

L.

OVERVIEW OF FORECASTING MODELS

This appendix explains and summarizes the long-term energy and demand forecasting models for Cheyenne Light and Black Hills Power.

Long-term energy forecasts use a combination of its billing data, weather data from the National Oceanic and Atmospheric Administration (NOAA), and economic and demographic data from Woods & Poole. Demand forecasts use a combination of hourly system demand data and the same weather, economic, and demographic data.

Forecasts are based on single sales or on separate use-per-customer sales for residential, commercial, industrial, and municipal models. Included are the formulas employed to develop these forecast models.

**Overview of Long-Term Energy and Demand Forecasting Models for
Black Hills Power, Inc. and Cheyenne Light Fuel and Power Company**

for

Black Hills Corporation

by

Christensen Associates Energy Consulting, LLC

800 University Bay Drive, Suite 400

Madison, WI 53705-2299

Voice 608.231.2266 Fax 608.231.2108

April 18, 2021

Table of Contents

1. Introduction.....1

2. Overview of the Forecast Development Process.....1

3. The Black Hills Power (South Dakota) Forecast6

 3.1 Residential6

 3.2 Commercial.....8

 3.3 Industrial.....9

 3.4 Municipal10

 3.5 System Peak Demand11

4. The Cheyenne Light Fuel and Power (Wyoming) Forecast12

 4.1 Residential12

 4.2 Commercial Non-Demand15

 4.3 Commercial General Service Secondary and Primary16

 4.4 Industrial.....18

 4.5 System Peak Demand19

Appendix: Estimated Models.....21

Overview of Long-Term Energy and Demand Forecasting Models for Black Hills Power, Inc. and Cheyenne Light Fuel and Power Company

for

Black Hills Corporation

by

CHRISTENSEN ASSOCIATES ENERGY CONSULTING, LLC

April 18, 2021

1. INTRODUCTION

Christensen Associates Energy Consulting, LLC (CA Energy Consulting) assisted Black Hills Corporation (Black Hills) in developing long-term energy and demand forecasts for Black Hills Power, Inc. and Cheyenne Light Fuel and Power Company. Black Hills is required to file an Integrated Resource Plan (IRP) in Wyoming and South Dakota before July 1, 2021.

Black Hills develops class-specific sales forecasts using a combination of its billing data, weather data from the National Oceanic and Atmospheric Administration (NOAA), and economic and demographic data from Woods & Poole. System demand is forecast using a combination of hourly system demand data with the weather and economic data listed above.

Section 2 provides a description of the principles we apply when developing forecasts. Section 3 describes the models developed for Black Hills Power, Inc. (BHP). Section 4 describes the models developed for Cheyenne Light Fuel and Power Company (CLFP).

2. OVERVIEW OF THE FORECAST DEVELOPMENT PROCESS

Selecting the dependent variable

Statistical forecast models begin by explaining historical variation in a dependent variable (e.g., class-level sales or use per customer (UPC)) with available explanatory variables.

Forecasts may be based on either a single sales model or separate UPC and customer count models. The latter method may be preferred for mass-market classes (e.g., residential), where intra-class customer differences are expected to be minor compared to, say, a large industrial class. Separately modeling UPC and customer counts for these classes can improve the estimates of the effect of the various explanatory variables. For example, the number of households may be a primary driver of the number of residential customers served, but not be strongly related to residential use per customer.

In each of the models presented here, we take the natural log of the dependent variable and the continuous explanatory variables.¹ This makes it easier to interpret and compare the estimated coefficients, as they represent percentage effects. If the models were instead estimated without logging the variables, the estimated coefficients would represent *level* effects whose interpretation is affected by the scale of the variables.

Selecting the explanatory variables

The explanatory variables may include the following categories:

- Weather;
- Economic conditions;
- Demographics;
- Seasonal indicators; or
- Time trends or shift variables.

Weather variables are typically based on temperatures and commonly expressed as cooling degree days (CDDs) or heating degree days (HDDs). CDDs are intended to reflect cooling-related usage and are calculated for day d as follows:

$$CDD_d = \text{MAX}\{0, (MaxTemp_d + MinTemp_d) / 2 - Threshold\}$$

The MAX function ensures that CDD values are always non-negative. $MaxTemp_d$ and $MinTemp_d$ represent the maximum and minimum temperatures for the day, respectively. $Threshold$ is the average daily temperature at which cooling load tends to begin (typically around 60°F).

HDDs reflect heating-related loads and are calculated in a similar manner as CDDs, but reversing the order of the average daily temperature and the $Threshold$:

$$HDD_d = \text{MAX}\{0, Threshold - (MaxTemp_d + MinTemp_d) / 2\}$$

Forecasting models often use monthly sales or UPC as the dependent variable, in which case the CDDs and HDDs are summed across the relevant days to form the variables used in the statistical model.

Economic factors reflect the effect of the economy on electricity use or the number of customers served. The relevant variables can vary with the customer class of interest and may include the following variables:

- Household income;
- Gross regional product (GRP); or
- Earnings, sales, or employment (total or by sector).

Demographic variables can reflect changes in the size or makeup of the utility's service territory, and may include the following variables:

- Number of households; or

¹ The exceptions for continuous variables are CDDs and HDDs, which are frequently zero and therefore drop out when logged.

- Persons per household.

Seasonal or monthly factors are “indicator” variables (sometimes called “dummy” variables)² that account for seasonal changes in usage that are not accounted for by other included explanatory variables (e.g., lighting-related usage that can vary with the hours of daylight).

Time trend and shift variables are sometimes needed to reflect changes in the dependent variable that are clearly visible in the data but are not explained by any available explanatory variables. For example, this may include changes in the definition of the customer class. Time trend variables reflect the rate of change in the dependent variable over time, while shift variables account for one-time changes in the dependent variable.

Note that any explanatory variable included in the statistical model must have both historical and forecast values to be of use in the development of the forecast. In the case of weather variables, the forecast values typically represent normal weather conditions (e.g., the average value over the previous 20 years). Economic and demographic variables are best employed when external forecasts of them are available. Black Hills uses data from Woods & Poole, which provides historical and forecast values for economic and demographic variables by county.

When evaluating explanatory variables for inclusion in statistical forecast models, we focus on the following factors:

1. Whether the included variables make intuitive sense.
2. Whether the estimated coefficients on the included variables make intuitive sense.
3. Whether the resulting forecast is a reasonable reflection of the past as well as expectations for the future.

Regarding the first point, we consider whether it’s plausible that the included economic and/or demographic variables have a causal effect on the dependent variable (e.g., use per customer, sales, or the number of customers). For example, farm employment is probably not going to drive outcomes for a customer class that does not consist of a high share of farm-related customers.

On the second point, the estimates should have the expected sign, a reasonable magnitude, and be statistically significantly different from zero.³ For example, we expect electricity use, use per customer, and the number of customers to increase as economic conditions improve. That would be reflected by a positive sign on the estimated coefficient on the economic variable.⁴

² For example, a March indicator variable would equal 1 for March observations and 0 for all other observations.

³ This is evaluated using the p-value associated with the estimate, which is based on a test that the estimated coefficient equals zero (the “null hypothesis”). If the estimated coefficient equals zero, it means that changes in the variable do not affect the dependent variable (e.g., sales). A point estimate that is not zero may be statistically equivalent to zero if the standard error associated with the estimate is sufficiently large. A low p-value (below 0.10 or 0.05) leads us to reject the null hypothesis that the variable has no effect.

⁴ A negative sign would be expected for economic variables for which higher values represent worsening conditions, such as the unemployment rate.

Demographic changes such as the change in the number of households are also expected to have specific signs. For example, an increase in the number of households should lead to increases in sales and increases in the number of customers served (a positive coefficient). Increases in persons per household may lead to increases in residential use per customer (also a positive coefficient).

Evaluating the magnitude of the coefficient requires some judgment and people may reach different conclusions. Economic variables shouldn't have outsized effects. For example, in models with logged dependent and explanatory variables, estimated coefficients larger than 1.0 mean the percentage change in the dependent variable will be larger than the percentage change in the economic variable (e.g., a 2 percent increase in GRP leads to a greater than 2 percent increase in sales). That threshold is a good starting point for judging the reasonableness of the variable, though the reasonableness of the coefficient may also be apparent in the forecast growth rate (i.e., an economic effect that is too large may lead to a growth rate in electricity sales that appears too high relative to historical rates).

Note that sometimes there are no economic variables that provide an intuitively appealing explanation of the dependent variable. This can arise when sales or use per customer are declining, perhaps due to conservation and improved energy efficiency (whether sponsored by the utility or as part of general economic or regulatory trends). In these cases, a time trend variable can be useful to allow the model to explain changes over time. In some cases, the introduction of a time trend allows the model to be able to estimate a separate and reasonable economic effect, but this is not always the case.

Finally, the model should produce a forecast that is a reasonable reflection of expectations given prior trends and Company information. For example, if sales declined steeply from 8 to 10 years ago but have remained relatively flat in more recent years, one might expect the forecast to place a higher weight on the recent (flat) trend. Applying this criterion involves exercising judgment and isn't necessarily a right vs. wrong issue (in contrast to evaluating the sign of a coefficient).

Accounting for serial correlation

Serial correlation is present when the statistical model's error (the difference between the observed value and the value predicted by the model) in a time period is related to the error in a prior time period. The presence of serial correlation does not produce biased coefficient estimates but may lead to incorrect inferences regarding a coefficient's statistical significance.

The presence of first-order serial correlation (when the current and previous observation's errors are related) is detected using the Durbin-Watson test. If the test indicates that serial correlation is present, we estimate the model using a Prais-Winsten method rather than traditional Ordinary Least Squares (OLS).

Developing High and Low Forecast Scenarios

Black Hills requested that we develop an 80 percent confidence interval around the demand and sales forecasts. That is, the forecast represents the sales and demand levels we expect to occur on average. However, considerable uncertainty remains regarding the economic conditions that will occur during the forecast period. For example, a recession could arise, or a period of sustained growth could occur. The confidence interval provides an indication of the extent to which demand and sales can vary due to such uncertainties.

In order to capture a wide range of economic conditions, we base our variability calculations on data beginning in 1969 and ending with the most recent observed data point. The data are provided by Woods & Poole and focus on the variable used in the forecast models. The variability calculation takes a mid- to long-term perspective, based on the average annual percentage change over ten-year period.

Specifically, we calculate the year-to-year percentage changes in the economic variable (e.g., gross regional product, total employment, or personal income) and then calculate 10-year moving averages of those percentage changes. The peak demand model provides us with an estimate of the effect of changes in the economic variable on changes in peak demand, along with a standard error associated with the estimate. These two uncertainties (in economic conditions over time and in the estimated effect of economic conditions on peak demand) are combined to produce the confidence interval around the demand and sales forecasts.

Here is a description of the steps we used to develop the confidence interval:

1. Calculate the average annual 10-year percentage change in the economic variable for each 10-year window between 1969 and 2017, producing 39 separate percentage change values.
2. Calculate the mean and standard deviation of the percentage changes across these 39 observations.
3. From the peak demand model, obtain the estimated coefficient and standard error associated with the included economic variable.
4. The mean expected growth rate of demand is estimated as the product of the estimated coefficient and the mean of the 39 percentage change observations.
5. The standard deviation of the growth rate of demand is estimated by combining the standard error of the estimated coefficient with the standard deviation of the historical percentage changes in the economic variable.⁵
6. The coefficient of variation (CV) of the economic-based variability is calculated as the standard deviation calculated in step 5 divided by the mean expected growth rate calculated in step 4.
7. For any given forecast value, the high and low scenarios are simulated as the 90th and 10th percentile values (respectively) from a normal distribution, with a mean equal to

⁵ This calculation is performed as the standard deviation of the product of two random variables, as follows:

$$\text{Var}(XY) = \text{Var}(X)\text{Var}(Y) + \text{Var}(X)(E(Y))^2 + \text{Var}(Y)(E(X))^2$$

the “base” forecast growth rate and the standard deviation equal to the absolute value of the base forecast growth rate⁶ multiplied by the CV calculated in Step 6.

8. These high and low percentages are applied to the demand and sales forecasts in each of the forecast months.

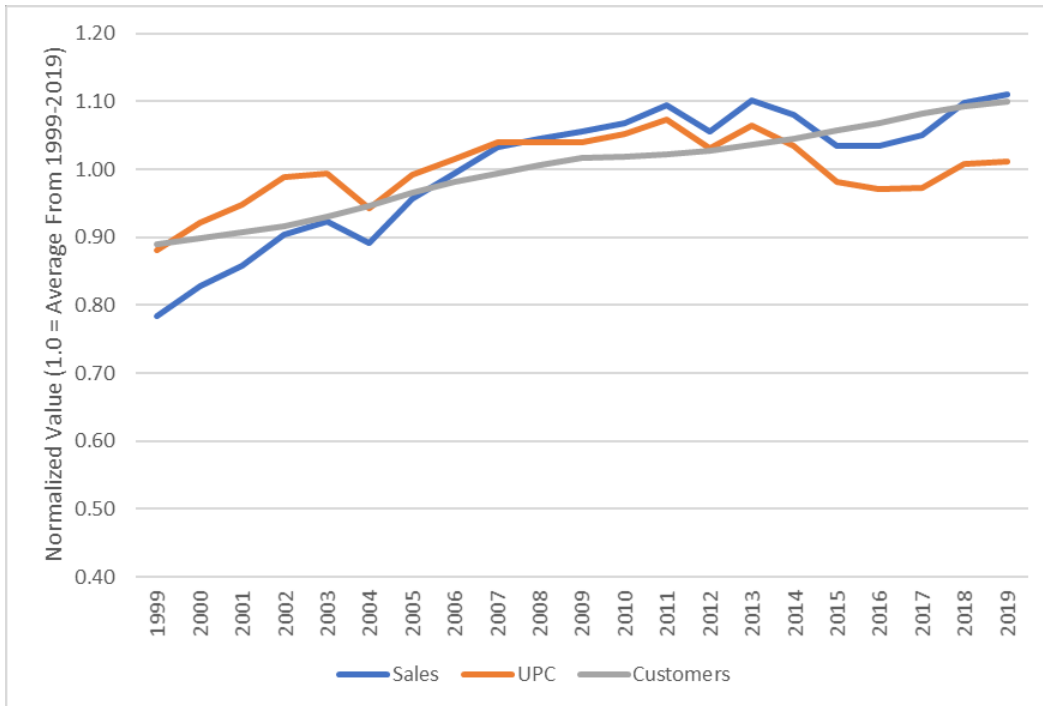
3. THE BLACK HILLS POWER (SOUTH DAKOTA) FORECAST

In this section, we describe the forecasting models for each customer class in the Black Hills Power (BHP) service territory. In each case, we show a graph reflecting historical annual sales, use per customer (UPC), and the number of customers over time. Each series is normalized to show a value that is indexed to the average across the time period shown (i.e., a value of 0.9 means that year’s value is 90 percent of the average over the time period shown). This normalization facilitates a comparison of trends in the outcomes across years, which naturally occur on different scales. The figures show observed (non-weather normalized) values. The Appendix contains detailed results for each forecast model.

3.1 Residential

Figure 3.1 shows the normalized sales, UPC, and customer counts for BHP’s Residential customer class. The upward trend in sales appears to be primarily driven by growth in customers served, while year-to-year variations in total sales are highly correlated with those of UPC. We estimate separate UPC and customer models to better account for these separate effects.

⁶ Taking the absolute value of the forecast growth rate is necessary because standard deviations cannot be negative.

Figure 3.1: BHP Residential Normalized Sales, UPC, and Customer Counts

The Residential UPC model is:

$$\ln(\text{upc}_t) = a + b^{CDD} \times CDD_t + b^{HDD} \times HDD_t + b^{Trend} \times Trend_t + \sum_m (b^m \times Month_{m,t}) + e_t$$

In this equation, a and the b 's are estimated parameters; e_t is the error term; t indexes time periods; and m indexes months. The explanatory variables are:

- CDD_t = CDD using a 60°F threshold
- HDD_t = HDD using a 60°F threshold
- $Trend_t$ = Time trend
- $Month_{m,t}$ = month dummies⁷

The model is estimated using data from 2007 through 2019 using the Prais-Winsten serial correlation correction. No available economic or demographic variables produced a reasonable coefficient estimate. Data prior to 2007 is excluded due to the high growth in UPC during that period vs. more recent years. The time trend accounts for the slight downward trend in UPC following 2007 (approximately 0.5 percent per year).

⁷ The model includes eleven month-specific dummies, with the January variable omitted to prevent perfect multicollinearity of the month variables. That is, the coefficients for the included months are interpreted as an effect relative to the omitted month.

The Residential customer model is:

$$\ln(\text{custs}_t) = a + b^{\text{Emp}} \times \ln(\text{TotEmp}_t) + \sum_m(b^m \times \text{Month}_{m,t}) + e_t$$

The explanatory variables are:

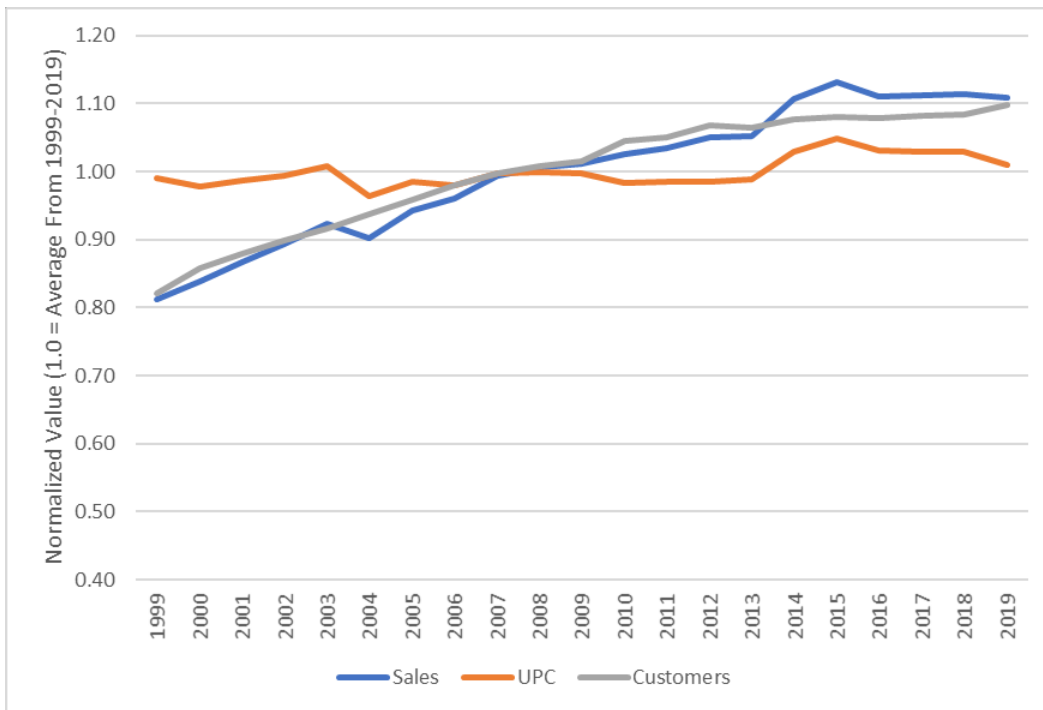
- $\ln(\text{TotEmp}_t)$ = the natural log of total employment (12-month moving average)
- $\text{Month}_{m,t}$ = month dummies

The model is estimated using data from 2007 through 2019 using the Prais-Winsten serial correlation correction. The estimated coefficient on the total employment variable reflects a positive relationship between economic conditions and the number of customers served.

3.2 Commercial

As Figure 3.2 shows, BHP’s Commercial UPC (and as a result, total sales) increased to a higher level beginning around 2014. This upward shift appears to be due to customers changing classes, resulting in an influx of customers that led to a one-time shift in UPC. Class sales, which had been increasing prior to 2014, were largely flat following the class shift.

Figure 3.2: BHP Commercial Normalized Sales, UPC, and Customer Counts



The Commercial UPC model is:

$$\ln(\text{upc}_t) = a + b^{\text{CDD}} \times \text{CDD}_t + b^{\text{HDD}} \times \text{HDD}_t + b^{\text{Shift}} \times \text{ClassShift}_t + b^{\text{Trend}} \times \text{Trend}_t + b^{\text{Emp}} \times \ln(\text{TotEmp}_t) + \sum_m(b^m \times \text{Month}_{m,t}) + e_t$$

The explanatory variables are:

- CDD_t = CDD using a 60°F threshold
- HDD_t = HDD using a 60°F threshold
- $ClassShift_t$ = a “class shift” indicator variable equal to 1 beginning in June 2014 and 0 prior to that month
- $Trend_t$ = Time trend
- $\ln(TotEmp_t)$ = the natural log of total employment (12-month moving average)
- $Month_{m,t}$ = month dummies

The model is estimated using data from 1999 through 2019 using the Prais-Winsten serial correlation correction. The estimated coefficients for the employment and time trend variables reflect offsetting effects. Commercial UPC increases with employment, with a separate downward trend of approximately 0.6 percent per year. The estimated coefficient for the class shift variable indicates 6.1 percent higher UPC during the post-June 2014 period.

The Commercial customer model is:

$$\ln(custs_t) = a + b^{Emp} \times \ln(TotEmp_t) + b^{Emp_Shift} \times (TotEmp_t \times ClassShift_t) + b^{Shift} \times ClassShift_t + \sum_m (b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- $\ln(TotEmp_t)$ = The natural log of total employment (12-month moving average)
- $ClassShift_t$ = A “class shift” indicator variable equal to 1 beginning in June 2014 and 0 prior to that month
- An interaction between the $\ln(\text{total employment})$ variable and the class shift variable
- $Month_{m,t}$ = month dummies

The model is estimated using data from 1999 through 2019 using the Prais-Winsten serial correlation correction. The interaction between the class shift variable and the total employment variable allows the effect of employment to differ before and after the class shift occurs. The estimates reflect a much higher employment effect in the pre-shift period.

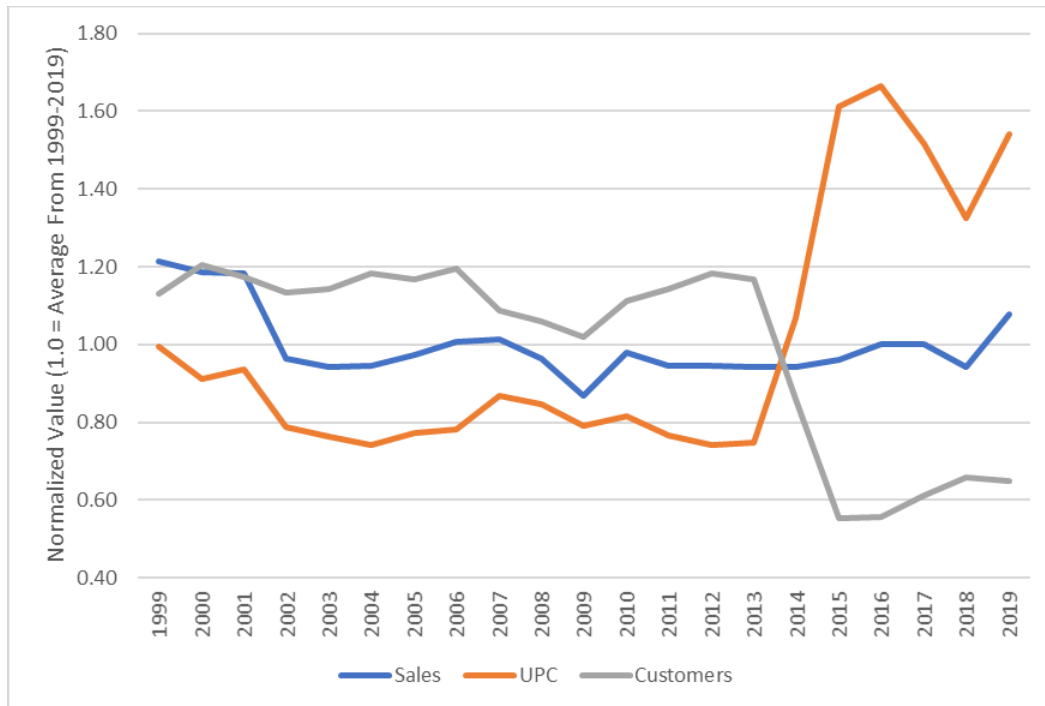
The forecast produced by this model had a reasonable annual growth rate but some prediction error in the final year that resulted in a forecast that started from a level that appeared to be too high. To remedy this, we applied the forecast percentage growth rate to the last year’s weather normalized sales. The weather normalization adjustment was developed as the difference between the model’s predicted sales at normal and observed weather. That difference was added to observed sales to arrive at weather-normalized sales.

3.3 Industrial

This class is not forecast using a statistical model, with flat sales (i.e., no growth) assumed during the forecast period. Figure 3.3 shows the reasonableness of this assumption. The class

shift described above for the Commercial class affected this class as well, resulting in a reduction in the number of customers and an increase in UPC. Note that the effect of the shift is more pronounced for this class, as it has fewer customers than the Commercial class (25 to 40 Industrial customers vs. more than 12,000 Commercial customers). Because the shifted customers represented a relatively low share of class sales, the class sales remained relatively flat through the 2014 class-shift period.

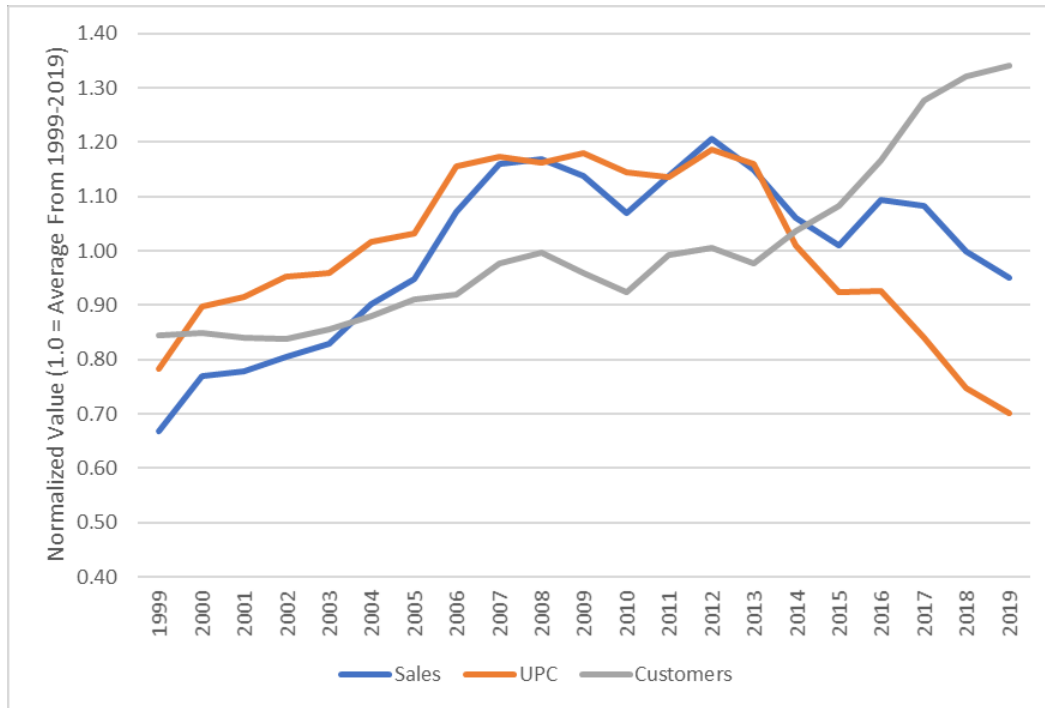
Figure 3.3: BHP Industrial Normalized Sales, UPC, and Customer Counts



3.4 Municipal

Sales to BHP's Municipal class have displayed varying dynamics from 1999 to 2019, with rapid increases through 2007 followed by a plateau and an eventual decline. No economic or demographic variables explain these changes over time. As a result, our forecasting model focuses on following the observed trends and basing the forecast on the post-2007 experience. Note that the Municipal class accounts for a small percentage of BHP's total sales (1.1% in 2019).

Figure 3.4: BHP Municipal Normalized Sales, UPC, and Customer Counts



The Municipal sales model is:

$$\ln(ups_t) = a + b^{CDD} \times CDD_t + b^{D2007} \times D2007_t + b^{Trend} \times Trend_t + b^{Trend07} \times (Trend_t \times D2007_t) + \sum_m(b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- CDD_t = CDD using a 60°F threshold
- $D2007_t$ = a2007 indicator variable equal to 1 beginning in January 2007 and 0 prior to that month
- $Trend_t$ = Time trend
- An interaction between the 2007 indicator variable and the time trend
- $Month_{m,t}$ = month dummies

The model is estimated using data from 1999 through 2019 using the Prais-Winsten serial correlation correction. The estimated time trends show approximately 5.3 percent per year growth through 2006, with a -1.3 percent per year change in sales from 2007 on.

3.5 System Peak Demand

Forecasting system peak demand presents different challenges than forecasting monthly sales. The objective of the statistical model is to explain the factors that contribute to the most extreme observed loads. To increase the sample size of “peak-like” hours, we include all hours that are within 1 percent of each month’s peak demand value.

The system demand model is:

$$\ln(MW_t) = a + b^{CDD} \times CDD_t + b^{HDD} \times HDD_t + b^{CDD-d} \times CDD_Day_t + b^{HDD-d} \times HDD_Day_t + b^{Wknd} \times Weekend_t + b^{PI} \times \ln(TotPI_t) + \sum_m (b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- CDD_t = the date's CDD using a 60°F threshold
- HDD_t = the date's HDD using a 60°F threshold
- CDD_Day_t = average CDD per day during the month
- HDD_Day_t = average HDD per day during the month
- $\ln(TotPI_t)$ = the natural log of total personal income
- $Weekend_t$ = a weekend indicator variable (equal to 1 on weekends and zero on weekdays)
- $Month_{m,t}$ = month dummies

The date specific CDD and HDD variables account for the effect of the day's temperatures on the peak day's loads. The monthly average CDD and HDD variables reflect the overall weather conditions (e.g., heat or cold buildup) surrounding the peak day. The personal income variable reflects the effect of economic conditions on peak demand. The weekend indicator variable allows the model to explain the fact that weekend peaks are lower than weekday peaks, all else equal (by approximately 2.3 percent, according to our estimate). The month dummies reflect seasonal patterns in peak demand.

The model is estimated using data from 2010 through 2019. No correction is made for serial correlation.⁸

4. THE CHEYENNE LIGHT FUEL AND POWER (WYOMING) FORECAST

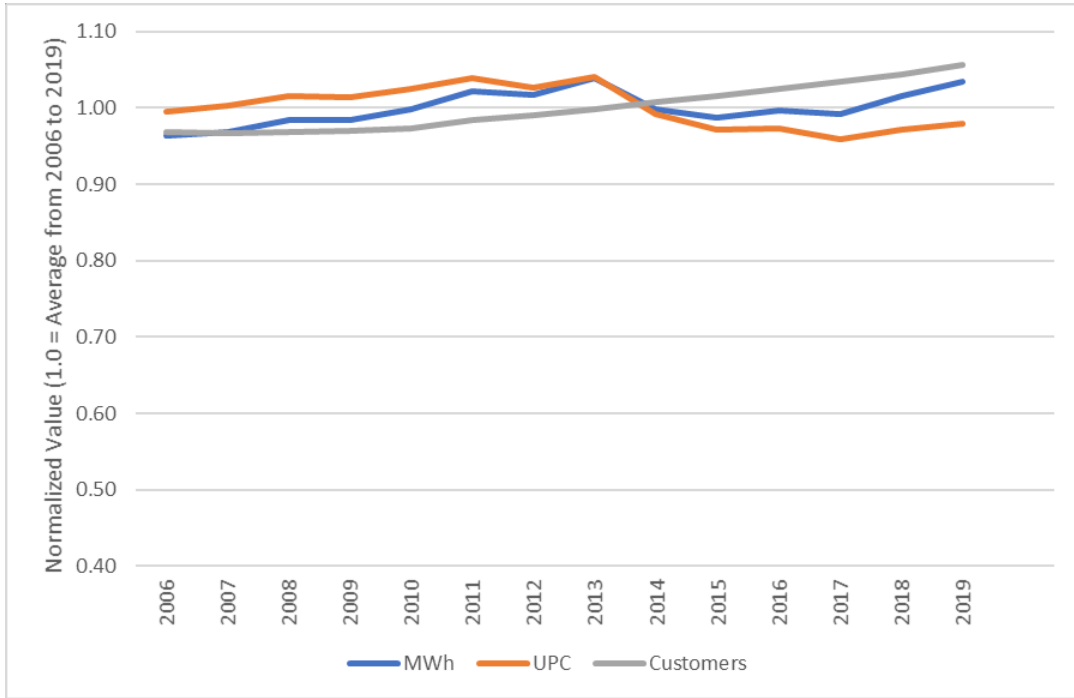
This section contains a description of each CLFP forecast model. The Appendix provides detailed results for each model.

4.1 Residential

Figure 4.1 shows the normalized sales, UPC, and customer counts for CLFP's Residential customer class. The overall upward trend in sales appears to be primarily driven by growth in customers served, while year-to-year variations in total sales are highly correlated with those of UPC. UPC (and therefore sales) drops in the years following 2013 but recovers somewhat in the most recent years. We estimate separate UPC and customer models to better account for these separate effects.

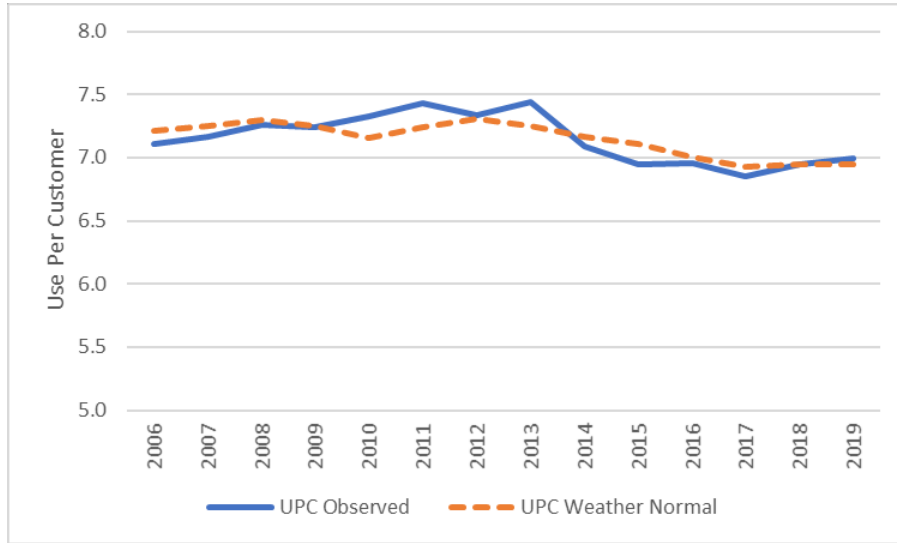
⁸ Unlike the monthly class sales models, the interval between observations can vary in the peak demand model. This makes it difficult to identify and correct for serial correlation.

Figure 4.1: CLFP Residential Normalized Sales, UPC, and Customer Counts



We examined weather normalized UPC to test whether the dip in UPC that occurs from 2014 through 2017 was due to mild weather. In Figure 4.2 below, the blue line represents observed UPC while the dashed orange line reflects weather normalized UPC. We conclude from this that some of the reduction in UPC was due to weather, but the decline was still somewhat steady through those years.

Figure 4.2: BHP Residential Observed vs. Weather Normalized UPC



The Residential UPC model is:

$$\ln(upc_t) = a + b^{CDD} \times CDD_t + b^{HDD} \times HDD_t + b^{Inc} \times \ln(HhldInc_t) + b^{Trend} \times Trend_t + \sum_m(b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- CDD_t = CDD using a 60°F threshold
- HDD_t = HDD using a 60°F threshold
- $\ln(HhldInc_t)$ = the natural log of real household total personal income (12-month moving average)
- $Trend_t$ = Time trend
- $Month_{m,t}$ = month dummies

The model is estimated using data from February 2005 through December 2019 using the Prais-Winsten serial correlation correction. The estimated time trend reflects a 0.7 percent per year decline in UPC, which is offset to some extent by the positive relationship between household income and UPC.

The Residential customer model is:

$$\ln(custs_t) = a + b^{Hhld_pre} \times \{\ln(Hhlds_t) \times Pre2010_t\} + b^{Hhld_CIS} \times \{\ln(Hhlds_t) \times CISplus_t\} + b^{CIS} \times CISplus_t + \sum_m(b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- $\ln(Hhlds_t) \times Pre2010_t$ = the natural log of the number of households interacted with a pre-2010 indicator variable

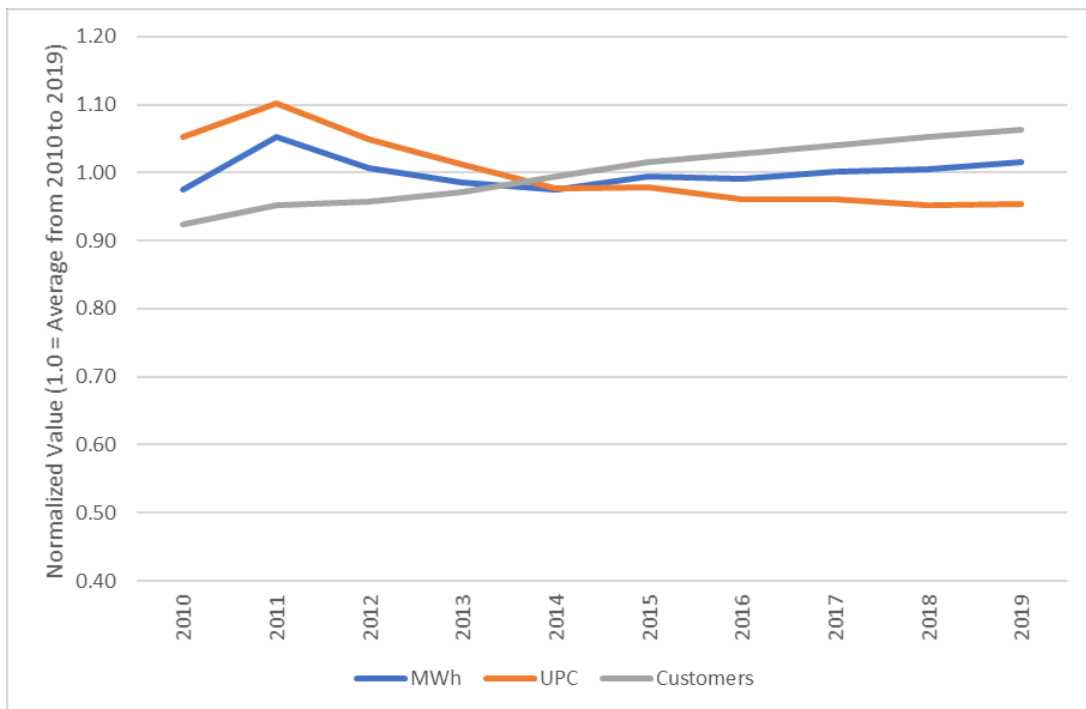
- $\ln(Hhlds_t) \times CISplus_t$ = the natural log of the number of households interacted with a 2010+ indicator variable
- $CISplus_t$ = a 2010+ indicator variable, reflecting the approximate date that CLFP’s new CIS system was implemented (and thus may have affected the recording of customer counts)

The model is estimated using data from February 2005 through December 2019 using the Prais-Winsten serial correlation correction. The number of households is positively related to the number of customers in the 2010+ period, with no statistically significant relationship estimated in the preceding years.

4.2 Commercial Non-Demand

Figure 4.3 shows large changes in sales and UPC for CLFP’s Commercial Non-Demand customers during the 2010 to 2013 period, followed by a more stable period through 2019. In contrast, the number of customers increases steadily through the 2010 to 2019 period.

Figure 4.3: CLFP Commercial Non-Demand Normalized Sales, UPC, and Customer Counts



The Commercial Non-Demand UPC model is:

$$\ln(upc_t) = a + b^{CDD} \times CDD_t + b^{HDD} \times HDD_t + b^{Trend} \times Trend_t + \sum_m (b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- CDD_t = CDD using a 60°F threshold
- HDD_t = HDD using a 60°F threshold
- $Trend_t$ = Time trend
- $Month_{m,t}$ = month dummies

The model is estimated using data from 2014 through 2019 using the Prais-Winsten serial correlation correction. No available economic or demographic variables produced a reasonable estimate. Data prior to 2014 is excluded due to the unexplained variability in UPC relative to more recent years. The time trend accounts for the slight downward trend in UPC from 2014 through 2019, at approximately 0.7 percent per year.

In the Commercial Non-Demand customer model, the sole explanatory variable is the natural log of total employment (12-month moving average).

$$\ln(custs_t) = a + b^{Emp} \times \ln(TotEmp_t) + e_t$$

We tested monthly indicator variables but found that they were not jointly statistically significant. The estimate on the employment variable indicates a positive relationship between economic conditions and the number of customers served.

4.3 Commercial General Service Secondary and Primary

Because of inter-class customer migrations during recent years, the forecast combines CLFP's General Service Secondary and Primary customers into a single forecast. Figures 4.4 and 4.5 show the changes in sales, UPC, and customer counts for each group. Notice how since 2013 sales have been persistently decreasing in the Secondary class and increasing in the Primary class. Customers re-classifying from Secondary to Primary are at least partially responsible for these trends. Figure 4.6 shows the corresponding values for the two classes combined, revealing a flatter sales trend since 2013.

Figure 4.4: CLFP GS Secondary Normalized Sales, UPC, and Customer Counts

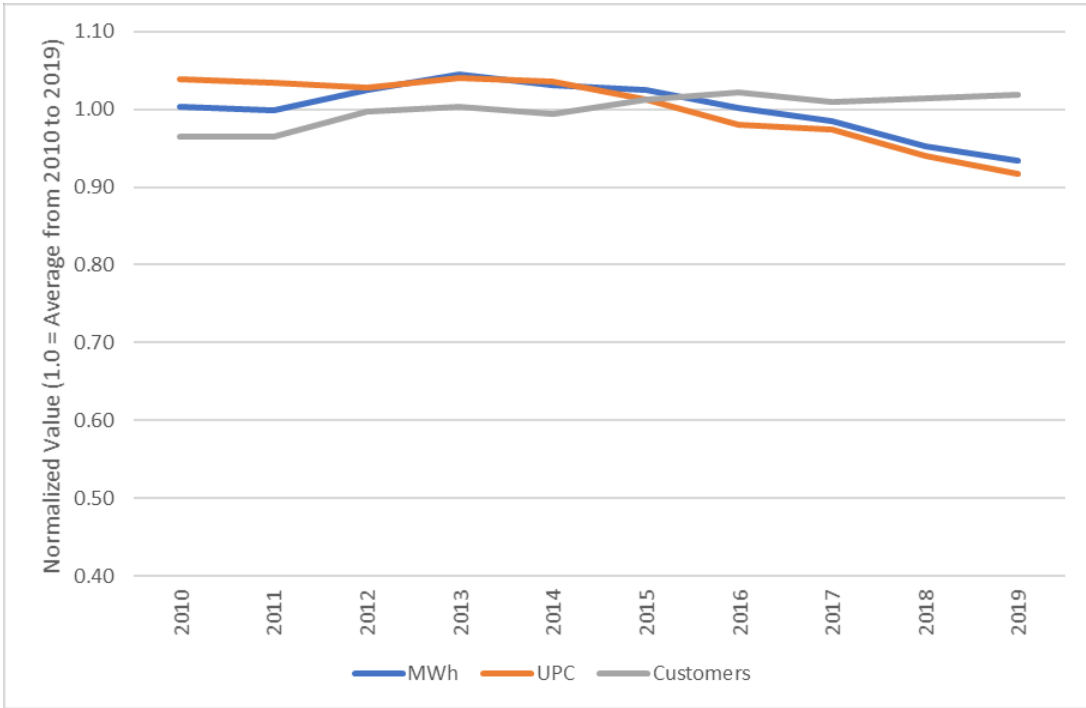


Figure 4.5: CLFP GS Primary Normalized Sales, UPC, and Customer Counts

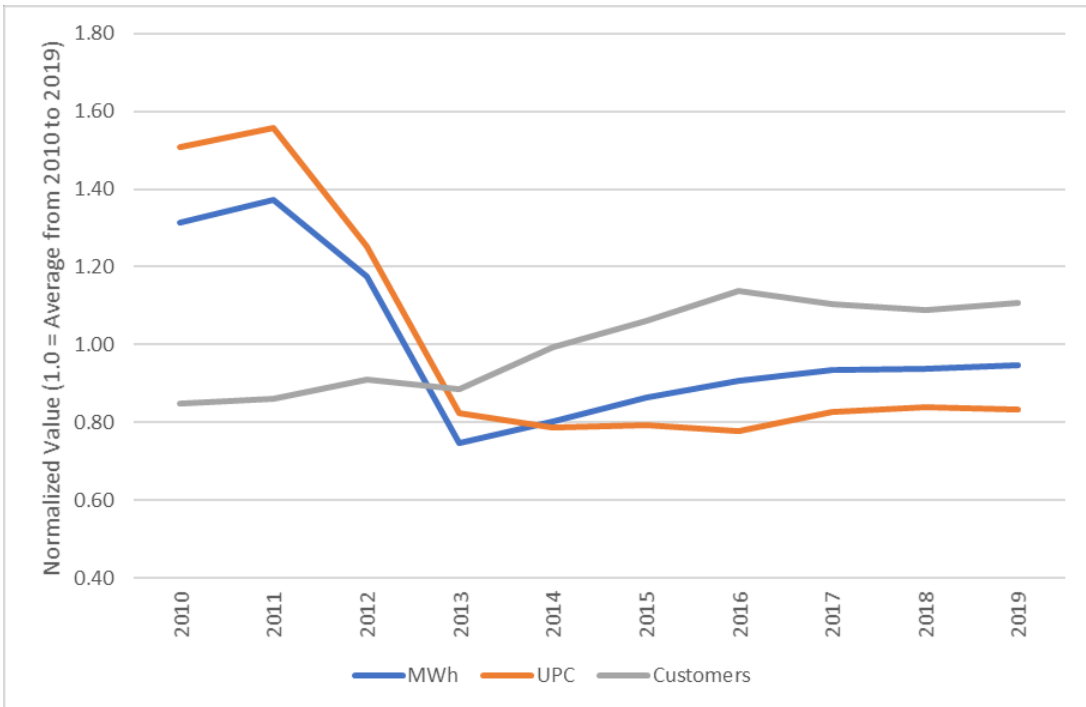
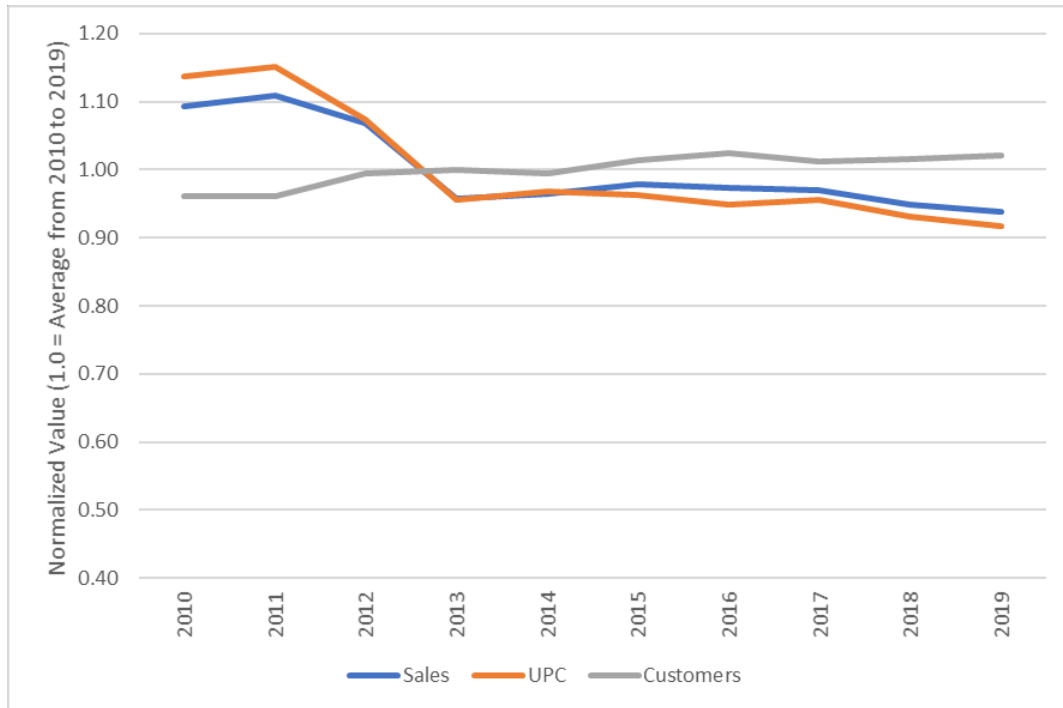


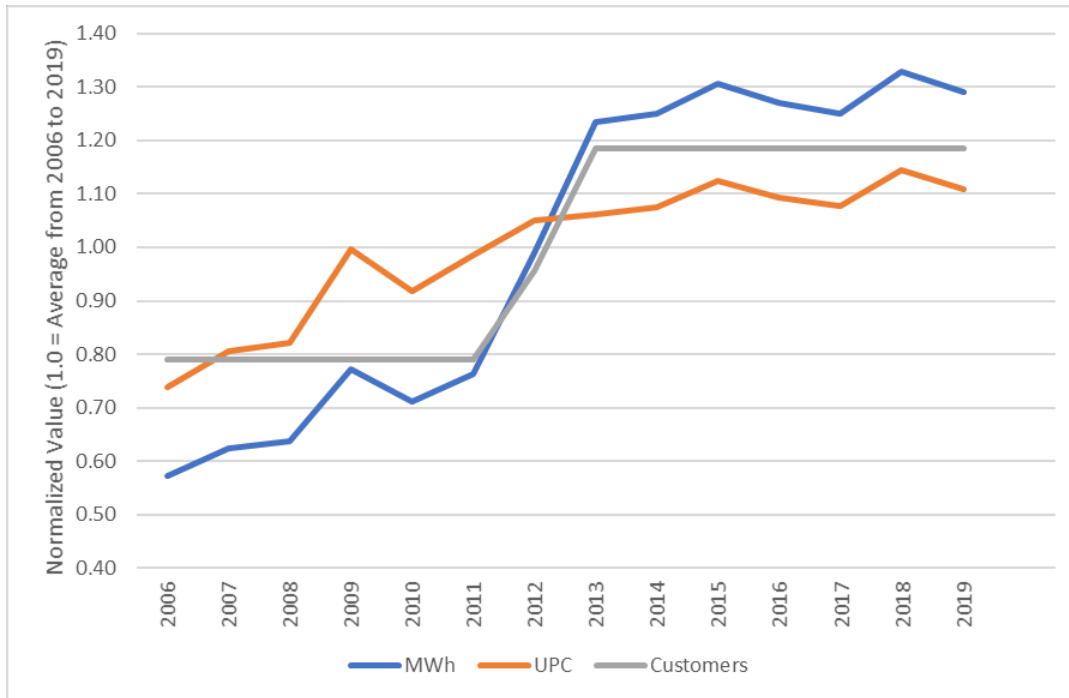
Figure 4.6: CLFP GS Secondary + Primary Normalized Sales, UPC, and Customer Counts

Statistical forecasting models developed for the combined General Service class produced declining sales, particularly in the 2030s and beyond. This contradicts Black Hills's expectations for this class, which is that sales will remain flat during the forecast period. Therefore, for this class Black Hills uses a forecast assumption of flat sales rather than a statistically based forecast.

4.4 Industrial

This class is not forecast using a statistical model, which is appropriate given that it only has two or three customers. With so few customers, variations in sales are likely more due to idiosyncratic effects on individual companies rather than reflections of widespread trends, making them difficult to explain using the data at hand. As Figure 4.7 shows, Industrial sales increase when a customer is added but have remained relatively constant in recent years.

Figure 4.7: CLFP Industrial Normalized Sales, UPC, and Customer Counts



4.5 System Peak Demand

As was the case for the BHP system peak demand model, the CLFP model includes all hours that are within 1 percent of each month's peak demand value.

The system demand model is:

$$\ln(MW_t) = a + b^{CDD} \times CDD_t + b^{HDD} \times HDD_t + b^{CDD_d} \times CDD_Day_t + b^{CDH} \times CDH_t + b^{HDH} \times HDH_t + b^{Emp} \times \ln(TotEmp_t) + \sum_m (b^m \times Month_{m,t}) + e_t$$

The explanatory variables are:

- CDD_t = the date's CDD using a 60°F threshold
- HDD_t = the date's HDD using a 60°F threshold
- CDD_Day_t = average CDD per day during the month
- CDH_t = Cooling degree hours (CDHs) during the peak hour⁹
- HDH_t = Heating degree hours (HDHs) during the peak hour¹⁰
- $\ln(TotEmp_t)$ = The natural log of total employment
- $Month_{m,t}$ = month dummies

⁹ $CDH_h = \text{MAX}\{0, Temp_h - 70\}$, where h is the hour in question.

¹⁰ $HDH_h = \text{MAX}\{0, 50 - Temp_h\}$, where h is the hour in question.

The date specific CDD and HDD variables account for the effect of the day's temperatures on the peak day's loads. The CDH and HDH variables reflect temperatures in the peak hour itself. The monthly average CDD variable reflects the overall weather conditions (e.g., heat buildup) surrounding the peak day. The total employment variable reflects the effect of economic conditions on peak demand. The month dummies reflect seasonal patterns in peak demand.

The model is estimated using data from 2008 through 2019. As with the BHP peak demand model, no correction is made for serial correlation.

APPENDIX: ESTIMATED MODELS

BHP Residential UPC Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	156
Model	6.59638489	14	.47117035	F(14, 141)	=	279.48
Residual	.23771102	141	.001685894	Prob > F	=	0.0000
				R-squared	=	0.9652
				Adj R-squared	=	0.9618
Total	6.83409591	155	.044090941	Root MSE	=	.04106

lupc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cdd60	.0007491	.0000792	9.46	0.000	.0005925 .0009057
hdd60	.0003371	.0000268	12.56	0.000	.000284 .0003901
trend	-.0053457	.0011228	-4.76	0.000	-.0075653 -.003126
m2	-.0809981	.0146294	-5.54	0.000	-.1099194 -.0520769
m3	-.0765306	.0176738	-4.33	0.000	-.1114705 -.0415906
m4	-.1715452	.0214173	-8.01	0.000	-.2138857 -.1292046
m5	-.2487149	.0261492	-9.51	0.000	-.3004101 -.1970197
m6	-.2838251	.0317412	-8.94	0.000	-.3465753 -.221075
m7	-.220695	.0402097	-5.49	0.000	-.3001869 -.1412031
m8	-.1982313	.0444751	-4.46	0.000	-.2861556 -.1103071
m9	-.2625503	.0377637	-6.95	0.000	-.3372066 -.187894
m10	-.3046907	.0290236	-10.50	0.000	-.3620684 -.2473129
m11	-.2245492	.0227044	-9.89	0.000	-.2694342 -.1796643
m12	-.0621735	.0159359	-3.90	0.000	-.0936777 -.0306693
_cons	6.704262	.0353275	189.77	0.000	6.634422 6.774102
rho	.2203338				

Durbin-Watson statistic (original) 1.688659
 Durbin-Watson statistic (transformed) 2.075256

BHP Residential Customer Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	156
Model	11.7509368	12	.979244731	F(12, 143)	>	99999.00
Residual	.000585916	143	4.0973e-06	Prob > F	=	0.0000
				R-squared	=	1.0000
				Adj R-squared	=	0.9999
Total	11.7515227	155	.075816275	Root MSE	=	.00202

lcust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lntotemp12	.8832825	.0454891	19.42	0.000	.7933645 .9732005
m2	-.0001767	.0005589	-0.32	0.752	-.0012813 .000928
m3	.0006558	.0007511	0.87	0.384	-.0008289 .0021405
m4	.0001208	.0008707	0.14	0.890	-.0016003 .0018418
m5	.0001964	.0009466	0.21	0.836	-.0016746 .0020675
m6	.0004601	.0009896	0.46	0.643	-.0014959 .0024161
m7	-.0002992	.0010043	-0.30	0.766	-.0022843 .0016859
m8	.001199	.0009921	1.21	0.229	-.0007621 .00316
m9	.0015761	.0009518	1.66	0.100	-.0003053 .0034574
m10	.0011851	.000879	1.35	0.180	-.0005525 .0029226
m11	.0006704	.0007636	0.88	0.381	-.000839 .0021799
m12	.0014634	.000579	2.53	0.013	.0003189 .0026079
_cons	7.853703	.158067	49.69	0.000	7.541253 8.166152
rho	.9162612				

Durbin-Watson statistic (original) 0.198497
 Durbin-Watson statistic (transformed) 2.673149

BHP Commercial UPC Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	252
Model	3.2015146	16	.200094662	F(16, 235)	=	152.52
Residual	.308297984	235	.001311906	Prob > F	=	0.0000
				R-squared	=	0.9122
				Adj R-squared	=	0.9062
Total	3.50981258	251	.013983317	Root MSE	=	.03622

lupc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cdd60	.0004078	.0000516	7.90	0.000	.0003062 .0005095
hdd60	.0000878	.0000188	4.68	0.000	.0000508 .0001248
m2	-.0520282	.011984	-4.34	0.000	-.075638 -.0284183
m3	-.0386922	.0118223	-3.27	0.001	-.0619834 -.015401
m4	-.0847384	.0139929	-6.06	0.000	-.1123059 -.0571708
m5	-.0978604	.0170247	-5.75	0.000	-.1314009 -.0643199
m6	-.0237413	.0208928	-1.14	0.257	-.0649023 .0174197
m7	.0054528	.0265259	0.21	0.837	-.0468061 .0577117
m8	.0245937	.0296813	0.83	0.408	-.0338816 .0830691
m9	-.0073286	.0249257	-0.29	0.769	-.056435 .0417778
m10	-.0483032	.0188734	-2.56	0.011	-.0854859 -.0111205
m11	-.0999738	.0147761	-6.77	0.000	-.1290844 -.0708633
m12	-.0056054	.0123565	-0.45	0.651	-.029949 .0187383
class_shift	.0605904	.0081693	7.42	0.000	.044496 .0766848
trend	-.0061663	.0028089	-2.20	0.029	-.01117 -.0006325
lntotemp12	.4757653	.236463	2.01	0.045	.0099071 .9416235
_cons	6.844035	.7777653	8.80	0.000	5.311752 8.376318

rho | -.1303725

Durbin-Watson statistic (original) 2.237867
 Durbin-Watson statistic (transformed) 1.976362

BHP Commercial Customer Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	252
Model	14.0120084	14	1.00085775	F(14, 237)	=	13152.44
Residual	.018034924	237	.000076097	Prob > F	=	0.0000
				R-squared	=	0.9987
				Adj R-squared	=	0.9986
Total	14.0300434	251	.055896587	Root MSE	=	.00872

lcust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
m2	-.0038088	.0020341	-1.87	0.062	-.007816 .0001983
m3	-.0015248	.0026583	-0.57	0.567	-.0067617 .0037121
m4	.0036183	.0030146	1.20	0.231	-.0023205 .0095571
m5	.0125077	.003226	3.88	0.000	.0061525 .0188629
m6	.0186227	.003345	5.57	0.000	.012033 .0252123
m7	.0214386	.0033801	6.34	0.000	.0147797 .0280976
m8	.0277128	.0033456	8.28	0.000	.0211218 .0343038
m9	.0211541	.0032357	6.54	0.000	.0147796 .0275286
m10	.0148503	.0030301	4.90	0.000	.008881 .0208197
m11	.0087593	.0026822	3.27	0.001	.0034752 .0140433
m12	.000228	.0020735	0.11	0.913	-.003857 .0043129
lntotemp12	1.490022	.0418228	35.63	0.000	1.407631 1.572414
lntotemp_shift	-1.292099	.2045018	-6.32	0.000	-1.694973 -.8892259
class_shift	4.491643	.7154039	6.28	0.000	3.08228 5.901006
_cons	4.291352	.1422813	30.16	0.000	4.011054 4.57165

rho | .7318557

Durbin-Watson statistic (original) 0.548459
 Durbin-Watson statistic (transformed) 2.435974

BHP Municipal Sales Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	252
Model	13.5463717	15	.903091449	F(15, 236)	=	82.04
Residual	2.5978093	236	.011007667	Prob > F	=	0.0000
				R-squared	=	0.8391
				Adj R-squared	=	0.8289
Total	16.144181	251	.064319446	Root MSE	=	.10492

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cdd60	.0010164	.0001617	6.29	0.000	.0006978 .001335
trend	.0533716	.0061685	8.65	0.000	.0412193 .0655239
trend_d2007	-.0661732	.0068731	-9.63	0.000	-.0797138 -.0526326
d2007	.7528875	.0587961	12.81	0.000	.6370553 .8687198
m2	-.1061182	.0289156	-3.67	0.000	-.1630839 -.0491526
m3	-.0802588	.032359	-2.48	0.014	-.1440082 -.0165094
m4	-.0846191	.0331771	-2.55	0.011	-.1499801 -.019258
m5	-.0232965	.0334686	-0.70	0.487	-.0892319 .0426389
m6	.1198732	.037011	3.24	0.001	.0469591 .1927873
m7	.1241075	.0616468	2.01	0.045	.0026591 .2455559
m8	.0635651	.0754108	0.84	0.400	-.0849993 .2121295
m9	.088161	.0548742	1.61	0.109	-.0199449 .1962669
m10	.0585193	.0343227	1.70	0.090	-.0090986 .1261372
m11	-.1457969	.0324917	-4.49	0.000	-.2098077 -.0817861
m12	-.0417205	.0292042	-1.43	0.154	-.0992546 .0158136
_cons	6.874499	.0398607	172.46	0.000	6.795971 6.953028
rho	.2525515				

Durbin-Watson statistic (original) 1.509298
 Durbin-Watson statistic (transformed) 2.016210

BHP System Peak Demand Model

Source	SS	df	MS	Number of obs	=	230
Model	3.03781492	17	.178694996	F(17, 212)	=	137.46
Residual	.275587388	212	.001299941	Prob > F	=	0.0000
				R-squared	=	0.9168
				Adj R-squared	=	0.9102
Total	3.31340231	229	.014469006	Root MSE	=	.03605

lmwnat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cdd60	.0086782	.0009586	9.05	0.000	.0067885 .0105678
hdd60	.0027718	.0004093	6.77	0.000	.001965 .0035785
mnthcdd60perday	.01034	.0020343	5.08	0.000	.0063298 .0143501
mnthhdd60perday	.0042651	.0008499	5.02	0.000	.0025897 .0059405
lntotPI	.3380472	.0384863	8.78	0.000	.2621824 .4139121
weekend	-.0233456	.0114565	-2.04	0.043	-.0459288 -.0007625
m2	-.0523044	.0116995	-4.47	0.000	-.0753668 -.0292421
m3	-.0416192	.0138153	-3.01	0.003	-.0688522 -.0143861
m4	-.080603	.0160687	-5.02	0.000	-.1122778 -.0489281
m5	-.0320146	.0203971	-1.57	0.118	-.0722217 .0081925
m6	.0932003	.0273189	3.41	0.001	.0393488 .1470519
m7	.0802717	.03224	2.49	0.014	.0167197 .1438237
m8	.0892537	.0299565	2.98	0.003	.0302028 .1483045
m9	.0623604	.025859	2.41	0.017	.0113867 .113334
m10	-.0339564	.0189731	-1.79	0.075	-.0713565 .0034436
m11	-.0216245	.0155354	-1.39	0.165	-.0522481 .008999
m12	-.027556	.011648	-2.37	0.019	-.0505167 -.0045953
_cons	2.867306	.2934284	9.77	0.000	2.288895 3.445717

CLFP Residential UPC Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	179
Model	2.06797158	15	.137864772	F(15, 163)	=	114.28
Residual	.196636554	163	.001206359	Prob > F	=	0.0000
				R-squared	=	0.9132
				Adj R-squared	=	0.9052
Total	2.26460814	178	.012722518	Root MSE	=	.03473

lupc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cdd60	.0006874	.0000816	8.42	0.000	.0005263 .0008485
hdd60	.0002053	.0000278	7.39	0.000	.0001504 .0002602
lnincome12	.4358274	.0999293	4.36	0.000	.2385045 .6331504
trend	-.0067252	.001048	-6.42	0.000	-.0087946 -.0046559
m2	-.1014795	.0126135	-8.05	0.000	-.1263864 -.0765726
m3	-.1002444	.014534	-6.90	0.000	-.1289436 -.0715453
m4	-.1743273	.017176	-10.15	0.000	-.2082435 -.1404111
m5	-.2145596	.0207049	-10.36	0.000	-.255444 -.1736752
m6	-.2040182	.0272919	-7.48	0.000	-.2579095 -.1501269
m7	-.1671883	.0341282	-4.90	0.000	-.2345787 -.0997979
m8	-.1762289	.0372566	-4.73	0.000	-.2497967 -.1026611
m9	-.2100186	.0318326	-6.60	0.000	-.272876 -.1471612
m10	-.2242816	.0248828	-9.01	0.000	-.2734158 -.1751474
m11	-.184847	.0182471	-10.13	0.000	-.2208783 -.1488158
m12	-.0530595	.013202	-4.02	0.000	-.0791285 -.0269905
_cons	-5.518172	1.151543	-4.79	0.000	-7.792037 -3.244308

rho | .0804959

Durbin-Watson statistic (original) 1.873708
 Durbin-Watson statistic (transformed) 2.014441

CLFP Residential Customer Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	179
Model	17.7840565	3	5.92801885	F(3, 175)	>	99999.00
Residual	.000921729	175	5.2670e-06	Prob > F	=	0.0000
				R-squared	=	0.9999
				Adj R-squared	=	0.9999
Total	17.7849783	178	.099915608	Root MSE	=	.00229

lcusts	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lhhd_pre10	-.0812966	.0565073	-1.44	0.152	-.19282 .0302268
lhhd_CISplus	.6734902	.0222876	30.22	0.000	.6295031 .7174773
CISplus	-2.733243	.239088	-11.43	0.000	-3.20511 -2.261376
_cons	10.761	.2030163	53.01	0.000	10.36033 11.16168

rho | .7863497

Durbin-Watson statistic (original) 0.433740
 Durbin-Watson statistic (transformed) 2.368500

CLFP Commercial Non-Demand UPC Model

Prais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	72
Model	.382507485	14	.027321963	F(14, 57)	=	26.09
Residual	.05968656	57	.001047133	Prob > F	=	0.0000
				R-squared	=	0.8650
				Adj R-squared	=	0.8319
Total	.442194045	71	.006228085	Root MSE	=	.03236

lupc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cdd60	.0004083	.0001487	2.75	0.008	.0001105	.0007061
hdd60	.0001288	.0000395	3.26	0.002	.0000496	.0002079
trend	-.0073876	.0027641	-2.67	0.010	-.0129226	-.0018527
m2	-.0487067	.0179134	-2.72	0.009	-.0845777	-.0128357
m3	-.0659655	.0209113	-3.15	0.003	-.1078397	-.0240913
m4	-.1196877	.0258481	-4.63	0.000	-.1714476	-.0679279
m5	-.1510517	.0305823	-4.94	0.000	-.2122917	-.0898118
m6	-.1560555	.0386278	-4.04	0.000	-.2334063	-.0787047
m7	-.1050686	.0509712	-2.06	0.044	-.2071366	-.0030006
m8	-.0874926	.0572705	-1.53	0.132	-.2021747	.0271895
m9	-.1392	.0479274	-2.90	0.005	-.235173	-.0432271
m10	-.1530321	.0363063	-4.22	0.000	-.2257342	-.0803299
m11	-.1492752	.0271702	-5.49	0.000	-.2036827	-.0948678
m12	-.0318684	.0196828	-1.62	0.111	-.0712825	.0075456
_cons	.0584828	.0505365	1.16	0.252	-.0427147	.1596803

rho | .184152
 Durbin-Watson statistic (original) 1.680528
 Durbin-Watson statistic (transformed) 1.996436

CLFP Commercial Non-Demand Customer Model

rais-Winsten AR(1) regression -- iterated estimates

Source	SS	df	MS	Number of obs	=	72
Model	.050031725	1	.050031725	F(1, 70)	=	232.39
Residual	.01507069	70	.000215296	Prob > F	=	0.0000
				R-squared	=	0.7685
				Adj R-squared	=	0.7652
Total	.065102415	71	.000916935	Root MSE	=	.01467

lcust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lntotemp12	1.224647	.0962394	12.73	0.000	1.032703	1.41659
_cons	3.202115	.4055417	7.90	0.000	2.393288	4.010943

rho | -.0108025
 Durbin-Watson statistic (original) 2.021164
 Durbin-Watson statistic (transformed) 1.993565

CLFP System Peak Demand Model

Source	SS	df	MS	Number of obs	=	382
Model	2.7075761	17	.159269182	F(17, 364)	=	259.88
Residual	.223080505	364	.000612859	Prob > F	=	0.0000
				R-squared	=	0.9239
				Adj R-squared	=	0.9203
Total	2.9306566	381	.007692012	Root MSE	=	.02476

lmWwoMS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cdh	.0019791	.0007083	2.79	0.005	.0005862	.0033719
cdd60	.0048258	.0009546	5.06	0.000	.0029485	.0067031
hdh	.0007178	.0003331	2.16	0.032	.0000628	.0013728
hdd60	.0006511	.0003106	2.10	0.037	.0000402	.001262
mnthcdd60perday	.0078491	.0016518	4.75	0.000	.0046009	.0110973
lntotemp	.6704658	.0296946	22.58	0.000	.6120712	.7288603
m2	-.0104943	.0068803	-1.53	0.128	-.0240245	.0030359
m3	-.0561029	.0074336	-7.55	0.000	-.070721	-.0414847
m4	-.1031688	.00717	-14.39	0.000	-.1172686	-.0890689
m5	-.1075987	.0081318	-13.23	0.000	-.1235899	-.0916074
m6	-.0583115	.0121855	-4.79	0.000	-.0822743	-.0343486
m7	-.0386051	.0172689	-2.24	0.026	-.0725645	-.0046456
m8	-.0577883	.0147434	-3.92	0.000	-.0867812	-.0287954
m9	-.066601	.0111045	-6.00	0.000	-.0884381	-.0447638
m10	-.0840245	.0082962	-10.13	0.000	-.1003389	-.06771
m11	-.0293447	.006979	-4.20	0.000	-.0430688	-.0156205
m12	.0229512	.0067193	3.42	0.001	.0097376	.0361648
_cons	2.265582	.1232518	18.38	0.000	2.023208	2.507957