ISO new england

memo

To: NEPOOL Markets Committee

From: Market Development

Date: *Revised* February 1, 2023 (Original version: January 5, 2023)[1]

Subject: Day-Ahead Ancillary Services Initiative – Statistical Model for Real-Time Energy Prices

As part of its Day-Ahead Ancillary Services Initiative (DASI), the ISO proposes to create a suite of ancillary service products that will be settled as energy call options. As discussed previously with the Markets Committee, to provide efficient incentives for competitive suppliers of these day-ahead ancillary services, the call-option strike price will be based, in part, on the expected value of the real-time energy price.[2] This technical memorandum discusses the ISO's proposed statistical model and algorithm for evaluating, prior to each day's Day-Ahead Market, the expected value of the hourly real-time electricity price.

In addition, this statistical model has a second, important purpose. It provides estimates of the expected value of the call option close-out costs for each hour of the next (operating) day. We expect this information may be useful to market participants, as expected close-out costs are an important component of suppliers' competitive day-ahead ancillary service offer prices.[3] In addition, the ISO proposes to use this statistical model to calculate the expected close-out cost component of suppliers' reference prices, an element of the market power mitigation procedures for the co-optimized day-ahead market.[4]

The ISO acknowledges that accurate statistical modeling of real-time electricity prices is mathematically involved; as such, this memorandum explains the key concepts and modeling in a readily accessible way. The use of sophisticated statistical models in practice has benefits, however: The proposed statistical model and algorithm discussed below performs favorably relative to alternative, commercially-available real-time price forecasts, and better than other statistical techniques in the academic and applied literature we have

[1] This revised version provides additional statistics on the model's performance in the top 1% and top 5% of hours; see revised Tables 4a and 6 in Sections 5.2 and 5.4 and discussion therein. It also includes additional information on the model input data sources in Appendix A3.

[2] *See* the "Strike Price Overview" of ISO's November 2022 Markets Committee presentation (slides 51-54) at https://www.iso-ne.com/static-assets/documents/2022/11/a08_mc_2022_11_08-10_dasi_presentation.pptx

[3] *See,* generally, the ISO's "Competitive Offer Price Formulations for Day-Ahead Ancillary Services as Options" memo, at https://www.iso-ne.com/static-assets/documents/2022/11/a07a_mc_2022_12_06_08_dasi_memo.pdf

[4] *See* the "DA A/S offer foundations, components and examples" module of the ISO's December 2022 Market's Committee presentation (slides 17-22), at https://www.iso-ne.com/static-assets/documents/2022/12/a07a_mc_2022_12_06_08_dasi_presentation_rev1.pptx

ISO New England Inc.
One Sullivan Road
Holyoke, MA 01040-2841

iso-ne.com
isonewswire.com
@isonewengland
iso-ne.com/isotogo
iso-ne.com/isoexpress

examined. In practice, the models we propose to employ also fare well with regard to estimating expected close-out costs, achieving an expected close-out within \$0.05/MWh of the average calculated realized closeout (on a forward-looking basis). Based on these findings, and as further detailed below, we expect the model will be satisfactory for its intended purposes in administering the ancillary services of the co-optimized Day-Ahead Market.

During prior stakeholder discussions of the ISO's Energy Security Improvements (ESI) project in 2019-2020, the ISO indicated that it would publicly provide the algorithm to be used for calculating option strike prices, including (by implication) the statistical model for estimating expected hourly Hub real-time energy prices. Accordingly, in addition to explaining the statistical model and reviewing its performance, a mathematical Appendix provides the algorithm's formulas. The ISO plans to continue to refine the statistical model described herein during the implementation phase of DASI (i.e., through 2024); if enhancements are identified that result in performance improvements to the algorithm or its data inputs, the ISO will provide participants with updated information detailing the statistical model to be used in production, as well the results of any ongoing periodic updates (i.e., standard "re-training" of models) following implementation.

In recent stakeholder discussions, the ISO noted that it is continuing to evaluate the potential for a strike price "adder" that, when combined with ("added to") the expected value of the real-time price, would determine the hourly energy call option strike price. Because evaluation of a strike price adder is continuing, the memorandum does not presently address a strike price "adder" component. Importantly, this memorandum should not be read to implicate an ISO position or proposal on such a strike price "adder"; rather, the purpose of this technical memorandum is simply to detail the proposed method to calculate the expected real-time energy price component of the strike price, as well as the (closely related) methods to calculate suppliers' hourly expected close-out costs.

We hope this memorandum proves informative, and we welcome questions and stakeholder input on the methodologies discussed below.

A note on terminology is useful before proceeding. Throughout this memorandum, when we refer to the "real-time energy price", we mean the real-time Locational Marginal Price (RT LMP) at the New England Internal Hub pricing point. This is because the energy call options for ancillary services are system-level products, and are settled at the common Hub pricing point.[5]

1. Practicalities and Objectives

In this section, we provide a brief overview of several key considerations in the development of an algorithm for timely (i.e., daily) calculation of the call option strike price and expected close-out costs. These considerations relate to the market's timing and input data, and the types of outputs (possible real-time price values *and* their chances) needed by participants and the ISO.

▶ **Timing.** Consider the timing of the DA A/S market, depicted in the following figure:

[5] *See* Section 4.3 of the ESI White Paper, at https://www.iso-ne.com/static-assets/documents/2020/04/esi-white-paper-final-with-cover-page-04152020.pdf

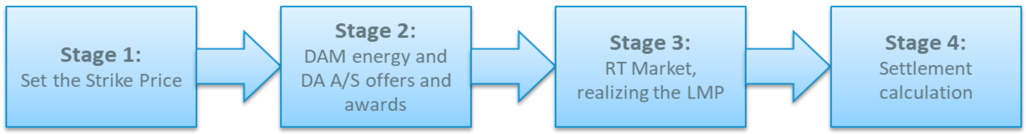

Figure 1. Timing in the co-optimized Day-Ahead Market

In advance of the day-ahead market bid and offer submission deadline (10:30 AM ET), the ISO will post the strike price for each hour of the next (operating) day. Consistent with existing day-ahead practices, the ISO will also make resource-specific reference prices (inclusive of expected close-out cost estimates) available to resource owners/operators, on a confidential basis, prior to the day-ahead market bid and offer submission deadline. These are the ISO actions depicted as Stage 1.

Stages 2 through 4 are similar to today. After participants submit their bids and offers, the Day-Ahead market is cleared to determine participants' awards; after the real-time market, participants' final settlements are calculated based on the totality of their day-ahead awards and prices and the real-time prices and awards.

The central timing point to note is that the strike price must be calculated *prior* to the bid/offer submission deadline, so that participants can incorporate its numerical value into their ancillary-service offer prices.

▶ **Input data.** To accurately predict both real-time energy prices *and* to model the uncertainty associated with that prediction, it is desirable to incorporate statistically useful predictors of the real-time price. Critically, such predictors must be available to the ISO at or before the time the strike price and expected close-out costs are computed, and reference prices disseminated to participants – that is, before bid/offer submission in the Day-Ahead Market.

In practice, both supply-side and demand-side information are available to the ISO in this timeframe that are useful for modeling hourly real-time prices the next (operating) day. These include, generally, fuel price information; forecast weather data; recent (past) real-time electricity prices for the same hour-of-day; and the ISO's existing forecast real-time system load. We provide additional detail on these modeling inputs in Section 3, below.

Importantly, the statistical model does not (and cannot) include the Day-Ahead Market clearing prices, or any other results of the clearing process, for the next (operating) day. This is because the Day-Ahead Market results are not available to the ISO, or to market participants, when Stage 1 is performed in Figure 1.

▶ **Output data: Confidence Intervals.** Importantly, the statistical models we propose are not simply single-value forecasts of the real-time energy price. Rather, they specify estimated *confidence intervals* that quantify the range of values where the real-time energy price will (is modeled to) occur with a given probability. In simple terms, the statistical models provide percentiles: the range of price values for the assigned hour in which the real-time price is estimated to fall with (e.g.) 90 percent probability, and the range with 80 percent probability, and so on. All such intervals are available from the trained model, for each real-time hour, and these intervals will typically differ in every hour.

There are three key reasons why the statistical models we propose are designed to estimate confidence intervals – or, stated in more technically-precise terms, we use statistical models of the *conditional probability distribution* of each hourly real-time price. First, the expected values that matter for participant offer formulation – that is, expected real-time energy prices and expected close-out costs – can be (readily) derived from these interval estimates, and timely provided to participants by the ISO, as Stage 1 in Figure 1 requires.

Second, interval estimates provide crucial information about the *uncertainty* associated with the real-time price for any hour in the next day. Intuitively, on some days the system is 'quiet', weather patterns are normal and little change in the weather is anticipated, and there is comparatively little uncertainty over the range of values where the day's real-time prices will fall. On such days, a 90 percent confidence interval for the real-time price each hour tends to be relatively narrow, and the statistical models we propose capture this property well. In contrast, on other days there can be great uncertainty over the real-time price in a particular hour of the next day, owing to a combination of high loads, high fuel prices and potential fuel-switching, and/or complex next-day forecast weather patterns (e.g., changeable weather or fronts over the course of the day). In such cases, the models we propose sensibly estimate correspondingly larger 90 percent confidence intervals for real-time prices the next day. In effect, the statistical models we propose not only convey that on such days the expected real-time energy price is high, the models can also convey whether or not the *uncertainty* associated with that forecasted price is large for the next day. We illustrate this property graphically with results for illustrative days in Section 2 presently.

Third, the uncertainty in the possible real-time price outcomes for a given hour of the next operating day has a direct impact on the estimated close-out costs that suppliers should incorporate into competitive ancillary service offer prices. Put in simpler terms, even if two different days have *identical* expected real-time energy prices for a given hour, greater *variation* in the possible outcomes on one day than the other will produce *different* expected close out costs. Intuitively, if there is a larger chance of an unusually high real-time price on one day than another, all else equal, then there will be a larger expected close-out cost for competitive suppliers to account for on the day with greater uncertainty.

▶ **A Simple Example.** This role of variation, and the need for statistical models to account for the full distribution (i.e., chances of realizing different values) of real-time prices, is usefully illustrated with a simple example. Consider two different hypothetical, simplified probability distributions over three potential RT LMP outcomes, as summarized in the following Table:

RT LMP ($/MWh)	Distribution A	Distribution B
$30	33.3%	50%
$60	33.3%	0%
$90	33.3%	50%

Table 1. Two Distributions Over RT LMP Outcomes

Under probability distribution A, all three possible real-time price outcomes are equally likely. Under distribution B, there is a greater likelihood of one of the two more extreme price values (e.g., $30 or $90).

The key point of this simple example is that distributions A and B both have the same expected real-time energy price (and therefore would lead to the same strike price), but distribution A and distribution B yield *different* expected close-out costs. Specifically, the expected RT LMP is $60/MWh in both cases.[6] However, each distribution leads to a different expected closeout charge. For simplicity, assume the strike price is $60/MWh; the expected value of the close-out cost (in mathematical terms, E[max($0, RT LMP – K)], evaluates for distribution A as:

$$Closeout_A = 33.3\% \times \max(0, 30 - 60) + 33.33\% \times \max(0{,}60 - 60) + 33.3\% \times \max(0{,}90 - 60) = \$10,$$

But the same calculation for distribution B yields a higher expected close out cost:

$$Closeout_B = 50\% \times \max(0, 30 - 60) + 0\% \times \max(0{,}60 - 60) + 50\% \times \max(0{,}90 - 60) = \$15.$$

While the two price probability distributions have the same expected RT LMP, Distribution B has more uncertainty (or variance) with regard to the RT LMP. That is, it has more 'weight' away from the expected real-time energy price than Distribution A, which results in a larger expected closeout for Distribution B. Competitive suppliers should account for the greater uncertainty present when a higher variance distribution, like B, is applicable to the next day's real-time price rather than a lower variance distribution, like A. The ISO's proposed statistical model is developed expressly to capture and model this type of uncertainty in the probability distribution of real-time prices, and in its impact on expected closeout costs, in a more sophisticated manner that in this simple numerical example – but the key idea, and its importance, is that illustrated with this example.

2. Conceptual Foundations: Mixtures of Days and Probability Distributions

The previous section highlighted that, in the New England system, there are some days and hours when there is considerable uncertainty over real-time energy price the next day; and there are other days and hours when there is less uncertainty. Moreover, modeling this variation is important to accurately estimate the (day-to-day and hourly changes in) expected close-out costs in an energy call option structure. Because quantitatively modeling this type of variation involves mathematically sophisticated probability models, we first explain the proposed model's key logic using simple graphs and price data.

Consider the histogram of real-time energy prices shown in Figure 2, below. These are *simulated* (hypothetical) data for a period of several days, and replicate key properties regarding uncertainty in actual real-time energy prices in New England, as explained next.[7] On the horizontal axis, the figure shows hourly RT LMPs, and the vertical axis shows their absolute frequency (number of hours). In these data, the most commonly observed price range is from $23/MWh to $42/MWh, which occurs in about half of all hours in these data. However, there are extremes in each direction, ranging from –$166/MWh to +$374/MWh. The mean RT LMP is equal to $43 MWh, with a decreasing frequency of RT LMPs that deviate significantly from this mean value, though there is a greater prevalence (frequency) of positive deviations (e.g. values up to about $250/MWh or so) than negative deviations. These data show visually what economists and

[6] In each case, this is calculated as the weighted sum across the distribution. For example, with Distribution A this is equal to 33.3% × $30 + 33.3% × $60 + 33.3% × $90 = $60/MWh.

[7] We use simulated data in this section to clearly present the main points of this section graphically. The concepts and methods described in this section apply quite generally to actual electricity price data.

statisticians call a "fat right-tail", meaning that the frequency of values does not decrease materially for a considerable range above the mean (e.g. from about \$68/MWh to about \$248/MWh or so).

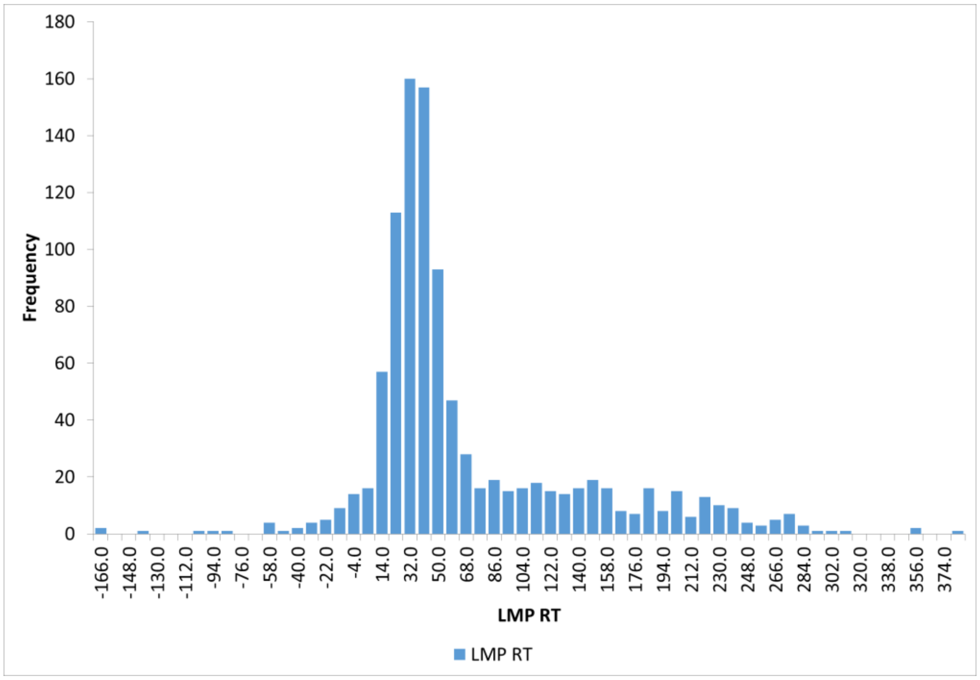

Figure 2. Histogram of the simulated RT LMP data

A standard approach to statistically modeling price data is to approximate their distribution using a familiar probability distribution, such as the bell-shaped *normal distribution*.[8] The following figure shows, superimposed with the same price histogram, the best possible normal distribution approximation, depicted as a solid line. The interpretation of the approximating normal distribution model is, for a given RT LMP value, the model's predicted frequency of that price (visually, the height of the solid line in Figure 3).

[8] The normal, or Gaussian, distribution is a statistical model for symmetric data centered on a mean, with rapidly decreasing tails. It is one of the most common statistical models, because it fits well (sometimes after simple transformations) with a wide variety of natural-world and financial data. (But not, as explained here, electricity price data).

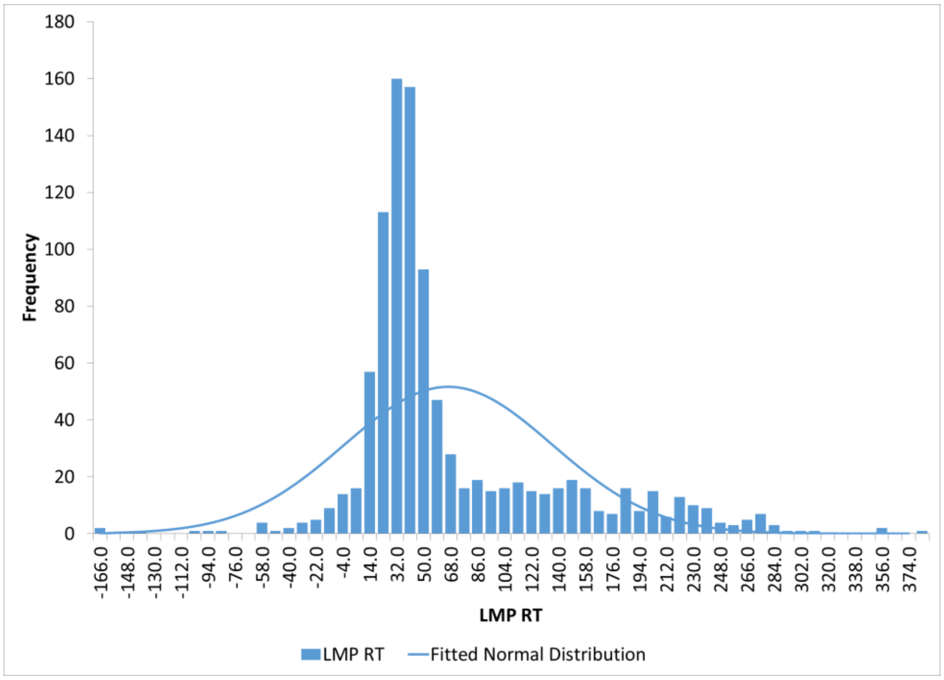

Figure 3. Normal Distribution Fit to the Simulated Data

In practical terms, a single normal distribution is a terrible approximation to the relative frequency of real-time price data. The approximation is centered around $75/MWh—well above these data's mean of $43/MWh. And it significantly underestimates the frequency of the most commonly occurring RT LMPs, between $14/MWh and $55/MWh, while significantly overestimates the frequency of RT LMPs between $68/MWh and $176/MWh. It also underestimates, but to a lesser degree, the likelihood of unusually high RT LMPs, i.e., those over about $200/MWh. As explained in the simple numerical example in the previous section, such over- or under-estimation of the likelihood of different real-time prices could yield poor estimates of expected real-time LMPs and the associated expected value of option close-out costs.

▶ **Mixtures of Distributions.** The reason the single normal distribution is a poor approximation to the data is that doing so ignores that electricity price data, when examined over a period of multiple days or more, inevitably comprise a mixture of different day types – and, accordingly, better approximated by a mixture of probability distributions. This idea can be made technically price in terms of what statisticians call, appropriately enough, *mixture distributions*. It sounds complicated, and some of the details are, but the idea is simple. It is to 'unpack', using appropriate statistical methods, the underlying price data and their frequencies into two (or more) underlying probability distributions.[9]

[9] Mixture distributions, or statistical *mixture models,* are a well-developed class of data modeling techniques. A readily-accessible, if mildly technical, summary is available on Wikipedia (as of this writing) at https://en.wikipedia.org/wiki/Mixture_model. A more complete technical discussion can be found in the references

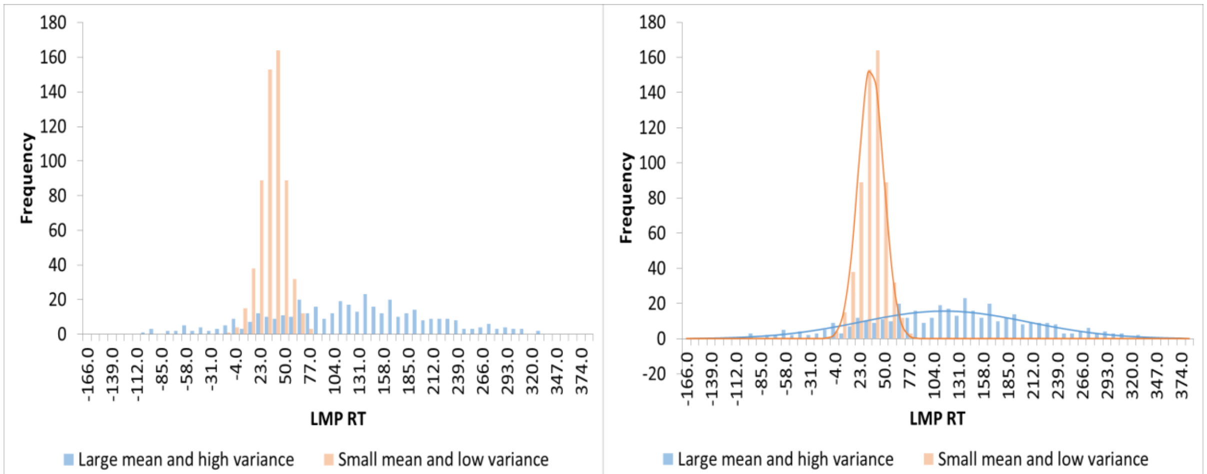

Figure 4. The RT LMP Simulated Data Separated Into, and Approximated Separately, by Two Distributions.

To illustrate the key ideas, consider Figure 4. It shows the same data as in Figures 2 and 3 above, but now with new color-coding of certain portions of the histogram. The left panel in Figure 4 shows the color-coding of particular subsets of the data in orange and blue bars: the orange-colored cluster of data has a mean RT LMP of around $31/MWh, with a relatively small variance. In contrast, the blue-colored price data have a much higher mean RT LMP of approximately $111/MWh, and a much higher variation (spread). We also show, in the right panel of Figure 4, normal distribution approximations (in solid lines) to each of the two distinct color-coded price distributions that comprise this overall price histogram.

Statistical mixture distribution models are a means to both identify and combine the constituent (called *component*) distributions that jointly characterize frequency data such as these. These models can sensibly capture the structure of electricity price data, their asymmetries (skewness), and the "fat right-tail" property that characterizes real-time electricity prices generally. This can be seen visually in Figure 5 below, which is the combined (i.e., mixture) distribution approximation to the full price data histogram in Figure 2 (i.e., the solid line in Figure 5 is the weighted sum of the two solid-line approximations in the right panel in Figure 4). The resulting approximating curve (the solid line) in Figure 5 is a proper probability distribution *per se* (that is, it integrates to one), so it can be used to calculate well-defined confidence intervals and the likelihood of each price outcome without advance knowledge of which component distribution applies to a given day.

---

therein and Johnson, Kotz, and Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, 2nd Ed. (John Wiley & Sons, 1994), Section 10.2.

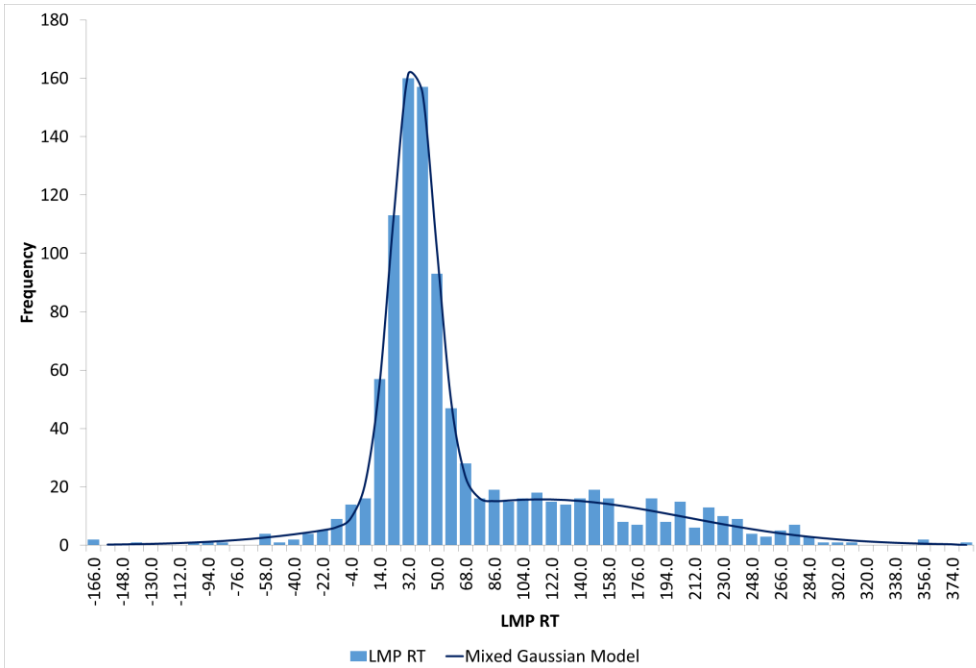

Figure 5. A Mixture Model's Fit to the Histogram of the Simulated Data.

▶ **The Key Points.** There are two key points to take away from this explanation of the data and modeling ideas. First, one can envision the statistical mixture distribution modeling technique we propose to employ as decomposing the underlying price data into different types of days (or hours), some with high price uncertainty (as in the blue histogram in Figure 4), and others with much less price uncertainty (as in the orange histogram in Figure 4).

The complicated part of such statistical modeling, stated in simple terms, is that real data do not come 'color coded'. Rather, we use the models and predictive input data (e.g., weather information, fuel prices, and such) to infer, statistically, both how to calibrate the different component distributions *and* how they should be 'mixed', or combined, on different hours and different days.

In effect, this calibration (or training) process performs two modeling functions simultaneously. One is that the model training process identifies both how spread-out and where each component distribution should be set, for each hour, as a function of all the (day-ahead) input data applicable to that hour. In addition, the model training process also identifies, in a statistically optimal way, how much 'weight' to put on each component distribution, where the appropriate weight on each will also (generally) vary hour-by-hour. For example, the model training process evaluates how much weight to place on a high-variance distribution (like the blue histogram in Figure 4) on a day with, say, high natural gas prices, and how much (less) weight to place on a lower-variance component distribution (like the orange histogram in Figure 4) on such days. And similarly for days with different gas prices, and different load or weather information, and so on.

The second point is that we have slightly over-simplified in our explanations so far, inasmuch as 'types' of days are not really discrete things. That is, electricity price variation is not easily (or reliably) categorized into simple binary or discrete day types, in a way one might say in informal parlance (e.g., as a 'high gas price', or a 'medium load level', day). Rather, there is a continuous range of fuel prices, forecast loads, and other input factors that impact real-time prices and their probabilities. Accordingly, in a statistical mixture model, the parameters (both the location and scale) of each component distribution continuously adjust as a function of all the input data values used to predict real-time prices and their probabilities.

Viewed from a practical standpoint, this means that with the statistical mixture modeling method, the data and days are not grouped into discrete categories *ex ante* to identify which underlying component distributions are more applicable to different days. In other words, there is no step in which the modeling team categorizes data into discrete day 'types' at all. Rather, the model training process itself optimally finds the component distribution shapes (in terms of location and scale), and simultaneously finds the optimal weights to place on each component distribution, both as continuous functions of all the available input data, in order to maximize the likelihood of observing (i.e., to "best fit") the historical price data with which it is trained.

This sounds mathematically involved, and some elements are. But the central idea is usefully illustrated in the example in Figures 2-5, above. The involved elements relate to the fact that a variety of statistically-useful data that helps predict real-time prices is employed in practice to determine how the shapes of those distributions should be adjusted, and how much each should be weighted, to calculate probabilities and expected values for each hour of the next (operating) day. We discuss these statistically-useful input data next.

3. Input Data and Considerations

A great deal of the variability in real-time electricity prices can be explained using predictive variables available at Stage 1 of the Day-Ahead Market timeframe. In this section, we summarize the main data types that are readily available each morning prior to the Day-Ahead Market that, in testing, we have found to be statistically meaningful for explaining next-day hourly real-time electricity prices in New England.

These predictive variables are employed both for training and predictive purposes. That is, the input data (or, rather, updated data for the same input variables) are employed for two purposes that take place on two different timeframes: (1) To train the statistical mixture model, which is done well in advance of its application each morning before the Day-Ahead Market; and (2) To predict the next (operating) day's expected hourly real-time energy prices and associated expected close-out cost values, using the trained model applied to each day's latest (updated) values of the input data, in the administration of the Day-Ahead Market (i.e., during Stage 1 shown in Figure 1).

3.1 Input Data Summary

After extensive testing, the ISO's modeling team determined that the most useful input data for predicting next-day hourly real-time electricity prices (and their confidence intervals) in the statistical mixture model are summarized in Table 2.[10] More detail about each of the relevant variables is included in Appendix A3.

Data	Units	Source
ISO's day-ahead system load forecast	MWh	ISO New England (existing public data)
Real-time hourly Hub LMP for the same hour of the prior day	$ per MWh	ISO New England (existing public data)
Natural gas price index (AGT-CG non-G)	$ per mmBTU	Intercontinental Exchange (commercial vendor service)
Fuel oil price (No. 6, 1%, NY Harbor)	$ per mmBTU	Argus Media (commercial vendor service)
Forecast mean cloud coverage	Percent	Several commercial weather forecasting services *Note:* Eight-city averages for New England. Hourly data.
Forecast mean temperature	°F	
Forecast mean dew point	°F	
Mean wind speed forecast	Miles per hour	
Winter season indicator	Value of 1 between November 1st and March 31st, and 0 otherwise	Constructed from date data

Table 2. Covariates Included in the Statistical Model

Although other related variables could be incorporated into the statistical model, the predictive informational value of other variables is already accounted for, to a large degree, in the information tabulated above. For example, weather information and its impact on net load is already incorporated into the system load forecast. As a result of such properties, we found that additional variables did not meaningfully improve the model's accuracy (the model's performance is reviewed in Section 5).

3.2. Input Data Impacts: Gas Prices

Gas prices are one of the more impactful predictors of real-time electricity prices. This should be expected, as gas-fired generation is on the margin (i.e., sets price) in approximately 80 percent of all hours annually. Importantly, the gas price not only helps to explain the *level* of real-time prices, it also explains the *variation* in possible real-time prices for any particular hour – and therefore characterizes the uncertainty in real-time prices as well.

We illustrate this dual role of the input data, such as gas prices, next. The following table groups all days over a two year period (2020-2021) by their gas price quartile, and shows the RT LMP's mean, standard

[10] The ISO can provide additional data descriptions as needed; all of the data indicated in Table 3 that are supplied by commercial data vendors are presently used in existing market administration and/or system operational processes at the ISO.

deviation, and inter-quartile range for the days in each gas-price quartile. For the first three quartiles of gas prices, the last column shows that the inter-quartile range of real-time electricity prices is relatively narrow (between $3 and $4/MWh in each case), and the third column shows that the standard deviation is similarly low (ranging from $2.61 to $4.15/MWh). However, when gas prices occur in their highest (fourth) quartile, the mean RT LMP and, importantly, the *variation* in RT LMPs, increases dramatically. The last row in Table 3 shows the standard deviation is much larger at $28.37/MWh, and the inter-quartile range expands similarly (to $75.67 - $47.93 = $27.74). This fact highlights the property we observe in the trained statistical model that there is a much wider (spread in the) distribution and confidence intervals for real-time prices on days with unusually high natural gas prices.

Gas Price Quartile	RT LMP Mean	RT LMP Standard Deviation	RT LMP Inter-Quartile Range
Q1	13.78	4.01	[12.81, 16.11]
Q2	22.08	2.61	[21.97, 24.24]
Q3	33.60	4.15	[33.30, 36.97]
Q4	66.91	28.37	[47.93, 75.67]

Table 3. Summary Statistics for Historical Hub RT LMPs per Gas Price Quartile

A related phenomenon is the variation in real-time prices that remains statistically ‘unexplained’ by the input data values on any given day. Unexplained variation arises because no predictive model of the future is a perfect predictor: some events that can cause large real-time price changes are inherently unpredictable (e.g., a sudden large generation source contingency (forced outage) that occurs in real-time). Larger such ‘unexplained’ variation in prices on some days will result in larger confidence intervals for the relevant hour(s), and larger expected close-out costs (generally).

Unexplained variation, and its impact and relation to gas prices, can be understood in terms of the standard errors (the ‘typical’ differences) between actual and expected real-time prices. The following graphs shows 2019 data in the left panel and 2021 data in the right panel. The standard error of the expected (predicted) RT LMP, which is a dynamic (hourly-varying) function when predicted with a statistical mixture model, is the higher-frequency data in blue, and measured on the left-side vertical blue axis in each panel. They day’s natural gas price is shown in orange, and measured on the right-side vertical orange axis in each panel.[11]

[11] The graphs in Figure 7 are based on a fully-trained model using historical data prior to these two years, as discussed in the next sections.

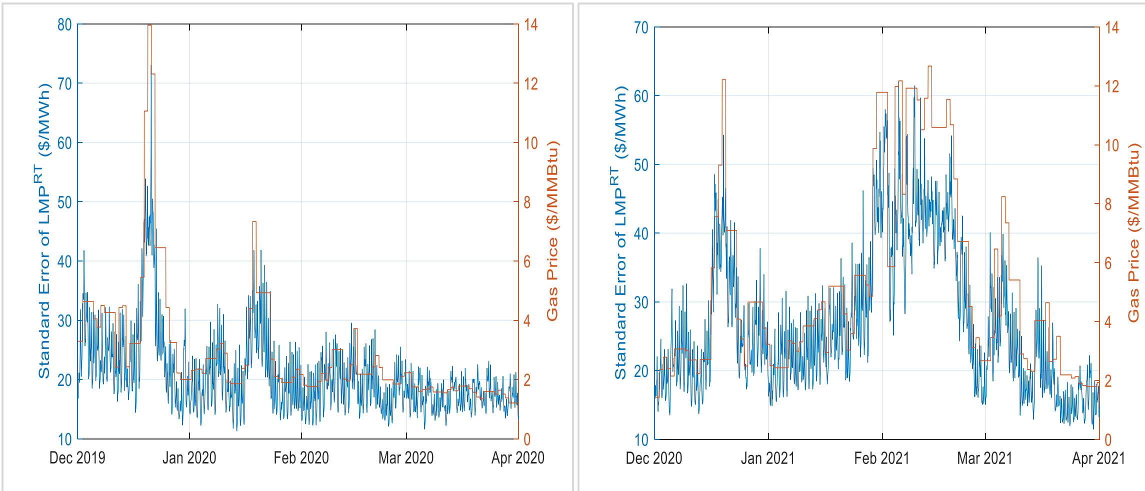

Figure 6. Gas Price and Standard Error of the GMM Model's Forecasted RT LMP

The point of Figure 6 is simple: During periods where natural gas prices are high, the *uncertainty* in the real-time electricity price is also much greater. Here uncertainty is measured by the standard error of the real-time electricity price (i.e., a measure of the 'typical' difference between the actual and the expected real-time price under those conditions). Similarly, Figure 6 also shows that when natural gas prices are lower, the uncertainty in real-time electricity prices is also lower.

We highlight these points for two reasons. First, they emphasize that key input factors, such as gas prices (and also load forecasts, among other things), not only help predict price levels, but are key determinants of the variation and uncertainty that remains regarding that prediction. Second, as noted earlier, expected close-out costs depend expressly on the uncertainty associated with the real-time electricity price; as shown in Figure 6, uncertainly can vary widely over time, even day-to-day. Accounting for that variation is important to provide expected close-out cost estimates that accurately model the uncertainties ancillary service sellers face at different times of the year, and from day-to-day.

4. Model Training and Testing

In this section we summarize the training and testing process performed to date, and upon which are based the model performance results (in Section 5). Formally, the statistical model we use is known as a *Gaussian Mixture Model,* or GMM for short. The mathematical specification of the GMM, the specific parameter values we have obtained, and the details of the optimization algorithm used to train and estimate these numerical parameters are provided in Appendices A1 through A5 of this memorandum.

In brief, there are several stages in the development and application of the GMM: First is the model's development, and second is its training (sometimes called the "fitting" process). Third is the testing and validation process. Last is the application of the trained and tested model for forecasting each day. We discuss each in turn next.

▶ **Model Development and Functionality.** The GMM we employ follows the logic discussed in Section 2, but with three component distributions instead of two. That is, we determined that a mixture of three different component distributions, each with its own mean and variance, provide the best fit to the historical data. Each component's mean and variance, in turn, are functions of the predictive variables and input data discussed in Section 3. This allows the location and scale of each component distribution to vary with day- and hour-specific data, such as fuel prices, the load forecast, and the weather information in Table 3. In addition, the GMM's relative weights associated with (that is, used to combine) each component distribution are (nonlinear) functions of the variables and input data discussed in Section 3. This enables the GMM to place greater, or lesser, weight on different component distributions in each forecast hour, as a function of (all of) the predictive input data in Table 3. For the functional specification of the GMM, see Appendix A1.

The formulated model is then trained (fitted) to historical data using an optimization algorithm. The objective function is standard maximum likelihood, that is, to find the set of all possible model sensitivities (numerical parameters) that, given the historical input data used as predictors (the explanatory variables), maximize the probability of realizing the real-time hourly prices that actually occurred during the training period. The optimization is performed using a standard iterative procedure, known as the E-M Algorithm, detailed in Appendix A2.

The trained model is, in simple terms, a (lengthy) mathematical formula. This formula takes as inputs values of all the explanatory data elements in Table 3. For the forecast hour of the next (operating) day, the formula then calculates the chance, or likelihood, of realizing each possible value of (i.e., the mathematical probability distribution of) the real-time price. The probability distribution of the real-time price for the forecast hour can then be (readily) used to estimate the percentiles of the real-time price, the expected value of the real-time price, and the expected value of the close-out costs (for any value of the strike price). The final formula that (technically) comprises the trained model does not change until the model is 're-trained'. All of the outputs – probabilities, percentiles, expected values, and so forth – that the model produces will vary each day and hour, as new input data values are used in the formula for each hour of each day.

▶ **Model Training**. For the evaluation and results provided in this memorandum, we trained the GMM with data covering the seven year period from January 1, 2012, to December 31, 2018. By "train" we mean "to estimate the parameters of the statistical model" in Appendix A1. In a nutshell, the training process consists of optimally fitting each component distribution's mean, variance, and weight functions using this lengthy period of historic data.

The total number of hours included in the training data is 61,368. During this training period, the New England system experienced a wide range of fuel prices, system load levels, weather conditions, and real-time price outcomes. The GMM's final estimated parameters applicable to this training period are provided in Appendix A3.

▶ **Model Testing**. We evaluate the trained model's performance during a subsequent (also called an "out of sample") testing period, using historical data that followed the years used for the model's training. More specifically, during the testing period, we used the trained model, each day, to forecast 'one day ahead' the hourly real-time price probabilities, expected value of the real-time price, and expected close-out costs (assuming, for simplicity in this evaluation, a strike price equivalent to the expected real-time price). For each hour during the testing period, we compare the historical realized real-time price and actual close-out cost with their expectations, as calculated ‘a day in advance’ by the trained model. The forecasting algorithm is enumerated in Appendix A4, and a simple example of its execution is provided in Appendix A5.

The goal of this phase is to evaluate the performance of the model in conditions analogous to how it will be used in practice: To apply the trained model with new daily inputs each day, and evaluate its accuracy. We sequentially tested the model, without retraining, each day for a three-year period from January 1, 2019 through December 31, 2021. The total number of hours in the test-period data is 26,304, and the performance results discussed in Section 5 are calculated from the model’s results during this testing period.

▶ **Re-Training Considerations.** Any statistical forecasting model requires periodic maintenance. Specifically, training a model on a given data set ensures that the forecasts account for structural factors and the market conditions that generated the historical data. Naturally, however, as time progresses, some of those factors may tend to become more or less important in predicting prices and their probabilities, due to the markets’ evolution and structural changes in the underlying variables. As this occurs, the model’s performance may degrade and necessitate re-training with updated data that incorporate more recent market and system conditions.

One way to see the need for periodic retraining is in the results over the three year testing period we have used to date. Figure 7 shows the trained GMM’s average monthly standard error (the monthly average ‘typical’ difference between the actual and expected real-time price), on the vertical axis, over the three-year testing period of 2019-2021. The dashed line in the figure is the trend, which is upward overall.

The figure shows that, during 2021 (i.e., toward the far right), the model’s typical error significantly increases and seems to follow a slightly different pattern. Before 2021, the monthly standard error remains typically under $15/MWh, with seasonal spikes observed during higher gas price winter months and higher load level summer months. Such seasonal spikes from 2019 through 2020 generally have a similar magnitude. However, in 2021 two things change. First, the standard error spikes become of a larger magnitude and more frequent. Second, the model’s typical error rises to over $15/MWh in the final year, from lower average levels in the earlier years, indicating a systematic increase in the magnitude of the forecasting errors of the model. Indeed, the latter point is confirmed by the dashed line representing the trend in the model’s typical errors, which shows that, as time passes after training the model, the model’s errors become larger.

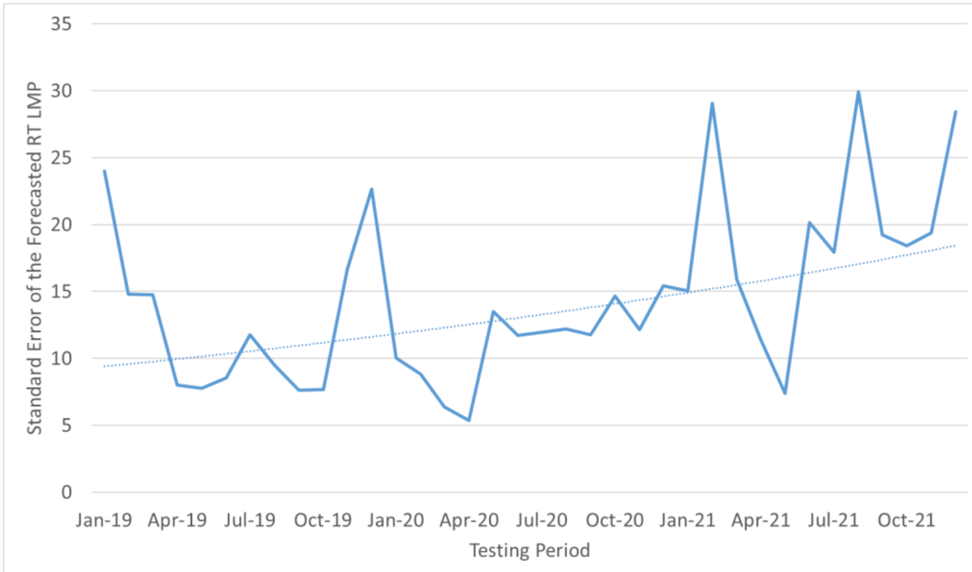

Figure 7. GMM Model's Estimated Standard Error Over the Three-Year Testing Period

This has two central implications. First, statistical models must be periodically retrained. Currently, the ISO envisions retraining the GMM, using the latest updated data, on a quarterly basis. However, during implementation (in 2024), the ISO will continue to evaluate its recommended retraining timeframe and, if a different frequency is recommended, can apprise stakeholders of the same.

Second, the performance results discussed in the next section, which are based on this three-year testing period without any retraining, are likely to be conservative assessments of model performance that we should observe in practice. That is, the results reported below for the three-year testing period incorporate the increasing model (standard) errors observed in the second half of 2021, as shown above. However, in practice, with a model that is retrained each quarter, one would not expect to observe the performance degradation that is evident without retraining (as seen in Figure 7). In sum, for the performance results below for a three-year period without retraining, the measures of model accuracy should tend to be conservative measures of the accuracy of a regularly-retrained model in practice.

5. Results and Comparative Performance

This section summarizes and interprets several performance measures of the trained GMM's forecasts, based on the testing period results. We provide statistics concerning the accuracy of the model's expected real-time price, when used as forecast, and compare its performance to that of a leading commercial vendor's real-time forecast electricity price. We then evaluate the model's distributional properties, in terms of the accuracy of the model's calculated confidence intervals in practice (using metrics known as *coverage probabilities*). Finally, we build on this evaluation of the distributional properties to compare the model's expected call-option closeout estimates with those that would have been observed historically.

To start, however, we first consider a simple two-day illustration of the model's estimated confidence intervals, expected real-time prices, and actual real-time prices, in order to connect the explanations and results of the trained GMM to the key ideas set out in Sections 1 and 2.

5.1 Confidence Intervals and Expected Prices: Example

As noted at the outset of this memorandum, a key feature of the statistical model we propose and have trained is that it provides estimated confidence intervals (probability distributions) for the real-time price, for each hour of the next (operating) day. To illustrate these and their related outputs, Figure 8 shows hourly results for two different days: July 17, 2019 and March 05, 2020. These illustrative days have significantly different real-time price distributions and diurnal load profiles.

In each graph in Figure 8, the red line shows the model's expected real-time hourly Hub LMP. The blue line shows the actual real-time hourly Hub LMP. Generally, the expected price series is smoother (over time) than the actual price series, which is typical of forecasting models. It occurs because forecasting tools cannot account for all possible factors that impact price, or for any unanticipated real-time system events (unexpected resource unavailabilities during the operating day, different weather than forecasted day-before, and so on).

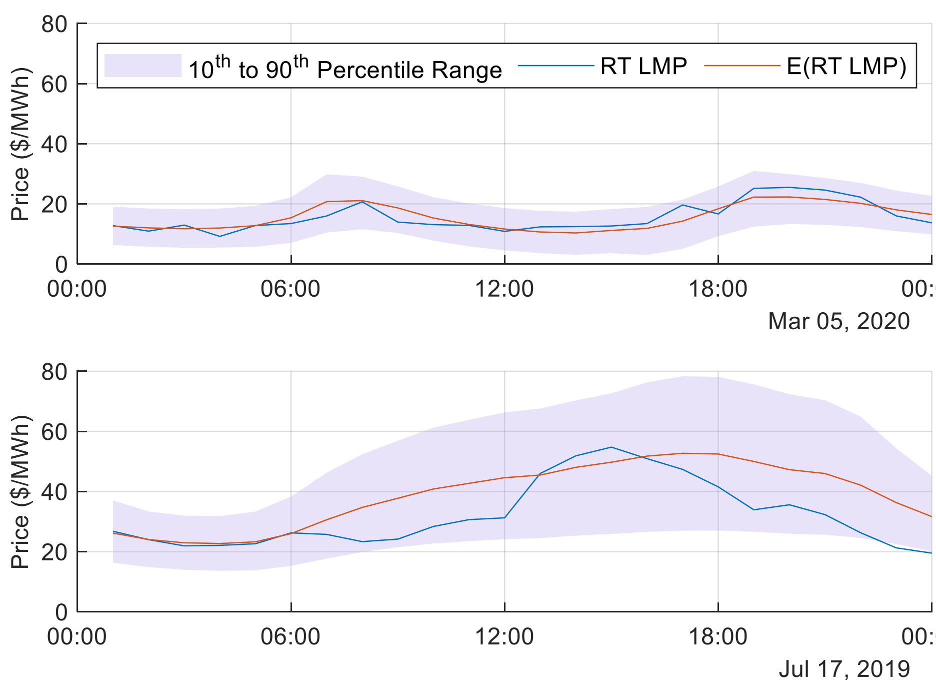

Figure 8. Realized and Expected RT LMP With The 80% Inter-Percentile Range

The shaded region in both panels of Figure 8 indicate the 10th and 90th percentiles of the estimated probability distribution of the real-time hourly Hub LMP, for each hour, as calculated by the GMM using the input data available prior to the Day-Ahead Market. The lightly shaded “band” in each graph therefore comprises an estimated 80 percent confidence interval for each hour’s real-time price. The term “80 percent confidence” interval reflects the fact that, when averaged over a span of time, intervals constructed in this manner from the GMM are expected to include the actual real-time price in 80% of all hours. In Figure 8, these confidence intervals include the actual real-time price slightly more than 80% of the hours shown (the actual price falls outside the bands only late in the day in the upper panel). In Section 5.3 below, we provide more comprehensive data on the accuracy of the model’s confidence intervals over the full testing period.

Note that the confidence intervals differ considerably over the course of the first day in the upper panel, and between the two days. For example, for July 17, 2019 (upper panel), the 80% confidence interval for the hour ending at 18:00 is from approximately $25/MWh to nearly $80/MWh, a nearly $55/MWh range. This is approximately twice the range of the 80% confidence interval for the same day during the pre-dawn hours (i.e., prior to hour ending 05:00), when the model’s expected real-time price was also a more accurate predictor of the actual real-time LMPs. Moreover, on the lower price (and lower load) day of March 5, 2020 in the lower panel, and also the hour ending 18:00, the 80% confidence interval runs from only $10/MWh to approximately $25/MWh, a $15/MWh range. In sum, the trained model adjusts both its expected real-time LMP calculations, and its probabilities and confidence intervals, in sensible ways over the course of the day, as the model is applied over time to very different day types and conditions.

To relate these results to the concepts in the simulated-data example in Section 2, Figure 9 below shows the full probability distribution of the real-time price calculated with the statistical model for two of the actual hours above: hour ending 18:00 on July 17, 2019 (in the upper panel of Figure 9), and hour ending 12:00 on March 5, 2020 (in the lower panel of Figure 9). The upper panel shows the model’s day-before estimated probability distribution for this hour’s real-time price has an asymmetric shape, with a fat right-tail, which is a common occurrence in the data and model results. This wide (spread-out) predicted probability distribution of the real-time price for that hour tends to pull up the expected real-time price, shown below and in Figure 8, of approximately $55/MWh.

In contrast, the lower panel of Figure 9 shows the model’s day-before estimated probability distribution for an hour with a lower expected real-time price. The estimated distribution for this hour has shifted and narrowed considerably (relative to that in the upper panel in Figure 9), illustrating the inherent shape and locational flexibility of the statistical mixture modeling technique. The distribution in the lower panel in Figure 9 is relatively narrow, nearly symmetric, and, in this hour, the actual and expected real-time price nearly coincide.

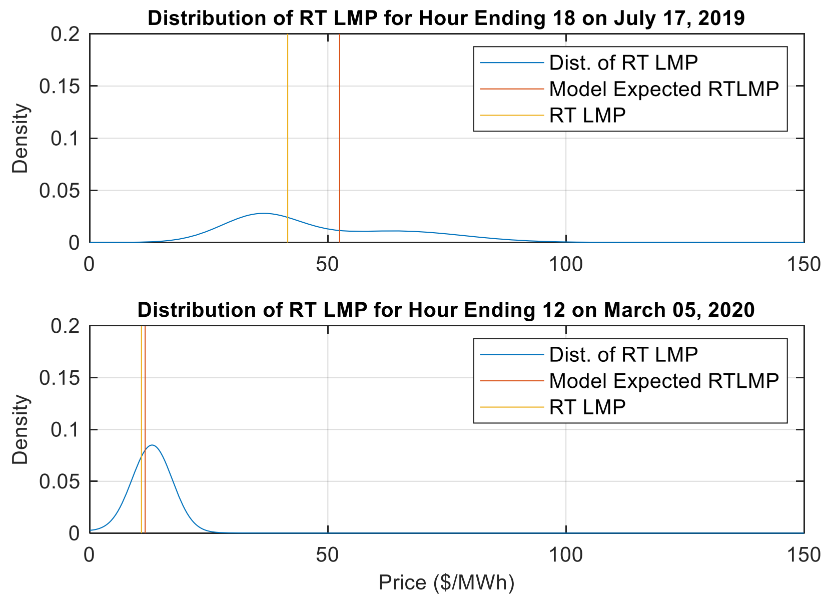

Figure 9. The Model's Day-Before Estimated Probability Distribution of the RT LMP for Two Different Hours

Overall, these figures highlight common features observed in the model's results. In many – but not all – system hours, the model calculates asymmetric probability distributions for the next day's real-time electricity prices, matching that common characteristic evident in the relative frequency distributions of actual hourly electricity price values. In addition, both the expected real-time prices, and the probabilities of different price outcomes, can shift substantially from day to day, and even across hours of the same day, with respect to the distribution's location, its shape (skew), and its overall scale (width). These features sensibly capture the key characteristics in the variation in electricity price data.

5.2 Overall Accuracy of the Expected RT LMPs

Below we provide three measures of the accuracy of the model's expected RT LMP, as a forecast of the actual RT LMP: the forecast's correlation with the actual RT LMP, the forecast's mean absolute error, and the forecast's root mean square error. We benchmark these results against an alternative New England real-time price forecast that the ISO procures from a leading commercial forecasting vendor, and that the vendor provides to the ISO daily.[12]

The overall results summarized in Table 4 show that the statistical mixture model, even without retraining over a three year testing period, performs well and is slightly superior to the forecast available from the

[12] For 2019, we use the commercial vendor's New England real-time electricity price forecast from 7:00 am on the day before the operating day; this is the latest forecast that would be available from it prior to the day-ahead bid/offer submission deadline. For 2020 and 2021, data availably limited the analysis here to using the vendor's real-time electricity price forecast provided at 6:00 pm the prior evening (i.e., the evening prior to the Day-Ahead Market).

commercial forecasting service. Here, higher correlation values indicate a more accurate forecast; lower mean absolute error and root mean square error values indicate a more accurate forecast.[13] (The formulas for these statistics are provided in Appendix A6).

Forecast	Correlation Coefficient	Mean Absolute Error	Root Mean Square Error
ISO's GMM real-time price forecast	.75	$9.27/MWh	$15.77/MWh
Commercial vendor real-time price forecast	.71	$9.74/MWh	$16.65/MWh
For Reference Purposes:			
DA LMP	.83	$7.48/MWh	$13.26/MWh

Table 4. Comparison of the RT LMP Forecasts

As a point of reference, we also include the same statistics for the day-ahead market's Hub LMP, which itself can be interpreted as a predictor of the real-time Hub LMP. In general, the DA LMP performs better than either of the model-based forecasts assembled prior to the Day-Ahead Market. This should not be surprising, inasmuch as the day-ahead market incorporates far more information into its calculations than does either of the two forecasts computed prior to the Day-Ahead Market. (Recall that the DA LMP cannot be used for setting the strike price, as it is only realized *after* resources submit their bids/offers for ancillary services into the DA market; see again Figure 1 in Section 1). Notably, however, the DA LMP itself is not all that more accurate than either of the two 'prior to' forecasts, differing in mean absolute error by approximately two dollars.

The DA LMP is a useful reference point. Economic arguments suggest that the accuracy of the market-derived DA LMP should be a limit on the accuracy of any real-time price forecasting method computed (using public information) *prior to* the Day-Ahead Market. This economic argument rests on arbitrage and information: If, *in arguendo,* there existed a systematically superior forecast of the real-time price than the DA LMP and that forecast was computable prior to the Day-Ahead Market, participants with knowledge of it would have a strong arbitrage incentive to make use of it in the Day-Ahead Market (i.e., by submitting profitable virtual transactions), until the additional supply and demand from those transactions improved the accuracy of the DA LMP down to the accuracy of the forecast. Viewed from that perspective—which suggests the DA LMP is (in theory) the best performance that any price forecasting model (calculable prior to the Day-Ahead Market) can achieve—Table 4 indicates that both the GMM method and the commercial vendor's price forecast perform relatively well.

To assess the model's accuracy in hours when the system is more likely to be stressed, Table 4a below reports the same two error metrics shown in Table 4, but now computed only for the hours in the top 5% and 1% load forecast and gas price days during the three-year testing period. We observe similar results:

[13] These statistics do not imply that either forecast was systematically smaller or bigger than another one. The statistics summarize the size (magnitude) of the errors, rather than their direction.

the GMM real-time price forecast model is more precise than the commercial alternative and is not much less accurate than the DA LMP theoretical benchmark, included here for reference as well.

In dollar terms, the mean absolute error and root mean square error values in Table 4a are larger (generally, roughly double for the top 5% of hours and roughly 3-4 times larger for the top 1% of hours) than their 'all hours' counterparts in Table 4, because the forecast errors here are tabulated for only a small number of more extreme condition days. That is, in the underlying data to which the model is trained, there are only 1/20th (for the top 5% of hours) and only 1/100th (for the top 1% of hours) as many observations with these more extreme conditions, relative to the 'all hours' results in Table 4 earlier. Given the exceptionally high RT LMPs that tend to prevail on these limited number of tighter system condition hours annually, however, the overall average error values in Table 4a indicate as before that the GMM method (and the commercial vendor's forecast) performs relatively well.

	Top 5% Load Forecast		Top 5% Gas Price	
Forecast	**Mean Absolute Error**	**Root Mean Square Error**	**Mean Absolute Error**	**Root Mean Square Error**
ISO's GMM real-time price forecast	$18.10/MWh	$32.31/MWh	$21.69/MWh	$28.10/MWh
Commercial vendor real-time price forecast	$18.99/MWh	$32.55/MWh	$25.85/MWh	$34.44/MWh
DA LMP (for reference purposes only)	$16.92/MWh	$28.54/MWh	$18.74/MWh	$26.10/MWh
	Top 1% Load Forecast		**Top 1% Gas Price**	
Forecast	**Mean Absolute Error**	**Root Mean Square Error**	**Mean Absolute Error**	**Root Mean Square Error**
ISO's GMM real-time price forecast	$31.36/MWh	$55.57/MWh	$28.02/MWh	$33.12/MWh
Commercial vendor real-time price forecast	$31.44/MWh	$53.53/MWh	$32.51/MWh	$41.29/MWh
DA LMP (for reference purposes only)	$31.78/MWh	$47.99/MWh	$22.23/MWh	$29.35/MWh

Table 4a. Comparison of the RT LMP Forecasts for Top 5% and 1% Load Forecast and Gas Price Hours

Last, to assess whether there is any "systematic" bias in the GMM-based forecasts, relative to the commercial vendor's forecast and also relative to the DA LMP, we compute the correlation of the GMM-based forecasts' errors with the same-hour error of the commercial vendor's forecast and of the DA LMP. We find a strong positive error correlation between the models' errors, in both cases. Put in simple terms, this result indicates that in hours the GMM-based forecast materially mis-estimates the actual RT LMP, the other two estimates tend to do so as well. The practical implication of this finding is that we do not expect the GMM-based forecast to materially mis-estimate the actual RT LMP when using the commercial vendor's model would have improved things. A more detailed explanation of this finding is provided in Appendix A7.1.

5.3 GMM Method Uncertainty Estimates: Coverage Probability Accuracy

As noted above, confidence intervals—quantifying a range in which we expect the actual RT LMP to fall with a given likelihood—are a useful means to convey the uncertainty of a forecast. They also provide a useful means to evaluate the performance of a statistical forecasting model, here and generally. We discuss their accuracy when calculated with the trained model next.

One statistical performance metric that assess a model's confidence intervals is known as the intervals' actual *coverage probability*. Actual coverage probability simply measures the relative frequency with which the actual real-time price falls within the associated hours' day-before calculated confidence intervals. In visual terms, this corresponds to more systematically evaluating the variation illustrated previously in the context of Figure 8, which showed how often the actual real-time price occurred within the hourly 80% confidence intervals over those two particular days.

Stated more precisely, when a model is correctly specified and estimated, we expect the coverage probability to coincide with the significance level of the confidence interval (i.e., the 90% confidence interval to contain the realized RT LMP 90% of the time, the 80% confidence interval to contain the realized RT LMP 80% of the time, and so on). If the coverage probabilities are significantly different from the significance level of the model's confidence intervals, we can interpret it as evidence that the model's estimated probability distribution of real-time prices does not accurately reflect the likelihood and uncertainty associated with the actual RT LMPs (e.g., the model predicts a narrower distribution than is actually observed, or vice versa).[14]

Year	50% Confidence Interval	80% Confidence Interval	90% Confidence Interval
2019	51%	84%	94%
2020	53%	86%	94%
2021	46%	78%	89%
2019-2021	**50%**	**83%**	**92%**

Table 5. Coverage Probabilities for the Trained Gaussian Mixture Model, by Year

To inform this, Table 5 reports the coverage probabilities of the GMM-based model, by testing-period year and overall, for three conventional confidence interval ranges: 50%, 80%, and 90%. Overall, the bottom row shows that the statistical mixture model produces predicted confidence intervals that are quite close to the measure coverage probabilities, accurately reflecting the price variation (at those percentiles) observed in practice. For example, the 50% confidence interval covers exactly 50% of all realized historic RT LMPs during the sample period (as shown in the last row of the second column).

[14] A coverage probability that is significantly greater than the significance level of the confidence interval (e.g., a coverage probability of 75% for an intended 50% confidence interval) is a sign of a mis-specified model, as it points to an over-estimated variance in the model.

Overall, the estimates for the 80% and 90% confidence intervals appear to be slightly conservative, in that the coverage probability values in the two right-hand columns in Table 5 are nearly all slightly greater than their corresponding confidence intervals' nominal significance level. In other words, on average the model estimates a slightly wider distribution of 80% and 90% confidence intervals for the RT LMPs than is observed in the historic real-time price outcomes.

However, taken together, Table 5 shows that the GMM model performs quite well regarding the coverage probabilities, for all of these confidence interval levels. On this basis, we conclude that the day-before GMM-based model's confidence intervals are quite accurate in quantifying the levels of uncertainty associated with the real-time prices, when assessed over the course of each year.

Comparing the distributional properties and coverage probabilities of the trained GMM model with other alternatives is not feasible, to our knowledge, as the GMM model is the only forecasting tool that estimates a full RT LMP distribution for each hour.

5.4 Overall Accuracy of the Expected Closeouts

Given that New England's day-ahead ancillary services market as energy options is proposed for future implementation, there are no actual historical closeout cost data with which to directly assess the model's performance in estimating suppliers' option closeout costs. However, we can still evaluate how the model performs along this dimension by combining historic RT LMP data with the model's results. In this section, we summarize the accuracy of the model's expected closeout costs, based on the results from the three-year testing period (2019-2021).

More specifically, we can use the trained model to forecast each day's hourly expected RT LMP, which in this exercise we will use to simulate hourly strike prices. We can then compare the actual RT LMP each hour of the testing period to the simulated strike price and compute the "simulated actual closeout" cost, using the standard call option closeout cost formula (that is, max{0, RT LMP – *strike price*}). We can then compare these "simulated actual closeouts" to the expected closeouts implied for the same hour by the model, as calculated sequentially one-day-before the operating day using the trained GMM.[15]

Table 6 reports how the expected closeout cost estimates compares to the actual (that is, "simulated actual") closeout costs during the three year testing period. The first numerical row of Table 6 shows that on average, the expected closeouts derived from the GMM are quite accurate, with precision to within 0.05 $/MWh: the average ("simulated actual") closeout in the testing period is $5.52/MWh, whereas the GMM's expected closeout is nearly the same, at $5.56/MWh.

[15] A comparison of the expected closeout costs' accuracy with other day-before forecasting models is not feasible, as (again, to our knowledge) such models do not estimate the RT LMP's distribution used to compute the expected closeouts.

Quintile	Realized RT LMP	Expected RT LMP	Average Realized Closeout	Expected Closeout
Overall (2019-2021)				
All	32.95	31.18	5.52	5.56
Gas Price				
Q1	19.96	17.75	4.14	3.98
Q2	22.10	22.51	2.81	4.31
Q3	26.71	24.61	4.67	4.36
Q4	37.20	33.23	7.22	5.75
Q5	59.09	58.04	8.82	9.42
Top 5%	76.85	81.72	8.41	13.71
Top 1%	92.18	97.29	11.45	17.53
Load Forecast				
Q1	20.29	18.51	3.49	3.29
Q2	26.71	25.15	4.36	4.07
Q3	31.58	29.67	5.29	4.94
Q4	37.40	35.90	6.07	6.29
Q5	48.87	46.76	8.40	9.22
Top 5%	50.97	47.18	10.94	10.22
Top 1%	72.38	55.88	23.93	13.71

Table 6. Realized vs. Expected RT LMP and Closeouts by Quintiles of Gas Price and Load Forecast

Table 6 also reports on the accuracy of the expected closeout costs as a function of fuel price and forecast system loads. This is useful for understanding the accuracy of the model in estimating expected closeout costs under different system conditions (and more stressed, high load or higher fuel price, conditions in particular). The remaining data in Table 6 compares, by days sorted into natural gas price percentiles and (separately by hours sorted into) load forecast quintiles, the average actual RT LMP with the model's expected RT LMPs, and the average ("simulated actual") closeout cost with the model's expected closeout costs. Specifically, we present the information for the gas price and load forecast quintiles (Q1-Q5) and observations in the top 5% and 1%.

The results show that across all five quintiles, the average actual and expected real-time prices are quite close, generally within two dollars (per MWh) or less. This is also true for the closeout costs, where across the five quintiles most differ by less than fifty cents (per MWh), and are less than a dollar in even the highest load and gas price quintile. When we consider more extreme system conditions, for instance the hours in the top 5% or 1% of all gas prices or load forecasts, we observe that the model's expected closeout cost precision decreases (more so for the top 1% than the top 5% of hours). In general, this is an expected property of forecasting models trained to best explain the varied outcomes in historical data: outlier and rare conditions tend to lead to less accurate forecasts than observations and conditions that are far more frequently observed. Overall, however, we do not find the differences observed for the small number of top 5% or top 1% condition hours in Table 6 to present significant concerns with respect to the suitability of the GMM for its intended purpose in administering the proposed DA A/S market. In addition, we expect the differences in Table 6 to decrease over time in implementation with more frequent model retraining, for the reasons explained in the last portion of Section 4.

Overall, we conclude the model fares reasonably at adapting its predicted prices and expected closeout costs to different ranges of gas prices and load forecast over the course of this three year testing period.

6. Conclusion

The statistical forecasting model proposed and discussed in this memo has the capability to estimate the dynamics of next-day real-time electricity prices, and their time-varying uncertainty, with considerable flexibility. The model incorporates readily-available day-before fuel price indices, hourly load forecasts, and forecast weather information to produce forecasts, confidence intervals, and expected close-out costs that vary by hour. While sophisticated in its mathematical formulations, it formulas are readily computable, the model can be periodically re-trained as appropriate, and it can be implemented by the ISO to produce option strike prices and expected closeout cost values that are needed to administer various elements of the proposed day-ahead ancillary services market.

After evaluating the models performance over a three-year testing period, we conclude that the proposed GMM statistical forecasting tool is capable of forecasting the real-time price and its distribution with reasonable accuracy. Moreover, the model's forecasts of the real-time price compares favorably (that is, is slightly superior to) that available from an established commercial vendor's real-time price forecasts. Based on these findings, we expect the proposed model will be satisfactory for its intended purposes in administering the ancillary services of the co-optimized Day-Ahead Market.

Technical Appendix

A1. Mathematical Formulas for the Gaussian Mixture Model (GMM)

This section describes the GMM mathematically. Let $N(\mu, \sigma^2)$ be the density function of a normal distribution with mean μ and variance σ^2:

$$N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)} \tag{1}$$

A GMM with k component distributions consists of a parametric distribution function with $3k$ meta-parameters (μ_i, σ_i, π_i) , denoted by

$$G\left(\{\mu_i\}_{i=1}^{k}, \{\sigma_i\}_{i=1}^{k}, \{\pi_i\}_{i=1}^{k}\right) \tag{2}$$

To allow the meta-parameters of the model to depend on the covariates of interest (denoted by X), we further make the following parametric modelling decisions:

$$\mu_i\left(\beta_i^{\mu} \mid X_j\right) = X_j \beta_i^{\mu}, for\ i = 1, \dots, k \tag{3}$$

$$\sigma_i^2\left(\beta_i^{\sigma} \mid X_j\right) = e^{X_j \beta_i^{\sigma}}, for\ i = 1, \dots, k \tag{4}$$

$$\pi_i\left(\beta_i^{\pi} \mid X_j\right) = \frac{e^{X_j \beta_i^{\pi}}}{\sum_{i=1}^{k} e^{X_j \beta_i^{\pi}}}, for\ i = 1, \dots, k-1 \tag{5}$$

where X_j is a $1 \times n$ vector containing one observations of our n covariates and β is a $n \times 1$ vector of n parameters, one for each covariate. These functional forms exhibited the best fit for the model among several others that were tested. In the model training stage, we are interested in determining the parameters—that is, the values of the $\beta's$—that comprise the meta-parameters of the GMM on the left-hand side of (3)-(5).

As it stands, the model is statistically identifiable up to a constant. As a result, without loss of generality, we assume that $\beta_k^{\pi} = 0$. Under this modelling specification, the conditional probability distribution over possible RT LMPs when the realization of the covariates is X_j is given by the following conditional GMM specification:[16]

$$G\left(\beta \mid X_j\right) = G\left(\left\{\mu_i\left(\beta_i^{\mu} \mid X_j\right)\right\}_{i=1}^{k}, \left\{\sigma_i^2\left(\beta_i^{\sigma} \mid X_j\right)\right\}_{i=1}^{k}, \left\{\pi_i\left(\beta_i^{\pi} \mid X_j\right)\right\}_{i=1}^{k}\right) = \sum_{i=1}^{k} \pi_i\left(\beta_i^{\pi} \mid X_j\right) N\left(\mu_i\left(\beta_i^{\mu} \mid X_j\right), \sigma_i^2\left(\beta_i^{\sigma} \mid X_j\right)\right). \tag{6}$$

[16] We call our model "conditional" (or "hybrid") GMM because it evaluates the price probabilities as a function (that is, conditional upon) the input data. This is a generalization of the standard statistical GMM formulation.

A2. Estimation Algorithm: Maximum-Likelihood Estimation.

This section describes the method to fit (train) the parameters of the model. $G(\beta \mid X_j)$ can be interpreted as the probability of observation j of the RT LMP being produced by our hybrid GMM model when the parameters are β. Thus, under (conditionally) independent and identically distributed observations, given the information in X, the probability that our data $(RT\ LMP, X)$ with N observations of the covariates would be generated by the model is

$$L(\beta \mid X) = \prod_{j=1}^{N} G(\beta \mid X_j). \tag{7}$$

In our case, we do not follow the tradition of the literature of working with the logarithm of the likelihood function L. In the literature, the goal of relying on the log-likelihood is to derive analytical optimization conditions for the likelihood function, which then are implemented in the numerical algorithm that finds the optimizing parameters, saving computational resources and time. In our case, such conditions are complex and not needed. To estimate the parameters β, we numerically optimize the likelihood function L using Matlab. The parameters that maximize the likelihood function are those that maximize the probability that our dataset was generated by a hybrid GMM data generating process. We use the Matlab's pre-built solver. In general, the estimation results with different solvers should be accurate within negligible variation, when the solver follows the typical optimization algorithm described below:

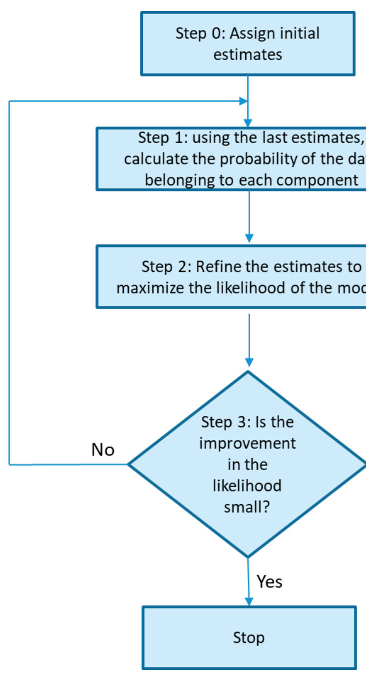

Figure 10. Optimization Algorithm

To determine the optimal number of components, we first optimize to maximize likelihood for models with a different number of component distributions ($k = 2, 3, \dots$). Then, we compute the likelihood for each specification. Finally, we compare the specifications' likelihoods to determine the model with the highest one. That model has the optimal number of components.

A3. Data, Model Training, and Parameter Values

We first detail the data sources and construction of the variables included in the model. For computational processes, some of these variables are scaled by constants (noted below) that enhance the estimation speed and reliability.

- ISO's day-ahead system load forecasts are provided under the "three-day system demand forecast" on the ISO's external website.[17] For a given hour, we use that hour's most recent next forecast, as posted on Monday 01/30 before 10:30 A.M. The information in the ISO's website is posted in MWh whereas the model is estimated using GWh units. Therefore, the load forecast must be divided by 1000 before estimating and executing the GMM model.

- Real-time hourly Hub LMP for the same hour of the prior day are the "Final Real-Time Hourly LMPs" on the ISO's external website.[18] To input the model and forecast a given hour of a given day's RT LMP, we use previous day's "Locational Marginal Price" (column G) of the CSV at the link below. For instance, to execute the GMM model and predict the hour ending 04's RT LMP on Tuesday 01/31, the model uses the Locational Marginal Price for hour ending 04 of Tuesday 01/31. For computational purposes, we divide the previous day's RT LMP in $/MWh by 100.

- Natural gas price index (AGT-CG non-G): The end-of-day (i.e., end of the trading day) price index for next-day delivery provided by the Intercontinental Exchange (ICE, via commercial subscription service).[19] On ICE's end, the price index is a MW-weighted average of all closed transactions by the end of the day. We are aligning the prices with the gas day. Therefore, to forecast day D-0 on D-1, we use the index from D-2's afternoon with next-day delivery. For instance, to forecast Wednesday, we need the gas price on Tuesday with next-day delivery, which is set on Monday's evening. The data source provides the price index in $/mmBTU and for computational purposes we divide it by 100. The same index is used for all hour's forecasts.

- Fuel Oil Price: "Argus, Fuel Oil, No 6, 1.0%, New York, Delivered, Waterborne" price is provided by Argus Media (a commercial subscription service).[20] For forecasting a day's RT LMP we use the price index at the closing business day two days before. For instance, to forecast Wednesday's RT LMP on Tuesday we use Monday's evening index. The data source provides

[17] Found at https://www.iso-ne.com/markets-operations/system-forecast-status/three-day-system-demand-forecast
[18] Final Real-Time Hourly LMPs is found at https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/lmps-rt-hourly-final
[19] https://www.theice.com/
[20] https://www.argusmedia.com/en

the price index in \$/mmBTU and for computational purposes we divide it by 100. The same index is used for all hour's forecasts.

- Weather forecast variables: the ISO internally constructs an index for each of the relevant weather variables (forecast mean cloud coverage (%), forecast mean temperature (^{o}F), forecast mean dew point (^{o}F), and mean wind speed forecast (mph)). The ISO relies on several vendors to provide the input information: Weather Bank (only for training period purposes), Weather Services Corporation, Weather Services International, Weather Bug, and Earth Network. For each variable of interest (e.g. cloud coverage), we consider an average of the reports of the weather vendors (which may not necessarily include the entire list above) across 8 cities in New England: Hartford CT, Boston MA, Bridgeport CT, Worcester MW, Providence RI, Concord NH, Burlington VT, and Portland ME. For computational purposes, we divide the resulting average by 100.

- Winter season indicator: the information on Table 2 is enough to construct this variable.

We summarize here the trained model parameter results from the estimation process (the algorithm described in Section A3), applied using the observable variables described in Section 3 for the model training period discussed in Section 4.

For computational purposes, we processed the observable variables through some simple transformations described below:

- Day-Ahead Load Forecast: Divide the value by 1000.

- 24-hour lagged LMP, Gas Price, No. 6 Fuel Price, Mean Cloud Coverage, Mean Temperature Forecast, Mean Dew-Point Forecast, and Mean Wind Speed Forecast: Divide each of these variables by 100.

Based on the training period data and its $n = 10$ covariates, we determined that the optimal number of components for our hybrid GMM is $k = 3$. As a result, we estimate a model with 80 parameters in total: Ten each for the three means and three variance, and ten for the first two of the three mixture weight functions. The result of our training process are the following numerical values for all of the fundamental sensitivity parameters of the model.

	k	Constant	DAM Load forecast	Lagged RT LMP	Gas Price	No. 6 Oil Price	Cloud Coverage	Temp.	Dew Point	Wind Speed	Winter
β_k^{μ}	1	-502	1.5	3.3	435	19.2	12.4	-15.5	5.3	-3.3	-3.1
	2	-1651	4.2	13.7	224	36.9	68.2	-41.5	4.4	-30.9	-8.1
	3	-8894	12.2	16.6	73	692.8	162.1	-238.6	79.5	-0.7	-37.3
β_k^{σ}	1	61	0.1	0.5	27	-2.8	4.0	-0.3	0.2	0.4	-0.1
	2	470	0.0	0.3	12	-0.2	1.5	-0.7	1.4	1.3	0.2
	3	509	0.1	0.1	0	1.1	-6.3	4.2	0.2	-0.3	0.9
β_k^{π}	1	562	-0.3	0.1	-33	9.1	1.6	6.0	-2.6	-7.3	0.7
	2	401	-0.2	-0.4	-10	4.0	0.1	5.0	-1.8	-1.8	0.6
	3	0	0	0	0	0	0	0	0	0	0

We do not place much emphasis on the direct interpretation of these coefficients, as this is a predictive model (the sensitivities measured by these parameters are not of primary interest; they are needed as intermediate values to perform *other* final predictive calculations). However, we note that our parameters are generally consistent with intuition. For instance, the first three values in the "gas price" column show that when the gas price increases, the model will predict a sharply higher mean RT LMP, and the next three values in this same column show it will predict a higher variance in the RT LMP as well.

A4. Forecasting Process Workflow and Algorithm, In Summary

The complete model training and (subsequent) daily forecasting algorithm is summarized as follows:

1. *Training stage.* Estimate a model with k components as described in section A2 to obtain the 80 maximum likelihood ("best fit") sensitivity parameters $\beta_i^{\mu}, \beta_i^{\sigma}, and\ \beta_i^{\pi}$.

2. *Forecasting general (static) formula.* Use the numerical parameter values obtained in step 1 to (substitute in) the equations in section A1 for each mean, variance, and weight.

3. *Forecasting formula daily update.* Use the forecasting day's data observations X_j to substitute in the equations found obtained in step 2, to find numerical values for the mean, variance, and weight of each component.

4. *Forecast the next day's price.* Forecast the expected RT LMP using the right-hand side of the formula

$$E\big(\, RT\ LMP \mid X_j \,\big) = \sum_{i=1}^{k} \pi_i\big(\beta_i^{\pi} \mid X_j\big)\mu_i\big(\beta_i^{\mu} \mid X_j\big). \qquad (8)$$

5. *Forecast the next day's probabilities of all other possible prices.* Forecast the distribution of the RT LMP using equation 6

$$G(\beta \mid X_j) = \sum_{i=1}^{k} \pi_i(\beta_i^{\pi} \mid X_j) N\left(\mu_i(\beta_i^{\mu} \mid X_j), \sigma_i^2(\beta_i^{\sigma} \mid X_j)\right). \tag{9}$$

A5. Executing the Forecasting Algorithm: A Numerical Example

1. Suppose we estimate a GMM with 2 components and two covariates, Y and Z, with the following parameters derived following section A2:

	$(\beta_Y^{\mu}, \beta_Z^{\mu})$	$(\beta_Y^{\sigma}, \beta_Z^{\sigma})$	$(\beta_Y^{\pi}, \beta_Z^{\pi})$
Component Distribution 1	(1, 2)	(0.05, 0.06)	(0.09, 0.1)
Component Distribution 2	(1.5, 2.5)	(0.07, 0.08)	(0.11, 0.12)

2. Substituting these parameters into the formulas for the mean, variance, and weights (equations 3, 4, and 5) shown in Appendix A1, we get the following equations that can be evaluated at any values of the covariates:

	μ	σ^2	π
Component Distribution 1	$Y + 2Z$	$e^{0.05Y+0.06Z}$	$\frac{e^{0.09Y+0.1Z}}{e^{0.09Y+0.1Z} + e^{0.11Y+0.12Z}}$
Component Distribution 2	$1.5Y + 2.5Z$	$e^{0.07Y+0.08Z}$	$\frac{e^{0.11Y+0.12Z}}{e^{0.09Y+0.10Z} + e^{0.11Y+0.12Z}}$

3. Finally, to forecast, we need to evaluate the equations in last table using the numerical values of the explanatory (covariate) data. If Y=10 and Z=25, we get the following estimates:

	μ	σ^2	π
Component Distribution 1	60	7.39	0.33
Component Distribution 2	77.5	14.88	0.67

4. The formula in equation 8 predicts an expected RT LMP of

$$0.33 \times 60 + 0.67 \times 77.5 = 71.69.$$

5. Finally, the formula in equation 9 predicts the distribution of the RT LMP to be

$$0.33 \times N(60{,}7.39) + 0.67 \times N(77.5{,}14.88).$$

A6. Statistics to Assess Pointwise Accuracy

For reference, we include here the standard formulas for the statistics reported in Section 5. Let "proxy" refer to one of the forecasting models (DA, GMM model, or Commercial Vendor). Let i denote a given observation. Let $\bar{X}$ be the average of variable X in the data set. When the dataset contains n observations, we compute:

$$Correlation\ Coefficient = \frac{\sum_{1=1}^{n}[\left(LMP_i^{RT} - \overline{LMP^{RT}}\right)\left(proxy_i - \overline{proxy}\right)]}{\sqrt{\sum_{i=1}^{n}(proxy_i - \overline{proxy})^2}\sqrt{\sum_{i=1}^{n}(LMP_i^{RT} - \overline{LMP^{RT}})^2}}. \tag{10}$$

$$Mean\ Absolute\ Error = \frac{\sum_{i=1}^{n}|proxy_i - LMP_i^{RT}|}{n}. \tag{11}$$

$$Root\ Mean\ Square\ Error = \sqrt{\frac{\sum_{i=1}^{n}\left(proxy_i - LMP_i^{RT}\right)^2}{n}}. \tag{12}$$

A7. Auxiliary Figures and Additional Performance Analysis Observations

A7.1 Correlation Across Models' Errors

Figure 11 shows on the vertical axis the errors of the GMM model with respect to the realized RT LMP. The horizontal axis of the left panel presents the same error for the DA hub LMP, whereas the horizontal axis of the right panel presents the same error for the commercial vendor's RT LMP forecast. Generally speaking, the point clouds form a positive pattern (pointing south-west to north-east) revealing that when the GMM error is small (large), the DA LMP and the commercial vendor's RT LMP forecast errors tend to be small (large) as well.

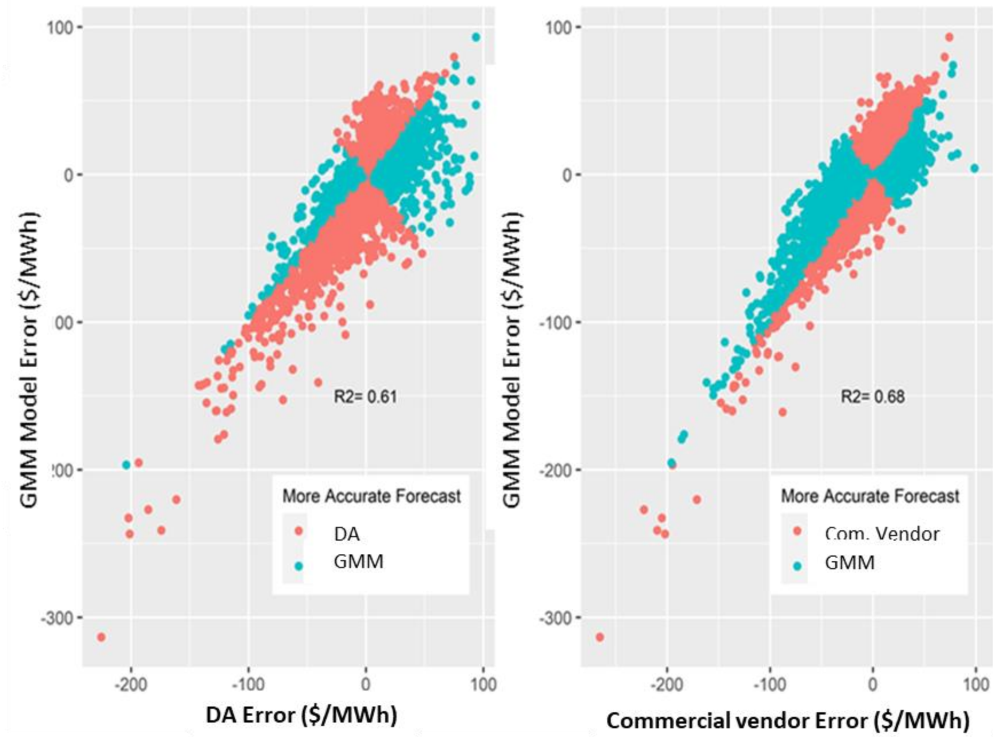

Figure 11. Correlation Between Models' Errors

The R^2 for the relationship is 0.61 between the DA errors and the GMM Model errors, and 0.68 between the commercial vendor's errors and the GMM Model errors. These numbers, in the context of linear regression, can be interpreted as (the squared) correlation coefficients, indicating that the correlation between errors is moderate to strong. Finally, the fact that there are no points far to the north-west or south-east direction means that there are no observations where there are material errors of the models in opposite directions (e.g., the DA Error was an overestimation and the GMM Model's error was an underestimation). In other words, the three models produce directionally similar forecasting errors.

A7.2. Correlation of the GMM model's errors with the load forecast and the gas price

To assess whether the GMM model's errors (variances) are intuitively consistent with high fuel prices and load levels, we examine the correlation of the GMM's standard errors with the two main components of the model, the load forecast and gas price. In general, we expect that the load forecast mainly drives the expected RT LMP's volatility during the summer and the gas price during the winter. This intuition is corroborated by the data in Figure 12.

The figure presents on the vertical axis the model's (hourly) standard errors, on the horizontal axis of the left panel the gas price, and on the horizontal axis of the right panel the load forecast. The figures show that during the summer, there is a strong positive correlation between the load forecast and the model's standard errors (the data cloud forms a south-west to north-east pattern). This pattern continues (significantly weaker) into the spring, and disappears in the fall and winter, where the data clouds seem to be unordered. Likewise, a similar positive relationship between the gas price and the model's standard errors appears in the winter and continues (weakly) into the spring, disappearing in the summer and fall.

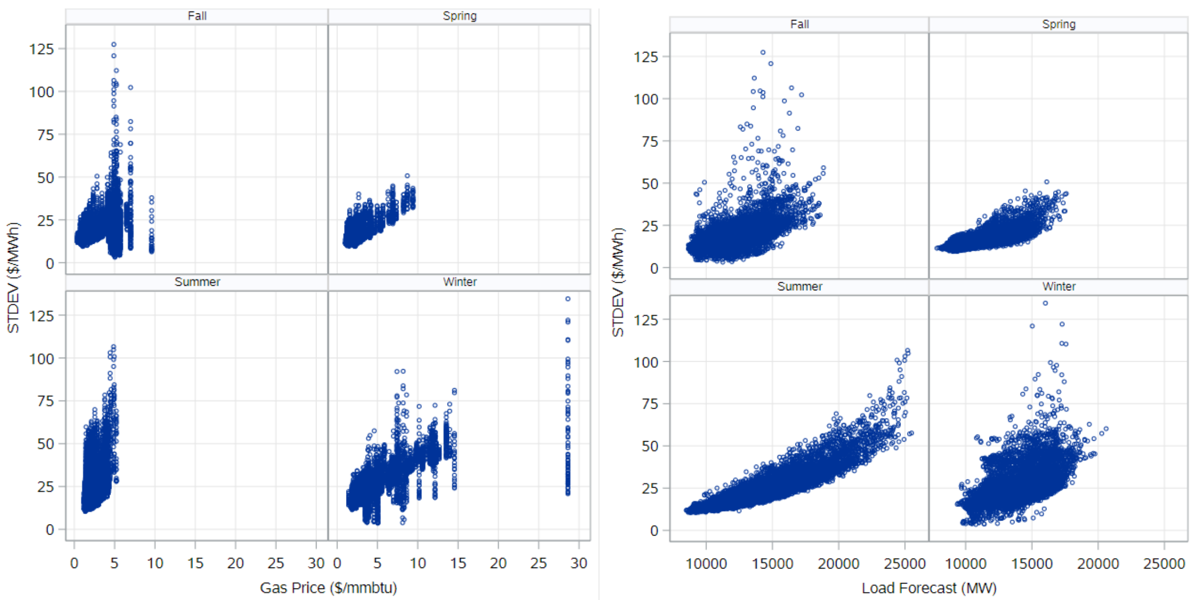

Figure 12. Correlation of the GMM Model's Standard Error with the Gas Price and Load Forecast by Season