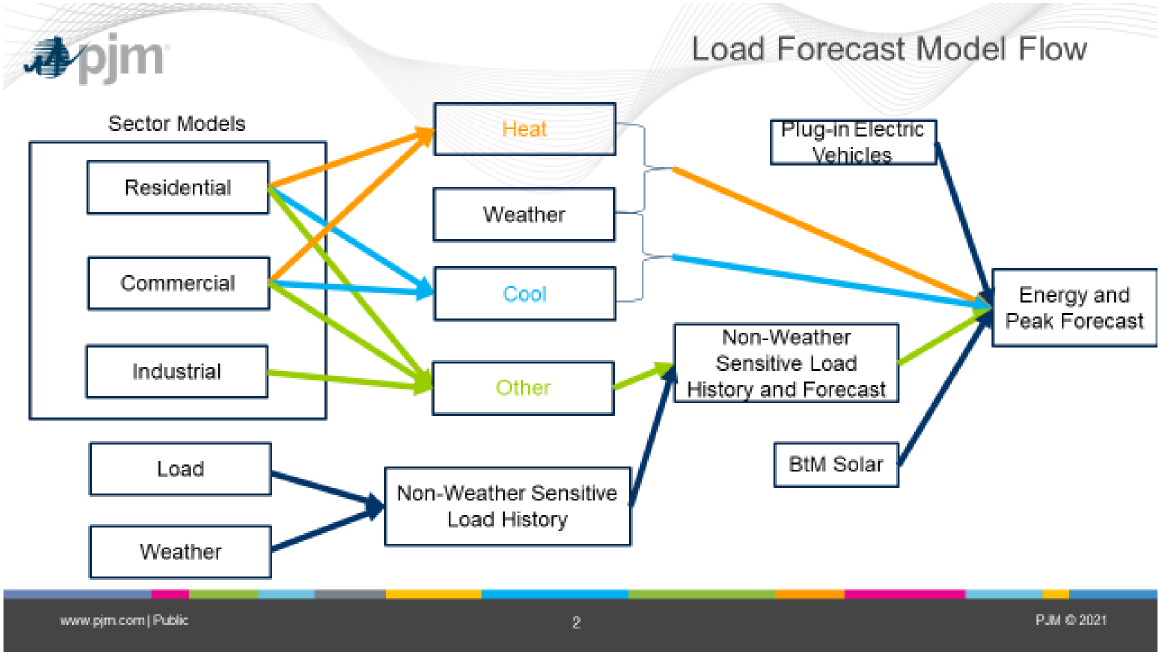
Sector Model Estimation Period Sensitivity

As indicated at the October 4th Load Analysis Subcommittee (LAS) meeting, at the request of a group of stakeholders, PJM agreed to conduct analysis using a 10 year estimation period for the sector models and non-weather sensitive models (see figure below for info on forecast model flow). The "Energy and Peak Forecast" is already modeled on an approximately 10 year estimation period. This document is meant as an accompaniment to the presentation and spreadsheets and should be used as a reference to help understand the outcome of the analysis.

Figure (slide from September 3, 2021 LAS Presentation: https://www.pjm.com/-/media/committees-groups/subcommittees/las/2021/20210903/20210903-item-04-forecast-methodology-updates.ashx)



Sector Models

Sector models take a look at historical energy data[1], and using economic and end-use information (includes trends of use intensity) try to forecast the path going forward. This information is also used to help set our understanding for how load is segmented into weather sensitive and non-weather sensitive uses.

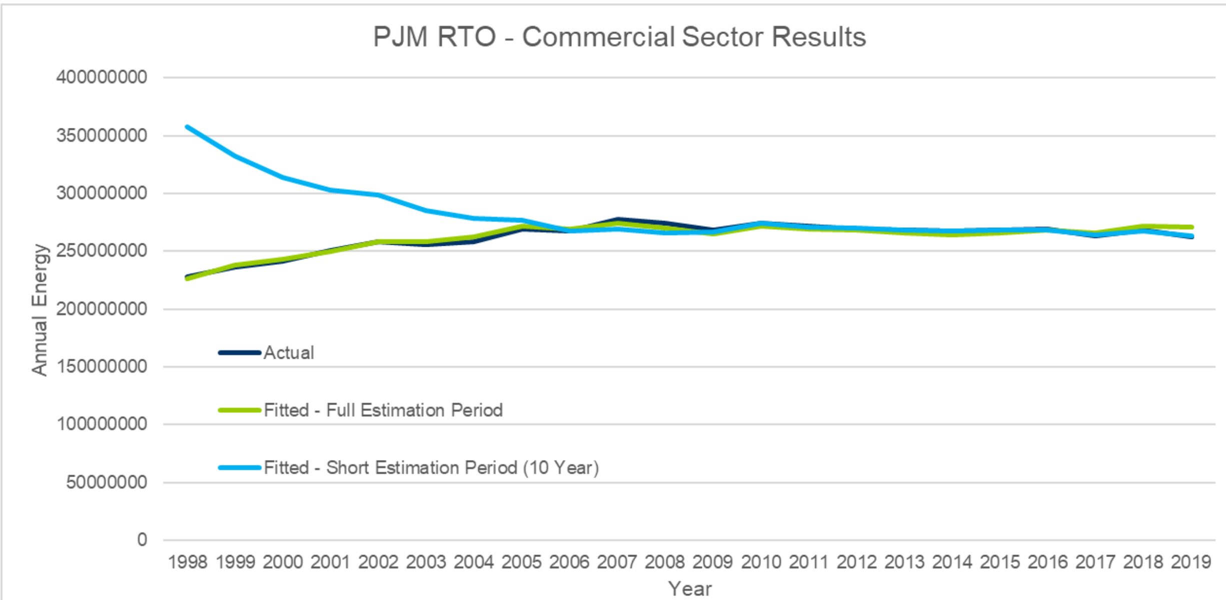
It has been PJM's contention that the paths of economics and end-use have largely explained how energy trends have evolved over time from the growth seen in 1998-2010 to the generally flat to declining trends witnessed in much of the post 2010 time period. This is the predominant reason why we have argued for defaulting to using data back to 1998. If the independent variables explain the observed trend, then it is not necessary (or advisable) to omit earlier

[1] In this case the historical energy data is annual data compiled from EIA 861 for the Residential, Commercial, and Industrial sectors.

data. Additionally, we have been concerned with the ability to produce quality results when placing strict restrictions on the number of observations.

When we run a model with a shortened estimation period (10 years of 2010-2019) and compute the backcasts for the pre-2010 time period we can conclude that this model does not fit the data well. This is best illustrated in the Commercial model results. In the figure below, it can be seen that the light blue line lies above the dark blue line (actuals) in earlier years. In contrast, the model with a full estimation period (the green line in the figure) does fit the data well, illustrating that the independent variables appropriately explain how load in more recent years differs with earlier periods. Forecasting is difficult in either case, but additional caution should be used in interpreting forecast results from a model that poorly fits the historic data.

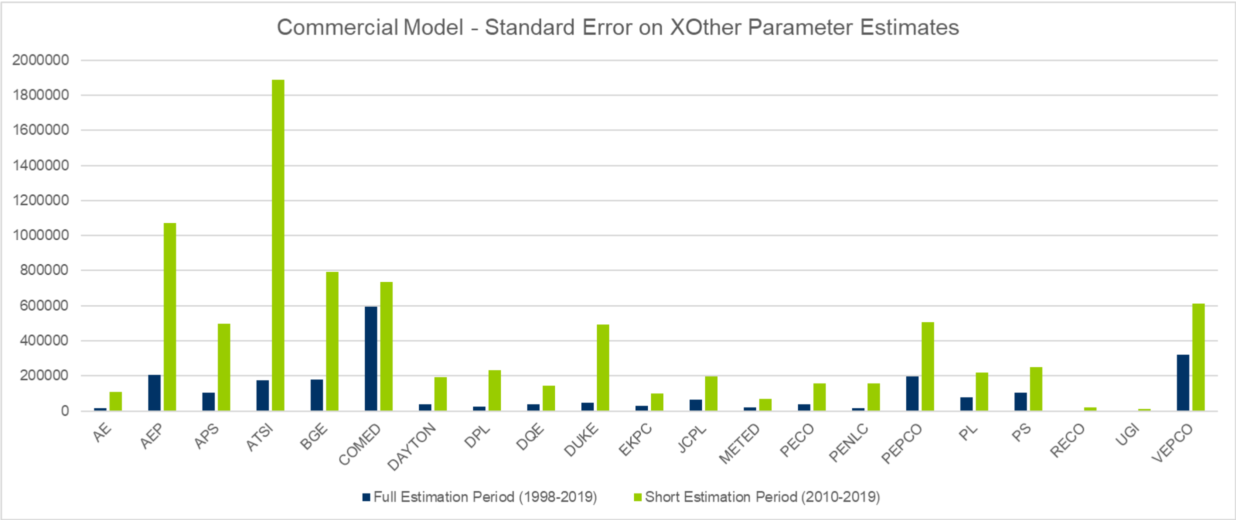
Figure (Comparing RTO level Results of Commercial Sector Model)



For the 10 year time period in the sector models, 2010 to 2019 is relatively stable both for the dependent variable (energy) and the independent parameters (economics, end-use intensity) in the multivariate regression. The goal of multivariate regression is to understand how the dependent and independent variables relate to one another. By both limiting the number of observations and limiting those observations to a period of stability, the model puts a lot of weight on the model intercept and makes uncertain or improper conclusions on the other coefficients. This is seen in the large standard errors on the coefficient estimates (an example of which is seen in the figure below[2] in which the green bars are significantly larger than the blue bars). This figure illustrates the large degree of uncertainty in estimating the relationships to the dependent variables. This has a compounding effect when looking into the forecast horizon, exhibited by forecast standard errors that increase as we move into the forecast period. When using the full estimation period, the standard errors on the coefficient estimates are smaller and consequently the forecast standard errors are more stable.

[2] Additional data on parameter estimates and the corresponding standard errors are available in the spreadsheet that accompanies the meeting materials.

Figure (Comparing Parameter Standard Errors for XOther term in Commercial Model)



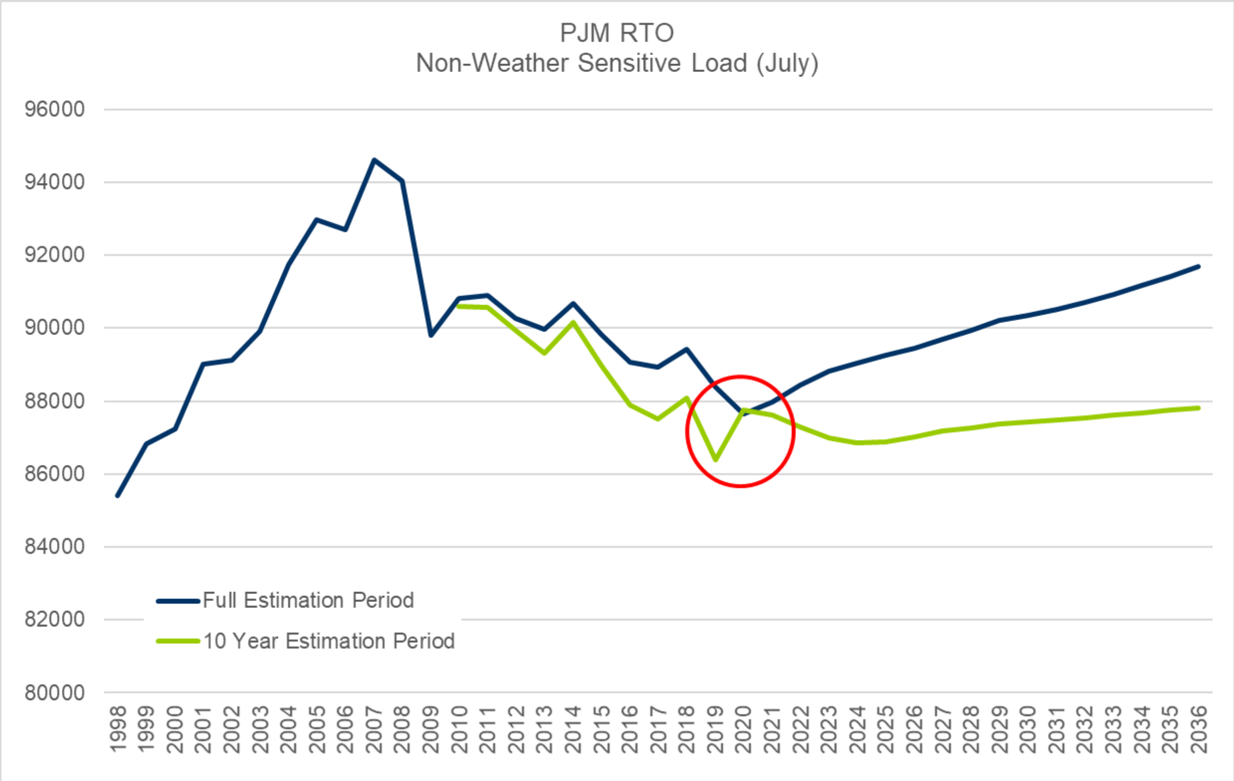
Shortening the time period in the sector models may be advisable in isolated cases. For instance, it was our recommendation in the case of the Industrial models at the September 3rd Load Analysis Subcommittee meeting that a 15 year estimation period should be considered in some cases should the coefficients have the improper sign. However, in the case of shortening the estimation period to 10 years the model is unable to establish reasonable relationships with confidence and thus any results gotten from such an approach should be considered imprecise.

Non-Weather Sensitive Model

Non-weather sensitive load is a modeled parameter that is an input into the final model. It relies on the output of the sector models, and thus any flaws present in the sector models will also be reflected in the fitted non-weather sensitive series. In the case of the sector results from the 10 year estimation period, the Other series[3] and consequently the non-weather sensitive series shows growth in 2020, a result we know to be inconsistent with what actually occurred. In the circled area in the figure below, it can be seen that the model backcast shows growth in 2020 when using the 10 year estimation period. Model results are only as good as the inputs used to drive them and in the case of the 10 year estimation period the inputs (the Sector model results) are not sound. These concerns are not present when using the full estimation period as we have more confidence in the output of the Sector models.

[3] The sum of non-weather sensitive energy from the Residential, Commercial, and Industrial models.

Figure (Comparison of Non-Weather Sensitive Results using Full Estimation vs 10 Year Estimation Period)



Final Model

The Sector models and the Non-Weather sensitive models influence the fit and results of the final model used to produce the Load Forecast. When the historical inputs are modified then the parameter estimates change as well, and so one cannot just look at the series in isolation. In this case the end result of inputs, in combination with the fit to those inputs, is a load forecast that is slightly higher using the 10 year estimation period than with the full estimation period for the sector models. In the figure below, the forecast using the 10 year estimation period (the gray line) is approximately 1.5% higher at the beginning of the forecast than with the full estimation period (the orange line) and the forecasts converge at the end of the 15 year horizon. This generally higher forecast is mostly a factor of the aforementioned issue with the non-weather sensitive series in the 10 year sensitivity resulting in the non-weather sensitive series getting assigned less weight in the production of the forecast and more weight being assigned to the trend variables.

Figure (Comparison of Summer Peak Forecast Results using Full Estimation vs 10 Year Estimation Period)