Itron

Knowledge to Shape Your Future

Electric | Gas | Water

information collection, analysis and application

Update on PJM Models Forecast Performance and Recommendations

Itron, Inc.

Forecasting and Load Research Solutions

11236 El Camino Real

San Diego, CA 92130-2650

November 4, 2011

Forecast Performance and Recommendations

This memo provides an update on Itron recommendations based on forecast performance analysis performed by PJM. First we provide some background for the original recommendations. This is followed by a summary of recent PJM analysis results. We conclude with updated recommendations.

Background on Recommendations 1, 3 and 4

In the Phase I report, Recommendation #1 was to implement the Index1 approach. This approach combines economic variables using a set of weights based on the industry survey. The following is an excerpt from the original September, 2010 Phase I report.

> *The two index approaches have about the same historical fit. Surprisingly, the simpler approach (Index1) actually does slightly better in all three cases than the sector -weighted approach. Index2 is more complicated to implement because of the need to develop sector weights for each zone. As a result, we recommend using Index1 for the current round of forecasts. Further development of Index2 may be appropriate in later phases, especially if it is decided to incorporate end use content into the models.*
>
> ***Recommendation:*** *Implement Index1 approach to combine economic variables using weights from the Survey analysis.*

Following this report, PJM performed additional studies using historical economic forecasts to generate peak forecasts based on data that would have been available at different points in time. This work allowed evaluation of forecast performance from one to three years ahead (looking at forecast years 2008, 2009, and 2010). This analysis was all performed with Moody's economic forecasts. These results were summarized by Itron in a report titled PJM Model Accuracy and Forecast Stability in July, 2011. The key conclusions from this report are as follows:

> *The index methods are consistently more accurate and give more stable forecasts than the GMP based method.*
>
> *The results do not indicate a strong advantage for one index approach over the other. Index1 appears to have a slight edge in terms of accuracy at the PJM level. Index2 appears to have a slight advantage in terms of stability.*

Recommendations 3 and 4 involved the choice of vendor to use as the source for economic data and forecasts. In the initial report, the utility survey clearly identified Moody's and Global Insight as the most frequently used vendors. About 44% of the respondents used Moody's and about 31% used Global Insight. Based on this finding and analysis of forecasts for the three study cases (BGE, ComEd, and PPL), the following recommendation was made. At the time of this recommendation, no data on forecast performance was available.

... we recommend the relatively simple approach of pre-processing forecasts from the two main economic vendors, using 50/50 weights initially. These combined inputs can then be run through the existing weather simulation process and post-processing steps.

Over time, PJM should track the forecast accuracy of successive forecasts from the two vendors. This will allow differential weighting over time (e.g., inversely proportional to forecast standard errors) based on vendor performance. The focus should be on long-run forecast accuracy, such as three years ahead. Tracking should be performed for the combined economic variable index at the regional level, giving a single measure of performance for each zone. These values can be combined across zones using NCP weights for an overall measure of performance.

Recommendation 3: *Use a weighted average of the baseline economic forecasts from Moody's Economy.com and Global Insight to generate the baseline peak forecasts. Initially, weight each vendor equally.*

Recommendation 4: *Track the vendor performance based on three-year ahead forecasts of the economic index used in the peak forecast models and adjust the relative vendor weighting to be inversely proportional to forecast standard errors.*

Updated Forecast Performance Analysis

Following the July 2011 report, PJM performed additional forecast performance analyses, with an expanded scope. For this work, historical forecasts of GDP, GMP, and other economic drivers were obtained from both Moody's and Global Insight. The data supported forecast tests covering 6 years (2006 through 2011). In each test, models were estimated with data up to the time of the forecast start, and forecasts were generated with the economic forecast that was available at that time.

Six combinations of model specifications were examined:

- GMP – Model with GMP as the only driver
- GDP – Model with GDP as the only driver
- Index1 – Model with Index1 drivers and weights
- Index2 – Model with Index2 drivers and weights by zone
- Index1GDP – Model with Index1 weights using GDP in place of GMP and Real Personal Income
- Index2GDP – Model with Index2 weights using GDP in place of GMP and Real Personal Income

Each model specification was run three times using different sources for the economic data and forecasts:

- Moody's – All economic history and forecasts from Moody's Economy.com
- Global Insight – All economic history and forecasts from Global Insight
- Average – Blended data and forecasts with 50/50 weights

Combining the three alternative data sources with the 6 alternative methods, a total of 18 forecast combinations were generated.

For each of the 18 combinations, 11 sets of forecasts were generated for each zone and for PJM totals. The first forecast was generated using data available in December of 2005. Forecasts as of June and December were generated for 2006 through 2010. Each forecast was generated for a wide variety of weather scenarios and the median value of the annual coincident peak forecasts was used as the PJM peak forecast and the median value of the annual zone peak forecasts was used as the non coincident peak forecast for each zone.

To generate aggregate ranking statistics the following steps were taken:

- For each forecast vintage, compare the forecast in each year for each zone to the weather normalized actual value for that year and zone.
- Compute the forecast error and percentage error for each forecast vintage, year and zone.
- For each forecast duration (1 year ahead, 2 years ahead, …) compute the average across vintages of the absolute error and absolute percent error for each zone.
- For each forecast duration, identify the best (lowest absolute percent error) method for each zone.
- For each forecast duration and zone, compute the relative accuracy score for each method as the ratio of the absolute percentage error of the method to the average absolute percentage error of the best method. Subtract 1 to give the best method a score of 0 and other methods a score that represents the relative percentage above the best method. (For example, if the best method has an error of 2% and Method A has an error of 3%, then the relative accuracy score for Method A is 50%.
- For each method and forecast duration, compute the weighted average of the relative accuracy scores across zones using zone CP values as weights.
- For each method, compute the average across years of the weighted average relative accuracy scores. Rank the results from best (lowest) to worst (highest).

Other simpler ranking approaches were also executed by PJM. Of the approaches provided, the approach described above was selected for two reasons. First, it weights the forecast accuracy results proportional to zone size, which is appropriate for forecasting PJM total peaks. Second, it counts large errors (relative to the best method) more heavily than small errors, which is not the case with simple ordinal ranking approaches.

In addition to the forecast accuracy statistics, a second set of statistics were generated for forecast stability. This was done by computing coefficient of variation statistics based on the variation across forecast vintages of forecasts for 2009, 2010, … 2015. For each zone, relative stability statistics were computed as follows:

- For each forecast year, method and zone, compute the standard deviation of forecasts across forecast vintages. Convert to a coefficient of variation (CV) by dividing the standard deviation by the average forecast value.
- For each forecast year and zone, identify the most stable method as the method with the lowest CV value.
- For each method and forecast year, compute the relative stability of each method as the ratio of the CV for that method to the CV for the most stable method. Subtract 1 to give the best method a score of 0 and other methods a score that represents the relative percentage above the most stable method. (For example, if the best method has a CV of .04 and Method A has a CV of .05, then the score for Method A is 25%.
- For each method and forecast year, compute the weighted average of the relative stability scores relative across zones using zone CP peak values as weights.
- For each method, compute the average across years of the weighted average relative stability scores. Rank the results from best (lowest) to worst (highest).

The results of the accuracy and stability analysis are summarized below.

Figure 1: Summary of Accuracy and Stability Rankings

		Accuracy		Stability		Combined Metric
Vendor	Method	MAPE % > Most Accurate Forecast	Accuracy Index	COV % > Most Stable Forecast	Stability Index	Forecast Score
Moody's	Index1	13%	1.00	14%	1.00	1.00
Moody's	Index2	18%	1.37	17%	1.20	1.29
Average	Index1	20%	1.50	25%	1.79	1.65
Moody's	Index1_GDP	25%	1.88	23%	1.66	1.77
Average	Index1_GDP	22%	1.61	29%	2.07	1.84
Global Insight	Index1_GDP	19%	1.41	36%	2.62	2.02
Average	Index2	26%	1.90	30%	2.17	2.04
Moody's	Index2_GDP	31%	2.63	29%	2.08	2.20
Average	Index2_GDP	28%	2.08	39%	2.79	2.44
Global Insight	Index1	29%	2.14	40%	2.90	2.52
Global Insight	Index2_GDP	25%	1.88	49%	3.56	2.72
Moody's	GMP	40%	3.00	40%	2.87	2.93
Global Insight	Index2	34%	2.52	47%	3.41	2.97
Average	GMP	52%	3.87	52%	3.72	3.79
Moody's	GDP	66%	4.87	62%	4.48	4.68
Average	GDP	59%	4.39	77%	5.54	4.96
Global Insight	GDP	53%	3.92	93%	6.70	5.31
Global Insight	GMP	71%	5.29	79%	5.72	5.50

In the table, the relative accuracy and relative stability scores are presented and are transformed into an index by dividing all results by the lowest value. The resulting indexed values are combined into a final score using 50/50 weights.

Using an Indexed Approach. As in the original analysis (July, 2010) of three zones and the follow up analysis of forecast accuracy and stability (June, 2011), it is clear that the index approaches provide better accuracy and stability than the use of GMP as a sole driver. The top ranked approaches are consistently index-based approaches. The bottom ranked approaches are consistently the GMP only and GDP only approaches. We continue to recommend an index approach.

Using GDP for GMP. Substituting GDP forecasts in the indexes for GMP and Real Personal Income has mixed results. For Moody's, the ranked results are consistently worse with GDP substituted in. For Global Insight, results are mixed. Based on these results, we recommend using a mix of GMP, GDP and other economic variables, as specified in the initial index recommendations.

Vendor Weights. Given that GMP is used as a forecast driver in the index approaches, the results consistently favor the Moody's economic forecasts over the Global Insights forecast. The forecasts with Moody's inputs are more accurate and more stable with both Index1 and Index2 approaches. As a result, we recommend for now that PJM continue to use Moody's economic forecasts to drive the zone models. We also recommend that PJM continue to monitor vendor performance using the same type of forecast performance analysis and implement appropriate weights as warranted by the results.

Choice of Index Approach. The aggregate ranking results place the use of Index1 and Index2 with Moody's economic drivers as the top two out of the 18 approaches. As in the earlier studies, Index1 provides slightly better accuracy. In this set of results, it also provides slightly better stability. However, we prefer Index2 on a conceptual basis, and PJM stakeholders have expressed support for Index2 because it provides economic factor weights that are zone specific. Also, now that Index2 has been implemented, the additional work involved is no longer a factor. We recommend that PJM implement Index2 and continue to work on ways to improve the end-use and load factor content of this index based on zone-specific information.