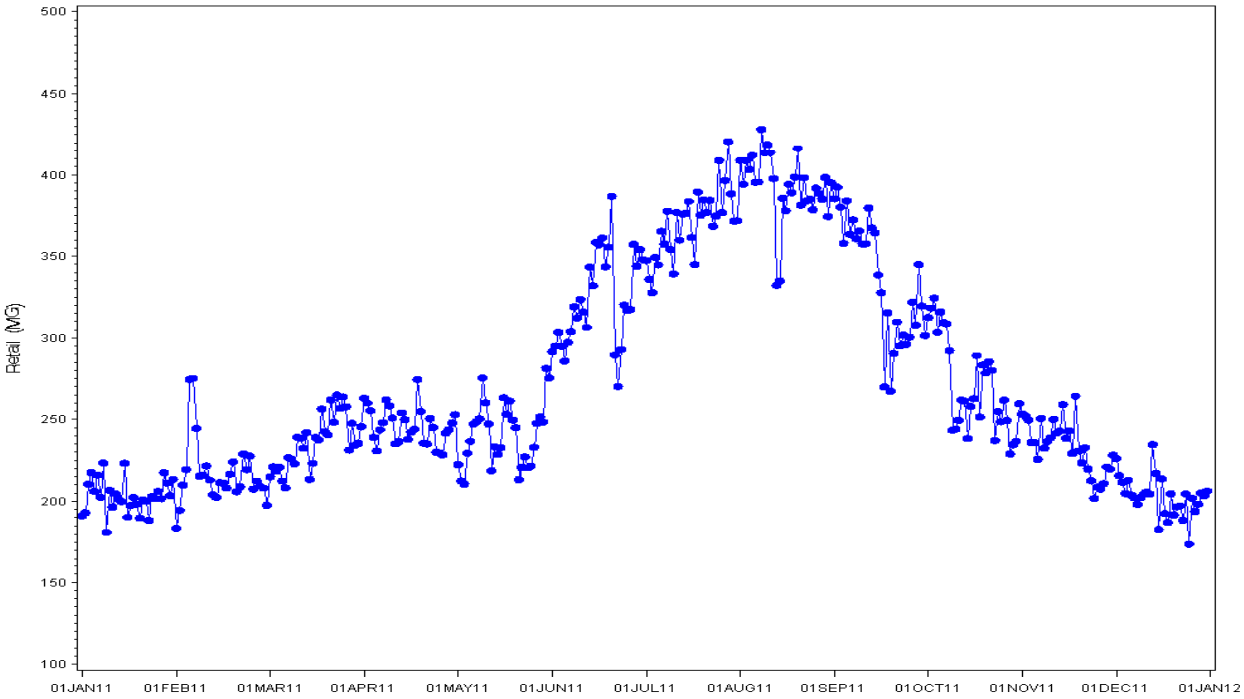
Water Restrictions in Action Since 2011

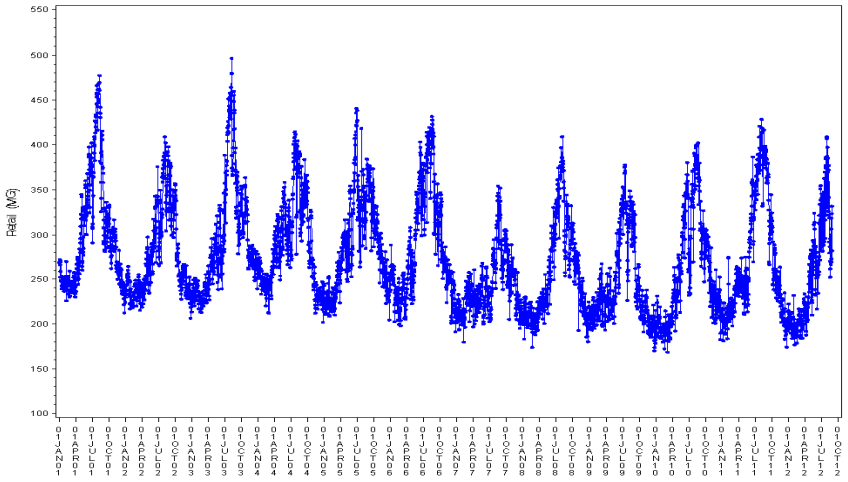


City of Dallas Maximum Twice Per Week Watering Schedule

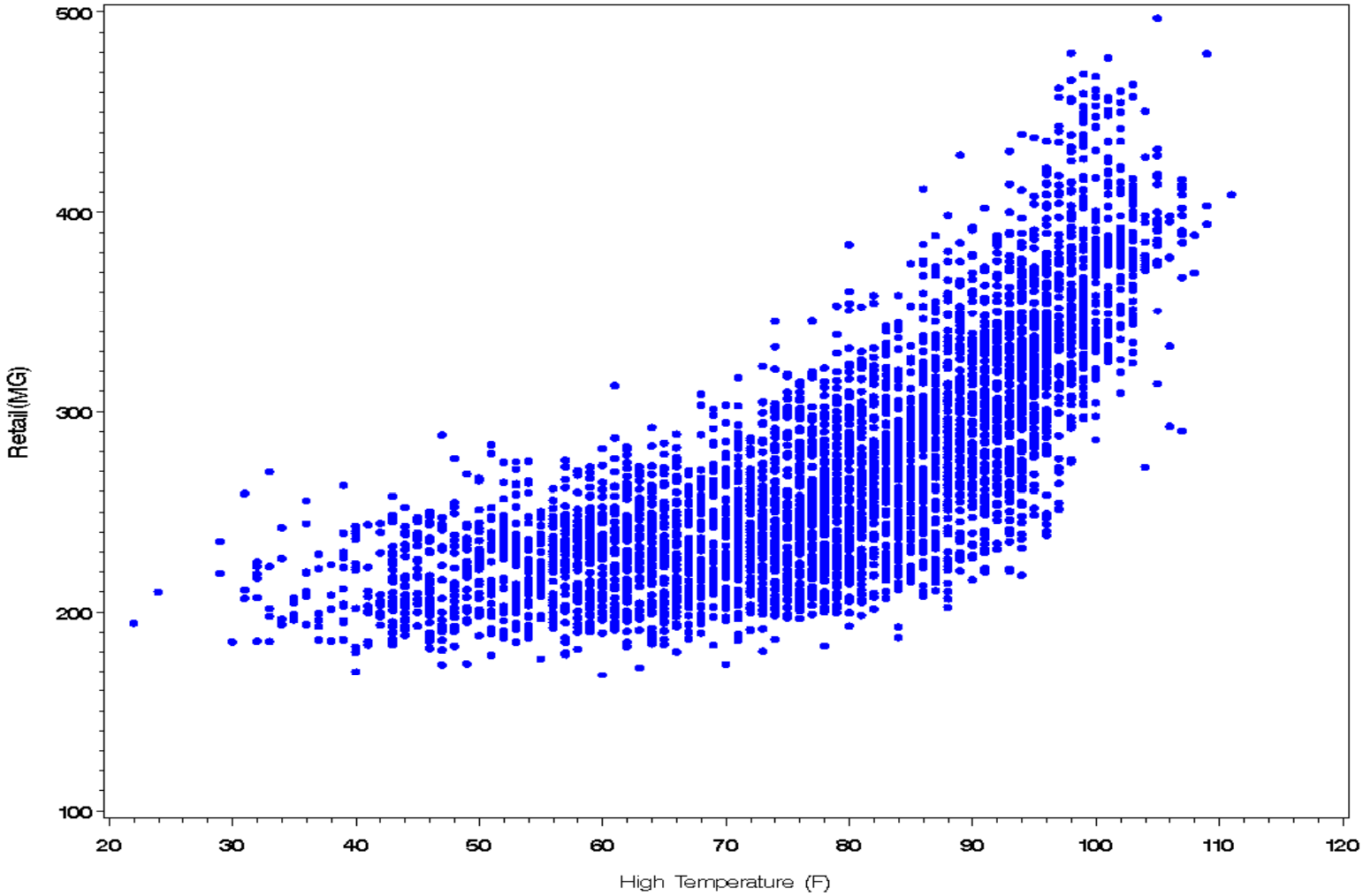
Last Digit of Address	Watering Days Allowed
0, 2, 4, 6 or 8 (even numbers)	Sundays and Thursdays
1, 3, 5, 7 or 9 (odd numbers)	Saturdays and Wednesdays
No address number	Sundays and Thursdays

Water Demand Seasonal Patterns

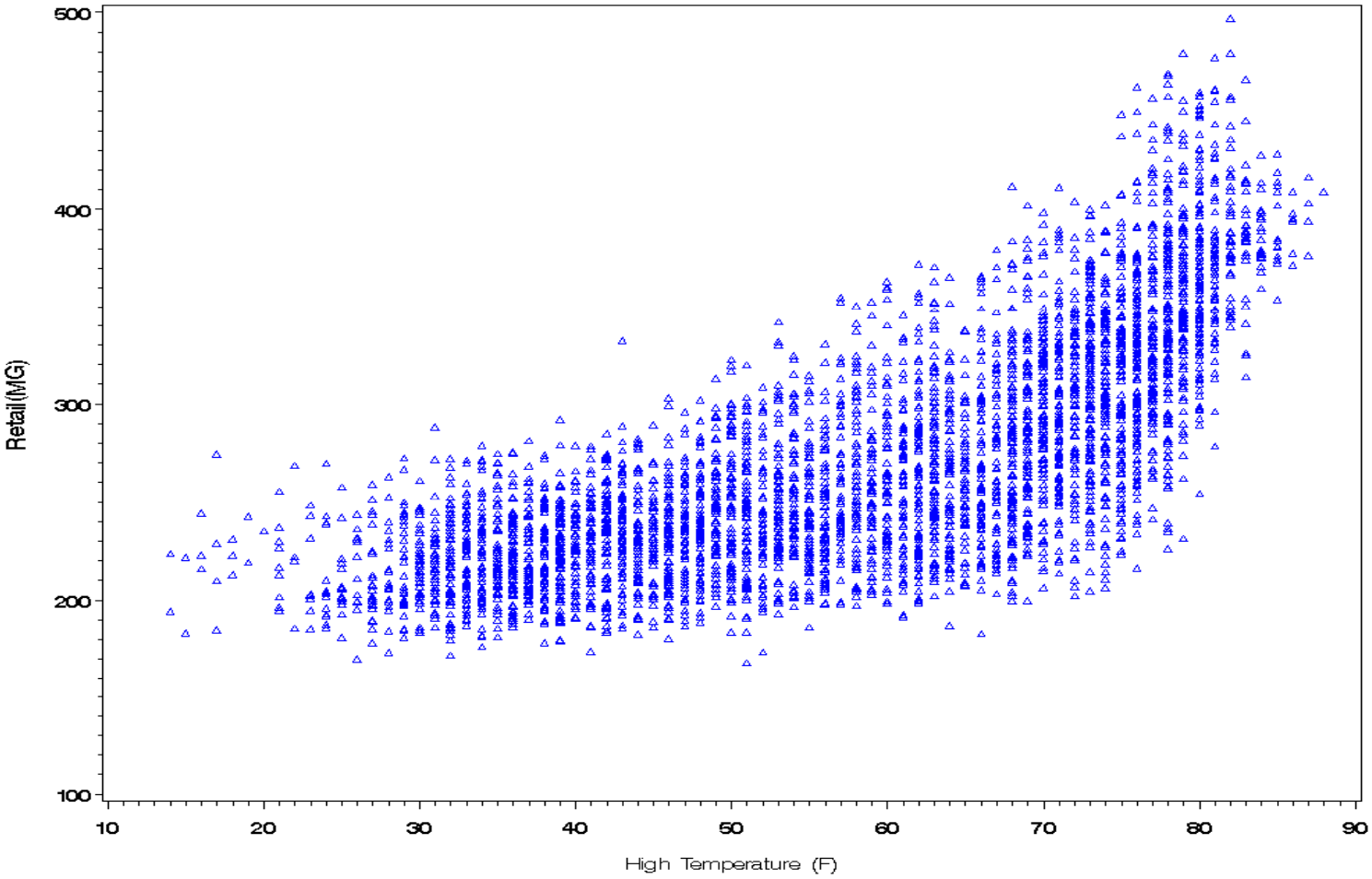




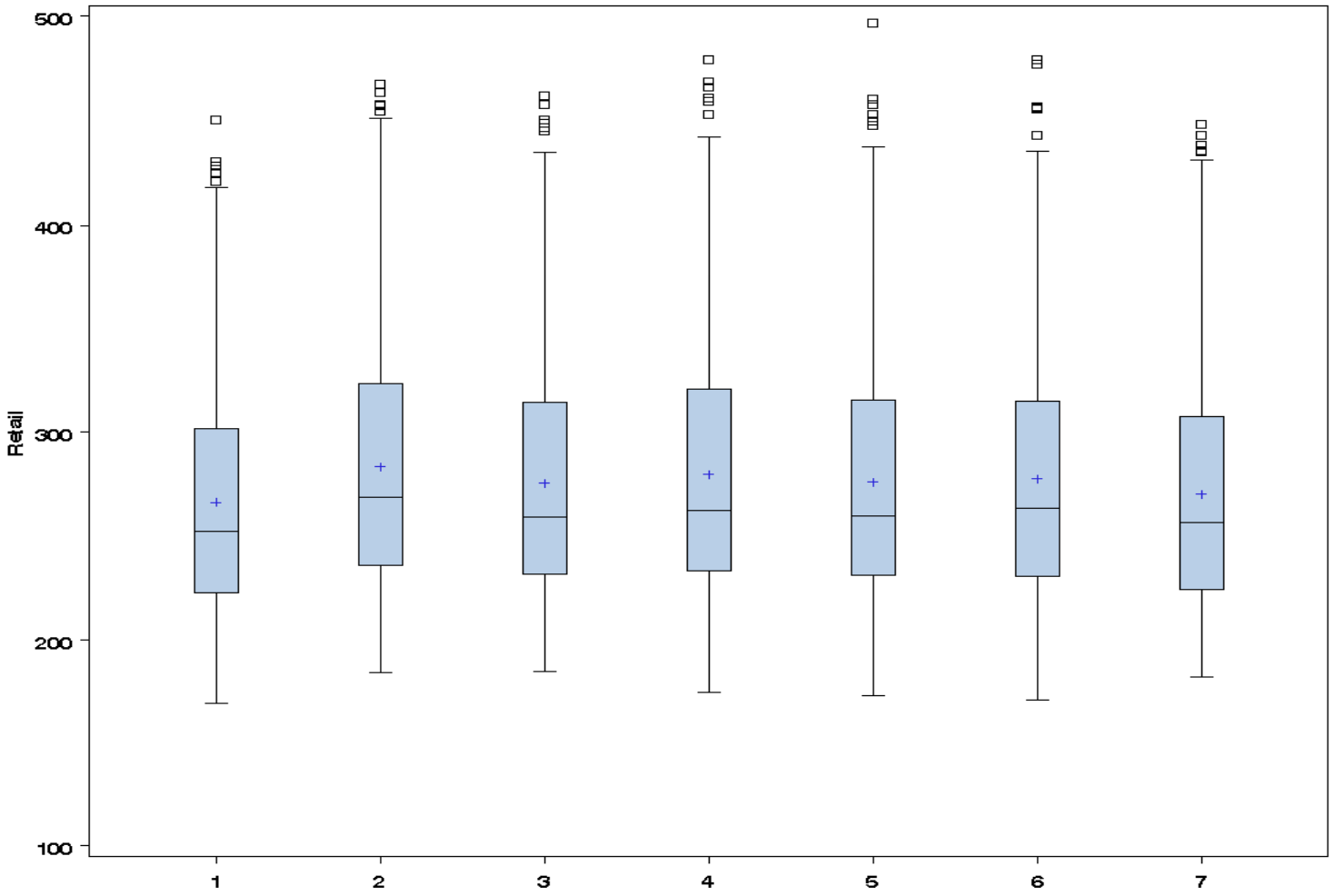
Demand vs High Temperature



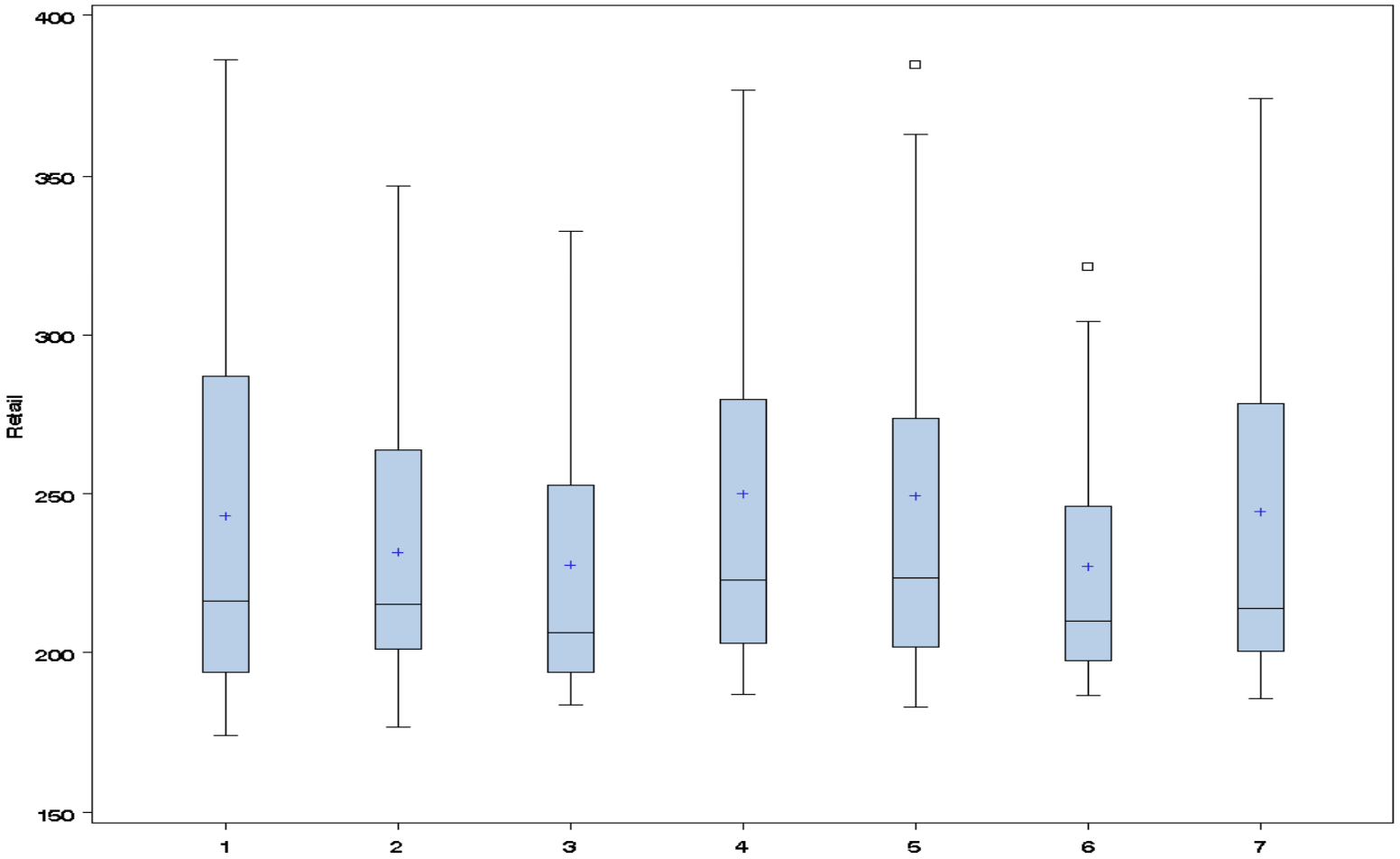
Demand vs Low Temperature



Weekday Demand before Restrictions



Weekday Demand during Restrictions



Interventional Effects

- **Input variables: Intervention event is an indicator variable coded with 0 and 1. in this case, restriction is an intervention factor**

  - **Impulse intervention:**

$$X_t = 1, \quad if \quad t = time$$
$$X_t = 0, \quad otherwise$$

  - **Continuing intervention:**

$$X_t = 0, \quad if \quad t < time$$
$$X_t = 1, \quad otherwise$$

Box-Jenkins Approach(1976)

ARIMA*(p, d, q)* Model

$$W_t = \mu + \sum \frac{\omega_i(B)}{\delta_i(B)} B^{k_i} X_{i,t} + \frac{\theta(B)}{\phi(B)} \alpha_i$$

- W_t is the response series Y_t or a difference of the response series. μ is the mean term;
- X_{it} is the *ith* input series or a difference of the ith input series at time *t*, B is the backshift operator, k_i is the pure time delay for the effect of the ith input series, $\omega_i(B)$ is the numerator polynomial of the transfer function for the ith input series, $\delta_i(B)$ is the denominator polynomial of the transfer function for the ith input series.
- *θ(B)* is the moving average operator, represented as a polynomial in the backshift operator with the order of ***q***: *Φ(B)* is the autoregressive operator, represented as a polynomial in the backshift operator with the order of ***p***: α_i is a random error.

Syntax: ARIMA Procedure

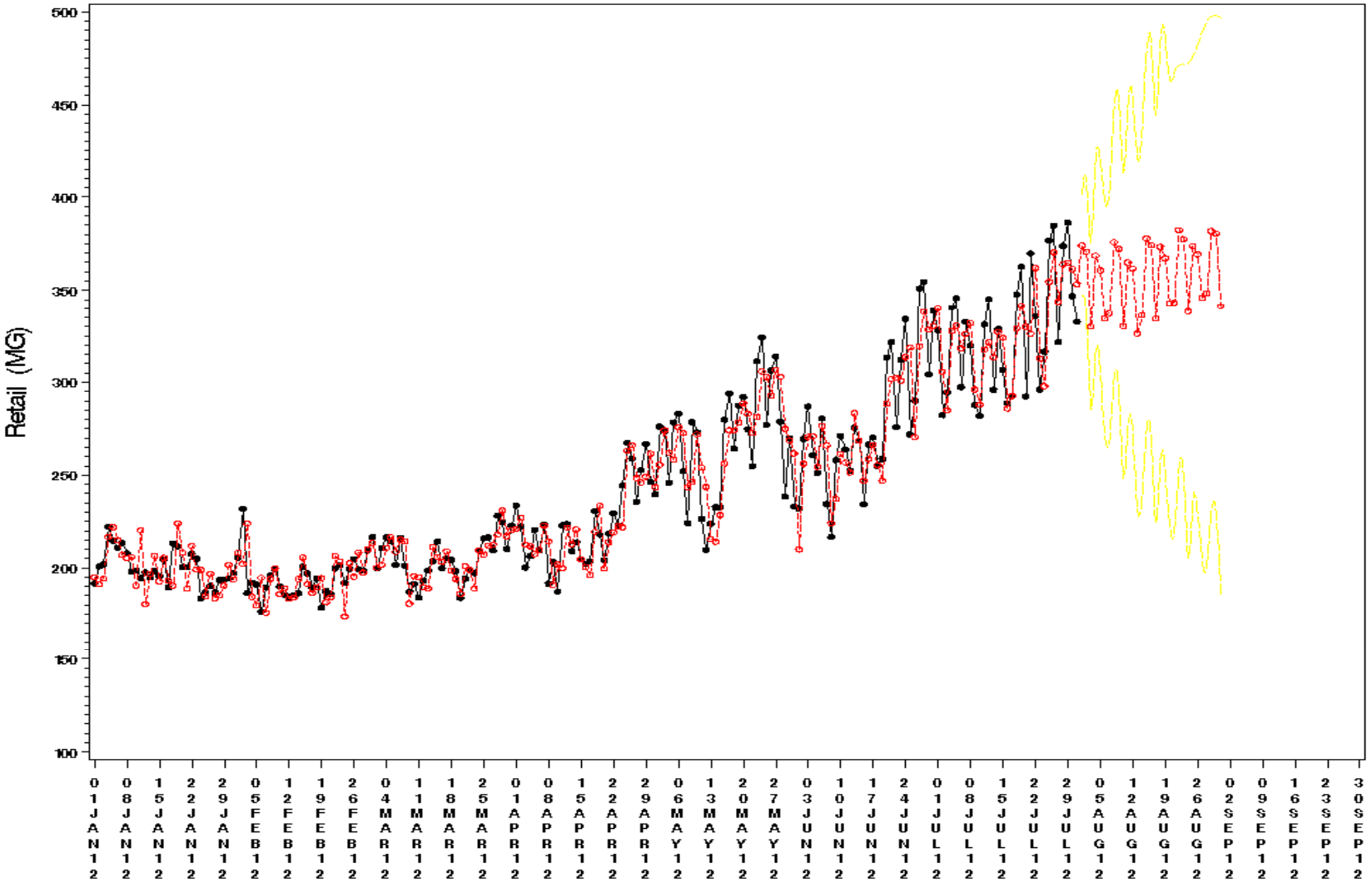
```
proc arima data=retail;
identify var= retail(1) crosscorr=(Temp_max(1)
             Temp_min(1) Prcp(1) Rstc(1));
estimate p=(7) q=(1 7) input=(Rstc Prcp Temp_max
             Temp_min) noconstant method=ml;
forecast lead=61 id=dates out=result_fc;
run;
```

Notice: Intervention event is an indicator variable as a regular regressor

Proc ARIMA

- Construct plots for the response series and its *ACF* to determine if the time series is stationary (a constant mean and covariance over time)
- Construct plots for the estimated *ACF* and *PACF* of the stationary series, and identify *ARIMA* model (*p, d, q*)
- Examine cross-correlations between the response series and input series to identify significant factors
- Estimate different models using maximum likelihood estimation and select the best model by *AIC* and *BIC*
- Forecast based on the best model selected

ARIMA Model Forecasts



Harvey's Approach (1989)

Unobserved Components Model (UCM)

$$Y_t = \mu_t + \gamma_t + \psi_t + \sum_{j=1}^{m} \beta_j x_{jt} + \varepsilon_t$$

- The terms μ_t, γ_t and ψ_t represent the trend, seasonal, and cyclical components, respectively.
- The regression term, $\sum_{j=1}^{m} \beta_j x_{jt}$, includes contribution of regression variables with fixed regression coefficients. A model can also contain regression variables that have time varying regression coefficients with a random regressor.
- The disturbance term ε_t, also called the irregular component, is usually assumed to be Gaussian white noise. In some cases it is useful to model the irregular component as a stationary ARMA process.

Syntax: UCM Procedure

```
Proc UCM data = retail;
id date interval=day;
model retail= Temp_max Temp_min prcp;
irregular /*p=1 q=0*/;
level /*variance=0 noest*/ plot=smooth;
slope /*variance=0 noest*/ plot=smooth;
season length=7 type=trig /*variance=0 noest*/ plot=smooth;
randomreg Rstc/plot=smooth;
cycle plot=smooth /*noest=variance var=0*/;
estimate plot=(residual normal acf);
forecast lead=61 plot=decomp outfor=results;
run;
```

Basic Structure Model(BSM)

```
Proc UCM data = retail;
id date interval=day;
model retail= Temp_max Temp_min prcp;
irregular ;
level  plot=smooth;
slope variance=0 noest plot=smooth;
season length=7 type=trig  plot=smooth;
randomreg Rstc/ plot=smooth;
run;
```

Basic Structural Model(BSM)

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	42.62745	7.31378	5.83	<.0001
Level	Error Variance	99.36026	10.87055	9.14	<.0001
Slope	Error Variance	4.461438E-8	9.27515E-6	0.00	0.9962
Season	Error Variance	0.03927	0.02172	1.81	0.0706
Temp_max	Coefficient	0.49440	0.07373	6.71	<.0001
Temp_min	Coefficient	0.36397	0.08262	4.41	<.0001
Prcp	Coefficient	-4.60895	1.05960	-4.35	<.0001
Rstc	Error Variance	3.20478	1.63061	1.97	0.0494

Likelihood Based Fit Statistics	
AIC (smaller is better)	7610.9
BIC (smaller is better)	7635.1

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.14	0.7045
Level	1	774.18	<.0001
Slope	1	0.27	0.6024
Season	6	9.62	0.1415
Rstc	1	48.52	<.0001

BSM2: Stochastic Level, Fixed Slope, Stochastic Dummy Seasonal

```
Proc UCM data = retail;
id date interval=day;
model retail= Temp_max Temp_min prcp;
irregular;
level plot=smooth;
slope variance=0 noest plot=smooth;
season length=7 type=trig plot=smooth;
randomreg Rstc/plot=smooth;
run;
```

BSM2

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	42.62745	7.31378	5.83	<.0001
Level	Error Variance	99.36026	10.87055	9.14	<.0001
Season	Error Variance	0.03927	0.02172	1.81	0.0706
Temp_max	Coefficient	0.49440	0.07373	6.71	<.0001
Temp_min	Coefficient	0.36397	0.08262	4.41	<.0001
Prcp	Coefficient	-4.60895	1.05960	-4.35	<.0001
Rstc	Error Variance	3.20478	1.63061	1.97	0.0494

Likelihood Based Fit Statistics	
AIC (smaller is better)	7608.9
BIC (smaller is better)	7628.2

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.14	0.7045
Level	1	774.18	<.0001
Slope	1	0.27	0.6024
Season	6	9.62	0.1415
Rstc	1	48.52	<.0001

BSM3: Stochastic Level, Non-Stochastic Dummy Season

```
Proc UCM data = retal_clmt_02;
id date interval=day;
model retail= Temp_max Temp_min prcp;
Irregular;
level  plot=smooth;
season length=7 type=trig variance=0 noest plot=smooth;
randomreg Rstc/plot=smooth;
run;
```

BSM3

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	52.59923	7.20789	7.30	<.0001
Level	Error Variance	96.39881	10.65710	9.05	<.0001
Temp_max	Coefficient	0.51153	0.07521	6.80	<.0001
Temp_min	Coefficient	0.35284	0.08413	4.19	<.0001
Prcp	Coefficient	-4.95189	1.08338	-4.57	<.0001
Rstc	Error Variance	3.21421	1.60183	2.01	0.0448

Likelihood Based Fit Statistics	
AIC (smaller is better)	7613.1
BIC (smaller is better)	7627.6

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.04	0.8486
Level	1	789.25	<.0001
Season	6	263.69	<.0001
Rstc	1	85.69	<.0001

BSM4 : Stochastic Level, Non-Stochastic Dummy Seasonal, Stochastic Cycle

```
Proc UCM data = retail;
id date interval=day;
model retail= Temp_max Temp_min prcp;
irregular ;
level  plot=smooth;
season length=7 type=trig variance=0 noest plot=smooth;
randomreg rstc/plot=smooth;
cycle plot=smooth;
estimate plot=(residual normal acf);
run;
```

BSM4

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	30.28257	12.19195	2.48	0.0130
Level	Error Variance	42.55691	11.91217	3.57	0.0004
Cycle	Damping Factor	0.70236	0.11486	6.11	<.0001
Cycle	Period	1516.11038	130200.1	0.01	0.9907
Cycle	Error Variance	81.70918	28.89956	2.83	0.0047
Temp_max	Coefficient	0.50470	0.07590	6.65	<.0001
Temp_min	Coefficient	0.35390	0.08581	4.12	<.0001
Prcp	Coefficient	-4.72885	1.08722	-4.35	<.0001
Rstc	Error Variance	3.09132	1.54631	2.00	0.0456

Likelihood Based Fit Statistics	
AIC (smaller is better)	7603.4
BIC (smaller is better)	7632.4

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.00	0.9551
Level	1	476.93	<.0001
Cycle	2	0.29	0.8637
Season	6	270.10	<.0001
Rstc	1	85.05	<.0001

BSM5 : Stochastic Level, Non-Stochastic Dummy Seasonal, ARMA(1,1) Process

```
Proc UCM data = retail;
id date interval=day;
model retail= Temp_max Temp_min prcp;
irregular p=1 q=0 plot=smooth;
level plot=smooth;
slope variance=0 noest plot=smooth;
season length=7 type=trig plot=smooth;
randomreg Rstc/plot=smooth;
estimate plot=(residual normal acf);
run;
```

BSM5

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	123.26280	16.71235	7.38	<.0001
Irregular	MA_1	0.17255	0.08343	2.07	0.0386
Irregular	AR_1	0.70236	0.11486	6.11	<.0001
Level	Error Variance	42.55674	11.91193	3.57	0.0004
Temp_max	Coefficient	0.50470	0.07590	6.65	<.0001
Temp_min	Coefficient	0.35390	0.08581	4.12	<.0001
Prcp	Coefficient	-4.72887	1.08722	-4.35	<.0001
Rstc	Error Variance	3.09133	1.54632	2.00	0.0456

Likelihood Based Fit Statistics	
AIC (smaller is better)	7601.4
BIC (smaller is better)	7625.6

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.25	0.6140
Level	1	476.92	<.0001
Season	6	270.10	<.0001
Rstc	1	85.05	<.0001

BSM6 : Fixed Level, Non-Stochastic Dummy Season(365), ARMA(1,1) Process

```
Proc UCM data = retail;
id date interval=day;
model retail= Temp_max Temp_min prcp;
irregular p=1 q=0 plot=smooth;
level variance=0 noest plot=smooth;
season length=365 type=trig plot=smooth;
randomreg rstc/plot=smooth;
estimate plot=(residual normal acf);
forecast lead=61 plot=decomp outfor=results;
run;
```

BSM6

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > |t|
Irregular	Error Variance	245.84318	14.64002	16.79	<.0001
Irregular	MA_1	0.39709	0.04559	8.71	<.0001
Irregular	AR_1	0.90283	0.02190	41.22	<.0001
Temp_max	Coefficient	0.51813	0.11687	4.43	<.0001
Temp_min	Coefficient	0.42958	0.12476	3.44	0.0006
Prcp	Coefficient	-5.33590	1.74457	-3.06	0.0022
Rstc	Error Variance	3.61854	2.32405	1.56	0.0456

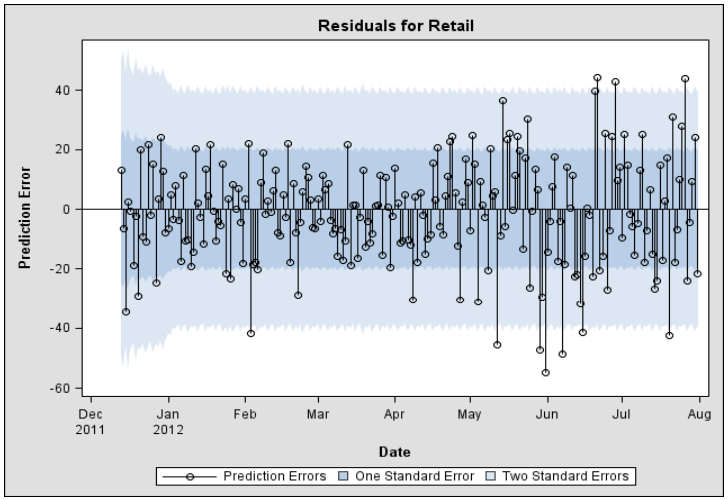
Likelihood Based Fit Statistics	
AIC (smaller is better)	7077.8
BIC (smaller is better)	7095.2

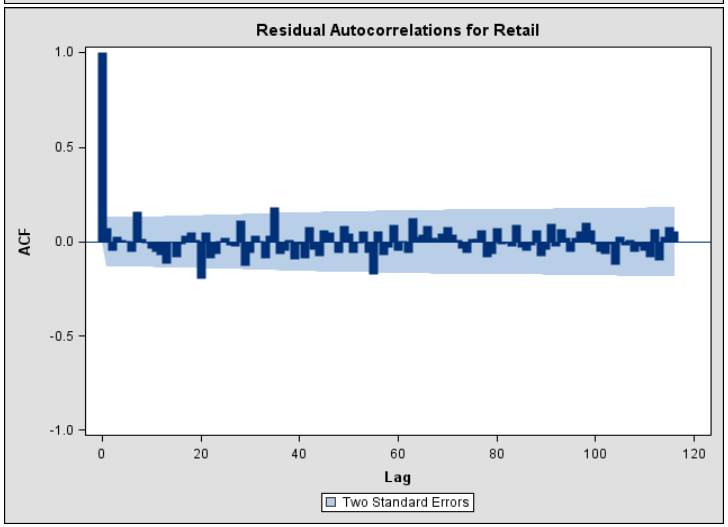
Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	0.03	0.8701
Level	1	546.14	<.0001
Season	364	455.71	0.0008
Rstc	1	42.50	<.0001

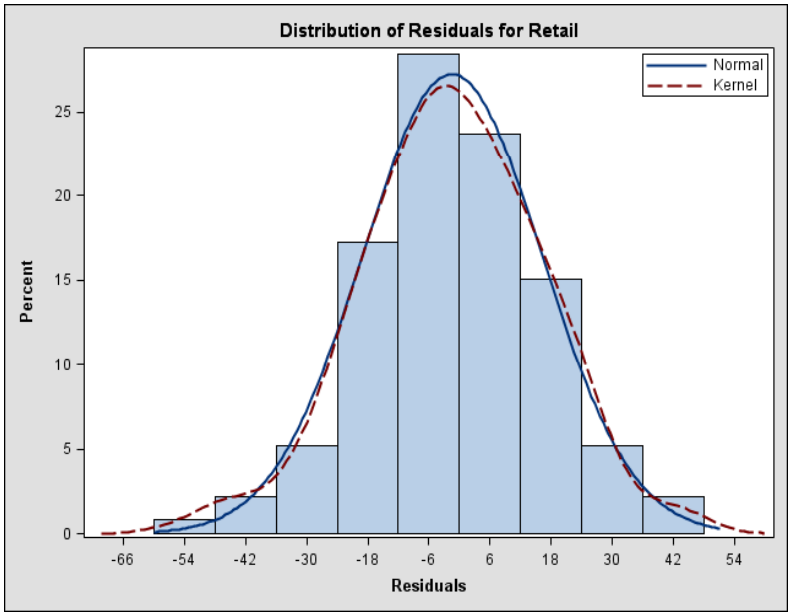
Likelihood-Based Fit Statistics

UCM	Description	Variance Components	AIC	BIC
BSM	Irregular	$\sigma_\varepsilon^2 > 0$	7610.9	7635.1
	Stochastic Level	$\sigma_\eta^2 > 0$		
	Stochastic Slope	$\sigma_\xi^2 > 0$		
	Stochastic dummy season(7)	$\sigma_\omega^2 > 0$		
BSM2	Irregular	$\sigma_\varepsilon^2 > 0$	7608.9	7628.2
	Stochastic Level	$\sigma_\eta^2 > 0$		
	Fixed Slope	$\sigma_\xi^2 = 0$		
	Stochastic dummy season (7)	$\sigma_\omega^2 > 0$		
BSM3	Irregular	$\sigma_\varepsilon^2 > 0$	7613	7627.5
	Stochastic Level	$\sigma_\eta^2 > 0$		
	Non-Stochastic dummy season (7)	$\sigma_\omega^2 = 0$		
BSM4	Irregular	$\sigma_\varepsilon^2 > 0$	7603.4	7632.4
	Stochastic Level	$\sigma_\eta^2 > 0$		
	Non-stochastic dummy season (7)	$\sigma_\omega^2 = 0$		
	One stochastic cycle	$\sigma_\nu^2 > 0$		
BSM5	Irregular	$\sigma_\varepsilon^2 > 0$, ARMA (1, 0, 1)	7601.4	7625.6
	Stochastic Level	$\sigma_\eta^2 > 0$		
	Non-stochastic dummy season (7)	$\sigma_\omega^2 = 0$		
BSM6	Irregular	$\sigma_\varepsilon^2 > 0$, ARMA (1, 0, 1)	7077.8	7095.2
	Fixed Level	$\sigma_\eta^2 = 0$		
	Non-stochastic dummy season (365)	$\sigma_\omega^2 = 0$		

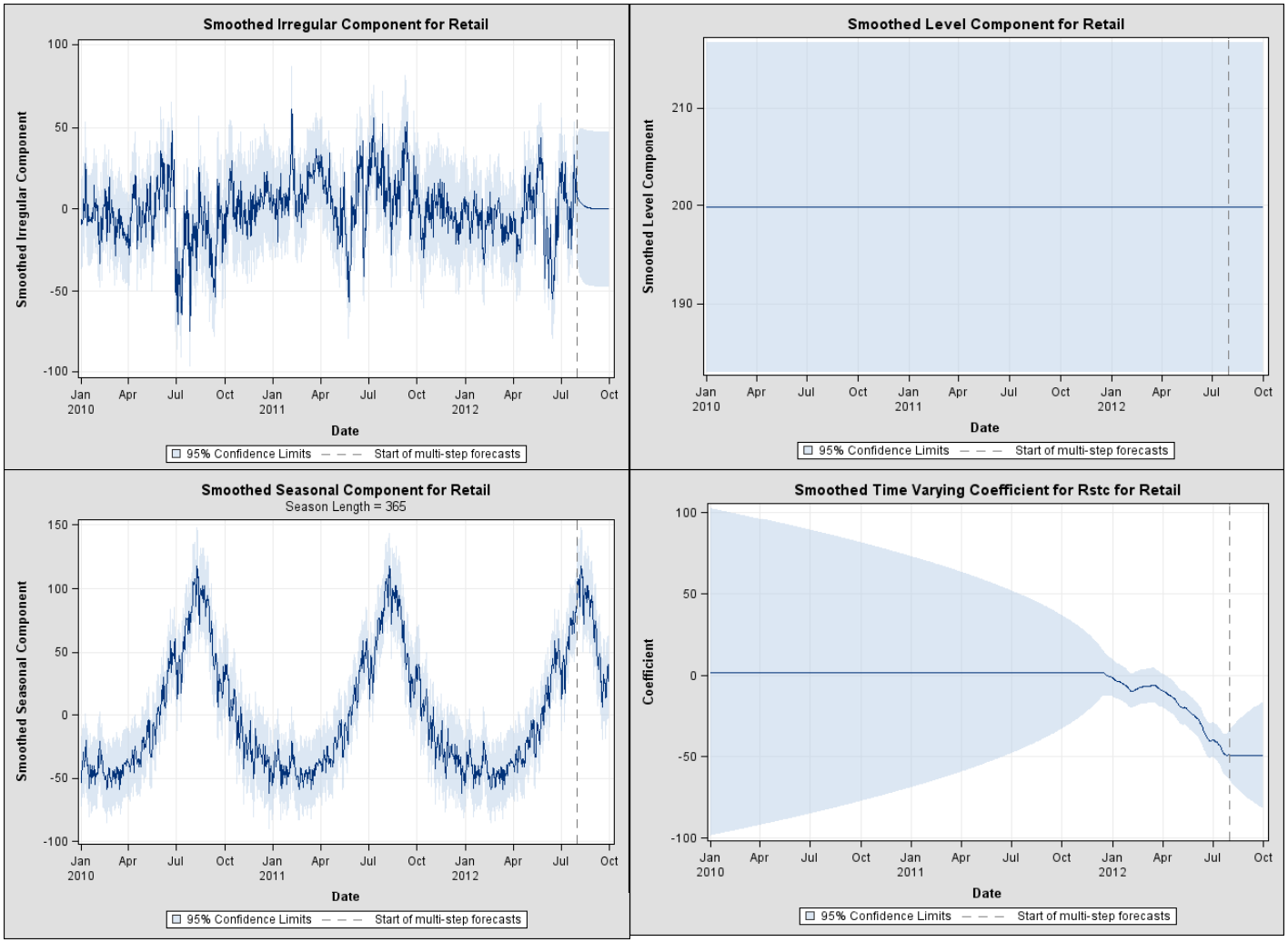
Residual Plots



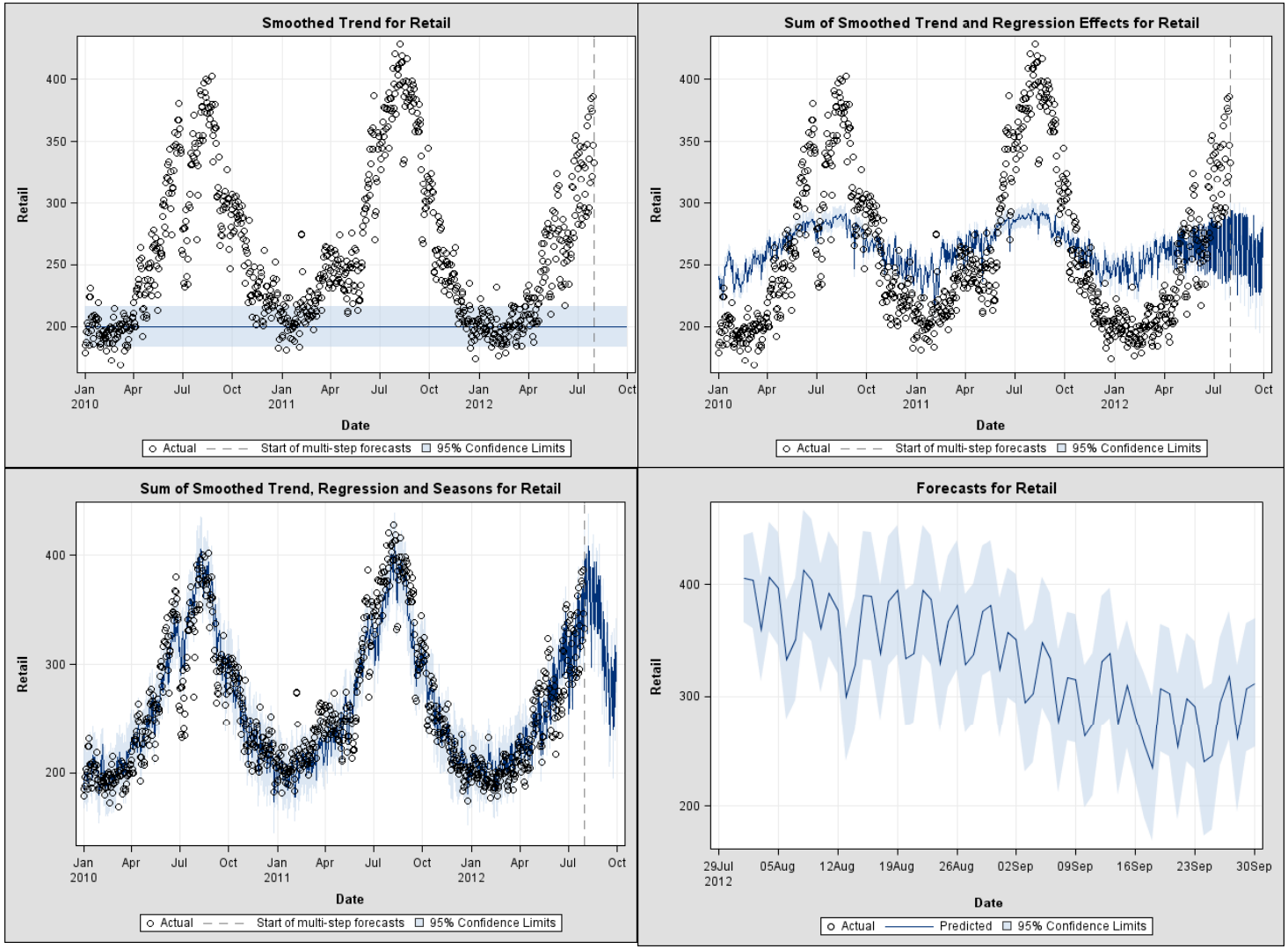




Component Estimates



Demand Estimates



Summary

- UCM is a very good tool for retail operation forecast
- Very convenient to decompose a time-series into several components of trend, season, cycle and irregular terms
- Residual and estimate plots help diagnose model fitting and interpret findings
- Meaningful interpretation for each component and more forecasting accuracy

Questions?

Fujiang Wen, Ph.D.
City of Dallas | Water Utilities
1500 Marilla St. 5AS | Dallas, TX 75201
☎ Office: 214.670.4022
✉ fujiang.wen@dallascityhall.com