In a time series, there can be a change in parameter coefficient or error variance at a time point $k$. Traditional simple regression modeling can limit the accuracy and predictability of this time series if this change is not accounted for. Utilizing PROC UCM in SAS/ETS®, the structural break at $k$ can be identified so that two separate regression lines can be modeled to better represent the data, and to provide us with better predictions for future time periods.

**Abstract**

**Introduction**

A simple walk-through of how data with structural break look like.

**Methods**

A detailed PROC UCM step-by-step guide to find the break point $k$ and develop better models for the Nile River water level data.

**Application 1**

Using the structural break analysis to identify a change in the stock market trend of the SPY ETF closing prices.

**Application 2**

Applying this structural break idea to traditional non-time series data of PROC SQL runtimes.

**Conclusion**

Summary and thoughts.
**Introduction**

A time series with a structural break exhibits changes in parameter coefficients and/or error variance at a time point \( k \) such that:

\[
y_t = \alpha_1 + \beta_1 x_t + \epsilon_t, \quad t=1, \ldots, k
\]

\[
y_t = \alpha_2 + \beta_2 x_t + \eta_t, \quad t=k+1, \ldots, T
\]

where \( \epsilon_t \sim i.i.d. N(0, \sigma_1^2) \)

\( \eta_t \sim i.i.d. N(0, \sigma_2^2) \)

\( \alpha_2 > \alpha_1 \)

\( \sigma_2^2 > \sigma_1^2 \)

\( \beta_2 > \beta_1 \)

**Examples**

**Nile River water levels** (see Methods section)
- Aswan dam was built in 1899, lowering the mean water level, \( \alpha_2 < \alpha_1 \)

**SPY ETF closing prices** (see Application 1 section)
- A break of an uptrend in closing prices, \( \beta_2 < \beta_1 \)

**SAS® Proc SQL runtimes** (see Application 2 section)
- Although not a time series, this data mimic \( \sigma_2^2 > \sigma_1^2 \)

**PROC UCM**

PROC UCM can be used to perform the structural break analysis.

**Syntax: UCM Procedure**

The UCM procedure uses the following statements:

```
PROC UCM <options>;
   AUTOREG <options>;
   BLOCKSEASON <options>;
   BY variables;
   CYCLE <options>;
   DEPLAG <options>;
   ESTIMATE <options>;
   FORECAST <options>;
   ID variable options;
   IRREGULAR <options>;
   LEVEL <options>;
   MODEL dependent variable =< regressors >;
   NLOPTIONS options;
   PERFORMANCE options;
   OUTLIER options;
   RANDOMREG regressors /< options>;
   SEASON <options>;
   SLOPE <options>;
   SPLINEREG regressor <options>;
   SPLINESEASON options;
   TF regressor <options>;
```

The PROC UCM and MODEL statements are required. In addition, the model must contain at least one component with nonzero disturbance variance.
Step-by-step Guide

1. Find components that are significant:

   ```
   procglm data=river;
   model level = year / solution;
   run;
   ```

   - Step 1: Include autoregressive and slop components.

2. **PROC REG** with RMSE = 150.55

   - Simple regression line resulted in heteroscedastic residuals

   ```
   procreg data=river plotting=(isolution); 
   model level = year / solution; 
   run;
   ```

   - Result: simple regression line had an insignificant error term.

   ```
   procreg data=river plotting=(isolution); 
   model level = year / solution ac=(0 1); 
   run;
   ```

   - Result: level component has an insignificant error term.

   ```
   procucm data=river plotting=(isolution); 
   model level / ar1=(0 1); 
   run;
   ```

   - Result: the forecasted water level using **PROC REG** would be significantly lower and less predictive than that of **PROC UCM**.

   ```
   procucm data=river plotting=(isolution); 
   model level / ar1=(0 1 2); 
   run;
   ```

   - More importantly, the forecasted water level using **PROC REG** would be significantly lower and less predictive than that of **PROC UCM**.

Nile River Water Level Data

- 100-year water level data of Nile River
- Aswan dam built in 1899, lowering water level
- **PROC REG** with RMSE = 150.55

- **PROC UCM** with RMSE = 128.26

Abstract
Introduction
Methods
Application 1
Application 2
Conclusion

Please use the headings above to navigate through the different sections of the poster.
**SPY ETF Closing Price Data**

- A real-time structural break analysis on Exchange Traded Funds (ETF) SPY daily closing prices starting from Aug 17th, 2017
- Daily analysis using PROC UCM performed to identify the break point indicating the end of an uptrend
- SPY closing prices on Jan 31st, 2018

**PROC UCM analysis on Jan 31st, 2018**

<table>
<thead>
<tr>
<th>Obs</th>
<th>Date</th>
<th>Break Type</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>117</td>
<td>2018-01-31</td>
<td>Additive Outlier</td>
<td>-2.76431</td>
<td>0.769458</td>
<td>12.91</td>
<td>1</td>
<td>0.0003</td>
</tr>
<tr>
<td>119</td>
<td>2018-01-31</td>
<td>Level</td>
<td>-3.52351</td>
<td>1.298002</td>
<td>11.73</td>
<td>1</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

Daily analysis results with Jan 30th closing price first identified as an outlier on Jan 30th, subsequently confirmed as a break point on Jan 31st.

- SPY closing prices on Feb 28th, 2018

**PROC UCM analysis on Jan 30th, 2018**

<table>
<thead>
<tr>
<th>Analysis On</th>
<th>Date</th>
<th>Break Type</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 26, 2018</td>
<td>Jan 30</td>
<td>Additive Outlier</td>
<td>2.80</td>
<td>0.78</td>
<td>12.84</td>
<td>1</td>
<td>0.0003</td>
</tr>
<tr>
<td>Jan 30, 2018</td>
<td>Jan 30</td>
<td>Additive Outlier</td>
<td>-3.54</td>
<td>1.04</td>
<td>11.66</td>
<td>1</td>
<td>0.0006</td>
</tr>
<tr>
<td>Jan 31, 2018</td>
<td>Jan 26</td>
<td>Additive Outlier</td>
<td>2.76</td>
<td>0.77</td>
<td>12.91</td>
<td>1</td>
<td>0.0003</td>
</tr>
<tr>
<td>Jan 31, 2018</td>
<td>Jan 30</td>
<td>Level</td>
<td>-3.52</td>
<td>1.03</td>
<td>11.73</td>
<td>1</td>
<td>0.0006</td>
</tr>
<tr>
<td>Feb 1, 2018</td>
<td>Jan 26</td>
<td>Additive Outlier</td>
<td>2.75</td>
<td>0.77</td>
<td>12.95</td>
<td>1</td>
<td>0.0003</td>
</tr>
<tr>
<td>Feb 1, 2018</td>
<td>Jan 30</td>
<td>Level</td>
<td>-3.51</td>
<td>1.03</td>
<td>11.69</td>
<td>1</td>
<td>0.0006</td>
</tr>
<tr>
<td>Feb 2, 2018</td>
<td>Feb 2</td>
<td>Additive Outlier</td>
<td>-6.49</td>
<td>1.19</td>
<td>29.63</td>
<td>1</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Feb 2, 2018</td>
<td>Feb 2</td>
<td>Level</td>
<td>-6.49</td>
<td>1.19</td>
<td>29.63</td>
<td>1</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Anecdotally, PROC SQL join operation can have runtimes that vary greatly depending on the size of the datasets. The runtimes are generally proportional to the size but only up to a certain point. Datasets that are much larger seem to have runtimes much longer than anticipated. This sounds very similar to a structural break analysis.

The runtime data generated have the following attributes:

- 173 simulated runtimes
- Primary dataset **heart** has varying simulated number of observations from 15K to 10M
- All four reference datasets have the same 50K observations and a range of 11 to 400 variables in all simulations

PROC UCM cannot be used in this case because the independent variable (number of observations in **heart**) does not have a regular interval like time series data. However, the modeling idea remains the same.

Visually, the break point is somewhere between 1 to 3 million observations, where the variability of runtimes is much greater after 3 million observations. A separate model can be developed using only the runtime data with 3+ million observations.
CONCLUSION

• Structural break can occur when parameter coefficient and/or error variance change at a particular time point.

• PROC UCM can be used to find the break point, and to model the time series with a structural break.

• Better models and predictions when structural break is accounted for.

• Real-time structural break analysis on financial data is also an application of PROC UCM.

• Awareness of possible structural break in traditional non-time series data can be helpful in selecting data for modeling and predictions.

References
