The Effects of Time-Variant Electricity Pricing on Commercial and Industrial Customers*

Reigner Kane

November 29, 2023

Abstract

As electricity systems decarbonize and shift towards intermittent forms of power generation such as solar and wind, managing system load relative to generation becomes critical for policymakers and utilities alike. With large-scale storage still in development, one tool already in use is pricing which varies throughout the course of a day. However, the empirical evidence remains limited around key questions such as the responsiveness of different classes of customers to such policies. This paper proposes and estimates a model of region-sector-level electricity demand, highlighting heterogeneity across regions and across sectors with respect to on- versus off-peak demand. We recover sector-specific price elasticities for electricity using utility-level data and find that industrial customers are far more heterogeneous across space than commercial firms in responsiveness time-variant pricing.

*Thank you to my MAPSS preceptor, Dr. Alessandra Gonzalez, for extensive guidance and encouragement through the initial formulation of this project; to my thesis advisor Dr. Christopher Dobronyi for input and discussion regarding the structural estimation; to Dr. Joseph Hardwick for input into the reduced-form estimation; and to the many other peer and faculty participants in the Applied Microeconomics and Econometrics & Theory MA seminars at the University of Chicago for their many suggestions. This current draft has been substantially updated since initial submission as an MA thesis, most notably with improved estimation in Section 4.4. reignerkane@uchicago.edu

1 Introduction

Governments across the world are now beginning to implement an array of policies aimed at reducing greenhouse gas emissions. Simultaneous broad electrification and decarbonization of electricity generation promise substantial gains in the fight against climate change Steinberg et al. (2017). As this project progresses, there remain numerous practical challenges facing the policymakers responsible for working towards the goal of low-carbon electricity generation and distribution. In particular, there is a divide between the marginal cost of supplying electricity and the price which customers are asked to pay. Importantly, this discrepancy has a temporal component, due both to variations in demand in concert with nonlinear costs, and to the variation in costs of supply throughout the day. In the simplest case, the marginal cost of solar generation is approximately zero during the sunny hours of midday, but effectively rises to infinity after sunset.

Such drastic differences do not accurately reflect the problem faced by electric utilities, as the current fossil-fuel dominated energy mix generally insulates against such fluctuations. Nonetheless, the effect of "peak" demand (the period of time corresponding to the greatest demand for electricity, and hence when the costliest methods of supply must be employed) has already compelled many utilities to implement time-variant pricing policies. Such policies have historically aimed to reduce consumption during peak hours, without regard to variations in energy mix throughout the day Lazar (2014). Now, time-variant pricing may play an important role in enabling the transition, enabling utilities to reshape intra-day demand fluctuations towards a more efficient outcome.

In this paper, we analyze the impact of differing severity of time-variant electricity pricing on commercial and industrial customers across the United States. In particular, we attempt to recover the effect of changes in pricing (e.g. changes in on- or off-peak prices) on output and electricity consumption. Additionally, by making some assumptions on the relevant production functions, we provide sector-specific estimates of the production elasticity of output, key parameters for informed policy decisions. Our estimation procedure also enables us to compare differences in the distribution of firms' electricity consumption patterns across regions.

Prior research on time-variant electricity pricing has largely focused on its applications to residential customers (e.g. Borenstein and Bushnell (2018), Borenstein (2012), Ito (2014)). Results from this literature indicate that households are not especially responsive to intra-day changes in the price of electricity. The existing literature on commercial and industrial customers finds a broadly similar story holds for firms; for example, Jessoe and Rapson (2015) find that time-of-use pricing "does not induce an economically significant change in usage or peak demand". Blonz (2016) finds that, in California, the only meaningful reduction in peak consumption in response to time-variant pricing was among firms in a particular geographic region, lending credence to the conclusion that climatological factors (particularly the extremes of temperature) are a substantial factor. These findings may be cause for concern among policymakers and utilities, who often adopt time-variant policies to encourage load-shifting away from periods of peak demand.

2 Model

2.1 The Firm's Problem

We now develop a simple model of production, with an emphasis on the choices firms must make concerning their electricity consumption. We do so specifically with an eye toward estimation on the aggregate data described in Section 3. Time-variant pricing is generally an opt-in program, allowing firms to select between a fixed-rate and time-variant price. Because of this, in general, firms must first choose whether to participate in a time-variant program. Then, they decide how much electricity to consume, and when to consume it. We consider the following Cobb-Douglas type production function for a representative firm and simplified case, where E_1 is the quantity of electricity consumed during off-peak hours, E_2 is the quantity consumed during on-peak hours, and L the quantity of labor employed: $F(E_1, E_2, L) = AE_1^{\alpha_1}E_2^{\alpha_2}L^{\alpha_L}$.

We take the output to be the numeraire, and denote the wage by w, time-variant price in each period as p_1 and p_2, respectively, and the fixed price as p_0. We assume that $p_1 \leq p_0 \leq p_2$, so that off-peak electricity is never more expensive than the fixed baseline, and on-peak is never less expensive. Additionally, we impose decreasing returns to scale, by restricting $\tilde{\alpha} := \alpha_1 + \alpha_2 + \alpha_L < 1$. Now, the firm is profit-maximizing, and considers its profits under both alternatives; if it selects into the time-variant pricing, it solves the problem: $\max\,[\pi_v = AE_1^{\alpha_1}E_2^{\alpha_2}L^{\alpha_L} - p_1E_1 - p_2E_2 - wL]$. Alternatively, were it to reject time-variant pricing, it instead solves: $\max\,[\pi_f = AE_1^{\alpha_1}E_2^{\alpha_2}L^{\alpha_L} - p_0(E_1 + E_2) - wL]$. We solve each problem to compare profits; denote the optimal level of production by Y^*. Under time-variant pricing, we have the first order conditions $Y^*/E_1^* = p_1/\alpha_1, Y^*/E_2^* = p_2/\alpha_2$, and $Y^*/L^* = w/\alpha_L$. Rearranging to express the other variables in terms of E_1^*, we have $E_2^* = \left(\frac{p_1\alpha_2}{p_2\alpha_1}\right)E_1^*$ and $L^* = \left(\frac{p_1\alpha_L}{w\alpha_1}\right)E_1^*$. Using the earlier first order condition with respect to E_1^*, we now see that

$$Y^* = A(E_1^*)^{\alpha_1}(E_2^*)^{\alpha_2}(L^*)^{\alpha_L}$$

$$\implies E_1^*\left(\frac{p_1}{\alpha_1}\right) = A(E_1^*)^{\tilde{\alpha}}\left(\frac{p_1}{p_2}\right)^{\alpha_2}\left(\frac{\alpha_2}{\alpha_1}\right)^{\alpha_2}\left(\frac{p_1}{w}\right)^{\alpha_L}\left(\frac{\alpha_L}{\alpha_1}\right)^{\alpha_L}$$

$$\implies (E_1^*)^{1-\alpha_1} = A\left(\frac{p_1}{p_2}\right)^{\alpha_2}\frac{\alpha_2^{\alpha_2}p_1^{\alpha_L-1}\alpha_L^{\alpha_L}}{\alpha_1^{\alpha_2+\alpha_L-1}w^{\alpha_L}}$$

$$\implies E_1^* = \left[A\left(\frac{p_1}{p_2}\right)^{\alpha_2}\frac{\alpha_2^{\alpha_2}p_1^{\alpha_L-1}\alpha_L^{\alpha_L}}{\alpha_1^{\alpha_2+\alpha_L-1}w^{\alpha_L}}\right]^{\frac{1}{1-\tilde{\alpha}}}.$$

This gives an explicit solution for the optimal consumption of electricity during off-peak hours, by the convexity of decreasing-returns-to-scale Cobb-Douglas. Therefore, we can now characterize the optimal level of production under time-variant pricing as follows:

$$Y_v^* = \left[A\left(\frac{p_1}{p_2}\right)^{\alpha_2}\frac{\alpha_2^{\alpha_2}p_1^{\alpha_L-1}\alpha_L^{\alpha_L}}{\alpha_1^{\alpha_2+\alpha_L-1}w^{\alpha_L}}\right]^{\frac{1}{1-\tilde{\alpha}}}\left(\frac{p_1}{\alpha_1}\right).$$

Now, we repeat a similar procedure for fixed-rate pricing, and arrive at optimal production

$$Y_f^* = \left[A\frac{\alpha_2^{\alpha_2}p_0^{\alpha_L-1}\alpha_L^{\alpha_L}}{\alpha_1^{\alpha_2+\alpha_L-1}w^{\alpha_L}}\right]^{\frac{1}{1-\tilde{\alpha}}}\left(\frac{p_0}{\alpha_1}\right).$$

Of course, firms will compare their prospective profits, not production, when deciding between pricing policies. However, we notice that the maximized profits from each take the form $\pi_v^* = Y_v^*(1-\tilde{\alpha})$ and $\pi_f^* = Y_f^*(1-\tilde{\alpha})$, and so it is sufficient to compare production; in particular, the firm will choose the time-variant pricing if and only if $Y_v^* \geq Y_f^*$. However, since each are positive values, this is equivalent to considering

$$\frac{Y_v^*}{Y_f^*} \geq 1 \iff \left(\frac{Y_v^*}{Y_f^*}\right)^{1-\tilde{\alpha}} \geq 1$$

$$\iff \left(\frac{p_1}{p_0}\right)^{1-\tilde{\alpha}}\left(\frac{p_1}{p_2}\right)^{\alpha_2} \geq 1$$

$$\iff \left(\frac{p_1}{p_2}\right)^{\alpha_2} \geq \left(\frac{p_0}{p_1}\right)^{1-\tilde{\alpha}}.$$

As intuition would suggest, then, the decision is made by comparing the relative losses from moving to time-variant pricing (that is, the higher price during on-peak hours, $p_2 \geq p_0$) to the gains (the lower prices during off-peak hours, $p_1 \leq p_0$).

Finally, we note that the total expenditure on electricity under each regime, E_v^* and E_f^*, are given by

$$E_v^* = \left(\frac{p_1\alpha_2 + p_2\alpha_1}{p_2\alpha_1}\right)\left[A\left(\frac{p_1}{p_2}\right)^{\alpha_2}\frac{\alpha_2^{\alpha_2}p_1^{\alpha_L-1}\alpha_L^{\alpha_L}}{\alpha_1^{\alpha_2+\alpha_L-1}w^{\alpha_L}}\right]^{\frac{1}{1-\tilde{\alpha}}}$$

and

$$E_f^* = \left(\frac{\alpha_2 + \alpha_1}{\alpha_1}\right)\left[A\frac{\alpha_2^{\alpha_2}p_0^{\alpha_L-1}\alpha_L^{\alpha_L}}{\alpha_1^{\alpha_2+\alpha_L-1}w^{\alpha_L}}\right]^{\frac{1}{1-\tilde{\alpha}}}.$$

The differences between these optimal consumption levels is what will enable our empirical characterization of the values (later, distribution) of α_1 and α_2.

2.2 Within-Sector Heterogeneity

Thus far, we have exclusively considered a representative firm, thereby eliding the heterogeneity that exists within and between sectors. In service of examining this heterogeneity, we will introduce some additional notation. We divide space into I regions and firms into industrial and commercial sectors. Then, for each region-sector, we allow for a distribution of values for α_1 and α_2. This represents production processes which place differing emphasis on on-peak or off-peak consumption. However, we assume that all types are homogeneous in the values of $\tilde{\alpha}$ (overall returns to scale) and α_L. This allows us to rewrite, for sector ς and a particular firm, $\alpha_{2,\varsigma} = \tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha_{1,\varsigma}$. That is, $\alpha_{2,\varsigma}$ is uniquely determined by $\alpha_{1,\varsigma}$ and sector-wide parameters, and so we can characterize their distribution by a univariate density f_α which is supported on (at most) $[0, \tilde{\alpha} - \alpha_L]$. Now, we can write the total consumption of electricity across the entire sector ς, observed after each firm has optimized, as

$$\frac{E_\varsigma^{\text{obs}}}{N_\varsigma} = \int_0^{\tilde{\alpha}-\alpha_L} f_\alpha(\alpha_1)E_\varsigma^*(\alpha_1)\,d\alpha_1,$$

where $E_\varsigma^*(\alpha_1)$ is the optimal consumption for firms with off-peak production elasticity α_1; the values for each E_ς^* are determined via the process described in Section 2.1. We are here concerned with the average per-firm expenditure.

Because our ultimate goal is to bring this model to panel data that contains observations at the county-year level, we introduce some additional assumptions on the production function. In particular, we will rewrite the per-firm total-factor productivity in region i and year t as $A_{\varsigma it} = \gamma_{\varsigma t}\delta_i\epsilon_{\varsigma it}$, where we interpret $\gamma_{\varsigma t}$ as capturing time variation in productivity, δ_i as the exogenous location-specific productivity, and $\epsilon_{\varsigma it}$ as a region-year idiosyncratic shock. We assume that $\epsilon_{\varsigma it}$ is identically and independent drawn from a distribution with corresponding distribution function G, support $\mathbb{R}_+$, and with $\mathbb{E}[\ln(\epsilon_{it})] = 0$. (The standard log-normal distribution, used in the later empirical analysis, conforms to this assumption.)

This allows us to, conditional on prices p_0, p_1, p_2, and w and fundamentals, determine the distribution of $\frac{E^{\text{obs}}_{\varsigma it}}{N_{\varsigma it}}$; namely, for any e, we have

$$
\begin{aligned}
&\mathbb{P}\left(\frac{E^{\text{obs}}_{\varsigma it}}{N_{\varsigma it}} \le e\right) = \mathbb{P}\left(\int_0^{\tilde{\alpha}-\alpha_L} f_\alpha(\alpha_1) E^*_\varsigma(\alpha_1)\, d\alpha_1 \le e\right) \\
&= \mathbb{P}\left[A_{\varsigma it}^{\frac{1}{1-\tilde{\alpha}_\varsigma}} \int_0^{\tilde{\alpha}-\alpha_L} \lambda_{0\varsigma it}(\alpha_1)\, \lambda_{v\varsigma it}(\alpha_1)^{\mathbb{1}(\alpha_1 \notin S(\mathbf{p}_{\varsigma it}))} \lambda_{f\varsigma it}(\alpha_1)^{\mathbb{1}(\alpha_1 \notin S(\mathbf{p}_{\varsigma it}))} f_\alpha(\alpha_1)\, d\alpha_1 \le e\right] \\
&= \mathbb{P}\left[\epsilon_{\varsigma it} \le \left(\frac{e^{1-\tilde{\alpha}_\varsigma}}{\gamma_{\varsigma t}\delta_i}\right) \left(\int_0^{\tilde{\alpha}_\varsigma-\alpha_{L\varsigma}} \lambda_{0\varsigma it}(\alpha_1)\, \lambda_{v\varsigma it}(\alpha_1)^{\mathbb{1}(\alpha_1 \notin S(\mathbf{p}_{\varsigma it}))} \lambda_{f\varsigma it}(\alpha_1)^{\mathbb{1}(\alpha_1 \notin S(\mathbf{p}_{\varsigma it}))} f_\alpha(\alpha_1)\, d\alpha_1\right)^{\tilde{\alpha}_\varsigma - 1}\right] \\
&= G\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right) x^{1-\tilde{\alpha}_\varsigma} \left(\int_0^{\tilde{\alpha}_\varsigma-\alpha_{L\varsigma}} \lambda_{0\varsigma it}(\alpha_1)\, \lambda_{v\varsigma it}(\alpha_1)^{\mathbb{1}(\alpha_1 \in S(\mathbf{p}_{\varsigma it}))} \lambda_{f\varsigma it}(\alpha_1)^{\mathbb{1}(\alpha_1 \notin S(\mathbf{p}_{\varsigma it}))} f_\alpha(\alpha_1)\, d\alpha_1\right)^{\tilde{\alpha}_\varsigma - 1}\right) \\
&= G\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right) \left(\frac{e}{\Lambda_{\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right),
\end{aligned}
$$

where

$$
\lambda_{0\varsigma it}(\alpha_1) = \left(\frac{(\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha_{1\varsigma})^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha_{1\varsigma}} \alpha_{L\varsigma}^{\alpha_{L\varsigma}}}{\alpha_{1\varsigma}^{\tilde{\alpha}_\varsigma - \alpha_{1\varsigma} - 1} w^{\alpha_{L\varsigma}}}\right)^{\frac{1}{1-\tilde{\alpha}_\varsigma}},
$$

$$
\lambda_{v\varsigma it}(\alpha_1) = \left(\frac{p_{1\varsigma it}(\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha_{1\varsigma}) + p_{2\varsigma it}\alpha_{1\varsigma}}{p_{2\varsigma it}\alpha_{1\varsigma}}\right) \left(\frac{p_{1\varsigma it}^{\tilde{\alpha}_\varsigma - \alpha_{1\varsigma} - 1}}{p_{2\varsigma it}^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha_{1\varsigma}}}\right)^{\frac{1}{1-\tilde{\alpha}_\varsigma}},
$$

$$
\lambda_{f\varsigma it}(\alpha_1) = \left(\frac{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha_{1\varsigma}}{\alpha_{1\varsigma}}\right) p_{0\varsigma it}^{\frac{\alpha_{L\varsigma}-1}{1-\tilde{\alpha}_\varsigma}},
$$

$$
\Lambda_{\varsigma it}(f_\alpha) = \int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} \lambda_{0\varsigma kit}\, \lambda_{v\varsigma kit}^{\mathbb{1}(\alpha_{1\varsigma k} \in S(\mathbf{p}_{\varsigma it}))} \lambda_{f\varsigma kit}^{\mathbb{1}(\alpha_{1\varsigma k} \notin S(\mathbf{p}_{\varsigma it}))} f_\alpha(\alpha_1)\, d\alpha_1,
$$

and

$$
S(\mathbf{p}_{\varsigma it}) = \left\{\alpha \,:\, \left(\frac{p_{1\varsigma it}}{p_{2\varsigma it}}\right)^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \alpha} \le \left(\frac{p_{0\varsigma it}}{p_{1\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right\}.
$$

Now, provided that G is absolutely continuous, we can derive the conditional density h associated with $\frac{E^{\text{obs}}_{\varsigma it}}{N_{\varsigma it}}$ given the relevant prices and parameters. In particular, we have

$$
h_\varsigma(e; \mathbf{p}_{\varsigma it}, \gamma_{\varsigma t}, \delta_{\varsigma i}) = \frac{d}{de} G\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right)\left(\frac{e}{\Lambda_{\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right) = \left(\frac{1-\tilde{\alpha}_\varsigma}{\gamma_{\varsigma t}\delta_i \Lambda_{\varsigma it}^{1-\tilde{\alpha}_\varsigma} e^{\tilde{\alpha}_\varsigma}}\right) g\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right)\left(\frac{e}{\Lambda_{\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right).
$$

This analysis lays the necessary groundwork for maximum likelihood estimation in Section 4.3.

3 Data

The core data set is compiled from publicly-available government data sets. Data on utilities' electricity sales and revenue is supplied by the results of the Energy Information Administration's form 861; the observations are at the utility-year-sector level, divided into "industrial" and "commercial" sectors U.S. EIA (2022). Data on the number of firms, employment, wages, on a county-sectoral basis is from the Census Bureau's County Business Patterns (CBP) database U.S. Census Bureau (2022). County-level GDP is provided by the Bureau of Economic Analysis's `CAGDP9` series U.S. BEA (2022). Data on utility-level pricing policies is compiled by the National Renewable Energy Laboratory's OpenEI U.S. Utility Rate Database OpenEI and NREL (2022); this last data set is incomplete, in particular having missing values for many utilities' industrial pricing.

Utilities' service regions do not necessarily coincide with county boundaries; to account for this fact, the utility-level data must coerced into county-level data. For non-price data, this is accomplished by splitting values across the counties in a given utility's service region proportionately to the total county employment. Then, county-level aggregate electricity sales, revenue, and average price are calculated across all relevant utilities by combining these utility-county components.

Another set of simplifications is made with respect to the rate database. All information about rates in a year-county-sector is reduced to two values: p_{high} and p_{low}, the maximum and minimum possible prices charged throughout the year. Additionally, there is substantial missing data, and so county-sector-year observations without rate information must be excluded from the analysis. Informally, utilities with missing data do not appear to differ systematically from counties with rate data, either in the size nor sectoral composition of the counties they service. For more details, see Appendix A.

4 Empirical Strategy and Results

4.1 Reduced Form Strategy

The empirical strategy for identifying the effect of a pricing policy changes on the outcome variables of interest will rest on solving the inherent problem of endogeneity arising from simultaneity, as we observe prices and quantities in equilibrium. Even if we were to relax the assumption of (local) monopoly profit-maximization (as was assumed in Section 2.1), we nonetheless must contend with the effect of changes in quantity demanded on prices, provided utilities do not entirely ignore market behavior. Given the quoted motivation for various policy changes, including changes to prices, it seems extremely likely that endogeneity would plague any naive identification strategy.

To make the above ideas more rigorous, fix a sector $i \in \{\text{res}, \text{ind}, \text{com}\}$ and assume for the moment that $T = 1$; that is, firms and utilities do not consider time-dependent factors in their decision-making process. Fixing quantities $\mathbf{q}^{-i}$ for all other sectors, we can assume simple functional forms for each side of the market to get:

$$\begin{cases} q_D^i & = {Q^i}^{-1}(p^i) + u^i \\ q_S^i & = f_i(p^i - C(q_S^i + \mathbf{q}_S^{-i})) + v^i \end{cases}$$

Notice that a solution to the above system exists and yields an equilibrium price and quantity for sector i, assuming the relevant functions are all sufficiently well-behaved (e.g. continuous, monotonic, etc.). Moreover, note that the cost function C is not indexed by i. This enforced commonality in supply costs across sectors. A shift in the cost function for one sector will affect all; there is no "industrial" electricity. Additionally, since C shows up with the same sign (positively) in all supply functions, it must be that, *ceteris paribus*, a change in C will affect all sectors' equilibrium prices in the same direction. This observation, together with the understanding that the demand side of the market is not directly affected by price changes for other sectors, is what will allow us to identify the effect of price changes at the sectoral level. Let us assume the following functional forms for ease of exposition:

$$\begin{cases} q_D^i & = \beta_0^i + \beta_1^i p^i + u^i \\ q_S^i & = \gamma_0^i + \gamma_1^i p^i + \gamma_2^i \tilde{q}_c + v^i, \end{cases}$$

for some random variable $\tilde{q}_c$ which stands in for exogenous shifts in the utility's overall cost function, and so is not indexed by i. This characterization should make clear that it is plausibly uncorrelated with u^i; this is the assumption that the remaining rests on. We solve the above system and find

$$p^{i*} = \frac{\beta_0^i - \gamma_0^i - \gamma_2^i \tilde{q}_c + u^i - v^i}{\gamma_1^i - \beta_1^i},$$

which makes obvious that $\mathbb{E}[p^{i*} u^i] \neq 0$. However, we also have , for some $j \neq i$,

$$\mathbb{E}[p^{j*} u^i] = -\frac{\gamma_2^j}{\gamma_1^j - \beta_1^j} \mathbb{E}[\tilde{q}_c u^i] + \frac{1}{\gamma_1^j - \beta_1^j} \mathbb{E}[u^i u^j - u^i v^j].$$

From our assumption that $\tilde{q}_c$ is a cost shifter for the utility, it follows that the first term is zero. The second term is zero if we believe that errors across sectors are uncorrelated; this requires further justification. In fact, it is unlikely to hold perfectly, primarily due to omitted variables bias. It is entirely plausible, even guaranteed, that multiple sectors' demand would be impacted by similar factors. For example, differing state regulatory frameworks are likely to have impacts for at least industrial and commercial customers. To address this concern, we include state-level fixed effects (among other covariates) in the regression model we estimate.

4.2 Reduced Form Results

As mentioned previously, the empirical strategy rests on the idea that we can use the prices facing sector j as instruments for the prices facing sector j'; that is, industrial customers' prices are instrumented for using commercial customers' prices, and vice versa. This inspires the following reduced form equation:

$$\begin{aligned} y_{i,t}^{j} = \beta_0 &+ \beta_1 \log\Big([\overline{p}]_{i,t}^{j}\Big) + \beta_2 \log\Big([p_{\text{off}}]_{i,t}^{j}\Big) + \beta_3 \log\Big([p_{\text{on}}]_{i,t}^{j}\Big) + \beta_4 \log\big(w_{i,t}^{j}\big) \\ &+ \beta_5 e_{i,t}^{j} + \beta_6 \text{urb}_i + \gamma_t + \delta_i + u_{i,t}^{j}, \end{aligned}$$

where y is the outcome variable (either total electricity consumption or sector-wide output), $\overline{p}$ is the average price for electricity, w is the average wage, e^j the total employment in industry or sector j, urb is an indicator variable which is 1 when the county is considered urban according to the National Center for Health Statistics classification, and γ_t and δ_i are year and state fixed effects, respectively. The average price, here, is the *observed* average price over all electricity sales, not analogous to a simple time average. It is recovered from revenue and sales data at the utility level, and is included only in some specifications of the model. To handle the issue of simultaneity, we introduce a first stage of the form

$$\begin{aligned} x_{i,t}^{j} = \gamma_0 + \gamma_1 \log\Big([\overline{p}]_{i,t}^{j}\Big) &+ \gamma_2 \log\Big([p_{\text{off}}]_{i,t}^{-j}\Big) + \gamma_3 \log\Big([p_{\text{on}}]_{i,t}^{-j}\Big) \\ &+ \gamma_4 \log(w_{i,t}) + \gamma_5 e_{i,t}^{-j} + \epsilon_{i,t}^{j}, \end{aligned}$$

where e^{-j} is the total employment in industries other than j, $x \in \{\overline{p},\, p_{\text{off}},\, p_{\text{on}},\, w_{i,t}\}$, and the other variables are defined similarly to before.

Before considering any coefficient estimates, we empirically examine our two-stage least-squares strategy, insofar as that is possible. For all regressions estimated, the Wu-Hausman test indicates that (contingent on the validity of our selected instruments) the second-stage would suffer from endogeneity bias if estimated according to ordinary least squares. We also find that the F-statistics for each endogenous regressor (considered independently as the outcome variable for the first stage) are all larger than 200, and so in addition to the theoretical arguments presented earlier, empirical evidence suggests that this approach will not suffer weak instruments. We also now state the identification assumption explicitly: we assume

$$\mathbb{E}\left[\left(\log\left([p_{\text{off}}]_{i,t}^{-j}\right),\, \log\left([p_{\text{on}}]_{i,t}^{-j}\right),\, \log(w_{i,t}),\, e_{i,t}^{-j}\right)' u_{i,t}^{j}\right] = 0.$$

Under this assumption, the parameters of interest β_1, β_2, and β_3 can be consistently estimated by two-stage least squares. Of course, this assumption only involves the cross-section, and so we should mention the restrictions required to handle issues arising from the panel nature of the data set. We will allow for arbitrary correlation between errors within states in any given year and for particular counties across time; that is, we cluster at the state and year levels. The results from these regressions are summarized below, in Table 1.

Table 1: TSLS Coefficient Estimates, Average Price Included

Outcome Variable	Elasticity Estimates				
	Average price included			*Average price excluded*	
	p_{off}	p_{on}	$\overline{p}$	p_{off}	p_{on}
Industrial GDP	-0.0525	0.4904	0.0492	-0.0528	0.5093
	(0.0261)	(0.2223)	(0.2221)	(0.0257)	(0.2127)
Industrial Quantity Demanded	-0.1337	1.365	-0.6525	-0.1291	1.1141
	(0.0537)	(0.5194)	(0.8071)	(0.0512)	(0.6411)
Commercial GDP	-0.0143	0.5109	-0.8340	-0.02838	0.1780
	(0.0192)	(0.2644)	(0.5680)	(0.0193)	(0.2089)
Commercial Quantity Demanded	-0.0967	1.794	-2.089	-0.1319	0.9597
	(0.0594)	(0.5088)	(0.7447)	(0.0552)	(0.3974)
Manufacturing GDP	-0.0318	0.1959	0.0308	-0.0318	0.1959
	(0.0146)	(0.1033)	(0.1247)	(0.0146)	(0.1033)

State and year clustered standard errors are provided below point estimates.

Of course, the estimates provided here must be very carefully interpreted; there is a nearly mechanistic relationship between p_{off} and p_{on}, and $\overline{p}$. We pursue a strategy intended to account for this issue in 4.3. As presented here, all estimates for the elasticities of p_{off} and p_{on} must be taken in context with the accompanying estimate for overall (average) price elasticity.

However, even taking the prior discussion into account, when the average price variable is omitted, the coefficient estimates do not change substantially. Most strikingly, under this specification, we retain positive elasticity estimates for the on-peak prices. Considering these are the prices which generally coincide with peak demand, it seems peculiar that higher prices would cause anything other than *lower* quantity demanded. One way to interpret these results is that, in spite of the earlier arguments, there is some underlying problem with the empirical strategy which results endogeneity that biases the coefficient estimates so severely as to change their sign. If this is in fact the case, there are a handful of potential causes. The most plausible, and indeed even likely, cause stems from the exact way that time-varying pricing is often implemented. In many jurisdictions, the time-varying pricing is not universally mandated. Sometimes, this means that only some customers are required to switch, and sometimes it is entirely voluntary. In these situations, it stands to reason that greater peak demand across all sectors may be associated with utilities' decisions to offer time-varying pricing (in an attempt to discourage such high peaks); however, if firms refuse to take up the time-varying pricing, there may be a statistical correlation between higher quantity demanded and the offering of time-varying prices.

This discussion highlights the importance of a structural approach as in Section 2.1, which makes explicit the discrete choice problem underlying firms' decisions. Making concrete the heterogeneity within a sector, and the resulting impact on aggregate measures of electricity consumption, is the aim of subsequent sections.

4.3 Structural Empirical Strategy

We now turn to recovering the parameters characterizing the distribution of α_1 (and hence α_2) on a sector-region level via estimation of the structural model from Section 2.2. To do so, we impose a distributional assumption on the density f_α, namely that

$$\alpha_{1\varsigma i} \sim \text{TruncNorm}(\mu, \sigma^2, 0, \tilde{\alpha}_\varsigma - \alpha_{\varsigma L}),$$

the truncated normal distribution with parameters μ and σ^2, supported on the interval $[0, \tilde{\alpha}_\varsigma - \alpha_{\varsigma L}]$. The explicit form for its density function is, for $X \sim \text{TruncNorm}(\mu, \sigma^2, 0, b)$,

$$f_X(x) = \frac{1}{\sigma} \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)}$$

(where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and CDF of the standard normal distribution, respectively).

We begin our estimation by first recovering estimates for the values of $\gamma_{\varsigma t}$, δ_i, $\tilde{\alpha}_\varsigma$, and $\alpha_{L\varsigma}$, applying instrumental variables on a restricted subsample. We consider only sector-county-year observations which had no time-variant pricing available, ensuring that all firms were subject to a fixed rate, regardless of type. Taking logs of our earlier derivation of optimal consumption restricted to fixed-rate pricing, we are left with (given region-sector fixed effects

$\delta_{\varsigma i})$,

$$\ln\left(\frac{E_{\varsigma it}}{N_{\varsigma it}}\right) = \ln\left(\mathbb{E}\left[\frac{\alpha_2^{\alpha_2}\alpha_1^{\alpha_1}}{\alpha_1^{\tilde{\alpha}_\varsigma}}\,\middle|\,\tilde{\alpha}_\varsigma\right]\right) + \ln(\tilde{\alpha}_\varsigma - \alpha_L) + \frac{1}{1-\tilde{\alpha}}\left[\ln(\gamma_{\varsigma t}) + \ln(\delta_{\varsigma i}) + \alpha_{2\varsigma k}\ln(\alpha_{2\varsigma k}) + \right.$$
$$\left. + (\alpha_{L\varsigma} - 1)\ln(p_{0\varsigma it}) + (-\alpha_{L\varsigma})\ln(w_{\varsigma it}) + \ln(\epsilon_{\varsigma it})\right],$$

which we estimate, instrumenting for $\ln(p_{0\varsigma it})$ and $\ln(w_{\varsigma it})$ with $\ln(p_{0\varsigma' it})$ and $\ln(w_{\varsigma' it})$ similarly to the procedure outlined in Section 4.1. (Here, ς' indicates the other sector, e.g. $\varsigma = \text{ind} \implies \varsigma' = \text{com}$.) This allows us to consistently estimate $\frac{\alpha_{L\varsigma}-1}{1-\tilde{\alpha}_\varsigma}$ and $-\frac{\alpha_{L\varsigma}}{1-\tilde{\alpha}_\varsigma}$, which together identify both $\tilde{\alpha}_\varsigma$ and $\alpha_{L\varsigma}$[1]. Additionally, from this same estimation procedure, we get consistent estimates of $\gamma_{\varsigma t}$, which we will make use of in later estimation.

Throughout this derivation, we have used "region" rather than "county" to describe both the fixed effects and electricity output elasticities. This is intentional, as we will impose that $\delta_i = \delta_j$ when i and j are counties within the same Census Division. This grouping intends to balance . Concretely, we understand the separation into regions as a partition of $\{1, 2, \ldots, J\}$, the list of all counties, into regions $R_1, R_2, \ldots, R_9$. In a slight abuse of notation, we will use as summation and product indices the expression $j \in R_k$, understood to mean all county-year observations whose county is included in region R_k. Associated with each of these regions we now estimate a total-region fixed effect for each R_k, using the formula

$$\hat{\delta'}_k = \frac{\sum_{j\in R_k}\log(x_j) - \log(\hat{x}_j)}{\sum_j \mathbb{1}(j \in R_k)},$$

which gives us an estimate for the regional TFP based on our earlier modelling assumption that $\mathbb{E}[\ln(\epsilon_{it})] = 0$. To feasibly compute this, we take x_j to be county j's electricity consumption per firm, and estimate the same quantity (given prices and year fixed-effects, but omitting the regional fixed effects) on the same fixed-price subsample that we used to estimate $\hat{\gamma}_t$, $\hat{\tilde{\alpha}}_\varsigma$, and $\hat{\alpha}_{\varsigma L}$.

Equipped with these parameter estimates, we return to the density proposed in Section 2.2, together with the explicit form for the density of α_1 and the standard log-normal density for g, so that

$$g(x) = \frac{1}{x\sqrt{2\pi}}\exp\left(-\frac{\ln(x)^2}{2}\right).$$

This finally yields

[1] For the industrial sector, the estimate of α_{LI} was found to be flawed, as it would be consonant only with a *positive* causal relationship between wages and output. As such, the estimate for α_{LI} is instead replaced with 0.6, roughly consistent the evidence presented in Autor et al. (2020)

$$h_\varsigma(x;\mathbf{p}_{\varsigma it},\gamma_{\varsigma t},\delta_{\varsigma i}) = \left(\frac{1-\tilde{\alpha}_\varsigma}{\gamma_{\varsigma t}\delta_i\Lambda_{\varsigma it}^{1-\tilde{\alpha}_\varsigma}x^{\tilde{\alpha}_\varsigma}}\right) g\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right)\left(\frac{x}{\Lambda_{\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right)$$

where we recall that $\Lambda_{\varsigma it}$ is a function of the density of α_1, and therefore of the parameters which describe its distribution, μ and σ^2. With the density g fixed and the parametric family for f_α assumed, we can now proceed with maximum likelihood estimation. For region k and sector ς, we have the likelihood function

$$\mathcal{L}(\mu_k,\sigma_k^2;\mathbf{p}_{\varsigma it},\gamma_{\varsigma t},\delta_{\varsigma k},\tilde{\alpha}_\varsigma,\alpha_{L\varsigma}) = \prod_{j\in R_k}\left(\frac{1-\tilde{\alpha}_\varsigma}{\gamma_{\varsigma t}\delta_k\Lambda_{\varsigma j}^{1-\tilde{\alpha}_\varsigma}x_j^{\tilde{\alpha}_\varsigma}}\right) g\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right)\left(\frac{x_j}{\Lambda_{\varsigma j}}\right)^{1-\tilde{\alpha}_\varsigma}\right),$$

with corresponding log-likelihood function $\ell(\mu_k,\sigma_k^2;\mathbf{p}_{\varsigma it},\gamma_{\varsigma t},\delta_{\varsigma t},\tilde{\alpha}_\varsigma,\alpha_{L\varsigma}) = \log(\mathcal{L})$. While this likelihood function is a continuous and differentiable function of both μ and σ^2, it is not obviously generally concave. It is this observation that poses the greatest challenge to the empirical results which follow, in Section 4.4. Without proof of concavity, the statistical guarantees often associated with maximum likelihood estimation can not be assumed[2]. Additionally, due to the computation cost of computing $\Lambda_{\varsigma it}$, which requires numerical integration, it is advantageous to employ a method that evaluates the objective function a relatively small number of times. However, because we can analytically compute the partial derivatives with respect to μ and σ^2 of ℓ, we can attempt to use a gradient-based approach such as a quasi-Newton's method.

In particular, we have

$$\frac{\partial\ell}{\partial a} = (1-\tilde{\alpha}_\varsigma)\sum_j\left[\ln\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right)\left(\frac{x}{\Lambda_{\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right)\left(\frac{1}{a}+\frac{\frac{\partial}{\partial a}\Lambda_{\varsigma j}}{\Lambda_{\varsigma j}}\right)\right]$$

and

$$\frac{\partial\ell}{\partial b} = (1-\tilde{\alpha}_\varsigma)\sum_j\left[\ln\left(\left(\frac{1}{\gamma_{\varsigma t}\delta_i}\right)\left(\frac{x}{\Lambda_{\varsigma it}}\right)^{1-\tilde{\alpha}_\varsigma}\right)\left(\frac{1}{b}+\frac{\frac{\partial}{\partial b}\Lambda_{\varsigma j}}{\Lambda_{\varsigma j}}\right)\right].$$

The explicit form of these partial derivatives, in their entirety, are included in Appendix B.

In practice, we find that numerical issues relating to the computation of the far tails of a normal distribution result in badly-behaved derivative calculation for many of the cases (Census Division $\times$ sector) we wish to estimate. As such, we take advantage of the low-

[2]Visual inspection of the objective function, indicates that it is indeed well-behaved (that is, smooth and not often stationary), if not globally concave. This point offers a substantial opportunity for future work, to provide modifications or adaptations to the model stated here while maintaining concavity and the associated statistical properties.

dimensionality of the objective function and perform iteratively restricted grid search, beginning with a large, coarse grid and progressively refining around regions of highest objective function value (this method is closely analogous to the deterministic focused grid search from **?**). While this method also lacks theoretical guarantees given the non-concavity of the objective function, the final estimates are produced from grids on the order of 10^{-1} to 10^{-2}, and so in the absence of extremely strong non-linearities our estimates should be close to the true maximizers.

4.4 Structural Estimation Results

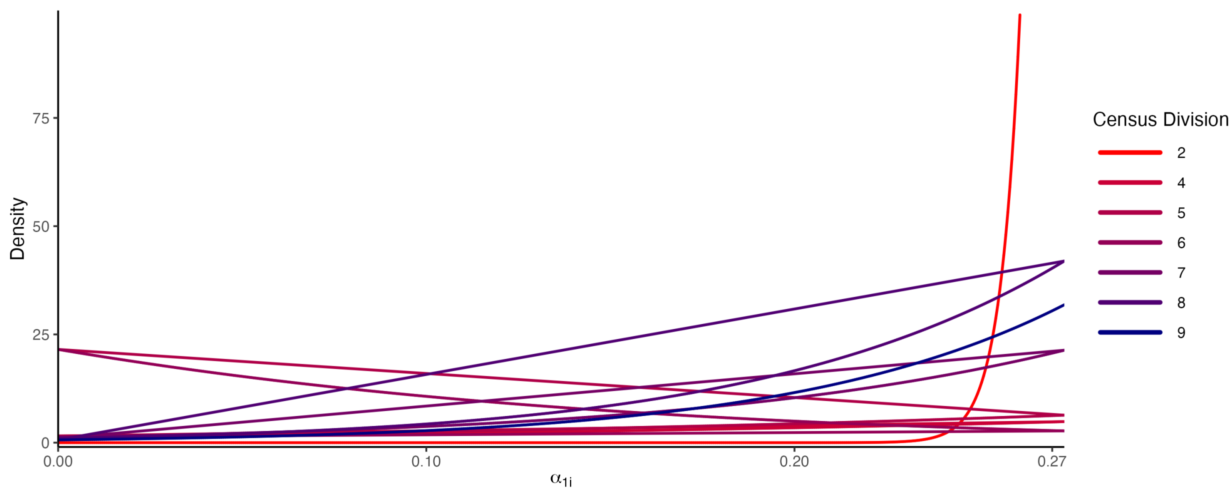
We first provide estimates for $\theta = (\mu, \sigma^2)$ for each sector, by Census Division. The results from the maximum likelihood estimation are presented in Table 2 and visualized in Figure 1. One finding which immediately stands out is that, while there is substantial heterogeneity across Census Divisions for the key parameters of interest among industrial customers, there is no estimated heterogeneity at all in the commercial sector. This statement must of course be understood in light of the discussion at the end of Section 4.3, namely that numerical difficulties make exact estimates unreliable as $\widehat{\sigma^2}$ approaches 0. This problem impacts the estimates from Census Divisions 1 and 3 for industrial customers, and all Census Divisions for commercial customers. However, the imperfect estimates for these cases strongly indicate point masses within the interior of the support (for both industrial estimates) and at the right edge of the support (for the commercial estimates); see Appendix C for further discussion.

Table 2: Estimated values for μ and σ^2

Industrial sector		
Census Division	$\hat{\mu}$	$\widehat{\sigma^2}$
1 (New England)	0.2	$\to 0$
2 (Middle Atlantic)	7.6	0.039
3 (East North Central)	0.1	$\to 0$
4 (West North Central)	8.7	1.80
5 (South Atlantic)	6.1	1.15
6 (East South Central)	-1.0	0.15
7 (West South Central)	2.7	0.25
8 (Mountain)	1.5	0.10
9 (Pacific)	3.0	0.20
Commercial sector		
Census Division	$\hat{\mu}$	$\widehat{\sigma^2}$
1-9 (All Divisions)	1.7	$\to 0$

The common pattern amongst all Divisions for the commercial sector implies a strong degree of similarity between the distribution of firms' relative elasticities of on- and off-peak electricity across space. In fact, this strongly aligns with the intuition that, in most areas of the country, broadly similar commercial establishments (e.g. retail stores and office buildings) are present; industrial activity, in contrast, features a much richer pattern of specialization along dimensions such as specific types of industrial processes which are relevant to this estimation. Of course, the exact economic activity being classified as commercial may vary drastically from place to place, but this exercise presents suggestive evidence that the pattern of electricity demand is relatively homogeneous.

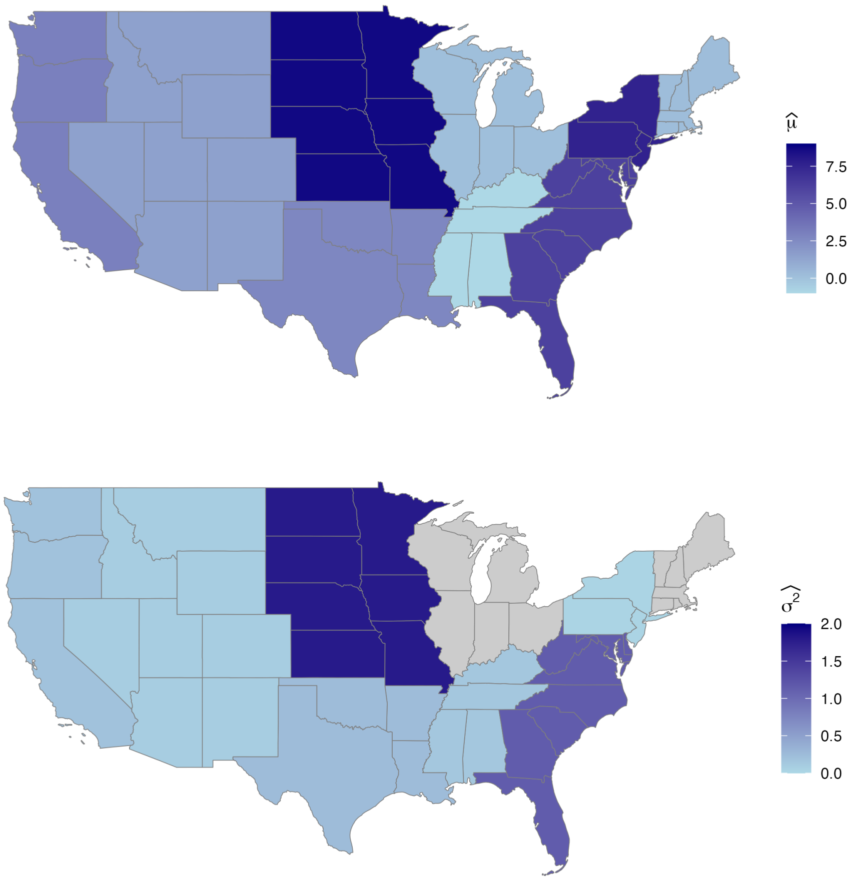
Figure 1: Estimated densities, off-peak electricity elasticity



Notes: Census Divisions 1 and 3 encountered numerical difficulties, but converged approximately to point masses at the corresponding value of $\hat{\mu}$ given in Table 2.

Turning our attention to the differences in estimates across Census Divisions for industrial customers, we see that there are essentially three classes of distributions: approximately uniform distributions (e.g. Division 5); distributions with interior modes, well-approximated by a point mass on the interior of the support (Divisions 1 and 3); and those well-approximated by a point mass at the right edge of the support (Division 2). For direct comparison purposes, see a visualization of all but Divisions 1 and 3 in Figure 1. Interestingly, almost all regions outside of the northeast have broadly similar distributions, falling into the aforementioned first class. In fact, only Census Division 2 (Middle Atlantic, which is comprised of New York, New Jersey, and Pennsylvania) has a distribution among its industrial customers which closely matches the estimated distribution for commercial customers across the country. We visualize this spatial heterogeneity in Figure 2.

Figure 2: Spatial heterogeneity in off- vs. on-peak electricity elasticity, industrial customers



Notes: Census Divisions 1 and 3 encountered numerical difficulties, but converged approximately to point masses at the corresponding value of $\hat{\mu}$ given in Table 2.

5 Conclusion

In this paper, we present a new model for characterizing a particular aspect of firms' production, explicitly separating out production elasticities for electricity between on-peak and off-peak consumption. Furthermore, we provide estimates for the regional distribution of these parameters. We also offer reasons why a naïve estimation strategy making use of two-stage least squares is unlikely to recover any particular parameter of interest, and thereby provide ample motivation for pursuing a structurally-grounded empirical strategy.

Although initially motivated by questions which face energy policymakers in guiding the transition to renewable electricity, the approach and estimates derived here have the potential for much broader relevance. In particular, should more granular contexts enable the use of micro data, the approach introduced here could allow for more fine-grained intrasectoral comparisons, rather than only commenting on the differences between industrial and commercial customers. Additionally, if this analysis was replicated at a finer geographical level (for example, at the neighborhood level), it could be used to inform the decisions of urban planners and regional development policymakers.

Particularly as substantial spatial heterogeneity was found along this on- vs. off-peak dimension of substitution, this work is also relevant to the active literature in spatial and environmental economics on optimal electricity networks as well as climate and economic-geographic research which takes into account shifting patterns of electricity generation and consumption over the coming decades.

References

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Blonz, J. A. (2016). Making the best of the second-best: Welfare consequences of time-varying electricity prices. *Working paper*.

Borenstein, S. (2012). The private and public economics of renewable electricity generation. *Journal of Economic Perspectives*, 26(1):67–92.

Borenstein, S. and Bushnell, J. (2018). Do two electricity pricing wrongs make a right? cost recovery, externalities, and efficiency. *NBER working paper*.

Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *American Economic Review*, 104(2):537–563.

Jessoe, K. and Rapson, D. (2015). Commercial and industrial demand response under mandatory time-of-use electricity pricing. *The Journal of Industrial Economics*, 63(3):397–421.

Lazar, J. (2014). Teaching the duck to fly. *Regulatory Assistance Project*.

OpenEI and NREL (2022). OpenEI U.S. Utility Rate Database. https://apps.openei.org/USURDB/. Accessed: 2 January 2022.

Steinberg, D., Bielen, D., Eichman, J., Eurek, K., Logan, J., Mai, T., McMillan, C., Parker, A., Vimmerstedt, L., Wilson, E., and et al. (2017). Electrification and decarbonization: Exploring u.s. energy use and greenhouse gas emissions in scenarios with widespread electrification and power sector decarbonization. *National Renewable Energy Laboratory*.

U.S. BEA (2022). CAGDP9 Data Set. https://apps.bea.gov/regional/downloadzip.cfm. Accessed: 2 January 2022, select “CAGDP9: Real GDP in Chained Dollars by County and MSA”.

U.S. Census Bureau (2022). County Business Patterns (CBP) Datasets. https://www.census.gov/programs-surveys/cbp/data/datasets.html. Accessed: 2 January 2022.

U.S. EIA (2022). Annual Electric Power Industry Report, Form EIA-861 detailed data files. https://www.eia.gov/electricity/data/861/. Accessed: 2 January 2022.

A Analysis of inclusion in estimation data

This appendix addresses the issue raised in Section 3, which noted that not all county-sector-year observations could be included in the estimation sample due to missing data for electricity prices. This is summarised in Figure A.1, which shows the geographic distribution of the number of years a (mainland US) county is in the estimation sample for each sector. We see that, while there is some concentration in certain states (for example, Washington state is over-represented relative to its neighbors), there is good coverage overall.

Furthermore, while we do not have electricity pricing at the appropriate level for some county-year-sector, we can nevertheless compare the values of other key county indicators, such as GDP and total electricity sales, to arrive at a rough comparison between those included and those not included. This comparison is shown visually in Figure A.2 and numerically in Table A.1, which provides t-tests for the same covariates. While some of the means are shown to be statistically significantly different in Table A.1, it remains true that the the distributions are qualitatively similar, and critically, that the apparent differences in means are economically insignificant (only representing a small proportion of the typical county's GDP and electricity sales).

Figure A.1: Comparison between observations included and missing from rate database

(a) Industrial (b) Commercial

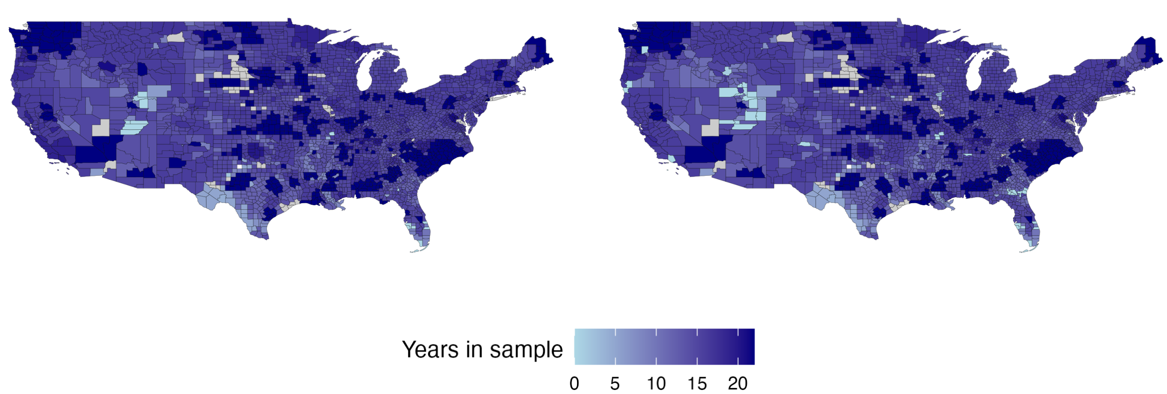


Figure A.2: Comparison between observations included and missing from rate database

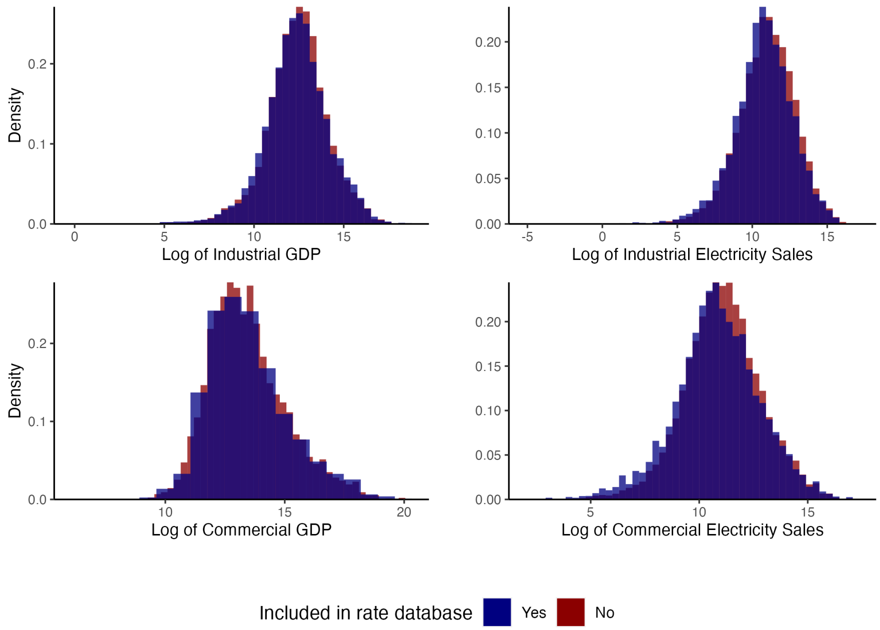


Table A.1: Comparison between county-sector-years included and not included in estimation sample

	Industrial		*Commercial*	
	GDP	Electricity sales	GDP	Electricity sales
Difference in means ($000s)	-69	36	-755	3
t-statistic	-2.03	6.85	-4.11	0.29

B Log-Likelihood Derivatives

From the log-likelihood function considered (and maximized) in Sections 4.3 and 4.4, we obtain the following first partial derivatives:

$$\frac{\partial l}{\partial \mu} = \frac{(1-\tilde{\alpha}_\varsigma)}{\sigma^2} \sum_j \left\{ \ln \left(\left(\frac{x_j}{\int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} E_j^*(\alpha_1) f_\alpha(\alpha_1; \mu, \sigma^2)\, d\alpha_1} \right)^{1-\tilde{\alpha}_\varsigma} \right) \right.$$
$$\left. \times \left[\frac{\sigma \left(\phi \left(\frac{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \mu}{\sigma} \right) - \phi \left(\frac{-\mu}{\sigma} \right) \right)}{\Phi \left(\frac{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \mu}{\sigma} \right) - \Phi \left(\frac{-\mu}{\sigma} \right)} + \frac{\int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} \hat{I}_j(\alpha_1; \mu, \sigma^2)(\alpha_1 - \mu)\, d\alpha_1}{\int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} \hat{I}_j(\alpha_1; \mu, \sigma^2)\, d\alpha_1} \right] \right\},$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{1-\tilde{\alpha}_\varsigma}{2\sigma^2} \sum_j \Bigg\{ \ln \left(\left(\frac{x_j}{\int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} E_j^*(\alpha_1) f_\alpha(\alpha_1; \mu, \sigma^2)\, d\alpha_1} \right)^{1-\tilde{\alpha}_\varsigma} \right)$$

$$\times \left[-1 + \frac{\phi\left(\frac{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \mu}{\sigma}\right)(\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \mu) - \phi\left(\frac{-\mu}{\sigma}\right)(-\mu)}{\sigma\left(\Phi\left(\frac{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma} - \mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)\right)} \right.$$

$$\left. - \frac{1}{\sigma^2} \frac{\int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} \hat{I}_j(\alpha_1; \mu, \sigma^2)(\alpha_1 - \mu)^2\, d\alpha_1}{\int_0^{\tilde{\alpha}_\varsigma - \alpha_{L\varsigma}} \hat{I}_j(\alpha_1; \mu, \sigma^2)\, d\alpha_1} \right] \Bigg\},$$

where

$$\hat{I}_j = E^*(\alpha_1) \exp\left(\frac{1}{2} \left(\frac{\mu - \alpha_1}{\sigma} \right)^2 \right),$$

and the exchange of the derivative and integral is justified by Leibnitz's Rule.

C Behavior of a Truncated Normal Distribution, $\sigma \to 0$

Consider a truncated normal distribution characterized as before by the corresponding normal distribution's mean μ, standard deviation σ, and the (truncated) support. As in Section 4.3, let the support be on $[0, b]$ for some $b > 0$.

Now, consider the following: fixing μ and b, what is the qualitative description of this distribution as $\sigma \to 0$? There are two primary cases to examine:

1. $\mu \in [0, b]$: much like with a non-truncated normal distribution, the distribution will converge to a point mass at μ.
2. $\mu \notin [0, b]$: if $\mu < 0$, the distribution will converge to a point mass at 0, and analogously if $\mu > b$ it will converge to a point mass at b.

We will prove the second case as follows: first, notice that, on either side of its mean, the normal distribution's density function is monotonic (i.e. for the standard normal distribution, $\phi(x) < \phi(x')$ for $x < x' < 0$, and $\phi(x) > \phi(x')$ for $0 < x < x'$). Assume first that $\mu > b$ (the case which is practically relevant for the estimation results in Section 4.4. Recall that the density of a truncated normal distribution is given by

$$f^{\text{T.Norm.}}(x; \mu, \sigma, b) = \frac{1}{\sigma} \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)}$$

Then, we know that our corresponding truncated normal density $f^{\text{T.Norm.}}$ must be monotonically increasing, as all terms in the definition apart from $\phi(\frac{x-\mu}{\sigma})$ are constant with respect to x. Because the truncated normal distribution has compact support, and we know that its density is monotonically increasing, we can compare the density value at any two points in the support and thereby get a sense of the distribution as $\sigma \to 0$.

In particular, fix some $b' \in [0, b)$, and see that

$$\begin{aligned}
\frac{f^{\text{T.Norm.}}(b; \mu, \sigma, b)}{f^{\text{T.Norm.}}(b'; \mu, \sigma, b)} &= \frac{\phi\left(\frac{b-\mu}{\sigma}\right)}{\phi\left(\frac{b'-\mu}{\sigma}\right)} \\
&= \exp\left[-\frac{1}{2\sigma^2}\left((b-\mu)^2 - (b'-\mu)^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(b^2 - 2b\mu - \mu^2 - b'^2 + 2b'\mu + \mu^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(b - b'\right)\left(b + b' - 2\mu\right)\right].
\end{aligned}$$

From this, we can see that

$$\lim_{\sigma \to 0} \frac{f^{\text{T.Norm.}}(b; \mu, \sigma, b)}{f^{\text{T.Norm.}}(b'; \mu, \sigma, b)} = \lim_{\sigma \to 0} \exp\left[-\frac{1}{2\sigma^2}\left(b - b'\right)\left(b + b' - 2\mu\right)\right].$$

Now, since $b' < b < \mu$, we see that the non-σ terms inside of the exponentiation collapse to a positive constant with respect to the limit. From this, it follows that the ratio diverges towards $+\inf$. This implies, as we have left our choice of b' arbitrary, that the density function grows without bound at the right edge of the support relative to any other point in the support; that is, the limiting distribution must be a point mass at the support's upper bound b.

An extremely similar argument shows that, for $\mu < 0$, the distribution converges to a point mass at the lower bound 0. In fact, even if we allowed μ to vary with σ, the same results would hold (so long as all values of μ along the limit path are on the appropriate side of $[0, b]$).