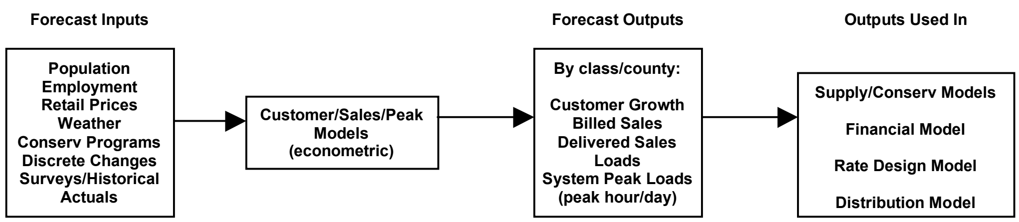
Load Forecasting Models

This appendix provides a more detailed technical description of the four econometric methodologies used to forecast (a) billed energy sales, (b) customer counts, (c) system peak loads for electricity and natural gas, and (d) hourly distribution of electric loads.

For the 2007 load forecast used in the 2009 IRP, we updated our key forecast driver assumptions and re-estimated the main equations. The diagram below shows the overall structure of the analysis.

Figure E-1
Econometric Model for Forecasts
of Energy Sales, Customer Counts, and Peak Loads



I. Electric and Gas Billed Sales and Customer Counts

PSE estimated the following use-per-customer (UPC) and customer count equations using varied sample dates from within a historical monthly data series from January 1989 to September 2007, depending on sector or class and fuel type. The billed sales forecast is based on the estimated equations, normal weather assumptions, rate forecasts, and forecast of various economic and demographic inputs. The variable "t" denotes a month within the sample, and is therefore unique. However, when we restrict a given month to be 1, 2,…,12 it is to be understood that we are talking about which monthly equivalence class it belongs to.

The UPC and customer count equations are defined as follows:

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

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

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

$UPC_{c,t}$ = 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”

$W_{c,t}$ = class-appropriate weather variable; cycle-adjusted HDD/CDD using base temperatures of 65, 60, 45, 35 for HDD and 65 and 75 for CDD; cycle-adjusted HDDs/CDDs are created to fit consumption period implied by the class billing cycles

$ED_{c,t(k)}$ = class-appropriate economic and demographic variables; variables include income, household size, population, employment levels or growth, and building permits

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 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, we use a lagged form of the dependent variable in many of the UPC equations. This lagged form could be as simple as a one month lag, or could be a more sophisticated time-series model, such as an ARIMA(p,q) model. This imposes a realistic covariant structure to the forecast equation.

Similar to UPC, PSE forecasts the customer count equations on a class level using several explanatory variables such as household population, total employment, manufacturing employment, or the retail rate. 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 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.

We generate customer forecasts by county by estimating an equation relating customer counts by class and county to population or employment levels in that county. Once the customer counts for each county are estimated, adjustments are made proportionally so that the total of all customer counts is scaled to the original service area forecast.

The billed sales forecast for each customer class 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 forecast is adjusted for 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. 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 each delivered sales for losses from transmission and distribution. This approach also enables computation of the unbilled sales each month.

II. Peak Load Forecasting

A. Electric Peak-hour Load Forecast

Based on the forecast delivered loads, we use hourly regressions to estimate a set of monthly peak loads for both residential and nonresidential sectors based on 3 specific design temperatures: “Normal”, “Power Supply Operations” (PSO) and “Extreme”. The “Normal” peak is based on the average temperature at the monthly peak during the historical time period, currently the past 27 years. The winter peaks are set at the highest Normal peak which is currently the December peak of 23° F. We estimated the PSO peak design temperatures to be a 1-in-20 year probability of exceedance. These temperatures were established by examining the minimum temperatures of each winter month. A function relating the monthly minimum temperature and the return probability was established. The analysis revealed the following design temperatures: 15°F for January and February, 17°F for November and 13°F for December. Finally, the "Extreme" peak design temperatures are estimated at 13°F 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 residential loads and commercial loads. The equations allow the impact of peak design temperature on peak loads to vary by month. This permits the weather-dependent effects of residential and nonresidential delivered loads on peak demand to vary by season. The sample period for this forecast utilized monthly data from January 1991 to December 2004.

In addition to the effect of temperature, the peak load is estimated by accounting for the effects of several other variables. A variable is used to account for the portion of monthly residential and nonresidential delivered loads which are non-weather dependent and affect the peak load. The peak forecast also depends on a number of other variables such as a dummy variable accounting for large customer changes, a day of the week variable, and a cold snap variable to account for when the peak day occurs following several cold days. The functional form of the electric peak-hour equation is

$$PkMW_t = \vec{\alpha}_{1,m} R_t + \vec{\alpha}_{2,m} NR_t + \vec{\alpha}_{3,m} \chi_1 \cdot \Delta T \cdot Ws + \vec{\alpha}_{4,m} \chi_2 \cdot \Delta T \cdot C + \alpha_{5,m} S48 + \vec{\beta}_d \cdot DD_d + \alpha_{6,m} CSnp$$

where:

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

$$\chi_2 = \begin{cases} 1, & Month = 7,8 \\ 0, & Month \neq 7,8 \end{cases}$$

$PkMW_t$ = monthly system peak-hour load in MW

R_t = residential delivered loads in the month in aMW

NR_t = commercial plus industrial delivered loads in the month in aMW

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

Ws = residential plus a % of commercial delivered loads

C **=** monthly delivered loads for the commercial class.

$S48$ = dummy variable for when customers in schedule 48 switched to transportation customers

DD_d = day of the week dummy

$CSnp$ = 1 if the minimum temperature the day before peak day is less than 32 degrees

χ_1, χ_2 = 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 and β_d are used to denote coefficient vectors; α_m denotes a monthly coefficient vector (12 coefficients) and β_d denotes a coefficient for the day of the week (7 coefficients). The difference between α_m and $\dot{\alpha}_m$ is that all values in α_m are constant, whereas $\dot{\alpha}_m$ can have unique values by month. That is to say, all "January" months will have the same coefficient. There are also two indicator variables that use a weather-sensitive combination of residential and some commercial loads for all months except for July and August, which use only commercial loads, to reflect the growing summer usage caused by increased saturation of air conditioning.

B. 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 June 2006 to represent peak day firm requirements:

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

where:

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

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

$PkDThm_t$ = 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 Nino is present during the winter
Win, Sum = winter or summer dummy variable to account for seasonal effects
$CSnp$ = 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 the ability to account for the effects of conservation on peak loads by using billed sales with conservation included as the forecast variable. It also 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°F average temperature for the day), based on the costs and benefits of meeting a higher or lower design day temperature. In the 2003 Least Cost Plan (LCP), PSE changed the gas supply peak-day planning standard from 55 heating degree days (HDD), which is equivalent to 10°F degrees or a coldest day on record standard, to 51 HDD, which is equivalent to 14°F degrees or a coldest day in 20 years standard. The Washington Utilities and Transportation Commission (WUTC) responded to the 2003 plan with an acceptance letter directing PSE to “analyze” the benefits and costs of this change and to “defend” the new planning standard in the 2005 LCP.

As discussed in our 2005 LCP, Appendix I, PSE completed a detailed, stochastic cost-benefit analysis that 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. This analysis determined that it would be appropriate to increase our planning standard from 51 HDD (14°F) to 52 HDD (13°F). PSE’s gas planning standard relies on the value our natural gas customers attribute to reliability and covers 98% of historical peak events. As such, it is unique to our customer base, our service territory, and the chosen form of energy. Thus, we use projected delivered loads by class and this design temperature to estimate gas peak-day load.

III. 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 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.

A. Data

We developed a representative distribution of hourly temperatures based on data from January 1, 1950 to December 31, 2003. Actual hourly delivered electric loads between January 1, 1994 and December 16, 2004 were used to develop the statistical relationship between temperatures and loads for estimating hourly electric demand based on a representative distribution of hourly temperatures.

B. 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 54 years was used to calculate average monthly temperature.
- A ranking of the times when these temperatures occurred by month, coldest to warmest; these rankings were averaged to provide an expected time of occurrence.

Next we 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 provides a representative hourly profile of temperature.

C. Methodology for Hourly Distribution of Load

For the time period January 1, 1994 to December 31, 2003, we 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 h from 1 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.