Model Documentation

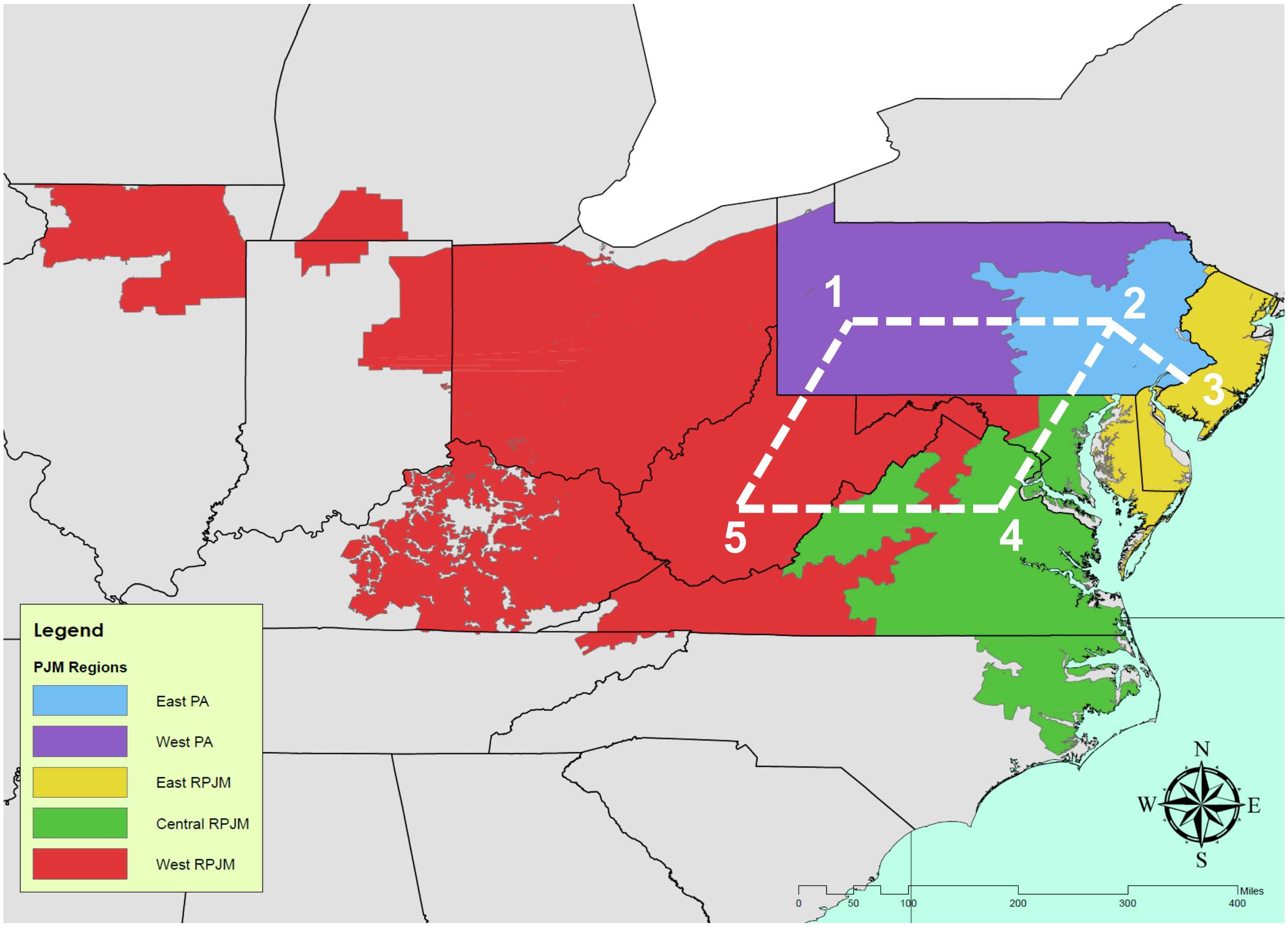
PJM Electricity Model for Policy Analysis (PEMPA)

Section 1 provides a detailed overview of the simulation model. Section 2 presents the functional forms used in the simulation model. Section 3 discusses the parameter values and data sources used for calibration. Section 4 outlines the assumptions and data sources used to generate emissions estimates. In Section 5 we outline the assumptions regarding the dynamic trends that underlie our simulation results. Pre-existing polices that we account for in our analysis are described in Section 6. Finally, section 7 validates our baseline against historical data.

1 Model Overview

The PJM Electricity Model for Policy Analysis (PEMPA) is a multi-market numerical simulation model that combines: 1. a direct current optimal power flow (DCOPF) model of the PJM power system; 2. the endogenous supply of new generation capacity within PJM; and 3. the importation of Renewable Energy Credits (RECs) from outside of PJM. The model is calibrated using data for 2016 and 2017 collected from over a dozen sources and is validated using 2018 data across several dimensions: REC prices, locational marginal prices (LMPs), predicted new capacity, and generation mix. PEMPA operates on an annual time-step and we simulate outcomes from 2018 to 2030. The PEMPA domain is depicted in Figure A.1.

Figure A.1: PJM Electricity Model for Policy Analysis Model (PEMPA) Domain



As shown in Figure A.1, PEMPA consists of five regions (purple, blue, yellow, green, and red) which comprise the wholesale electricity market operated by PJM.

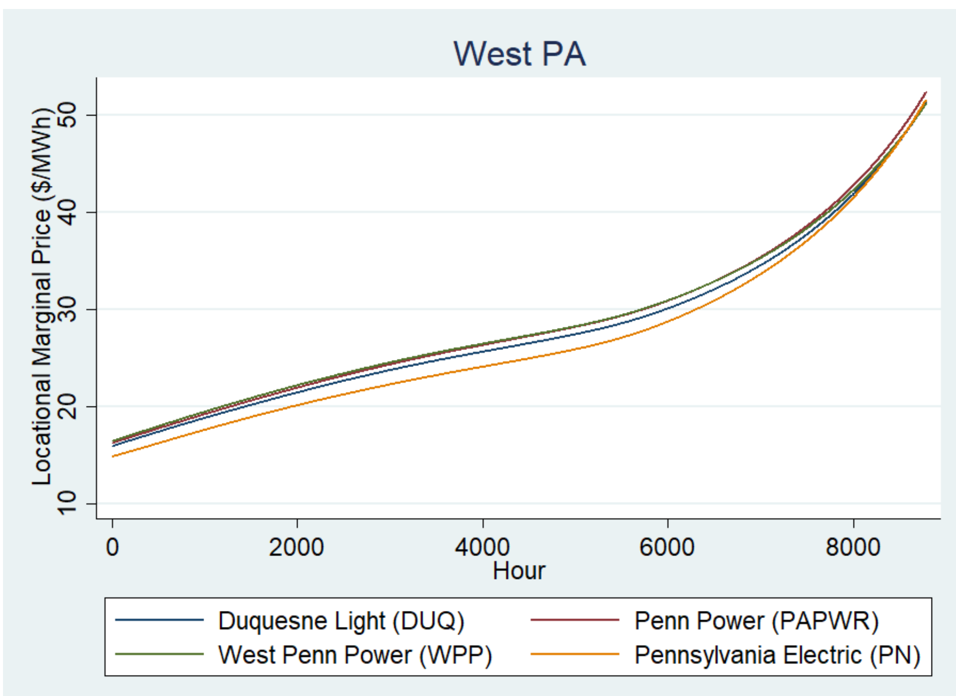
The five regions that comprise PJM in PEMPA are constructed by aggregating load zones in PJM based on similarity across in real-time hourly locational marginal prices (LMPs)[1] and geographical proximity in Pennsylvania and Rest of PJM (RPJM). The five regions are: 1. West Pennsylvania (purple), which includes West Penn Power (the Pennsylvania part of Allegheny Power), Penn Power (the Pennsylvania part of American Transmission Systems Inc), Duquesne Light Company, and the Pennsylvania Electric Company; 2. East Pennsylvania (blue), which includes the Metropolitan Edison Company, PPL Electric Utilities Corporation, and the PECO Energy Company; 3. East RPJM

[1]The five figures below show real-time hourly LMPs for these load zones based upon the PEMPA PJM regions they are grouped into, given a curve smoothing bandwidth of 0.8, and after sorting hours in the year from lowest to highest load.

(yellow), which includes Atlantic City Electric Company, Jersey Central Power and Light Company, Public Service Electric and Gas Company, Delmarva Power and Light Company and Rockland Electric Company; 4. Central RPJM (green), which includes Baltimore Gas and Electric, Dominion, and Potomac Electric; and, 5. West RPJM (red), which includes the non-Pennsylvania part of Allegheny Power, American Electric Power Company, the Ohio part of American Transmission Systems Inc, the Commonwealth Edison Company, Duke Energy Ohio and Kentucky, East Kentucky Power Cooperative Inc, and the Dayton Power and Light Company.

Figure A.2 shows the sorted real-time LMPs in 2016 for the four load zones that make up the West Pennsylvania region in PEMPA. The LMP in West Pennsylvania in the lowest demand hour is as low as approximately \$ 15/MWh and the LMP in the highest demand hour is more than \$50/MWh. The median LMP is around \$27/MWh.

Figure A.2: Real-time LMPs in West Pennsylvania in 2016



Similarly, Figure A.3 provides the sorted real-time LMPs of the three load zones that make up East Pennsylvania. The LMP in the lowest demand hour is close to \$ 5 per MWh and LMP in the highest demand hour is over \$ 50. The median LMP is approximately \$ 22. Overall, West Pennsylvania observes higher LMPs than East Pennsylvania in 2016, although there is some convergence in LMPs during high demand periods.

Figure A.3: Real-time LMPs in East Pennsylvania in 2016

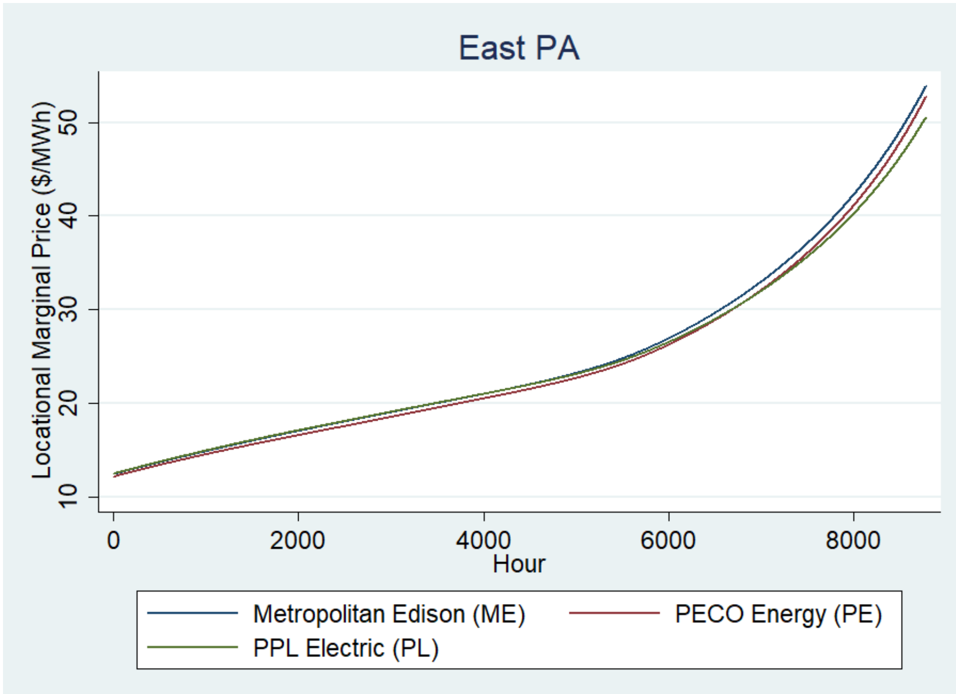
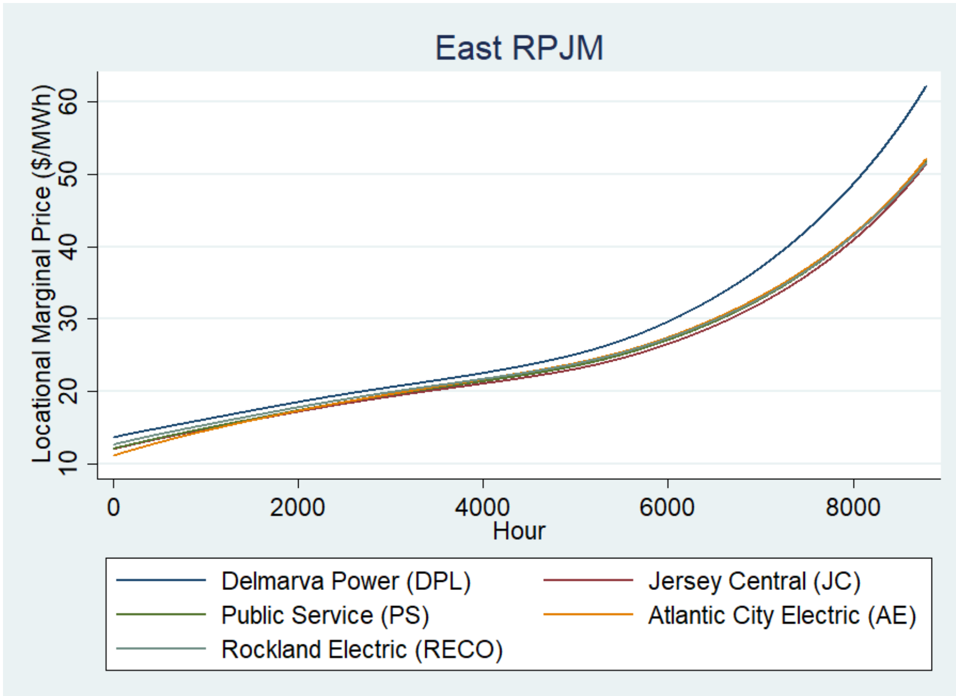


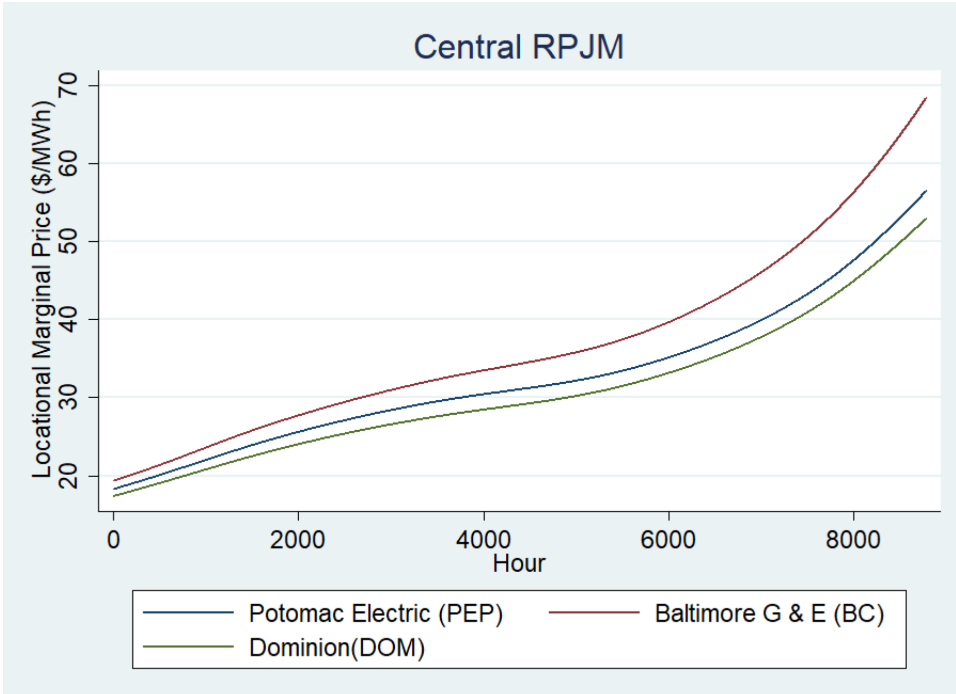
Figure A.4 present the sorted real-time LMPs of the five load zones which constitute East RPJM. LMPs in the lowest demand hour is as low as $ 11 per MWh and LMPs in the highest demand hour is more than $ 60 for the DPL zone and more than $ 50 for the other zones. The median LMP is around $ 22.

Figure A.4: Real-time LMPs in East RPJM in 2016



Similarly, Figure A.5 shows sorted real-time LMPs for the three load zones which comprise Central RPJM. LMPs in this region are the most spread out with a LMP in the low demand hour of approximately $18 per MWh and a LMP near $ 69 in the highest demand hour. The median LMP lies just above $ 30.

Figure A.5: Real-time LMPs in Central RPJM in 2016



Finally, Figure A.6 shows sorted real-time LMPs of the seven load zones which make up West RPJM. The LMP in the lowest demand hour is nearly $ 15 per MWh. The LMPs in the 2,000 highest demand hours in West RPJM are fairly spread out; at the highest demand hour the LMP is approximately $ 55 in Ohio Edison and $ 43 in East Kentucky. The median LMP is around $ 27. Overall, Central RPJM observes higher LMPs than West RPJM, which in turns observes higher LMPs than East RPJM in 2016.

Figure A.6: Real-time LMPs in West RPJM in 2016

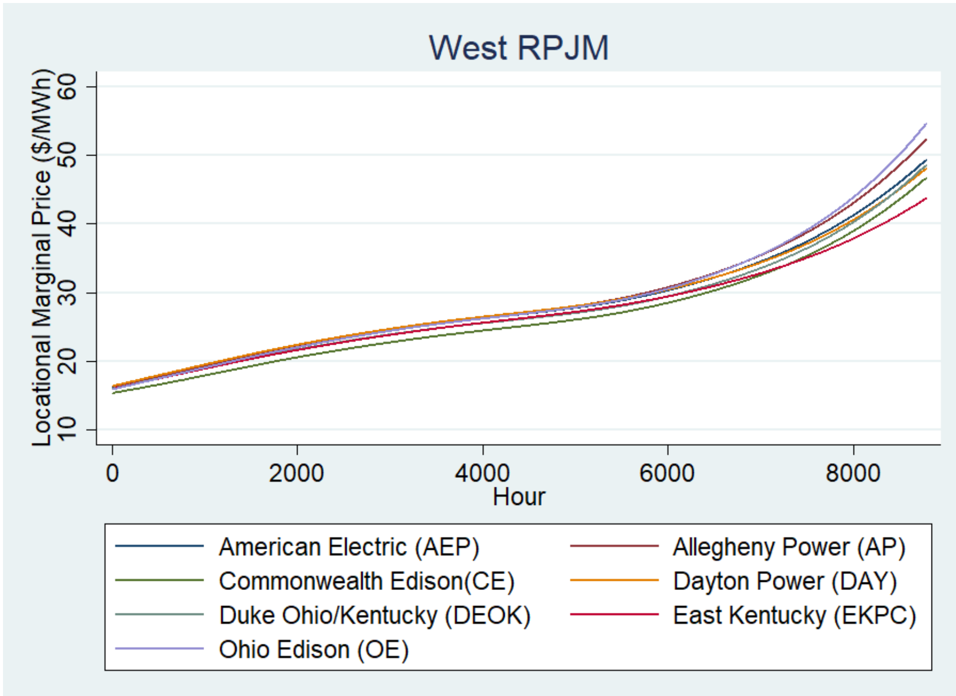
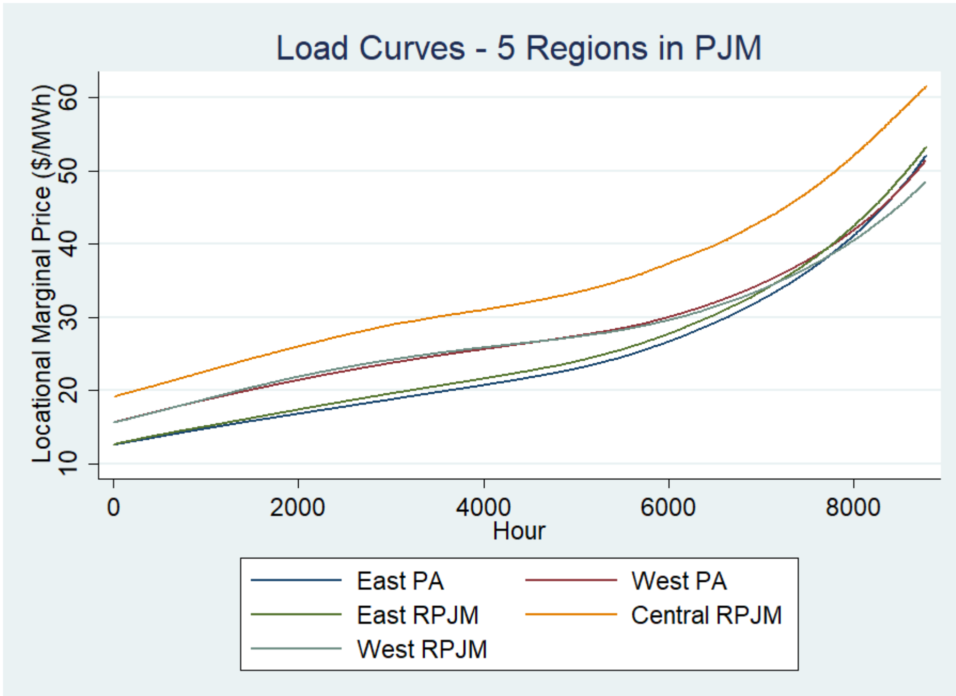


Figure A.7 summarizes the average real-time LMPs for the five PJM regions in PEMPA, sorted from the lowest demand hour to the highest demand hour. Central RPJM has higher average LMPs in 2016 than all other regions. West Pennsylvania and West RPJM have the next highest average LMPs in low to moderate demand hours, followed by East Pennsylvania and East RPJM. While there appears to be some convergence—east and west–between Pennsylvania and RPJM at low and moderate demand hours, this breaks down in high demand hours. In higher demand periods West Pennsylvania observes higher average LMPs than West RPJM whereas East RPJM observes higher average LMPs than East Pennsylvania. In totum, in high demand hours, excluding Central RPJM, East RPJM has higher average LMPs, followed by East Pennsylvania, West Pennsylvania, and West RPJM. In general, the eastern part of PJM observes lower LMPs and overall is an exporter of electricity compared to the central and western parts of PJM.

Figure A.7: Average Real-time LMPs Across Five PJM Regions in 2016



The resulting five regions which comprise PJM in PEMPA are shown in Figure A.1. The black dashed lines depict the five aggregate transmission lines which link these five regions: A. line 12 connects West Pennsylvania and East Pennsylvania; B. line 23 connects East Pennsylvania and East RPJM; C. line 24 connects East Pennsylvania and Central RPJM; D. line 45; and, E. line 15 connects West Pennsylvania and West RPJM. These five lines are constructed based upon visual inspection of the transmission linkages between the five regions. There are no direct links between East Pennsylvania and West RPJM, West Pennsylvania and East RPJM, West Pennsylvania and Central RPJM, East RPJM to Central RPJM, or East RPJM to West RPJM. While these lines capture the aggregate physical location of transmission lines across PJM, the transmission constraints are numerically calibrated to replicate the generation-weighted average inter-regional difference in LMPs; see Section **??** for additional details.

Within each PJM region, we assume that demand for electricity is partially inelastic across 96 load segments, or 24 load segments for each of four seasons. The supply side considers the economic decisions of 843 representative existing generation units (EGUs) which have been aggregated by state, fuel type, and technology type from a population of 3,095 EGUs located within PJM and whose location is known. Fuel costs facing generators vary across region and are allowed to vary daily reflecting correlation with load. Given predicted generation from representative EGUs, generation and emissions can be descaled to the full population of EGUs across the landscape. PEMPA also allows for endogenous new capacity expansion in natural gas combined cycle (NGCC), solar, and wind.

2 Functional Forms

2.1 Regional Demand for Electricity

The regional direct economic benefit from consuming electricity in region i and load segment l is given by:

$$DB_{ilj}\left(d_{il}\right) = c_{il}d_{il} - \frac{n_{il}}{2}d_{il}^{2}, \tag{1}$$

where n_{il} and c_{il} are the slope and intercept of the linear inverse private demand curve implied by the first-order conditions for the numerical model with respect to d_{il}: $p_{il} = c_{il} - n_{il}d_{il}$.

2.2 Supply of Electricity from Existing Generation

The direct economic costs from representative electric generating unit (EGU) j located in state s and supplying electricity in load segment l is given by:

$$DC_{slj}\left(g_{slj}^{E}\right) = b_{slj}^{E}g_{slj}^{E} + \frac{m_{slj}^{E}}{2}\left(g_{slj}^{E}\right)^{2}, \tag{2}$$

where m_{slj}^{E} and b_{slj}^{E} are the slope and intercept of the linear inverse private supply curve for existing EGU j in state s at load segment l $\left(g_{slj}^{E}\right)$ implied by the first-order conditions for the numerical model with respect to g_{slj}^{E}: $p_{sl} = b_{slj}^{E} - m_{slj}^{E}g_{slj}^{E}$.

In addition, each representative existing EGU has a limit on the amount of power they can supply in load segment l reflecting their capacity constraint:

$$g^E_{slj} \leq \bar{g}^E_{slj}, \tag{3}$$

where $\bar{g}^E_{slj}$ is the representative EGU's available capacity given its published capacity constraint after accounting for its annual utilization factor.

2.3 Supply of Electricity from New Generation

The direct economic costs from a new EGU j located in state s and supplying electricity in load segment l is given by:

$$DC^N_{slj}\left(g^N_{slj}\right) = b^N_{slj} g^N_{slj} \tag{4}$$

where b^N_{slj} is the private exogenous total marginal cost for a new EGU j in state s at load segment l for providing g^N_{slj} MWh of new generation.

Additionally. each new EGU j has a limit in total generation it can produce in every load segment, reflecting its capacity constraint:

$$g^N_{slj} \leq \bar{g}^N_{slj}$$

where $\bar{g}^N_{slj}$ is the maximum capacity that new EGU j in state s can produce at load segment l. Note that $\bar{g}^N_{slj} = (\gamma_{slj})K^N_s$, in which γ_{slj} is the utilization factor for new EGU j in state s at load segment l that varies across different technologies.

2.4 External RECs

The total external RECs provided by an eligible EGU j outside of PJM to state s within PJM on an annual basis must not exceed the number of RECs that that EGU has for that year:

$$g_j^R = \sum_s g_{js}^R \leq \bar{g}_j^R$$

where g_{js}^R is the total annual amount of RECs provided by EGU j outside of PJM to a PJM state s, g_j^R is the total number of RECs that external EGU j provides to all states in PJM and $\bar{g}_j^R$ is the total amount of RECs that said EGU j has in possession for that year times the percentage of said total RECs that is eligible to provide RECs to PJM as mentioned later in 3.4

2.5 Transmission Network

Net power flow into region i from another region h at any given load segment l must not exceed the capacity constraint of the transmission line between the two regions at that load segment:

$$-\bar{f}_{ihl} \leq f_{ihl} \leq \bar{f}_{ihl}$$

where f_{ihl} is the net flow of electricity into region i from region h at load segment l and $\bar{f}_{ihl}$ is the maximum transmission capacity between region i and h at load segment l. The sign indicates the direction of the power flow, with negative (-) sign meaning power is flowing away from i and positive sign (+) meaning power is flowing into i.

2.6 Market Operator Solution

2.6.1 Market Clearing Condition

The electricity market is cleared in region i at load segment l when the demand at region i at load segment l is met, i.e. the total sum of existing and new generation and net flow at region i and load segment l is at least equal to demand plus transmission and distributional loss and virtual bids:

$$\sum_{s\in i}\sum_{j} g^{E}_{slj} + \sum_{s\in i}\sum_{j} g^{N}_{slj} + \sum_{h} f_{ihl} \geq d_{il}(1+\epsilon_l) \quad \forall i, l$$

where ϵ_l is the distribution and transmission loss and virtual bids as percentage of of demand at load segment l.

2.6.2 Pre-existing Policies

Renewable Portfolio Standards (RPS) Total RPS-eligible generation by RPS category (Tier 1, tier 2, SRPS) in state s plus total external RECs imported by state s from EGUs outside of PJM must be at least equal to a set percentage (state's set RPS) of the state's total generation in a year:

$$\frac{\sum_l \sum_{j\in q} \left(g^{E}_{slj} + g^{N}_{slj} + g^{R}_{js}\right)}{\sum_l \sum_j \left(g^{E}_{slj} + g^{N}_{slj}\right)} \geq \bar{r}_{sq}$$

where q indicates RPS categories (Tier 1, tier 2, SRPS), $j \in q$ indicates the EGUs that are eligible to provide generation to comply with RPS standards by tier 1, tier 2 or SRPS, g^{E}_{slj} is existing generation from existing EGU j in state s in load segment l, g^{N}_{slj} is new generation from new EGU j in state s in load segment l, g^{R}_{js} is the number of RECs provided by external EGU j to state s in that year, and $\bar{r}_{sq}$ is the RPS target in state s of RPS categories $q = \text{tier 1, tier 2, solar RPS}$.

Other Policies Title IV of the Clean Air Act (imposing a cost on SO_2 and NOx) and nuclear subsidies are directly added into the total marginal costs of all eligible EGUs.

2.6.3 Central Operator's Objective Function

The central operator maximizes total social welfare in PJM:

$$\sum_i \sum_s \sum_l \delta_l \left[\left(-\frac{n_{il}}{2} d_{il}^2 + c_{il} d_{il}\right) - \sum_j \left(\frac{m^{E}_{slj}}{2}\left(g^{E}_{slj}\right)^2 + b^{E}_{slj} g^{E}_{slj}\right) - \left(C_{sj} K_{sj} + b^{N}_{slj} g^{N}_{slj}\right) \right]$$

A numerical solution to the *baseline* in year y involves maximizing 2.6.3 choosing variables $g^E_{slj}, g^N_{slj}, g^R_{sj}, f_{ihl}, K_{sj}$, given constraints 2.2, 2.3, 2.5, 2.4, and 2.6.2.

3 Data and Calibration

3.1 Demand

For demand side calibration, we use hourly metered load data for all load zones from PJM. We also divide 8784 hours in year 2016 into 96 load segments using three load segment cuts. The first cut is based on seasons: Winter (December 20 to March 21), Spring (March 22 to June 20), Summer (June 21 to September 20) and Fall (September 21 to December 21). We also divide each season into 6 smaller segments based on the descending order of load, following the rule used by the EIA's Integrated Planning Model (IPM) v.5.13: The first segment is the hours with 1% highest loads, the second segment is the next 4%, the third segment is the next 10%, the fourth segment is the next 30%, the fifth segment is the next 30% and finally, the sixth segment is the lowest 25%. After doing this, we now have 24 load segments, based on seasons and loads. Next, we divide each of these 24 load segments into 4 smaller segments based on the descending order of gas price in each segment, with the highest gas-price based load segment is the highest 10% gas price, the second load segment is the next 20% gas price, the third load segment is the next 30% gas price and the final gas-price based load segment in each 24 segment is the lowest 40% gas price. In total, we categorize all 8784 hours in 2016 into $L = 96$ load segments. We do perform the similar processof for 2017 load and LMP data but instead of 8,784 hours, we only have 8,760 hours that make up of the 96 load segments.

Assuming demand elasticity (η^D) of -0.05 (Bushnell et al., 2017), we used the observed load (L_{il}) and LMP in each load segment and PJM region to establish the slopes and intercepts of the quadratic demand functions for five established PJM regions:

$$\text{Demand curve intercepts } c_{il} = \left(\frac{1}{\eta^D}\right) \times \left(\frac{LMP_{il}}{L_{il}}\right)$$
$$\text{Demand curve slopes } n_{il} = \left(1 + \frac{1}{\eta^D}\right) \times LMP_{il}$$

Details on the functional forms and model formulations are detailed in Appendix 2. The functional form assumed for demand is not consequential as our 2016 and 2017 models both predict demand matches actual loads with a high R^2 of 0.998.

The next four figures below show the load curves in each season in 2016. As expected, summer has the highest LMPs, followed by winter. Spring and fall are shoulder months and thus have lower LMPs.

Figure A.8: RT LMPs in PJM in Winter 2016

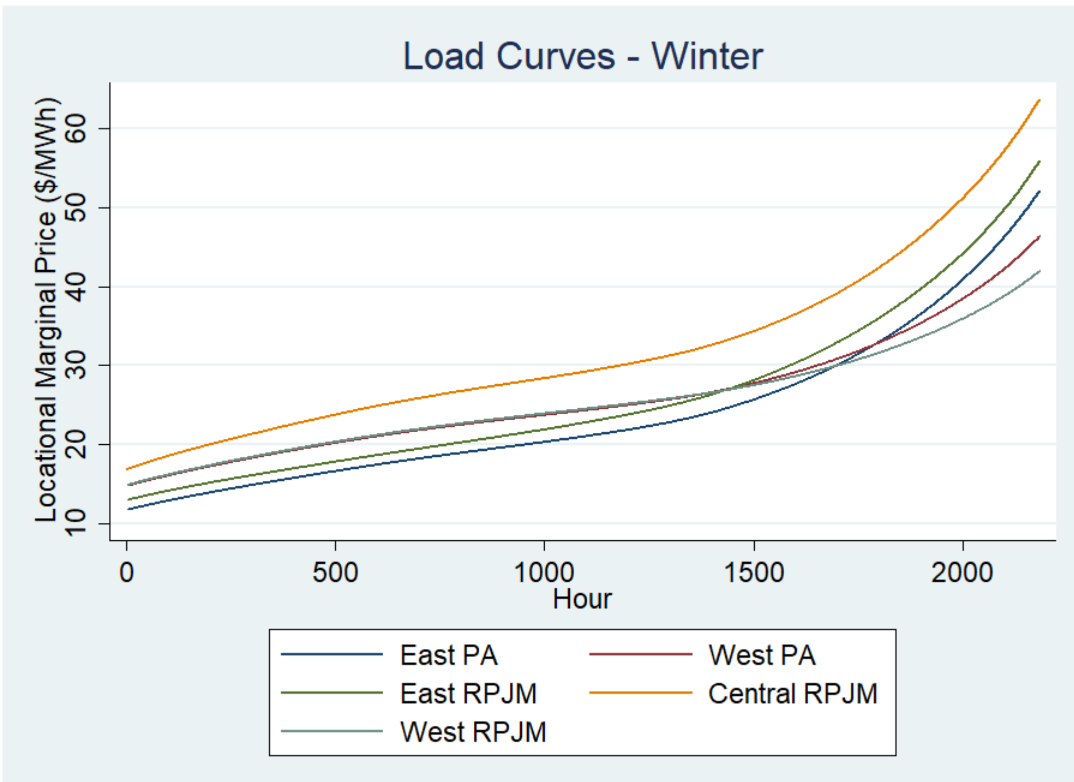


Figure A.9: RT LMPs in PJM in Spring 2016

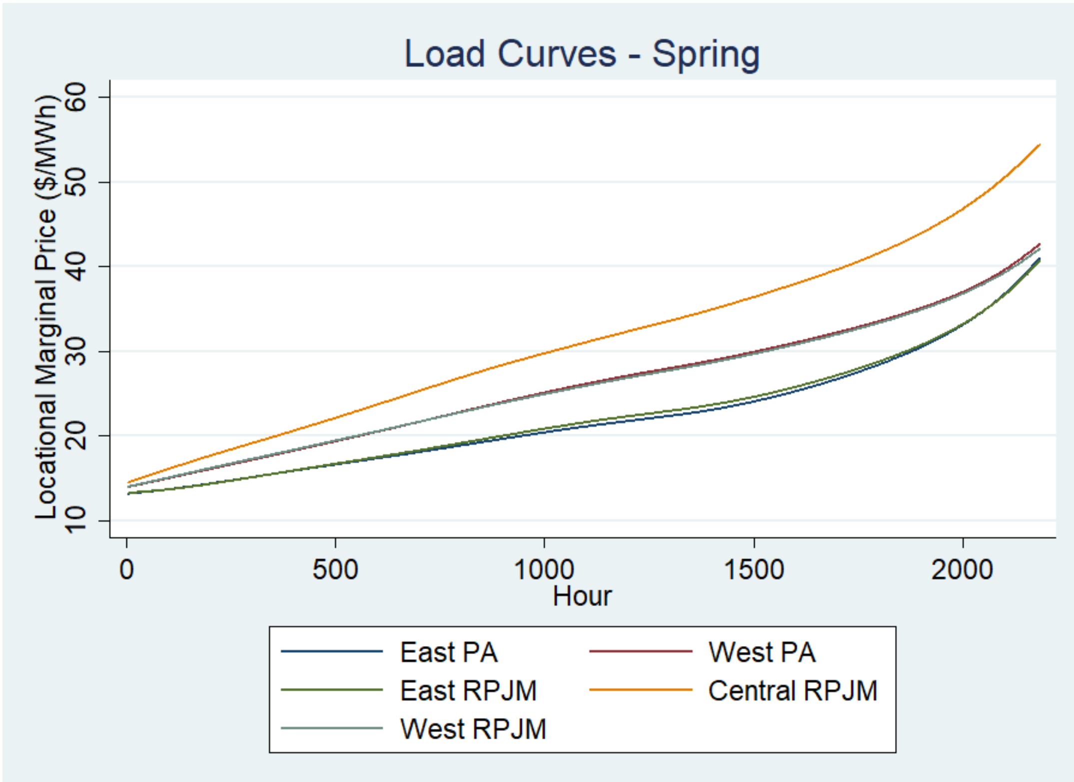


Figure A.10: RT LMPs in PJM in Summer 2016

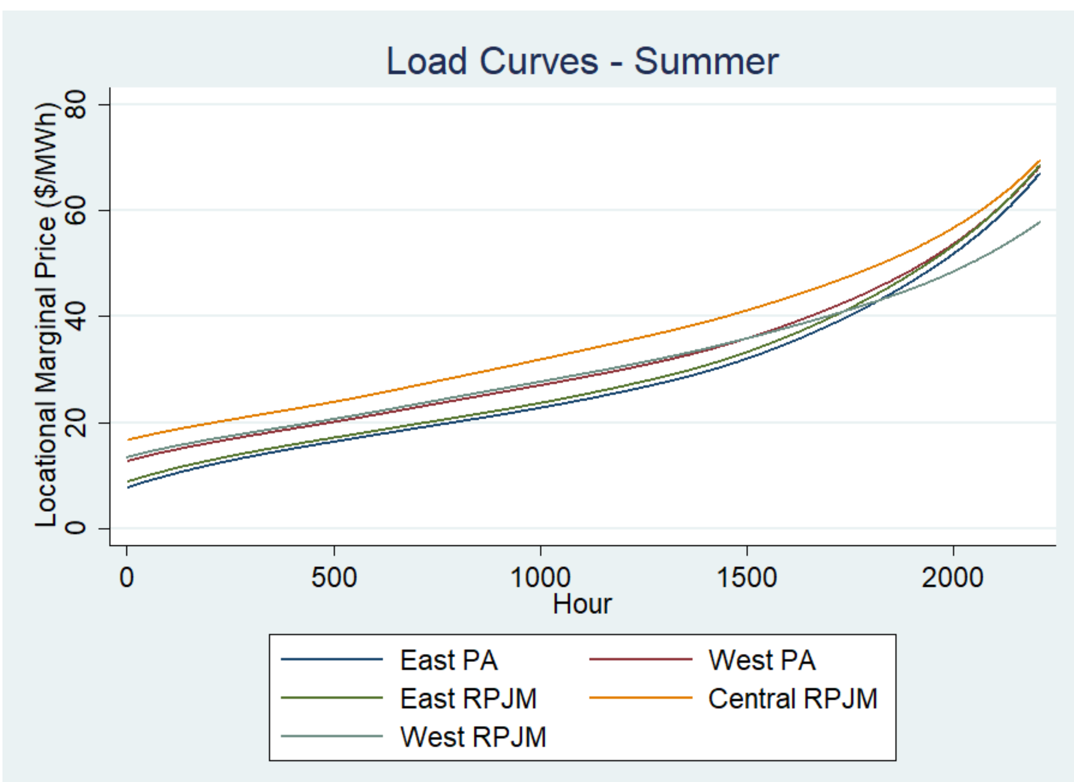
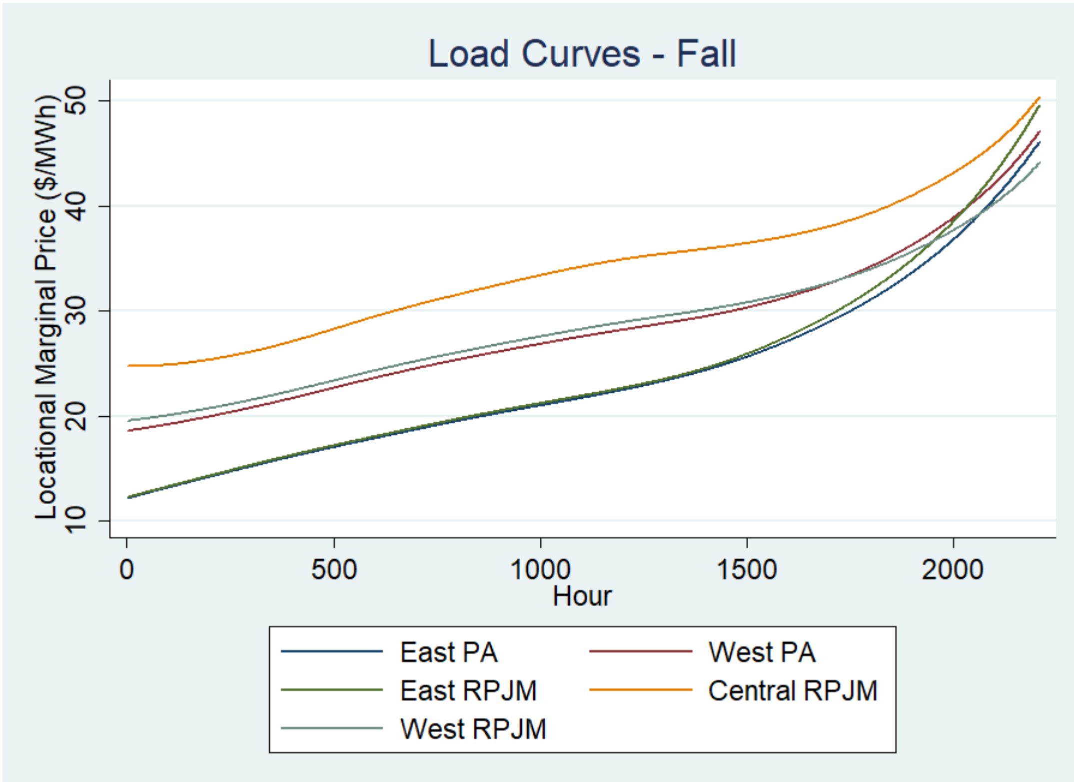


Figure A.11: RT LMPs in PJM in Fall 2016



We calibrate the model using observed market data for 2016 (for the no capacity expansion base case) and 2017 (for the base case with capacity expansion).

3.2 Supply of Electricity from Existing Generation

To calibrate generation, we use five different datasets: The latest National Electric Energy Data System (NEEDS) dataset version 5.15 (updated in August 2015) (accessed 05/21/2017), the SNL summer capacity dataset (accessed 05/21/2017), the EPA's Emissions & Generation Resource Integrated Database (eGrid 2014) (accessed 06/15/2017), the EPA's Continuous Emission Monitoring System dataset (CEMS) and Form EIA-923, which is a survey form collecting detailed electric power data on electricity generation, fuel consumption, fuel,... from power plants, (Schedules 3A & 5A - generator data including generation, fuel consumption and stocks) (accessed 07/18/2017), all of which are for the PJM control area. The reason we need all five dataset to create a full PJM generation dataset is because each dataset has its own shortcomings: The NEEDs dataset might not have updated with the newly added generators, the SNL dataset does not have information on locations of the generators, eGrid was last updated in 2014 by the time we last accessed it and thus does not have the most up-to-date data on all PJM generators, EIA-923 only includes a subset of PJM generators taken from a randomized procedure, and CEMS only includes generators of bigger size that are dispatched during the year. Our general method to establish a full PJM generator dataset is to first merge the NEEDs dataset and the SNL dataset to get the first version of the final dataset. Second, we manually check the plants which are in counties that we think might not be in PJM and remove them from this final dataset if necessary. Then, EIA 923 form, eGrid and CEMS datasets are used to match the plants in the final dataset to reasonable heat rates, capacities and emission rates.

The final NEEDS dataset that is used here consists of both active plants and retired plants and covers 10 NERC regions that roughly coincide with the PJM control area as of 2016: *PJM_AP*, *PJM_ATSI*, *PJM_COMD*, *PJM_DOM*, *PJM_EMAC*, *PJM_PENE*, *PJM_SMAC*, *PJM_WMAC*, *PJM_West*, and *S_C_KY*. The NEEDS dataset includes plant names, plant types, unit IDs, oris plant codes (Office of Regulatory Information System codes, which are the unique identifiers assigned to power plants in NEED), geographical locations of the plants at county level, fuel types, plant capacities, heat rates, and plants' available information (online years, retirement years). The SNL dataset includes plant names, plant types, unit IDs, plant capacities, fuel types, net annual CO_2 emissions, net annual generations, fuel costs and variable operation & maintenance costs (VOM costs). The NEEDS dataset is slightly out of date compared to the SNL dataset, so we combined it to the SNL dataset to obtain a final dataset that will be used for our model calibration. The final dataset consists of the active units located in counties of PJM control area that are in both NEEDS and SNL and the active units that are in NEEDS but not in SNL and those that are in SNL but not in NEEDS.

The NEEDS dataset has 3,071 units in total and the SNL dataset has 3,522 units in total. Most of the plant names, plant types, capacities and unit IDs in the two datasets are similar but not quite exactly the same, therefore, to find the matched and unmatched units in both datasets, we have to merge the two datasets multiple times using both regular merge (merge based on perfectly matched plant names, plant types and unit IDs) and probabilistic merge (merge based on similar plant names, plant types and unit IDs). We first find all the one-to-one unit matches, remove them from the

NEEDS and SNL and then find the one-to-many, many-to-one, and many-to-many unit matches from the remaining units in both datasets. These non one-to-one matches occur because in some cases, the NEEDS report aggregate units made up of multiple boilers of the same plant types while the SNL reports each individual boilers. After finding all these matches, the leftover units from both datasets are the unmatched units, of which only those that are active, are included in the final dataset. Detailed processes of the one-to-one units matching and non-one-to-one units matching are described in the following subsections. Table A.1 and Table A.2 are summaries of our merge results for the NEEDS and the SNL.

3.2.1 Construction of Population of Existing Units

One-to-One Matching NEEDS and SNL Units To find one-to-one unit matches between NEEDS and SNL, we merge the two datasets four times.

First, we match the units that have exactly the same plant names, plant types and unit IDs in NEEDS and SNL. This merge results in 456 one-to-one matched units between the two datasets. These matched units are set aside and removed from the original NEEDS and SNL datasets, leaving the remaining NEEDS dataset to now have 2,615 units and the SNL dataset to now have 3,066 units.

Second, for the remaining units in the two datasets, we use a probabilistic merge (reclink2) in Stata, again using plant names, plant types and unit IDs with the additional requirement that unit IDs be matched exactly. The probabilistic merge gives us a match score for each unit in a range from 0 (not a match at all) to 1 (a perfect match). A match score of 0.9 or above provides a pretty good one-to-one match between the two datasets. We hand check those with a match score of less than 0.9 to filter out the wrong matches.

To do hand-checking, we consider a unit in NEEDS a match for a unit in SNL if they have similar plant names, the exact same unit ID, similar plant type, similar fuel type and similar capacity. For example, unit "Homer City" in SNL with plant type of Steam Turbine, unit ID of 2, fuel type of Coal and capacity of 617.5 MW is a one-to-one match for unit "Homer City Station" in NEEDS with plant type of Coal Steam, unit ID of 2, fuel type of Bituminous and capacity of 614 MW. This provides another batch of 1,577 one-to-one matched units. We again remove these additional one-to-one matched units from the two datasets.

Third, for the remaining units from the two datasets (1,038 units from NEEDS and 1,489 units from SNL), we again use reclink2, this time only using plant names and unit IDs as identifiers since the plant capacities for this subset of units maybe different. We notice several instances where an unit in SNL matches with more than one unit in NEEDS or several units in SNL match with one unit in NEEDS or several units in SNL match with several units in NEEDS. For example, unit "Covanta Plymouth (Montenay Montgomery)" in SNL with fuel type of Biomass, unit ID of 1 and capacity of 28 MW is a match for two units in NEEDS combined, which are "Montenay Montgomery LP" unit ID of 1 and unit ID of 2 with fuel type of Municipal Solid Waste (MSW), capacity of 14 MW each. But these types of matching units are not one-to-one matches and thus we rule them out for now and will only match them in the next subsection. We also relax the requirement of exactly matched unit IDs since the two datasets can have different unit IDs to mean the same units; for example, unit 1 of the same plant in the NEEDS dataset has unit ID of GEN1 (generator 1) but in the SNL dataset has unit ID of BOIL1 (boiler 1). They, however, both mean the same unit (unit 1). Hand-checking and correcting the match results from this round of merge, we have another 263 one-to-one matched units, bringing the total one-to-one matched units so far to 2,296 units.

Last, for the remaining units in the two datasets (775 units from NEEDS and 1,226 units from SNL), we use reclink2 once more, this time only using plant names. Hand-checking and correcting the match results again, we end up with another 147 one-to-one matched units, bringing the total one-to-one matched units so far to 2,443 units. The breakdown of these units and their total capacities are shown in Table A.1 and Table A.2, rows 2-6. Of these 2,443 units, 2,367 units are active in both datasets, 36 units are only active in NEEDs, 27 units are only active in the SNL and 13 units are inactive in both datasets. We only include those units that are active in SNL in the final dataset, which means only 2,394 (2,367+27) are included in the final dataset out of the 2,443 one-to-one matched units.

After completing the one-to-one units matching, we have 628 units in the NEEDS that cannot be matched one-to-one to the units in SNL and 1,109 units in SNL that cannot be matched one-to-one to units in the NEEDS. These two sets of units are next to be matched together either one-to-many,

many-to-one or many-to-many.

Non-One-to-One Matching NEEDS and SNL Units We can also match several units in NEEDS to one unit in SNL, or one unit in NEEDS to several units in SNL, or many units in NEEDS to many units in SNL, as long as these matching units have the same plant names in both datasets. To do this, we find plants in NEEDS and SNL that have the same or similar plant names, same or similar plant types and the same or close to the same total capacity across all the units in the plants of the same fuel types. For example, two coal units in SNL, Joliet 29 unit 7 and unit 8 with capacity of 518 MW each, combined together is a many-to-many match for four bituminous units in NEEDS, Joliet 29 units 71, 72, 81, 82 with capacity of 259 MW each. We find 66 SNL units that can be matched to 118 NEEDS units that can be collapsed down to just 41 one-to-one common plants or common sub-plants of the same fuel types between the two datasets. The breakdown of these units and their total capacities are shown in Table A.1 and Table A.2, rows 7-11. Of these 41 common plants, 40 plants are active in both datasets, corresponding to 113 units in the NEEDS and 65 units in the SNL, 1 plant is active in NEEDS but inactive in the SNL, corresponding to 5 units in the NEEDS and 1 unit in the SNL. There is no units inactive in the NEEDS but active in SNL and no units inactive in both datasets. We again only include those units that are active in SNL in the final dataset, which means only 65 additional units are included in the final dataset out of the 66 one-to-one matched units. Note that we only include the units from one dataset (SNL) to avoid double counting. We choose the SNL because random unit checks online show the SNL to have more accurate capacities.

We now have 2,509 units in SNL matching with 2,561 units in NEEDS. Removing these unit matches from the original NEEDS and SNL datasets, we end up with 510 unmatched units in NEEDS (Table A.2, row 12) (of which 240 units are retired) and 1,013 unmatched units in SNL (of which 23 units are retired) (Table 2, rows 12-14).

Checking Unmatched NEEDS and SNL Units After matching is completed, there remain 270 active units in NEEDS that cannot be matched to the SNL. Since the NEEDS is slightly outdated compared to the SNL, it is possible some of the 270 units are already retired. We search each of these units on the web to remove those units that are no longer available as of 2016. Of these 270 units, we find 34 units are actually retired, closed, withdrawn, demolished, decommissioned, shuttered, forced to stop due to regulatory violations, or not yet operating in 2016, leaving us with only 236 active unmatched units in NEEDS. Of these 236 units, 53 are believed to be aggregated units over small capacity units of the same fuel types, regions and states.

We believe these 53 units are disaggregated in the SNL dataset and thus they are not included in the final dataset to avoid double-counting. Of the 183 remaining units, 62 are believed to not be actually in PJM since the 10 PJM NERC regions do not exactly match the PJM control area, 2 units had capacity of 0 and did not generate during 2016, 14 are small solar PV units, 24 (most of them are combustion turbine units) are small generators suspected to provide power to local facilities and non-dispatchable by PJM, 21 are small units of less than 3 MW that we cannot find information about operating status or capacities. We exclude these units mentioned above (34+53+62+2+14+24+21=210) from the final dataset. Therefore, only 60 units from the 270 unmatched units in the NEEDS are included in the final dataset (Table A.1, row 13). These are the units that are still operating and in PJM control area but are not in the SNL.

For the remaining unmatched units in SNL (1,013 units), we filter out the 23 units that are inactive and hand-check the remaining 990 units to make sure they are indeed in PJM control area. After manually checking these units, we only keep 576 units that we believe to belong in PJM and integrate these units into our final dataset (Table 2, row 13).

Table A.1: NEEDS Dataset v5.15 Unit Breakdown.

	Number of Units	Capacity of Units (MW)	Included in Final Dataset?
Total Number of Units in NEEDS v5.15 Dataset	3,071	219,511	—
One to One Matched with SNL Units	2,443	178,307	—
Active Units In NEEDS and SNL	2,367	173,503	Y
Active in NEEDS and Inactive in SNL	36	2,092	N
Inactive in NEEDS and Active in SNL	27	2,451	Y
Inactive in both NEEDS and SNL	13	261	N
Collapsed to Merge with SNL Units	118	5,308	—
Active Units In NEEDS and SNL	113	5,212	Y
Active in NEEDS and Inactive in SNL	5	96	N
Inactive in NEEDS and Active in SNL	0	0	Y
Inactive in both NEEDS and SNL	0	0	N
Unmatched with SNL	510	35,896	—
Units in PJM and Still Active, but Not in SNL	60	2,470	Y
Units Not Included in Final Dataset	450	33,426	N
Unmatched Aggregate Units	53	1,650	N
Not in PJM	109	11,908	N
In PJM	288	19,868	N
Listed As Inactive by NEEDS	202	16,865	N
Identified As Inactive Via Web Search	25	2,893	N
Otherwise Removed Via Web Search	61	109	N

Table A.2: SNL Dataset Unit Breakdown as of 2016.

	Number of Units	Capacity of Units (MW)	Included in Final Dataset?
Total Number of Units in SNL Dataset	3,522	193,500	—
One to One Matched with NEEDS Units	2,443	176,635	—
Active Units In NEEDS and SNL	2,367	161,420	Y
Active in NEEDS and Inactive in SNL	36	11,643	N
Inactive in NEEDS and Active in SNL	27	2,266	Y
Inactive in both NEEDS and SNL	13	1,306	N
Collapsed to Merge with NEEDS Units	66	5,138	—
Active Units In NEEDS and SNL	65	4,771	Y
Active in NEEDS and Inactive in SNL	1	367	N
Inactive in NEEDS and Active in SNL	0	0	Y
Inactive in both NEEDS and SNL	0	0	N
Unmatched with NEEDS	1,013	11,727	—
Not in PJM	414	850	N
In PJM	599	10,877	—
Units Included in Final Dataset	576	10,646	Y
Units in SNL, Marked as Inactive	23	231	N

Final Dataset Unit Breakdown After matching the NEEDs and the SNL, we have 3,509 units in total. Table A.3 summarizes the breakdown of these units and their total capacity (rows 2-4) as well as the breakdown of fuel types (rows 5-18). This dataset consists of the matched units that are active in both NEEDS and SNL or active in SNL and inactive in NEEDS and the unmatched units in SNL and in NEEDS. As mentioned earlier, of the matched units, 2,394 one-to-one matched units are added to the final dataset and additional 65 non-one-to-one matched units are also added in the final dataset, summing to 2,459 units. Finally, we also include the 60 unmatched units in NEEDS and the 576 unmatched in SNL in the final dataset.

All the units in this dataset make up total capacity of 181,573 MW, of which 168,457 MW are in both NEEDS and SNL, 2,320 MW are in NEEDS only and 10,796 MW are in SNL only. The biggest fuel sources are gas, coal and nuclear, with natural gas units make up 38% of total capacity in PJM, followed by coal units with 32% and nuclear units with 17%. Renewables units and oil units are only 9% and 4% of total PJM capacity, respectively. Biomass, landfill gas, and other fuel types are negligible, making up of only 1% of total capacity in PJM. The final dataset unit breakdown is shown in table A.3.

Table A.3: Final Dataset Unit Breakdown in 2016.

	Number of Units	Capacity of Units
Total Number of Units in Final Dataset	3,095	181,573
Included in both NEEDS and SNL	2,459	168,457
Included in NEEDS and Not Included in SNL	60	2,470
Not Included in NEEDS and Included in SNL	576	10,646
Natural Gas	836	68,688
Combustion Turbines	597	35,468
Combined Cycle	231	33,190
Other	8	30
Coal	181	57,359
Oil	548	7,023
Nuclear	32	31,244
Biomass & Landfill Gas	765	1,547
Renewables	723	15,479
Solar	340	1,544
Wind	78	5,317
Hydro	305	8,616
Other Fuel	10	233

Final Sample of Existing Representative EGUs To establish the supply curve, the next step is to group these units in the final dataset into bins of the same or similar attributes. These attributes include total marginal cost, which is the sum of fuel cost, transportation cost and emission cost, fuel type, emission intensity, and location by state and region. After the binning process, we reduce 3,095 units in our final dataset to 843 representative units. We then aggregate the capacity of all the units by bin, as well as calculate average heat rate and emission rate by bin and assign them as heat rate and emission for each bin. The transmission unconstrained PJM supply curve is the merit order of these 843 aggregated units.

3.2.2 Unit Characteristics

Capacity For those units that are in both the NEEDs and SNL dataset, we use SNL's capacities for them. For the rest of the units, we use capacities from the NEEDs.

Capacity factor/Plant availability To impute capacity factor for each fuel type, we use three datasets - the Generating Availability Data System (GADS) and EIA's Daily U.S. Nuclear Outage for 2016 and the PJM Generation Outage Daily in 2016. Subtracting the daily nuclear outage from the PJM's total daily outage, we have the daily outage of non-nuclear generators in PJM. Now, we use GADS to find the percentage of outage for each non-nuclear fueltype monthly. Combining the PJM's generation outage data and the GADS together, assuming the same outage for non-nuclear fueltypes in each hour of the same month, we can calculate the percentage of outage for each fueltype in each load segment.

Private Marginal Costs Marginal cost of each power plant is calculated as MC = Fuel Cost (including Transportation Cost) + Regulatory Costs + 10% Mark Up, each of these component costs except for the 10% Mark Up is calculated as detailed above. The 10% Mark Up is mentioned in the Monitoring Analytics' 2016 State of the Market as a component of LMP.

Heat rate To assign heat rate to each unit in our final dataset, we merge the final dataset with eGrid, EIA form 923 and CEMS. We start with the units that are in CEMS because CEMS collects actual emission rates and heat rate reported by units that were dispatched in 2016. First, we use heat rate from CEMS for units in the final dataset that are also in the CEMS in 2016 (374 units). After this step, we have 2,721 units remaining in our final dataset.

The next reliable dataset we use for assigning heat rate to the rest of the units is eGrid. However, there are a few instances when eGrid reports a negative total generation, which coincide with the units that are marked “Data from EIA-923 Generator File overwritten with distributed data from EIA-923 Generation and Fuel", so we flag units in eGrid with this description to be excluded from our calculation of heat rate and emission rate. The average heat rate in eGrid is 13,147.53 with standard deviation of 14,698.89. We use eGrid's average heat rate plus two times its standard deviation (=42,545.31) as eGrid's threshold to assign heat rate from eGrid to the remaining units in our final dataset. This means, we assign the units in the remaining 2,721 units in our final dataset that are also in eGrid their corresponding eGrid's heat-rate as long as the eGrid's heat rate is no more than 42,545.31. After this step, we assign 1,490 units more.

We now have 1,231 units left to assign heat rate to. Next, we use the heat data from SNL to assign to these remaining units. Since the SNL did their own heat rate calculation using EIA 923, in some cases when some units generated a very small amount of energy through the entire year, their estimates of heat rate for these units are so unrealistically high that they had to cap them at 100,000. We believe these estimates are not reasonable and want to exclude these units from our heat rate calculation. To do so, we establish a heat rate threshold for SNL of its average heat rate plus two times its standard deviation (8,914.102+2*8,779.87 = 26,473.842). We assign any remaining units that are in SNL their corresponding SNL heat rate as long as they are no more than 26,473.842. After this step, we assign 801 more units. Thus, now we only have 430 units to assign heat rate to.

To assign these 430 units, we use the EIA 923 form, which reports heat rate for a subset of these units (411) for five years 2012-2016. We calculate the average heat rate across five years and assign them accordingly to these 411 units. The 19 remaining units are assigned heat rates from the NEEDS. Summary of heat rate assignment is shown in table A.4.

Table A.4: Heat Rate Assignments to Units in Final Dataset.

Dataset	Number of Units Assigned Heat Rate to
Final Dataset	3,095
CEMS	374
eGrid	1,490
SNL	801
EIA 923	411
NEEDS	19

Fuel Prices

Natural Gas Prices: We get daily gas spot price from Bloomberg for seven gas hubs: Alliance, Dominion North Point, Chicago City Gate, Lebanon OH, TETCO Zone M3, Tennessee Gas Zone 4 - Marcellus and Transco Leidy in 2016, 2017 (for calibration purpose) and 2018 (for validation purpose). Assuming that each power plant buy gas from the hub that is closet to their location in 2016, we map these hubs in ArcGIS and assign each power plant to the gas hub closest to it. We assume power plants buy gas from the same gas hubs in 2017 and 2018 and going forward.

To calculate gas transportation price, we use natural gas delivered price to the electric power sector from the EIA's Annual Energy Outlook 2017 (AEO 2017) for 2016 prices by census region and aggregate the Bloomberg gas spot prices above also by census region. The gas transportation price for each region is then determined by subtracting the aggregated regional Bloomberg gas spot prices from the EIA's gas delivered price.

The final gas fuel price for each natural gas-fired power plant is determined by the the gas spot price associated with that plant plus the transportation gas price to the census region in which the plant is located.

Coal Prices: For coal fuel prices, we use weekly coal spot prices for five main coal basins in 2016 from the EIA: Central Appalachia Basin, Northern Appalachia Basin, Illinois Basin, Powder River Basin and Unita Basin. Coal transportation prices from each coal basin to the states that it historically shipped to are taken from the EIA's Coal transportation rates database, then weighted averaged in each PJM regions (defined in section 2) by modes of transportation (railway, train and truck). We then calculate five weekly total coal price for each plant assuming they could potentially buy coal from the five basins by adding the weekly coal spot prices from the respective basin to the weighted coal transportation price from that basin to the state in which the power plant located. Next, we average these five total coal prices by load segments as defined in section 2 to get the five potential total coal prices for each coal-fired power plant by each segment. The final total coal price for each power plant in each segment is then determined by the minimum total coal prices for that plant in that segment.

Oil Prices and Nuclear Prices: Quarterly oil prices for electric power sector by census region are taken from the EIA for actual 2016 data. Annual uranium prices by census region are also taken from the EIA.

Other Fuel Prices: Other fuel prices are taken from SNL power plant data, associated with each power plant.

Other Pre-existing Regulatory Costs Emission cost for each power plant is the sum of SO_2 emission cost and NOx emission cost. SO_2 emission cost is calculated in $\$/\mathrm{MWh}$ as SO_2 current permit (from EIA) $\times$ SO_2 permit rate (in \$/mmBTu, from NEEDs data) $\times$ heat rate of the associated power plant $\times \dfrac{1}{2000,000}$. Similarly, NOx emission cost is calculated in \$/MWh as NOx current permit (from EIA) $\times$ NOx permit rate (in \$/mmBTu, from NEEDs data) $\times$ heat rate of the associated power plant $\times \dfrac{1}{2000,000}$.

Marginal fuel cost for each of EGU is calculated as the product of its fuel cost and heat rate. The private marginal cost of said EGU is then calculated as the sum of its marginal fuel cost and regulatory cost.

Variable Operation and Maintenance Costs Variable Operation and Maintenance Costs (VOC) are taken from the SNL dataset for all units.

Total Marginal Costs Total Marginal Costs (TMC) are calculated as the summation of private marginal cost (PMC) and variable operation and maintenance costs (VOC).

Calibrate Total Marginal Cost Curves for 843 Aggregate EGUs

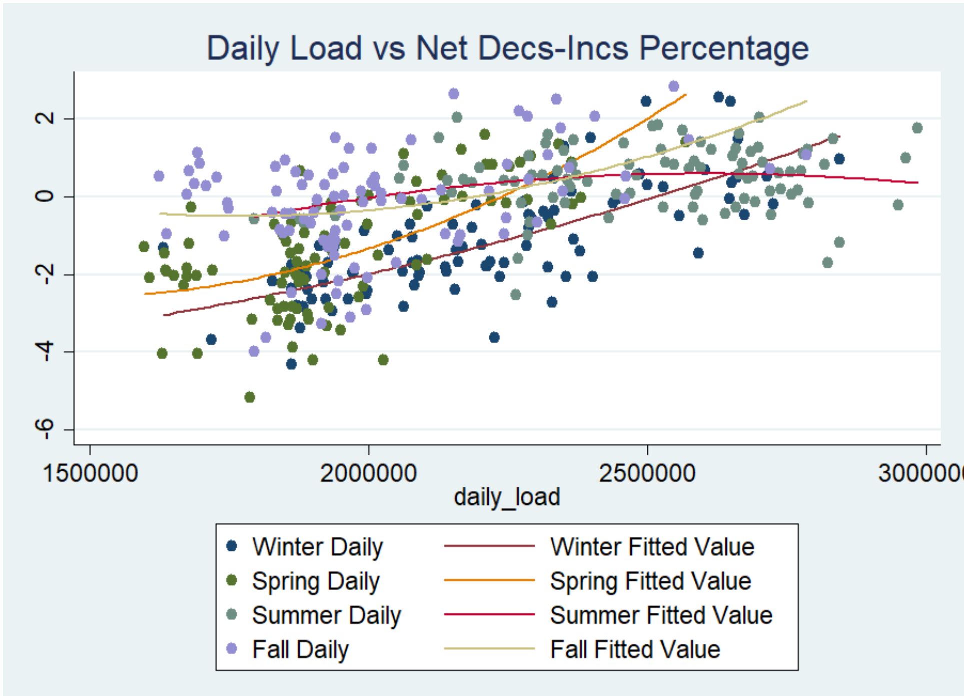
As mentioned in previous section, we bin 3,095 EGUs in PJM into 843 aggregate units based on similarities in fuel types, marginal costs, heat rates, emission rates, locations by state and region. The private marginal cost of each of these 843 unit bins in each load segment l is determined by fitting a linear regressions to all the EGUs in that bin for each load segment, $p_{jl} = m_{jl}g_{jl} + b_{jl}$, in which, m_{jl} and b_{jl} are the slope and intercept of aggregate EGU j at load segment l.

3.3 Virtual Bidding

We account for the amount of cleared daily net virtual bid (decs - incs) into our segment loads in our model. We detect a quadratic relationship between daily net virtual bids and daily load in each season. These correlations are shown in the figure below. We use quadratic regression with seasonal fixed effects (W = Winter, S = Spring, Su = Summer, F = Fall) to predict the hourly cleared virtual bid (NVB) and net cleared virtual bid in the load segment level across PJM (L). The daily cleared incs and decs are taken from the dataset “Daily cleared INCs, DECs, and UTCs" from PJM’s data miner 2.

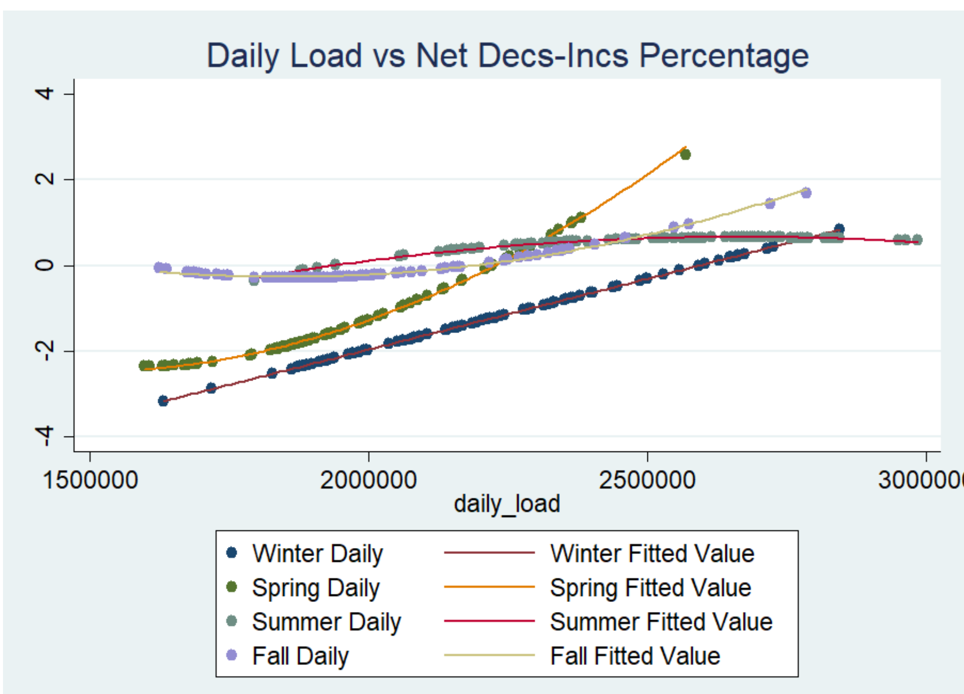
$$
\begin{aligned}
\mathrm{NVB}_{daily} = {} & \beta_1\left(\mathrm{L}_{daily}\right) + \beta_2\left(\mathrm{L}^2_{daily}\right) + \delta_1\left(\mathrm{W}\right) + \delta_2\left(\mathrm{S}\right) + \delta_3\left(\mathrm{Su}\right) + \delta_4\left(\mathrm{F}\right) \\
& + \sigma_1\left(\mathrm{W} \times \mathrm{L}_{daily}\right) + \sigma_2\left(\mathrm{S} \times \mathrm{L}_{daily}\right) + \sigma_3\left(\mathrm{Su} \times \mathrm{L}_{daily}\right) + \sigma_4\left(\mathrm{F} \times \mathrm{L}_{daily}\right) \\
& + \gamma_1\left(\mathrm{W} \times \mathrm{L}^2_{daily}\right) + \gamma_2\left(\mathrm{S} \times \mathrm{L}^2_{daily}\right) + \gamma_3\left(\mathrm{Su} \times \mathrm{L}^2_{daily}\right) + \gamma_4\left(\mathrm{F} \times \mathrm{L}^2_{daily}\right)
\end{aligned}
$$

Figure A.12: Daily Load vs Daily Net Cleared Virtual Bid



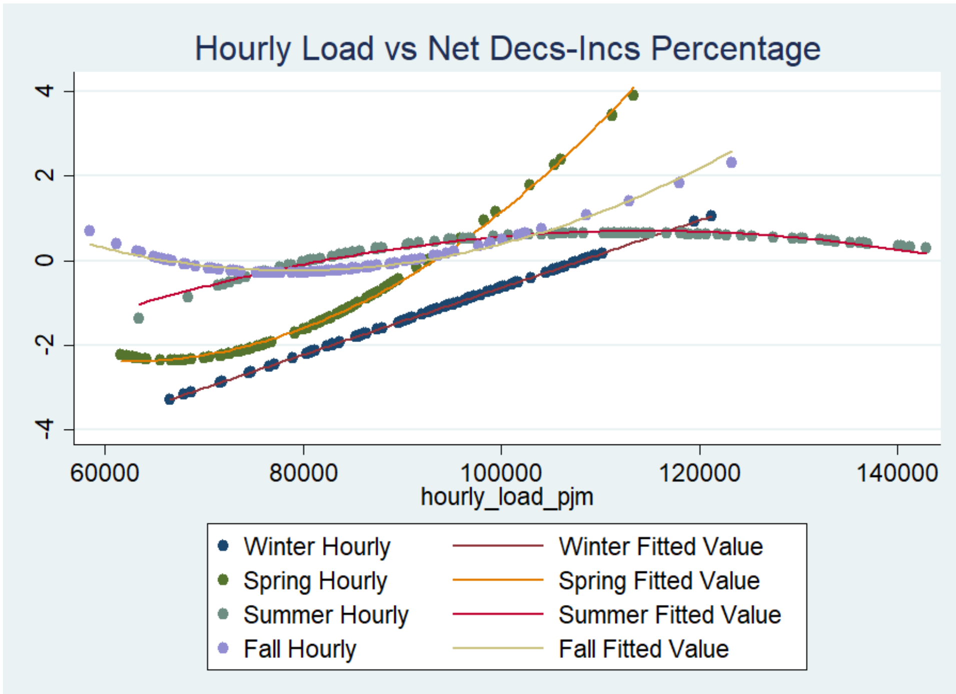
The predicted daily virtual bids from the above regression are shown in the figure below. We can see that the quadratic relationship is a reasonable assumption.

Figure A.13: Daily Load vs Predicted Daily Net Cleared Virtual Bid



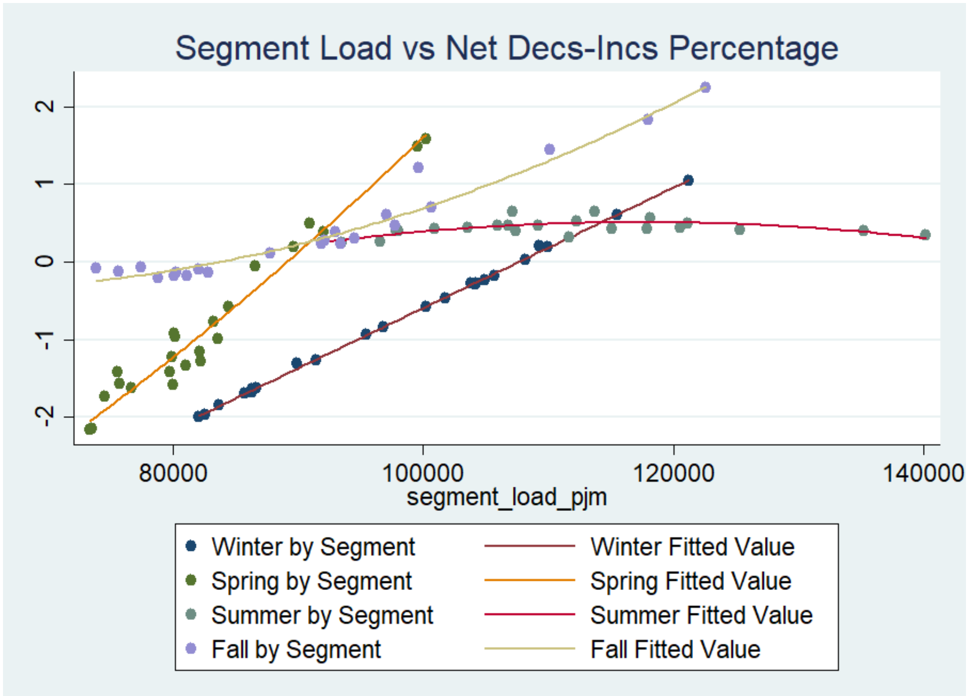
Now, apply the parameters $\beta, \delta, \sigma, \gamma$ found in the regression above (with adjustment to units from daily to hourly) to hourly PJM load, we can predict the net virtual bid cleared every hour across PJM.

Figure A.14: Daily Load vs Predicted Hourly Net Cleared Virtual Bid



Summing all the hours in each load segment, we have a predicted cleared net virtual bid in every load segment as a function of segment load:

Figure A.15: Daily Load vs Predicted Segment Net Cleared Virtual Bid



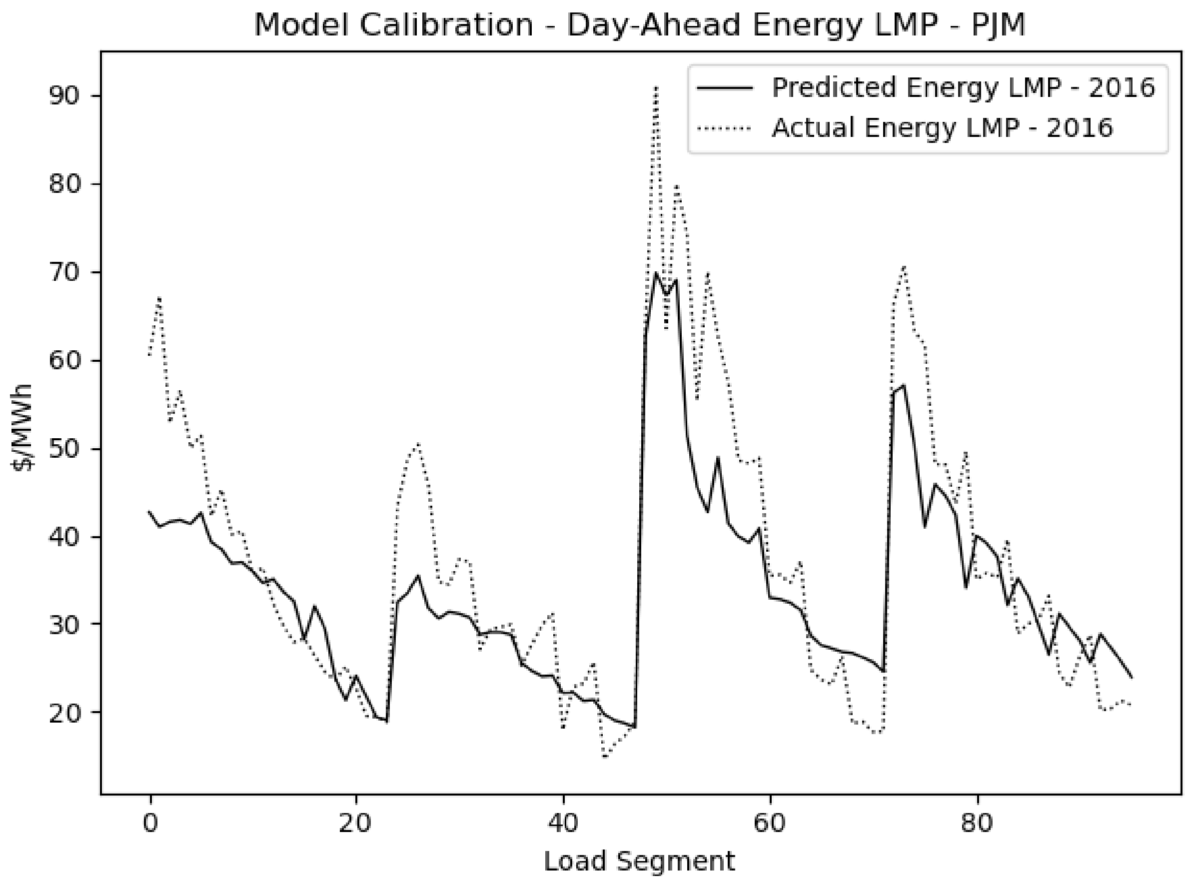
These predicted net cleared virtual bids by segment are divided by the total PJM load in corresponding load segments to find the percentage of net cleared bid in total load in that segment, denoted as z_l, where $l = 1, 2, 3, ..., 96$. The final segment loads used in the model is defined as:

$$L_{il}^{*} = L_{il}\left(1 + z_l\right)$$

in which, $i = 1, 2, 3, 4, 5$ are the index for 5 regions in our model and $l = 1, 2, ..., 96$ are the index for 96 load segments in our model. L_{il} is the original load in each load segment in each region and L_{il}^{*} is the final load in each load segment in each region after accounting for net virtual bid in that segment.

With the addition of virtual bids, the model's prediction of Energy LMP has $R^2 = 0.72$, as shown in Figure A.16.

Figure A.16: Calibration Results for Day-Ahead Energy LMP for PJM in 2016.



3.4 External RECs

We allow for import of external RECs from outside of PJM to help states to comply with their RPS targets. Data of units that are eligible to provide RECs to PJM States are taken from PJM Environmental Information Services (EIS)'s tracking system Generation Attribute Tracking System (GATS), or PJM-GATS for nine states: DC, DE, IL, MD, NJ, OH, PA, VA, WV; and from North Carolina Renewable Energy Tracking System (NC-RETS) for NC, and from Michigan Renewable Energy Certificates System (MIRECS) for MI. By 2017, there are in total 262 units outside of PJM that are eligible to supply RECs to states in PJM.

Since not all of the RECs supplied by these 262 units are used by PJM states to satisfy with their RPS targets, we also limit number of total RECs that can be imported into PA and RPJM so that our modeled REC prices match well with actual REC prices in PA and RPJM in 2016. This results in allowing only 62% of the RECs that are eligible to be imported to PA to actually be bought by PA, and only 81% of the RECs that are eligible to be imported to RPJM to be bought by RPJM states.

3.5 Transmission Network

Transmission line data are taken from SNL's Operating Transmission Projects Map for 2016. Transmission lines that connect every two zones within two PJM regions are aggregated into single aggregated transmission lines, which reflect the net maximum MWhs that can be transferred in between two PJM regions, as shown in Figure $A.1$. We consider a line to connect between two regions if it connects two ISO market hubs, one in each region. We have in total five aggregated transmission lines among our five PJM regions. The table below shows the five aggregate links among our five LMP regions and their respective aggregate KV.

Table A.5: PJM Transmission Networks

Transmission Line Between	Total KV (kv_i)
PA East and PA West	960
PA East and RPJM East	1,880
PA East and Central RPJM	690
PA East and RPJM West	N/A
PA West and RPJM East	N/A
PA West and Central RPJM	N/A
PA West and RPJM West	5,208
RPJM East and Central RPJM	N/A
RPJM East and RPJM West	N/A
Central RPJM and RPJM West	3,460

To know the actual transmission capacities of these lines, we are still missing the impedance in each line above, which is the aggregation of all individual lines between every two PJM regions. Therefore, there is no one realistic impedance we can assume for each aggregate line. We also do not observe the true transmission capacities which can vary in real time due to many climate factors such as wind speed and direction, solar radiation, temperature, etc... Furthermore, our model is an aggregate model and our objective is to calibrate average congestion LMPs for the five links above, simply adding the transmission capacities between PJM regions does not capture well these aggregate price effects as the true congestion LMPs reflect complicated grid realities such as the complicated and largely unobserved disaggregated transmission system between zones within PJM regions which may vary across seasons and correlate with load. As a result, to calibrate these five aggregate transmission line capacities we search for the optimal values of five scalers in each load segment (sc^*_{il}), with which $sc^*_{il} \times kv_i \times 1,000$ represents the transmission capacities in each load segment that minimizes the LMP congestion loss function across PJM in that load segment, using the method used in Ferreyra (2007).

By performing a grid search of 100,000 iterations of combinations of scalers among the five LMP regions in the range of 0.2 to 2, with increment of 0.2, similar to Ferreyra (2007), in the first stage, we find the optimal model predicted congestion LMPs in 5 PJM regions, $prLMP^{cong,stage1}_{ilc^*}$ where c^* is the index number of the iteration that yields the optimal solution, which is the solution to the the first stage loss minimization:

$$Loss^{stage1}_l = \min_{prLMP^{cong}_{ilc}} \left(\sum_i (daLMP^{cong}_{il} - prLMP^{cong}_{ilc})^2 \right)$$

Next, we can calculate the standard deviations across the five PJM regions of the congestion LMP residuals:

$$std^{congLMP}_l = std \left(\sum_i (daLMP^{cong}_{il} - prLMP^{cong,stage1}_{ilc^*}) \right)$$

In the second stage, we re-perform the grid search that we did in stage 1, only this time, the loss function is weighted by the standard deviations across five PJM regions calculated above. The solution to this grid search is the second stage optimal model predicted LMPs, $prLMP^{cong,stage2}_{ilc^*}$.

$$Loss^{stage2}_l = \min_{prLMP^{cong}_{ilc}} \left(\sum_i \frac{(daLMP^{cong}_{il} - prLMP^{cong}_{ilc})^2}{std^{congLMP}_l} \right)$$

Finally, sc_{ilc^*} are used as initial values for an fmincon search in MATLAB to find the exact scalers (sc^*_{il}) that minimize the aggregate gap between the DA congestion LMPs and the model predicted congestion LMPs across the five PJM regions, using the second stage loss function and limiting the lower and upper bounds of these scalers in between the range of $[0.01, 5]$ for all five transmission links. These optimal scalers give us the final optimal model predicted congestion LMPs, $prLMP^*_{il}$.

These optimal results are shown in the figures below.

Figure A.17: Calibration Results for Day-Ahead Congestion LMP in West Pennsylvania (Top Row, Left), East Pennsylvania (Top Row, Right), East RPJM (Second Row, Left), Central RPJM (Second Row, Right) and West RPJM (Bottom Row) in 2016. $R^2 = 0.13$

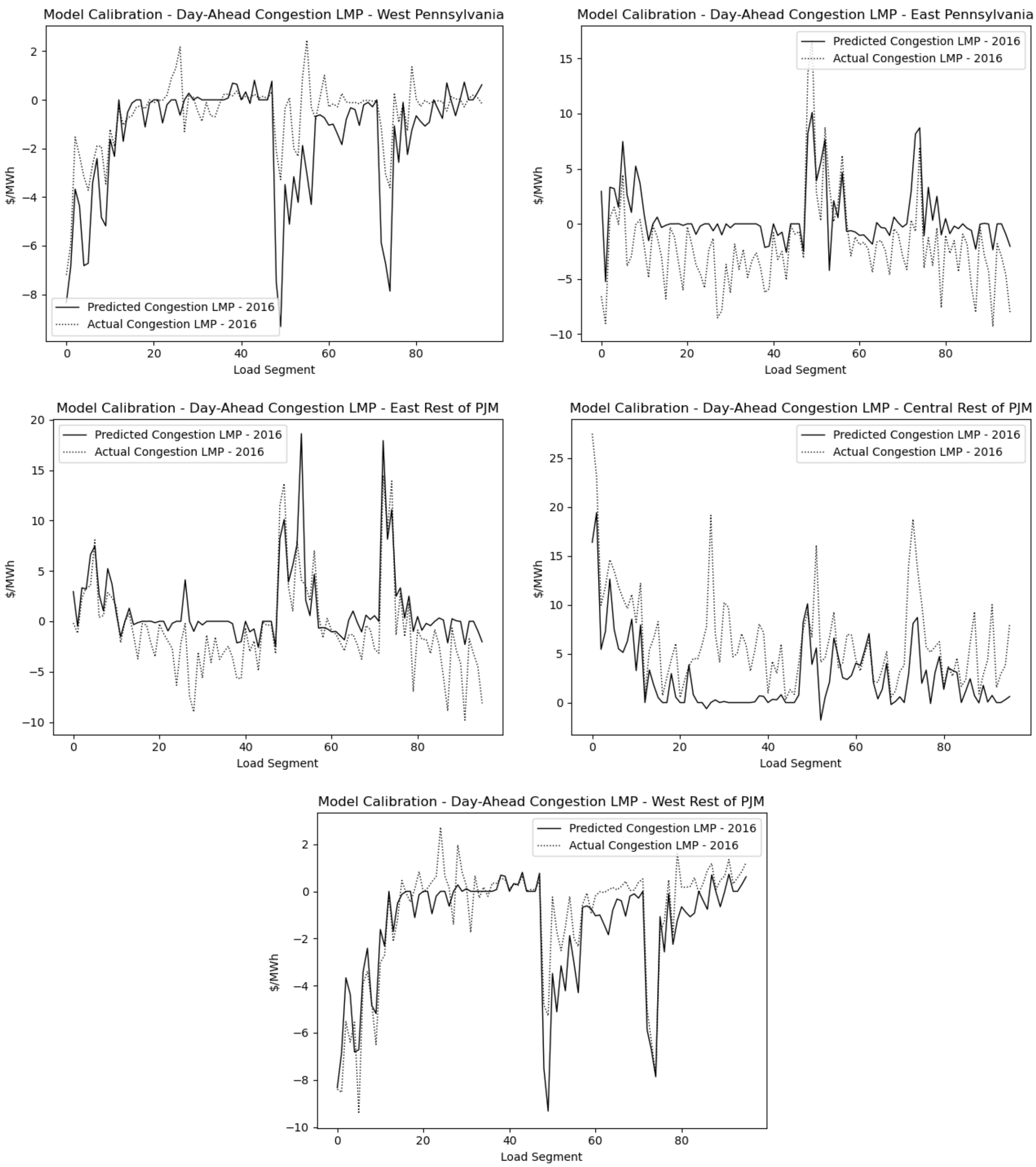


Figure A.18: Calibration Results for Day-Ahead LMP in West Pennsylvania (Top Row, Left), East Pennsylvania (Top Row, Right), East RPJM (Second Row, Left), Central RPJM (Second Row, Right) and West RPJM (Bottom Row) in 2016. $R^2 = 0.70$

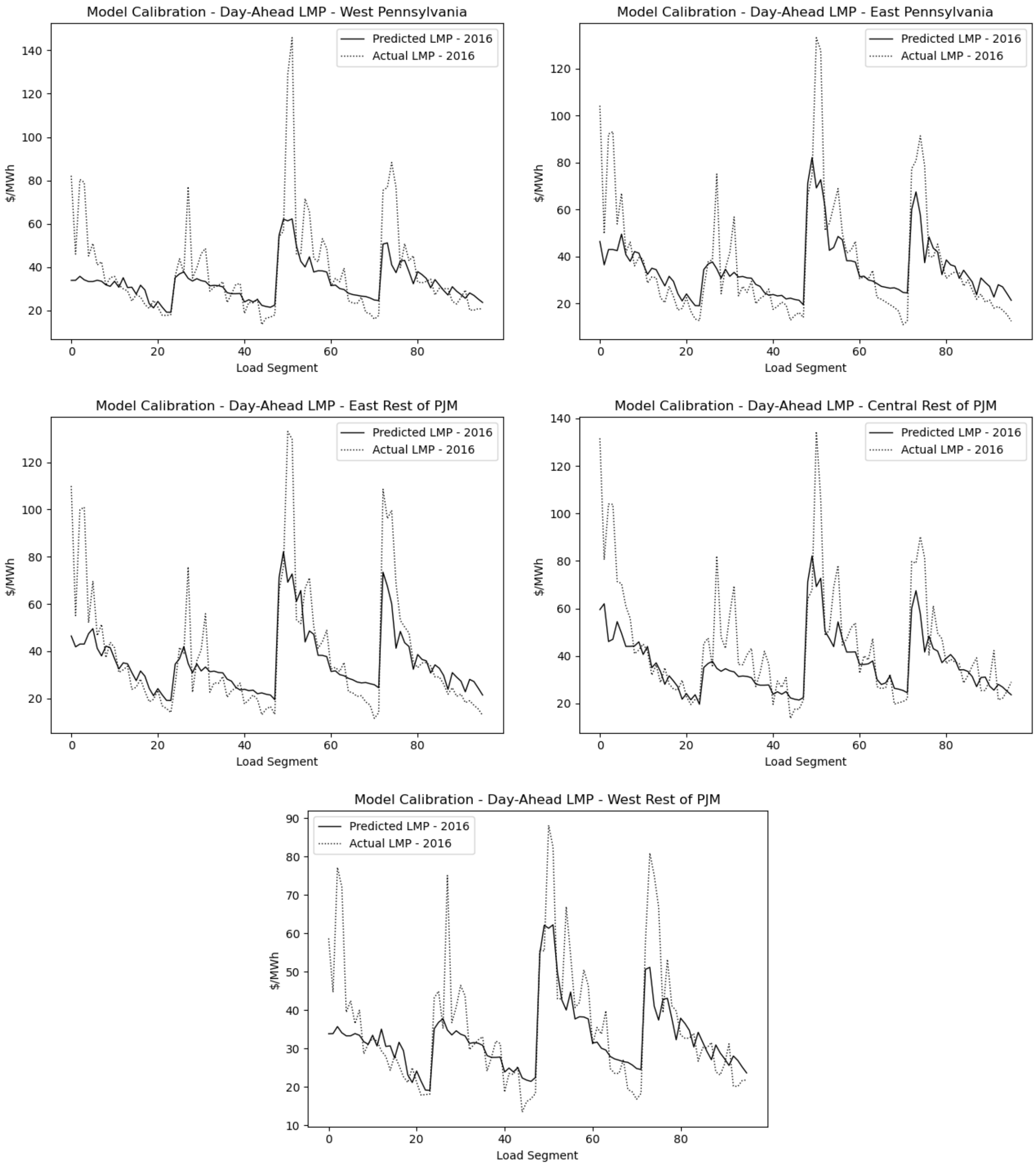


Figure A.19: Calibration Results for Day-Ahead LMP in West Pennsylvania (Top Row, Left), East Pennsylvania (Top Row, Right), East RPJM (Second Row, Left), Central RPJM (Second Row, Right) and West RPJM (Bottom Row) in 2017. $R^2 = 0.36$

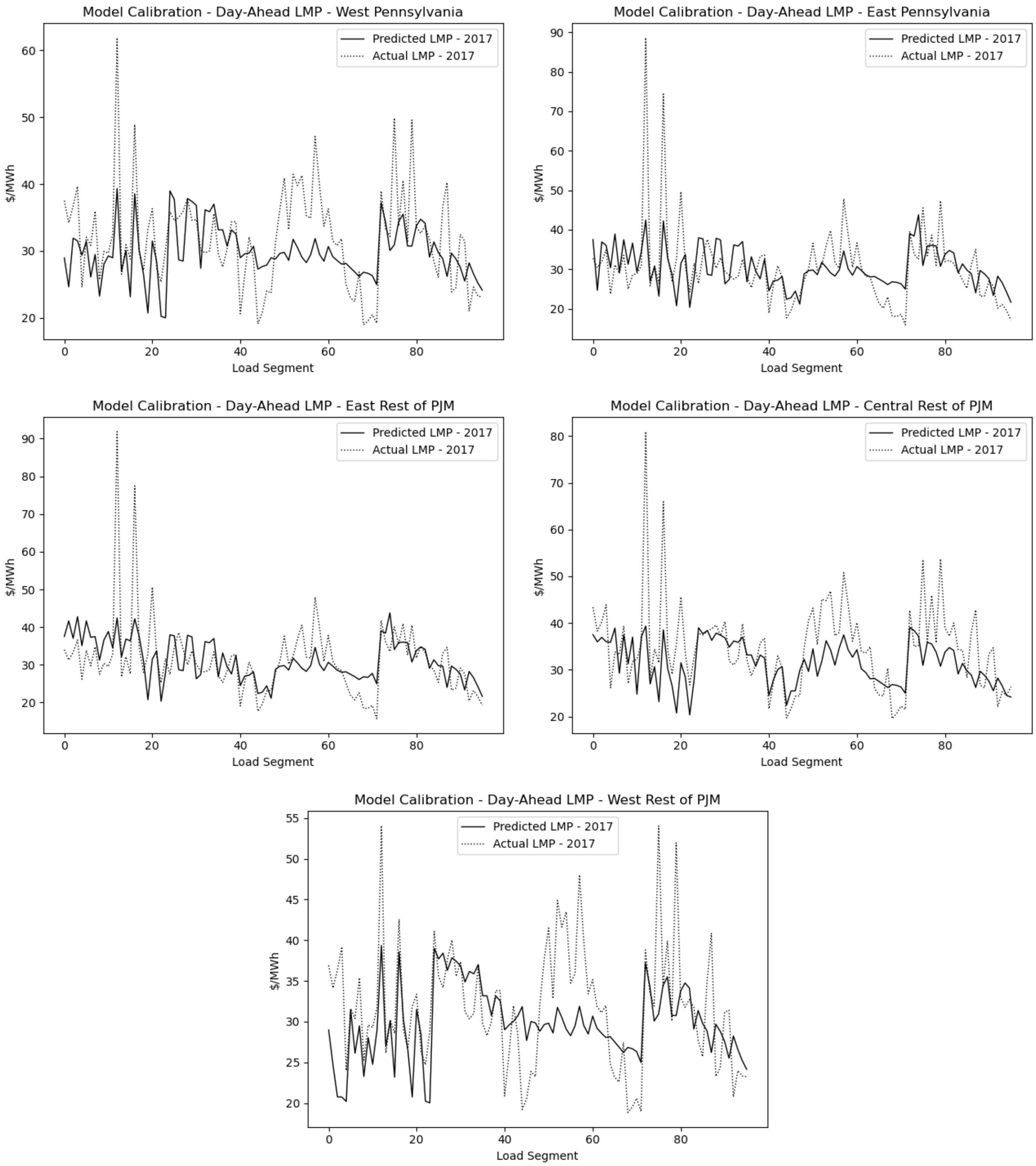


Table A.6: Calibration Results for Total Generation in PA and RPJM in 2016 (GWh).

Fuel	PA-Actual	PA-Calibrated	PA-Error	RPJM-Actual	RPJM-Calibrated	RPJM-Error
Coal	54,672	54,310	0.66%	220,609	219,465	0.52%
Nuclear	82,924	82,924	0.00%	196,622	196,662	0.00%
Gas	68,048	68,321	-0.40%	146,974	146,104	0.60%
Hydro	2,374	2,375	0.00%	11,312	11,312	0.00%
Wind	3,476	3,476	0.00 %	14,240	14,240	0.00%
Oil	363	355	2.51 %	1,800	1,780	1.15 %
Solar	75	75	0.00%	945	944	0.03%
Biomass	1,883	1,857	1.40%	4,017	4,035	-0.46%
Other	1,250	699	44.00%	958	1,311	-36.85%
Total	215,067	213,392	0.31%	597,478	595,813	0.28%

3.6 New Capacity and Supply of Electricity from New Generation

We assume that when new capacity C_{sj}^N of new EGU j in state s is added, it supplies power g_{slj}^N at constant private marginal costs b_{slj}^N and it is expanded with capital costs K_{sj}^N.

There are three possible technologies (units) j that can be expanded each year in each state in our model: NGCC, wind and solar. These are the technologies that are most likely to be developed due to low gas prices and increase in demand for clean energy. We assume the average utilization factors across all gas, wind and solar units in our model in 2016 for these technologies. Specifically, the assumed utilization factor γ_{slj} for a new natural gas unit is 0.961, for a new wind unit is 0.295, and for a new solar unit is 0.17. We also assume the new wind and solar capacities within each state are available to be used to satisfy the RPS requirement for that state for that year.

Marginal cost for the new natural gas unit in a state is based on the average gas price in that state. This marginal cost is calculated as average gas price in a state times assumed heat rate for new gas EGU, which is 9000 khw/BTU. VOC costs for the new units of all three types of technologies are based on the state's average VOC costs for existing EGUs of the same fuel type in 2016. Marginal costs for all new wind and solar units are zeros. Total marginal costs b_{slj}^N for the newly expanded units equal the sum of marginal costs and VOC costs.

Capital costs C_{sl}^N for the three technologies for the 2017 baseline are determined by solving the error minimization problem in which the error is the difference between the actual capacities expanded by technology for PA and RPJM in 2017 and the predicted capacities expanded by technology from our model. Data from new capacities in PJM are taken from PJM's New Service Queue database (https://www.pjm.com/planning/services-requests/interconnection-queues.aspx). In 2017, there are 1,340 MW of new natural gas capacity, 0 MW of new wind capacity and 0 MW of new solar capacity in PA. In the same year, RPJM observed 1,626 MW of new natural gas capacity, 126 MW of new wind capacity and 204 MW of new solar capacity. Our model's calibration results for new capacity by technology in 2017 are as follows.

Table A.7: Capacity Expansion Calibration Results in PJM in 2017.

Technology	PA-Model	PA-Actual	PRJM-Model	RPJM-Actual
Natural Gas	1,270	1,340	1,626	1,766
Wind	0	0	156	126
Solar	0	0	194	204

The total cost of expanding and running a new EGU is, therefore:

$$\sum_s \sum_j C_{sj}^N K_{sj}^N + \sum_s \sum_j \sum_l \delta_l \left(b_{slj}^N g_{slj}^N \right)$$

where C_{sj}^N is the capital cost of new EGU j at state s of expanding K_{sj}^N MW of new capacity for that EGU, g_{slj}^N is total generation provided by new EGU j in state s at load segment l, and b_{slj}^N is the private exogenous total marginal cost in \$/MWh of these new EGUs for providing g_{slj}^N MWh of new generation.

3.7 List of Calibrated Parameters

Table A.8: List of Calibrated Parameters.

Parameters	Explanation
Sets	
J	Number of aggregate EGUs (index j)
I	Number of regions (index i)
S	Number of states (index s)
L	Number of load segment (index l)
q	RPS categories
δ_l	Number of hours in load segment l
Supply Side	Explanation
η^S	Supply elasticity
$\bar{g}_{sjl}$	Capacity of EGU j in state s
ϕ_{sj}	Emission rate of EGU j in state s
m^E_{slj}	Supply function slope in state s and load segment l for existing EGU j
b^E_{slj}	Supply function intercept in state s and load segment l for existing EGU j
b^N_{slj}	Marginal cost in state s and load segment l for new EGU j
K^N_{sj}	Capital cost for EGU j in state s
$\bar{g}^E_{slj}$	Capacity of existing EGU j in state s at load segment l
γ_{slj}	Utilization factor of new EGU i in state s at load segment l
$\bar{g}^R_j$	Total amount of external RECs available to satisfy RPS in PJM
Demand Side	
n_{il}	Demand function slope in region i and load segment l.
c_{il}	Demand function intercept in region i and load segment l.
ϵ_l	Percentage of network loss and net cleared virtual bid in total load for load segment l.
Transmission	
$\bar{f}_{ihl}$	Transmission constraint in region i and load segment l
Policies	
$\bar{r}_{sq}$	RPS target in state s of category q
$\bar{E}_{PA}$	Emission cap in Pennsylvania
$\bar{E}_{RPJM}$	Emission cap in in Rest of PJM

4 Emissions Calculations

4.1 CO_2, SO_2, NOx

To assign CO_2, SO_2, NOx emission rate to each unit, we also perform a similar procedure as the procedure to assign heat rate above, but for emission rate. This means, first, we assign the CEMS's emission rates to the units that are in CEMS in the final dataset (374 units). Then we use the eGrid's threshold for emission rate (eGrid's average emission rate plus two times its standard deviation) to make a cut-off and assign emission rate for units in eGrid (1,519). We finally repeat the same same for SNL dataset, after which, all units are assigned emission rates. Summary of emission rate assignment is shown in table A.9.

We use quadratic regression to assign emission rates to the remaining units, after assigning emission rate of zero for units that have fuel-types of wind, hydro, solar and nuclear. R^2 for the regressions for CO_2 emission rates, SO_2 emission rates, and NOx emission rates of these remaining units are 0.84, 0.72 and 0.67 respectively.

$$\begin{aligned} emission - rate_i^{CO_2,SO_2,NO_x} &= \beta_0 + \mu(fueltype_i) + \beta_1(heatrate_i) + \beta_2(heatrate_i)^2 \\ &+ \beta_3(capacity_i) + \beta_4(capacity_i)^2 + \epsilon_i \end{aligned}$$

Table A.9: Emission Rate Assignments to Units in Final Dataset.

Dataset	Number of Units Assigned Emission Rate to
Final Dataset	3,095
CEMS	374
eGrid	1,519
SNL	1,201
EIA 923	—
NEEDS	—

4.2 PM, NH_3, VOC, CO

To impute emission rates for PM, NH_3, VOC, and CO, we follow the structure of the EPA's flat file methodology v514 with slight modifications regarding assignments of sulfur and ash contents. We first use sulfur content and ash content (generation weighted) by state and PJM region from the *"EPA 5-13_Base_Case RPE Replacement File"* from the IPM v513 base case output first, then we link the heat content with these sulfur and ash contents from the IPM v513 base case assumptions (Table 9-5). After that we use the flat file input (*PMAshSulfurContent* tab) to assign sulfur and ash contents for the units that are still not assigned ones from the previous step. Finally, we use table 12 in the flat file methodology to assign sulfur and ash contents for the remaining units.

To validate the above methodology of imputing emission rates, we use three data files - the NEEDS dataset, the *“Web-Ready_Parsed_File_EPA5-13_Base_Case_2018"* file and the *“FlatFile_EPA513_BC_7c_2018_* file. We first merge the NEEDS dataset with the *“Web-Ready_Parsed_File_EPA5-13_Base_Case_2018"* to link total annual generation to each unit in the NEEDS. Using the methodology of calculating emission rates established above, we take our calculated emission rates times the total annual generation to get the total annual emissions. Finally, we merge the combined data to the *“FlatFile_EPA513_BC_7c_2018_20131108"* which has total annual emissions by pollutants calculated by the EPA. We use these emission data to validate our calculated total emissions.

Our imputed unit-level total emissions match well with the emissions calculated by the EPA. For pollutants NH_3, VOC, CO, our imputed emissions are matched almost exactly with EPA's calculated emissions with R^2 of 0.99. For PM_{25}, and PM_{10}, our imputed emissions from natural gas units are matched almost exactly with EPA's calculated emissions, with $R^2 = 1$. For all fuel types, our imputed emissions are all matched well with EPA's calculated ones, with $R^2 > 0.91$.

5 Intertemporal Dynamics

5.1 Load Growth

Annual load growth rate for future load projection in the model is 0.42% increase from the previous year, which is the result of a linear forecast based on the actual loads of the past five years in PJM as reported by Market Monitoring Analytics in their State of the Market Reports 2013-2018.

5.2 Fuel Prices

Fuel prices are grown from the 2018 fuel price baseline. We use data for fuel price growth rates from the 2018 Annual Energy Outlook (AEO 2018) for the Mid-Atlantic region published by the U.S. Energy Information Administration (EIA). We apply price growth rates for natural gas, coal, oil and uranium and assume that prices for other fuel types stay the same as in 2018. Compared to the 2018 baseline, the fuel price growth rates from 2019-2030 used in our model are as follows:

Table A.10: Fuel Price Growth Rates 2019 - 2030 (in %).

	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Natural Gas	9.03	14.71	11.94	12.18	14.69	16.38	18.89	18.98	19.77	19.36	21.82	21.63
Coal	0.84	1.92	2.07	2.02	2.00	2.42	2.82	2.86	2.63	2.66	3.18	3.40
Distillate Oil	8.98	33.91	46.75	53.93	59.90	61.07	61.73	61.88	63.76	65.23	67.84	69.27
Uranium	0.15	0.46	0.62	0.77	1.08	1.23	1.54	1.69	2.00	2.16	2.31	2.62

5.3 External RECs

Observing the total number of RECs that are eligible to be supplied to PA and RPJM states in 2016 and 2017, we assume an annual growth rate of 4% in number of RECs that are allowed to be imported into PA and an annual growth rate of 5% in number of RECs that are allowed to be imported into RPJM states.

5.4 Capital Costs

For new natural gas unit's capital cost, we assume no growth rate, which means capital cost for new natural gas unit is still kept at $\$73,000$ MW-year. For new wind unit's capital cost, Wiser et al (2016) surveys different wind technology experts and finds that onshore wind costs would decline by 24% by 2030 relative to 2014, which is translated to an annual average reduction rate of 1.7%. For new solar PV system's capital cost, we look at the EIA Sunshot Initiative which has targeted a 50%reduction in utility scale solar PV costs from their 2020 target, which was achieved in 2017, to their 2030 target. We think this may be too ambitious so we cut it in half and assume a 25% reduction in PV costs by 2030 relative to 2017. This implies an annual average PV reduction rate in PV costs of 2.19%.

6 Additional Details on Pre-existing Policies

6.1 The Acid Rain Program

The Acid Rain Program (ARP) was established under Title IV of the Clean Air Act (CAA) Amendments, which established SO_2 and Nox emission allowance trading program for EGUs. To model this, we add onto the total marginal costs of EGUs regulatory costs of buying SO_2 and Nox emission allowance to comply with the ARP. These pre-existing regulatory costs are discussed in 3.2.2.

6.2 Nuclear Subsidies

In the states of Illinois and New Jersey, eligible nuclear EGUs are eligible for the Zero Emission Certificates (ZECs).

6.2.1 Nuclear Subsidies in New Jersey

In New Jersey, the program starts in 2019. The formula for calculation of ZEC price is by dividing the full recovery of all costs associated with the electric public utility's required procurement of ZECs at the end of the prior energy year by the greater of: 40 % of the total number of MWh of electricity distributed by the electric public utilities in the State in the prior energy year, or the number of megawatt-hours of electricity generated in the prior energy year by the selected nuclear power plants. New Jersey approved to subsidize 3 nuclear EGUs so far, which are Hope Creek and Salem 1 and 2, which total 28,441,726 MWh which is less than 40 % of total New Jersey's generation which is $0.4 * 74,908,250 = 29,963,300$ MWh.

In simpler terms, the formula for calculation of ZEC price is New Jersey is given by:

$$\frac{min\{p^{RP}g^{RS}; 3,000,000\}}{29,963,300}$$

where p^{RP} is electricity price charged to rate payers in \$/MWh in and g^{RS} is the total electricity sold to consumer in MWh. In 2019, $p^{RP} = \$4$ and $g^{RS} = 76,016,762$ MWh. Therefore, the total electricity retail sale in 2019 in New Jersey is $\$4 * 76,016,762 = \$304,067,048$, which exceeds the $\$3,000,000$ threshold. Therefore, the ZEC price in New Jersey in 2019 is $\frac{\$3,000,000}{29,963,300}$ = \$10.012.

6.2.2 Nuclear Subsidies in Illinois

In Illinois, ZEC price is \$16.5 unless the subtraction of the average of the Social Cost of Carbon (SCC) and difference between the average of the market price indices and the baseline market price index is less than zero. In that case, ZEC price is zero.

The average of the Social Cost of Carbon is \$16.5/MWh, the average of the market price indices taken from my model for year 2019 is \$35.35 and the baseline market price index is \$31.40 as defined in subparagraph (B) of paragraph (1) of this subsection (d-5) of the Illinois Power Agency Act: Section 1-75(d-5), known as the Zero Emission Standard. Since $16.5 - (35.35 - 31.40) = 12.55 > 0$. As a result, ZEC price in Illinois is \$16.5.

6.3 Renewable Portfolio Standards (RPS)

Many states in PJM have legislation on a defined percentage of supplied generation be served by renewable resources, for which definitions vary by state. These are called renewable portfolio standards (RPS). In 2018, there are 9 states that have RPS. They are Delaware, Illinois, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, and Washington, DC. There are two states that have voluntary RPS (Virginia and Indiana) and two states that do not have RPS (Kentucky, Tennessee and West Virginia). We only model RPS in the 9 states that have required RPS, of which 4 states (DC, MD, NJ and PA) classify their RPS into tier 1 and tier 2 RPS, under which different eligible renewable energy technologies are clearly defined in each state. The other 5 states (DE, IL, MI, NC and OH) do not classify their RPS into different tiers in 2017 but their eligible technologies are for the most part identical to tier 1 resources and will be modeled as tier 1 resource. By 2030, DC's tier 2 RPS will go down to 0 % and thus in 2030 only tier 1 RPS is modeled for DC. Details on RPS percentages and eligible technologies by tier in each state are shown in table below.

Table A.11: List of RPS regulations by State in PJM in 2017.

PJM State	State Number	RPS Tier 1	RPS Tier 2	Solar RPS
DC	1	13%	1.5%	0.98%
DE	2	16%	N/A	1.5%
IL	3	11.5%	N/A	6%
IN	4	N/A	N/A	N/A
KY	5	N/A	N/A	N/A
MD	6	13.1%	2.5%	1.15%
MI	7	10%	N/A	N/A
NC	8	6%	N/A	0.14%
NJ	9	13.5%	2.5%	3%
OH	10	3.5%	N/A	0.22%
PA	11	6%	8.2%	0.2933%
TN	12	N/A	N/A	N/A
VA	13	N/A	N/A	N/A
WV	14	N/A	N/A	N/A

Table A.12: List of RPS regulations by State in PJM in 2030.

PJM State	State Number	RPS Tier 1	RPS Tier 2	Solar RPS
DC	1	42%	0%	4.5%
DE	2	25%	N/A	3.5%
IL	3	25%	N/A	6%
IN	4	10%	N/A	N/A
KY	5	N/A	N/A	N/A
MD	6	20%	2.5%	2.5%
MI	7	35%	N/A	N/A
NC	8	12.5%	N/A	0.2%
NJ	9	50%	2.5%	2.21%
OH	10	12.5%	N/A	0.5%
PA	11	8%	10%	0.5%
TN	12	N/A	N/A	N/A
VA	13	N/A	N/A	N/A
WV	14	N/A	N/A	N/A

Table A.13: List of state-level RPS eligible technologies by tier.

State	Tier 1 RPS	Tier 2 RPS	Eligible Location
DC	(1) Solar PV, (2) solar thermal, (3) wind, (4) biomass (>65% efficiency),(5) methane from a landfill or wastewater treatment plant, (6) geothermal, (7) ocean including energy from waves, tides, currents, and thermal differences, (8) fuel cells that produces electricity from a Tier 1 renewable source.	(1) Hydroelectric power other than pump storage generation. The facility must have existed and been operational as of January 1, 2004	Sources must be located (1) within the PJM region or (2) an adjacent state to the PJM region or (3) outside the PJM region or adjacent state but in a control area that is adjacent to the PJM Region, if the electricity is delivered into the PJM Region.
DE	(1) Solar, (2) wind, (3) ocean, (4) geothermal, (5) fuel cell powered by renewable fuels, (6) combustion of gas from the anaerobic digestion of organic material, (7) small hydroelectric facility ($\leq$ 30 MW), (8) sustainable biomass excluding waste to energy, (9) landfill methane gas	(1) Units in commercial operation after 12/31/1997. No more than 1 percent of each year's sales may come from resources that are not new	Sources must be located (1) within or (2) imported into the PJM region.
IL	(1) Wind, (2) solar thermal energy, (3) PV cells and panels, (4) bio-diesel, (5) anaerobic digestion, (6) crops and untreated and unadulterated organic waste biomass, (7) tree waste, in-state landfill gas, (8) hydro-power that does not involve new construction or significant expansion of hydropower dams, (9) other alternative sources of environmentally preferable energy.		Sources must be located (1) in IL or (2) from adjoining states if approved by the Illinois Power Agency, or (3) within portions of the PJM and MISO footprint in the US.

Continued on next page

Table A.13 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	Eligible Location
IN* 4	(1) Solar energy, (2) PV cells and panels, (3) dedicated crops grown for energy production, (4) organic waste biomass, (5) hydro-power, (6) fuel cells, (7) hydrogen, (8) energy from waste to energy facilities including energy derived from advanced solid waste conversion technologies, (9) energy storage systems or technologies, (10) geothermal energy, (11) coal bed methane, (12) industrial byproduct technologies that use fuel or energy that is a byproduct of an industrial process, (13) waste heat recovery from capturing and reusing the waste heat in industrial processes for heating or for generating mechanical or electrical work, (14) landfill methane recovery, (15) demand side management or energy efficiency initiatives, (16) a clean energy project described in the statute, (17) nuclear energy, (18) distributed generation connected to the grid, (19) combined heat and power, (20) electricity that is generated from natural gas at a facility constructed in Indiana after July 1, 2011 which displaces electricity generation from an existing coal fired generation facility.		At least 50 percent of RECs must be purchased from resources located within Indiana.
KY	No RPS.		

Continued on next page

Table A.13 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	Eligible Location
MD	(1) Solar, (2) wind, (3) qualifying biomass, (4) methane from a landfill or wastewater treatment plant, (5) geothermal, (6) ocean, (7) fuel cell powered by methane or biomass, (8) small hydroelectric plant (< 30 MW), (9) poultry litter incineration facilities in Maryland, (10) waste-to-Energy facilities in Maryland, (11) certain geothermal heating and cooling systems and biomass systems that generate thermal energy.	(1) Hydroelectric power other than pumped storage generation	Source must be (1) located in the PJM Region; or (2) outside the area described in item (1) but in a control area that is adjacent to the PJM service territory, if the electricity is delivered into the PJM service territory. Solar resources must be connected to the distribution grid serving Maryland.
MI	(1) Biomass, (2) solar PV, (3) solar thermal, (4) wind, (5) geothermal, (6) municipal solid waste (MSW), (7) landfill gas, (8) existing hydroelectric, (9) tidal, wave, and water current (e.g., run of river hydroelectric) resources.		Resources must be located within Michigan or anywhere in the service territory of retail electric provider in Michigan that is not an alternative electric supplier. There are many exceptions to these requirements.
NC	(1) Solar-electric, (2) solar thermal, (3) wind, (4) hydro-power ($\leq$ 10 MW), (5) ocean current or wave energy, (6) biomass that uses Best Available Control Technology (BACT) for air emissions, (7) landfill gas, (8) combined heat and power (CHP) using waste heat from renewables, (9) hydrogen derived from renewables, (10) and electricity demand reduction. Up to 25% of the requirement may be met through energy efficiency technologies, including CHP systems powered by non-renewable fuels. After 2021, up to 40% of the standard may be met through energy efficiency.		Dominion, the only utility located in both the state of North Carolina and PJM, may purchase RECs from anywhere.

Continued on next page

Table A.13 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	Eligible Location
NJ	(1) Solar technologies, (2) PV technologies, (3) wind energy, (4) fuel cells powered by renewable fuels, (4) geothermal technologies, (5) wave or tidal action, (6) methane gas from landfills or a biomass facility provided that the biomass is cultivated and harvested in a sustainable manner, (7) hydroelectric facilities ($\leq$ 3 MW) that are located in NJ and placed in service after July 23, 2012.	(1) Resource recovery facility (subject to qualifications), (2) small hydroelectric power facility (< 30 MW)	Source must be (1) within or (2) delivered into the PJM region. If the latter, the energy must have been generated at a facility that commenced construction on or after January 1, 2003
OH	(1) Solar photovoltaics (PV), (2) solar thermal technologies used to produce electricity, (3) wind, (4) geothermal, (5) biomass, (6) biologically derived methane gas, landfill gas, certain non-treated waste biomass products, (7) solid waste (as long as the process to convert it to electricity does not include combustion), (8) fuel cells that generate electricity, certain storage facilities, and qualified hydroelectric facilities, (9) certain co-generation and waste heat recovery system technologies that meet specific requirements, (10) distributed generation systems used by customers to generate electricity using the aforementioned eligible renewable resources, (11) run-of-the-river hydroelectric systems on the Ohio River ($>$ 40 MW).		Source must be (1) in-state facilities or (2) can be shown to be deliverable into the state.

Continued on next page

Table A.13 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	Eligible Location
PA	(1) Solar PV and solar thermal energy, (2) wind power, (3) Low-impact hydro-power, (4) geothermal energy, (5) biologically derived methane gas, (6) generation of electricity utilizing by-products of the pulping process and wood manufacturing process including bark, wood chips, sawdust and lignin in spent pulping liquors (in-state resources only), (7) biomass energy, (8) coal mine methane.	(1) Waste coal, (2) distributed generation systems, (3) demand-side management, (4) large-scale hydro-power (including pumped storage), (5) municipal solid waste, (6) generation of electricity utilizing by-products of the pulping process and wood manufacturing process including bark, wood chips, sawdust and lignin in spent pulping liquors, (7) integrated combined coal gasification technology.	Source must be (1) located inside the geographical boundaries of this Commonwealth or (2) within the service territory of any regional transmission organization that manages the transmission system in any part of this Commonwealth.
TN	No RPS		
VA*	(1) Solar, (2) wind power, (3) geothermal energy, (4) hydro-power, (5) wave, (6) tidal, (7) biomass energy.		Electricity must be generated or purchased in (1) Virginia or (2) in the PJM service territory.
WV	No RPS		

End of long table.

Notes: * States with voluntary RPS

7 Model Validation

We validate our model using 2018 data across a few categories: load, generation mix, capacity expansion, CO_2 emissions, weighted REC prices and region-level Locational Marginal Prices (LMPs). We use PJM's actual total hourly load in 2018 and actual fuel prices from the same sources described in 3.2.2 for year 2018. Our model predicted load matches actual 2018 load with an $R^2 = 0.948$.

7.1 Generation Mix Validation

We validate the generation mix from out model using data on actual generation by fuel type reported by the Monitoring Analytics in their 2018 State of the Market report. The validation results are shown in the table below.

Table A.14: Validation Results for Total Generation in PJM in 2018 by fuel type (GWh).

Fuel	PJM-Actual	PJM-Model
Coal	239,612	229,691
Nuclear	286,115	287,678
Gas	256,701	267,483
Hydro	19,416	19,441
Wind	21,628	22,596
Oil+Other	3,581	2,480
Solar	2,111	3,192
Biomass	8,390	9,044
Total	837,594	841,605

7.2 Capacity Expansion Validation

Table A.15: Validation Results for Capacity Expansion (MW) and New Generation (GWh) in Pennsylvania and Rest of PJM in 2018 by technology.

	PA-Actual	PA-Model	RPJM-Actual	RPJM-Model
New Capacity				
NGCC	5,112	3,407	4,330	4,443
Wind	0	0	762	987
Solar	0	0	265	804
New Generation				
NGCC	43,093	28,002	36,455	36,593
Wind	0	0	1,969	2,550
Solar	0	0	395	1,198

7.3 CO_2 Emissions Validation

Total PJM's CO_2 Emission for 2018 is taken from the Monitoring Analytics' 2018 State of the Market, which reports total CO_2 Emission of 370.1 MMT. Using CO_2 emission data from 2018 EIA's State Electricity Profiles for Pennsylvania of 69.9 MMT, we can impute rest of PJM's 2018 CO_2 emission of 300.2 MMT. Our model predicts CO_2 emission 60.9 MMT in Pennsylvania, and of 291.0 MMT in rest of PJM, totaling 351.3 MMT, which yields a 5% error. We slightly under-predict CO_2 emissions in both Pennsylvania and rest of PJM.

7.4 REC prices Validation

Table A.16: Validation Results for Generation Weighted REC prices in Pennsylvania and Rest of PJM in 2018.

	PA-Actual	PA-Model	RPJM-Actual	RPJM-Model
Tier 1 Price	\$4.35	\$4.60	\$46.00	\$36.39
Tier 2 Price	\$12.45	\$13.17	\$3.99	\$3.18
SREC Price	\$0.0028	\$0.0029	\$0.48	\$0.38

7.5 Locational Marginal Prices (LMPs) Validation

Figure A.20: Validation Results for Day-Ahead LMP in West Pennsylvania (Top Row, Left), East Pennsylvania (Top Row, Right), East RPJM (Second Row, Left), Central RPJM (Second Row, Right) and West RPJM (Bottom Row) in 2018. $R^2 = 0.28$