Appendix

Modeling the Impact of RGGI on Pennsylvania's Power Grid:

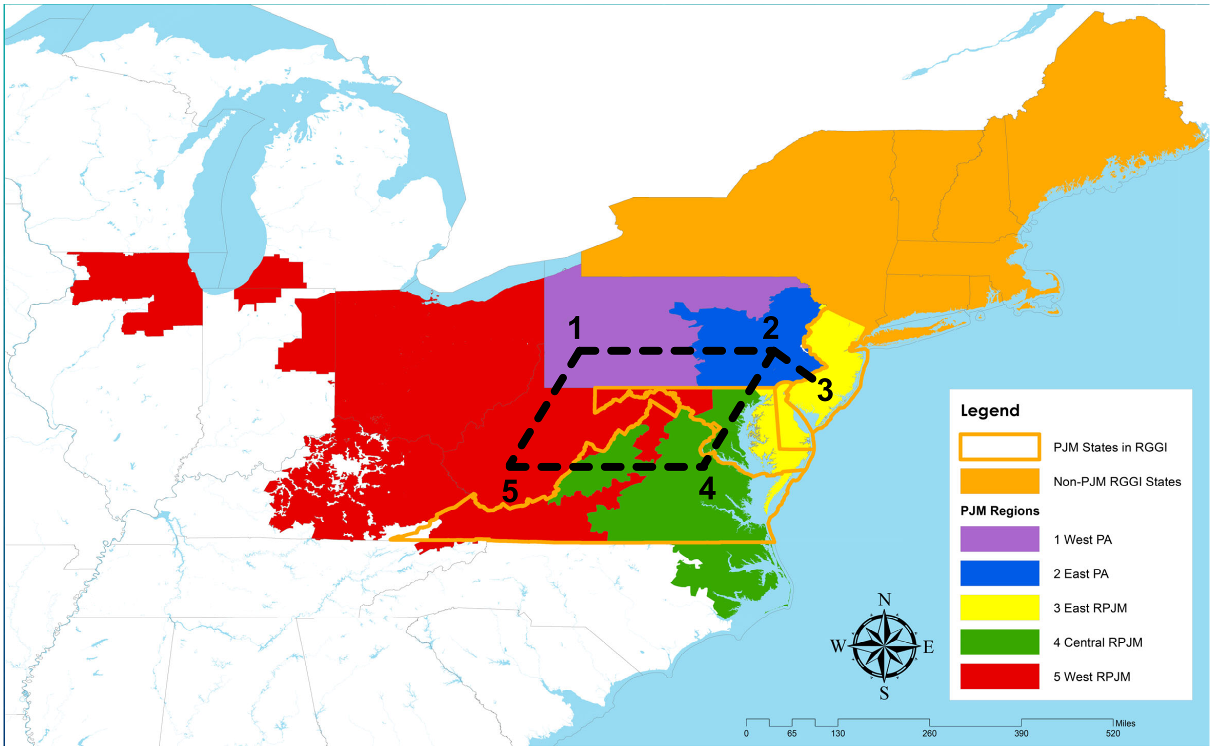
Costs, Emissions, and Leakage

Section A provides a detailed overview of the simulation model. Section B presents the functional forms used in the simulation model. Section C discusses the parameter values and data sources used for calibration, the assumptions and data sources used to generate emissions estimates, and our model calibration results. In Section D we outline the assumptions regarding the dynamic trends that underlie our simulation results. Section E validates our baseline against historical data. Pre-existing polices that we account for in our analysis are described in Section F. Finally, Section G presents additional results not reported in the main text.

A Model Overview

The RGGI+PJM Policy Analysis Model (RPAM) is a multi-market numerical simulation model that combines: 1. a transportation model of the PJM power system; 2. the endogenous supply of new generation capacity within PJM; 3. the importation of Renewable Energy Credits (RECs) from outside of PJM; 4. the supply of CO_2 abatement from non-PJM Regional Greenhouse Gas Initiative (RGGI) states; and 5. the supply/demand of banked CO_2 allowances from current RGGI market participants. Parts 1-3 of the model are 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. Parts 4 and 5 are jointly econometrically estimated using historical data on emissions, caps, permits sold at auction, and allowance price data from RGGI. RPAM operates on an annual time-step and we simulate outcomes from 2018 to 2030. The RPAM domain is depicted in Figure A.1.

Figure A.1: RGGI+PJM Policy Analysis Model (RPAM) Domain



As shown in Figure A.1, RPAM consists of five regions (purple, blue, yellow, green, and red) which comprise the wholesale electricity market operated by PJM as well as the CO_2 emissions released by RGGI states that are not in PJM (solid orange). In addition, a subset of states are members of RGGI and are also wholly or partly contained with PJM's system boundaries (orange outline) and allowances added to the RGGI allowance bank prior to 2022 are assumed to be held by market participants within all current RGGI states (solid orange or outline orange).

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. RPAM also has 96 load segments, constructed from 8,784 hours in 2016 (a leap year) and 8,760 hours in 2017. Within each PJM region, we assume that demand for electricity is partially inelastic across 96 load segments. Details on how we construct the five aggregate regions, 96 load segments, demand elasticity and transmission lines are discussed in Sections B.5 and C.1.

The supply side of the PJM wholesale electricity market considers the economic decisions of 843 representative existing generation units (EGUs) which have been aggregated from a population of 3,095 EGUs located within PJM and whose locations are known. Fuel costs facing EGUs 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. RPAM also allows for endogenous new capacity expansion in natural gas combined cycle (NGCC), solar, and wind. Details on calibrations for the supply side can be found in Section C.2.

B Functional Forms

B.1 Regional Demand for Electricity

The total benefit or willingness to pay by load serving entities in region i at any given hour in load segment l are given by:

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

where d_{il} is the electricity demand in region i at a representative hour in load segment l, and n_{il} and c_{il} are the slope and intercept, respectively of the linear inverse demand curve implied by the first-order conditions for the numerical model with respect to d_{il}: $p_{il} = c_{il} - n_{il}d_{il}$, where p_{il} is the price of electricity in region i at load segment l.

B.2 Supply of Electricity from Existing Generation

The variable costs from existing representative electric generating unit (EGU) j located in region i and supplying electricity at any given hour in load segment l are given by:

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

where g_{jl}^{E} is the electricity supplied by existing EGU j in a representative hour in load segment l, m_{jl}^{E} and b_{jl}^{E} are, respectively, the slope and intercept of the linear inverse supply curve for existing EGU j in region i at any given hour in load segment l implied by the first-order conditions for the numerical model with respect to g_{jl}^{E}: $p_{il}=b_{jl}^{E}-m_{jl}^{E}g_{jl}^{E}$. Variable costs include fuel costs, operation and maintenance costs, and, for some existing EGUs, the costs of complying with Title IV of the Clean Air Act and state nuclear subsidies.

In addition, each representative existing EGU j has a limit on the amount of power it can supply in a representative hour in each load segment l reflecting its available effective capacity:

$$g_{jl}^{E}\leq K_{jl}^{E}, \tag{3}$$

where K_{jl}^{E} is EGU j's effective capacity at any given hour in load segment l.

B.3 New Capacity and Supply of Electricity from New Generation

Each year, we allow as many as 42 ($=3\times 14$) new EGUs j to be added to the model. In particular, we consider three different technologies that can be added in each of the 14 states within PJM in each year. The technologies we consider are NGCC, wind, and solar, which we believe are the technologies most likely to be added in the next few years due to current market and regulatory drivers. We allow these to enter by state since state Renewable/Alternative Energy Portfolio Standards (RPSs) are important determinants of renewable expansion across states within PJM. The variable costs from operating a new EGU j and supplying electricity at any given hour in load segment l are given by:

$$VC_{jl}^N\left(g_{jl}^N\right) = b_{jl}^N g_{jl}^N, \tag{4}$$

where g_{jl}^N is the electricity supplied by new EGU j in a representative hour in load segment l, b_{jl}^N is the private marginal operating cost for new EGU j in a representative hour in load segment l.

Additionally, each new EGU j has a limit on the amount of electricity it can produce at any given hour in a load segment, reflecting its available effective capacity:

$$g_{jl}^N \leq K_j^N, \tag{5}$$

where K_j^N is the effective capacity of new EGU j at any given hour. This is calculated as follows: $K_j^N = \bar{K}_j^N \gamma_j^N$, where $\bar{K}_j^N$ is the total amount of new nameplate capacity that is expanded for new EGU j, measured in MW, and γ_j^N is the utilization factor for new EGU j.[5]

Finally, new generation also accrues a capacity cost from adding each MW of new capacity for EGU j in a given year that is given by:

$$CC_j^N\left(\bar{K}_j^N\right) = C_j^N \bar{K}_j^N, \tag{6}$$

where C_j^N is the annual cost of adding one MW of new capacity to new EGU j per year. These annualized fixed costs accrue in years subsequent to the vintage year in which new capacity is added. Capacity costs includes the purchase costs of capital, the costs from purchasing or leasing land, financing costs, search costs, and all costs associated with new capacity permitting and approval.

[5]Some states have EGUs that span multiple load regions in the model (Pennsylvania, Maryland, New Jersey, and Virginia). For these states we assume that generation from new EGUs is allocated between the multiple regions in proportion to the 2016 share of a state's generation within each load region of which it is a member to that state's total generation.

B.4 External Renewable Energy Certificates

We allow Renewable Energy Certificates (RECs) generated by EGUs outside of PJM to be used for compliance with various state Alternative Energy/Renewable Portfolio Standards (RPSs), which often have multiple tiers, within PJM. Let r_{qst} be the amount of tier t RECs that an EGU q outside of PJM supplies to state s in PJM in a given year. The total amount of external RECs that EGU q can supply in a given year across all the states with RPSs within PJM must satisfy:

$$\sum_{st \in \mathcal{ST}_q} r_{qst} \leq r_q, \tag{7}$$

where r_q is the total amount of external RECs that external EGU q can supply to a subset of states with RPSs in PJM in a given year across all tiers and $\mathcal{ST}_q$ is the subset of states and tiers for which external EGU q can supply external RECs. $r_q = \bar{r}_q \gamma_q$, where $\bar{r}_q$ is the total amount of external RECs that external EGU q can supply to all states (including states that are outside of PJM) in a given year, and γ_q is the percentage of $\bar{r}_q$ that indicates the total amount of external RECs that external EGU q can supply only to states in PJM.

The constraints in (7) ensure that external RECs are distributed to states within PJM with the highest REC prices, second highest, and so on, until r_q is exhausted. After the model is solved, aggregate surplus by state is adjusted by the costs of external RECs purchased by the state in light of the REC prices predicted by the model.

B.5 Transmission Network

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

$$-\bar{f}_{ihl} \leq f_{ihl} \leq \bar{f}_{ihl} \tag{8}$$

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 effective transmission capacity between region i and h at load segment l, all of which are measured in MWh. The sign indicates the direction of the power flow, with a negative sign denoting power is flowing away from i and a positive sign that power is flowing into i. $\bar{f}_{ihl}$ is calculated as follows: $\bar{f}_{ihl} = V_{ih}A_{ihl}$, where V_{ih} is the sum of voltage (measured in volts) across all transmission lines connecting region i and region h, and A_{ihl} is the effective current between region i and region h during load segment l in amperes.

B.6 CO_2 Emissions from RGGI States Outside of PJM

The total emissions benefit from covered EGUs in RGGI states that are not in PJM is given by:[6]

$$TB^{NPJM}\left(E^{NPJM}\right) = c^{NPJM}E^{NPJM} - \frac{n^{NPJM}}{2}\left(E^{NPJM}\right)^2, \tag{9}$$

where E^{NPJM} is the CO_2 emissions from covered EGUs in RGGI states outside of PJM, n^{NPJM} and c^{NPJM} are the slope and intercept, respectively, of the linear inverse demand curve for covered emissions from non-PJM RGGI states in a given year implied by the first-order condition with respect to E^{NPJM}: $p^{RGGI} = c^{NPJM} - n^{NPJM}E^{NPJM}$, where p^{RGGI} is the RGGI allowance price in that year.

B.7 Supply/Demand of Banked Allowances

The total benefit to holders of allowances in a given year from allowances that have been banked from all previous years is given by:

$$TB^{B}\left(B, \bar{B}\right) = c^{B}\left(\bar{B}\right)B - \frac{n^{B}}{2}\left(B\right)^2, \tag{10}$$

[6]Observe that if the marginal covered emissions benefit (the derivative of (9) with respect to E^{NPJM}) is set equal to marginal costs of covered emissions of zero, a ceteris paribus unregulated or baseline covered emissions level can be obtained, E_0^{NPJM}. Abatement can then be defined as $A^{NPJM} = E_0^{NPJM} - E^{NPJM}$ and the marginal costs of abatement from covered EGUs in non-PJM RGGI states can be defined as: $\left(c^{NPJM} - n^{NPJM}E_0^{NPJM}\right) + n^{NPJM}A^{NPJM}$. We report (9) here as E^{NPJM} is what is solved for in the model and because the marginal benefit of covered emissions is what we directly estimate from historical data as further discussed below.

where B is the total amount of RGGI allowances in the RGGI allowance bank in a given year, n^B and $c^B \equiv c^B\left(\bar{B}\right)$ are the slope and intercept, respectively, of the linear inverse demand curve for RGGI banked allowances, implied by the first-order condition for the numerical model with respect to B: $p^{RGGI} = c^B - n^B B$. As discussed further below, n^B is fixed across years whereas $c^B \equiv c^B\left(\bar{B}\right)$ is a function of the bank account balance at the end of the previous year, $\bar{B}$. Permits are withdrawn from (added to) the bank in a given year if $\bar{B} - B > 0$ ($\bar{B} - B < 0$). The starting 2020 value for $\bar{B}$ is calculated from historical RGGI auction and cap data prior to 2020. After 2020, the prior bank account balance is defined as: $\bar{B} = B_{y-1}$, where B_{y-1} is the bank account balance at the end of the previous year.

B.8 Market Clearing Conditions

B.8.1 Electricity Market

The electricity market clears in region i at a representative hour in load segment l when the following constraint is satisfied:

$$\sum_{j\in\mathcal{J}_i^E} g_{jl}^E + \sum_{j\in\mathcal{J}_i^N} g_{jl}^N + \sum_{h\in\mathcal{L}_i} f_{ihl} \geq d_{il}\left(1+\epsilon_l\right) \quad \text{for all } i, l, \tag{11}$$

where $\mathcal{J}_i^E$ is the set of all existing EGUs in PJM region i, $\mathcal{J}_i^N$ is the set of all new EGUs in PJM region i, $\mathcal{L}_i$ is the set of all nodes that are connected to i, and ϵ_l is aggregate loss in any given hour in load segment l as a percentage of demand in region i for that hour, which reflects transmission and distribution system losses as well as the difference in virtual increment offers and decrement bids, as discussed further below.

B.8.2 Renewable Energy Credit Markets

Several states s within PJM have multiple tiers t (e.g., tier 1, tier 2 and solar RPS) of Renewable/Alternative Energy Portfolio Standards (RPS) which mandate that at least a certain fraction of generation in a given year come from 'numerator EGUs' (eligible EGUs) for a given state-tier st relative to total generation in that state. Similar to other mandate and trade systems (e.g., the Renewable Fuel Standard, Corporate Average Fuel Economy Standards), state-tier RPSs allow for restricted trade in virtual renewable energy credits (RECs) among market participants that are eligible (and in some cases, required) to participate under each state-tier RPS st. The REC market clears for each RPS state-tier combination st in a given year according to:

$$\frac{\sum_{j \in \mathcal{J}_{st}} \sum_l \delta_l g_{jl} + \sum_{q \in \mathcal{Q}_{st}} r_{qst}}{\sum_{j \in \mathcal{J}_s} \sum_l \delta_l g_{jl}} \geq \bar{R}_{st}, \tag{12}$$

where δ_l is number of hours in load segment l, g_{jl} (no superscript) is generation from new or existing EGU j in a representative hour in load segment l, $\bar{R}_{st}$ is the RPS target of tier t in state s in a given year, $\mathcal{Q}_{st}$ is the subset of external EGUs which can supply external RECs for compliance with the st RPS standard, $\mathcal{J}_{st}$ is the subset of eligible new and existing EGUs in PJM under state-tier RPS constraint st, and $\mathcal{J}_s$ is the subset of eligible and ineligible new and existing EGUs for each state s (which is common across tiers for all states in PJM and includes all the existing and new EGUs in that state); see Section F.5 for further details.

B.8.3 RGGI Allowance Market

So far all states that have joined RGGI have chosen to allow inter-state allowance trading among RGGI market participants. As a result there is a single market for RGGI allowances which clears in a given year according to:

$$\sum_{j \in \mathcal{J}_{RGGI}} \sum_l \delta_l \phi_j^{\mathrm{CO}_2} g_{jl} + E^{NPJM} + \left(B - \bar{B}\right) \leq \sum_{s \in \mathcal{S}_{RGGI}} \bar{E}_s^{RGGI}, \tag{13}$$

where $\bar{E}_s^{RGGI}$ is the adjusted allowance budget assigned to state s consistent with its membership in RGGI, $\mathcal{J}_{RGGI}$ is the subset of covered new and existing EGUs in PJM that are also in states that are members of RGGI, $\mathcal{S}_{RGGI}$ is the subset of states that are members of RGGI in a given year, and $\phi_j^{\mathrm{CO}_2}$ is the CO_2 emissions factor of existing or new EGU j.

B.8.4 Characterization of the Competitive Equilibrium

There are multiple ways to numerically solve for the competitive equilibrium solution which effectively reflects: 1. the PJM system operator's optimal hourly dispatch decision reflecting a transportation model; 2. the decentralized annual power system capacity investment equilibrium;

3. the decentralized REC equilibrium; and 4. the decentralized RGGI allowance market equilibrium, conditional on all other (exogenous) pre-existing policies. According to the First Fundamental Welfare Theorem and the assumptions of our model the most direct solution method involves solving 1 across all hours of the year, but allowing for new capacity investment and conditional on market clearing in the REC and RGGI allowance markets. In this case the objective function to maximize is:

$$\sum_i \sum_l \delta_l \left(c_{il} d_{il} - \frac{n_{il}}{2} (d_{il})^2 \right) - \sum_j \left[C_j^N \bar{K}_j^N + \sum_l \delta_l \left(b_{jl}^E g_{jl}^E + \frac{m_{jl}^E}{2} \left(g_{jl}^E\right)^2 + b_{jl}^N g_{jl}^N \right) \right] + \left[c^{NPJM} E^{NPJM} - \frac{n^{NPJM}}{2} \left(E^{NPJM}\right)^2 \right] + \left[c^B B - \frac{n^B}{2} (B)^2 \right] \tag{14}$$

given (1), (2), (4), (6), (9), and (10). Thus, a *competitive equilibrium* is the quantities and (shadow) prices that are returned from maximizing (14) (choosing d_{il} for all i, l, g_{jl}^E for all j, l, g_{jl}^N for all j, l, r_{qst} for all q, s, t, $\bar{K}_j^N$ for all j, E^{NPJM}, B and f_{ihl} for all i, h, l) subject to (3) for all j, l, (5) for all j, l, (7) for all q, s, t, (8) for all i, h, l, (11) for all i, l, (12) for all s, t, (13), and non-negativity constraints on d_{il}, g_{jl}^E, g_{jl}^N, r_{qst}, $\bar{K}_j^N$, E^{NPJM}, and B.

The *competitive equilibrium prior to Pennsylvania joining RGGI*, which provides a common baseline prior to 2022 and serves as our baseline counterfactual between 2022 and 2030, is the competitive equilibrium as defined above where, in equation (13), the set $\mathcal{J}_{RGGI}$ includes covered existing and new EGUs within RGGI member states prior to Pennsylvania's entry and the the set $\mathcal{S}_{RGGI}$ includes RGGI member states prior to Pennsylvania's entry. The set of RGGI member states differs across time. Prior to 2020, RGGI member states include: Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New York, Delaware, and Maryland. In 2020 New Jersey rejoins RGGI and in 2021 Virginia also joins.

The *competitive equilibrium with Pennsylvania joining RGGI* is the competitive equilibrium as defined above where, in equation (13), the set $\mathcal{J}_{RGGI}$ includes covered existing and new EGUs within RGGI member states, including Pennsylvania, the set $\mathcal{S}_{RGGI}$ also includes Pennsylvania beginning in 2022.

C Data and Calibration

In this section, we discuss the data and intermediate steps to calibrate the parameters. We calibrate our model primarily using 2016 and 2017 historical data. Since we use up-to-date EGUs data in 2016 from various data sources, our model does not include capacity expansion in 2016. In 2017, our model allows capacity expansion of new NGCC, wind and solar EGUs to be added to the set of EGUs that can be dispatched by PJM. Finally, the non-PJM RGGI marginal benefits from emissions and the marginal benefits of banked allowances are estimated using more recent data from the inception of RGGI in 2009 through 2019. The parameters of the model can be divided into three types: 1. those that are *analytically calibrated* given analytic expressions and collected data, 2. those that are *estimated* given statistical models and collected data, and 3. those that are *numerically calibrated* by minimizing the Euclidean distance between model predictions and collected data. In this section, we begin by explaining all of the parameters that are either analytically calibrated or estimated for each of the sectors/agents specified in the proceeding section. Following, this we discuss the multi-step numerical calibration method we use to recover the remaining model parameters. We then discuss how emissions are calculated from predicted model output. Finally, we compare the results from our calibrated baseline for the years 2016 and 2017 against observed 2016 and 2017 data.

C.1 Regional Demand for Electricity

C.1.1 Load Regions

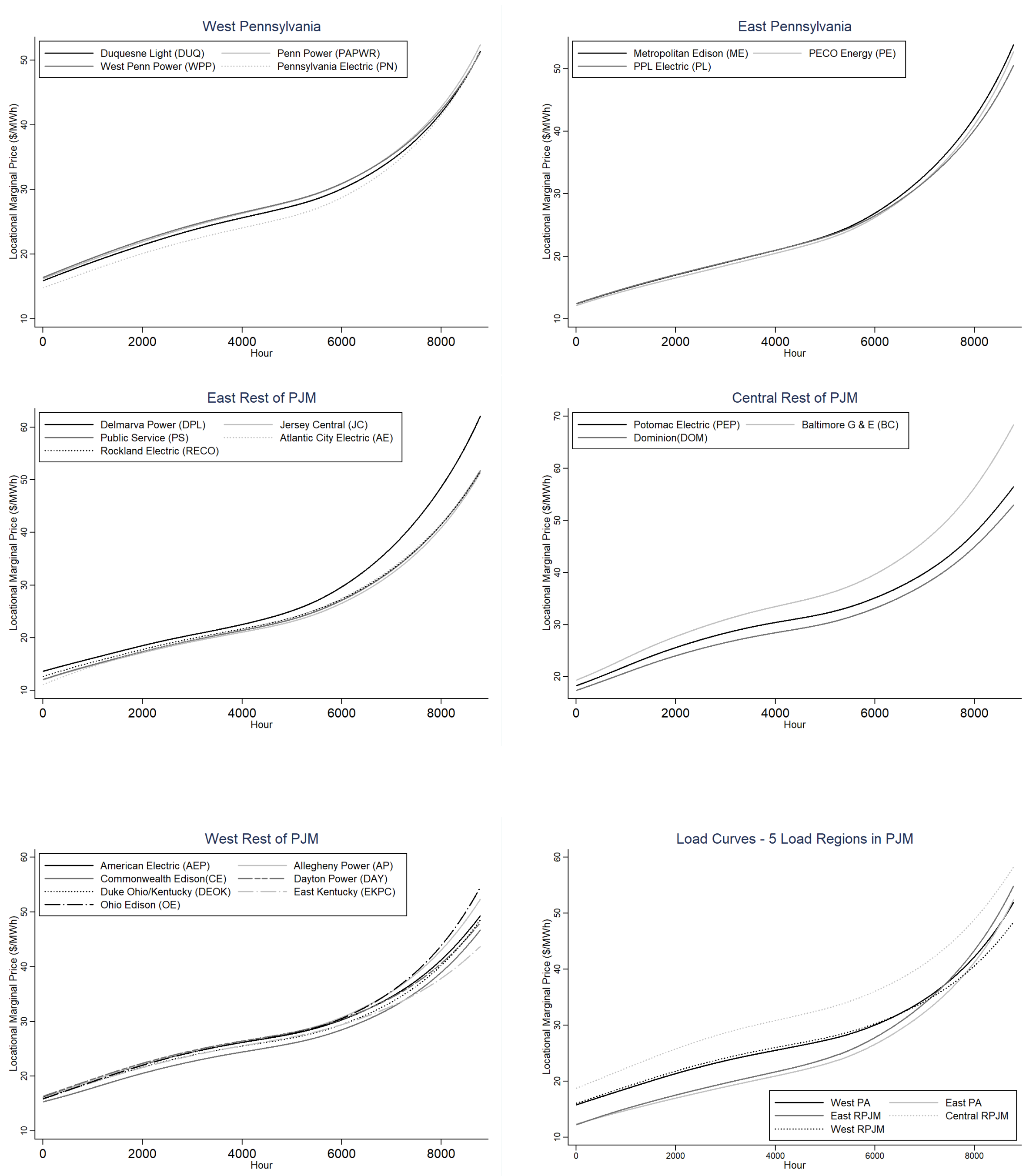
The five regions that comprise PJM in RPAM are constructed by aggregating load zones in PJM based on similarity in day-ahead hourly locational marginal prices (LMPs) in 2016 from the *Day-Ahead Hourly LMPs for 2016* dataset (last accessed 05/15/2017 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/da_hrl_lmps), and geographical proximity. These five regions are: 1. West Pennsylvania (purple coloration in Figure A.1), 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 (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.

The first five panels of Figure A.2 report the results of locally weighted regressions (assuming a bandwidth of 0.8) of hours in the year sorted from lowest to highest load (see below for data on load) against day-ahead hourly LMPs for each of the original load zones that form the five RPAM PJM regions. Panel A shows that the average LMP in West Pennsylvania in the lowest demand hour is as low as approximately \$15/MWh and the average LMP in the highest demand hour is more than \$50/MWh. The median LMP is around \$27/MWh. Similarly, panel B shows that the average LMP in the lowest demand hour is close to \$5 per MWh and average LMP in the highest demand hour is over \$50. The median LMP is approximately \$22. Overall, West Pennsylvania observes higher LMPs than East Pennsylvania, although there is some convergence in LMPs during high demand periods. Panel C shows that the average LMP in the lowest demand hour is as low as \$11 per MWh and the average LMP 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. Panel D shows that LMPs in the Central RPJM region are the most spread out with an average LMP in the low demand hour of approximately \$18 per MWh and an average LMP near \$69 in the highest demand hour. The median LMP lies just above \$30. Finally, panel E shows that the average 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.

Panel F of Figure A.2 reports the results of locally weighted regressions (assuming a bandwidth of 0.8) of hours in the year sorted from lowest to highest load against load weighted average day-ahead hourly LMPs for the final five PJM regions in RPAM. Central RPJM has higher average LMPs 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—across 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.2: Day-Ahead locational marginal prices in 2016 Across PJM. Panels from top-left: A. West Pennsylvania, B. East Pennsylvania, C. East Rest of PJM, D. Central Rest of PJM, E. West Rest of PJM, and F. Load Weighted Average for Final Five PJM RPAM Regions



C.1.2 Load Segments

We consider 96 load segments consisting of representative hours that allow us to capture inter-temporal variability in demand across and within seasons and the correlation of demand with natural gas prices across days of the year. Load segments are defined using a three step-process.

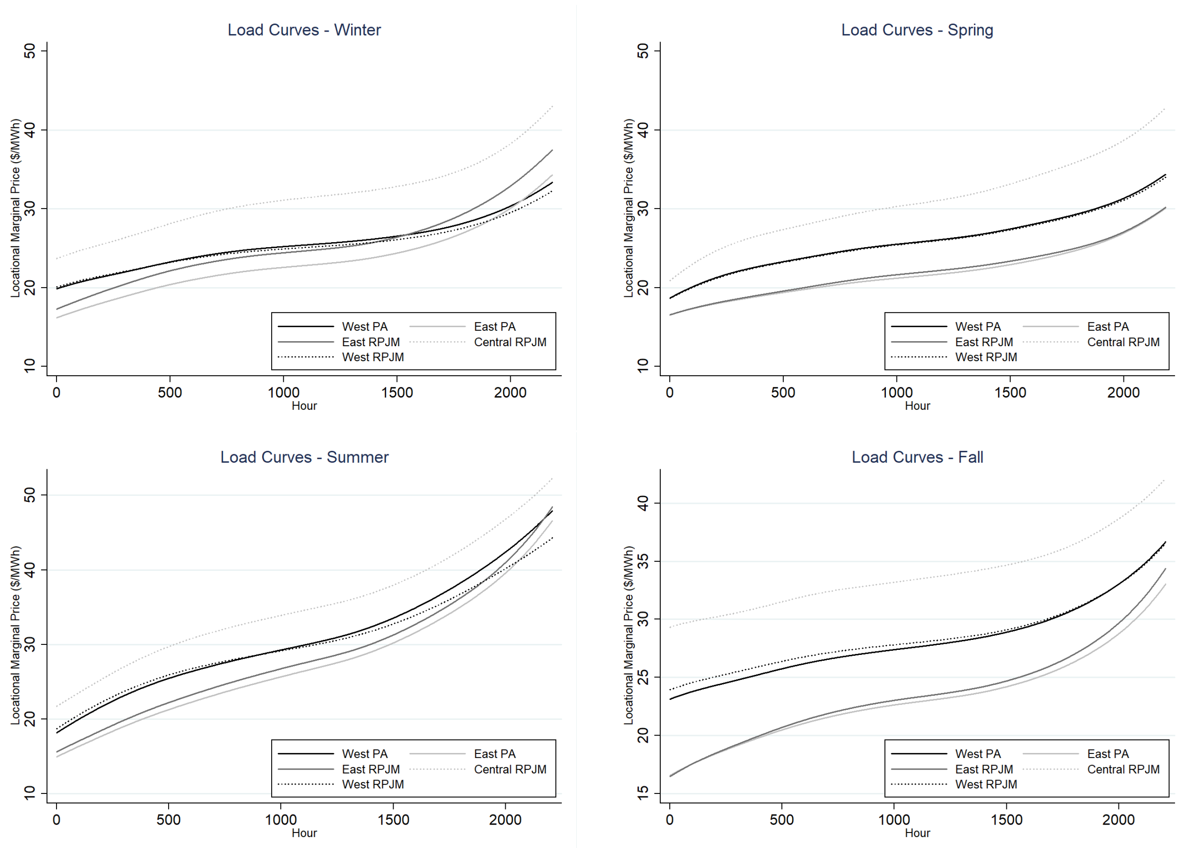
In the first step, we divide the 8,784 hours in 2016 into four seasons: Winter (December 20 to March 21), Spring (March 22 to June 20), Summer (June 21 to September 20) and Fall (September 21 to December 21).

In the second step, we divide each season into six bins based on the variation in load over each season. To this end, we first calculate hourly metered load for all of PJM. This is the sum of hourly metered load from the original load regions located within PJM, where load data for the original zones comes from the *Hourly Load: Metered for 2016* dataset (last accessed 05/15/2017 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/hrl_load_metered). We then sort total PJM hourly metered load from lowest to highest across the all hours within each season to construct seasonal Load Duration Curves. For each season, we define all hours that are within the top 1% of total PJM hourly metered load as the first bin. The second bin for each season is defined as all hours between the 2nd and 5th percentiles of total PJM hourly metered load. The third bin for each season is defined as all hours between the 6th and 15th percentiles of total PJM hourly metered load. The fourth bin for each season is defined as all hours between the 16th and 45th percentiles of total PJM hourly metered load. The fifth bin for each season is defined as all hours between the 46th and 75th percentiles of total PJM hourly metered load. Finally, the sixth bin for each season is defined as all hours between the 76th and 100th percentiles of total PJM hourly metered load. This partition of within seasonal variation in load is the same as IPM (2013), except here we consider four seasons and not two, and we go one step further.

Finally, in the third step we divide each of these 24 bins into four final segments. To this end, we sort the PJM average hourly natural gas spot price from lowest to highest within each of the 24 prior bins. The PJM average hourly natural gas spot price equals the PJM average daily natural gas spot price for which an hour is matched. The PJM average daily natural gas spot price equals the average across all existing natural gas EGU's within PJM of the daily natural gas spot price that has been spatially linked to that EGU, as discussed further below. For each bin, the first segment is defined as those hours with the 10% of highest prices. The second segment in each bin is defined as those hours with prices between the 11th and 30th percentiles. The third segment in each bin is defined as those hours with prices between the 31st and 60th percentiles. Finally, the fourth segment in each bin is defined as those hours with prices between the 61st and 100th percentiles. This leaves us with 96 ($= 4 \times 6 \times 4$) final load segments, where each load segment maps several (possibly non-temporally contiguous) to specific hours of 2016. This one-to-many mapping between hours and load segments is used across all years (although for non-leap years, February 29 does not occur and there are 24 fewer hours in the Winter season).

Given the final construction of the 96 load segments, we can define the sets $\mathcal{H}_l$ for each l which specifies the hours in 2016 that map to load segment l. The size of each of these sets equals δ_l. Figure A.3 depicts the hours within each season sorted from lowest to highest load-weighted average Locational Marginal Price in 2016, the latter which is calculated as described in the previous section. As expected, Summer has the highest LMPs, followed by Winter. Spring and Fall are shoulder months and thus have lower LMPs.

Figure A.3: Day-Ahead locational marginal prices by Season and Five PJM RPAM Regions in 2016. Panels from top-left: A. Winter, B. Spring, C. Summer, and D. Fall



C.1.3 Parameters

We select n_{il} and c_{il} using an iterative process that relies on several full model solutions and which allows us to replicate the quantity demanded and the own-price elasticity of demand in region i and load segment l with high precision. In the first iteration, we obtain a solution to the full model assuming that demand is perfectly inelastic at the observed 2016 aggregate demand levels, L_{il}, for the regions and load segments defined above. This returns our first predicted value of the Locational Marginal Price (LMP) in region i and load segment l, LMP_{il}, conditional on the assumption of perfectly inelastic demand. Using this we then select n_{il} and c_{il} as follows:

$$n_{il} = \left(1 + \frac{1}{\eta_{il}^{D}}\right) LMP_{il}, \text{ and}$$
$$c_{il} = \left(\frac{1}{\eta_{il}^{D}}\right)\left(\frac{LMP_{il}}{L_{il}}\right), \tag{15}$$

where η_{il}^{D} is the own-price elasticity of electricity demand for region i and load segment l. We then re-run the full model assuming elastic demand reflecting the n_{il} and c_{il} selected above to obtain a more accurate prediction of LMP_{il} that is consistent with our calibrated demand elasticity. This new prediction is then used to select our final n_{il} and c_{il} following (15), which both replicates our baseline aggregate demand levels and the own-price elasticity of demand with only minimal numerical error remaining. Following Bushnell et al. (2017), we select $\eta_{il}^{D} = -0.05$ for all i and l. This is also consistent with Ito (2014), which concludes that the own-price elasticity of electricity demand is less than 0.10.

Following the same procedure outlined above for 2016, we also calibrate n_{il} and c_{il} for 2017 using the same elasticities, 2017 day-ahead hourly LMPs from the *Day-Ahead Hourly LMPs for 2017* dataset (last accessed 03/21/2018 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/da_hrl_lmps), and 2017 hourly load data from the *Hourly Load: Metered for 2017* dataset (last accessed 03/21/2018 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/hrl_load_metered).

C.2 Supply of Electricity from Existing Generation

To calibrate the supply of existing generation in 2016 we use a multi-step approach. First, as described in Section C.2.1, we aggregate $k = 1, ..., 3,095$ existing electric generation units (EGUs) in PJM into $j = 1, ..., 843$ representative existing EGUs based upon similarity in: location (state and load region), fuel type, heat rate, CO_2 emissions factor, and marginal costs. Second, as described in Section C.2.2, we calculate aggregate effective capacity, K_{jl}^{E}. Third, as described in Section C.2.3, we estimate m_{jl}^{E} and b_{jl}^{E} for each representative EGU based on the $k(j) \in \mathcal{K}_j$ subset of EGUs in the full sample that are aggregated to construct each representative EGU j.

C.2.1 Characterization of the Sample of Representative Existing EGUs

To construct our sample of representative existing EGUs, we first meticulously identify the entire population of active existing EGUs in PJM in 2016. To do so, we combine EGU level data on existing EGUs from *The National Electric Energy Data System* (NEEDS) dataset version 5.13 (updated in August 2015) (accessed 05/21/2017) for which a PJM identifier is present in the dataset and the S&P Global's *Summer Capacity* dataset (accessed 05/21/2017) for the subsample of EGUs that S&P Global reports are within PJM. 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 NEEDS), geographical locations of the plants at county level, fuel types, plant capacities, heat rates, and plants' available information (online years, retirement years). The S&P Global 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). We start with the NEEDS sample of existing EGUs because it is more commonly used, better documented, and is not behind a paywall. However, the NEEDS dataset is more dated and tends to miss newer EGUs and some smaller EGUs which are present in the S&P Global dataset. In addition, the two datasets are often not in agreement as to the present operating status of EGUs and the number of EGUs within a facility and the classification of physical location of EGUs within PJM in each of the two datasets is sometimes inconsistent.

To construct our population of active existing EGUs in PJM we use a two-step process that can be broadly outlined as follows. First, we merge existing EGUS reported as being in PJM from the NEEDs dataset with the S&P Global dataset, using a series of sequential merges. Of these, only those EGUs listed as active in the S&P Global dataset are kept. At the end of these merges, there remains a subsample of EGUs from either the NEEDS or S&P Global that remain classified as being located within PJM. Second, of this subsample, we manually check the plants which are in counties that we think might not be in PJM and remove them if necessary. Details on each of these two steps follow.

1. Merging the NEEDS and S&P Global EGUs

The NEEDS dataset reports 3,071 EGUs in PJM and the S&P Global dataset 3,522 EGUs. Most of the plant names, plant types, capacities and unit IDs in the two datasets are similar but not quite exactly the same. Therefore, we complete a sequential series of merges: A. find all one-to-one matches between the NEEDS and S&P Global datasets, B. after removing the matches from A, find all one-to-many, many-to-one, and many-to-many EGU matches in both datasets.[7]

A. One-to-One Matching NEEDS and S&P Global EGUs: To find one-to-one EGU matches between NEEDS and S&P Global, we merge the two datasets four times. First, we match the EGUs that have exactly the same plant names, plant types and unit IDs in NEEDS and S&P Global. This merge results in 456 one-to-one matched EGUs between the two datasets. These matched EGUs are set aside and removed from the original NEEDS and S&P Global datasets, leaving the remaining NEEDS dataset to now have 2,615 EGUs and the S&P Global dataset to now have 3,066 EGUs.

Second, for the remaining EGUs 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 EGU 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 conduct the hand-checks, we consider an EGU in NEEDS a match for an EGU in S&P Global if they have similar plant names, the exact same unit ID, similar plant type, similar fuel type and similar capacity. For example, EGU "Homer City" in S&P Global 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 EGU "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

[7] These non one-to-one matches occur because in some cases, the NEEDS report aggregate EGUs made up of multiple boilers of the same plant types while the S&P Global reports each individual boilers.

batch of 1,577 one-to-one matched EGUs. We again remove these additional one-to-one matched EGUs from the two datasets.

Third, for the remaining EGUs from the two datasets (1,038 EGUs from NEEDS and 1,489 EGUs from S&P Global), we again use reclink2, this time only using plant names and unit IDs as identifiers since the plant capacities for this subset of EGUs maybe different. We notice several instances where an EGU in S&P Global matches with more than one EGU in NEEDS or several EGUs in S&P Global match with one EGU in NEEDS or several EGUs in S&P Global match with several EGUs in NEEDS. For example, EGU "Covanta Plymouth (Montenay Montgomery)" in S&P Global with fuel type of Biomass, unit ID of 1 and capacity of 28 MW is a match for two EGUs 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 EGUs 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 EGU IDs since the two datasets can have different EGU IDs to mean the same EGUs; for example, EGU 1 of the same plant in the NEEDS dataset has unit ID of GEN1 (generator 1) but in the S&P Global dataset has unit ID of BOIL1 (boiler 1). They, however, both mean the same EGU (EGU 1). Hand-checking and correcting the match results from this round of merge, we have another 263 one-to-one matched EGUs, bringing the total one-to-one matched EGUs so far to 2,296 EGUs.

Lastly, for the remaining EGUs in the two datasets (775 EGUs from NEEDS and 1,226 EGUs from S&P Global), 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 EGUs, bringing the total one-to-one matched EGUs so far to 2,443 EGUs. The breakdown of these EGUs and their total capacities are shown in Table A.1 and Table A.2, rows 2-6. Of these 2,443 EGUs, 2,367 EGUs are active in both datasets, 36 EGUs are only active in NEEDs, 27 EGUs are only active in the S&P Global and 13 EGUs are inactive in both datasets. We only include those EGUs that are active in S&P Global 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 EGUs. After completing the one-to-one EGUs matching, we have 628 EGUs in the NEEDS that cannot be matched one-to-one to the EGUs in S&P Global and 1,109 EGUs in S&P Global that cannot be matched one-to-one to EGUs in the NEEDS.

B. Non-One-to-One Matching NEEDS and S&P Global EGUs: We also match several EGUs in NEEDS to one EGU in S&P Global, or one EGU in NEEDS to several EGUs in S&P Global, or many EGUs in NEEDS to many EGUs in S&P Global, as long as these matching EGUs have the same plant names in both datasets. To do this, we find plants in NEEDS and S&P Global 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 EGUs in the plants of the same fuel types. For example, two coal EGUs in S&P Global, Joliet 29 EGU 7 and EGU 8 with capacity of 518 MW each, combined together is a many-to-many match for four bituminous EGUs in NEEDS, Joliet 29 EGUs 71, 72, 81, 82 with capacity of 259 MW each. We find 66 S&P Global EGUs that can be matched to 118 NEEDS EGUs 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 EGUs 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 EGUs in the NEEDS and 65 EGUs in the S&P Global, 1 plant is active in NEEDS but inactive in the S&P Global, corresponding to 5 EGUs in the NEEDS and 1 EGU in the S&P Global. There is no EGUs inactive in the NEEDS but active in S&P Global and no EGUs inactive in both datasets. We again only include those EGUs that are active in S&P Global in the final dataset, which means only 65 additional EGUs are included in the final dataset out of the 66 one-to-one matched EGUs. Note that we only include the EGUs from one dataset (S&P Global) to avoid double counting. We choose the S&P Global because random EGU checks online show the S&P Global to have more accurate capacities.

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

2. Processing Remaining Unmatched NEEDS and S&P Global EGUs

After the series of matches is concluded, there remain 270 active EGUs in NEEDS that cannot be matched to the S&P Global. Since the NEEDS is slightly outdated compared to the S&P Global, it is possible some of the 270 EGUs have already retired. We conduct individual web searches for each of these EGUs to remove those EGUs that are no longer available as of 2016. Of these 270 EGUs, we find 34 EGUs 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 EGUs in NEEDS. Of these 236 EGUs, 53 are believed to be aggregated EGUs over small capacity EGUs of the same fuel types, regions, and states.

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

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

Table A.1 summarizes the construction of the final population of active existing EGUs in PJM in 2016 starting with the NEEDS sample whereas Table A.2 reports the same starting with the S&P Global sample. After combining the NEEDs and the S&P Global datasets, we have 3,509 EGUs in the final population of active existing EGUs in PJM. Table A.3 breaks down the number and capacity of these EGUs by dataset and fuel type. The combined capacity of this population of EGUs is 181,573 MW, of which 168,457 MW are in both NEEDS and S&P Global, 2,320 MW are in NEEDS only and 10,796 MW are in S&P Global only. The biggest fuel sources are gas, coal and nuclear, with natural gas EGUs make up 38% of total capacity in PJM, followed by coal EGUs with 32% and nuclear EGUs with 17%. Renewables EGUs and oil EGUs 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.

Table A.1: Population Construction Starting with the NEEDS Dataset v5.13

	Number of EGUs	Nameplate Capacity of EGUs (MW)	Included in Final Dataset?
Total Number of EGUs in NEEDS v5.13 Dataset	3,071	219,511	–
One to One Matched with S&P Global EGUs	2,443	178,307	–
Active EGUs In NEEDS and S&P Global	2,367	173,503	Y
Active in NEEDS and Inactive in S&P Global	36	2,092	N
Inactive in NEEDS and Active in S&P Global	27	2,451	Y
Inactive in both NEEDS and S&P Global	13	261	N
Collapsed to Merge with S&P Global EGUs	118	5,308	–
Active EGUs In NEEDS and S&P Global	113	5,212	Y
Active in NEEDS and Inactive in S&P Global	5	96	N
Inactive in NEEDS and Active in S&P Global	0	0	Y
Inactive in both NEEDS and S&P Global	0	0	N
Unmatched with S&P Global	510	35,896	–
EGUs in PJM and Still Active, but Not in S&P Global	60	2,470	Y
EGUs Not Included in Final Dataset	450	33,426	N
Unmatched Aggregate EGUs	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: Population Construction Starting with the S&P Global Dataset

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

Table A.3: Final Population of Active Existing Electric Generation EGUs

	Number of EGUs	Nameplate Capacity of EGUs
Total Number of EGUs in Final Dataset	3,095	181,573
Included in both NEEDS and S&P Global	2,459	168,457
Included in NEEDS and Not Included in S&P Global	60	2,470
Not Included in NEEDS and Included in S&P Global	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

3. Constructing the Sample of Representative Existing EGUs

To construct the sample of 843 representative existing EGUs from the population of 3,095 active existing EGUs in PJM in 2016 we assign the 3,095 EGUs into bins based upon similar attributes. These attributes include annual average marginal costs, fuel type, technology type, 25 MW-capacity cut-off, emissions factor, heat rate and location by state s and load region i, the data for which comes from several sources as described below. First, we divide up the population of 3,095 active existing EGUs in PJM in 2016 into bins based on state, fuel type, technology type, and the final five PJM RPAM load regions. Second, we divide each of these sub-samples into as many as seven equal groups based upon variation in emissions rates within each sub-sample.[8] Third, we divide each of these new sub-samples into as many as seven equal groups based upon variation in the annual average hourly marginal costs in 2016 weighted by the number of hours in each load segment (δ_l) within each new sub-sample. This process yields 843 final groups which are used to characterize our sample of representative existing EGUs. This process generates a unique mapping between each representative EGU j to a subset of the population of 3,095 active existing EGUs in PJM in 2016, $k(j) \in \mathcal{K}_j$, where $\mathcal{K}_j$ is the subset of k EGUs assigned to bin j.

C.2.2 Effective Capacity of Representative Existing EGUs

The effective capacity K^E_{jl} of each representative existing EGU j is given by:

$$K^E_{jl} = \sum_{k \in \mathcal{K}_j} K^E_k \gamma^E_{kl}, \tag{16}$$

where K^E_k is the nameplate capacity for original existing EGU k and γ^E_{kl} is the final capacity factor after adjustments for EGU k in load segment l. For the subset of original existing EGUs that are in both the NEEDS and S&P Global datasets, we use the reported nameplate capacity for K^E_k reported in S&P Global whereas for those EGUs that are in NEEDS but not in S&P Global, we use the reported nameplate capacity from NEEDS; we give preference to the nameplate capacity from S&P Global as the S&P Global dataset has been more recently updated.

The final capacity factor after adjustments for EGU k for any hour in load segment l is calculated as: $\gamma^E_{kl} = \tilde{\gamma}^E_{kl} \nu_{fuel(k)}$, where $\tilde{\gamma}^E_{kl}$ is the percent of reported nameplate capacity for original existing EGU k in an hour in load segment l that is lost and $\nu_{fuel(k)}$ is a scalar corresponding to original existing EGU k's fuel type $fuel(k)$.

[8]Note, that if there is no variation between subsets of groups, they are collapsed into fewer than seven groups in each of these steps.

The percent of lost reported nameplate capacity is calculated as: $\tilde{\gamma}^E_{kl} = pa_{kl} cf^{E,M}_{fuel(k)}$. This expression captures capacity that is lost due to plant outages and maintenance through pa_{kl} as well as capacity that is withheld from dispatch for unobserved technical reasons (e.g., thermal constraints) and/or that is underutilized due to unavailability of renewable resources through $cf^{E,M}_{fuel(k)}$. pa_{kl} is availability by original existing EGU k for an hour in load segment l after accounting for plant outages and maintenance. For original existing EGUs whose fuel type is nuclear, $pa_{kl} = ar_{kl}$, where ar_{kl} is the average nuclear availability rate for nuclear original existing EGU k for an hour in load segment l. This is derived from the daily availability rates for each nuclear original existing EGU k from the Energy Information Administration (EIA)'s *Status of US. Nuclear Outages, 2016* dataset (see, *Daily U.S. Nuclear Outage for 2016*, accessed 05/21/2017 at: https://www.eia.gov/nuclear/outages/#/?day=1/1/2016) given the hours mapped to each day in 2016. For original existing EGUs whose fuel type $fuel$ is not nuclear: $pa_{kl} = (1 - or_{fuel(k)l})$, where $or_{fuel(k)l}$ is the average outage rate in 2016 for EGUs of fuel type $fuel$ in an hour in load segment l. $or_{fuel(k)l}$ is derived from outage rates by fuel type and month in 2016 from the *Generating Availability Data System* for 2016 (accessed on 05/17/2017 at https://www.nerc.com/pa/RAPA/gads/Pages/GeneratingAvailabilityDataSystem-(GADS).aspx) given the hours mapped to each month in 2016. pa_{kl} is kept constant for 2017. $cf^{E,M}_{fuel}$ is the capacity factor by fuel type $fuel$ in PJM in 2016 provided in the Market Monitoring Analytics' *2016 State of the Market Report* (see, *Table 5-26 PJM capacity factor (By unit type (GWh): January through December, 2015 and 2016, Section 5 - Capacity Market*, page 250). It captures the percentage of nameplate capacity for EGUs of a particular fuel type $fuel$ that is withheld from dispatch for unobserved technical reasons and/or that is underutilized due to unavailability of renewable resources averaged across 2016. We use update values of these capacity factors for 2017 from Market Monitoring Analytics' *2017 State of the Market Report* (see, *Table 5-31 PJM capacity factor (By unit type (GWh)): 2016 and 2017, Section 5 - Capacity Market*, page 277).

Since we observe a small deviation in capacity by fuel type reported by the PJM market monitor (Market Monitoring Analytics), and the capacity inferred from our bottom-up calibration, we introduce ν_{fuel} to ensure that total capacity by fuel type $fuel$ matches published data on total PJM capacity by fuel type in 2016. It is given by: $\nu_{fuel} = \frac{K^{E,M}_{fuel}}{\sum_{k(fuel)} K^E_k}$, where $K^{E,M}_{fuel}$ is total nameplate capacity by fuel type $fuel$ in PJM in 2016 from Market Monitoring Analytics' *2016 State of the Market Report* (see, *PJM Installed Capacity by Fuel Type, Section 5 - Capacity Market* on page 214) and $k\,(fuel)$ denotes the subset of original existing EGUs of fuel type $fuel$.

C.2.3 Marginal Costs of Existing Generation

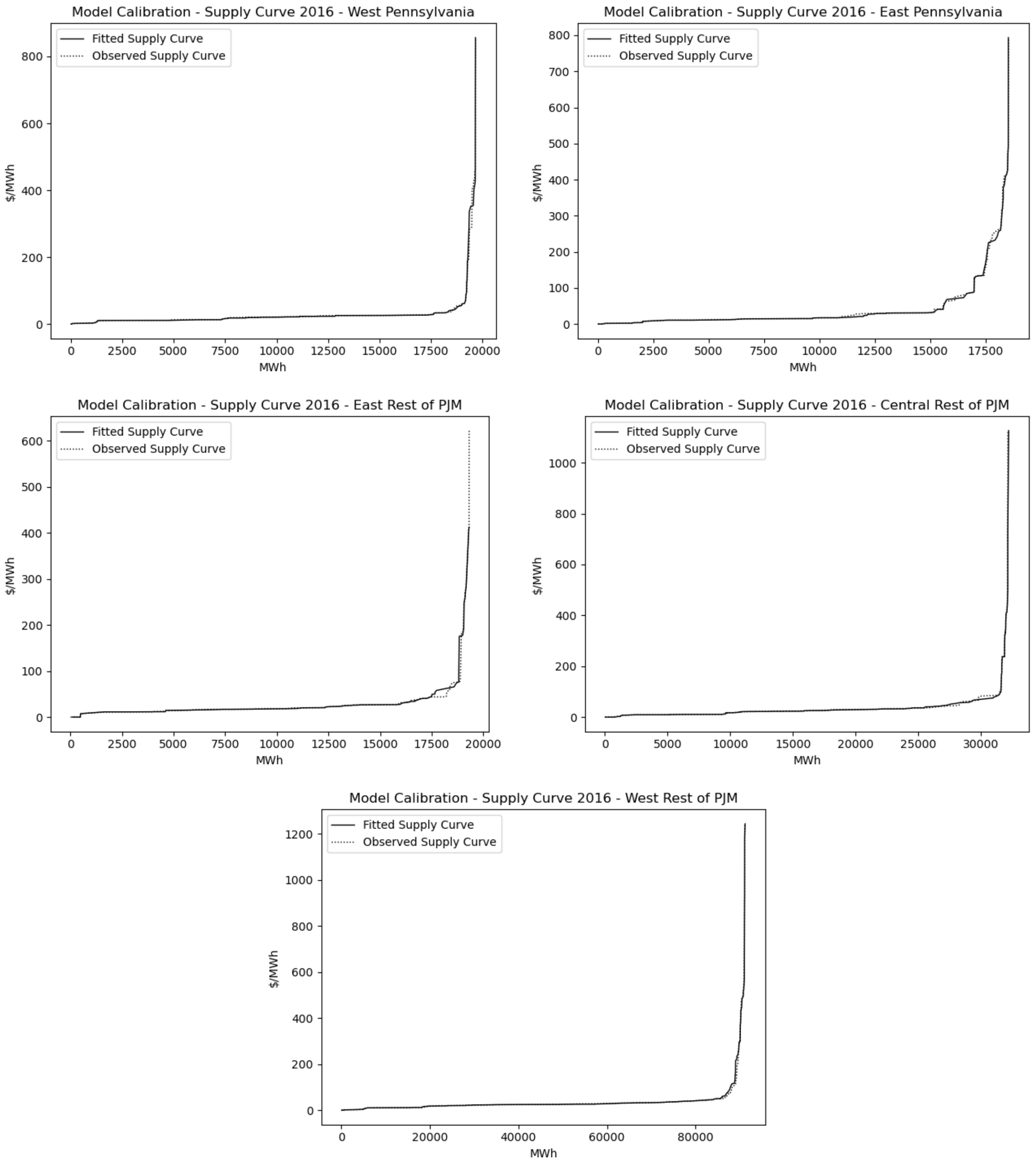
For each j and l, m^E_{jl} and b^E_{jl} are the point estimates from running 80,928 $(= 843 \times 96)$ marginal variable cost OLS regressions:

$$MC^E_{k(j)l} = m^E_{jl}\tilde{K}^E_{k(j)l} + b^E_{jl} + \epsilon^E_{k(j)l}, \tag{17}$$

where $MC^E_{k(j)l}$ is the marginal variable costs of producing electricity from existing EGU $k\,(j)$ during load segment l, $\tilde{K}^E_{k(j)l}$ is the cumulative effective capacity of original existing EGU $k\,(j)$ for representative existing EGU j upon sorting $MC^E_{k(j)l}$ from lowest to highest across all $k\,(j) \in \mathcal{K}_j$ for a given hour in load segment l, and $\epsilon^E_{k(j)l}$ is the error term.

Intuitively, the objective of each of these regression is to identify the line of best fit between marginal variable costs and cumulative capacity for each representative existing EGU j. For many representative existing EGUs, m^E_{jl} is close to zero in which case the EGU has constant mean marginal variable costs b^E_{jl} across the effective capacity for that EGU, K^E_{jl}. In other cases, when m^E_{jl} is positive, marginal variable costs for that representative EGU reflects an upward sloping curve whose generation may be constrained by K^E_{jl}. Figure A.4 illustrates the final result of these 80,928 regressions. Each panel reports the annual average total supply curve across all EGUs within each of the five load regions in PJM in RPAM, where the y-axis reports the annual average hourly marginal costs weighted by the number of hours in each load segment. These curves reflect the merit order in which EGUs are likely to be dispatched within each load region, in which the capacity of EGUs are sorted from lowest to highest annual average hourly marginal costs. These curves clearly show that our rich regression approach from (17), coupled with our construction of representative EGUs as described in the previous section, does an exceptional job of characterizing marginal costs across the original population of existing EGUs. The R^2 for each region reports the amount of variation from our calibrated regional average total supply curves that are explained by regional average total supply curves constructed using the marginal costs and effective capacity from the original population of existing EGUs; these are: West Pennsylvania, $R^2 = 0.998$ (top row, left); East Pennsylvania, $R^2 = 0.979$ (top row, right); East Rest of PJM, $R^2 = 0.858$ (second row, left); Central Rest of PJM, $R^2 = 0.982$ (second row, right); and West Rest of PJM, $R^2 = 0.944$ (bottom row).

Figure A.4: Comparison of Actual and Representative Annual Average Hourly Marginal Costs Weighted by the Number of Hours in each Load Segment for each Load Region in 2016



Cumulative effective capacity is calculated as:

$$\tilde{K}^E_{k(j)l} = \sum_{x=1}^{x(k(j))} K^E_x \gamma^E_{xl}, \tag{18}$$

where K^E_x and γ^E_{xl} are defined as above. Here, $x(k\,(j))$ reflects a re-mapping of k to rank x of original existing EGU k within each bin j after sorting $MC^E_{k(j)l}$ from lowest to highest.

The marginal variable costs of producing electricity from existing EGU $k\,(j)$ during load segment l is calculated according to:

$$MC^E_{k(j)l} = fc^E_{k(j)l} + rc^E_{k(j)} + omc^E_{k(j)}, \tag{19}$$

where $fc^E_{k(j)l}$ is the marginal fuel costs of existing EGU $k\,(j)$ for an hour in load segment l, $rc^E_{k(j)l}$ is the marginal pre-existing regulatory costs (excluding the marginal costs from RPS and RGGI compliance, which are endogenous in the model) for existing EGU $k\,(j)$, and $omc^E_{k(j)}$ is marginal operation and maintenance costs for existing EGU $k\,(j)$, all measured in $\$/MWh$. For EGUs in the original population, marginal variable operation and maintenance costs are taken from the S&P Global dataset. For EGUs not in S&P Global, we assume that marginal operation and maintenance costs equal the average of marginal operation and maintenance costs of EGUs with the same fuel type and in the same state that are reported in S&P Global. We next describe how marginal fuel and regulatory costs are calculated.

Marginal Fuel Costs

Marginal fuel costs are given by: $fc^E_{k(j)l} = \psi^E_{k(j)} p^E_{k(j)l}$, where $\psi^E_{k(j)}$ is the heat rate of existing EGU $k\,(j)$ measured in $mmBTU/MWh$ (converted from BTU/kWh) and $p^E_{k(j)l}$ is the delivered price of fuel to existing EGU $k\,(j)$ for any hour in load segment l measured in $\$/mmBTU$ in light of the fuel type $fuel(k)$ corresponding to that EGU.

The heat rates for our population of original existing EGUs, $\psi^E_{k(j)}$, use the best available estimates across five datasets (from most to least accurate): *Continuous Emission Monitoring System* dataset (CEMS) published by the Environmental Protection Agency (EPA) (accessed

06/28/2017), EPA's *Emissions & Generation Resource Integrated Database* (eGrid) (accessed 06/15/2017), S&P Global, EIA's *Form EIA-923 Schedules 3A & 5A*, and NEEDS. A detailed summary of heat rates across our original population of EGUs is provided in Table A.4.

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

Dataset	Number of EGUs Assigned Heat Rate to
Final Dataset	3,095
CEMS	374
eGrid	1,490
S&P Global	801
EIA 923	411
NEEDS	19

CEMS reports measured real-time heat rates which we believe are superior to engineering derived estimates. Thus, for the 374 EGUs in our original population that are also in CEMS, we first assign CEMS heat rates for those 374 EGUs. Second, we assign heat rates to 1,490 EGUs in our original population using the heat rates reported in eGrid that we have reason to believe are credible. We do not assign heat rates to all remaining 2,721 EGUs (3,095-374), since we observe a fair amount of measurement error in the heat rates and generation data that eGrid reports.[9] Third, we assign heat rates to an additional 801 EGUs using the heat rates reported in S&P Global. We do not assign heat rates to all remaining 1,231 EGUs (2,721-1,490), since we again observe a fair amount of measurement error in the heat rates that S&P Global reports.[10] Fourth, we assign heat rates to an additional 411 EGUs using the average heat rates between 2012-2016 reported in *Form EIA-923*. These 411 EGUs represent the subset of remaining EGUs after the last step for which *Form EIA-923* reports annual heat rates between 2012-2016. Finally, the 19 remaining EGUs (430-411) are assigned heat rates from the NEEDS.

[9] We exclude from this step EGUs with eGrid heat rates that are above 42,545.31 BTU/kWh (the mean eGrid heat rate plus 1.96 times the standard deviation of eGrid heat rates) as well as EGUs marked "Data from EIA-923 Generator File overwritten with distributed data from EIA-923 Generation and Fuel." We do this to avoid assigning unrealistically high heat rates to EGUs in our final dataset and to avoid assigning heat rates from EGUs in eGrid that report negative annual generation, which coincide with the EGUs that are marked "Data from EIA-923 Generator File overwritten with distributed data from EIA-923 Generation and Fuel."

[10] We exclude from this step EGUs with S&P Global heat rates that are above 26,473.842 BTU/kWh (the mean S&P Global heat rate plus 1.96 times the standard deviation of S&P Global heat rates). S&P Global performs their own heat rate calculation using *Form EIA-923*. For some EGUs that generate a very small amount of electricity in a year, their estimates of heat rates for these EGUs are so unrealistically high that they cap them at 100,000 BTU/kWh. We believe these estimates are not reasonable and want to exclude these EGUs from our heat rate calculation.

The delivered price of fuel to existing EGU $k\,(j)$ for any hour in load segment l depends on the fuel type corresponding to each existing EGU, $fuel(k)$, as well as the physical location of each existing EGU.

For natural gas fired existing EGUs, the delivered price of fuel equals the natural gas spot price assigned to existing EGU k plus transportation costs. In our analysis natural gas spot prices are allowed to vary across load segments to capture the strong observed correlation between load and exogenous drivers of natural gas price variability. Lacking data at the EGU-level on contract arrangements for securing natural gas deliveries, we assign existing natural gas fired EGUs to one of seven gas spot pricing hubs: Alliance, Dominion North Point, Chicago City Gate, Lebanon OH, TETCO Zone M3, Tennessee Gas Zone 4 - Marcellus and Transco Leidy. We assign each EGU to the nearest hub using the Euclidean distance (as the crow flies) between the latitude and longitude of the existing EGU and centroids for each of the seven spot pricing hubs. For a few natural gas fired existing EGUs for which we do not have latitude and longitude, we assign those EGUs the average daily natural gas spot prices of all the natural gas fired EGUs in the states that they are located in. Spot prices for each of these hubs come for 2016 come from Bloomberg's *Daily Natural Gas Spot Price* dataset. Given the hours in each day and their assignment to the load segments across the year, the load segment average spot price assigned to each natural gas fired existing EGU provides the natural gas spot price assigned to existing EGU k for load segment l. Transportation costs are constant across load segments but are allowed to vary across the census regions in which natural gas fired existing EGUs are geographically located. Transportation costs equal the difference between the 2016 average natural gas delivered price to the electric power sector for each census region in PJM, taken directly from the EIA's *Annual Energy Outlook 2017* (AEO) (*Table: Energy Prices by Sector and Source, Case: Reference case, AEO 2017*), and the annual EGU-weighted average natural gas spot price derived from the Bloomberg data. To calculate the latter we first create annual averages of the daily natural gas prices for each gas hub. Then, given the spatial assignment of these annual averages to each natural gas fired existing EGU, we take the average across all natural gas fired existing EGUs located within a given census region.

For coal fired existing EGUs, the delivered price of fuel assigned to EGU k equals the coal spot price plus transportation costs. Similar to natural gas, coal spot prices vary across load segments, however, in this case, we exploit variation in spot prices across weeks and not days. Transportation costs are a larger share of marginal fuel costs for coal fired existing EGUs than for natural gas fired existing EGUs, so we anticipate more opportunities for input price arbitrage across coal basins than across natural gas hubs. In addition, we lack data at the EGU-level on contract arrangements for securing coal deliveries and restrictions at the EGU level as to the types of coal that the EGU can utilize.[11] For these reasons, the fuel cost we assign to each EGU k in a given hour in load segment l is the lowest delivered price of coal to EGU k in a given hour in load segment l from one of five coal basins: Central Appalachia Basin, Northern Appalachia Basin, Illinois Basin, Powder River Basin, and the Unita Basin. To determine this, we first calculate the delivered price of coal to EGU k in week w for each of the five coal basins. This equals the week w coal spot price in each coal basin plus annual average transportation costs from each basin to the state $s(k)$ in which coal fired existing EGU k is based in. Weekly coal spot prices from each of the five coal basins are taken directly from the EIA's *Coal Markets Archive* (see, https://www.eia.gov/coal/markets/includes/archive2.php). Average annual transportation costs from each coal basin to each state equal the average annual transportation costs across three modes of transportation: truck, waterway, and railroad, respectively taken from the EIA's *Coal Transportation Rates to the Electric Power Sector*, *Table 3a. Average annual coal transportation costs from coal basin to state by truck*, *Table 3b. Average annual coal transportation costs from coal basin to state by waterway*, and *Table 3c. Average annual coal transportation costs from coal basin to state by railroad.* Finally, the delivered price of coal to EGU k for each of the five coal basins in week w is converted to a price that varies across load segments $l(w)$ given the mapping between hours and weeks.

Delivered oil and uranium prices for existing EGUs k that use these fuel types are assumed to only vary by the census region in which that existing EGU is located and are constant across load

[11] We requested access to this confidential information from the EIA, but they denied our request.

segments within each year. The prices we use are the annual delivered distillate oil prices and uranium prices for electric power sector by census region for 2016 from the EIA's *Annual Energy Outlook 2017, Table: Energy Prices by Sector and Source, Case: Reference case.* We do not consider within-year variability in distillate oil prices since most of the oil fired existing EGUs we observe are infrequently dispatched only during peak demand periods, and we assume these EGUs have access to local storage sufficient to fully price arbitrage across a year. Nuclear prices simply do not exhibit large variation in prices within a year, given the energy density of uranium fuel and its capacity to be stored on site at low cost. Other delivered fuel prices (biomass and other fuels) are taken from S&P Global's *Summer Capacity* dataset for each remaining existing EGU k and which are also constant across load segments within each year.

Marginal Pre-existing Regulatory Costs

Marginal pre-existing regulatory costs (excluding the marginal costs from RPS and RGGI compliance, which are endogenous in the model) for existing EGU k are given by: $rc^{E}_{k(j)} = p^{SO_2}\phi^{E,SO_2}_{k(j)} + p^{NOx}\phi^{E,NOx}_{k(j)}$, where p^{SO_2} and p^{NOx} are, respectively, the annual SO_2 and NOx allowance prices measured in $\$/lb$ and $\phi^{E,SO_2}_{k(j)}$ and $\phi^{E,NOx}_{k(j)}$ are, respectively, the emissions intensity of $\mathrm{SO_2}$ and NOx for existing EGU $k\,(j)$ measured in lbs/MWh as discussed further below.[12] p^{SO_2} is the sum of the annual SO_2 allowance prices for the Acid Rain Program and the Cross-State Air Pollution Rule (CSAPR) in 2016. p^{NOx} is the annual NOx CSAPR allowance price in 2016. These three prices are taken directly from S& P Global's *NOx/SO2 Allowances*.

C.3 New Capacity and Supply of Electricity from New Generation

We assume that utilization factors for new EGU j, γ^{N}_{j}, vary only by technology types and not across states. These utilization factors by technology type are taken directly from Monitoring Analytics' *2016 State of the Market Report, Table 5-26 PJM capacity factor (By unit type (GWh)): January through December, 2015 and 2016, on page 250* and are 0.961, 0.295, and 0.177 for NGCC,

[12] In addition, starting in 2019, Illinois and New Jersey implement nuclear subsidies. These subsidies are included in marginal pre-existing regulatory costs in future years, although here we seek to characterize marginal pre-existing regulatory costs in the years of calibration, 2016 and 2017; see Section D for a complete description of how marginal pre-existing regulatory costs vary in future years and Section F for additional details on these subsidies.

wind, and solar, respectively. Likewise, we also assume that C_j^N only varies by technology type. While estimates of certain components of the annual cost of adding new capacity costs are pervasive (e.g., the purchase costs of capital, the costs from purchasing or leasing land, financing costs, and some costs associated with new capacity permitting and approval) they tend to be based largely on engineering estimates rather than revealed preferences. Moreover, data on other cost components are unobserved (e.g., search costs, costs of new technology adoption and innovation, costs costs arising from regulatory uncertainty). As such, we numerically calibrate C_{tech}^N for each technology type $tech$ as described below.

The marginal variable costs for new EGU j for an hour in load segment l, b_{jl}^N, is calculated as:

$$b_{jl}^N = fc_{jl}^N + rc_j^N + omc_j^N, \tag{20}$$

where fc_{jl}^N is the marginal fuel costs of new EGU j for an hour in load segment l, rc_j^N is the marginal pre-existing regulatory costs (excluding the marginal costs from RPS and RGGI compliance, which are endogenous in the model) for new EGU j, and omc_j^N is marginal operation and maintenance costs for new EGU j, all measured in $\$/MWh$.

For new wind and solar EGUs, $fc_{jl}^N = 0$. For new NGCC EGUs, $fc_{jl}^N = \psi_j^N p_{jl}^N$, where ψ_j^N is the heat rate of new NGCC EGU j measured in $mmBTU/MWh$ (converted from BTU/kWh) and p_{jl}^N is the delivered price of natural gas to new EGU j) for any hour in load segment l measured in $\$/mmBTU$. ψ_j^N is set to equal the average heat rate of all existing NGCC EGUs in PJM. $p_{jl}^N = \frac{1}{J_s}\sum_{j\in s} p_{jl}^{NGCC}$, where J_s is the number representative EGUs in state s and $p_{jl}^{NGCC} = \frac{1}{J_j^{NGCC}}\sum_{k\in\mathcal{K}_j^{NGCC}} p_{kl}^E$, given that $\mathcal{K}_j^{NGCC}$ is the set of NGCC existing EGUs in the original population that correspond to NGCC representative existing EGU j of size J_j^{NGCC} and p_{kl}^E is the delivered price of fuel (here restricted to just natural gas fired existing EGUs) for existing EGU k in load segment l, as defined in the previous section. Marginal pre-existing regulatory costs (excluding the marginal costs from RPS and RGGI compliance, which are endogenous in the model) for new EGU j are given by: $rc_j^N = p^{SO_2}\phi_j^{N,SO_2} + p^{NOx}\phi_j^{N,NOx}$, where p^{SO_2} and p^{NOx} are, respectively, the annual SO_2 and NOx allowance prices as discussed in the previous section and ϕ_j^{N,SO_2} and $\phi_{k(j)}^{N,NOx}$ are, respectively, the emissions intensity of $\mathrm{SO_2}$ and NOx for new EGU j measured in lbs/MWh as discussed further below. $omc_j^N = \frac{1}{J_s^{tech}} \sum_{j \in s} omc_j^{E,tech}$, where J_s^{tech} is the number representative EGUs in state s of technology type $tech$ and $omc_j^{E,tech} = \frac{1}{J_j^{tech}} \sum_{k \in \mathcal{K}_j^{tech}} omc_k^E$, given that $\mathcal{K}_j^{tech}$ is the set of existing EGUs in the original population of technology type $tech$ that correspond to representative existing EGU j of technology type $tech$ of size J_j^{tech} and omc_k^E is the marginal operation and maintenance costs for existing EGU k, as defined in the previous section.

C.4 External Renewable Energy Certificates

For those state-tier RPSs that allow it, we allow for the importation of external RECs from outside of PJM to assist those states with complying with their RPS targets. The sets $\mathcal{ST}_q$ and $\mathcal{Q}_{st}$ are constructed based on the state in which external REC supplier q is located and importation constraints specified in individual state's RPS standards. The annual amount of RECs that external REC supplier q can supply to all states including states that are outside of PJM), $\bar{r}_q$, are taken directly from three sources: PJM Environmental Information Services' *Generation Attribute Tracking System* (PJM-GATS), the *North Carolina Renewable Energy Tracking System* (NC-RETS), and the *Michigan Renewable Energy Certificates System* (MIRECS). PJM-GATS only tracks RECs that can be used to comply with the RPSs of nine states within PJM: District of Columbia, Delaware, Illinois, Maryland, New Jersey, Ohio, Pennsylvania, Virginia, and West Virginia and therefore we supplement using data on RECs tracked by the state tracking systems for North Carolina and Michigan, NC-RETS and MIRECS. In 2016 there are 243 EGUs outside of PJM that can supply RECs for compliance with state RPSs within PJM and in 2017, 262.

While we have data on the amount of total external REC supply available, $\bar{r}_q$, data on the actual amount of external RECs supplied does not exist, r_q. As many of these external REC suppliers can likely supply to many other states outside of PJM that also have RPSs, we cannot assume that all of $\bar{r}_q$ is supplied to PJM states exclusively. For this reason we have introduced the parameter γ_q in (7). In order to calibrate this vector of parameters, we divide the entire set of external REC suppliers into two types based upon importation constraints specified in individual state's RPS standards: those that supply exclusively to Pennsylvania, $\mathcal{Q}_{PA,t}$, and those that supply to states in PJM other than Pennsylvania exclusively, $\mathcal{Q}_{\neg PA,t}$. For external REC suppliers in the first set we assume that $\gamma_q = \gamma_{PA}$ and for those in the second set we assume that $\gamma_q = \gamma_{\neg PA}$. As discussed further below, we numerically calibrate these two scalars twice for 2016 and 2017, leaving us with four scalars: γ_{PA}^{2016}, $\gamma_{\neg PA}^{2016}$, γ_{PA}^{2017}, and $\gamma_{\neg PA}^{2017}$.

C.5 Transmission Network

Transmission line data is taken from S&P Global's *Operating Transmission Projects Map* for 2016. Upon visual inspection, transmission lines that connect any two zones within two PJM RPAM regions are aggregated into a single aggregated transmission lines with voltage V_{ih} equal to the sum of the maximum voltage across all lines connecting the two PJM RPAM regions, as shown in Figure A.1. To be precise, a line is presumed to connect two regions if it connects at least two ISO market hubs with at least one ISO market hub in each region.[13] We identify five aggregate transmission lines that connect our five PJM RPAM regions. The table below depicts all possible combinations of aggregate links between our five PJM RPAM regions and their respective aggregate voltage in KV. V_{ih} equals the number provided divided by 1,000 (in MV).

Table A.5: PJM Transmission Networks

Transmission Line Between	Total KV
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

While its possible to construct an estimate of the aggregate voltage of the five lines considered in our analysis based upon the data on the actual voltage of individual lines to which each aggregate line corresponds, data on the actual effective amperage of lines does not exist for

[13]According to S&P Global, an ISO market hubs is "trading location at which settlement data is available and verifiable." In PJM, this reflects a point where locational marginal prices are specified for the wholesale market.

which to perform a similar aggregation nor, given the physical properties of transmission, is it clear that aggregate effective amperage can be aggregated in a similar fashion as voltage. Moreover, what we characterize as the effective amperage of a particular line in a given hour in our simplified representation of the physical power system (e.g., we abstract from conductor emissivity, conductor absorptivity, resistance, and other factors which would be need to formulate a complete engineering description of a transmission line) is likely to vary with temperature, solar flux, wind speed, and wind direction in that hour. Across an aggregate transmission line, effective amperage may also vary across hours depending upon how the ensemble of actual transmissions lines that correspond to that aggregate line are interconnected and physically interact with the aggregate flow of power that moves through them reflecting, possibly, localized congestion and losses which are likely highly correlated with load. Thus, while many numerical power system modelers make assumptions about the amperage of the aggregate lines that they model (and which are typically assumed to be constant over time), we not feel comfortable making similar assumptions. These assumptions have critical implications for the predictions generated by these models which may make it difficult for these models to accurately predict the inter-regional flow of power (and therefore, the within region flow of power), regional locational marginal prices, and the resulting implications for the welfare of economic agents throughout the power system. For these reasons we do not adopt a similar approach, but instead introduce novel numerical calibration methods often used in other modelling domains to numerically calibrate A_{ihl} for all $ih = 1, ..., 5$ and load segments $l = 1, ..., 96$, as discussed further below.

C.6 Estimation of CO_2 Emissions by RGGI States Not in PJM and RGGI Allowance Bank

The transportation model of the PJM power system included within RPAM was originally designed to reflect the scale in which electricity dispatch and new capacity expansion decisions are actually made within the wholesale electricity market operated by PJM. RGGI of course encompasses states located within two other significantly sized wholesale electricity markets: the wholesale electricity market that includes Maine, Vermont, New Hampshire, Massachusetts, Connecticut, and Rhode Island which is managed by the New England Independent System Operator (ISONE), and the wholesale electricity market encompassing New York operated by the New York Independent System Operator (NYISO). We ultimately chose not to expand the transportation model of the PJM power system to individually account for the economic decisions made in these two additional wholesale markets. In addition, to not being feasible due to time constraints, such an expansion would require us to make strong assumptions regarding how power flows economically between these two regional wholesale electricity markets and PJM.

Since our focus is on understanding Pennsylvania's entry into RGGI and Pennsylvania is a critical member of PJM, all that is really necessary is for us to proceed is to account for CO_2 emissions from RGGI states not in PJM and changes in allowances that have been banked by RGGI market participants historically (in PJM or otherwise). Given the rich data available to us from RGGI, we are able to do so quite credibly reflecting the revealed economic decisions made by RGGI market participants from the beginning of RGGI's first allowance market in 2009. We treat these economic decisions in a reduced form manner, meaning we do not explicitly characterize how changes in electricity dispatch in PJM affects auction prices–for which we do not have data–but instead identify the RGGI market equilibrium using plausibly exogenous instrumental variables. We then use these two reduced form equations to extend the transportation model of the PJM power system to allow us to characterize market clearing in the RGGI cap and trade system in (13).

To this end, we use three stage least squares (3SLS) to simultaneously estimate the following system of equations using quarterly data from 2009 to 2019, where t denotes quarters:

$$B_t = \beta_0 + \beta_1 P_t + \beta_2 B_{t-1} + \beta_3 D_{NJinRGGI,t} + \beta_4 D_{q2,t} + \beta_5 D_{q3,t} + \beta_6 D_{q4,t} + \varrho_t, \tag{21}$$

$$E_t = \gamma_0 + \gamma_1 P_t + \gamma_2 D_{CP2,t} + \gamma_3 D_{CP3,t} + \gamma_4 D_{q2,t} + \gamma_5 D_{q3,t} + \gamma_6 D_{q4,t} + \gamma_7 R_t + \gamma_8 D_{CP4,t} + \nu_t,$$

$$P_t = \delta_0 + \delta_1 Q_t + \delta_2 D_{CPA1,t} + \delta_3 D_{CPA2,t} + \delta_4 D_{q2,t} + \delta_5 D_{q3,t} + \delta_6 D_{q4,t} + \delta_7 R_t + \varepsilon_t.$$

where: B_t is the amount of banked permits as of t, B_{t-1} is the amount of banked permits as of

$t-1$, P_t is the market clearing price of permits at auction, $D_{NJinRGGI,t}$ is a dummy equal to one if New Jersey is in RGGI in t and zero otherwise, $D_{q2,t}$ is a dummy equal to one if t corresponds to quarter two, $D_{q3,t}$ is a dummy equal to one if t corresponds to quarter three, $D_{q4,t}$ is a dummy equal to one if t corresponds to quarter four, Q_t is the quantity of permits offered at auction in t, $D_{CPA1,t}$ is a dummy that equals one if the first control period adjustment is in effect, $D_{CPA2,t}$ is a dummy that equals one if the second control period adjustment is in effect, R_t is the quarterly average of daily bank discount rates for treasury bills with four week maturation, E_t is total CO_2 emissions from covered EGUs in RGGI states that are not in PJM, $D_{CP2,t}$ is a dummy equal to one if t is in control period two, $D_{CP3,t}$ is a dummy equal to one if t is in control period three, and $D_{CP4,t}$ is a dummy equal to one if t is in control period four. ϱ_t, ν_t, and ε_t are the modelled error terms and the β's,γ's, and δ's are the parameters to be estimated.

R_t is calculated from the daily bank discount rates for treasury bills with four week maturation as reported in the U.S. Department of the Treasury's *Daily Treasury Bill Rates Data*.

E_t is calculated from quarterly level data on CO_2 emissions from covered EGUs in RGGI from RGGI's *CO_2 Allowance Tracking System* (COATS) for the subset of EGUs in RGGI states that are not in PJM. We assume that if a EGU is "not operating" that it releases zero CO_2 emissions.

The amount of permits added to the permit bank in a quarter are assumed to equal the number of permits sold in the quarterly RGGI auction as reported in RGGI's *Allowance Prices and Volumes* dataset plus the amount of permits not-auctioned in the quarter less the CO_2 emissions from covered EGUs in all RGGI states aggregated from EGU level quarterly emissions from COATS. B_t is the cumulative sum of these additions and subtractions to the bank (restricted to be non-negative) as of quarter t and B_{t-1} is the one quarter lag of B_t.

The amount of permits not-auctioned in the quarter equals the quantity of permits offered at auction in the quarter multiplied by the share of non-auctioned permits in the associated year to the total amount of permits offered in the associated year.[14] The quantity of permits offered at auction in the quarter is taken from RGGI's *Allowance Prices and Volumes* dataset. The total

[14]Permits sold in 2008 before the beginning of the first control period in 2009 are added to the permits offered in 2009.

amount of permits offered in a year is calculated from the quarterly amount of permits offered. The amount of non-auctioned permits by year equals the CO_2 Allowance Adjusted Budget plus the amount of permits released by the Cost Containment Reserve (CCR) less the quantity of permits offered at auction that year. The CO_2 Allowance Adjusted Budget and the amount of permits released by the CCR are taken from RGGI's Distribution of 2009-2020 Allocation Year CO_2 Allowances datasets. The quarterly allowance price at auction, P_t is also taken from RGGI's *Allowance Prices and Volumes* dataset.

All remaining dummy variables are defined based upon the years associated with each of the four control periods and the years in which the CO_2 Allowance Budget was modified by the first and second control period adjustments.

The first equation in (21) is a reduced form representation of the first-order condition of an economic agent's inter-temporal profit maximization decision to bank permits. The second equation is a reduced form representation of a representative generator's (here, reflecting generation across all non-PJM RGGI states) profit maximizing decision to reduce CO_2 emissions for compliance with RGGI. Concurrently, this also reflects the representative generator's demand for permits at auction. The third equation reflects market clearing in the RGGI primary allowance auction, where Q_t reflects the supply of new permits offered at auction.

Taken together the last two equations in (21) reflect the conventional structural estimation of the supply and demand of permits. Given the exogenous variables in these equations, these two equations identify the endogenous market clearing price for permits sold at auction, which, together with additional exogenous variables, is used to estimate the decision to withdraw (demand) or add (supply) permits in t to the bank. We use quarterly level data for which we have a complete series from quarter 1 of 2009 to quarter 4 of 2019.

After some manipulation the estimated parameters from (21), given that the original data is by quarter and in short tons, the parameters estimated from the first equation imply that: $c^b\left(\bar{B}\right) = -1.102311\left(\frac{\frac{1}{4}\left(4\hat{\beta}_0+\hat{\beta}_4+\hat{\beta}_5+\hat{\beta}_6\right)+\hat{\beta}_2\bar{B}}{\hat{\beta}_1}\right)$ and $n^B = -\frac{1}{\hat{\beta}_1}$, where $\hat{\cdot}$ denotes a parameter estimate. $\bar{B}$ is initialized to equal the amount of observed banked permits at the end of quarter four in 2019 and updated each year thereafter as described above. Likewise, the parameters estimated from the second equation imply that: $c^{NPJM} = -1.2150895 \left(\frac{4\left(\hat{\gamma}_0 + \hat{\gamma}_7 \tilde{R}_t + \hat{\gamma}_8\right) + \hat{\gamma}_4 + \hat{\gamma}_5 + \hat{\gamma}_6}{4\hat{\gamma}_1} \right)$ and $n^{NPJM} = -\frac{1}{4\hat{\gamma}_1}$, where $\tilde{R}_t$ is the average of the quarterly average daily bank discount rates for treasury bills with four week maturation for quarters in the fourth control period.

C.7 Aggregate Loss

Aggregate loss in any given hour in load segment l as a percentage of demand in region i for that hour is given by: $\epsilon_l = z_0 + z_l$, where $z_0 = 0.034$ is the annual average transmission and distribution system losses across all of PJM, and z_l is the difference in virtual increment offers and decrement bids in an hour in load segment l as a fraction of total PJM load in load segment l. z_0 reflects annual average transmission and distribution system across PJM of 3.4%, calculated for 2016 as the difference between total PJM generation and total PJM load all normalized by total PJM load in 2016, where total generation and load in PJM in 2016 are taken from Market Monitoring Analytics' *2016 State of the Market Report*.

Economic participants within PJM are allowed to inter-temporally arbitrage in the day ahead market at each load node by taking long (decs) or short (incs) positions through a virtual bidding process on the amount of load expected at each hour and at each load node. If participants make an offer to sell MWh in a given hour and at a given load node, they make a virtual increment offer or 'inc bid' in the day ahead market. Conversely, if participants make an offer to buy MWh in a given hour and at a given load node, they make a virtual decrement offer or 'dec bid' in the day ahead market. We account for the amount of cleared daily net virtual bids (total decs - total incs) by load segment in our model through z_l, but not across load nodes as data on inc-decs by load zone is not publicly available. Accounting for virtual bidding behavior in our model is economically significant as we find that z_l ranges from -0.0518 to 0.0283, and thus for some load segments, virtual bidding in PJM is of the same order of magnitude as losses from the transmission and distribution system within PJM.

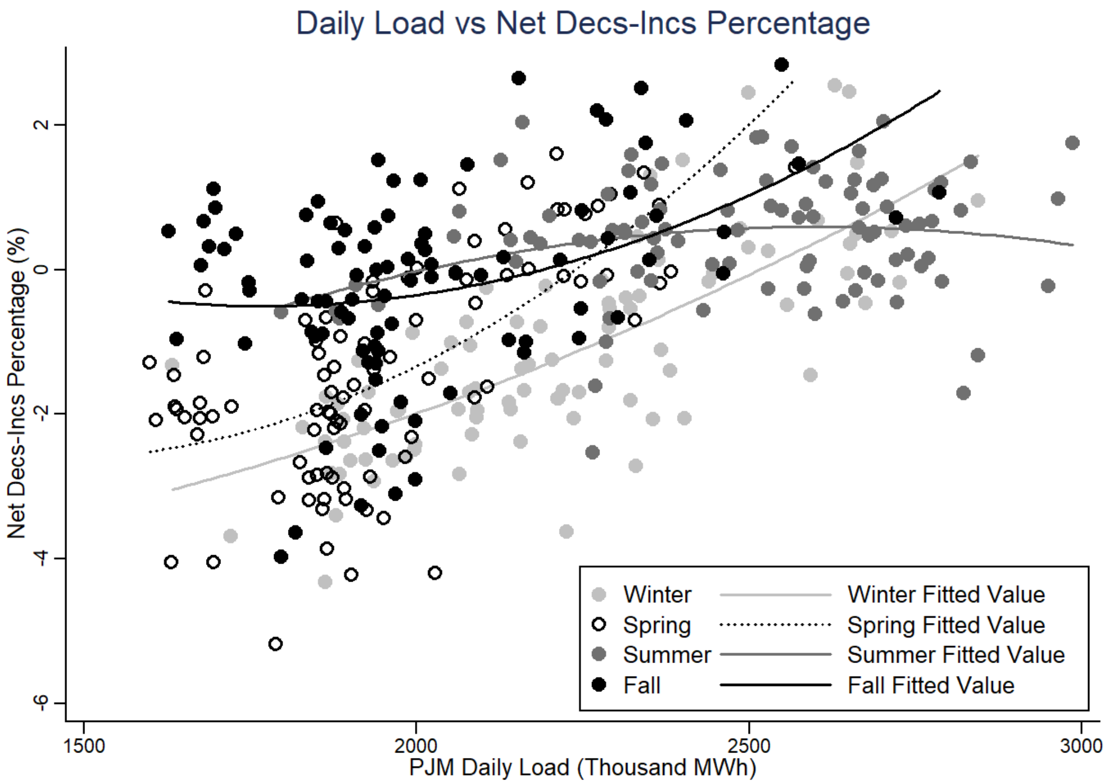
To assess how virtual bidding effects market clearing in PJM, we first estimate the following

OLS regression using daily data from 2016, where t denotes days:

$$
\begin{aligned}
NVB_t = {} & \beta_1 L_t + \beta_2 L_t^2 + \beta_3 D_{W,t} + \beta_4 D_{Sp,t} + \beta_5 D_{Su,t} + \beta_6 D_{F,t} \\
& + \beta_7 D_{W,t} \times L_t + \beta_8 D_{Sp,t} \times L_t + \beta_9 D_{Su,t} \times L_t + \beta_{10} D_{F,t} \times L_t \\
& + \beta_{11} D_{W,t} \times L_t^2 + \beta_{12} D_{Sp,t} \times L_t^2 + \beta_{13} D_{Su,t} \times L_t^2 + \beta_{14} D_{F,t} \times L_t^2 + \varepsilon_t,
\end{aligned} \tag{22}
$$

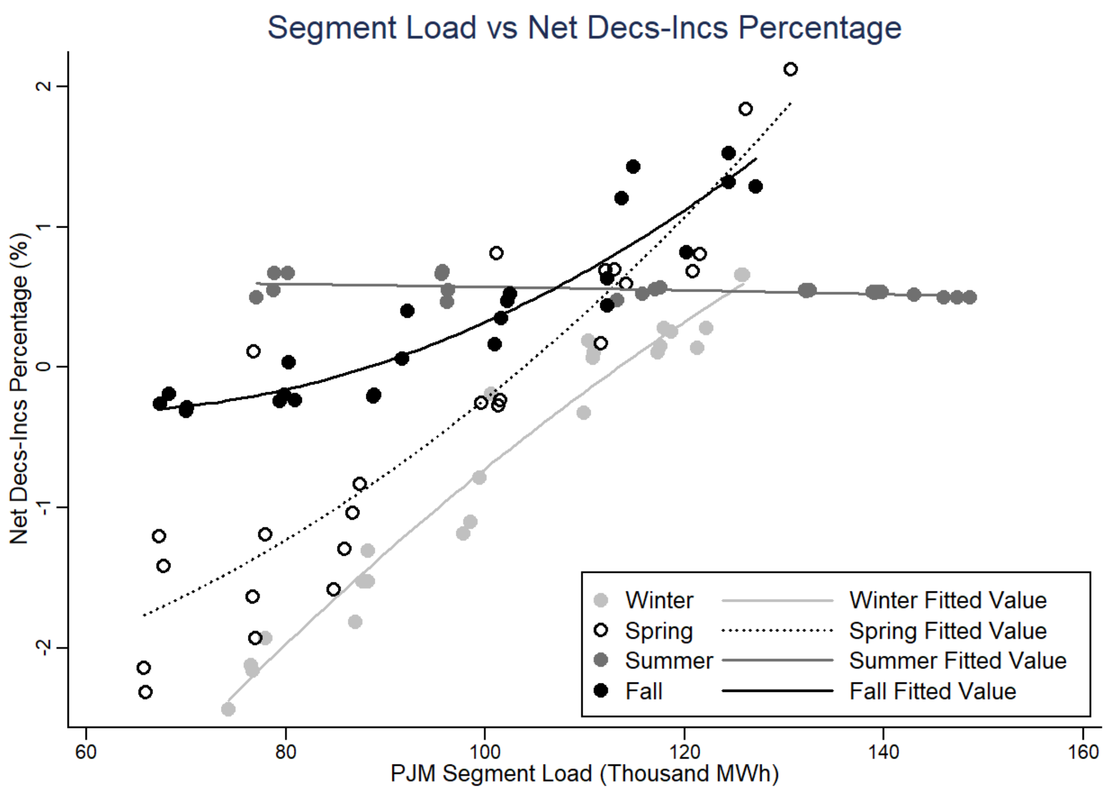
where: NVB_t is the net virtual bid (total number of incs less total number of decs) in day t; L_t is the total PJM load in day t; and $D_{W,t}$, $D_{Sp,t}$, $D_{Su,t}$, and $D_{F,t}$ are seasonal dummies by t. ε_t is the modelled error term and the β's are the parameters to be estimated. NVB_t is calculated using total daily cleared incs and decs taken directly from the *Daily cleared INCs, DECs, and UTCs* dataset from *PJM's Data Miner 2* (accessed May 19, 2019 at https://dataminer2.pjm.com/feed/day_inc_dec_utc/definition). Seasonal dummies across days correspond to our definitions of seasons stated in Section C.1. Total PJM load in day t is the sum across all load zones in PJM and all hours in each day t of the hourly load data by load zone (from the *Hourly Load: Metered for 2016* dataset, discussed above). The R^2 from this regression is 0.57. As seen in Figure A.5, which reports the results of four locally weighted regressions (assuming a bandwidth of 0.8) for each season, our quadratic specification with interaction terms with seasonal dummies is well-justified by the data.

Figure A.5: Daily Load and Daily Net Cleared Virtual Bids as a Percentage of Daily Load



Given the estimated coefficients from (22), we next obtain point estimates of net virtual bids by hour h, $N\hat{V}B_h$, using data on PJM load by hour L_h and seasonal dummies by hour: $D_{W,h}$, $D_{Sp,h}$, $D_{Su,h}$, and $D_{F,h}$. Finally, $z_l = \frac{\sum_{h \in \mathcal{H}_l} N\hat{V}B_h}{\sum_{h \in \mathcal{H}_l} L_h}$, where $\mathcal{H}_l$ is the set of hours within the year 2016 that are assigned to load segment l, as defined above. These values are plotted (in percentage terms) against total load segment load in PJM in Figure A.6 for each of the four seasons.

Figure A.6: Segment Load and Predicted Daily Net Cleared Virtual Bids as a Percentage of Daily Load, $z_l \times 100$



C.8 Numerical Calibration

Numerical calibration amounts to a type of nonlinear least squares estimation (see, e.g., Amemiya, 1977) across the numerical simulation model specified above whereby identification of a vector of numerically calibrated parameters relies on matching, as closely as possible, various moments predicted by the numerical model against observed data using a single instantiation of data. The method we adopt to numerically calibrate our model is in the spirit of the one-step non-linear least squares estimation procedure developed in Ferreyra (2007). We select parameters for numerical calibration (γ_{PA}^{2016}, $\gamma_{\neg PA}^{2016}$, A_{ihl} for all $ih = 1, ..., 5$ and all $l = 1, ..., 96$ in 2016, γ_{PA}^{2017}, $\gamma_{\neg PA}^{2017}$, and C_{tech}^{N} for all $tech \in \{NGCC, solar, wind\}$ for 2017) for which the data necessary to analytically calibrate or estimate the parameter is scarce or incomplete. Our numerical calibration approach proceeds in three stages. The first stage relies entirely on 2016 data and the analytically calibrated and/or estimated parameters from above for 2016. The second stage uses the same as well as the numerically calibrated parameters for 2016 from the first stage. The third stage, uses the same as well as 2017 data and the analytically calibrated and/or estimated parameters from above for 2017. In the first stage, we search for γ_{PA}^{2016} and $\gamma_{\neg PA}^{2016}$, assuming that transmission is unconstrained and that new generation cannot be added[15] (and so the parameters C_{tech}^{N} and A_{ihl} vector are not needed), so that our vector of predicted annual REC prices in Pennsylvania and Rest of PJM in 2016 match observed 2016 REC prices. Second, we search over A_{ihl} for all $ih = 1, ..., 5$ and all $l = 1, ..., 96$, such that our vector of predicted locational marginal prices by region and load segment match observed 2016 LMPs by region and load segment. Third, we search for γ_{PA}^{2017}, $\gamma_{\neg PA}^{2017}$, C_{NGCC}^{N}, C_{solar}^{N}, and C_{wind}^{N} so that so that our vector of predicted annual REC prices in Pennsylvania and Rest of PJM in 2017 match observed 2017 REC prices and so that our vector of total new capacity by technology matches observed total new capacity by technology in 2017. We next detail the precise methodology used to perform numerical calibration in each stage.

C.8.1 Share of External RECs Supplied in 2016

We search for $\mathbf{\Theta}_1 = \{\gamma_{PA}^{2016}, \gamma_{\neg PA}^{2016}\}$ using Feasible Weighted Least Squares to account for heteroskedasticity across predicted outputs. The first stage of Feasible Weighted Least Squares minimizes the following loss function which captures the Euclidean distance between the model's predicted REC prices by tier in 2016 and observed REC prices by tier in 2016 in Pennsylvania (actual) and Rest of PJM (generation weighted):

$$L\left(\mathbf{\Theta}_1\right) = \sum_{t \in PA} \left(y_t^{PA,2016} - \hat{y}_t^{PA,2016}\left(\mathbf{\Theta}_1\right)\right)^2 + \sum_{t \in RPJM} \left(y_t^{RPJM,2016} - \hat{y}_t^{RPJM,2016}\left(\mathbf{\Theta}_1\right)\right)^2, \quad (23)$$

where $\hat{y}_t^{PA,2016}\left(\mathbf{\Theta}_1\right)$ are the model predicted REC prices by tier t under Pennsylvania's AEPS in 2016 and $\hat{y}_t^{RPJM,2016}\left(\mathbf{\Theta}_1\right)$ are the model predicted state-generation weighted average REC prices by tier t across states in Rest of PJM with RPSs in 2016, which are calculated from the unweighted REC prices by tier and state in PJM predicted by the model, i.e., the Lagrange multipliers/shadow prices on (12) after solving for the competitive equilibrium prior to Pennsylvania joining RGGI in 2016. $y_t^{PA,2016}$ and $y_t^{RPJM,2016}$ are exogenous data that are calculated using the REC prices in 2106

[15] Recall, that our population of existing EGUs contains all EGUs as of 2016.

by tier for all states with RPS's in PJM taken from S&P Global's *Market Intelligence - California Carbon/RGGI Allowances* (accessed on 02/19/2018 at https://platform.marketintelligence.spglobal.com/web/client?auth=inherit#markets/co2AndRGGIAllowances). Data on generation used to construct the generation-weights for RPJM in this first phase are calculated from first running the model for 2016 without any RPS constraints. The residuals from the $\mathbf{\Theta}_1$ that solves this regression are used to calculate $\hat{\sigma}^{PA,2016}$ and $\hat{\sigma}^{RPJM,2016}$ or the standard deviation in residuals for PA and RPJM, given variation across tiers within each Pennsylvania/RPJM. We also construct new generation weights from PJM given the $\mathbf{\Theta}_1$ that solves this regression.

The second stage runs nonlinear least squares on variables transformed to account for heteroscedasticity. That is, we search for $\mathbf{\Theta}_1$ that minimizes the following loss function of the transformed variables:

$$\begin{aligned} \tilde{L}\left(\mathbf{\Theta}_1\right) = & \sum_{t \in PA} \left(y_t^{*PA,2016} - \hat{y}_t^{*PA,2016}\left(\mathbf{\Theta}_1\right)\right)^2 \\ & + \sum_{t \in RPJM} \left(y_t^{*RPJM,2016} - \hat{y}_t^{*RPJM,2016}\left(\mathbf{\Theta}_1\right)\right)^2, \end{aligned} \tag{24}$$

where * denotes normalization by $\hat{\sigma}^{PA,2016}$ on the first term on the right-hand side of (24) and normalization by $\hat{\sigma}^{RPJM,2016}$ on the second term on the right-hand side of (24). Since model predictions appear to be well-behaved in the γ_{PA}^{2016} and $\gamma_{\neg PA}^{2016}$ parameter space, we solve both stages of this regression using MATLAB's *fmincon* function subject to constraints that γ_{PA}^{2016} and $\gamma_{\neg PA}^{2016}$ must individually be on the interval $[0,1]$ and with starting values in the first regression of one in both Pennsylvania and RPJM and starting values in the second regression equal to the solution from the first regression.

The final γ_{PA}^{2016} and $\gamma_{\neg PA}^{2016}$ that solve (24) are 0.615 and 0.814, respectively. This means that 61.5% of all the available external RECs from Pennsylvania exclusive external REC suppliers are supplied to Pennsylvania, and 81.4% of all the available external RECs from the remaining PJM external REC suppliers are supplied to all PJM states (including, possibly, Pennsylvania).

C.8.2 Effective Aggregate Amperage in 2016

For each load segment $l = 1, \ldots, 96$ we search for $\mathbf{\Theta}_2^l = \{A_{ihl}\}_{ih=1}^{5}$ using a similar approach as discussed in the previous section. However, in contrast to the parameters calibrated numerically in the previous section, our model is highly non-linear with respect to amperage. As such, we first perform an exhaustive grid search that relies on 100,000 annual model predictions for the year 2016. For each model solution, we specify $A_{ihl} = \bar{A}_{ih}$ for all $l = 1, \ldots 96$. We consider ten values of $\bar{A}_{ih}$ for each ih, from 200 to 2,000 amperes with increments of 200. This provides 100,000 ($= 10^5$) combinations of $\left\{\bar{A}_{ih}\right\}_{ih}^{5}$.[16] Given these model predictions, the first stage of Feasible Weighted Least Squares minimizes the following loss functions for each $l = 1, \ldots, 96$ which captures the Euclidean distances between the model's predicted nodal congestion LMPs by PJM RPAM model region i and load segment l and observed congestion LMPs by region i and load segment l (load weighted):

$$L_l\left(\mathbf{\Theta}_2^l\right) = \sum_{i=1}^{5}\left(y_{il} - \hat{y}_{il}\left(\mathbf{\Theta}_2^l\right)\right)^2, \tag{25}$$

where $\hat{y}_{il}\left(\mathbf{\Theta}_2^l\right)$ are the model predicted congestion LMPs in region i and load segment l in 2016, which are calculated from the LMPs in region i and load segment l predicted by the model, i.e., the Lagrange multipliers/shadow prices on (11) after solving for the competitive equilibrium prior to Pennsylvania joining RGGI in 2016. y_{il} are exogenous day-ahead congestion LMPs in region i and load segment l in 2016 that are load weighted twice, first based on load by load zone within each PJM RPAM load region and second by load across hours within each segment, as done similarly in Section C.1. y_{il} is calculated using observed day-ahead congestion LMPs by load zone and hour

[16] As each annual model solution determines market clearing for each of 96 load segments and five load regions and the solution for load segment l is independent of the solution for load segment $l' \neq l$, given our model assumptions, this is equivalent to solving 9.6 million ($= 100,000 \times 96$) market clearing solutions for all five load regions. One drawback of our approach is that our annual model solution, conditional on a particular $\left\{\bar{A}_{ih}\right\}_{ih}^{5}$ draw imposed across all load segments, also determines equilibrium annual RGGI allowance and REC prices which affects market clearing and therefore congestion LMPs in each load segment. An alternative would be to directly specify all possible combinations of A_{ihl} across both ih and l; for ten values of amperage, this would entail $10^{5\times 96} = 10^{4}90$ model evaluations which is computationally prohibitive. Moreover, we do not expect the predicted annual RGGI allowance and REC prices across the 100,000 $\left\{\bar{A}_{ih}\right\}_{ih}^{5}$ draws to vary substantially, and possibly impact the congestion LMPs we recover. As such, we believe that our current approach is sufficient.

from PJM's *Day-Ahead Hourly LMPs for 2016* dataset and using load by load zone and hour from PJM's *Hourly Load: Metered for 2016* dataset. The residuals from the $\mathbf{\Theta}_2^l$ that solves this regression for each $l = 1, ..., 96$ are used to calculate $\hat{\sigma}^l$ or the standard deviation in residuals across regions for load segment l.

The second stage runs nonlinear least squares on variables transformed to account for heteroscedasticity. That is, we search for $\mathbf{\Theta}_2^l$ that minimizes the following loss functions of the transformed variables:

$$\tilde{L}_l\left(\mathbf{\Theta}_2^l\right) = \sum_{i=1}^{5}\left(y_{il}^* - \hat{y}_{il}^*\left(\mathbf{\Theta}_2^l\right)\right)^2, \tag{26}$$

where * denotes normalization by $\hat{\sigma}^l$.

Once candidate solutions to the second stage have been identified for all $l = 1, ..., 96$, we then further refine our solutions by minimizing $\tilde{L}_l\left(\mathbf{\Theta}_2^l\right)$ for all $l = 1, ..., 96$ using MATLAB's *fmincon* function. In each case, we specify lower bounds that are equal to 10 amperes and upper bounds that are equal to 5,000 amperes for all links, and use the candidate solutions from minimizing (26) as starting values. The initial grid search, the solutions to the two stage Feasible Weighted Least Squares regressions for all $l = 1, ..., 96$, and the fmincon searches for all $l = 1, ..., 96$ are performed in parallel across 100 cores with access to ample memory and storage from Penn State's Roar supercomputer (formerly known as the Institute for Computational and Data Sciences Advanced CyberInfrastructure). Total computational time for this step is roughly three days.

C.8.3 Share of External RECs Supplied and Capacity Costs For New Generation in 2017

We search for $\mathbf{\Theta}_3 = \left\{\gamma_{PA}^{2017}, \gamma_{\neg PA}^{2017}, C_{NGCC}^N, C_{solar}^N, C_{wind}^N\right\}$ using a similar two stage approach as discussed in the previous sections. In this case, the first stage of Feasible Weighted Least Squares minimizes the following loss function which captures the Euclidean distance between the model's predicted REC prices by tier in 2017 and observed REC prices by tier in 2017 in Pennsylvania (actual) and Rest of PJM (generation weighted) as well as model predicted total PJM capacities expansion by technology in 2017 and observed total nameplate PJM capacities expansion by technology in 2017:

$$
\begin{aligned}
L\left(\boldsymbol{\Theta}_3\right) = & \sum_{tech}\left(y_{tech}^{N}-\hat{y}_{tech}^{N}\left(\boldsymbol{\Theta}_3\right)\right)^2 \\
& + \sum_{t\in PA}\left(y_t^{PA,2017}-\hat{y}_t^{PA,2017}\left(\boldsymbol{\Theta}_3\right)\right)^2 \\
& + \sum_{t\in RPJM}\left(y_t^{RPJM,2017}-\hat{y}_t^{RPJM,2017}\left(\boldsymbol{\Theta}_3\right)\right)^2,
\end{aligned} \tag{27}
$$

where $\hat{y}_{tech}^{N}\left(\boldsymbol{\Theta}_3\right)$ is the model predicted estimate of the total amount of new nameplate capacity of technology type $tech$ added to PJM in 2017 ($\bar{K}_{tech}^{N} = \sum_{s=1}^{14} \bar{K}_{s,tech}^{N}$), $\hat{y}_t^{PA,2017}\left(\boldsymbol{\Theta}_3\right)$ are the model predicted REC prices by tier t under Pennsylvania's AEPS in 2017 and $\hat{y}_t^{RPJM,2017}\left(\boldsymbol{\Theta}_3\right)$ are the model predicted state-generation weighted average REC prices by tier t across states in Rest of PJM with RPSs in 2017, which are calculated from the unweighted REC prices by tier and state in PJM predicted by the model, i.e., the Lagrange multipliers/shadow prices on (12) after solving for the competitive equilibrium prior to Pennsylvania joining RGGI in 2017. y_{tech}^{N}, $y_t^{PA,2017}$, and $y_t^{RPJM,2017}$ are exogenous data. Data on new capacity in 2017 by technology type in PJM, y_{tech}^{N}, is taken from PJM's *New Service Queue database* (accessed on 03/21/2018 at https://www.pjm.com/planning/services-requests/interconnection-queues.aspx). $y_t^{PA,2017}$ and $y_t^{RPJM,2017}$ are calculated using the REC prices in 2017 by tier for all states with RPS's in PJM taken from S&P Global's *Market Intelligence - California Carbon/RGGI Allowances* (accessed on 02/19/2018 at https://platform.marketintelligence.spglobal.com/web/client?auth=inherit#markets/co2AndRGGIAllowances).

Data on generation used to construct the generation-weights for RPJM in this third phase are calculated based upon our predicted model generation for 2016 at the conclusion of the second phase of the numerical calibration. The residuals from the $\boldsymbol{\Theta}_3$ that solves this regression are used to calculate $\hat{\sigma}^{tech}$, $\hat{\sigma}^{PA,2017}$, and $\hat{\sigma}^{RPJM,2017}$, given variation across technologies/tiers within Pennsylvania/RPJM.

As before, in the second stage we search for $\boldsymbol{\Theta}_3$ that minimizes the following loss function of the transformed variables:

$$\tilde{L}\left(\mathbf{\Theta}_3\right) = \sum_{tech}\left(y_{tech}^{*N} - \hat{y}_{tech}^{*N}\left(\mathbf{\Theta}_3\right)\right)^2 + \sum_{t\in PA}\left(y_t^{*PA,2017} - \hat{y}_t^{*PA,2017}\left(\mathbf{\Theta}_3\right)\right)^2 + \sum_{t\in RPJM}\left(y_t^{*RPJM,2017} - \hat{y}_t^{*RPJM,2017}\left(\mathbf{\Theta}_3\right)\right)^2, \tag{28}$$

where * denotes normalization by $\hat{\sigma}^{tech}$ on the first term on the right-hand side of (28), $\hat{\sigma}^{PA,2017}$ on the second term on the right-hand side of (28) and normalization by $\hat{\sigma}^{RPJM,2017}$ on the third term on the right-hand side of (28). Since model predictions appear to be well-behaved in the C_{tech}^N, γ_{PA}^{2017}, and $\gamma_{\neg PA}^{2017}$ parameter space, we solve both stages of this regression using MATLAB's *fmincon* function subject to constraints that γ_{PA}^{2017} and $\gamma_{\neg PA}^{2017}$ must individually be on the interval $[0,1]$ and with starting values in the first regression of one in both Pennsylvania and RPJM and starting values in the second regression equal to the solution from the first regression. The final γ_{PA}^{2017}, $\gamma_{\neg PA}^{2017}$, C_{NGCC}^N, C_{solar}^N, and C_{wind}^N that solve (28) are 0.651, 0.869, \$73,900, \$135,000, and \$135,580, respectively.

C.9 Emissions Calculations

Source emissions for pollutant $P = \mathrm{CO_2, SO_2, NOx, PM, VOC, NH_3, CO}$ from existing EGUs in our model for an hour in load segment l are calculated after first descaling predicted generation from representative existing EGU j (g_{jl}^E) to the subset of original existing EGUs $k(j)$ that correspond to it ($g_{k(j)l}^E$). This is then combined with linear marginal emissions rates for each original existing EGU k, $\phi_k^{E,P}$ to calculate the source emissions released by original EGU k in an hour in load segment l, according to:

$$e_{kl}^{E,P} = \phi_k^{E,P} g_{kl}^E. \tag{29}$$

Source emissions from new NGCC EGUs [17] predicted by our model for an hour in load segment

[17] Source emissions from new solar and wind EGUs are assumed to equal zero.

l are calculated using predicted generation from representative new EGU j in l, g_{kl}^{N}, and linear marginal emissions rates for each new EGU j, $\phi_j^{N,P}$, according to:

$$e_{jl}^{N,P} = \phi_j^{N,P} g_{jl}^{N}. \tag{30}$$

In addition, both (29) and (30) can be summed across all hours in all load segments within a year to obtain annual source emissions estimates:

$$\begin{aligned} e_k^{E,P} &= \sum_l \delta_l \phi_k^{E,P} g_{kl}^{E}, \text{ and} \\ e_j^{N,P} &= \sum_l \delta_l \phi_j^{N,P} g_{jl}^{N}. \end{aligned} \tag{31}$$

These estimates can be reported directly, regionally aggregated, and/or possibly combined with data on latitude and longitude to infer destination emissions across space and/or time. Details on how generation is down-scaled from representative existing EGU j to original existing EGU $k\left(j\right)$, how emissions rates by pollutant for existing and new EGUs are calculated, and how the latitude and longitude of existing and new EGUs are determined are next discussed in turn.

C.9.1 Downscaling of Existing Generation

In order to accurately calculate source emissions for original existing EGU k, we downscale generation predicted by the model for each representative existing EGU j in an hour in load segment l, g_{jl}^{E}, to generation from existing EGU k in an hour in load segment l, g_{kl}^{E}, for all original existing EGUs assigned to representative EGU j, $k\left(j\right) \in \mathcal{K}_j$. To do so, we first sort all original existing EGUs $k\left(j\right)$ in the set $\mathcal{K}_j$ from lowest to highest marginal costs, $MC_{k(j)l}^{E}$. Based upon this sort, we determine the merit order in which original units k in bin j are dispatched in l and calculate cumulative effective capacity for l across all k in bin j (see, equation (18)). Where cumulative effective capacity equals g_{jl}^{E} determines which subset of the original existing EGUs in j are pseudo-dispatched. Define the marginal costs when this occurs as $MC_{k^*(j)l}^{E}$. Those original EGUs k with $MC_{k(j)l}^{E} < MC_{k^*(j)l}^{E}$, are assigned generation equal to their effective capacity. For the binding original EGU k^* their generation is set equal to the difference between g^E_{jl} and cumulative effective capacity evaluated at $k^* - 1$. Finally, those original EGUs k with $MC^E_{k(j)l} > MC^E_{k^*(j)l}$, are assigned generation equal to zero.

C.9.2 Emissions Rates

CO_2, SO_2, and NOx

Similar to the procedure used to assign heat rates to existing EGUs within our original population of units as discussed in Section C.2.3, we first assign CO_2, SO_2, and NOx emissions rates to each existing EGU $k(j)$ using the best available estimate across four datasets (from most to least accurate): CEMS, EPA's eGrid, and S&P Global's *PJM Summary Capacity*.[18] For a small subset of remaining existing EGUs for which data on emissions rates does not exist across these four datasets, we impute emissions rates. Each of these steps is discussed in detail below, although the assignment of emissions rates by dataset/method is provided in Table A.6.

Table A.6: CO_2, SO_2, and NOx Emissions Factor Assignments to Existing EGUs

Dataset/Method	EGUs Assigned Emissions Rates
Final Dataset	3,095
CEMS	374
eGrid	1,490
S&P Global	801
Imputed	369
Remaining Solar/Wind/Hydro	61

To assign CO_2 emissions rates to original existing EGU k, we first use the 2016 CEMS. CEMS reports measured real-time annual CO_2 emissions and annual generations used for regulatory compliance and which we suspect are the most accurate. We calculate annual CO_2 emission factors using the 2016 CEMS data by dividing the total annual CO_2 emissions for each CEMS EGU by the EGU's total annual generation. For the 374 EGUs in our original population that are also in CEMS, we assign those units CEMS CO_2 emissions rates. Next, we assign CO_2 emissions rates to 1,490 EGUs in our original population using the CO_2 emissions rates reported in eGrid.[19] Third, we

[18] S&P Global's *PJM Summary Capacity* SO_2 and NOx emissions rates are similar to those reported in NEEDS. However, NEEDS does not report CO_2 emissions rates and so we use S&P Global's *PJM Summary Capacity* emissions rates for CO_2, SO_2, and NOx emissions.

[19] We exclude from this step EGUs in our population and eGrid whose heat rates are above 42,545.31 BTU/kWh

assign CO_2 emissions rates to an additional 801 EGUs using the CO_2 emissions rates reported in S&P Global's *PJM Summer Capacity* dataset.[20] Finally, 430 remaining EGUs are assigned imputed CO_2 emissions rates equal to the average CO_2 emissions rates of all existing EGUs with assigned CO_2 emissions rates of the same fuel type and located in the same state as the remaining EGU.

To assign SO_2 and NOx emissions rates to each existing EUG k, we first assign CEMS, eGrid, S&P Global, and NEEDS SO_2 and NOx emissions rates to 2,665 existing EGUs in our original population the same way we assign CEMS, eGrid, and S&P Global CO_2 emissions rates previously (374 EGUs get assigned SO_2 and NOx emissions rates from CEMS; 1,490 EGUs get assigned SO_2 and NOx emissions rates from eGrid; and 801 EGUs get assigned SO_2 and NOx emissions rates from S&P Global). Finally, we use OLS regressions to impute SO_2 and NOx emissions rates for 369 of the remaining 430 existing EGUs in our original population. We exclude from these regressions EGUs that have fuel-types of wind, hydro, solar (61 EGUs), whose SO_2 and NOx emissions rates are assumed to be zero. For $P = \mathrm{SO_2}, \mathrm{NOx}$, we estimate the following two OLS regressions using data on assigned existing EGUs:

$$\begin{aligned} \phi_k^{E,P} \quad = \quad & \beta_0 + \beta_1 fuel_k + \beta_2 \psi_k^E + \beta_3 \bar{K}_k^E + \beta_4 fuel_k \times \psi_k^E + \beta_5 fuel_k \times \bar{K}_k^E \\ & + \beta_6 \left(\psi_k^E\right)^2 + \beta_7 \left(\bar{K}_k^E\right)^2 + \beta_8 fuel_k^E \times \left(\psi_k^E\right)^2 + \beta_9 fuel_k^E \times \left(\bar{K}_k^E\right)^2 + \varepsilon_k, \end{aligned} \tag{32}$$

where $\phi_k^{E,P}$ is the emission factor of existing EGU k for pollutant P, $fuel_k$ are dummy variables indicating the fuel type of existing EGU k, ψ_k^E is the heat rate of existing EGU k, $\bar{K}_k^E$ is the nameplate capacity of existing EGU k, ε_k is the regression error term, and the $\boldsymbol{\beta}$ are the regression coefficients to be estimated. The R^2 for the SO_2 emissions rates and NOx emissions rates regressions, are 0.612 and 0.572, respectively. Given the estimated parameters from (32), we then assign emissions rates for pollutant P equal to the predicted values for those remaining existing EGUs given data on their fuel types, heat rates, and capacities.

and/or whose annual generation is reported to be negative since errors in CO_2 emissions rates across units appear to be correlated with errors in these other observables. We parse heat rates in a similar fashion as discussed SectionC.2.3.

[20] Again, we exclude from this step EGUs with S&P Global heat rates that are above 26,473.842 BTU/kWh; see, Section C.2.3.

The CO_2, SO_2, and NOx emissions rates assigned to new NGCC EGUs located in a particular state s equal the average of the CO_2, SO_2, and NOx emissions rates from existing NGCC EGUs in state s.

PM, NH_3, VOC, CO

To calculate the PM (PM_{25} and PM_{10}), NH_3, VOC, and CO emissions rates for the original population of existing EGUs, we use the EPA's *Flat File Generation Methodology v514* (accessed on 05/19/2020 at https://www.epa.gov/sites/production/files/2015-07/documents/flatfile_methodology.pdf) with slight modifications regarding assignments of sulfur and ash contents. The calculation steps outlined in these documents are used by the EPA for post-processing PM (PM_{25} and PM_{10}), NH_3, VOC, and CO emissions from model output generated by the Integrated Planning Model (IPM).

We first use generation weighted sulfur content and ash contents by state and PJM region from the *EPA 5-13_Base_Case RPE Replacement File* from the *Results Using EPA's Base Case v.5.13, Detail IPM output files*, (accessed on 05/27/2020 at https://www.epa.gov/sites/production/files/2015-08/epa_base_case_v513_data_files.zip), then we link the heat contents with these sulfur and ash contents from *Table 9-5, page 276 (9-13), Documentation for EPA Base Case v.5.13 Using the Integrated Planning Model* (accessed on 04/05/2020 at https://www.epa.gov/sites/production/files/2015-07/documents/documentation_for_epa_base_case_v.5.13_using_the_integrated_planning_model.pdf). After that we use the *PMAshSulfurContent* tab in the *flatfile_inputs_1* (accessed on 04/27/2020 on https://www.epa.gov/sites/production/files/2015-07/flatfile_inputs_1.xls) to assign sulfur and ash contents for the EGUs that are still not assigned sulfur and ash contents from the previous step. Finally, we use *Table 12, Flat File Generation Methodology v514* to assign sulfur and ash contents to the remaining EGUs.

After assigning sulfur and ash contents for the all the EGUs in our original population, we follow the *Flat File Generation Methodology v514* using the imputed sulfur and ash contents to calculate PM (PM_{25} and PM_{10}), NH_3, VOC, and CO emissions rates for all the EGUs. While the *Flat File Generation Methodology v514* directly calculates total annual emissions for these pollutants, we instead calculate linear emissions rates for these pollutants by using heat rates instead of heat inputs in all of the formulas in *flat File Generation Methodology v514*.

The PM, NH_3, VOC, and CO emissions rates assigned to new NGCC EGUs located in a particular state s equal the average of the PM, NH_3, VOC, and CO emissions rates from existing NGCC EGUs in state s.

Emissions Calculation Verification

To verify the reliability our PM, NH_3, VOC, and CO emissions calculations, we calculate the 2016 annual emissions for each of these $P = \mathrm{PM}, \mathrm{NH}_3, \mathrm{VOC}, \mathrm{CO}$ pollutants using the methodology detailed in Section C.9.2 above and EPA model predictions on generation which we then compare against the EPA's own post-processing prediction of annual emissions based on the same underlying generation data. To do this, we use three data files: NEEDS v513 (2016), the *Web-Ready_Parsed_File_EPA5-13_Base_Case_2018*, and the *FlatFile_EPA513_BC_7c_2018_20131108* file (both accessed on 05/27/2020 at https://www.epa.gov/sites/production/files/2015-08/epa_base_case_v513_data_files.zip). We first merge the NEEDS dataset with the *Web-Ready_Parsed_File_EPA5-13_Base_Case_2018* to link total annual generation from the *Web-Ready_Parsed_File_EPA5-13_Base_Case_2018* to EGUs in NEEDS. Second, we multiply our calculated emissions rates times annual generation to calculate our estimate of imputed annual emissions for each pollutant P for all of the EGUs in the NEEDS. We then merge the combined dataset to the *FlatFile_EPA513_BC_7c_2018_20131108* which has EPA's calculations of annual emissions. For the sample of NEEDS units our predicted emissions estimates for pollutant P match almost exactly with the EPA's calculated emissions for those same units.

C.9.3 Latitude and Longitudes of Generation Sources

In order to calculate destination emissions at a particular location from source emissions released by original existing EGUs, we need to know the latitude and longitude of each existing EGU in our original population. We assign geographical locations to all existing EGUs k in our original population. To do so, we use the EIA's *United States Power Plant Map* (accessed on 04/23/2020 at https://www.eia.gov/maps/map_data/PowerPlants_US_EIA.zip). The dataset includes attribute information (e.g., ORIS plant codes, unit IDs, physical street addresses, and longitudes and latitudes) on all operable EGUs in the United States by energy source, including all EGUs that are operating, on standby, or are out of service in the short or long term, with a nameplate capacity of 1 MW or more. Using ORIS plant codes and unit IDs, we are able to match 2,882 EGUs from this dataset to the same number of existing EGUs in our original population of units. Only 213 existing EGUs from our original population remain unmatched. These are either existing EGUs that are newer and might not yet be included in the *United States Power Plant Map* dataset, and/or are smaller than 1 MW. For 201 of these 213 remaining EGUs, we manually perform web searches and assign street addresses and coordinates to them. The remaining 12 unmatched existing EGUs from our original population are small solar PVs, hydro and biomass EGUs, which we assign coordinates equal to the centroids of the counties in which they are located. Finally, new EGUs are added at the state level and thus are assigned latitude and longitude coordinates equal to the centroids of the states in which they are located.

C.10 Calibration Results

This section presents the results from our replicated calibrated baseline for 2016 and 2017 across several dimensions: generation (Tables A.7 and A.8), new capacity (Table A.9 for 2017 only), locational marginal prices (Figures A.7 through A.10), renewable energy credit prices (Tables A.10 and A.11), and CO_2 emissions.

Table A.7 reports our prediction of 2016 generation in Pennsylvania and Rest of PJM by fuel type against observed 2016 data on generation from Monitoring Analytics' *2016 State of the Market Report, Table 3-8 PJM Generation (By fuel source (GWh)): 2015 and 2016*, page 105 (accessed on 09/27/2017 at https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2016/2016-som-pjm-volume2.pdf). In 2016, we slightly under-predict generation in Pennsylvania and Rest of PJM by -0.3%. Much of this is driven by small under-predictions of coal in both regions of -0.7% and -0.5% in Pennsylvania and Rest of PJM, respectively. We also slightly over-predict gas and under-predict generation from other sources in Pennsylvania and observe the opposite in Rest of PJM. Table A.8 presents a similar comparison of predicted and observed generation in

2017. The model does a slightly better job of predicting total generation in both Pennsylvania and Rest of PJM although we under-predict coal by slightly more than in 2016.

Table A.7: Comparison of Calibrated Baseline Model Prediction of PJM Generation by Fuel Type (GWh) to 2016 Historic Data

		Pennsylvania			Rest of PJM	
Fuel Type	Observed	Model Predicted	% Difference	Observed	Model Predicted	% Difference
Coal	54,672	54,310	-0.7%	220,609	219,465	-0.5%
Nuclear	82,924	82,924	0.0%	196,622	196,662	0.0%
Gas	68,048	68,321	0.4%	146,974	146,104	-0.6%
Hydro	2,374	2,375	0.0%	11,312	11,312	0.0%
Wind	3,476	3,476	0.0%	14,240	14,240	0.0%
Oil	363	355	-2.5%	1,800	1,780	-1.2%
Solar	75	75	0.0%	945	944	-0.0%
Biomass	1,883	1,857	-1.4%	4,017	4,035	0.5%
Other	1,250	699	-44.0%	958	1,311	36.9%
Total	215,067	213,392	-0.3%	597,478	595,813	-0.3%

Table A.8: Comparison of Calibrated Baseline Model Prediction of PJM Generation by Fuel Type (GWh) to 2017 Historic Data

		Pennsylvania			Rest of PJM	
Fuel Type	Observed	Model Predicted	% Difference	Observed	Model Predicted	% Difference
Coal	47,634	46,649	-2.1%	208,980	206,627	-1.1%
Nuclear	83,200	83,384	0.2%	204,376	204,447	0.0%
Gas	72,503	73,388	1.2%	144,255	142,576	-1.2%
Hydro	2,518	2,499	-0.8%	12,350	12,497	1.2%
Wind	3,591	3,575	-0.4%	17,124	18,315	7.0%
Oil+Other	1,692	879	-45.1%	738	1,572	113.0%
Solar	70	69	-0.4%	1,399	1,933	38.2%
Biomass	1,916	2,292	19.7%	5,974	5,924	-0.8%
Total	213,034	212,736	-0.1%	595,196	593,890	-0.2%

Table A.9 presents our prediction of new capacity expansion in 2017 (the first year in which new capacity can be added) in PJM by technology type against observed 2017 data on new capacity expansion from *New Service Queue database* (accessed on 03/21/2018 at https://www.pjm.com/planning/services-requests/interconnection-queues.aspx). Even though we only consider variability in capacity costs by technology and not across regions, the model does a remarkable job predicting new capacity additions in both Pennsylvania and Rest of PJM for all technology types.

Table A.9: Comparison of Calibrated Baseline Model Prediction of PJM New Capacity Expansion (MW) by Technology Type and Region to 2017 Historic Data

	Pennsylvania			Rest of PJM		
Technology Type	Observed	Model Predicted	% Difference	Observed	Model Predicted	% Difference
NGCC	1,340	1,270	-5.2%	1,766	1,626	-7.9%
Wind	0	0	0.0%	126	156	23.8%
Solar	0	0	0.0%	204	194	-4.4%

Figure A.7 graphically compares our prediction of the 2016 energy locational marginal price (LMP) against observed 2016 data on the energy LMP from textitDay-Ahead Hourly LMPs for 2016 dataset (last accessed 05/15/2017 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/da_hrl_lmps) across seasons and load segments (first peak is winter, second is spring, third is summer, and fourth is fall). The energy LMP reflects the marginal fuel costs from the marginal EGU that clears the PJM market. The energy LMP reflects the market clearing price prior to the realization of congestion mark-ups captured in the congestion LMP and as such is driven by seasonal variation across load segments arising from variation in load, fuel prices, effective capacity, and virtual bids. In general, these attributes of the model explain almost all of the energy LMP variation across seasons and load segments in the observed energy LMPs in 2016.

Figure A.7: Comparison of Calibrated Baseline Model Prediction of Energy LMP to 2016 Historic Data

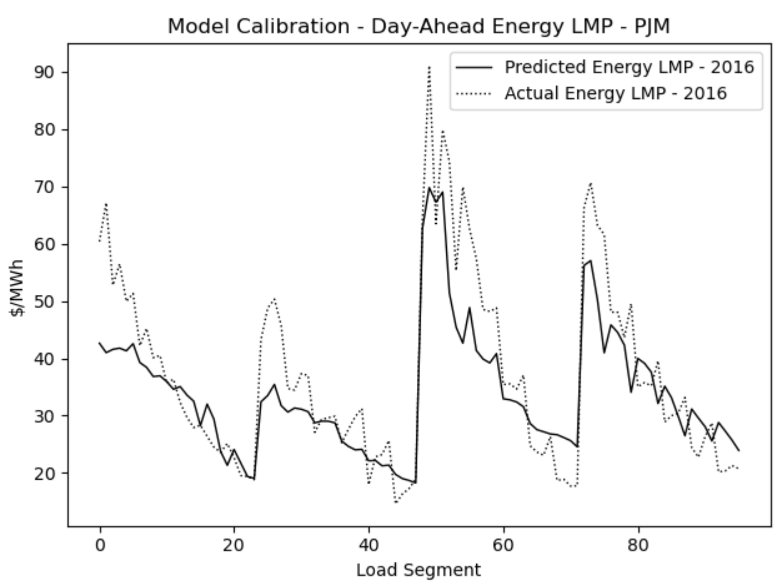


Figure A.8 depicts our prediction of 2016 congestion LMPs against observed 2016 data on congestion LMPs from textitDay-Ahead Hourly LMPs for 2016 dataset (last accessed 05/15/2017 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/da_hrl_lmps) across regions, seasons, and load segments. Variation in congestion LMPs across model regions is driven largely by the numerically calibrated effective amperages across the five RPAM model links. As is clearly shown, the numerical calibration of effective amperages does an exceptional job of explaining the variation in congestion LMPs observed in 2016. Moreover, as a share of the total LMPs the variation in congestion LMPs explains a substantial proportion of the variation in the total LMPs discussed next.

Figure A.8: Comparison of Calibrated Baseline Model Prediction of Congestion LMPs ($/MWh) by PJM Region to 2016 Historic Data

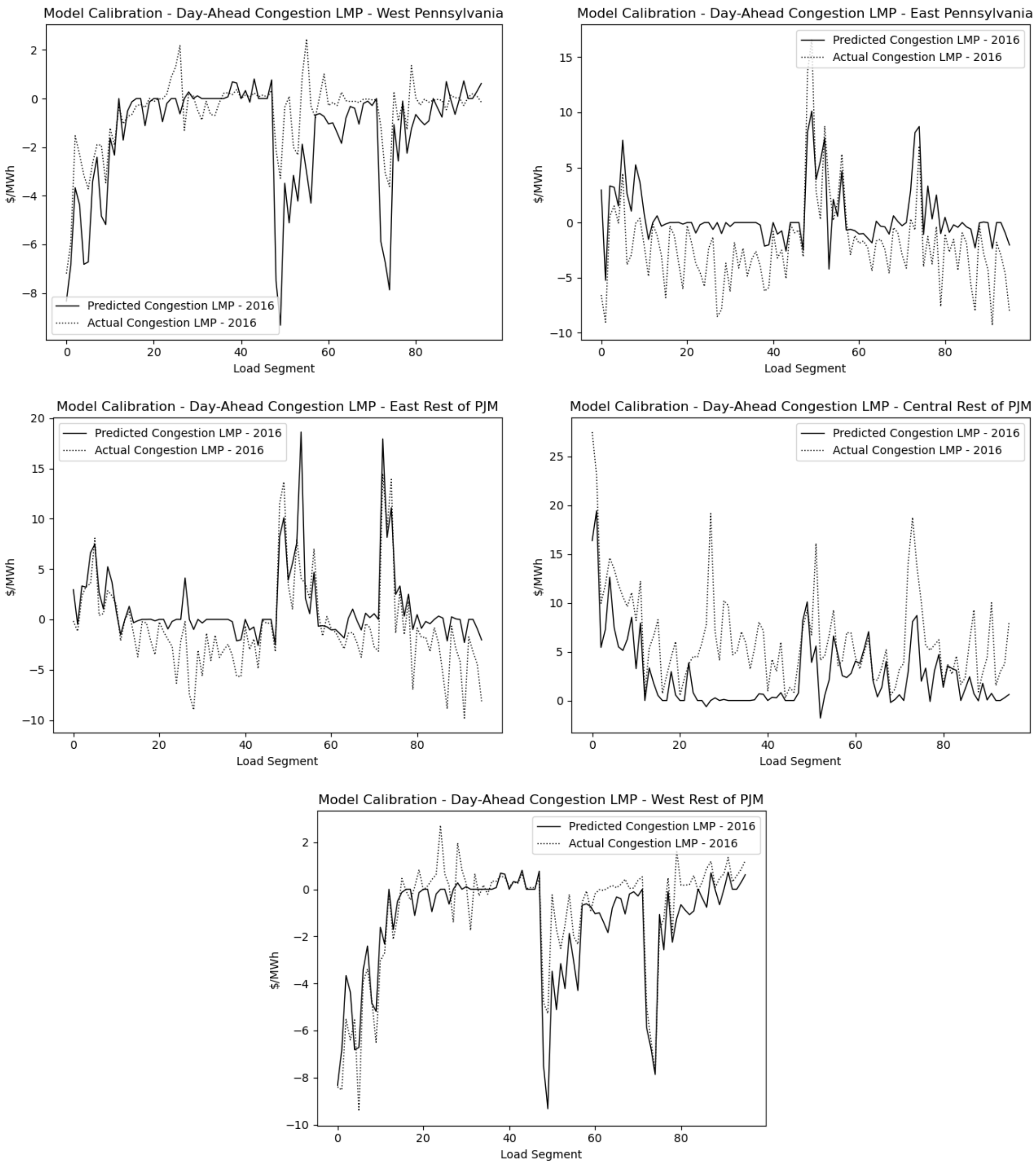


Figure A.9 depicts our prediction of 2016 day-ahead LMPs against observed 2016 data on day-ahead LMPs from textitDay-Ahead Hourly LMPs for 2016 dataset (last accessed 05/15/2017 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/da_hrl_lmps) across regions, seasons, and load segments. The day-ahead LMP reflects the sum of the energy LMP, congestion LMP, and a small component arising from the marginal pre-existing regulatory costs of the marginal EGU. The model under-predicts observed LMPs during peak demand periods and slightly over-predicts during off-peak periods in each season and region. However, the bulk of the variation in observed LMPs across seasons and regions in 2016 is well explained by the model. Figure A.10 repeats this analysis for 2017. Again the model consistently explains a great deal of the observed variation in LMPs across regions, seasons, and load segments, although there is greater within season variability since load segments are sorted based upon lowest to highest load in 2016.

Figure A.9: Comparison of Calibrated Baseline Model Prediction of Day-Ahead LMPs ($/MWh) by PJM Region to 2016 Historic Data

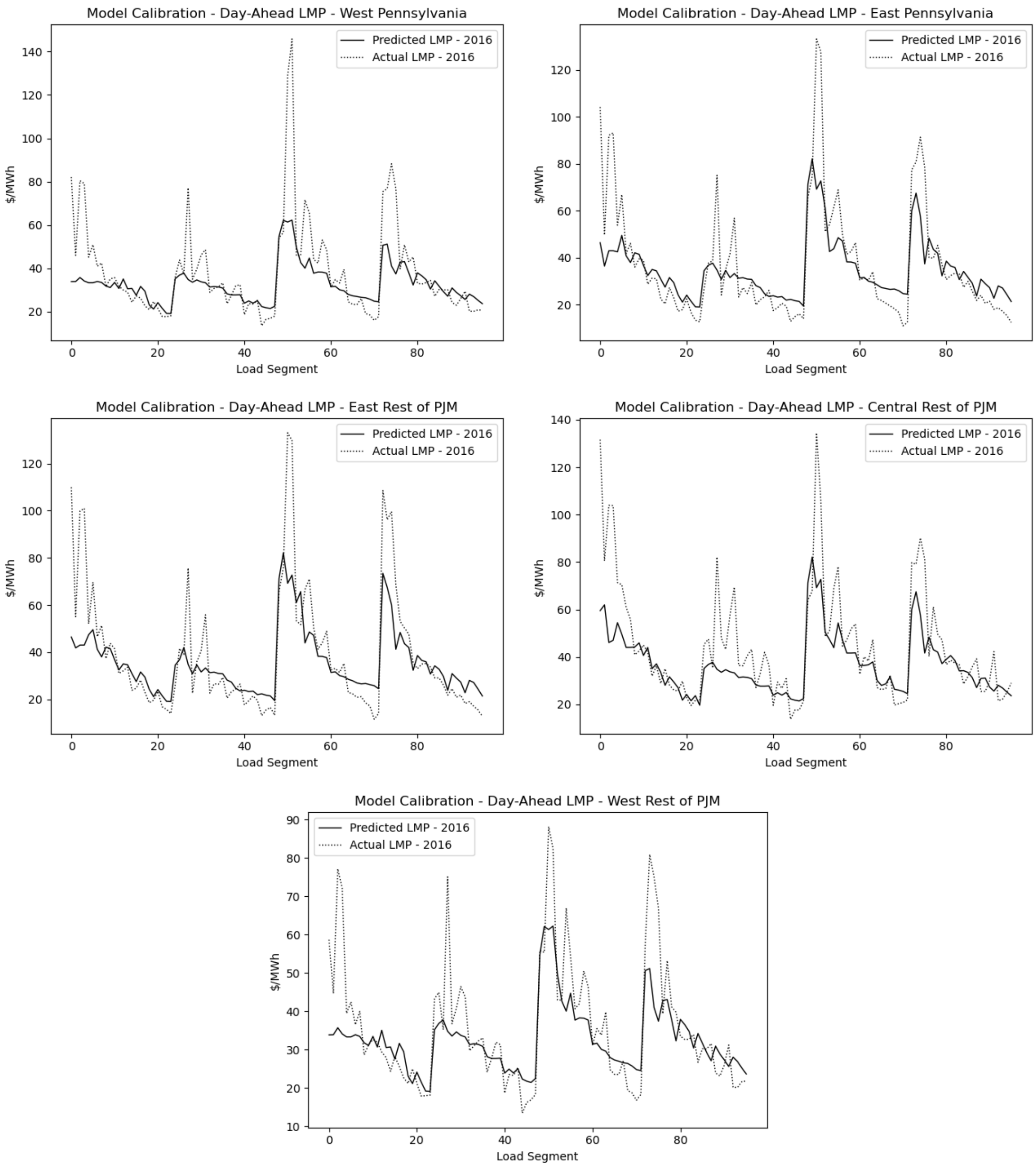


Figure A.10: Comparison of Calibrated Baseline Model Prediction of Day-Ahead LMPs ($/MWh) by PJM Region to 2017 Historic Data

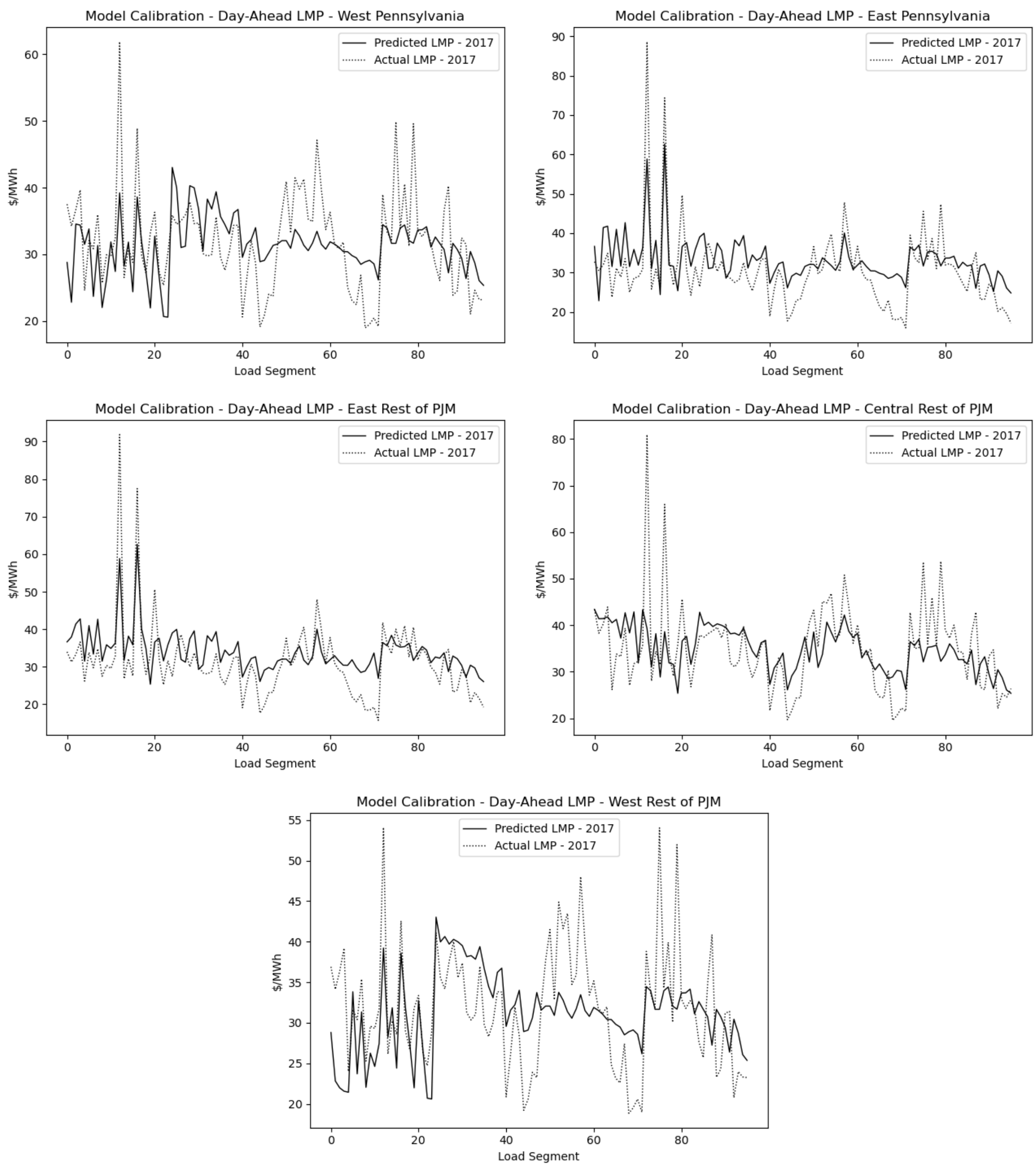


Table A.10 reports our prediction of 2016 generation weighted REC prices against observed 2016 data, using 2016 REC prices from S&P Global's *Market Intelligence - California Carbon/RGGI Allowances* (accessed on 07/02/2017 at https://platform.marketintelligence.spglobal.com/web/client?auth=inherit#markets/co2AndRGGIAllowances) and our predicted 2016 generation. In 2016 we slightly over-predict REC prices under Pennsylvania's AEPS across all three tiers, whereas we slightly under-predict across all three tiers the generation weighted average REC price across states in Rest of PJM with RPSs. Table A.8 repeats this analysis for 2017. Pennsylvania over-predicts by slightly more across all three tiers and Rest of PJM also over-predicts slightly across all three tiers.

Table A.10: Comparison of Calibrated Baseline Model Prediction of Generation Weighted REC prices ($/MWh) in Pennsylvania and Rest of PJM to 2016 Historic Data

	Pennsylvania			Rest of PJM		
AEPS/RPS Tier	Observed	Model Predicted	% Difference	Observed	Model Predicted	% Difference
Tier 1	$4.35	$4.60	5.7%	$46.00	$36.39	-20.9%
Tier 2	$12.45	$13.17	5.8%	$3.99	$3.18	-20.3%
SREC	$0.0028	$0.0029	3.6%	$0.48	$0.38	-20.8%

Table A.11: Comparison of Calibrated Baseline Model Prediction of Generation Weighted REC prices ($/MWh) in Pennsylvania and Rest of PJM to 2017 Historic Data

	Pennsylvania			Rest of PJM		
AEPS/RPS Tier	Observed	Model Predicted	% Difference	Observed	Model Predicted	% Difference
Tier 1	$2.95	$4.12	39.7%	$33.38	$35.53	6.4%
Tier 2	$8.44	$11.81	39.9%	$2.89	$3.08	6.6%
SREC	$0.0019	$0.0027	42.1%	$0.35	$0.37	5.7%

In 2016, RPAM predicts total CO_2 emissions of 359.6 MMT or just 0.3% more CO_2 emissions than the 358.6 MMT reported by Monitoring Analytics' *2016 State of the Market Report.* In contrast in 2017, RPAM under-predicts CO_2 emissions by -3.0%, or 364.4 MMT relative to the 375.7 MMT in Monitoring Analytics' *2017 State of the Market Report.* This is largely attributable to the slightly greater under-prediction of coal generation in 2017.

D Intertemporal Dynamics

The numerical model generates a time path of economic outcomes at one year intervals between 2018 and 2030. To account for underlying dynamic trends that alter our model predictions outside of the economic phenomena explicitly represented in RPAM, we allow for load, fuel prices, external RECs, and the costs of adding new capacity to adjust exogenously each year.

We assume that annual load grows 0.42% per year in PJM. This equals the average annual growth rate in PJM load observed between 2013-2018, calculated from annual loads provided in the *PJM Load Forecast Report*.

Daily natural gas spot price for 2017 is taken from Bloomberg's *Daily Natural Gas Spot Price* dataset for 2017 and converted into natural gas spot price by load segment for 2017 using the same method mentioned in Section C.2.3 for year 2016. Transportation prices for natural gas for 2017 and going forward are kept constant at the 2016-level. Coal, oil and uranium in 2017 are calculated as coal, oil and uranium in 2016 as assigned in C.2.3 times their 2016-2017 growth rates by census region. 2016-2017 delivered growth rates for these fuels are calculated from the EIA's *Annual Energy Outlook 2018, Table: Energy Prices by Sector and Source, Case: Reference case*, which has observed coal, oil and uranium prices for the electric power by census region in 2016 and 2017 in 2017 \$. We calculate 2016-2017 growth rates for these coal, oil and uranium based on observed coal, oil and uranium prices in 2016 and 2017 after converting them to 2016\$. We assume that prices for other fuel types (biomass, other fuel) in 2017 and going onward stay the same as in 2016.

Similarly, delivered natural prices by load segment for 2018 are calculated from daily natural gas spot price for 2018 is taken from Bloomberg's *Daily Natural Gas Spot Price* dataset for 2018 and natural gas transportation prices fixed at the 2016-level. Coal, oil and uranium in 2018 are calculated as coal, oil and uranium times their 2017-2018 growth rates by census region. 2017-2018 delivered growth rates for these fuels are calculated from EIA's *Annual Energy Outlook 2019, Table: Energy Prices by Sector and Source, Case: Reference case*, which has observed coal, oil and uranium prices for the electric power by census region in 2017 and 2018 in 2018 \$. We calculate 2017-2018 growth rates for these coal, oil and uranium based on observed coal, oil and uranium prices in 2017 and 2018 after converting them to 2016\$.

From 2019 onward, natural gas, coal, oil and uranium fuel prices are assumed to grow following the fuel price growth rates by census region imputed from the fuel price projections by census region provided in the *Annual Energy Outlook 2018, Table: Energy Prices by Sector and Source, Case: Reference case*. These annual growth rates for each year after 2018 are calculated using the 2018 fuel price projections in AEO 2018 and the fuel price projections for each year after that, after converting them to 2016$. Compared to the 2018 baseline, the fuel price growth rates from 2019-2030 used in our model are as follows:

Table A.12: Fuel Price Growth Rates 2019 to 2030 compared to 2018 (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

The amount of external RECs available to supply to Pennsylvania exclusively (e.g., $\gamma_{PA}\sum_{q\in PA}\bar{r}_q$) are assumed to grow by 4% per year whereas the amount of external RECs available to supply all states in PJM with RPSs (e.g., $\gamma_{\neg PA}\sum_{q\in\neq PA}\bar{r}_q$) are assumed to grow by 5% per year. For $w=\{PA,\neg PA\}$ and $y=\{2016,2017\}$, these growth rates reflect the growth between 2016 and 2017 in external RECs available for supply to states in PJM with RPSs given the numerically calibrated γ_w^y's and data on the total amount of external RECs available (i.e., $\sum_{q\in w}\bar{r}_q^y$); that is, the growth rates equal: $\frac{\gamma_w^{2017}\sum_{q\in w}\bar{r}_q^{2017}-\gamma_w^{2016}\sum_{q\in w}\bar{r}_q^{2016}}{\gamma_w^{2016}\sum_{q\in w}\bar{r}_q^{2016}}$.

We assume a growth rate in the annual cost of adding new wind capacity of -1.7%. This comes from Wiser et al. (2016), in which the authors survey different wind technology experts and find that onshore wind costs would decline by 24 % by 2030 relative to 2014, which translates into an average annual growth rate of -1.7 %. We assume a growth rate in the annual cost of adding new solar capacity of -2.2%. The U.S. Department of Energy's *SunShot Initiative* is targeting a 50% reduction in utility scale solar photovoltaic costs between 2020 and 2030. Although the 2020 SunShot target for utility scale solar was achieved in 2017, this 2030 target is much more ambitious (SETO, 2017), so we instead assume a 25% reduction in costs by 2030 relative to 2017, which translates into an average annual growth rate of -2.2 %. We assume a growth rate of zero in the annual cost of adding new NGCC capacity.

E Model Validation

Comparison of Model Predictions to Historic Data

We calibrate the model to 2016 and 2017 so we are able to compare our model's predictions for 2018 against observed historic data for 2018 across several dimensions: generation, new capacity, locational marginal prices, and CO_2 emissions. RPAM is calibrated to capture average annual economic adjustments and it does not capture all idiosyncratic exogenous shocks unique to 2018 which explain the 2018 allocation observed in the historic data. As such, expecting RPAM to exactly predict a single year of observed data across all dimensions is not within the scope of this validation exercise. Moreover, it does not make sense to design the model to achieve such an outcome as it would likely do a less satisfactory job of explaining average annual outcomes.

Table A.13 presents our out-of-sample model prediction of generation in PJM by fuel type against observed 2018 data on generation from Monitoring Analytics' *2018 State of the Market Report, Table 3-9 Generation (By fuel source (GWh)): 2017 and 2018 on page 122* (accessed on 01/11/2019 at https://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2018/2018-som-pjm-volume2.pdf). Overall, we over-predict total generation by 0.5% reflecting a slight difference between our 2018 load given the intertemporal dynamics discussed in Section D and observed 2018 load in PJM. For the fuel types which contribute the most to generation in 2018, coal, nuclear, and natural gas, our predictions are off less than five percent in 2018. We under-predict Oil+Other and over-predict solar and biomass by more than five percent, but these errors are small in absolute terms.

Table A.13: Comparison of Out of Sample Model Prediction of PJM Generation by Fuel Type (GWh) to 2018 Historic Data

Fuel Type	Observed	Model Predicted	% Difference
Coal	239,612	229,691	-4.1%
Nuclear	286,115	287,678	0.6%
Gas	256,701	267,483	4.2%
Hydro	19,416	19,441	0.1%
Wind	21,628	22,596	4.5%
Oil+Other	3,581	2,480	-30.8%
Solar	2,111	3,192	51.2%
Biomass	8,390	9,044	7.8%
Total	837,594	841,605	0.5%

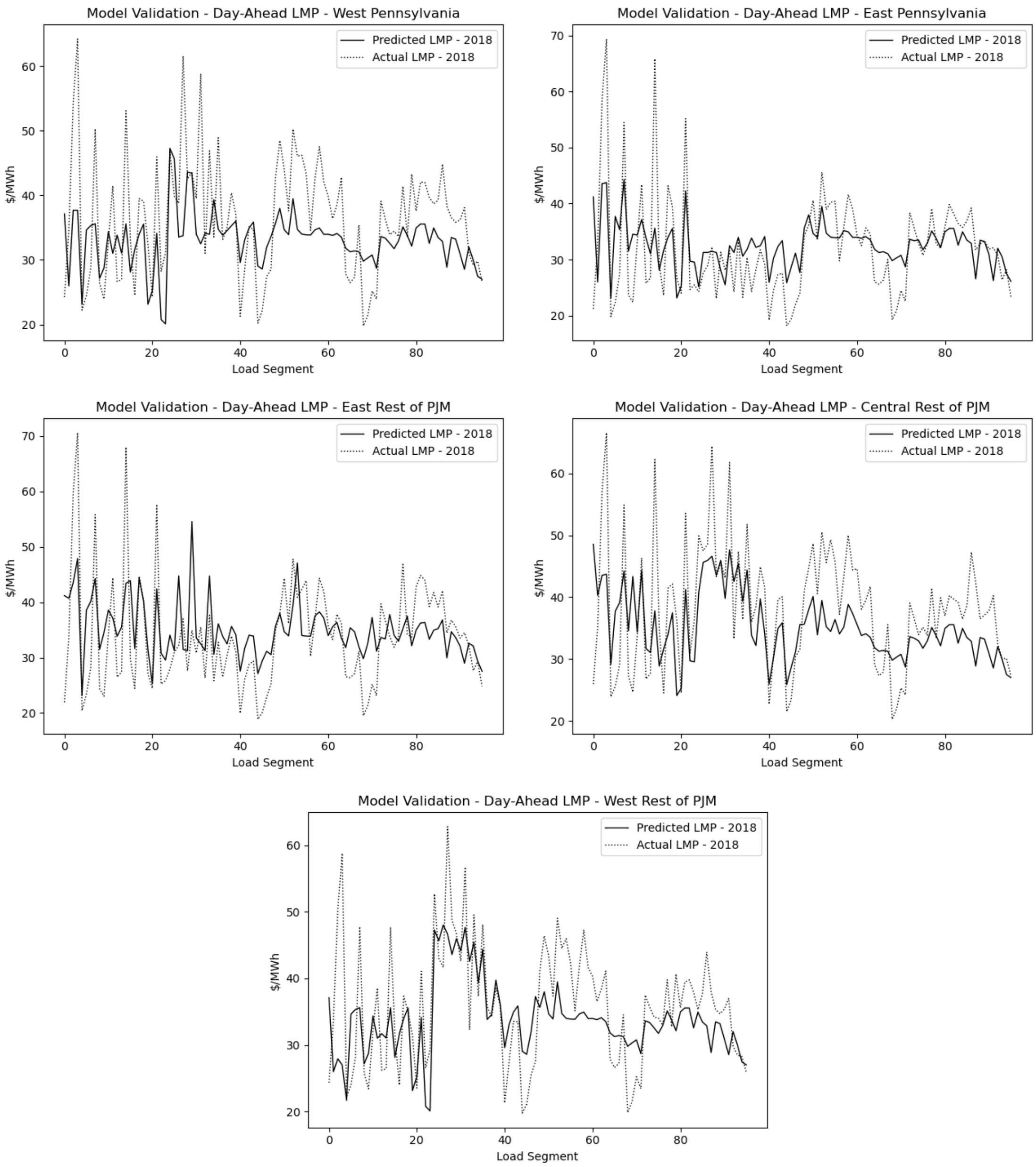
Table A.14 presents our out-of-sample model prediction of new capacity expansion in PJM by technology type against observed 2018 data on new capacity expansion from PJM's *New Service Queue database* (accessed on 04/21/2019 at https://www.pjm.com/planning/services-requests/interconnection-queues.aspx). In 2018 we over-predict new solar and wind capacity, but under-predict new NGCC capacity with total new capacity under-predicted slightly by -7.9%.

Table A.14: Comparison of Out of Sample Model Prediction of PJM New Capacity Expansion (MW) by Technology Type and Region to 2018 Historic Data

Technology Type	Observed	Model Predicted	% Difference
NGCC	9,442	7,850	-16.9%
Wind	762	987	29.5%
Solar	265	804	203.4%
Total	10,469	9,641	-7.9%

Figure A.11 graphically compares our out-of-sample model predictions of locational marginal prices in PJM by region and load segment against observed 2018 data on day-ahead locational marginal prices from textitDay-Ahead Hourly LMPs for 2018 dataset (last accessed 03/19/2019 online using *PJM Data Miner 2* at: https://dataminer2.pjm.com/feed/da_hrl_lmps), averaged based on variation in within region load (across load zones within each PJM RPAM model region) in each 2016 load segment. Overall, our out-of-sample predicted LMPs match remarkably well although with more variation than observed in the calibrated outcomes in 2016 and 2017 (see, Figures A.9 and A.9), likely driven by idiosyncratic shocks unique to 2018 which also drive a lot of LMP variation across region and load segments and which are exogenous to our model. While we initially had some reservations regarding the granularity in which we characterize load segments using 2016 load, these concerns do not appear to emerge in the out-of-sample model predictions of LMPs. While there is more variability across the load segments across regions and for each season, reflecting the fact that a peak load hour in 2018 need not be the same as in 2016, many of the factors correlated with LMPs in that load segment hour are captured by our model (e.g., load, fuel prices, effective capacity, virtual bidding, and effective amperage).

Figure A.11: Comparison of Out of Sample Model Prediction of Day-Ahead LMPs ($/MWh) by PJM Region to 2018 Historic Data



Finally, we compare our out-of-sample model prediction of total CO_2 emissions released by EGUs in PJM against observed 2018 data on CO_2 emissions from Monitoring Analytics' *2018 State of the Market Report.* Our estimate of 351.3 MMT CO_2 under-predicts the value reported by Monitoring Analytics by just -5.1%.

F Additional Policy Details

F.1 Clean Air Act

The federal Acid Rain Program (ARP) was established under Title IV of the 1990 Clean Air Act (CAA) Amendments to reduce SO_2 and NOx emissions that cause acid rain from the U.S. power sector. It includes a SO_2 program, which introduced a SO_2 cap-and-trade system, and a NOx program, which does not cap NOx emissions nor introduce a NOx trading program, but instead targets certain major NOx emitting coal-fired EGUs, limiting their total annual NOx emissions or their emission factors.

The ARP SO_2 program establishes a cap-and-trade system which clears annually, under which affected EGUs have to possess SO_2 allowances equal to their annual SO_2 emissions. Affected EGUs are free to choose how to reduce SO_2 emissions to meet targets, including buying, selling, and banking SO_2 allowances. The ARP SO_2 program's goal is to annually reduce SO_2 emissions by 10 million tons relative to the 1980 SO_2 emissions level. This translates to setting annual caps that vary in different phases of the program on affected EGUs. The ARP SO_2 program has two phases. Phase 1 was from 1995 to 1999, targeting 263 EGUs at 110 facilities in 21 eastern and midwestern states. Towards the end of phase 1 of the program, an additional 182 EGUs joined, bringing the total number of affected EGUs in phase 1 of the program to 445 affected EGUs. During phase 1, the EPA set an annual SO_2 emission cap at 5.7 million tons which equals the number of SO_2 allowances it issues. Phase 2 began in 2000 and is still ongoing, adding more EGUs into the program including smaller fossil fuel EGUs. By this time, the ARP SO_2 program targets all existing and new fossil fuel EGUs with capacity of 25 MW or more in all of the U.S., except Alaska, Idaho (which has no fossil fuel EGUs), and Hawai'i. During phase 2, the EPA set an annual SO_2 emissions cap of 8.95 million tons which equals the number of SO_2 allowances it issues. After the end of each year, all affected EGUs submit their SO_2 allowances that cover their SO_2 emissions for that year to the EPA, who then cancels all the submitted allowances. We do not consider the ARP NOx program in our model.

F.2 Cross-State Air Pollution Rule

The Cross-State Air Pollution Rule (CSAPR) was issued by the EPA in 2011, requiring 23 states in the eastern half of the United States to reduce annual SO_2 and NOx emissions to help downwind areas maintain or attain the annual $PM_{2.5}$ levels of the 2006 and/or 1997 national ambient air quality standards (NAAQS). CSAPR also requires 25 states to reduce ozone season NOx emissions to help downwind areas maintain or attain the annual ozone level of the 2008 and/or 1997 ozone NAAQS. Similar to the ARP, CSAPR establishes SO_2 and NOx cap-and-trade systems to reduce SO_2 and NOx emissions. So far, there are three cap-and-trade programs under CSAPR: the annual SO_2 program (which establishes separate trading regimes for SO_2 group 1 and SO_2 group 2 states), the annual NOx program, and the NOx ozone season program (which also establishes separate trading regimes for NOx group 1 and NOx group 2 states). The CSAPR NOx ozone season program was only recently finalized in 2016 and we do not consider it in our analysis. As such, we only consider the CSAPR annual SO_2 andNOx programs. To incorporate the ARP SO_2 program, the CSAPR annual SO_2 program, and the CSAPR annual NOx program into the model, marginal pre-existing regulatory costs (see equation (19)) includes the costs to covered EGUs of buying SO_2 and NOx emission allowances necessary for compliance with the ARP and CSAPR, where the annual allowance prices from these programs are assumed to vary exogeneously across years; see Section C.2.3.

F.3 Nuclear Subsidies

In Illinois and New Jersey, eligible existing nuclear EGUs qualify for nuclear excise subsidies. These excise subsidies are statutorily authorized by each state with the total amount of subsidy disbursement capped each year. In New Jersey this is achieved through the confusingly titled Zero Emission Certificate program, which is not to be confused with the trade of renewable energy certificates as permitted New Jersey's Renewable Portfolio Standard, from which nuclear units are excluded. Moreover, in Illinois this is achieved through the equally confusingly titled Zero Emission Standard, which is not to be confused with Illinois' Renewable Portfolio Standard, from which nuclear units are also excluded.

F.3.1 New Jersey

In New Jersey, the nuclear subsidy program started in 2019. The nuclear subsidy program is intended to subsidize eligible nuclear EGUs which are at risk of early retirement in order to maintain the environmental benefits that the EGUs provide. Eligible nuclear EGUs in New Jersey can collect one zero emission certificate (ZEC) for each MWh of energy produced. Unlike REC and RGGI allowance prices which are determined *ex post* through virtual market clearing, ZEC prices are established *ex ante* by the New Jersey Board of Public Utilities (NJBPU). The ZEC 'price' amounts to an excise subsidy per MWh of electric power produced by eligible existing nuclear EGUs that is scaled up or down each year by the regulator such that total subsidy payments does not exceed $300 million per year.

To calculate the ZEC price in New Jersey, we follow the ZEC price calculation method provided in Chapter 16 of New Jersey's *ZEC Act Legislation* (accessed on 02/12/2020 at //www.njleg.state.nj.us/2018/Bills/AL18/16_.HTM). To calculate the ZEC price in New Jersey, we divide 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 MWhs of electricity distributed by the electric public utilities in New Jersey in the prior energy year, or the number of MWhs of electricity generated in the prior energy year by the nuclear power plants authorized to receive the subsidy. New Jersey has approved three existing eligible nuclear EGUs to receive subsidies thus far (Hope Creek and Salem 1 and 2), and whose total generation in 2018 is 28,441,726 MWh, which is less than 40% of New Jersey's total generation in 2018, 29,963,300 MWh ($= 0.4 \times 74,908,250$), where 74,908,250 MWh is the total New Jersey generation in 2018 according to the EIA's *New Jersey 2018 State Electricity Profile*.

Thus, the formula for calculating the excise subsidy in 2019 is given by:

$$p_{NJ}^{ZEC} = \frac{\min\{p^{RP} g^{RS}, 300000000\}}{29963300} \tag{33}$$

where p^{RP} is the electricity price charged to ratepayers in \$/MWh to finance the nuclear subsidy program, and g^{RS} is the total electricity sold to consumers in MWh in the previous year (2018). p^{RP} has been set by the NJBPU to \$4/MWh. g^{RS} equals 76,016,762 MWh in 2018 according to the EIA's *New Jersey 2019 State Electricity Profile*. Since $p^{RP} g^{RS} = \$304,067,048 > \$300,000,000$, the excise subsidy in 2019 is $p_{NJ}^{ZEC} = \frac{\$300,000,000}{29,963,300} = \$10.012$ per MWh. Starting in 2019, this is also deducted from rc_k^E for the three approved nuclear units in New Jersey and thus accounted for in the producer surplus estimated for those EGUs. In addition, the total value of subsidy payments are deducted from annual consumer surplus calculated for East RPJM. We assume that this subsidy remains in place through 2030.

F.3.2 Illinois

In Illinois, all existing nuclear EGUs are eligible to receive a nuclear subsidy or ZEC price in 2019 of \$16.50/MWh, unless \$16.50 minus the difference in the annual average and baseline market price indices is less than zero, in which case the subsidy is zero. The baseline market price index is \$31.40 as defined in subparagraph (B) of paragraph (1) of subsection (d-5) of the *Illinois Power Agency Act*, Section 1-75(d-5) (also known as the Zero Emission Standard). Using data on locational marginal prices predicted by our model, we calculate annual average market price index in 2019 of \$35.35. Since $16.50 - (35.35 - 31.40) = 12.55 > 0$, the nuclear excise subsidy to Illinois units equals \$16.50/MWh. Starting in 2019, this is deducted from rc_k^E for all nuclear units in Illinois and PJM and thus accounted for in the producer surplus estimated for those EGUs. In addition, the total value of subsidy payments are deducted from annual consumer surplus calculated for West RPJM. We assume that this subsidy remains in place through 2030.

F.4 Regional Greenhouse Gas Initiative

The Regional Greenhouse Gas Initiative (or RGGI) is a regional cap-and-trade program on CO_2 emissions from affected (with capacities of 25 MW or more) fossil-fuel EGUs in the northeastern United States in which states voluntarily elect to have their affected fossil-fuel EGUs participate. Currently ten states participate in RGGI (e.g., $\mathcal{S}_{RGGI}$): Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, and Vermont. New Jersey left the program in 2011 but re-entered the program in January 2020. Virginia recently passed the *Clean Economy Act* and is scheduled to join RGGI in 2021. Governor Wolf issued Executive Order 2019-07 in October 2019 (last amended June 22nd, 2019) authorizing Pennsylvania to join RGGI by 2022. Although we do not explicitly track the capacity of individual new NGCC units, we assume that all new NGCC capacity are subject to RGGI.

The RGGI program sets annual caps on the total amount of CO_2 emissions that can be released from affected fossil-fuel EGUs in RGGI participating states. These caps decline over time. Affected EGUs are required to hold allowances equal to their CO_2 emissions over a three-year control period. The control periods are for every three years. The first four control periods are: 2009-2011, 2012-2014, 2015-2017, and 2018-2020. Information on RGGI caps and adjusted allowance budgets assigned to states (e.g., $\bar{E}_s^{RGGI}$ for all s in RGGI over time) are taken from RGGI and state authorizing legislation. Through 2020, adjusted allowance budgets reflect the adjustments made to RGGI allowance budgets as a result of the first and second control period interim adjustments for banked allowances made by RGGI. Using historical data on banked permits discussed in Section C, we also assume that there will be a future third control period interim adjustment for banked allowances impacting caps from 2021 to 2025 that reflects the formulas reported in recent authorizing legislation in Virginia and New Jersey and which is similar to the second control period adjustment. We currently do not consider interactions from Pennsylvania's entry into RGGI with Pennsylvania's Act 129 program, nor explicitly represent the waste-coal provisions under consideration by the Pennsylvania Department of Environmental Protection.

Every quarter, participating states issue a number of allowances equal to their cap, which declines over the years as the cap declines. After every three years, each EGU must report its total carbon emissions and submit an equal number of emission allowances that cover these emissions. Starting in the third control period (2015), each EGU must hold a number of allowances equal to 50% of its CO_2 emissions during the first two calendar years of each three-year control period. Each EGU then must hold the number of allowances equal to 100% of its remaining emissions for the three-year control period at the end of the three year control period.

Aside from the annual available allowance budget, RGGI also has a number of allowances under the Cost Containment Reserve (CCR) that it can auction whenever the clearing price at auction exceeds the CCR trigger price. The CCR budget replenishes annually, starting at 5 million allowances in 2014 (the year the CCR started) and 10 million allowances each year after. In 2020 and onward, CCR budget will be 10% of the regional cap each year. Once released, CCR allowances are released until the final clearing price equals the CCR trigger price or the CCR is exhausted, in which case the clearing price can be higher than the CCR trigger price. The CCR trigger price is $\$4$ in 2014, $\$6$ in 2015, $\$8$ in 2016, $\$10$ in 2017 and increases 2.5% each year after. Then the CCR trigger price will be $\$13.00$ in 2021 and will increase by 7% per year thereafter. Through 2030, we find in our model simulations that the soft price ceiling established by CCR trigger prices is never binding, with or without Pennsylvania's entry into RGGI.

RGGI also includes a hard price floor or minimum reserve price. In 2017, RGGI had a minimum reserve price of \$2.15 per allowance. Each calendar year thereafter, the minimum reserve price is 1.025 multiplied by the minimum reserve price from the previous calendar year, rounded to the nearest whole cent. In addition, beginning in 2021, RGGI states can also access an Emissions Containment Reserve (ECR), which provides a secondary soft price floor. States who implement ECR will withhold allowances equal to 10% of the allowances in their annual base budget from circulation to secure additional emission reductions if allowance prices fall below ECR trigger prices. The ECR trigger price will be $\$6.00$ in 2021, and rise at 7% per year thereafter. Through 2030, we find in our model simulations that hard and soft price floors established by the minimum reserve price and ECR trigger prices are never binding, with or without Pennsylvania's entry into RGGI.

F.5 Alternative Energy/Renewable Portfolio Standards

Many states in PJM alternative energy/renewable portfolio standards (generally, RPSs) which specify that a minimum fraction (e.g., $\bar{R}_{st}$) of that state's annual total generation that must come from various subsets of generation sources, typically renewable generators. In 2018, there are nine states in PJM that have RPSs: Delaware, Illinois, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, and the District of Columbia. There are two states in PJM that have voluntary RPS (Virginia and Indiana) and three states in PJM that do not have an RPS (Kentucky, Tennessee, and West Virginia). We only model RPS in the nine states that have mandatory RPSs, of which four states (District of Columbia, Maryland, New Jersey and Pennsylvania) 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 five states (Delaware, Illinois, Michigan, North Carolina and Ohio) 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 are modeled as tier 1 resources. By 2030, District of Columbia's tier 2 RPS is projected to decline to 0 % and thus in 2030 only the tier 1 RPS is modeled for the District of Columbia. Most states with required and voluntary RPS standards also have an additional, separate tier that targets generation from solar power (Solar RPS or SRPS). RPS shares (e.g., $\bar{R}_{st}$, here reported in percentage terms) are provided for the years 2017 and 2030 in Tables A.15 and A.16.

Eligible units are those EGUs in state s whose generation is targeted for expansion by the state-tier RPS constraint st (see, equation (12)), as summarized in Table A.17. Ineligible units are those EGUs in state s whose generation is not targeted for expansion by the state-tier RPS constraint st, but which must be considered for purposes of evaluating the state's RPS constraint. The precise subset of eligible new and existing EGUs under state-tier RPS constraint st, $\mathcal{J}_{st}$, differ across state RPS's and tier t according to each state's RPS authorizing statute. Likewise, the subset of eligible and ineligible new and existing EGUs for each state s, $\mathcal{J}_{st}$, may differ across state RPSs, but generally reflects the total annual generation dispatched in that state. Generation from eligible units generates renewable energy certificates (RECS, or for the SRPS, SRECS) which are used to evaluate compliance with the state RPS in a given year and can be freely traded among eligible and ineligible EGUs within each state. In addition, for some state-tier RPS constraints st, we also allow external RECs to be purchased by eligible EGUs outside of PJM, as discussed above. Finally, we assume that generation from new wind and solar capacity expansions in each state are eligible EGUs for states with RPSs in PJM.

Table A.15: RPS Targets by State and Tier 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.16: RPS Targets by State and Tier 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.17: Characterization of Eligible Generation Units by State RPS 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) EGUs 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 hydro-power 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.17 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	REC 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.17 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	REC 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.

Continued on next page

Table A.17 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	REC Eligible Location
NC	(1) Solar-electric, (2) solar thermal, (3) wind, (4) hydropower ($\leq$ 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.
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

Continued on next page

Table A.17 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	REC Eligible Location
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.
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		

Continued on next page

Table A.17 – *Continued from previous page*

State	Tier 1 RPS	Tier 2 RPS	REC Eligible Location
VA*	(1) Solar, (2) wind power, (3) geothermal energy, (4) hydropower, (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

References

Amemiya, T. (1977). The Maximum Likelihood and the Nonlinear Three-Stage Least-Squares Estimator in the General Nonlinear Simultaneous Equation Model. *Econometrica 45*, 955–968.

Bushnell, J. B., S. P. Holland, J. E. Hughes, and C. R. Knittel (2017). Strategic Policy Choice in State-Level Regulation: the EPA's Clean Power Plan. *American Economic Journal. Economic Policy 9*, 57–90.

Ferreyra, M. (2007). Estimating the Effects of Private School Vouchers in Multi-District Economies. *Econometrica 97*(3), 789–817.

IPM (2013). Documentation for EPA Base Case v.5.13 Using the Integrated Planning Model. Technical report, United States Environmental Protection Agency.

Ito, K. (2014). Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. *American Economic Review 104*(2), 537–563.

SETO (2017). The SunShot Goals. Technical report, Solar Energy Technologies Office, U.S. Department of Energy.

Wiser, R., K. Jenni, J. Seel, E. Baker, M. Hand, E. Lantz, and A. Smith (2016). Expert Elicitation Survey on Future Wind Energy Costs. *Nature Energy 16135.*

G Additional Figures and Tables

Table A.18: Impact of Pennsylvania Joining RGGI on Capacity Expansion

	Base Case	Central Case	No RPS
Pennsylvania:			
Natural Gas - 2022 (GW)	0.00	0.00	0.00
Natural Gas - 2023 to 2025	0.00	0.00	0.00
Natural Gas - 2026	0.00	0.00	0.00
Natural Gas - 2027 to 2029	0.00	0.00	0.00
Natural Gas - 2030	0.00	0.00	0.00
Wind - 2022	0.92	0.10	0.00
Wind - 2023 to 2025	0.47	0.16	0.00
Wind - 2026	0.00	0.00	0.00
Wind - 2027 to 2029	0.00	0.00	0.00
Wind - 2030	0.00	0.05	0.00
Solar - 2022	0.00	0.00	0.00
Solar - 2023 to 2025	0.00	0.00	0.00
Solar - 2026	0.00	0.00	0.00
Solar - 2027 to 2029	0.00	0.00	0.00
Solar - 2030	0.01	0.00	0.00
Rest of PJM:			
Natural Gas - 2022 (GW)	0.00	0.00	0.00
Natural Gas - 2023 to 2025	0.00	0.00	0.00
Natural Gas - 2026	0.00	0.00	0.00
Natural Gas - 2027 to 2029	0.00	0.00	0.00
Natural Gas - 2030	0.00	0.00	0.00
Wind - 2022	0.32	0.72	0.18
Wind - 2023 to 2025	4.85	7.43	7.77
Wind - 2026	2.00	2.19	2.10
Wind - 2027 to 2029	2.18	2.26	2.45
Wind - 2030	0.94	0.50	0.66
Solar - 2022	0.15	0.19	0.18
Solar - 2023 to 2025	2.55	0.68	0.67
Solar - 2026	0.21	0.23	0.23
Solar - 2027 to 2029	0.18	0.14	0.14
Solar - 2030	0.08	0.03	0.05

Table A.19: Impact of Pennsylvania Joining RGGI on Transmission

	2022	2026	2030	Cumulative
Flow into East PA (2) from West PA (1) (1,000 GWh)	22.7	22.3	23.0	202.3
Change	2.2	3.8	5.2	35.7
Flow into Central RPJM (4) from West RPJM (5)	19.2	21.4	25.6	196.7
Change	-8.7	-6.1	-5.8	-61.9
Total Cross-Border	-69.9	-71.7	-73.3	-650.9
Change	38.4	41.6	50.6	398.1
Flow into East PA (2) from East RPJM (3)	-50.5	-50.8	-50.9	-456.9
Change	1.3	0.2	0.5	3.6
Flow into East PA (2) from Central RPJM (4)	-12.3	-12.5	-13.2	-114.8
Change	8.7	8.5	9.7	83.2
Flow into West PA (1) from West RPJM (5)	-7.2	-8.5	-9.1	-79.3
Change	28.3	33.0	40.5	311.3
Congestion Price between East PA (2) and West PA (1) - Base Case ($/MWh)	$ -0.47	$ -0.06	$ 0.02	$ -0.18
Congestion Price between East PA (2) and West PA (1) - Central Case ($/MWh)	$ -0.31	$ 0.40	$ 0.92	$ 0.31
Change	$ 0.16	$ 0.45	$ 0.90	$ 0.48
Congestion Price between Central RPJM (4) and West RPJM (5) - Base Case	$ 0.59	$ 1.34	$ 2.27	$ 1.41
Congestion Price between Central RPJM (4) and West RPJM (5) - Central Case	$ 0.04	$ 0.72	$ 1.39	$ 0.69
Change	$ -0.55	$ -0.62	$ -0.88	$ -0.72
Congestion Price Cross Border - Base Case	$ -2.76	$ -3.55	$ -4.50	$ -3.72
Congestion Price Cross Border - Central Case	$ -1.27	$ -2.24	$ -2.04	$ -2.02
Change	$ 1.50	$ 1.31	$ 2.46	$ 1.70
Congestion Price between East PA (2) and East RPJM (3) - Base Case	$ -5.30	$ -8.11	$ -9.92	$ -8.18
Congestion Price between East PA (2) and East RPJM (3) - Central Case	$ -2.61	$ -6.20	$ -5.94	$ -5.49
Change	$ 2.69	$ 1.91	$ 3.98	$ 2.69
Congestion Price between East PA (2) and Central RPJM (4) - Base Case	$ -2.02	$ -1.97	$ -2.92	$ -2.28
Congestion Price between East PA (2) and Central RPJM (4) - Central Case	$ -0.77	$ -0.42	$ -0.32	$ -0.48
Change	$ 1.25	$ 1.55	$ 2.60	$ 1.80
Congestion Price between West PA (1) and West RPJM (5) - Base Case	$ -0.96	$ -0.57	$ -0.67	$ -0.70
Congestion Price between West PA (1) and West RPJM (5) - Central Case	$ -0.42	$ -0.09	$ 0.15	$ -0.10
Change	$ 0.54	$ 0.48	$ 0.81	$ 0.60