

# **Econometric Model Design, Approach and Methodology Report**

**By Assistant Professor Wang Wenjie, Li Wenze and Cai Jinbo**

## **Contents**

<b>1 Executive Summary .....</b>	<b>2</b>
<b>2 Data .....</b>	<b>4</b>
2.1 Data Description .....	4
2.2 Trending Plot.....	6
2.3 Imputation .....	9
2.4 Summary Statistics.....	9
<b>3 Methodology .....</b>	<b>12</b>
3.1 Previous Model .....	12
3.2 Revised Model .....	12
3.2.1 OLS.....	12
3.2.2 ARIMA .....	12
3.2.3 LASSO.....	12
<b>4 Results .....</b>	<b>14</b>
<b>5 Outliers .....</b>	<b>21</b>
<b>6 Conclusion .....</b>	<b>22</b>
<b>7 Reference .....</b>	<b>23</b>
<b>8 Appendix.....</b>	<b>24</b>
8.1 Imputation Methods .....	24
8.1.1 Linear Interpolation .....	24
8.1.2 Spline Interpolation.....	24
8.1.3 Stineman Interpolation.....	24
8.1.4 Structural Model & Kalman Smoothing .....	24
8.1.5 ARIMA State Space Representation & Kalman Smoothing.....	25
8.1.6 Last Observation Carried Forward (LOCF).....	26
8.1.7 Next Observation Carried Backward (NOCB) .....	26
8.1.8 Mean Imputation.....	26
8.1.9 Median Imputation.....	26
8.1.10 Mode Imputation.....	26
8.2 Further Tables and Figures .....	26

## **1. Executive Summary**

The Market Surveillance and Compliance Panel (MSCP) Annual Report, prepared and submitted by the MSCP to the Energy Market Company Pte Ltd (EMC), comprises an analysis of outliers of the Uniform Singapore Energy Price (USEP) in the National Electricity Market of Singapore (NEMS) which are identified by an econometric model.

The econometric model was last reviewed in 2020, and proposed a dynamic regression model that incorporated market demand and supply, production costs, and time effects. Utilising this regression, an upper bound was constructed to identify outliers. Analysis of a unique dataset spanning from 2003 to 2019 revealed a decrease in outliers over time, suggesting increased efficiency in Singapore's electricity market in recent years. Consistent results were obtained through alternative model specifications, including hybrid models and the Least Absolute Shrinkage and Selection Operator (LASSO). These findings remained robust across different subsamples and data frequencies. Additionally, the review introduced a more comprehensive set of criteria for model selection. However, it was noted that the explanatory power of the previous econometric model could be significantly enhanced by incorporating more empirical features and utilising more detailed datasets, such as periodic-level data.

This report introduces a revised econometric model that improves upon the previous model in three key aspects:

- a. Utilisation of a more comprehensive dataset with finer detail; and
- c. Incorporation of several important new explanatory variables; and
- c. Implementation of imputation methods to address unavailable data.

Firstly, a more granular dataset was used for the revised model. Specifically, there was a transition from utilising daily-level data to periodic-level (i.e. 30-minute) data. This shift enabled the ability to capture more nuanced information and bolster the explanatory power of the model, as evidenced by a significant increase in the  $R^2$  value.

Secondly, the model was expanded to incorporate additional variables, including reserve requirement (MW), reserve availability (MW), reserve price (\$/MWh), regulation requirement (MW), regulation price (\$/MWh), Injection Energy Quantity (MW), Other (OT) Facilities Generation (MW), Withdrawal Energy Quantity (MW), and Offers Priced Below \$200/MWh. This augmentation enhanced the explanatory capability of the model. Additionally, a time-based dummy variable was introduced to further refine the analysis.

Thirdly, the previous model did not address the issue of unavailable data (i.e. data series that do not cover the full date range under analysis). Consequently, the data analysis software (named Stata) automatically discarded these observations, potentially compromising the reliability of the results. To rectify this, various methods were employed to infer the unavailable data, including interpolation techniques (Linear, Spline, Stineman), advanced statistical models (Structural Model & Kalman Smoothing, ARIMA State Space Representation & Kalman Smoothing), and simpler approaches (Last Observation Carried Forward, Next Observation Carried Backward, and imputation using Mean, Median, and Mode values). The imputed values generated by these methods were then averaged. Through these efforts, the accuracy of the results was enhanced compared to the previous model.

In order to compare the results from both models, an analysis was conducted on the sample period from 2003 to 2019, using the following explanatory variables in the revised model:

- a. lagged Uniform Singapore Energy Price (USEP);
- b. system demand;
- c. reserve requirement;
- d. reserve availability;
- e. reserve price;
- f. regulation requirement;

- g. regulation price;
- h. Injection Energy Quantity (IEQ);
- i. Combined-Cycle Gas Turbine (CCGT) generation;
- j. Steam Turbine (ST) generation;
- k. Other (OT) Facilities generation;
- l. Withdrawal Energy Quantity (WEQ);
- m. offers priced below \$200/MWh;
- n. reserve cushion;
- o. supply cushion;
- p. lagged fuel oil price;
- q. year dummies;
- r. month dummies;
- s. weekday dummies; and
- t. period dummies.

To pinpoint outliers based on the regression results, upper and lower bounds were established at three standard deviations (SD) above and below the predicted USEP in the analysis. Next, the actual USEP from the dataset was compared to the upper bound. If the actual USEP exceeded the upper bound, it was classified as a price outlier. Refer to Figure 6 for a graphical representation of these outliers.

In order to provide clear criteria for the model and variable selection process, several measures were introduced:

- a. R-squared ( $R^2$ );
- b. Mean Absolute Percentage Error (MAPE);
- c. Mean Absolute Error (MAE);
- d. Root Mean Square Error (RMSE);
- e. Akaike Information Criterion (AIC); and
- f. Bayesian Information Criterion (BIC).

Based on these measures, the updated model consistently outperformed the previous one in identifying USEP outliers, indicating greater reliability in both USEP prediction and outlier detection. Furthermore, the robustness of the revised model was bolstered through several ways:

- a. results were analysed using subsamples from 2003 to 2020 and 2021 to 2023<sup>1</sup>;
- b. alternative model specifications were explored, such as ARIMA;
- c. the choice of explanatory variables with the Least Absolute Shrinkage and Selection Operator (LASSO) method were compared; and
- d. daily observation randomness was smoothed out by averaging periodic observations to a daily level.

Notably, the revised model yielded robust results across different sample periods and model specifications, with fewer outliers detected compared to prior models.

Overall, the revised model offers several advantages:

- a. enhanced explanatory power of independent variables;
- b. improved accuracy through imputation methods for unavailable data;
- c. greater precision with more detailed periodic data; and
- d. clearer criteria for variable selection, facilitating adaptation to future data availability.

---

<sup>1</sup> These two time periods were analysed separately as a significant fundamental shift appeared in the USEP trend prior to 2021, compared to 2021 and after.

## 2. Data

### 2.1 Data Description

The dataset contains 368,160 observations, spanning from 2003 to 2023. Detailed definitions for each variable can be found in Table 1. In this table, USEP is identified as the dependent variable, while the other variables are classified as explanatory variables.

Table 1. Variable Definitions

Variables	Definition
$USEP_t$	The Uniform Singapore Energy Price [1] (USEP) is recorded at half-hourly intervals. For the analysis, USEP values below \$50/MWh and above \$4,000/MWh are excluded to remove any outliers that could potentially skew the model estimation.
$USEP_{t-1}$	One-period time lag of USEP
$Demand_t$	Forecasted half-hourly electricity demand for Singapore.
$PriResReq_t$	The system-level requirement from primary reserve providers, who must be available for activation within a 9-second response time and sustained for a minimum duration of 10 minutes [1].
$PriResPrice_t$	The market-determined price for the primary reserve product [2].
$Pri\_Avail_t$	The total capacity of primary reserves available.
$ConResReq_t$	The system-level requirement from contingency reserve providers, who must be available for activation within a 10-minute response time and maintained for a minimum duration of 30 minutes.
$ConResPrice_t$	The market-determined price for the contingency reserve product [2].
$Con\_Avail_t$	The total capacity of contingency reserves available.
$SecResReq_t$	The system-level requirement from secondary reserve providers. However, it should be noted that the secondary reserve was phased out in 2017 [3].
$SecResPrice_t$	The market-determined price for the secondary reserve product [2].
$TotalReq_t$	The total reserve requirement is calculated as the sum of PriResReq, ConResReq, and SecResReq.
$RegReq_t$	The system-level regulation requirement, where regulation refers to the standby generation tasked with fine-tuning or correcting frequency variations and imbalances between power demand and supply in the power system [1].
$RegPrice_t$	The market-determined price for the regulation product [1].
$IEQ_t$	Injection Energy Quantity pertains to the electricity amount injected by generation facilities, supplied by the Market Support Services Licensee (MSSL), SP Services [1].
$CCGTGen_t$	Combined-Cycle Gas Turbine (CCGT) Generation refers to the energy offered by CCGT units.
$STGen_t$	Steam Turbine (ST) Generation refers to the energy offered by ST units.
$OTGen_t$	Other Facilities (OT) Generation refers to the energy offered by OT units.
$OCGTCGen_t$	Open Cycle Gas Turbine (OCGT) Generation refers to the energy offered by OCGT units.
$Solar_t$ <b>(non-dispatchable)</b>	While there are Solar generation facilities connected to the grid in Singapore, solar generation is not scheduled for dispatch by the Power System Operator (PSO) in the wholesale market because the power output cannot be controlled or varied at will.
$Import_t$	Import registered facility or IRF means a facility, installation and/or apparatus used

<b>(dispatchable energy offers)</b>	for the purposes connected with import of electricity that has been registered as a registered facility to provide one or more of energy, reserve, regulation or contracted ancillary services. The information in respect of IRF can be found as a subtype of generation registered facility in the publication of Capacity for Registered Facilities.
<b><math>ESS_t</math></b> <b>(dispatchable energy offers)</b>	Energy Storage facilities are considered as generation facilities under the Market Rules as they produce electricity when discharging.
<b>Total supply<sub>t</sub></b>	Total supply is calculated as the aggregate of generation from CCGT, ST, OT, OCGT, Solar, Import, and ESS sources.
<b><math>WEQ_t</math></b>	Withdrawal Energy Quantity denotes the volume of electricity withdrawn by load facilities, as reported by the Market Support Services Licensee (MSSL), SP Services [1].
<b>OffersPricedbelow200<sub>t</sub></b>	The cumulative volume of energy offers priced at or below \$200 per MWh.
<b>VestingQuantityMW<sub>t</sub></b> <b>VestingQuantity%<sub>t</sub></b>	The quantity (in MW and % terms) for the purpose of hedging non-contestable consumer (NCC) load under the Vesting Contract [1].
<b>VestingPrice<sub>t</sub></b>	The vesting price is determined based on the long run marginal cost of the most efficient technology that accounts for at least 25% of the total electricity demand in Singapore [4].
<b>ResCushion<sub>t</sub></b>	Reserve Cushion refers to the excess reserve capacity available after dispatch, quantified on a half-hourly basis.
<b>Supplycushion<sub>t</sub></b>	Supply Cushion represents the percentage of total supply available after matching off demand, measured every half hour.
<b>Fueloilprice_lag<sub>t</sub></b>	One-month Lagged Oil Price reflects the impact of fuel costs on electricity production expenses. This variable is distinct in that it is collected monthly, in contrast to other variables that are recorded more frequently.
<b>Fueloilprice_lag2<sub>t</sub></b>	The square of the one-month lagged oil price.
<b>TPCAppliedYN<sub>t</sub></b>	The Temporary Price Cap (TPC), introduced on 1 July 2023, was implemented to stabilise the Singapore Wholesale Electricity Market (SWEM). TPC Y/N denotes whether the TPC is active.
<b>TPCPriReserves<sub>t</sub></b>	Cap level of the primary reserve price under the TPC [6].
<b>TPCContingencyReserves<sub>t</sub></b>	Cap level of the contingency reserve price under the TPC [6].
<b>TPCRegulations<sub>t</sub></b>	Cap level of the regulation price under the TPC [6].
<b>MAPT</b>	The Moving Average Price Threshold under the TPC [6].
<b>SpotLRMC<sub>t</sub></b>	The spot long-run marginal cost will be derived from the publicly available and published Japan-Korea Marker (JKM) prices, which are accessible to generation companies [6].
<b>TermLRMC<sub>t</sub></b>	The term long-run marginal cost [6].
<b>LCP<sub>t</sub></b>	Load Curtailment Price (in \$/MWh) is used to pay load registered facility (LRF) for its scheduled load curtailment quantity in each dispatch period.
<b>TCL<sub>t</sub></b>	Total curtailed load is the sum of the load scheduled to be curtailed (in MW) across all load registered facilities for each half-hour period.
<b>GC_Magnitude</b>	Gas curtailment magnitude is the magnitude of piped gas curtailment severity: ‘0’ means 0% curtailment; ‘1’ means 1-5% curtailment; ‘2’ means 6-10% curtailment; ‘3’ means 11-15% curtailment; ‘4’ means 16-20% curtailment; ‘5’ means 21-25% curtailment; ‘6’ means 26-30% curtailment; ‘7’ means 31-35% curtailment.
<b>EMAdirections<sub>t</sub></b>	EMA standing directions represents the total generation quantity directed by EMA for facilities to generate for a particular day and period.

## 2.2 Trending plot

Figure 1 shows daily average USEP trends from 2003 to 2023, displaying a generally stable pattern over two decades, with notable spikes from 2021 to 2023. These spikes suggest factors such as the global energy crunch from late 2021, increased electricity demand, maintenance of generation facilities, and tight supply conditions influencing energy price outliers in recent years.

Figures 2, 3 and 4 provide detailed plots for these peak years, where it can also be observed that some price spikes align with seasonal trends, particularly during May to July, the hottest months in Singapore.

Figure 5 shows a plot of the average demand for each period. The peak demand typically occurs from around 10:30 am to 9:00 pm, coinciding with the hottest hours of the day in Singapore.

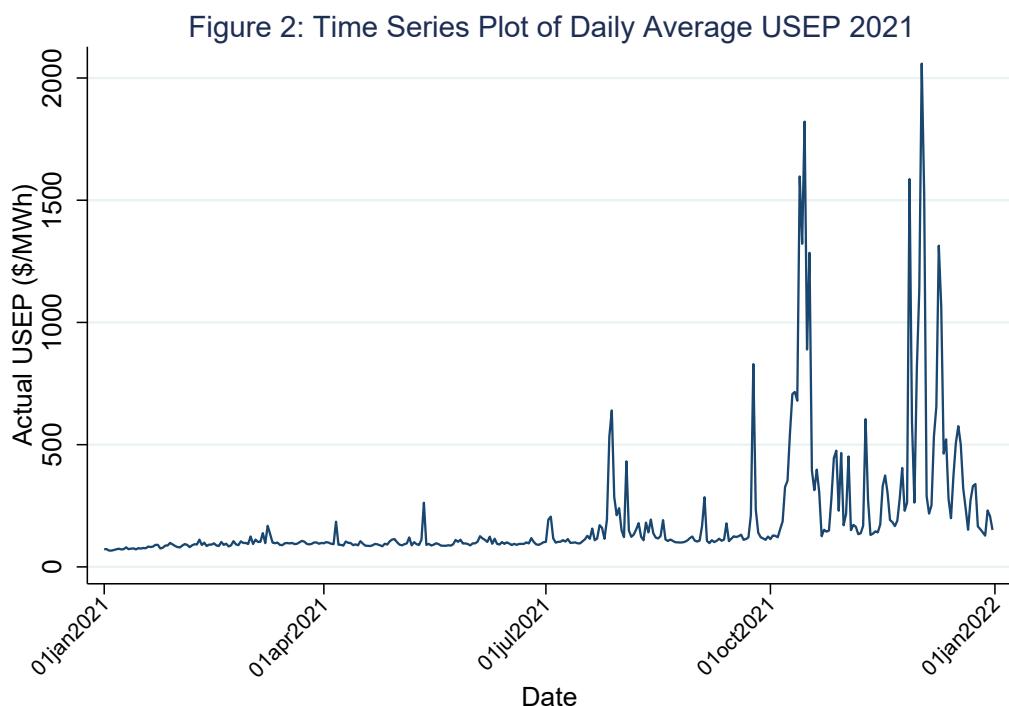
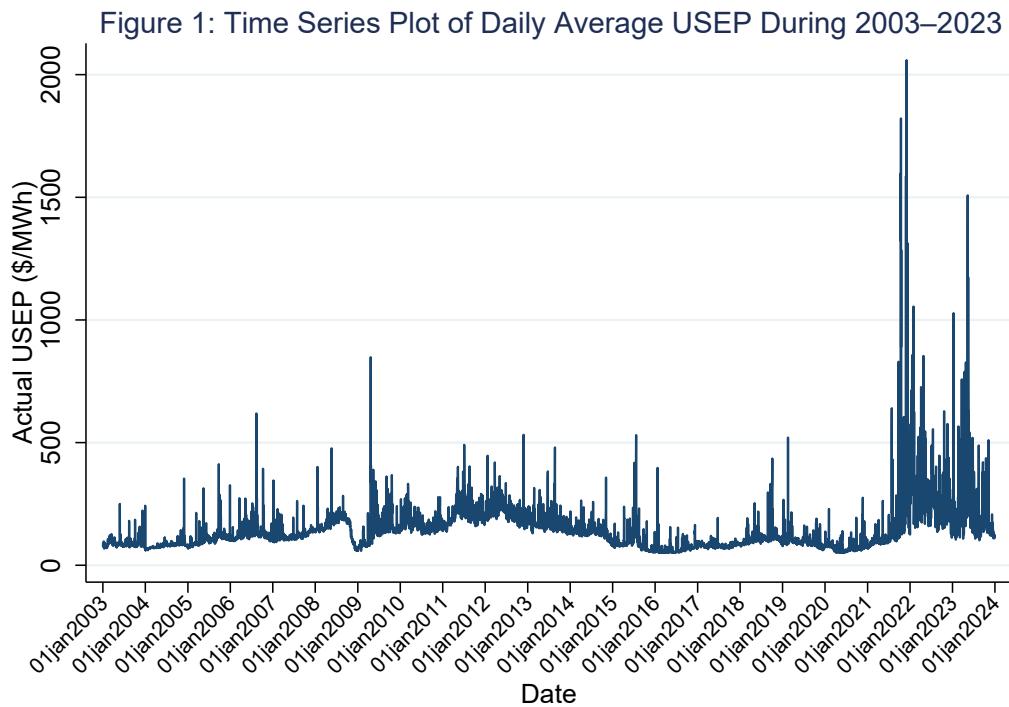


Figure 3: Time Series Plot of Daily Average USEP 2022

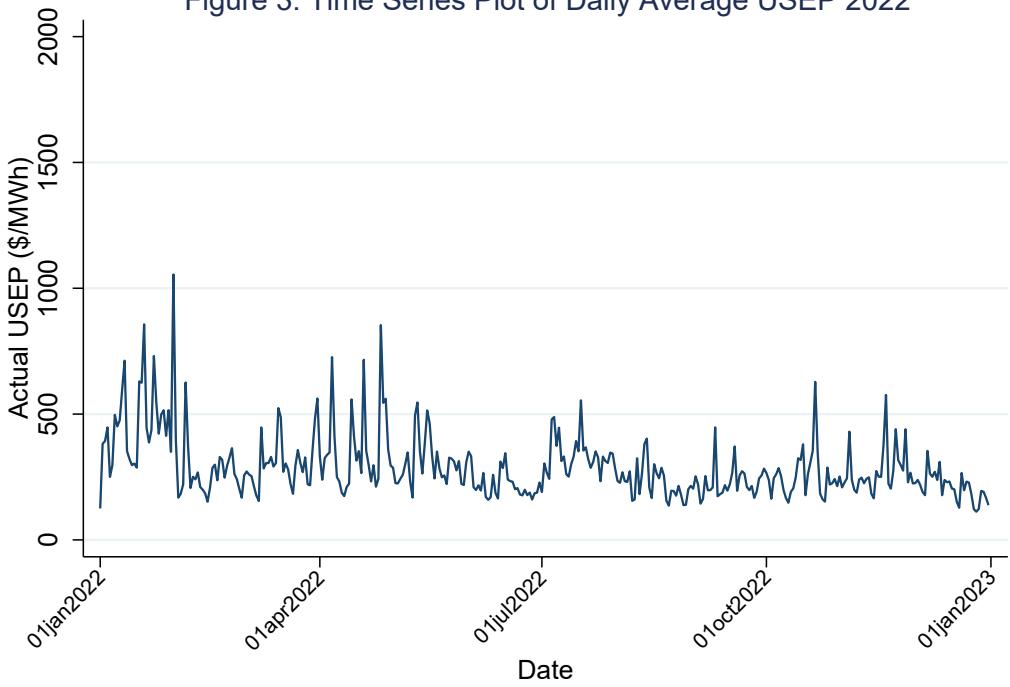


Figure 4: Time Series Plot of Daily Average USEP 2023

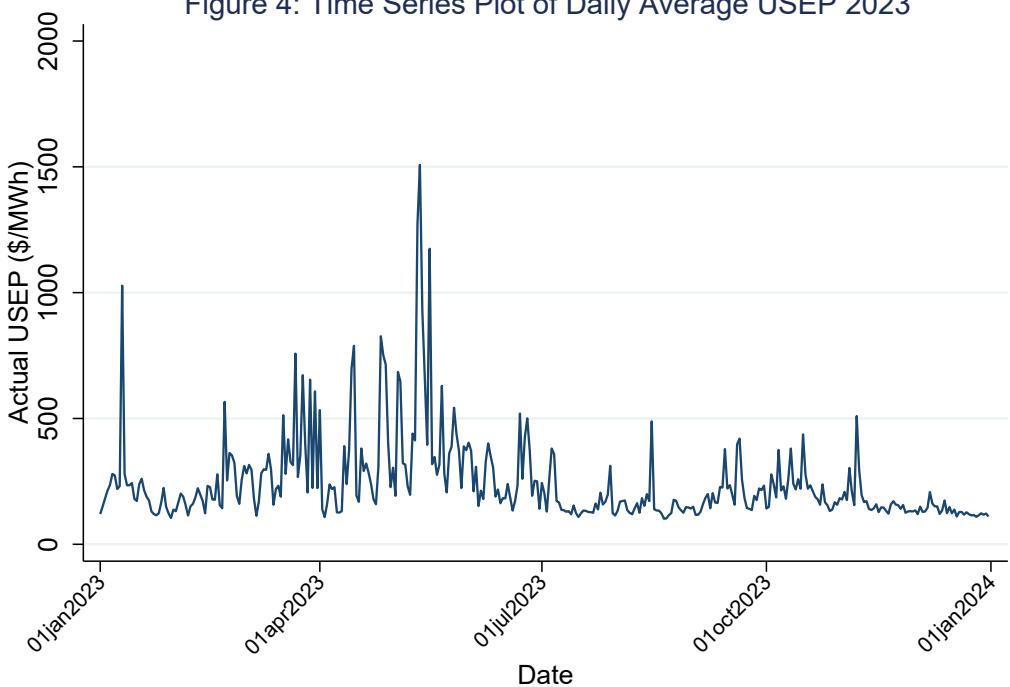
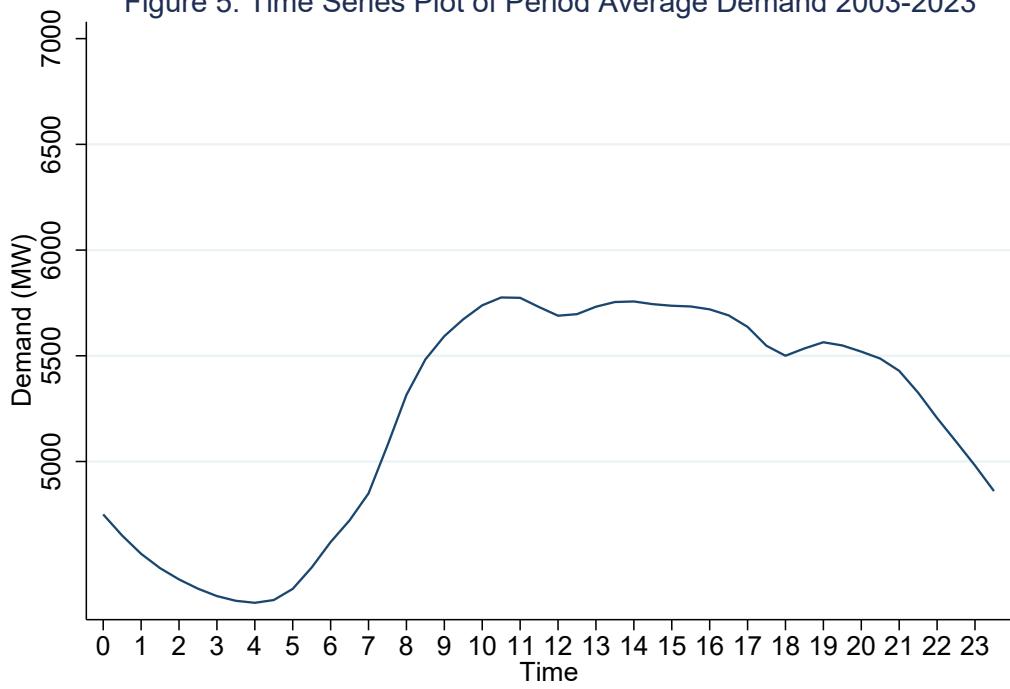


Figure 5: Time Series Plot of Period Average Demand 2003-2023



## 2.3 Imputation

The strategy for imputing unavailable values in variables involves a comprehensive multi-method approach. A suite of ten distinct imputation techniques were employed, encompassing various interpolation methods (Linear, Spline, Stineman), advanced statistical models (Structural Model & Kalman Smoothing, ARIMA State Space Representation & Kalman Smoothing), and simpler techniques (Last Observation Carried Forward, Next Observation Carried Backward, and imputation using Mean, Median, and Mode values). The specifics of these methods are detailed in the appendix. After applying these methods individually, the average of the imputed values from all these techniques was calculated. This averaged value is used to replace the unavailable values in the dataset, leveraging the combined strengths of each method for a more robust and reliable imputation. This general method was utilised for any data series that do not cover the full date range under analysis.

## 2.4 Summary Statistics

Tables 2 and 3 display the summary statistics of the daily pooled and daily average variables after imputation and before omitting USEP values that fall below \$50/MWh and above \$4,000/MWh.

Table 2: Descriptive Statistics: Daily Pooled

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max	(6) p25	(7) p50	(8) p75
Uniform Singapore Energy Price (\$/MWh)	368,160	144.0	157.3	-10.08	4,500	84.26	116.1	161.4
System Demand (MW)	368,160	5,226	972.5	2,535	7,569	4,505	5,240	5,987
Primary Reserve Requirement (MW)	368,160	201.1	42.57	95.28	418.4	171.1	190.8	225.9
Primary Reserve Price (\$/MWh)	368,160	3.679	44.63	0	4,250	0.0100	0.0500	1.020
Primary Reserve Availability (MW)	368,160	414.7	102.5	0	1,410	342.5	403.8	473.5
Contingency Reserve Requirement (MW)	368,160	542.6	76.93	252.6	838.0	515.3	543.1	606
Contingency Reserve Price (\$/MWh)	368,160	11.93	54.28	0	3,250	0.320	2	8.880
Contingency Reserve Availability (MW)	368,160	1,128	283.5	0	2,802	930.0	1,081	1,282
Total Reserve Requirement (MW)	368,160	929.8	164.9	384.3	1,487	809.4	967.3	1,060
Regulation Requirement (MW)	368,160	104.2	21.32	50	163.7	91.46	100	119.7
Regulation Price (\$/MWh)	368,160	45.39	112.1	0	2,750	8.210	22.45	49.14
Injection Energy Quantity (MW)	368,160	5,114	995.1	2,387	8,043	4,375	5,112	5,881
Combined-Cycle Gas Turbine (CCGT) Generation (MW)	368,160	4,717	1,333	1,259	7,869	3,651	4,696	5,841
Steam Turbine (ST) Generation (MW)	368,160	438.2	543.3	0	2,768	3	114.5	812
Other (OT) Facilities Generation (MW)	368,160	122.3	18.27	34	214	110	122	134
Open-cycle Gas Turbine (OCGT) Generation (MW)	368,160	202.5	74.55	0	408	160	180	218
Withdrawal Energy Quantity (MW)	368,160	5,099	1,002	2,429	7,633	4,353	5,095	5,869
OffersPricedbelow200	368,160	82.82	5.635	61.60	99.11	78.91	83.10	87.06
Reserve Cushion	368,160	48.43	12.44	-38.92	104.1	39.95	48.69	57.24
Supply Cushion	368,160	0.245	0.0706	-0.0445	0.560	0.201	0.242	0.293
Fuel Oil Price	368,160	408.5	157.6	153.1	748.2	282.6	401.8	503.5
Vested Quantity (kWh)	350,640	981,757	442,683	355,307	2.032e+06	548,713	867,158	1.324e+06
Vesting Price (MWh)	350,640	160.7	34.08	94.24	238.6	137.3	159.9	189.8
Vesting Quantity (%)	350,640	0.401	0.197	0.124	1.075	0.212	0.406	0.572
Secondary Reserve Requirement (MW)	258,576	265.0	23.48	165.9	414.2	251.3	264.1	279.8
Secondary Reserve Price (\$/MWh)	258,576	2.322	21.00	0	3,750	0.0200	0.140	1.040
Secondary Reserve Availability (MW)	243,331	470.3	100.2	0	784.3	390.9	453.7	538.0
Gas Curtailment Magnitude	210,379	0.446	0.926	0	7	0	0	1
Solar (MW)	153,024	14.67	34.08	0	513.8	0	0.00700	8.491

Import (MW)	26,736	29.51	44.91	-216.3	100	0	0	90.22
Energy Storage Systems (ESS) Generation (MW)	19,317	10.48	13.06	-0.200	116	0	2	18
MAPT (\$/MWh)	8,832	556.2	30.13	500.9	606.2	532.7	552.0	583.5
Spot LRMC (MWh)	8,832	197.6	22.24	165.1	233.2	177.6	195.5	220.1
Term LRMC (MWh)	8,832	181.4	8.705	166.9	191.2	173.6	183.6	190.2
Multiplier	8,832	2.834	0.235	2.500	3	2.500	3	3
Temporary Price Cap Applied (Y/N)	8,832	0.0212	0.144	0	1	0	0	0
TPC Primary Reserves	8,832	525.3	28.45	473.0	572.5	503.1	521.3	551.0
TPC Contingency Reserves	8,832	399.5	23.51	361.7	437.8	379.7	398.6	421.4
TPC Regulations	8,832	37.08	2.008	33.39	40.41	35.52	36.80	38.89
Load Curtailment Price (\$/MWh)	2,052	2,280	1,961	0	4,500	60.08	2,116	4,500
Total Curtailed Load (MW)	2,052	22.15	16.08	0.0300	63.90	6	12.60	35
EMA directions	1,774	169.5	88.07	17	400	125	180	215
CUSEP (MWh)	2,052	956.4	988.0	202.3	5,000	393.6	506.4	1,001

Note: Table 2 displays summary statistics for data post-imputation and before the exclusion of USEP values below \$50/MWh and above \$4,000/MWh.

Table 3: Descriptive Statistics: Daily Average

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max	(6) p25	(7) p50	(8) p75
Uniform Singapore Energy Price (\$/MWh)	7,670	144.0	104.2	34.48	2,059	85.86	119.4	172.0
System Demand (MW)	7,670	5,226	802.3	2,782	6,846	4,572	5,305	5,898
Primary Reserve Requirement (MW)	7,670	201.1	39.03	119.4	372.3	174.1	188.0	218.0
Primary Reserve Price (\$/MWh)	7,670	3.679	19.49	0.00792	963.9	0.0381	0.231	1.344
Primary Reserve Availability (MW)	7,670	414.7	94.78	76.55	1,409	344.3	406.5	475.8
Contingency Reserve Requirement (MW)	7,670	542.6	68.85	304.9	646.3	522.6	558.9	593.6
Contingency Reserve Price (\$/MWh)	7,670	11.93	24.33	0	659.9	1,455	4,732	12.52
Contingency Reserve Availability (MW)	7,670	1,128	219.7	308.0	2,801	988.0	1,122	1,252
Total Reserve Requirement (MW)	7,670	929.8	154.2	491.1	1,287	796.0	978.3	1,047
Regulation Requirement (MW)	7,670	104.2	14.71	80.84	126.5	94.17	100	117.2
Regulation Price (\$/MWh)	7,670	45.39	69.62	0.0929	1,246	13.07	28.12	58.43
Injection Energy Quantity (MW)	7,670	5,114	826.5	2,685	6,877	4,436	5,182	5,801
Combined-Cycle Gas Turbine (CCGT) Generation (MW)	7,670	4,717	1,259	1,505	7,397	3,698	4,940	5,797
Steam Turbine (ST) Generation (MW)	7,670	438.2	501.3	0	2,193	3	132.7	846.7
Other (OT) Facilities Generation (MW)	7,670	122.3	17.72	34	207.9	110.8	121.2	133.2
Open-cycle Gas Turbine (OCGT) Generation (MW)	7,670	202.5	70.28	0	408	160	180	244.3
Withdrawal Energy Quantity (MW)	7,670	5,099	831.0	2,689	6,923	4,412	5,162	5,791
Offers Priced Below \$200/MWh	7,670	82.82	4.981	66.01	97.31	79.08	83.01	86.27
Reserve Cushion	7,670	48.43	9.472	6.627	104.0	42.32	49.72	54.83
Supply Cushion	7,670	0.245	0.0513	0.0531	0.463	0.225	0.249	0.277
Fuel Oil Price	7,670	408.5	157.7	153.1	748.2	282.6	401.8	503.5
Vested Quantity (kWh)	7,305	981,757	380,270	468,771	1.671e+06	615,906	1.029e+06	1.323e+06
Vesting Price (MWh)	7,305	160.7	34.08	94.24	238.6	137.3	159.9	190.0
Vesting Quantity (%)	7,305	0.401	0.189	0.139	0.832	0.214	0.425	0.572
Secondary Reserve Requirement (MW)	5,387	265.0	15.34	209.0	352.2	254.4	263.8	274.6
Secondary Reserve Price (\$/MWh)	5,387	2.322	7.471	0	235.0	0.0950	0.507	1.501
Secondary Reserve Availability (MW)	5,104	470.7	91.69	271.2	731.9	393.5	454.4	541.2
Gas Curtailment Magnitude	4,383	0.446	0.902	0	7	0	0	0.854
Solar (MW)	3,188	14.67	16.91	0	75.55	1.486	7.912	23.54
Import (MW)	557	29.51	32.97	0	100	0	7.500	60.29
Energy Storage Systems (ESS) Generation (MW)	403	10.48	6.553	0	25.50	5.667	10.08	13.50
MAPT (\$/MWh)	184	556.2	30.21	500.9	606.2	532.7	552.0	583.5
Spot LRMC (MWh)	184	197.6	22.30	165.1	233.2	177.6	195.5	220.1
Term LRMC (MWh)	184	181.4	8.728	166.9	191.2	173.6	183.6	190.2
Multiplier	184	2.834	0.236	2.500	3	2.500	3	3
Temporary Price Cap Applied (Y/N)	184	0.0212	0.117	0	1	0	0	0
TPC Primary Reserves	184	525.3	28.53	473.0	572.5	503.1	521.3	551.0
TPC Contingency Reserves	184	399.5	23.57	361.7	437.8	379.7	398.6	421.4
TPC Regulations	184	37.08	2.013	33.39	40.41	35.52	36.80	38.90
Load Curtailment Price (\$/MWh)	334	2,010	1,350	0	4,500	800.5	2,095	3,000
Total Curtailed Load (MW)	334	16.55	13.04	2,990	52.37	6	10.55	30.02
EMA directions	113	150.1	90.58	20	400	53.50	180	197.4
CUSEP (MWh)	334	768.9	686.5	202.3	4,609	379.2	476.0	863.5

Note: Table 3 displays summary statistics for data post-imputation and before the exclusion of USEP values below \$50/MWh and above \$4,000/MWh.

### 3. Methodology

#### 3.1 Previous Model

$$\begin{aligned} \log(usep_t) = & \alpha + \beta_1 \log(usep_{t-1}) + \beta_2 trend + \beta_3 \log(ccgtsupply_t) + \beta_4 \log(stsupply_t) + \\ & \beta_5 \log(supplycushion_t) + \beta_6 \log(offers_t) + \beta_7 \log(demand_t) + \beta_8 \log(reservecushion_t) + \\ & \beta_9 \log(lagoilprice) + \beta_{10} \log(lagoilprice)^2 + \beta_{11} (\log(forceoutage_t) * \log(demand_t)) + \\ & \beta_{12} (\log(ccgtplannedoutage_t) * \log(demand_t)) + \beta_{13} (\log(stplannedoutage_t) * \log(demand_t)) + \\ & \gamma_k \sum_{k=1}^K year_k + \delta_m \sum_{m=1}^{12} month_m + \pi_p \sum_{p=1}^7 weekday_p + \varepsilon_t \end{aligned} \quad (1)$$

#### 3.2 Revised Model

##### 3.2.1 OLS

$$\begin{aligned} \log(usep_t) = & \alpha + \beta_1 \log(usep_{t-1}) + \beta_2 \log(ccgtsupply_t) + \beta_3 \log(stsupply_t) + \beta_4 \log(supplycushion_t) + \\ & \beta_5 \log(demand_t) + \beta_6 \log(reservecushion_t) + \beta_7 \log(lagoilprice) + \eta_1 X_{1t} + \beta_8 \log^2(lagoilprice) + \\ & \gamma_k \sum_{k=1}^K year_k + \delta_m \sum_{m=1}^{12} month_m + \pi_p \sum_{p=1}^7 weekday_p + \omega_n \sum_{n=1}^{48} period_n + \varepsilon_t \end{aligned} \quad (2)$$

Where  $X_{1t}$  includes Primary Reserve Requirement (MW), Primary Reserve Availability (MW), Contingency Reserve Requirement (MW), Contingency Reserve Price (\$/MWh), Contingency Reserve Availability (MW), Total Reserve Requirement (MW), Regulation Requirement (MW), Regulation Price (\$/MWh), Injection Energy Quantity (MW), Other (OT) Facilities Generation (MW), Withdrawal Energy Quantity (MW), and Offers Priced Below \$200/MWh.

In the revised model, the following adjustments were made:

- 1) Additional Explanatory Variables: The model was enhanced through the inclusion of new explanatory variables in  $X_{1t}$ , aiming to increase its comprehensiveness.
- 2) Incorporation of 48 periodic dummy variables  $period_n$  (where n ranges from 1 to 48) have been introduced to account for time-specific effects.
- 3) Omission of Terms with Minimal Impact: The study excludes both interaction terms and individual variables with negligible effects on USEP. This includes interaction terms such as  $\log(forceoutage_t) * \log(demand_t)$ ,  $\log(ccgtplannedoutage_t) * \log(demand_t)$ , and  $\log(stplannedoutage_t) * \log(demand_t)$ , as well as the trend variable, all defined in Equation 1.

##### 3.2.2 ARIMA

An ARIMA (p,d,q) model was also included in the analysis, where 'p' represents the lag of Autoregressive term, 'd' is the order of differencing, and 'q' is the lag of Moving Average. The ARIMA model is well-regarded for its forecasting capabilities, particularly in accounting for trends and seasonality. Considering that most of the variables are stationary, an ARIMA model with parameters p=1, d=0, and q=1 was selected. The ARIMA (1,0,1) model that was utilised is described as follows:

$$\log(usep_t) = \alpha + \beta_1 \log(usep_{t-1}) + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (3)$$

##### 3.2.3 LASSO

Additionally, LASSO (Least Absolute Shrinkage and Selection Operator) was employed in the analysis. As a popular machine learning method, LASSO improves model accuracy and interpretability by introducing an L1 penalty into

the regression equation:

$$\text{Minimize } (\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|) \quad (4)$$

Where,  $y_i$  represents the response variable,  $x_{ij}$  are the predictors,  $\beta_j$  are the coefficients,  $n$  is the number of observations,  $p$  is the number of predictors, and  $\lambda$  is a tuning parameter that controls the strength of the penalty (L1 regularisation).

This approach effectively selects important features and reduces overfitting by shrinking some coefficients to zero. LASSO is particularly useful in high-dimensional data for its simplicity and feature selection capabilities.

#### 4. Results

Table 4 presents the results of OLS, ARIMA, and LASSO analyses for daily pooled data.

Table 4: Regression Results (Daily Pooled)

	2003-2023			2003-2020			2021-2023		
	(1) OLS	(2) ARIMA	(3) LASSO	(4) OLS	(5) ARIMA	(6) LASSO	(7) OLS	(8) ARIMA	(9) LASSO
One Time Period Lag USEP (MWh)	0.765*** (154.608)		0.765*** (155.347)	0.683*** (86.999)		0.685*** (88.143)	0.746*** (109.232)		0.749*** (111.513)
L.ar		0.964*** (3,672.593)			0.965*** (3,226.689)			0.946*** (793.934)	
L.ma		-0.080*** (-208.589)			-0.115*** (-289.318)			-0.002 (-1.439)	
System Demand (MW)	0.053** (2.559)			0.258*** (12.923)			0.268*** (2.927)		
Primary Reserve Requirement (MW)	-0.030*** (-10.472)		-0.031*** (-10.649)	-0.034*** (-11.389)		-0.080*** (-24.711)	-0.443*** (-6.530)		-0.146*** (-13.644)
Primary Reserve Price (\$/MWh)			0.001 (1.151)			0.003*** (5.232)			0.009*** (3.155)
Primary Reserve Availability (MW)	-0.022*** (-7.230)		-0.020*** (-6.649)	-0.012*** (-4.345)		-0.006** (-2.158)	0.012 (0.876)		
Contingency Reserve Requirement (MW)	-0.000 (-0.028)		0.002 (0.295)	-0.028*** (-4.573)		-0.075*** (-14.065)	-1.193*** (-6.366)		
Contingency Reserve Price (\$/MWh)	0.025*** (53.356)		0.025*** (52.764)	0.032*** (50.192)		0.031*** (50.018)	0.023*** (23.458)		0.020*** (20.035)
Contingency Reserve Availability (MW)	0.149*** (29.039)		0.143*** (28.295)	0.149*** (24.865)		0.148*** (24.782)	0.283*** (16.230)		0.193*** (16.269)

Total Reserve Requirement (MW)	-0.099*** (-21.710)	-0.096*** (-20.910)	-0.111*** (-25.394)		1.356*** (5.327)		
Regulation Requirement (MW)	0.022*** (11.108)		0.011*** (6.146)		-0.049* (-1.883)		
Regulation Price (\$/MWh)	0.011*** (34.984)	0.012*** (39.782)	0.008*** (29.847)	0.008*** (31.779)	0.018*** (12.364)	0.020*** (13.318)	
Injection Energy Quantity (MW)	-0.503*** (-12.797)		-0.383*** (-10.536)		-2.813*** (-14.246)		
Combined-Cycle Gas Turbine (CCGT) Generation (MW)	-0.064*** (-12.738)	-0.066*** (-15.013)	-0.087*** (-15.794)	-0.063*** (-14.373)	0.006 (0.179)	-0.151*** (-6.572)	
Steam Turbine (ST) Generation (MW)	-0.002*** (-5.919)	-0.002*** (-6.298)	0.001** (2.399)	0.002*** (7.733)	-0.009*** (-6.799)	-0.009*** (-6.784)	
Other (OT) Facilities Generation (MW)	0.020*** (8.135)	0.019*** (7.932)	0.006*** (2.899)		-0.028*** (-4.253)	-0.021*** (-3.330)	
Open-cycle Gas Turbine (OCGT) Generation (MW)						-0.018*** (-5.780)	
Withdrawal Energy Quantity (MW)	0.443*** (11.780)		0.216*** (6.445)		2.337*** (14.071)		
Offers Priced Below \$200/MWh	-0.872*** (-55.743)	-0.868*** (-55.197)	-0.773*** (-46.369)	-0.764*** (-45.947)	-2.297*** (-37.443)	-2.324*** (-38.165)	
Reserve Cushion (%)	-0.074*** (-14.022)	-0.070*** (-13.478)	-0.106*** (-17.932)	-0.102*** (-17.269)	-0.135*** (-4.589)	-0.059*** (-3.097)	
Supply Cushion (%)	-1.598*** (-54.422)	-1.586*** (-54.875)	-1.424*** (-44.267)	-1.489*** (-44.230)	-4.011*** (-37.594)	-3.859*** (-40.399)	
fueloilpriceusmt_lag	0.479*** (15.412)	0.475*** (15.173)	-0.095*** (-3.518)		4.174*** (12.681)		
fueloilpriceusmt_lag2	-0.030*** (-11.450)	-0.029*** (-11.248)	0.026*** (11.238)	0.018*** (37.477)	-0.339*** (-12.896)	-0.007*** (-8.374)	
Constant	4.262***	4.830***	4.299***	5.440***	4.764***	5.481***	0.946    5.213***    13.737***

	(35.128)	(680.835)	(35.554)	(37.933)	(742.152)	(41.155)	(0.819)	(195.256)	(31.604)
Observations	356,492	358,004	356,492	304,019	305,531	304,019	52,473	52,473	52,473
R-squared	0.939		0.939	0.943		0.942	0.922		0.921
adjusted_R2	0.939		0.939	0.943		0.942	0.922		0.921

Note: Table 4 presents regression results using data post-imputation and after omitting USEP values outside the \$50/MWh to \$4,000/MWh range. Heteroskedasticity-robust t-statistics are shown in parentheses (\*\* p<0.01, \*\* p<0.05, \* p<0.1).

Overall, the coefficients associated with the explanatory variables exhibited consistency with the expected outcomes. Notably, there were significant negative coefficients linked to supply-side variables such as CCGT supply, ST supply, supply cushion, and reserve cushion. Conversely, the coefficient for demand was notably positive, aligning with standard demand and supply theory. Furthermore, the coefficient of 0.765 for the lagged USEP indicated its substantial influence in determining the current period's USEP, highlighting strong serial correlation across periods.

Additionally, the significantly positive coefficients of lag fuel oil price were reasonable, as fuel oil price represents a cost of generation; hence, as fuel oil price rises, so does USEP. Conversely, the negative coefficient for the square of the fuel oil price suggests a nonlinear relationship; initially, as fuel oil price increases, USEP rises, but beyond a certain threshold, USEP decreases. This phenomenon may stem from consumer behavior; when fuel oil prices reach a critical level, high USEP prompts consumers to reduce demand, subsequently leading to a decrease in USEP.

Table 5 presents the results of OLS, ARIMA, and LASSO analyses for daily average data.

Table 5: Regression Results (Daily Average)

	2003-2023			2003-2020			2021-2023			Previous Results by Feng and Zhou
	(1) OLS	(2) ARIMA	(3) LASSO	(4) OLS	(5) ARIMA	(6) LASSO	(7) OLS	(8) ARIMA	(9) LASSO	(10) OLS
One Time Period Lag USEP (MWh)	0.380*** (19.487)		0.384*** (19.747)	0.245*** (12.586)		0.248*** (12.821)	0.212*** (6.994)		0.210*** (7.137)	0.388*** (15.635)
L.ar		0.982*** (544.471)			0.989*** (504.576)			0.918*** (64.364)		
L.ma		-0.576*** (-108.326)			-0.654*** (-115.937)			-0.365*** (-13.570)		
System Demand (MW)	-1.065*** (-3.947)			-0.675*** (-3.034)			-2.312** (-2.026)		-2.099** (-2.210)	0.819*** (10.600)
Primary Reserve Requirement (MW)	-0.048* (-1.761)		-0.078*** (-2.800)	-0.044* (-1.924)		-0.118*** (-5.665)	-1.739* (-1.921)		0.003 (0.018)	
Primary Reserve Price (\$/MWh)		0.008** (2.054)			0.015*** (3.833)				0.030* (1.767)	
Primary Reserve Availability (MW)	-0.062*** (-3.122)		-0.056*** (-2.860)	-0.050*** (-3.359)		-0.042*** (-2.737)	-0.270** (-2.038)		-0.381*** (-3.379)	
Contingency Reserve Requirement (MW)	0.131** (2.219)			-0.120** (-2.172)		-0.131*** (-2.817)	-5.571** (-2.291)			
Contingency Reserve Price (\$/MWh)	0.058*** (22.115)		0.059*** (22.693)	0.070*** (27.896)		0.068*** (27.382)	0.059*** (5.201)		0.042*** (3.691)	
Contingency Reserve Availability (MW)	0.268*** (6.988)		0.224*** (9.084)	0.241*** (7.146)		0.129*** (7.226)	0.767*** (3.785)		0.410*** (4.147)	
Total Reserve Requirement (MW)	-0.268*** (-0.268***)		-0.130*** (-0.130***)	-0.263*** (-0.263***)			6.793** (6.793**)			

	(-8.037)	(-3.111)	(-9.401)		(2.031)		
Regulation Requirement (MW)	0.050 (0.950)	-0.005 (-0.093)	0.084* (1.666)		-0.581 (-0.980)	-1.215** (-2.407)	
Regulation Price (\$/MWh)	0.052*** (13.554)	0.050*** (13.246)	0.032*** (10.333)	0.031*** (9.822)	0.101*** (4.551)	0.112*** (5.201)	
Injection Energy Quantity (MW)	0.432 (0.570)		-0.772 (-1.257)		-3.384 (-1.386)		
Combined-Cycle Gas Turbine (CCGT) Generation (MW)	-0.166*** (-4.314)	-0.093*** (-2.714)	-0.210*** (-6.418)	-0.132*** (-4.422)	-0.145 (-0.441)	-0.282 (-0.893)	-0.234*** (-6.082)
Steam Turbine (ST) Generation (MW)	-0.008** (-2.159)	-0.005 (-1.409)	-0.001 (-0.346)	0.002 (0.592)	-0.035*** (-3.299)	-0.030*** (-2.868)	0.014*** (3.952)
Other (OT) Facilities Generation (MW)	0.082*** (3.459)	0.098*** (4.070)	0.048*** (2.678)	0.065*** (3.537)	-0.156*** (-2.805)	-0.148*** (-2.779)	
Open-cycle Gas Turbine (OCGT) Generation (MW)						-0.093** (-2.150)	
Withdrawal Energy Quantity (MW)	0.668 (0.936)		1.679*** (2.961)		5.829*** (2.856)	2.656*** (3.209)	
Offers Priced Below \$100/MWh							-0.568*** (-12.517)
Offers Priced Below \$200/MWh	-2.707*** (-27.883)	-2.670*** (-28.756)	-1.925*** (-25.582)	-1.915*** (-26.724)	-8.402*** (-16.426)	-8.594*** (-17.686)	
Reserve Cushion (%)	-0.054 (-1.289)		-0.138*** (-3.666)		0.230 (0.721)	0.672*** (3.381)	-0.248*** (-13.304)
Supply Cushion (%)	-5.075*** (-26.324)	-5.067*** (-28.771)	-3.875*** (-24.924)	-4.161*** (-28.131)	-13.395*** (-18.348)	-13.092*** (-17.922)	-1.595*** (-13.010)
fueloilpriceusmt_lag	1.768*** (7.187)	0.335*** (18.463)	0.188 (0.959)	0.000 (0.001)	10.879*** (3.331)		-0.109 (-0.524)
fueloilpriceusmt_lag2	-0.122*** (-5.801)		0.026 (1.562)	0.042** (2.488)	-0.896*** (-3.434)	-0.023*** (-2.954)	0.054** (2.221)

Trend										-0.000 (-1.212)	
Constant		10.594*** (11.556)	4.838*** (68.216)	14.424*** (23.097)	12.215*** (15.435)	4.756*** (58.189)	13.177*** (18.464)	9.070 (0.671)	5.283*** (51.567)	46.601*** (7.698)	7.579** (2.323)
Observations		7,612	7,642	7,612	6,517	6,547	6,517	1,095	1,095	1,095	6,032
R-squared		0.896		0.896	0.920		0.919	0.874		0.874	0.864
adjusted_R2		0.896		0.895	0.919		0.918	0.869		0.870	

Note: Table 5 presents regression results using data post-imputation and after omitting USEP values outside the \$50/MWh to \$4,000/MWh range. Heteroskedasticity-robust t-statistics are shown in parentheses (\*\* p<0.01, \*\* p<0.05, \* p<0.1).

Separate OLS regressions were also conducted for several variables . The available observations of these variables are significantly fewer than the full date range under analysis, typically because they no longer exist in the market, or commenced in more recent years. This necessitated a separate regression to be run for each variable. The results are presented in Table 6.

Table 6: Separate Analysis Results

Reg No.	Observations	Variables	Coefficients	Lasso Select YN
1	340,483	Vested Quantity (kWh)	-0.072*** (-33.278)	Y
2	340,483	Vesting Price (MWh)	0.001 (0.432)	N
3	340,483	Vesting Quantity	-0.217*** (-32.690)	N
4	250,389	Secondary Reserve Requirement (MW)	-0.285*** (-13.253)	Y
5	250,389	Secondary Reserve Price (MWh)	0.008*** (10.018)	Y
6	236,562	Secondary Reserve Availability (MW)	-0.036*** (-4.002)	N
		Gas Curtailment Magnitude=1	0.017*** (11.920)	Y
		Gas Curtailment Magnitude=2	0.002* (1.714)	N
		Gas Curtailment Magnitude=3	0.004** (2.393)	N
7	200,367	Gas Curtailment Magnitude=4	0.004 (0.942)	N
		Gas Curtailment Magnitude=5	-0.019*** (-6.913)	Y
		Gas Curtailment Magnitude=6	0.057*** (6.373)	N
		Gas Curtailment Magnitude=7	-0.023** (-2.304)	N
8	143,422	Solar (MW)	0.006*** (9.001)	Y
9	25,800	Import (MW)	0.004*** (5.617)	N
10	19,295	ESS (MW)	0.005*** (2.812)	Y
11	8,828	MAPT (MWh)	-0.086 (-1.156)	N
12	8,828	Spot LRMC (MWh)	-0.078 (-1.111)	Y
13	8,828	Term LRMC (MWh)	6.799*** (2.617)	N
14	8,828	Multiplier	6.876*** (2.618)	Y
15	8,828	TPCAppliedYN	-0.037*** (-2.658)	Y
16	8,828	TPC Primary Reserve (\$/MWh)	-0.086	N

				(-1.157)		
17	8,828	TPC Contingency Reserve (\$/MWh)	-0.111** (-2.042)		N	
18	8,828	TPC Regulation (\$/MWh)	-0.089 (-1.171)		N	
19	2,030	LCP (MWh)	-0.012*** (-4.666)		Y	
20	2,030	TCL (MW)	-0.022* (-1.953)		Y	
21	1,757	EMA Directions (MW)	0.005 (0.583)		N	

Note: Table 6 presents regression results using data post-imputation and after omitting USEP values outside the \$50/MWh to \$4,000/MWh range.

Additionally, the models demonstrate an improved performance relative to previous ones, which is evidenced through a higher  $R^2$ . For more detailed information, please refer to Table 7.

Table 7: Performance of Models

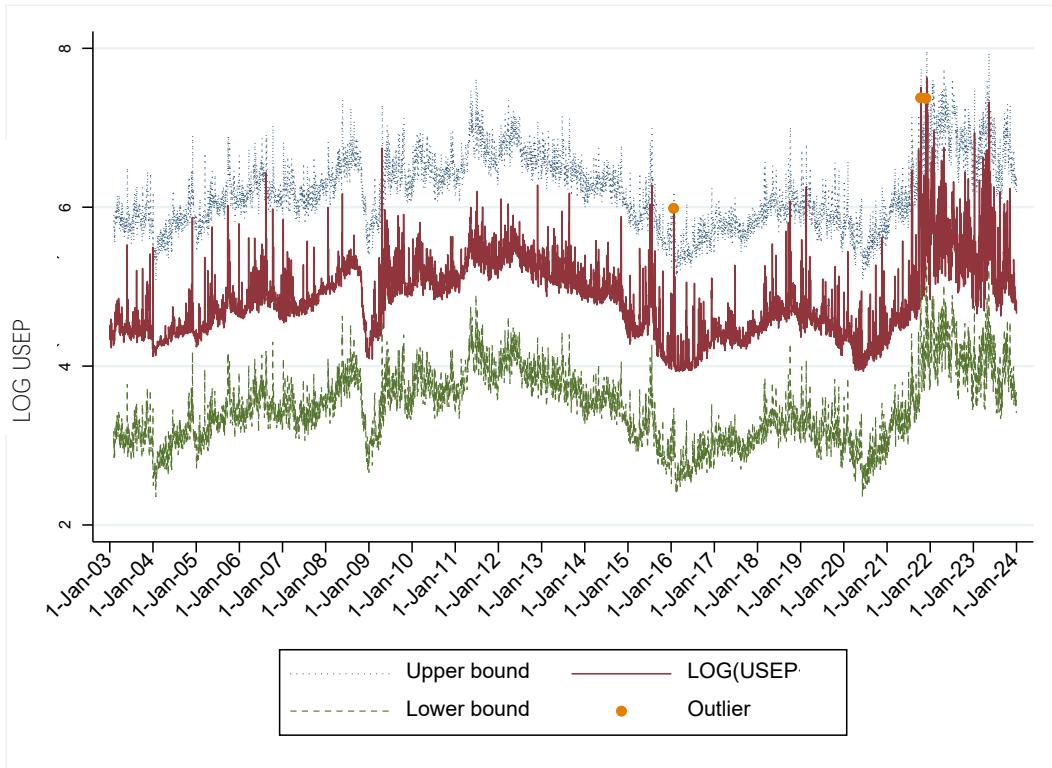
			R2	MAPE	MAE	RMSE	AIC	BIC
Revised Model (Daily Pooled)	2003-2023	OLS	0.939	1.214	6.11	0.121	-490073.4	-488941
		ARIMA		1.781	9.035	0.188	-374368.5	-374325.3
		LASSO	0.939	1.215	6.117	0.121	-489732.3	-488632.3
		Hybrid1 (OLS+ARIMA)		1.28	6.527	0.139		
		Hybrid2 (OLS+LASSO)		1.214	6.111	0.121		
	2003-2020	Hybrid3 (ARIMA+LASSO)		1.281	6.534	0.139		
		OLS	0.943	1.096	5.369	0.103	-516739.1	-515655.4
		ARIMA		1.601	7.899	0.164	-385122.7	-385080.2
	2021-2023	LASSO	0.943	1.101	5.387	0.103	-515964	-514944
		OLS	0.922	0.273	1.484	0.068	-31698.89	-30927.36
		ARIMA		0.418	2.308	0.109	-16062.25	-16026.78
		LASSO	0.921	0.274	1.488	0.069	-31342.14	-30623.83
Previous Model	2003-2019	OLS	0.864	1.82	8.96	0.15	-5,942	-5,627

Note: For the revised model (daily pooled) section of Table 7, the data shown is post-imputation and after omitting USEP values outside the \$50/MWh to \$4,000/MWh range.

## 5. Outliers

Instances where the actual daily average USEP exceeds the predicted USEP by more than 3 standard deviations are considered as outliers. Using this criterion, three outliers were detected. These occurred on October 12, 2021, and November 26, 2021, which could be attributed to the year-end global energy crunch. These outliers surpassed the upper bound of predicted USEP by \$35.80/MWh and \$371.13/MWh, respectively. The graph depicting the outliers is showcased in Figure 6.

Figure 6: Outliers



## 6. Conclusion

In conclusion, addressing several key shortcomings of the previous model has led to significant improvements in the model's accuracy and explanatory power. Firstly, transitioning to a more granular dataset at the periodic level allowed for a deeper understanding of underlying trends and relationships, as evidenced by a notable increase in the model's  $R^2$  value. Secondly, the expansion of the model to include additional variables further enhanced its explanatory capability, providing a more comprehensive analysis of the factors influencing the outcomes. Thirdly, by implementing various methods to handle unavailable data, such as interpolation techniques and imputation approaches, the risk of bias introduced by discarded observations was mitigated, resulting in more reliable results. Overall, these refinements collectively contributed to a more robust and insightful model compared to the previous iteration.

## 7. Reference

1. <https://www.home.emcsg.com/glossary>
2. <https://www.nems.emcsg.com/Guide-To-Prices>
3. <https://www.ema.gov.sg/content/dam/corporate/consultation-data/spinning-reserve-requirements-for-the-singapore-power-system/decision/EMA-Consultations-Final-Determination-Enhancements-to-Spinning-Reserves-Requirements-for-Power-System.pdf.coredownload.pdf>
4. <https://www.ema.gov.sg/regulations-licences/regulations/policies-frameworks/vesting-contracts>
5. <https://www.ema.gov.sg/resources/singapore-energy-statistics/chapter5>
6. <https://www.ema.gov.sg/partnerships/consultations/2023/consultation-on-temporary-price-cap>
7. Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: time series missing value imputation in R. *R J.*, 9(1), 207.

## 8. Appendix

### 8.1 Imputation Methods

#### 8.1.1 Linear Interpolation

Linear interpolation estimates unavailable values by assuming a linear relationship between consecutive observed data points. Specifically, given a set of observed values  $(x_1, y_1), (x_2, y_2)$ , the linear interpolation formula for estimating  $y$  at a point  $x$  between  $x_1$  and  $x_2$  is:

$$y = y_1 + (y_2 - y_1) \times \frac{x - x_1}{x_2 - x_1}$$

#### 8.1.2 Spline Interpolation

Spline interpolation is a mathematical technique used to construct a smooth curve (spline) that passes through a set of given data points. There are different types of spline interpolation, with cubic spline interpolation being one of the most popular. Here's a brief overview of cubic spline interpolation:

Suppose we have a set of data points  $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$  that we want to interpolate. Then, dividing the range of data into intervals, typically between adjacent data points. In each interval, use a cubic polynomial to interpolate the data. For the interval  $[x_i, x_{i+1}]$ , the cubic polynomial is expressed as:<sup>2</sup>

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

Solve for the coefficients  $a_i, b_i, c_i, d_i$  in each interval using the given conditions. This often involves setting up and solving a system of linear equations. Once the coefficients are determined for each interval, the overall cubic spline function is constructed by piecing together the individual cubic polynomials. Use the resulting cubic spline function to interpolate values within the given range.

The detailed example can refer to the Wikipedia<sup>3</sup>.

#### 8.1.3 Stineman Interpolation

Stineman interpolation is an enhancement of spline interpolation, particularly useful in handling noisy data by incorporating weighted residuals. The Stineman interpolation algorithm considers both the function value and the first derivative, assigning weights to minimise the influence of outliers. Instead of fitting a single polynomial through the entire dataset, Stineman interpolation fits a different polynomial in each subinterval between two adjacent data points.

Relative to the Spline one, Stineman interpolation uses a quadratic polynomial for each interval, rather than cubic. The main procedures and principles are quite similar to the Spline.

#### 8.1.4 Structural Model & Kalman Smoothing

This method involves fitting a structural model to the data and applying Kalman smoothing to estimate unavailable values, considering the underlying structure. Kalman smoothing combines current observations with prior estimates, adjusting the imputation based on the system dynamics and measurement errors.

Structural Model:

---

<sup>2</sup> Ensure that the spline is smooth by setting up conditions. The cubic spline satisfies the following conditions: 1)  $S_i(x_i) = y_i$ ; 2)  $S_i(x_{i+1}) = y_{i+1}$ ; 3)  $S'_i(x_{i+1}) = S'_{i+1}(x_{i+1})$ ; 4)  $S''_i(x_{i+1}) = S''_{i+1}(x_{i+1})$ .

<sup>3</sup> [https://en.wikipedia.org/wiki/Spline\\_interpolation](https://en.wikipedia.org/wiki/Spline_interpolation)

Let  $Y_t$  be the observed time series at time  $t$ .

The structural model can be represented as:  $Y_t = f(\theta_t) + \epsilon_t$

where  $f(\theta_t)$  is a function of parameters  $\theta_t$  representing the structural components, and  $\epsilon_t$  is the error term. The parameters  $\theta_t$  are estimated by maximising the likelihood function or minimising the sum of squared errors. For example, in an ARIMA model :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

The parameters  $\phi_1, \phi_2, \dots, \phi_p$  are estimated to capture the autoregressive components.

Kalman Smoothing:

State-Space Representation:

Represent the structural model in state-space form:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{w}_t \\ \mathbf{y}_t &= \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t\end{aligned}$$

where  $\mathbf{x}_t$  is the state vector,  $\mathbf{y}_t$  is the observation vector,  $\mathbf{F}_t$  and  $\mathbf{H}_t$  are state and observation transition matrices,  $\mathbf{B}_t$  is the input control matrix, and  $\mathbf{u}_t$  is the control vector.  $\mathbf{w}_t$  and  $\mathbf{v}_t$  are process and observation noise.

Kalman Filter Equations:

The Kalman filter updates the state estimate based on the observed data:

$$\begin{aligned}\hat{\mathbf{x}}_{t|t-1} &= \mathbf{F}_t \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_t \mathbf{u}_t \\ \mathbf{P}_{t|t-1} &= \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^T + \mathbf{Q}_t\end{aligned}$$

where  $\hat{\mathbf{x}}_{t|t-1}$  is the predicted state estimate,  $\mathbf{P}_{t|t-1}$  is the predicted state error covariance, and  $\mathbf{Q}_t$  is the process noise covariance.

Kalman Smoother Equations:

The Kalman smoother refines the filtered estimates using future observations:

$$\begin{aligned}\mathbf{G}_t &= \mathbf{P}_{t|t-1} \mathbf{F}_t^T (\mathbf{F}_t \mathbf{P}_{t|t-1} \mathbf{F}_t^T + \mathbf{Q}_t)^{-1} \\ \hat{\mathbf{x}}_{t|T} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{G}_t (\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t})\end{aligned}$$

where  $\mathbf{G}_t$  is the Kalman gain,  $\hat{\mathbf{x}}_{t|T}$  is the smoothed state estimate, and  $\hat{\mathbf{x}}_{t+1|T}$  is the state estimate at the next time step based on future data.

### 8.1.5 ARIMA State Space Representation & Kalman Smoothing

Utilising the ARIMA model and Kalman smoothing, this method imputes unavailable values in time series data by considering autoregressive and moving average components. ARIMA models capture temporal dependencies, and Kalman smoothing optimally estimates unavailable values by incorporating the evolving time series structure.

The mathematical principle is the same as the method above.

### **8.1.6 Last Observation Carried Forward (LOCF)**

LOCF imputes unavailable values by carrying forward the last observed value.

### **8.1.7 Next Observation Carried Backward (NOCB)**

NOCB imputes unavailable values by carrying backward the next observed value.

### **8.1.8 Mean Imputation**

Mean imputation replaces unavailable values with the mean of the observed values.

### **8.1.9 Median Imputation**

Median imputation replaces unavailable values with the median of the observed values.

### **8.1.10 Mode Imputation**

Mode imputation replaces unavailable values with the mode (most frequently occurring value) of the observed values.

Table A1: Separate Analysis Results (Group)

Reg No.	Observations	Variables	Coefficients	Lasso Select YN
		Vested Quantity (kWh)	-0.092*** (-13.294)	N
1	340,483	Vesting Price (\$/MWh)	0.005 (1.596)	Y
		Vesting Quantity (MW)	0.065*** (3.107)	Y
		Secondary Reserve Requirement (MW)	-0.303*** (-10.023)	Y
2	236,562	Secondary Reserve Price (\$/MWh)	0.007*** (9.159)	Y
		Secondary Reserve Availability (MW)	-0.037*** (-4.108)	N
		Gas Curtailment Magnitude=1	0.017*** (11.920)	Y
		Gas Curtailment Magnitude=2	0.002* (1.714)	N
		Gas Curtailment Magnitude=3	0.004** (2.393)	N
3	200,367	Gas Curtailment Magnitude=4	0.004 (0.942)	N
		Gas Curtailment Magnitude=5	-0.019*** (-6.913)	Y
		Gas Curtailment Magnitude=6	0.057*** (6.373)	N
		Gas Curtailment Magnitude=7	-0.023** (-2.304)	N
		Solar (MW)	-0.010 (-1.411)	N
4	18,391	Import (MW)	0.003** (2.246)	Y
		ESS (MW)	0.005*** (2.728)	Y
		MAPT (MWh)	211.504*** (6.118)	N
		Spot LRMC (MWh)	1.261 (1.370)	N
		Term LRMC (MWh)	3.312 (0.523)	N
		TPC Applied YN	-0.030** (-2.093)	Y
5	8,828	TPC Primary Reserve (\$/MWh)	-	N
		TPC Contingency Reserve (\$/MWh)	-0.197 (-1.175)	N
		TPC Regulation (\$/MWh)	-217.957*** (-6.094)	N
		Multiplier	1.585 (0.368)	Y

		LCP (MWh)	-0.012***	
6	2,030		(-4.217)	Y
		TCL (MW)	-0.006	
			(-0.473)	N
7	1,757	EMA Directions (MW)	0.005	
			(0.583)	N

Note: Table A1 presents regression results using data post-imputation and after omitting USEP values outside the \$50/MWh to \$4,000/MWh range.

Table A2: Complete Results from Separate Analysis (1 by 1)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
	TPC	Vested	Vesting	Vesting	Secondary	Secondary	Secondary	GC	Solar	Import	ESS	MAPT	Spot	Term	Multiplier	TPC	TPC	TPC	LCP	TCL	EMA
	Applied	Quantity	Price	Quantity	Reserve	Reserve	availability	Magnitude					LRMC	LRMC	Primary	Contingency	Regulations			directions	
One Time Period Lag USEP (MWh)	0.621***	0.766***	0.769***	0.766***	0.667***	0.666***	0.672***	0.779***	0.791***	0.612***	0.636***	0.621***	0.621***	0.621***	0.621***	0.620***	0.621***	0.552***	0.555***	0.446***	
	(22.005)	(157.020)	(158.584)	(157.452)	(73.668)	(73.411)	(74.449)	(138.720)	(165.598)	(44.234)	(44.105)	(21.962)	(21.961)	(21.975)	(21.975)	(21.962)	(21.955)	(21.962)	(22.268)	(22.357)	(9.895)
System Demand (MW)	-2.742***	0.224***	0.036*	0.181***	0.223***	0.222***	0.216***	-0.042	0.731***	1.085***	1.297***	-2.767***	-2.765***	-2.848***	-2.848***	-2.767***	-2.757***	-2.766***	6.455***	6.839***	4.018***
	(-4.231)	(10.483)	(1.652)	(8.377)	(10.335)	(10.281)	(9.299)	(-1.301)	(13.666)	(7.680)	(6.800)	(-4.175)	(-4.165)	(-4.385)	(-4.385)	(-4.175)	(-4.200)	(-4.173)	(5.979)	(6.161)	(3.623)
Primary Reserve Requirement (MW)	-0.280	-0.019***	-0.027***	-0.021***	-0.238***	-0.106***	-0.071***	-0.033***	-0.023***	0.290**	-0.275	-0.249	-0.251	-0.271	-0.271	-0.249	-0.236	-0.249	-1.121	-1.073	-0.504
	(-0.565)	(-6.592)	(-9.234)	(-7.358)	(-17.364)	(-11.023)	(-6.477)	(-8.689)	(-4.963)	(2.299)	(-0.755)	(-0.500)	(-0.503)	(-0.548)	(-0.548)	(-0.500)	(-0.474)	(-0.500)	(-1.352)	(-1.291)	(-0.726)
Primary Reserve Availability (MW)	0.101**	0.003	-0.000	0.002	-0.013***	-0.010***	0.049***	0.063***	0.070***	0.070***	0.092***	0.109***	0.109***	0.103**	0.103**	0.109***	0.116***	0.110***	0.120	0.125	0.167
Contingency Reserve Requirement (MW)	(2.476)	(0.453)	(-0.043)	(0.357)	(-4.609)	(-3.407)	(5.406)	(10.192)	(8.910)	(4.187)	(3.507)	(2.668)	(2.662)	(2.515)	(2.515)	(2.668)	(2.823)	(2.670)	(1.062)	(1.096)	(1.600)
	-1.018	-0.021***	-0.006	-0.021***	-0.503***	-0.067***	-0.005	-0.060***	-0.096***	0.565*	-1.280	-0.912	-0.916	-0.993	-0.993	-0.912	-0.873	-0.911	-2.109	-2.017	-0.793
Contingency Reserve Price (\$/MWh)	(-0.832)	(-2.865)	(-0.855)	(-2.944)	(-13.984)	(-3.358)	(-0.185)	(-7.181)	(-8.483)	(1.737)	(-1.427)	(-0.741)	(-0.744)	(-0.811)	(-0.811)	(-0.741)	(-0.711)	(-0.740)	(-0.830)	(-0.789)	(-0.389)
	0.014***	0.024***	0.024***	0.024***	0.033***	0.032***	0.032***	0.020***	0.021***	0.019***	0.019***	0.014***	0.014***	0.014***	0.014***	0.014***	0.014***	0.014***	-0.050***	-0.050***	0.036***
Contingency Reserve Availability (MW)	(6.434)	(52.721)	(53.752)	(52.727)	(46.251)	(45.740)	(46.189)	(37.044)	(31.653)	(16.876)	(12.859)	(6.573)	(6.565)	(6.520)	(6.520)	(6.573)	(6.671)	(6.574)	(-4.033)	(-4.073)	(3.444)
	0.285***	0.173***	0.163***	0.174***	0.120***	0.114***	0.133***	0.255***	0.279***	0.284***	0.303***	0.287***	0.287***	0.284***	0.284***	0.287***	0.291***	0.287***	0.137	0.132	0.608***
Total Reserve Requirement (MW)	(4.962)	(26.556)	(26.114)	(26.541)	(20.595)	(19.736)	(17.515)	(31.000)	(26.285)	(12.052)	(8.338)	(5.028)	(5.030)	(4.942)	(4.942)	(5.028)	(5.103)	(5.029)	(0.704)	(0.675)	(3.771)
	0.915	-0.111***	-0.105***	-0.109***	0.863***	0.010	-0.103***	-0.122***	-0.137***	-1.212***	0.941	0.784	0.790	0.888	0.888	0.783	0.729	0.782	3.665	3.509	0.246
Regulation Requirement (MW)	(0.534)	(-24.261)	(-22.662)	(-23.678)	(13.196)	(0.346)	(-2.978)	(-21.588)	(-21.459)	(-2.705)	(0.753)	(0.454)	(0.458)	(0.518)	(0.518)	(0.454)	(0.424)	(0.453)	(1.077)	(1.026)	(0.091)
	-0.512**	0.009***	0.021***	0.010***	0.024***	0.025***	0.026***	-0.001	-0.009*	0.047	0.135**	-0.531**	-0.531**	-0.528**	-0.528**	-0.531**	-0.531**	-0.531**	-0.121	-0.135	-0.079
Regulation Price (\$/MWh)	(-2.224)	(4.722)	(10.281)	(4.793)	(11.280)	(12.072)	(11.831)	(-0.184)	(-1.905)	(1.015)	(2.379)	(-2.302)	(-2.301)	(-2.284)	(-2.284)	(-2.302)	(-2.316)	(-2.302)	(-0.412)	(-0.458)	(-0.345)
	0.015**	0.011***	0.011***	0.011***	0.006***	0.006***	0.006***	0.016***	0.018***	0.018***	0.019***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***	0.123***	0.125***	0.033
Injection Energy Quantity (MW)	(2.510)	(36.910)	(35.970)	(37.311)	(23.412)	(22.146)	(22.961)	(31.028)	(27.334)	(8.565)	(6.797)	(2.743)	(2.740)	(2.740)	(2.740)	(2.743)	(2.741)	(2.743)	(6.773)	(6.859)	(1.508)
	-2.439***	-0.546***	-0.530***	-0.534***	-0.355***	-0.346***	-0.384***	-0.449***	-1.414***	-2.721***	-3.568***	-2.432***	-2.428***	-2.435***	-2.435***	-2.432***	-2.462***	-2.432***	-8.730***	-8.778***	-3.439**
Combined-Cycle Gas Turbine (CCGT) Generation (MW)	(-4.513)	(-13.960)	(-13.498)	(-13.641)	(-9.479)	(-9.274)	(-10.410)	(-6.706)	(-13.464)	(-11.012)	(-10.080)	(-4.502)	(-4.499)	(-4.508)	(-4.508)	(-4.502)	(-4.537)	(-4.502)	(-3.748)	(-3.731)	(-2.258)
	3.535***	-0.067***	-0.054***	-0.069***	-0.095***	-0.082***	-0.075***	-0.120***	-0.378***	-0.414***	-0.431***	3.536***	3.530***	3.617***	3.617***	3.536***	3.548***	3.534***	-3.003***	-3.132***	-3.136***



Magnitude = 1		
Gas Curtailment	(11.920)	0.002*
Magnitude = 2	(1.714)	0.004**
Gas Curtailment	(2.393)	0.004
Magnitude = 3	(0.942)	-0.019***
Gas Curtailment	(-6.913)	0.057***
Magnitude = 5	(6.373)	-0.023**
Gas Curtailment	(-2.304)	0.006***
Magnitude = 7	(9.001)	0.004***
Solar (MW)	(5.617)	0.005***
Import (MW)	(2.812)	-0.086
Energy Storage Systems (ESS) Generation (MW)	(-1.156)	(-1.111)
MAPT (\$/MWh)	6.799***	6.876***
Spot LRMC (MWh)	6.799***	6.876***
Term LRMC (MWh)	(2.617)	(2.618)
Multiplier	-0.086	-0.086
TPC Primary Reserve (\$/MWh)	(-1.157)	-0.111**
TPC Contingency Reserve (\$/MWh)	(-2.042)	

TPC Regulation (\$/MWh)																	-0.089				
																	(-1.171)				
Load Curtailment Price (\$/MWh)																	-0.012***				
																	(-4.666)				
Total Curtailed Load (MW)																	-0.022*				
																	(-1.953)				
EMA directions (MW)																	0.005				
																	(0.583)				
Constant	75.601**	4.047***	4.295***	3.811***	4.834***	5.593***	5.714***	3.932***	3.539***	47.752***	-88.111**	38.090	-6.621	1,113.573	***	2,865.737		0.005			
	(2.149)	(32.010)	(33.937)	(30.105)	(31.828)	(35.692)	(34.076)	(21.041)	(14.680)	(9.050)	(-2.135)	(1.038)	(-0.102)	(2.988)	(-2.498)	(1.038)	(0.307)	(1.022)	(-2.473)	(-2.449)	(-1.547)
Observations	8,828	340,483	340,483	340,483	250,389	250,389	236,562	200,367	143,422	25,800	19,295	8,828	8,828	8,828	8,828	8,828	8,828	8,828	2,030	2,030	1,757
R-squared	0.868	0.940	0.940	0.940	0.942	0.942	0.942	0.946	0.946	0.918	0.915	0.868	0.868	0.868	0.868	0.868	0.868	0.868	0.784	0.782	0.844
adjusted_R2	0.867	0.940	0.940	0.940	0.942	0.942	0.942	0.946	0.946	0.918	0.915	0.866	0.866	0.866	0.866	0.866	0.866	0.866	0.775	0.773	0.836

Table A3: Pairwise Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)																						
(1) USEPMWh	1.000																																									
(2) usepmwh_lag		<b>0.961***</b> (0.000)	1.000																																							
(3) SystemDemandMW			0.261*** (0.000)	0.257*** (0.000)	1.000																																					
(4) PriResReqMW				-0.217*** (0.000)	-0.215*** (0.000)	-0.562*** (0.000)	1.000																																			
(5) Pri_AvailMW					-0.250*** (0.000)	-0.244*** (0.000)	0.341*** (0.000)	-0.466*** (0.000)	1.000																																	
(6) ConResReqMW						-0.057*** (0.000)	-0.058*** (0.000)	0.398*** (0.000)	-0.028*** (0.000)	0.064*** (0.000)	1.000																															
(7) ConResPriceMWh							0.377*** (0.000)	0.349*** (0.000)	0.384*** (0.000)	-0.113*** (0.000)	-0.116*** (0.000)	0.293*** (0.000)	1.000																													
(8) Con_AvailMW								-0.423*** (0.000)	-0.415*** (0.000)	-0.223*** (0.000)	-0.073*** (0.000)	0.610*** (0.000)	-0.021*** (0.000)	-0.435*** (0.000)	1.000																											
(9) TotalReqMW									-0.134*** (0.000)	-0.134*** (0.000)	-0.354*** (0.000)	0.610*** (0.000)	-0.422*** (0.000)	0.476*** (0.000)	-0.057*** (0.000)	0.033*** (0.000)	1.000																									
(10) RegReqMW										-0.087*** (0.000)	-0.095*** (0.000)	0.495*** (0.000)	-0.067*** (0.000)	0.217*** (0.000)	0.233*** (0.000)	0.259*** (0.000)	-0.144*** (0.000)	-0.187*** (0.000)	1.000																							
(11) RegPriceMWh											0.456*** (0.000)	0.437*** (0.000)	-0.098*** (0.000)	0.166*** (0.000)	-0.428*** (0.000)	-0.046*** (0.000)	0.308*** (0.000)	-0.373*** (0.000)	0.193*** (0.000)	0.070*** (0.000)	1.000																					
(12) IEQMW												0.263*** (0.000)	0.262*** (0.000)	<b>0.997***</b> (0.000)	-0.566*** (0.000)	0.346*** (0.000)	0.392*** (0.000)	0.387*** (0.000)	-0.224*** (0.000)	-0.365*** (0.000)	0.494*** (0.000)	-0.099*** (0.000)	1.000																			
(13) CCGTGenMW													0.183*** (0.000)	0.181*** (0.000)	<b>0.910***</b> (0.000)	-0.655*** (0.000)	0.459*** (0.000)	0.393*** (0.000)	0.352*** (0.000)	-0.152*** (0.000)	-0.443*** (0.000)	0.456*** (0.000)	-0.185*** (0.000)	<b>0.914***</b> (0.000)	1.000																	
(14) STGenMW														0.143*** (0.000)	0.142*** (0.000)	-0.590*** (0.000)	0.518*** (0.000)	-0.597*** (0.000)	-0.229*** (0.000)	-0.212*** (0.000)	-0.140*** (0.000)	0.481*** (0.000)	-0.436*** (0.000)	0.356*** (0.000)	-0.599*** (0.000)	-0.789*** (0.000)	1.000															
(15) OTGenMW															0.208*** (0.000)	0.207*** (0.000)	0.320*** (0.000)	-0.383*** (0.000)	0.220*** (0.000)	0.072*** (0.000)	0.092*** (0.000)	0.011*** (0.000)	-0.179*** (0.000)	0.074*** (0.000)	0.000	0.322*** (0.875)	0.338*** (0.000)	-0.290*** (0.000)	1.000													
(16) WEQMW																0.261*** (0.000)	0.259*** (0.000)	<b>0.997***</b> (0.000)	-0.563*** (0.000)	0.345*** (0.000)	0.391*** (0.000)	0.387*** (0.000)	-0.225*** (0.000)	-0.364*** (0.000)	0.497*** (0.000)	-0.100*** (0.000)	<b>0.999***</b> (0.000)	<b>0.913***</b> (0.000)	-0.600*** (0.000)	0.319*** (0.000)	1.000											
(17) OffersPricedBelow200																	-0.131*** (0.000)	-0.129*** (0.000)	0.060*** (0.000)	0.283*** (0.000)	-0.203*** (0.000)	-0.255*** (0.000)	0.059*** (0.000)	-0.339*** (0.000)	-0.227*** (0.000)	0.310*** (0.000)	0.048*** (0.000)	0.056*** (0.000)	-0.095*** (0.000)	0.152*** (0.000)	-0.229*** (0.000)	0.059*** (0.000)	1.000									
(18) ResCushion																	-0.228*** (0.000)	-0.218*** (0.000)	0.048*** (0.000)	-0.465*** (0.000)	0.757*** (0.000)	-0.296*** (0.000)	-0.371*** (0.000)	0.760*** (0.000)	-0.443*** (0.000)	-0.045*** (0.000)	-0.375*** (0.000)	0.052*** (0.000)	0.164*** (0.000)	-0.364*** (0.000)	0.184*** (0.000)	0.050*** (0.000)	-0.260*** (0.000)	1.000								
(19) supplycushion																	-0.528*** (0.000)	-0.518*** (0.000)	-0.578*** (0.000)	0.114*** (0.000)	0.219*** (0.000)	0.068*** (0.000)	-0.411*** (0.000)	0.675*** (0.000)	0.325*** (0.000)	-0.329*** (0.000)	-0.253*** (0.000)	-0.577*** (0.000)	-0.394*** (0.000)	0.009*** (0.000)	-0.065*** (0.000)	-0.577*** (0.000)	-0.557*** (0.000)	0.370*** (0.000)	1.000							
(20) fueloilpriceusmt_lag																	0.596*** (0.000)	0.594*** (0.000)	0.373*** (0.000)	-0.540*** (0.000)	0.139*** (0.000)	0.178*** (0.000)	0.150*** (0.000)	-0.119*** (0.000)	-0.188*** (0.000)	-0.130*** (0.000)	0.153*** (0.000)	0.373*** (0.000)	0.472*** (0.000)	-0.217*** (0.000)	0.342*** (0.000)	0.369*** (0.000)	-0.452*** (0.000)	0.124*** (0.000)	-0.120*** (0.000)	1.000						

