

To: PJM Load Analysis Subcommittee
From: Audrey Lyke – Exelon
Date: November 18, 2010
Subject: Comments on Itron PJM Load Forecasting Model Change Proposals

I appreciate the work and thought PJM and Itron put into these analyses. I have some questions and concerns which are outlined below. Up front, bear in mind this is only the third meeting of LAS this year, and only the second meeting since Itron completed its study. I would like to caveat all these remarks in that several materials were received by LAS yesterday for PJM to review with LAS this afternoon, therefore I have not fully absorbed all the materials nor has PJM had the opportunity to address my comments. There may be questions below that are already answered in the materials provided or will be answered today. Conversely, there may be additional questions that arise after today's discussion and further review.

- 1) Goodness of Fit has been measured based on in-sample MAPEs rather than on MAPEs for a hold-out sample. Model evaluation should be based on out-of-sample MAPES. Otherwise a) one can and will over-fit, and b) one has no basis for judging that one model is better than another. This is particularly true if the difference between MAPES across methods is small, as is the case for these models. I would like to see out-of-sample MAPES before making a judgment as to whether the proposed changes improve or weaken existing models.
- 2) At the previous LAS, I believe the reason for using in-sample MAPES was given as a consequence of the Price Shift variable leading to difficulties creating out-of-sample MAPES. I may be remembering that wrong; if it is a real concern I would be content with comparing the model out-of-sample MAPES without the additional modifier for Price Shift.
- 3) The components of Index1 are correlated and in some cases redundant. What is the covariance matrix of the x-variables, and should some be dropped due to collinearity.
- 4) The collinearity issue in index1 is probably worsened due to measurement interval. While load and weather are measured at daily or smaller intervals for the models, GMP, population, employment, GDP and number of households are typically measured quarterly or annually. What method is used to convert these values to daily? Whether it is spline, moving average or some other technique, in essence I suspect trend variables and their interactions with CDD and HDD are unintentionally being put into the models for every quarter and arbitrarily improving model fit. In the extreme it is just levels with step changes between quarters; in that case one is putting in variables that move together every quarter without exception. I welcome comments from other members of the committee as to whether this is a valid concern; is one over-counting the improvement to a daily model with measurements that are taken at quarterly or annual intervals.
- 5) In the MAPE comparison of Index1 components, the column having lowest MAPE is highlighted.
 - a. With several multicollinear variables and a concern about over-fitting, we would actually like to successively eliminate those components that contribute least to goodness of fit.

I believe this is often going to be NME but it would be better to see the comparisons systematically to potentially throw out the weakest components. PJM should evaluate the column having the highest MAPE to throw out the weakest component(s).

- b. A weighted combination of the factors (Index1) is non-comparable to the individual components. Basically one is constructing a more flexible vector for fitting to the (in-sample) data; mathematically the MAPEs have to be smaller than if a single component is used. The Index1 MAPE has no bearing on the out-of-sample performance of Index1 versus GMP or any other single component measure.
 - c. Could I see the MAPES for changing GMP to Index1 without changing from Moody's to the Average for the forecast vendors. Although several combinations were released for today, this is an additional useful comparison.
- 6) Some references have been made to greater comfort with Indices rather than GMP based not on the specific PJM analysis or results, but on Itron's overall experience. I also appreciate Itron's experience with load forecasting, but PJM should derive security only from analyses that are directly comparable to its forecasting problem definition. It is not clear that studies comparing Indices versus GMP based primarily on Retail Sales, can be carried over to wholesale load forecast needs without modification. Only analyses that are directly comparable to PJM's problem definition should be included in judging use of Indices rather than GMP.
 - 7) Use of the same weighting of the index components in every zone obscures zonal differences in the contributions of Industrial, Commercial and Residential loads. There should be study as to how each zone's weights should be constructed to see if varying the weights is preferable to the one-size-fits-all approach (related to recommendation 6). It is not appropriate to use the same weights across zones as diverse as those included in PJM's footprint.
 - 8) Both Moody's Economy.com and Global Insight provide economic forecasts at the appropriate regional MSA level for the needed number of years out. I have no overall preference of vendor for such forecasts between Moody's Economy.com and Global Insight. However there hasn't been any testing based on the Global Insight forward view; the older vintage forecasts were not available for backcasting. In a growth year (e.g. 2005 going into 2006-2007), would inputs from Global Insight performed as well as inputs from Moody's with respect to the load forecast? We don't know.
 - 9) It is a good idea to track Moody's and Global Insight performance of their three-year ahead forecast with respect to the load forecast outcomes to then adjust the relative vendor weighting of the composite economic driver. As someone else commented, it is preferable to backcast forecast performance in advance of making a switch. In combining the forecasts, PJM should be alert that the vendors are measuring different things for some variables that have the same or similar concept name.
 - 10) With respect to Recommendation 6, "Explore adjusting load forecast method further, in future, to better weight residential and commercial sector loads and their drivers", it should be done instead of using the same weights for every zone.
 - 11) Again with respect to Recommendation 6, it is important not to inadvertently drop migration loads from the analysis of load needing to be served. Reporting of sales by sector to DOE may

not include migration loads, and analysis based on these data would be incorrect for projecting load needing to be served.

- 12) If recommendations 1, 2 and 3 are implemented, **the peak load forecast for 2014 (143,903 MW excluding Duke and ATSI) will be 1,000 MW lower than the peak actually experienced eight years prior, in 2006. Is everyone comfortable with that?**
- 13) The price-shift variable overlooks whether for that year price was really the cause of the variation in peak load. It would be better to explore this dependency after some years of dynamic pricing data become available. PJM also needs to avoid double-counting price response on both the demand side and the supply side.
- 14) Across the RTO, the forecast changes from PJM's 2010 Load Forecast report are about 40% vintage, 60% model and vendor change. In some zones, such as the AECO forecast, almost all the change is attributable to the update in the Moody's forecast.
- 15) I did some back-of-the envelope calculations as to the implied elasticity of peak load with respect to the Avg_Ind1_PS economic variables. Across the RTO, it is about 0.44 which seems to lie within a reasonable range. However in a few zones the results look very odd, in particular Atlantic Electric, AEP, Dayton and ATSI where the back-of-the-envelope results are elasticity greater than 1. In the case of AEP the implied elasticity is greater than 2 and is suspect.
- 16) PJM should be consistent on allocating RTO peaks by zone. The current calculations obscure real differences between zonal contribution to Coincident Peak and should be corrected.
- 17) As mentioned up front, additional time is needed to absorb the study outcomes to date and to evaluate the proposed changes. The point is to improve the models, and not to replace working models just for the sake of change. It is clearly a compressed timeline for a new process, and I would ask that stakeholders and the modelers be given additional time to evaluate the changes.