2017 PSE Integrated Resource Plan

E

Demand Forecasting Models

This appendix describes the econometric models used in creating the demand forecasts for PSE's 2017 IRP analysis.

Contents

1. ELECTRIC BILLED SALES AND CUSTOMER COUNTS

PSE estimated use-per-customer (UPC) and customer count econometric equations using sample dates from a historical monthly data series that extends from January 1990 to December 2015; the sample dates varied depending on sector or class. Electric classes include residential, commercial, industrial, streetlights, resale and transport. The billed sales forecast is based on these estimated econometric equations, normal weather assumptions, rate forecasts, and forecasts of various economic and demographic inputs.

The UPC and customer count equations are defined as follows:

$$UPC_{c,t} = f(RR_{c,t(k)}, W_{c,t}, EcoDem_{c,t(k)}, MD_m)$$

$$CC_{c,t} = f(EcoDem_{c,t(k)}, MD_m)$$

$$MD_i = \begin{cases} 1, Month = i \\ 0, Month \neq i \end{cases} \quad i \in \{1,2,\ldots 12\}$$

$$t \in \{1,\ldots,nobs\}$$

$UPC_{c,t}$ = use (billed sales) per customer for class “c”, month “t”

$CC_{c,t}$ = customer counts for class “c”, month “t”

$—_{t(k)}$ = the subscript $t(k)$ denotes either a lag of “k” periods from “t” or a polynomial distributed lag form in “k” periods from month “t”

$RR_{c,t(k)}$ = effective real 12-month moving average of retail rates for class “c” in polynomial distributed lagged form

$W_{c,t}$ = class-appropriate weather variable; cycle-adjusted HDD/CDD using base temperatures of 65, 60, and 45 for HDD and 70 and 75 for CDD; cycle-adjusted HDDs/CDDs are created to fit consumption period implied by the class billing cycles

$EcoDem_{c,t(k)}$ = class-appropriate economic and demographic variables; variables include income, household size, population, and employment levels or growth in polynomial distributed lagged form

MD_i = monthly dummy variable that is 1 when the month is equal to "i", and zero otherwise for "i" from 1 to 12

UPC is forecast monthly at a class level using several explanatory variables including weather, retail rates, monthly effects, and various economic and demographic variables such as income, household size and employment levels. Some of the variables, such as retail rates and economic variables, are added to the equation in a lagged or polynomial lagged form to account for both short-term and long-term effects of changes in these variables on energy consumption. Finally, depending on the equation, an ARMA(p,q) structure is imposed to acknowledge that future values of the predicted variables could be a function of its lag value or the lags of forecast errors.

Similar to UPC, PSE forecasts the customer count equations on a class level using several explanatory variables such as household population, total employment, or manufacturing employment. Some of the variables are also implemented in a lagged or polynomial distributed lag form to allow the impact of the variable to vary with time. Many of the customer equations use monthly customer growth as the dependent variable, rather than totals, to more accurately measure the impact of economic and demographic variables on growth, and to allow the forecast to grow from the last recorded actual value. ARMA(p,q) could also be imposed on certain customer counts equations.

The billed sales forecast for each customer class before new conservation is the product of the class UPC forecast and the forecasted number of customers in that class, as defined below.

$$Billed\ Sales_{c,t} = UPC_{c,t} \times CC_{c,t}$$

The billed sales and customer forecasts are adjusted for known, short-term future discrete additions and subtractions not accounted for in the forecast equations, such as major changes in energy usage by large customers. These adjustments may also include fuel and schedule switching by large customers. The forecast of billed sales is further adjusted for new programmatic conservation by class using the optimal conservation bundle from the most recent IRP.

Total billed sales in a given month are calculated as the sum of the billed sales across all customer classes:

$$Total\ Billed\ Sales_{t} = \sum_{c} Billed\ Sales_{c,t}$$

PSE estimates total system delivered loads by distributing monthly billed sales into each billing cycle for the month, then allocating the billing cycle sales into the appropriate calendar months using degree days as weights, and adjusting delivered sales for company own use and losses from transmission and distribution. This approach also enables computation of the unbilled sales each month.

2. ELECTRIC PEAK HOUR LOAD FORECASTING

Peak load forecasts are developed using econometric equations that relate observed monthly peak loads to weather-sensitive delivered loads for both residential and non-residential sectors. They also account for deviations of actual peak hour temperature from normal peak temperature for the month, day of the week effects, and unique weather events such as a cold snap or an El Niño season.

Based on the forecasted delivered loads, we use regression equations to estimate a set of hourly peak loads each month for the system based on three specific design temperatures: "Normal," "Power Supply Operations" (PSO) and "Extreme."

The "Normal" peak is based on the average temperature at the monthly peak during a historical time period, currently 30 years. The winter peaks are set at the highest Normal peak, which is currently the December peak of 23 degrees Fahrenheit. We estimated the PSO peak design temperatures to have a 1-in-20 year probability of occurring. These temperatures were established by examining the minimum temperature of each winter month during heavy load hours. An extreme value distribution function relating the monthly minimum temperature and the return probability was established. The analysis revealed the following design temperatures: 15 degrees Fahrenheit for January and February, 17 degrees Fahrenheit for November, and 13 degrees Fahrenheit for December. Finally, the "Extreme" peak design temperatures are estimated at 13 degrees Fahrenheit for all winter months.

Weather dependent loads are accounted for by the major peak load forecast explanatory variable, the difference between actual peak hour temperature and the average monthly temperature multiplied by system loads. The equations allow the impact of peak design temperature on peak loads to vary by month. This permits the weather-dependent effects of system delivered loads on peak demand to vary by season. The sample period for this forecast utilized monthly data from January 2002 to December 2015.

In addition to the effect of temperature, peak load estimates account for the effects of several other variables, among them the portion of monthly system delivered loads that affects peak loads but is non-weather dependent; a dummy variable that accounts for large customer changes; and a day of the week variable. The functional form of the electric peak hour equation is

$$PkMW_t = \vec{\alpha}_{1,m} \cdot MD_i \cdot S_t + \vec{\alpha}_{2,m} \chi_1 \cdot \Delta T \cdot MD_i \cdot S_t + \beta_{1,d} DD_d + \delta_1 \cdot LT_t$$

where:

$$\chi_1 = \begin{cases} 1, & Month = 6,7,8 \\ 0, & Month \neq 6,7,8 \end{cases}$$

$$MD_i = \begin{cases} 1, Month = i \\ 0, Month \neq i \end{cases} \quad i \in \{1,2,....12\}$$

$PkMW_t$ = monthly system peak hour load in MW

S_t = system delivered loads in the month in aMW

MD_i = monthly dummy variable

ΔT = deviation of actual peak hour temperature from monthly normal temperature

DD_d = day of the week dummy

LT_d = late hour of peak dummy, if the peak occurs in the evening

χ_1 = dummy variables used to put special emphasis on summer months to reflect growing summer peaks.

To clarify the equation above, when forecasting we allow the coefficients for loads to vary by month to reflect the seasonal pattern of usage. However, in order to conserve space, we have employed vector notation. The Greek letters α_m, β_d, and δ_d denote coefficient vectors; there are also indicator variables that account for air conditioning load, to reflect the growing summer electricity usage caused by increased saturation of air conditioning.

The peak load forecast is further adjusted for the peak contribution of future conservation based on the optimal DSM bundle derived from the IRP.

3. GAS BILLED SALES AND CUSTOMER COUNTS

At the gas system level, PSE forecasts use-per-customer (UPC) and customer counts for each of the customer classes it serves. The gas classes include firm classes (residential, commercial, industrial, commercial large volume and industrial large volume), interruptible classes (commercial and industrial) and transport classes (commercial firm, commercial interruptible, industrial firm and industrial interruptible). Energy demand from firm, interruptible and transport classes is summed to form the 2017 IRP Gas Base Demand Forecast.

PSE estimated the following UPC and customer count econometric equations using sample dates from a historical monthly data series that extends from January 1990 to December 2015; the sample dates varied depending on sector or class. The gas billed sales forecast is based on the estimated equations, normal weather assumptions, rate forecasts, and forecasts of various economic and demographic inputs.

The UPC and customer count equations are defined as follows:

$$UPC_{c,t} = f(RR_{c,t(k)}, W_{c,t}, EcoDem_{c,t(k)}, MD_m)$$
$$CC_{c,t} = f(EcoDem_{c,t(k)}, MD_m)$$
$$MD_i = \begin{cases} 1, Month = i \\ 0, Month \neq i \end{cases} \quad i \in \{1,2,\ldots 12\}$$
$$t \in \{1,\ldots,nobs\}$$

$UPC_{c,t}$ = use (billed sales) per customer for class “c”, month “t”

$CC_{c,t}$ = customer counts for class “c”, month “t”

$\_\__{t(k)}$ = the subscript $t(k)$ denotes either a lag of “k” periods from “t” or a polynomial distributed lag form in “k” periods from month “t”

$RR_{c,t(k)}$ = effective real retail rates for class “c” in polynomial distributed lagged form

$W_{c,t}$ = class-appropriate weather variable; cycle-adjusted HDDs using the base temperature of 35 or 65; cycle-adjusted HDDs are created to fit consumption period implied by the class billing cycles

$EcoDem_{c,t(k)}$ = class-appropriate economic and demographic variables; variables include unemployment rate, household size, non-farm employment levels and growth, manufacturing employment levels and growth, and building permits. Economic and demographic variables may be used in lag form or in polynomial distributed lag form.

MD_i = monthly dummy variable that is 1 when the month is equal to "i", and zero otherwise for "i" from 1 to 12

UPC is forecast monthly at a class level using several explanatory variables including weather, retail rates, monthly effects, and various economic and demographic variables such as unemployment rate, non-farm employment and manufacturing employment. Some of the variables, such as retail rates and economic variables are added to the equation in a lagged or polynomial lagged form to account for both short-term and long-term effects of changes in these variables on energy consumption. Finally, depending on the equation, an ARMA(p,q) structure could be imposed to acknowledge that future values of the predicted variables could be a function of its lag value or the lags of forecast errors.

Similar to UPC, PSE forecasts the gas customer count equations on a class level using several explanatory variables such as household size, building permits, total employment and manufacturing employment. Some of the variables are also implemented in a lagged or polynomial distributed lag form to allow the impact of the variable to vary with time. Many of the customer equations use monthly customer growth as the dependent variable, rather than totals, to more accurately measure the impact of economic and demographic variables on growth, and to allow the forecast to grow from the last recorded actual value. ARMA(p,q) could also be imposed on certain customer counts equations. In addition, some of the smaller customer classes are not forecast using equations; instead, those current customer counts are held constant throughout the forecast period. This is done for the transport classes, industrial interruptible class and industrial large volume class. These classes have low customer counts and are not expected to change significantly over the forecast period.

The billed sales forecast for each customer class, before new conservation, is the product of the class UPC forecast and the forecasted number of customers in that class, as defined below.

$$Billed\ Sales_{c,t} = UPC_{c,t} \times CC_{c,t}$$

The gas billed sales and customer forecasts are adjusted for known, short-term future discrete additions and subtractions not accounted for in the forecast equations, such as major changes in energy usage by large customers. These adjustments may also include fuel and schedule switching by large customers. The forecast of billed sales is further adjusted for new programmatic conservation by class using the optimal conservation bundle from the most recent IRP.

Total billed sales in a given month are calculated as the sum of the billed sales across all customer classes:

$$Total\ Billed\ Sales_{t} = \sum_{c} Billed\ Sales_{c,t}$$

PSE estimates total gas system delivered loads by distributing monthly billed sales into each billing cycle for the month, then allocating the billing cycle sales into the appropriate calendar months using heating degree days as weights, and adjusting delivered sales for company own use and losses from distribution. This approach also enables computation of the unbilled sales each month.

4. GAS PEAK DAY LOAD FORECAST

Similar to the electric peaks, the gas peak day is assumed to be a function of weather-sensitive delivered sales, the deviation of actual peak day average temperature from monthly normal average temperature and other weather events. The following equation used monthly data from October 1993 to December 2014 to represent peak day firm requirements:

$$PkDThm_t = \vec{\alpha}_{1,m} Fr_t + \vec{\alpha}_{2,m} \Delta T_g \cdot Fr_t + \alpha_{3,m} EN + \alpha_{4,m} M_t + \alpha_{5,m} Sum + \alpha_{6,m} Csnp$$

$$Win = \begin{cases} 1, & Month = 1,2,11,12 \\ 0, & Month \neq 1,2,11,12 \end{cases}$$

$$Smr = \begin{cases} 1, & Month = 6,7,8,9 \\ 0, & Month \neq 6,7,8,9 \end{cases}$$

where:

$PkDThm_t$ = monthly system gas peak day load in dekatherms

Fr_t = monthly delivered loads by firm customers

ΔT_g = deviation of actual gas peak day average daily temperature from monthly normal temperature

EN = dummy for when El Niño is present during the winter

M_t = dummy variable for month of the year

$CSnp$ = indicator variable for when the peak occurred within a cold snap period lasting more than one day, multiplied by the minimum temperatures for the day

As before, the Greek letters are coefficient vectors as defined in the electric peak section above.

This formula uses forecasted billed sales as an explanatory variable, and the estimated model weighs this variable heavily in terms of significance. Therefore, the peak day equation will follow a similar trend as that of the billed sales forecast with minor deviations based on the impact of other explanatory variables. An advantage of this process is that it helps estimate the contribution of distinct customer classes to peak loads.

The design peak day used in the gas peak day forecast is a 52 heating degree day (13 degrees Fahrenheit average temperature for the day). This standard was adopted in 2005 after a detailed, cost-benefit analysis requested by the WUTC. The analysis considered both the value customers place on reliability of service and the incremental costs of the resources necessary to provide that reliability at various temperatures; it is presented in Appendix I of the 2005 LCP. We use projected delivered loads by class and this design temperature to estimate gas peak day load. PSE’s gas planning standard covers 98 percent of historical peak events, and it is unique to our customer base, our service territory and the chosen form of energy.

5. MODELING UNCERTAINTIES IN THE LOAD FORECAST

Load forecasts are inherently uncertain, and to acknowledge this uncertainty, high and low load forecast scenarios are developed. To create high and low forecasts, uncertainty in both weather and long-term economic and demographic growth in the service territory were included.

The econometric load forecast equations depend on certain types of economic and demographic variables; these may vary depending on whether the equation is for customer counts or use per customer, and whether the equation is for a residential or non-residential customer class. In PSE's load forecast models, the key service area economic and demographic inputs are population, employment, unemployment rate, personal income, and building permits. These variables are inputs into one or more load forecast equations.

To develop the stochastic simulations of loads, a stochastic simulation of PSE's economic and demographic electric and gas models is performed to produce the distribution of PSE's economic and demographic forecast variables. Since these variables are also a function of key U.S. macroeconomic variables such as population, employment, unemployment rate, personal income, personal consumption expenditure index and long-term mortgage rates, we utilize the stochastic simulation functions in EViews[1] by providing the standard errors for the quarterly growth of key U.S. macroeconomic inputs into PSE's economic and demographic models. These standard errors were based on historical actuals from 1980 to 2015. The stochastic simulation of PSE's economic and demographic models from 1,000 draws provides the basis for developing the distribution of the relevant economic and demographic inputs for the load forecast models over the forecast period. Based on these draws, standard errors were estimated for PSE service area population, employment, unemployment rate, personal income and building permits for each year over the forecast horizon. In a similar manner, these standard errors were used in producing the 250 stochastic simulations of PSE's load forecasts within EViews.

1 / EViews is a popular econometric, forecasting and simulation tool.

Additionally, we introduced weather variability into these 250 stochastic simulations using weather between 1929 and 2015 by creating 87 weather scenarios, each with 20 consecutive years of weather data. For weather strips starting after 1996 there are not 20 years of consecutive weather data available. Therefore, after 2015 in the data series, the data wraps around to weather from January 1, 1989. The last weather scenario year starts in 2015. Random weather strips were assigned to each of the 250 stochastic simulations created with the economic and demographic model uncertainties to create the range of uncertainty used for both the gas and electric model.

The high and low load forecasts are defined in the IRP as the 95th and 5th percentile, respectively, of the 250 stochastic simulations of the loads based on uncertainties in the economic and demographic inputs and the weather inputs.

6. HOURLY ELECTRIC DEMAND PROFILE

Because temporarily storing large amounts of electricity is costly, the minute-by-minute interaction between electricity production and consumption is very important. For this reason, and for purposes of analyzing the effectiveness of different electric generating resources, an hourly profile of PSE electric demand is required.

We use our hourly (8,760 hours) load profile of electric demand for the IRP for the stochastic analysis in the Resource Adequacy Model (RAM), for our power cost calculation and for other AURORA analyses. The estimated hourly distribution is built using statistical models relating actual observed temperatures, recent load data and the latest customer counts.

Data

Actual hourly delivered electric loads between January 1, 1994 and December 31, 2015 were used to develop the statistical relationship between temperatures and loads for estimating hourly electric demand based on a representative distribution of hourly temperatures. Based on this relationship, PSE developed a representative distribution of hourly temperatures based on data from January 1, 1950 to December 31, 2015

Methodology for Distribution of Hourly Temperatures

The above temperature data were sorted and ranked to provide two separate data sets:
For each year, a ranking of hourly temperatures by month, coldest to warmest, over 60 years was used to calculate average monthly temperature. A ranking of the times when these temperatures occurred, by month, coldest to warmest, was averaged to provide an expected time of occurrence. Next PSE found the hours most likely to have the coldest temperatures (based on observed averages of coldest-to-warmest hour times) and matched them with average coldest-to-warmest temperatures by month. Sorting this information into a traditional time series then provided a representative hourly profile of temperature.

Methodology for Hourly Distribution of Load

For the time period January 1, 1994 to December 31, 2015, PSE used the statistical hourly regression equation:

$$\hat{L}_h = \vec{\beta}_{1,d} \cdot DD_d + \alpha_1 L_{h-1} + \alpha_2 \left(\frac{L_{h-2} + L_{h-3} + L_{h-4}}{3} \right) + \left(\vec{\alpha}_{3,m} T_h + \vec{\alpha}_{4,m} T_h^2 \right) + \vec{\beta}_{2,d} Hol + \alpha_5 P^{(1)}(h)$$

for hours from one to 24 to calculate load shape from the representative hourly temperature profile. This means that a separate equation is estimated for each hour of the day.

$\hat{L}_h$ = Estimated hourly load at hour "h"

L_h = Load at hour "h"

L_{h-k} = Load "k" hours before hour "h"

T_h = Temperature at time "h"

T_h^2 = Squared hourly temperature at time "h"

$P^{(1)}(h)$ = 1st degree polynomial

Hol = NERC holiday dummy variables

All Greek letters again denote coefficient vectors.