Design of a Weather-Normalization Forecasting Model

Final Briefing

09 May 2014

Sponsor:
Northern Virginia Electric Cooperative

Abram Gross
Jedidiah Shirey
Yafeng Peng
OR-699

Agenda

- Background
- Problem Statement
- Sponsor, Purpose, Objective
- Key Definitions
- Scope
- Assumptions
- Limitations
- Methodology
- Excursions
- Recommendations

Background

- Northern Virginia Electric Cooperative (NOVEC) is an energy distributor serving parts of 6 Northern Virginia counties.
  - Mandated to meet all customer energy requests.
  - Service provided by energy market purchases from regional providers:
    - 1) Bulk purchases via contracts 5 years in advance.
    - 2) Spot purchases up to one day prior to delivery.

- NOVEC conducts analysis to inform energy purchases.
  - Forecasts of energy demand over a 30-year horizon.
  - Decomposes forecasted demand into a base load and seasonal load.
    - Base-load: average demand from customer base.
    - Seasonal-load: weather's impact on base consumption.

- Predicted energy consumption determines bulk purchase quantity.
  - Large error leads to increased costs.
  - Recently requested by Sales Department to improve forecast accuracy.

Problem Statement

- Climate changes have caused NOVEC to question whether the current weather-normalization methodology can be improved.

- NOVEC requests a review of weather-normalization methods that account for changing weather trends:
  - Evaluate new methodologies to compare to existing procedure.
  - Improve estimates of customer base trends.

- NOVEC needs a model to accurately predict total energy consumption.
  - Forecast over 30 year horizon; energy demand reported monthly.
  - Emphasis on first 5 years to align with bulk purchase contracts.
  - Normalizing for weather to quantify customer growth, including impact of economic factors.

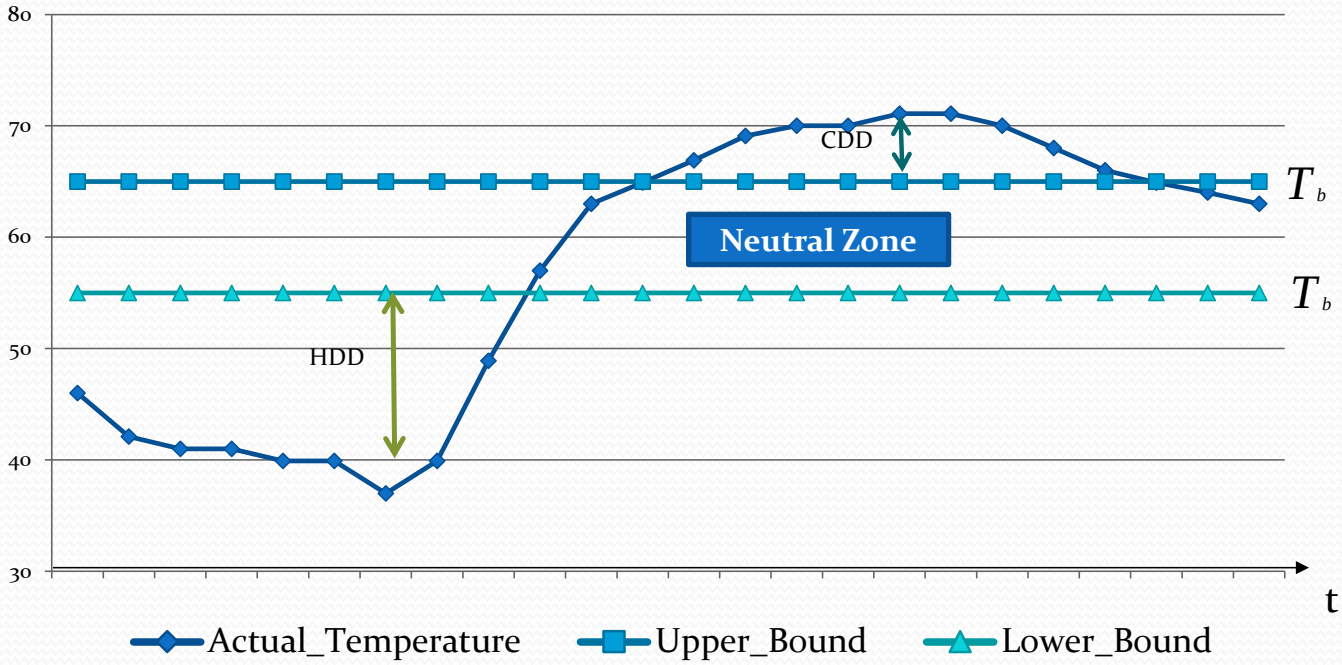
Sponsor, Purpose, Objectives

- Sponsor
  - NOVEC.
- Purpose
  - Provide a candidate methodology to normalize weather impact on monthly energy purchases.
- Objectives:
  - 1) Assess historic relationships between economic, weather, and power data.
  - 2) Develop a forecast model to test normalization methodologies.
  - 3) Test weather-normalization methodologies; recommend one for implementation based on accuracy and robustness.

Key Definitions

- Heating Degree-Day (HDD): Measure used to indicate amount of energy need to heat during cold weather.
- Cooling Degree-Day (CDD): Measure used to indicate amount of energy need to cool during hot weather.

$$CDD = \sum_{i=1}^{N} (\bar{T}_i - T_b)^+ \qquad HDD = \sum_{i=1}^{N} (T_b - \bar{T}_i)^+$$



Scope

- Data:
  - NOVEC monthly energy purchases data since 1983.
  - Dulles Airport weather data since 1963.
  - Historic economic factors data since 1980s metro D.C.; state and county data not evaluated.
  - 30 years of Moody's Analytics forecasted economic factors.

- Model:
  - Inputs: historic energy purchases, weather data, economic variables, and customer-base.
  - Output: Monthly predictions for energy consumption over a 30-year horizon.
  - Regression: characterize dynamics between parameters.
  - Weather-normalization: remove seasonal weather impacts on NOVEC's load.
  - Forecast: facilitate testing of varied normalization methodologies.
  - Ensure synergy with NOVEC's existing models (regression, weather-normalization, forecast).

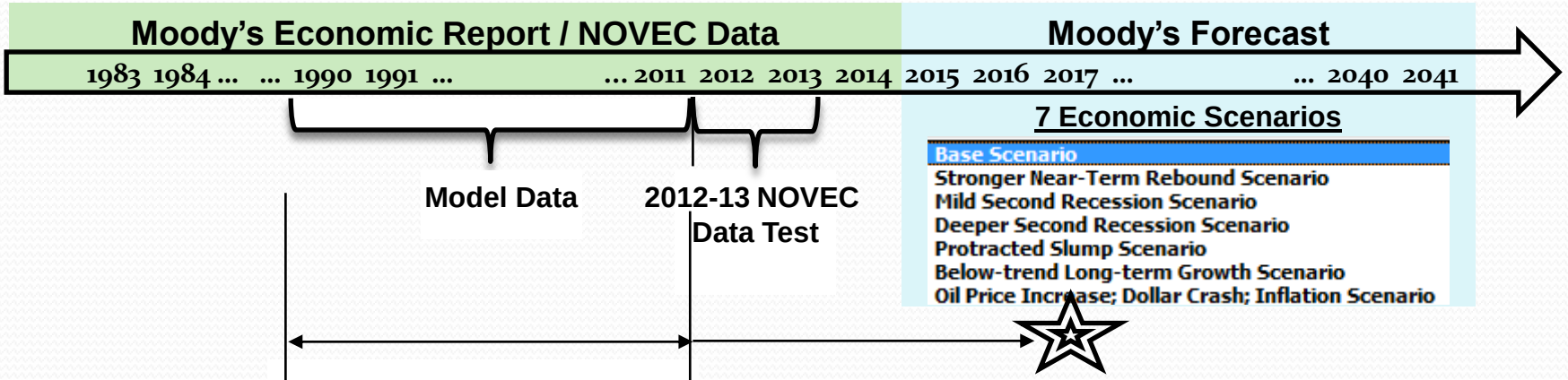
Assumptions

- Neutral zone between HDD/CDD has insignificant impact on energy consumption.
  - 55 and 65 degrees are the lower and upper bounds.
- Economic variables currently utilized provide proper indicators for gauging future power demand:
  - Employment: Total Non-Agricultural
  - Gross Metro Product: Total
  - Housing Completions: Total
  - Households
  - Employment (Household Survey): Total Employed
  - Employment (Household Survey): Unemployment Rate
  - Population: Total

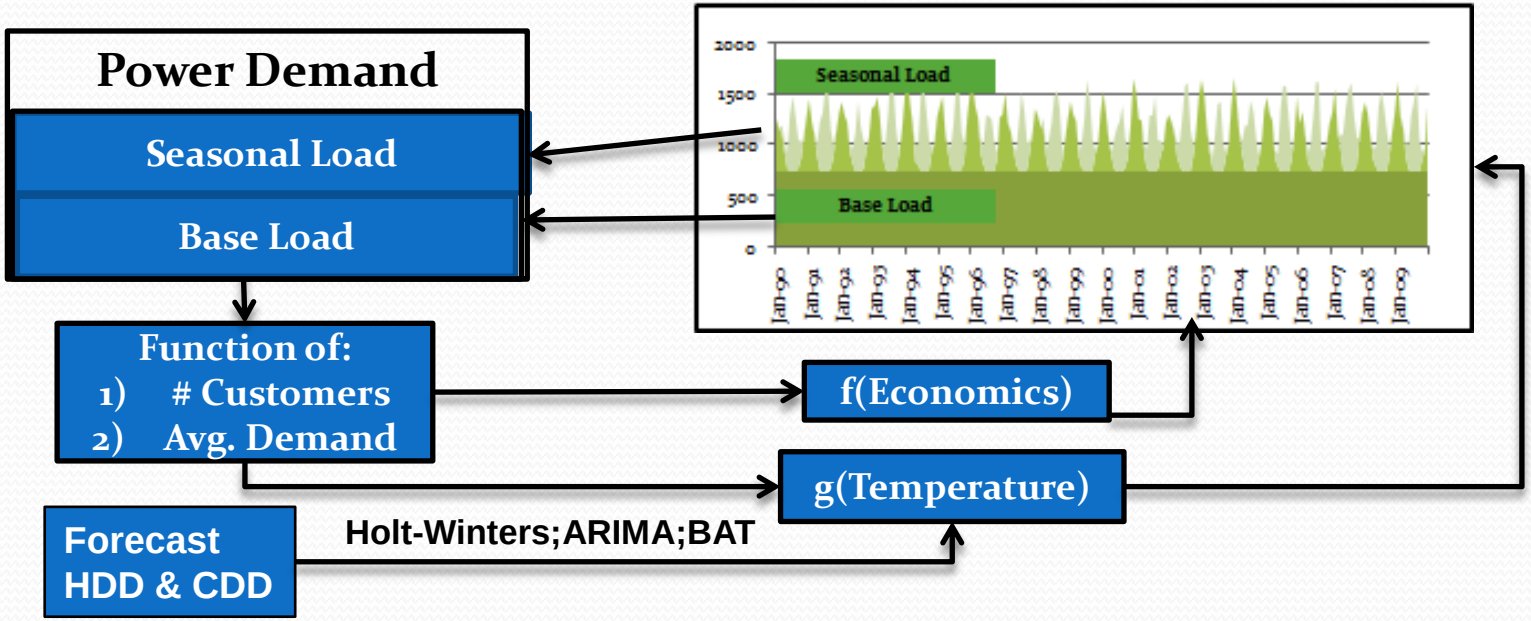
Limitations

- Unable to develop deep understanding of NOVEC's current forecast model due to complexity and time constraints:
  - Hinder adopting into existing model.
  - Skew comparisons of forecast accuracy.
- Economic regression model determines customer base; potential for inconsistent forecast comparisons of this and NOVEC's current model output.
- Baseline economic scenario only scenario evaluated.

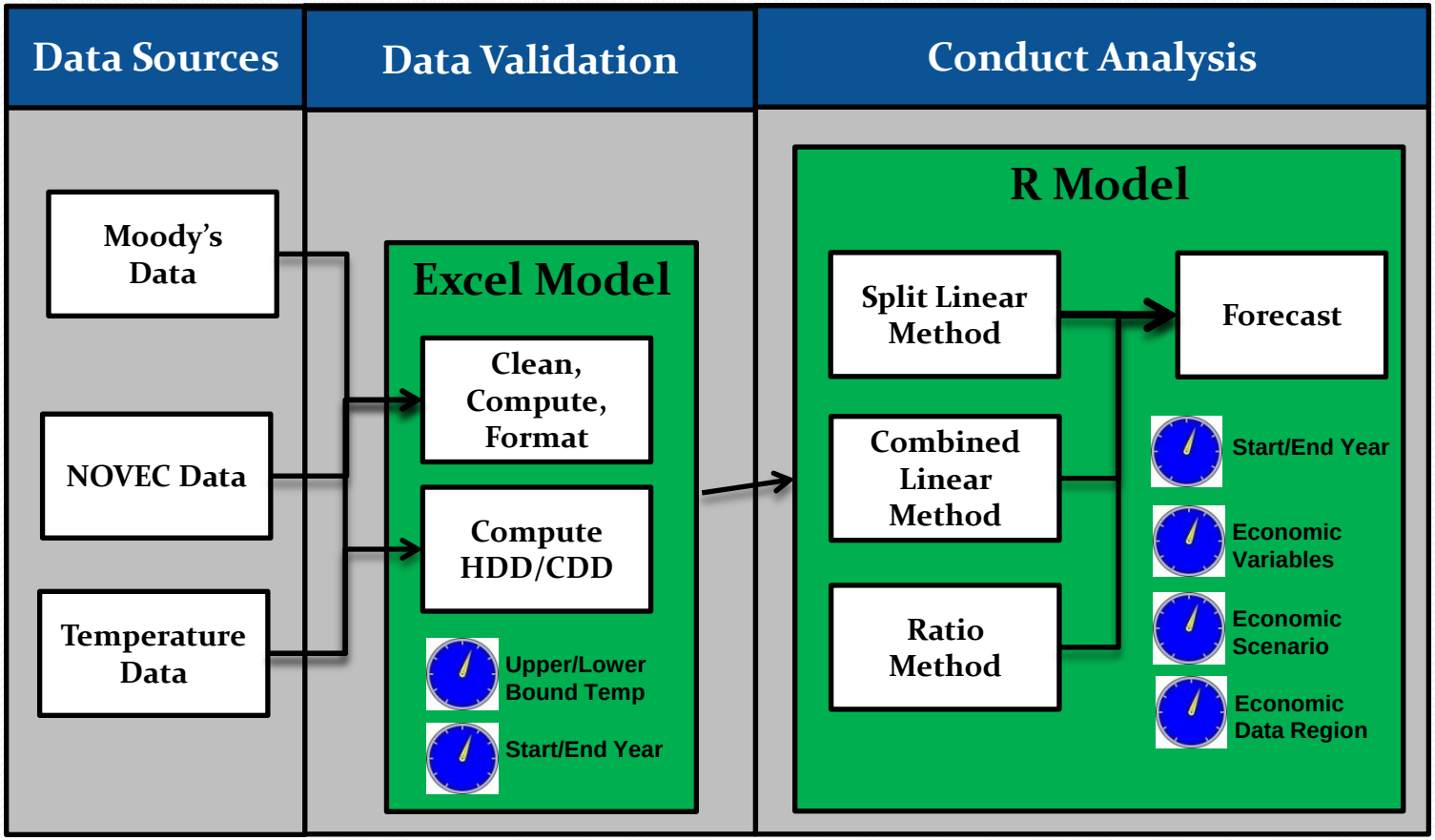
Approach



Determine Historic Relationships



Methodology



- Data processing included linear interpolation for data gaps, disaggregating quarterly economic data, as well as aggregating hourly weather data, up to monthly resolution.
- Primary methodologies: Split Regression Model, Combined Regression Model, and Ratio Model

Combined Linear Regression Model

- *Intuition:*
  - *Usage should be a function of economic contributions and weather contributions.*

$$y = f(x_{1..i}, HDD, CDD)$$

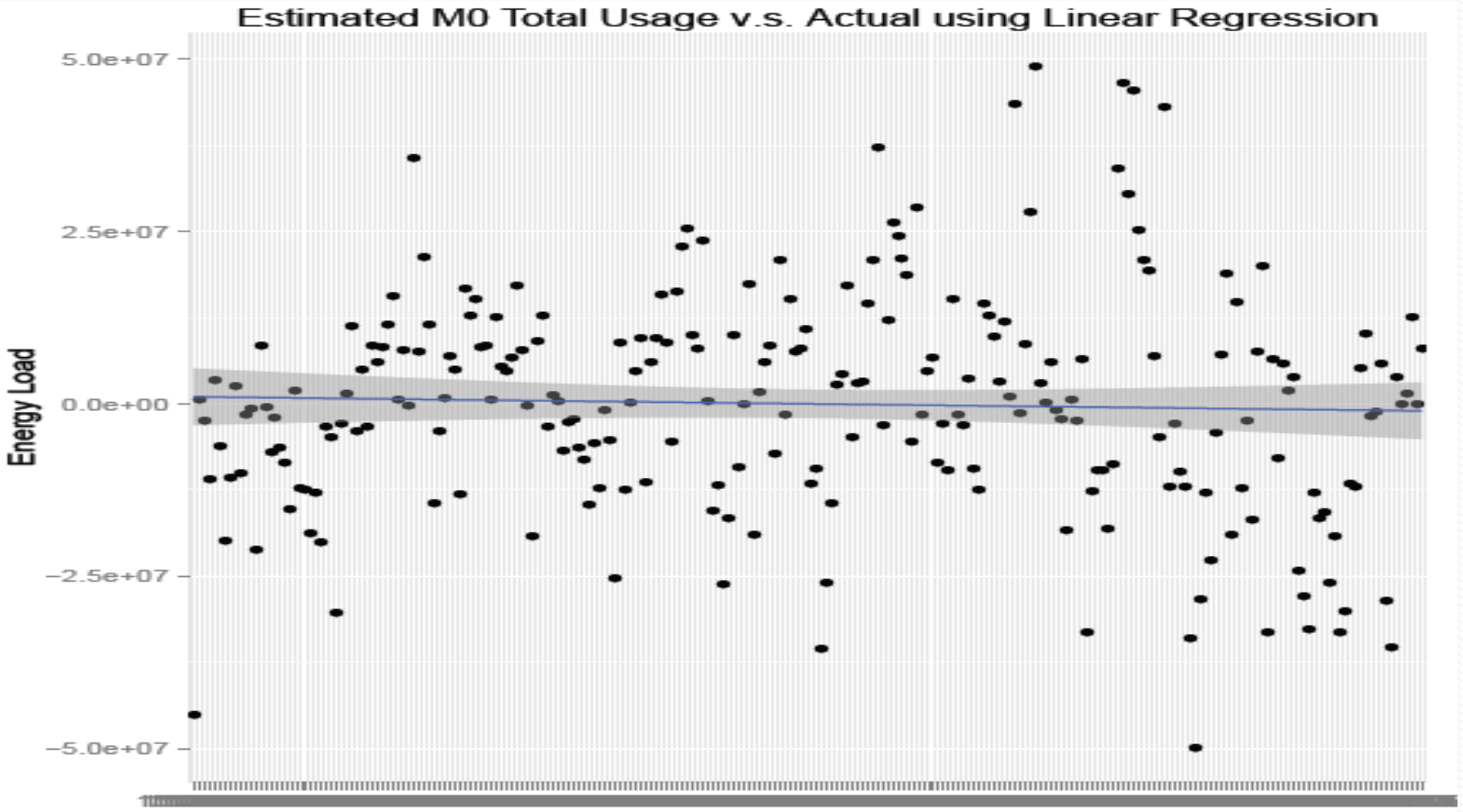
y: total observed load

x_i: economic variables

HDD: Heating degree day; CDD: Cooling degree day

Accuracy of Linear Regression Model

- Adjusted R-square = 0.925

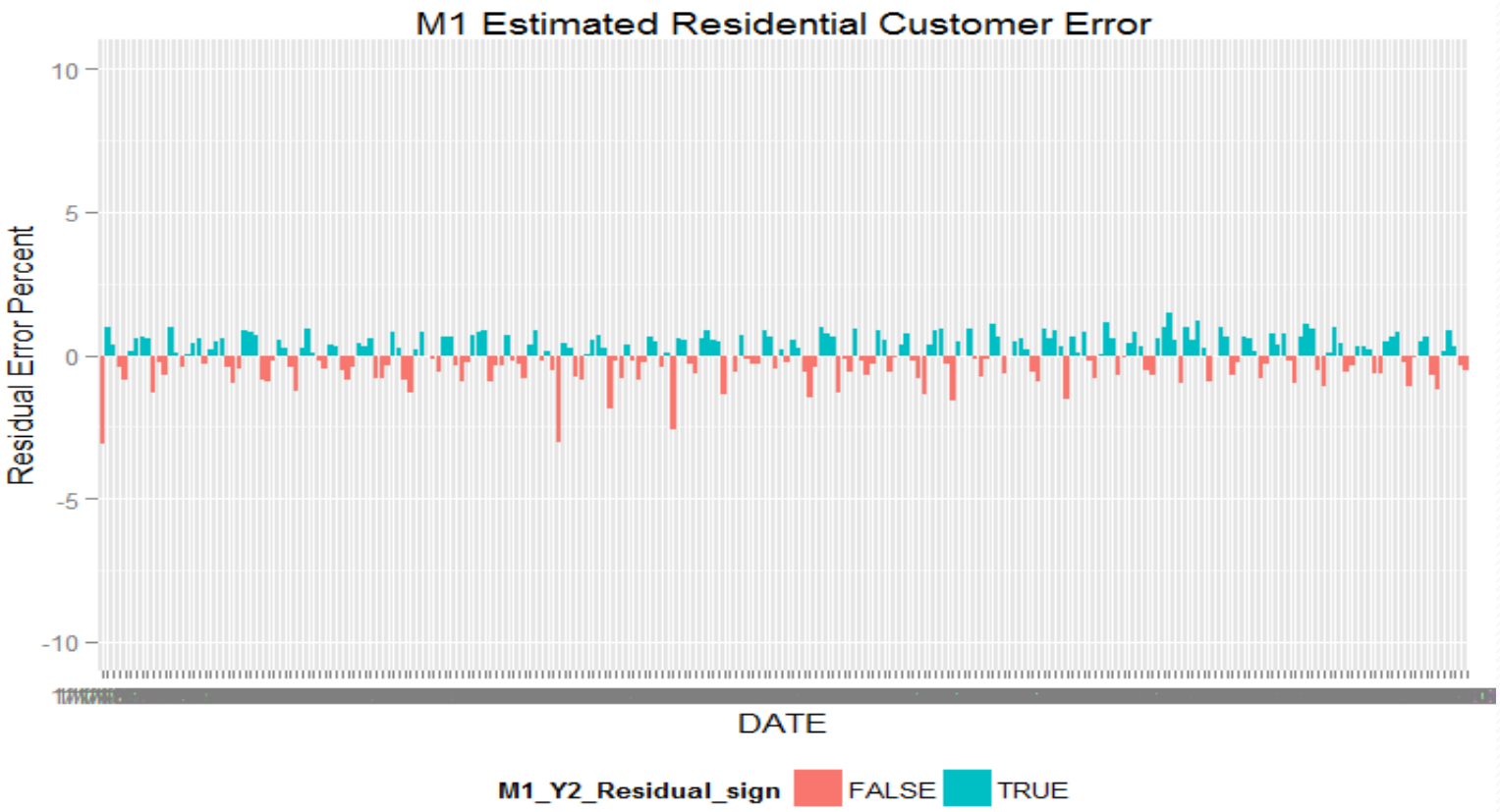


Additional Forecasting Methods

- Split Regression
  - total load = residential load + non-residential load
    - residential load = # of residential customer * avg residential
    - non-residential load = # of non-residential customer * avg non-residential
- Customer Ratio Method
  - total load = residential load + non-residential load
    - residential load = # of residential customer * avg residential
    - non-residential load = residential load * ratio

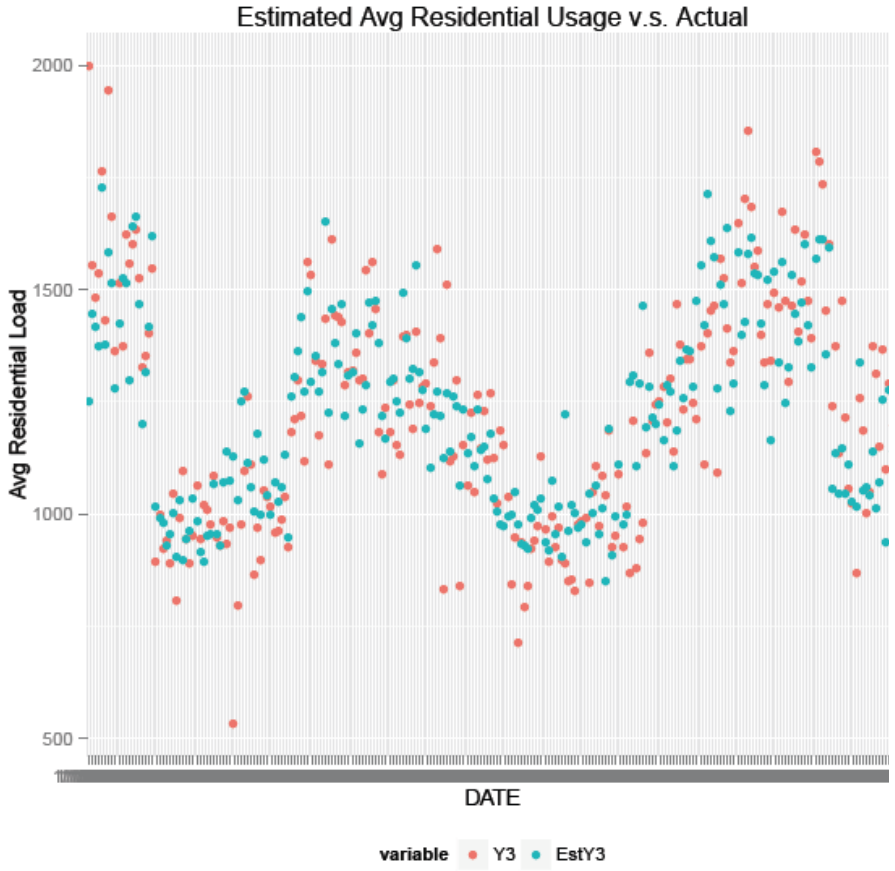
Estimate of Customer Base

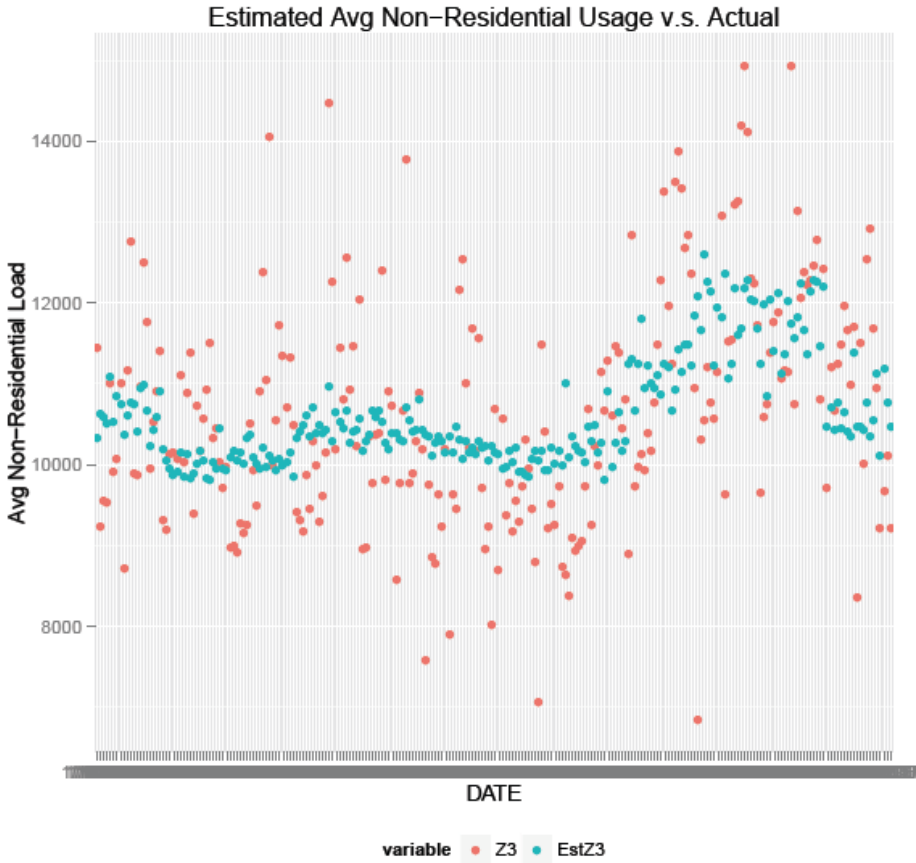
- Customers are categorized as either residential or non-residential.
- >99% adjusted R square based on the 7 econometric variables from 1990-2011.
  - Linear regression model provides sufficient accuracy for predicting customer base.



Estimate Average Customer Usage

- Tested linear regression on 3 similar models
  - 7 Econometric Variables + HDD + CDD
  - 7 Econometric Variables only
  - HDD + CDD only

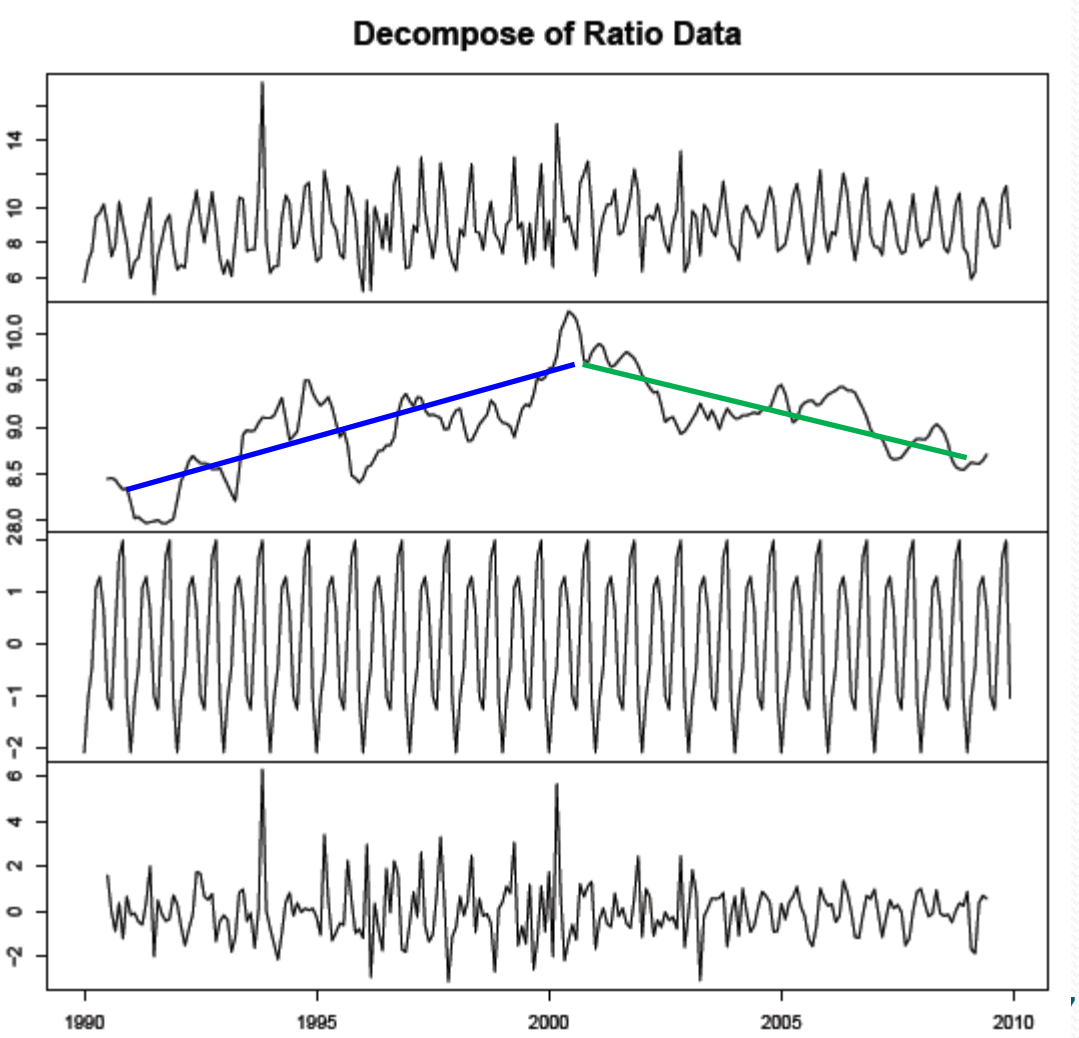




Customer Usage Ratio Method

- Ratio of average usage between non-residential vs. residential

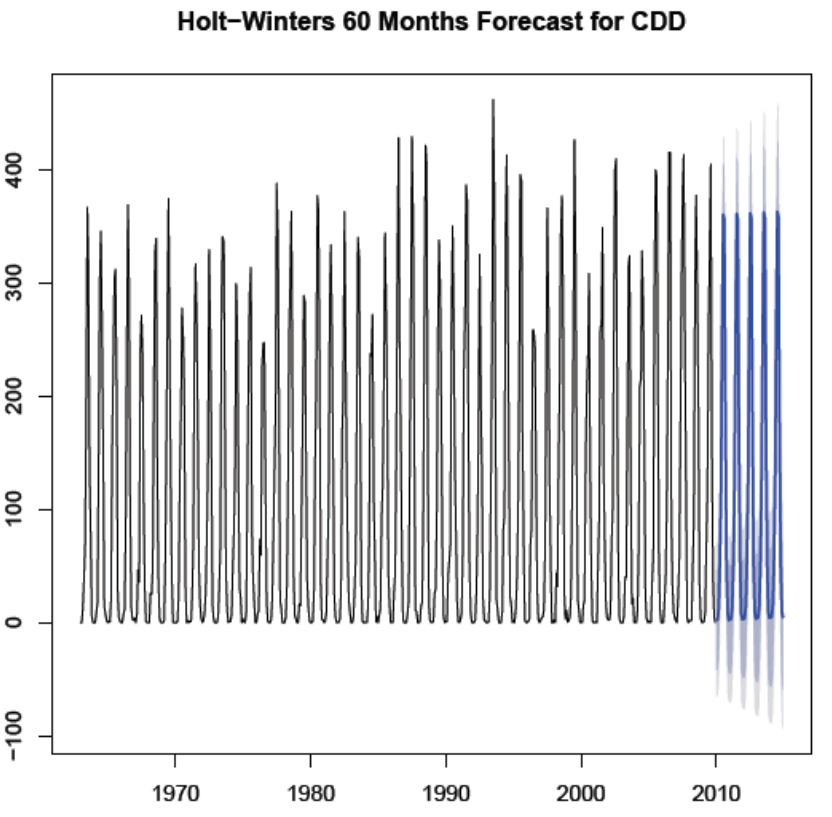
- Actual Data
- Trend
- Seasonality
- Random Error

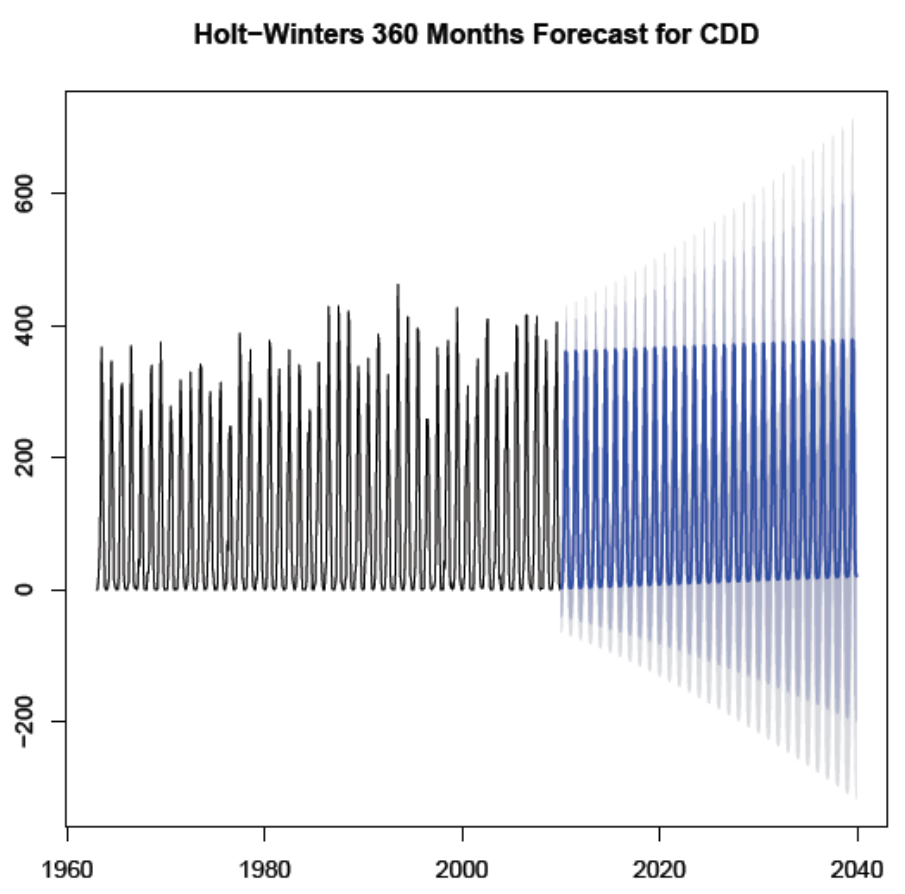


HDD/CDD Forecasting Methods

- Holt-Winters Method
- ARIMA method
  - Not good as correlogram violates control limit
- BAT Method
  - Basically a superset of Holt-Winters. No improvement over Holt-Winters Method.

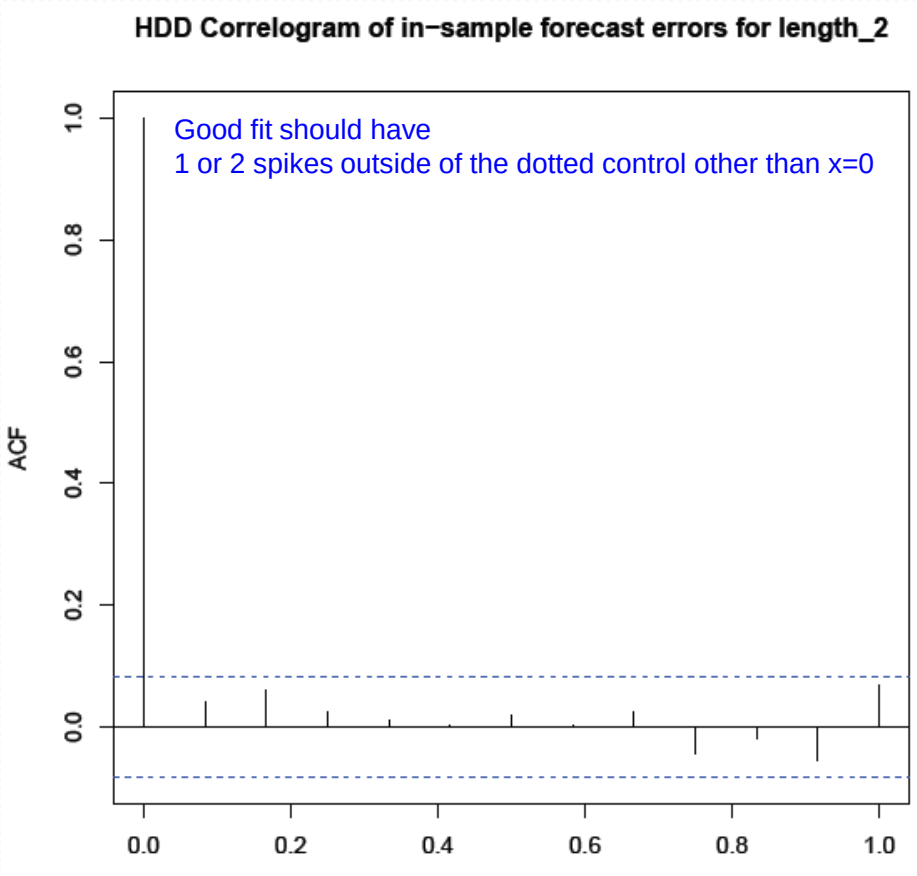
Holt-Winter Method for HDD/CDD Forecasting



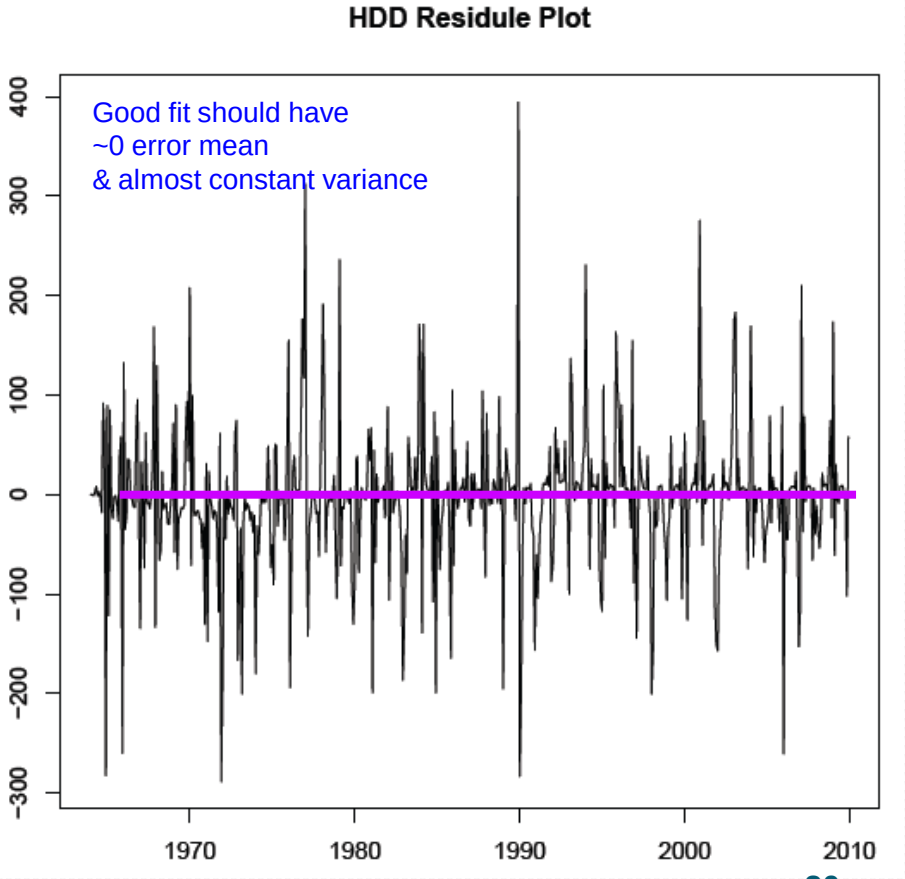


Accuracy of Holt-Winters Method

- Correlogram Plot

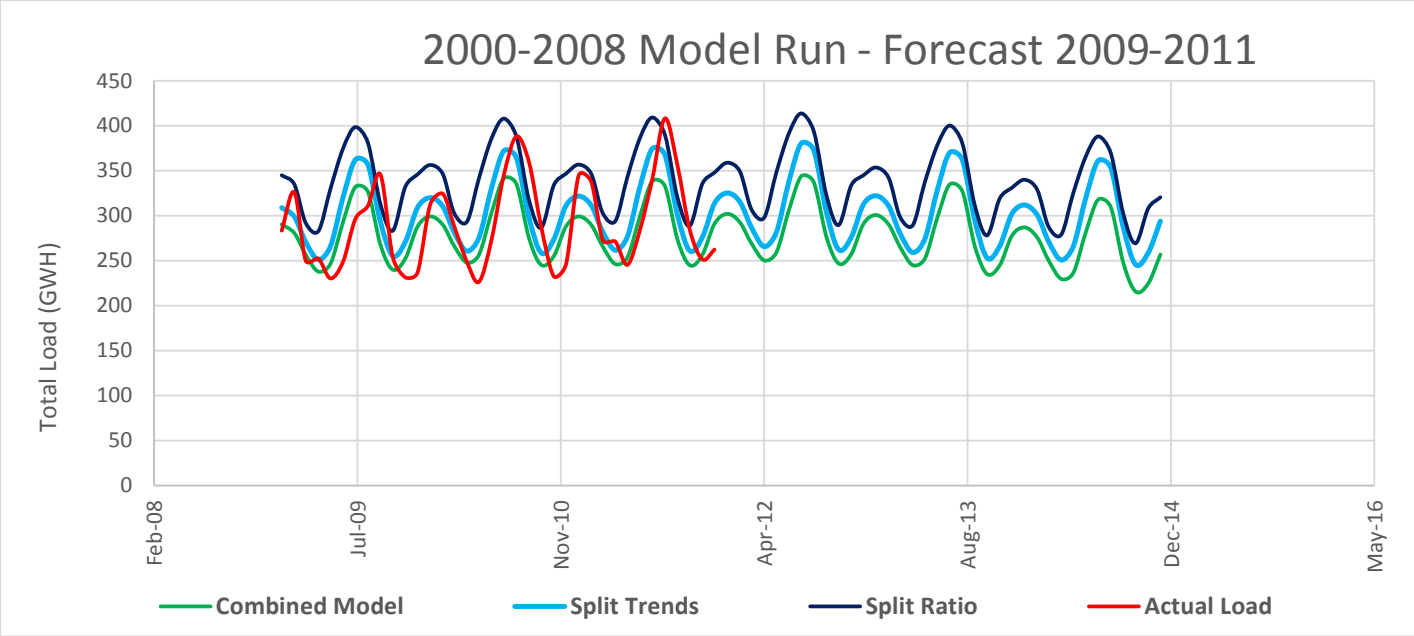


Residual Plot



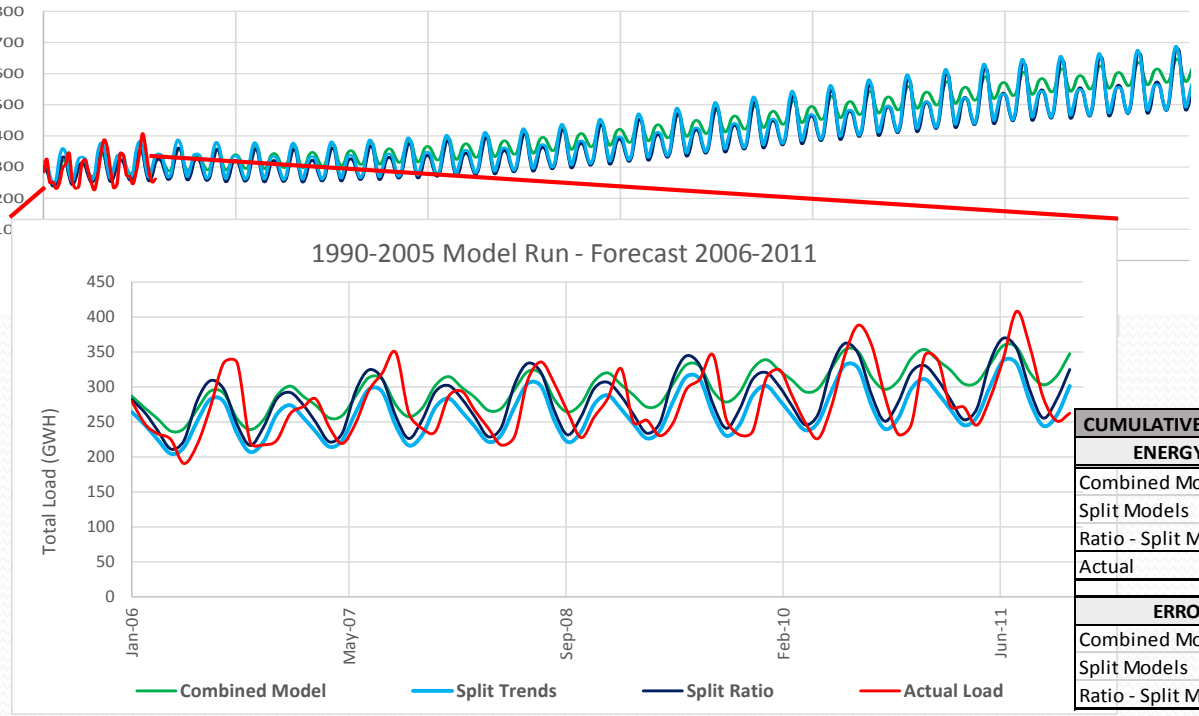
Modeling Excursions

- Tested sensitivity using different domains of time:
  - Regression models inform forecast output.
  - All historic economic variables are actual records:
    - Same for all scenarios.
    - Serve as starting point for forecast.



Model Selection

- Merit given to the balance between:
  - Bulk-energy error for first five years of forecast.
  - Robust to changes; back-tested on historic data.



- Using 1990-2005 data for regression modeling:
  - Informs monthly forecasts for 30 years.
  - Cumulative relative error assessed.
  - Robust to changes; back-tested on historic data.

CUMULATIVE LOAD (kWH)	FORECAST HORIZON				
ENERGY (kWH)	1	2	3	4	5
Combined Model	3.18E+09	6.57E+09	1.01E+10	1.37E+10	1.75E+10
Split Models	2.89E+09	5.91E+09	9.02E+09	1.22E+10	1.56E+10
Ratio - Split Models	3.09E+09	6.33E+09	9.65E+09	1.31E+10	1.67E+10
Actual	3.00E+09	6.23E+09	9.46E+09	1.27E+10	1.63E+10
ERROR (%)	1	2	3	4	5
Combined Model	6%	5%	6%	8%	8%
Split Models	-4%	-5%	-5%	-4%	-4%
Ratio - Split Models	3%	2%	2%	3%	2%

Modeling Excursion - Results

- Candidate model selection: split-model with seasonal trend forecast.

(Data Domain)	YEARLY %-ERROR in CUMULATIVE LOAD Modeling Approach	FORECAST HORIZON 1	2	3	4	5
1990-1995	Combined Model	10%	12%	14%	17%	27%
	Split Models	-2%	0%	1%	2%	3%
	Ratio - Split Models	4%	6%	6%	6%	7%
1990-2000	Combined Model	10%	12%	14%	17%	27%
	Split Models	-2%	0%	1%	2%	3%
	Ratio - Split Models	4%	6%	6%	6%	7%
1990-1995	Combined Model	4%	2%	3%	3%	2%
	Split Models	-4%	-7%	-9%	-11%	-13%
	Ratio - Split Models	4%	1%	-1%	-3%	-6%
1990-2005	Combined Model	6%	5%	6%	8%	8%
	Split Models	-4%	-5%	-5%	-4%	-4%
	Ratio - Split Models	3%	2%	2%	3%	2%
1990-2008	Combined Model	3%	3%	2%	2%	3%
	Split Models	3%	2%	0%	0%	0%
	Ratio - Split Models	10%	8%	7%	6%	6%

Recommendations

- We recommend that NOVEC use the "Split-Trend Model":
  - Mirrors methodology currently employed.
  - Potential for improvements with quantifying trends to seasonal loads (Holt-Winters method).
  - Supplement their current weather-normalization forecasting model.
- Suggestions for future analysis:
  - Determine best set of economic variables to predict future customer base;
    - Population size, and thus tested methodologies, are sensitive to this parameter.
    - Best subset is likely to change over time.
  - Determine best set of neutral zone boundaries to compute CDD/HDD.
  - Perform further analysis on the ratio method and the ARIMA HDD/CDD forecasting with preliminary de-trending.