NBER WORKING PAPER SERIES

REGULARIZATION FROM ECONOMIC CONSTRAINTS:
A NEW ESTIMATOR FOR MARGINAL EMISSIONS

Stephen P. Holland
Erin T. Mansur
Valentin Verdier
Andrew J. Yates

Working Paper 32065
http://www.nber.org/papers/w32065

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2024

We thank James Archsmith, Andrii Babii, Alex Belloni, Stéphane Bonhomme, Allan Collard-Wexler, Max Farell, Iván Fernández-Val, Chris Hansen, Soonwoo Kwon, Lihua Lei, Tianqi Li, Adam McCloskey, Chuck Manski, Matt Masten, Chris Muris, Ingmar Prucha, Pedro Sant'Anna, Frank Wolak, Jeff Wooldridge, Kaspar Wüthrich, Yinchu Zhu, and Jiacheng Zou for many helpful comments and suggestions, as well as seminar participants at the University of Maryland, Duke University, the University of North Carolina - Greensboro, the Triangle Resource and Environmental Economics Workshop, and EI@Haas Energy Camp. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Regularization from Economic Constraints: A New Estimator for Marginal Emissions
Stephen P. Holland, Erin T. Mansur, Valentin Verdier, and Andrew J. Yates
NBER Working Paper No. 32065
January 2024
JEL No. C23,Q42,Q48,Q53

ABSTRACT

Environmental policy is increasingly concerned with measuring emissions resulting from local changes to electricity consumption. These marginal emissions are challenging to measure because electricity grids encompass multiple locations and the information available to identify the effect of each location's consumption on grid-wide emissions is limited. We formalize this as a high-dimensional aggregation problem: The effect of electricity consumption on emissions can be estimated precisely for each electricity generating unit (generator), but not precisely enough to obtain reliable estimates of marginal emissions by summing these effects across all generators in a grid. We study how two economic constraints can address this problem: electricity supply equals demand and an assumption of monotonicity. We show that these constraints can be used to formulate a 'naturally regularized' estimator, which implements an implicit penalization that does not need to be directly tuned. Under an additional assumption of sparsity, we show that our new estimator solves the high-dimensional aggregation problem, i.e., it is consistent for marginal emissions where the usual regression estimator would not be. We also develop an asymptotically valid method for inference to accompany our estimator. When applied to the U.S. electricity grid with 13 separate consumption regions, our method yields plausible patterns of marginal generation across fuel types and geographic location. Our estimates of region-level marginal emissions are precise, account for imports/exports between regions, and allow for all fuel types to potentially be on the margin.

Stephen P. Holland
Bryan School of Business and Economics
University of North Carolina, Greensboro
P.O. Box 26165
Greensboro, NC 27402-6165
and NBER
sphollan@uncg.edu

Erin T. Mansur
Dartmouth College
100 Tuck Hall
Hanover, NH 03755
and NBER
erin.mansur@dartmouth.edu

Valentin Verdier
Department of Economics,
The University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
vverdier@email.unc.edu

Andrew J. Yates
Department of Economics
and Curriculum for the
Environment and Ecology
The University of North Carolina at Chapel Hill
CB 3305
Chapel Hill, NC 27599
ajyates@email.unc.edu

1 Introduction

Environmental policies addressing climate change and air pollution rely increasingly on electrification. The environmental effects of the resulting changes in electricity consumption are captured by marginal emissions. Marginal emissions are used by researchers for policy design and evaluation;[1] by policy makers for policy implementation;[2] and by the private sector to report or reduce the environmental effects of their electricity consumption.[3]

A widely used approach to estimate marginal emissions relies on a regression analysis, which identifies marginal emissions from the correlation between plausibly exogenous variation in electricity consumption and emissions across the grid.[4] However, this approach may be subject to excessive estimation noise, especially when estimating location-specific marginal emissions, for instance at the level of regions that partition the grid. This problem originates from the interconnected nature of electricity grids. As electricity grids encompass multiple regions, there is limited information available to separately identify the effect of each region's electricity consumption on grid-wide emissions. For example, we show that a regression of emissions on the electricity consumption of the entire eastern U.S. electricity grid yields a precise estimate of grid-level marginal emissions whereas breaking up eastern consumption into nine regions leads to unreliable estimates of region-level marginal emissions.

In this paper, we propose a new estimator for marginal emissions that utilizes generator-level data and augments a regression model with additional restrictions derived from the properties of electricity grids. We show, both theoretically and empirically, that this estimator can yield reliable estimates of marginal emissions where existing methods could not.

To understand the intuition behind our approach, consider the case where electricity consumption increases in a particular location of the grid. One or more electricity generators (i.e., electricity generating units) increase their generation to meet this increase in consumption. When these generators use fossil fuels, this leads to an increase in each of these generators'

[1] See, for instance, electricity pricing in Borenstein and Bushnell (2022), electrification of transportation in Gillingham, Ovaere, and Weber (2021), and renewable generation in Callaway, Fowlie, and McCormick (2018).

[2] Region-specific marginal emissions are to be reported per the Infrastructure Investment and Jobs Act, 117 USC § 40412 (2021), and subsidies for hydrogen production created by the Inflation Reduction Act, 117 USC § 13204 (2022) are determined by marginal emissions (see Bergman, Prest, and Palmer (2022)).

[3] See Buchanan et al. (2023) on carbon-aware computing.

[4] Marginal emissions are distinct from average emissions which can be calculated exactly instead of being estimated, even at local levels (de Chalendar, Taggart, and Benson (2019)). This corresponds to the EIA's reported emission intensities of load, `https://www.eia.gov/electricity/gridmonitor/about`.

emissions. Marginal emissions sum these effects across all generators. Instead of estimating marginal emissions directly, our method first relies on estimating marginal generation (the effect of consumption on generation) for each generator. This change of perspective reveals a set of intuitive restrictions. The first restriction follows from a grid's balancing condition. Since electricity consumption must equal generation at all times, the sum of marginal generation across all generators must be equal to one. The second restriction is an assumption of monotonicity which we argue is plausible in our setting. When electricity consumption increases, we assume that any particular generator may, on average, not respond or increase its generation, but it does not systematically decrease its generation. Our estimator simply imposes these economic constraints (balancing and monotonicity) on a regression to estimate each generator's marginal generation. After combining these estimates of marginal generation with the emissions profile of each generator, we estimate marginal emissions by aggregating the resulting generator-level responses.

From a theoretical standpoint, we formalize the issues that appear with existing regression approaches as a high-dimensional aggregation problem: generator-level effects can be estimated precisely, but not precisely enough to obtain reliable estimates of marginal emissions by summing these effects across all generators. We show that our new estimator can address this problem under a final assumption of sparsity: for each consumption region, there is a subset of electricity generators (unknown to the researcher) that responds to changes in this region's electricity consumption, while generators outside of that subset do not respond. We show that our constrained regression estimator performs a model selection that can yield consistency even in high-dimensional settings. We call this estimator 'naturally regularized' because it implicitly relies on a penalization, but the penalization is determined by the data and constraints (balancing and monotonicity) instead of being tuned by the researcher. We also provide an asymptotically valid method for inference to accompany our estimator.

To evaluate the performance of our new method from an empirical standpoint, we apply it to data from the U.S. electricity grid and report region-specific marginal emissions. We find that the model selection performed by our estimator yields plausible patterns in fuel type and geographic location of marginal generation, and that it yields precise estimates of marginal emissions. Previous studies that use regression methods to estimate marginal emissions tend either to report results that impose homogeneity across regions (e.g., Holland et al. (2022b) report marginal emissions for the eastern U.S. electricity grid), or impose a specific form of sparsity by assuming that all of the electricity generators that respond to a region's changes in electricity consumption are contained within that region (e.g., Callaway, Fowlie, and McCormick (2018)). While the first approach would mask heterogeneity in marginal emissions, we find that the second approach tends to exaggerate the degree of heterogeneity in marginal emissions across regions.[5] For example, New England does not contain almost any coal generators but we find that, through imports, coal generation in other regions accounts for a significant share of its marginal generation. This can lead to our estimates of the environmental effect of electricity consumption in New England being over 50% greater than when imports are ignored. In general, our method is able to use the data to uncover heterogeneity and sparsity patterns without assuming that they are known to researchers ex ante.

The ability to quantify variation in marginal emissions across regions more flexibly while retaining precise results is important because the locational variation in marginal emissions plays a central role in many environmental policies. For instance, even with a nationally uniform subsidy for purchasing electric vehicles, the environmental gain is determined by heterogeneous marginal emissions because the adoption of electric vehicles varies across the country (EIA (2023) and Department of Energy (2023)).[6] Location-specific marginal emissions are also required to optimize the process of replacing electricity generation with cleaner technologies (e.g., renewable energy) by prioritizing locations with high marginal emissions to maximize environmental gains.[7] Conversely, one could choose locations with the lowest marginal emissions to locate activities that consume electricity heavily, such as hydrogen producing plants, to minimize the resulting environmental cost.[8]

As a methodological contribution, our study can be seen as combining two literatures —

[5]This is because, with a region-by-region analysis, marginal emissions are entirely determined by the fleet of electricity generators contained within each region. If a region, say, tends to have high-emissions electricity generators, a region-by-region analysis would conclude that its marginal emissions are inordinately high, while in reality this difference may be compensated if the region imports some of its marginal electricity from regions with lower emissions electricity generators.

[6]An example of such subsidy is the $7,500 tax credit made fully refundable by the Inflation Reduction Act, 117 USC § 13401 (2022).

[7]Davis, Holladay, and Sims (2022), Butters, Dorsey, and Gowrisankaran (2023), and Johnson and Yang (2023) document that the process of upgrading the grid is costly and time consuming. It can therefore be optimized by prioritizing certain locations.

[8]This is implemented by the conditional subsidies for hydrogen production of the Inflation Reduction Act discussed in footnote 2 above.

one on estimating functionals of interest from noisy measurements, the other on regularized estimation — to study the estimation of a novel object of interest: the aggregation of many heterogeneous effects. Because we consider a panel data setting where noisy measurements of electricity generator responses can be obtained, our work is related to the literature on heterogeneous effects and noisy measurements. Our regression model will be a correlated random coefficient model of panel data, as in Chamberlain (1984) and Chamberlain (1992), and a large subsequent literature (see Sasaki and Ura (2022) for a recent review).[9] In our setting, regularized estimation will need to be considered because our object of interest is a high-dimensional linear functional, i.e., a sum of many heterogeneous effects, so that typical results on consistency (e.g., Chamberlain (1992) for average effects) do not apply.

Our analysis differs from much of the previous work on regularization methods because our estimator is naturally regularized, i.e., tuning parameter-free. Typical results on the consistency of LASSO (see, e.g., Hansen (2022) for a textbook treatment, and Manresa (2015) or Belloni et al. (2016) for work in panel data settings) generally impose a restriction on the convergence rate of the penalization parameter set by the researcher.[10] Our estimator implicitly relies on a similar penalization parameter (the Lagrange multiplier of summing-to-one constraints), but this quantity is determined by the constraints and the data, so that its convergence rate cannot be assumed and must instead be characterized.

This issue has been studied before in the context of constrained least-squares (see, e.g., Meinshausen (2013), Slawski and Hein (2013), Hastie, Tibshirani, and Wainwright (2015), Vershynin (2015), Zerbib et al. (2016), or Plan, Vershynin, and Yudovina (2017)) or, as a special case, synthetic control estimation (see, e.g., Abadie, Diamond, and Hainmueller (2010), Doudchenko and Imbens (2016), Athey et al. (2021), Arkhangelsky et al. (2021), Cattaneo, Feng, and Titiunik (2021), Chernozhukov, Wuthrich, and Zhu (2022), Spiess, Imbens, and Venugopal (2023)).[11] Relatedly, the regularizing power of restricting a high-dimensional pa-

[9]Other work on noisy measurements has studied the bias that arises when learning about non-linear functionals (e.g., distribution functions) based on noisy measurements (see, e.g., Efron (2011) and Jochmans and Weidner (2022) for reviews of a large literature, see Kwon (2021) for a related discussion of this issue in panel data settings).

[10]LASSO is only one approach to regularization in high-dimensional settings. There is a growing literature on high-dimensional panel data models. See, e.g., Bonhomme and Manresa (2015), Chernozhukov et al. (2019), Liu, Moon, and Schorfheide (2020), Arkhangelsky et al. (2021), Chetverikov and Manresa (2022), Viviano and Bradic (2023), and Moon and Weidner (2023) for new results and reviews of the literature.

[11]In many implementations of synthetic control, a researcher wishes to estimate (positive and summing to one) weights from panel data on outcomes.

rameter space to the standard simplex or, more generally, to a L_1-ball, has also been studied before (e.g., Ye and Zhang (2010), or Raskutti, Wainwright, and Yu (2011)). As discussed below, existing consistency results in this previous work do not apply in our setting because we consider a high-dimensional functional: a weighted sum of regression coefficients rather than a weighted average or a single coefficient. We will show that it is the combination of our specific panel data setting with a restriction to the standard simplex that yields positive results on consistency and asymptotically valid inference under a final assumption of sparsity.[12]

In section 2, we set up our problem formally, before presenting our main theoretical results in section 3 and applying our method to data from the U.S. electricity grid in section 4.

2 Estimating Marginal Emissions and the High-Dimensional Aggregation Problem

2.1 Regression-based approaches for estimating marginal emissions

Previous work — Siler-Evans, Azevedo, and Morgan (2012), Graff Zivin, Kotchen, and Mansur (2014), Holland et al. (2016), Callaway, Fowlie, and McCormick (2018), Clinton and Steinberg (2019), Fell and Johnson (2021), Sexton et al. (2021), Dauwalter and Harris (2023) among others — has relied on regression analysis to estimate marginal emissions.[13] Among this work, one approach uses data on electricity consumption across various regions that partition an electricity grid as well as data on emissions across all generators that make up the grid.[14] A regression is then used to measure the association of exogenous changes in electricity consumption (the "right-hand side" variables) with changes in emissions across all

[12]The same characteristics of our problem make existing methods for inference with high-dimensional models and model selection inapplicable (see, e.g, van de Geer et al. (2014), Belloni, Chernozhukov, and Hansen (2014), Farrell (2015), Cattaneo, Jansson, and Newey (2017), van de Geer (2019)). We propose an approach to inference based on debiasing and sample-splitting that makes use of the summing-to-one constraint and the panel structure of our data.

[13]The main alternative to regression methods is the use of engineering dispatch models (see Deetjen and Azevedo (2019) for a recent review), where the researcher simulates the allocation of electricity generation across the grid by minimizing the cost of electricity generation subject to the physical constraints of the grid. The main advantage of this approach is to directly account for the physical constraints of the electrical network, while its main shortcomings are the absence of uncertainty and the difficulty of accounting for economic and administrative complexities (e.g., market power, division of the grid into local administrative units). Our general view is that improved regression estimates can provide useful directions to improve dispatch models, and vice versa. In ongoing work, we compare the behavior of our method with that of a dispatch model by using both methods concurrently.

[14]An alternative regression-based approach (e.g., Callaway, Fowlie, and McCormick (2018)) estimates marginal emissions for each region independently, as discussed below.

generators (the "left-hand side" variable).[15]

Let $Emissions_t$ denote the total emissions on an electricity grid at time $t \in \mathcal{T}$.[16] Index consumption regions by $k \in \mathcal{K}$, define $x_{t,k}$ to be the electricity consumption of region k at time t, and $x_t = [x_{t,1}, ..., x_{t,K}]$ to collect consumption across all regions, where $K = |\mathcal{K}|$. The marginal emissions from a change in electricity consumption at time $t \in \mathcal{T}$ in region $k \in \mathcal{K}$ are given by $\frac{\partial Emissions_t}{\partial x_{t,k}}$.[17] A regression analysis estimates this object of interest using the regression model:

$$Emissions_t = x_t \Delta + z_t \boldsymbol{\gamma} + \mathbf{v}_t, \qquad E(\mathbf{v}_t | X, Z) = 0,$$

where $\Delta = [\Delta_1, ..., \Delta_K]'$ collects marginal emissions across all regions, z_t is a vector of control covariates that isolates plausibly exogenous variation in electricity consumption, $\mathbf{v}_t$ is an error term, and $X = \{x_t : t \in \mathcal{T}\}$, $Z = \{z_t : t \in \mathcal{T}\}$.

Marginal emissions are then estimated by the OLS regression estimator $\hat{\Delta} = [\hat{\Delta}_1, ..., \hat{\Delta}_K]'$ obtained from a regression of $Emissions_t$ on x_t (and control covariates) using all observations indexed in $\mathcal{T}$. Section 4.2 documents the unreliability of this approach when estimating region-level marginal emissions with data from the U.S. electricity grid.

To explain this issue and our solution, we first explicitly account for the fact that a grid's emissions are determined by the aggregation of emissions across all electricity generators on the grid. Index each generator on the electricity grid by $i \in \mathcal{N}$. We then have $Emissions_t = \sum_{i \in \mathcal{N}} w_{it}$, where w_{it} is generator i's emissions at time t. Therefore, we can write marginal emissions as the sum of the generator-specific marginal emissions, $\frac{\partial Emissions_t}{\partial x_{t,k}} = \sum_{i \in \mathcal{N}} \frac{\partial w_{it}}{\partial x_{t,k}}$. We could estimate generator-specific marginal emissions with a regression model:

$$w_{it} = x_t \theta_i + z_t \gamma_i + v_{it}, \qquad E(v_{it} | X, Z) = 0, \tag{1}$$

from which OLS regression estimators $\hat{\theta}_i$ can be obtained by regressing generator-specific emissions w_{it} on x_t (and control covariates) using all observations indexed in $\mathcal{T}$. Marginal emissions can be represented as the sum of the generator-specific marginal emissions, $\Delta =$

[15]We concentrate on the effect of changes to electricity consumption on emissions throughout this paper for simplicity since changes in electricity generation using renewable energy (e.g., wind or solar) can be seen as negative changes in net electricity consumption.

[16]In practice most researchers consider emissions of CO_2, NO_X, and SO_2, either separately or aggregated into a single measure of damages. This is discussed in further detail in section 4 below.

[17]Note that marginal emissions intend to capture short run effects of changes in electricity consumption or generation, see Holland et al. (2022b), Gagnon et al. (2022), and Holland et al. (2022a) for a discussion.

$\sum_{i\in\mathcal{N}}\theta_i$, and could be estimated by a simple plug-in estimator, $\sum_{i\in\mathcal{N}}\hat{\theta}_i$. One can easily show that an algebraic equivalence holds, and we obtain the same estimator for marginal emissions as when using a single regression with grid-level emissions, i.e., $\hat{\Delta} = \sum_{i\in\mathcal{N}}\hat{\theta}_i$. With this representation, we can formalize the issues that arise with this method as a high-dimensional aggregation problem.

2.2 High-dimensional aggregation

Intuitively, θ_i, the effect of consumption x_t on generator-specific emissions w_{it}, may be estimated relatively precisely by $\hat{\theta}_i$ for each generator $i \in \mathcal{N}$, as long as the researcher observes relatively many time periods in $\mathcal{T}$. However, our object of interest requires that these estimates be summed across all generators, $\hat{\Delta} = \sum_{i\in\mathcal{N}}\hat{\theta}_i$. Estimation error will generally accumulate in a sum so that, with many generators, it is possible that the estimator $\hat{\Delta}$ is excessively noisy even though each estimator $\hat{\theta}_i$ is relatively precise.[18]

To see this, consider a special case where the error term v_{it} in (1) is uncorrelated and homoscedastic. Let $n = |\mathcal{N}|$ be the number of generators in the data, $T = |\mathcal{T}|$ be the number of time periods in the data. When the covariates x_t have bounded variation, we have:[19]

$$\lambda_{\min}(\mathrm{Var}(\hat{\Delta})) \geq c \cdot \frac{n}{T},$$

where $\lambda_{\min}(.)$ denotes the minimum eigenvalue of a matrix and $c > 0$ is a positive constant. When the number of generators (n) is large relative to the number of observations available over time (T), i.e., $\frac{n}{T} \not\to 0$, marginal emissions Δ cannot be estimated consistently by the plug-in estimator $\hat{\Delta}$. We call this case a high-dimensional aggregation problem.[20]

[18]In our empirical application, the U.S. grid has over 1,500 power plants. While we are able to split the U.S. grid into three approximately independent interconnections, the largest interconnection (eastern interconnection), has over 1,000 power plants.

[19]Bounded variation in x_t here means that $\lambda_{\max}(\frac{1}{T}\sum_{t\in\mathcal{T}} x_t' x_t) \leq C \ \forall T$ for some constant $C > 0$, where $\lambda_{\max}(.)$ denotes the largest eigenvalue of a matrix. We therefore have $\mathrm{Var}(\hat{\Delta}) = \sum_{i\in\mathcal{N}}\mathrm{Var}(\hat{\theta}_i) = n\cdot\frac{\sigma^2}{T}(\frac{1}{T}\sum_{t\in\mathcal{T}} x_t' x_t)^{-1}$, and $\lambda_{\min}(\mathrm{Var}(\hat{\Delta})) = \frac{n}{T}\frac{\sigma^2}{\lambda_{\max}(\frac{1}{T}\sum_{t\in\mathcal{T}} x_t' x_t)} \geq c\cdot\frac{n}{T}$, where $\sigma^2 = \mathrm{Var}(v_{it}) > 0$.

[20]Siler-Evans, Azevedo, and Morgan (2012) and Callaway, Fowlie, and McCormick (2018) address the issues that arise with OLS regressions to estimate region-level marginal emissions by restricting marginal emissions effects to only occur within regions (e.g., changes to New England electricity consumption only lead to changes in electricity generation among New England plants). Fell and Johnson (2021) and Dauwalter and Harris (2023) use a similar approach. Since electricity grids encompass multiple regions, this approach may miss "imports" and "exports" in electricity between regions. As will be discussed in the next section, our approach can be seen as a data-driven "coarsening" of the power plants that respond to each region. We will impose an assumption of sparsity but will not assume that the researcher knows the structure of this sparsity (e.g., we will not require that generation response only occurs within consumption regions). Section 4 will provide empirical evidence of the biases that may arise when using a region-by-region analysis.

2.3 Regularized aggregation

Our approach to address the high-dimensional aggregation problem uses two main steps. First, we decompose marginal emissions into the product of marginal generation (a generator's change in generation due to changes in consumption across the grid) and emission factors (a generator's change in emissions due to a change in this generator's generation). Second, we show that the balancing properties of the electricity grid can be used to precisely estimate the first element of the decomposition, marginal generation, by making use of a natural regularization that results from imposing balancing constraints onto our estimator.

Formally, the decomposition of marginal emissions into marginal generation and emission factors is $\frac{\partial w_{it}}{\partial x_{t,k}} = \frac{\partial w_{it}}{\partial y_{it}} \frac{\partial y_{it}}{\partial x_{t,k}}$, where y_{it} is generator i's electricity generation at time t. We then apply the same regression model as (1) but to electricity generation:

$$y_{it} = x_t b_i + z_t \zeta_i + u_{it}, \qquad E(u_{it}|X, Z) = 0, \ \forall\, i \in \mathcal{N}, \tag{2}$$

where $b_i = [b_{i,1}, ..., b_{i,K}]'$ captures the effect of changing consumption $x_{t,k}$ for each region $k \in \mathcal{K}$ on the electricity generation of generator $i \in \mathcal{N}$.

Throughout the paper we assume a simple linear relationship between a generator's emissions and its generation, so that $\frac{\partial w_{it}}{\partial y_{it}} = \omega_i$.[21] Marginal emissions are therefore given by $\Delta = \sum_{i\in\mathcal{N}} \omega_i b_i$.

For now consider the case where emission factors ω_i are known for simplicity. Without further restrictions, one could estimate marginal generation effects based on the regression model (2) by a generator-specific regression, $\hat{b}_i = \text{argmin}_{\beta_i \in \mathbb{R}^K} \sum_{t\in\mathcal{T}} (y_{it} - x_t \beta_i)^2$ $\forall\, i \in \mathcal{N}$, which can be equivalently represented by a pooled regression across all generators: $\min_{\{\beta_i \in \mathbb{R}^K : i \in \mathcal{N}\}} \sum_{(i,t)\in\mathcal{N}\times\mathcal{T}} (y_{it} - x_t\beta_i)^2$. One could then use these estimated marginal generation effects to estimate marginal emissions by $\sum_{i\in\mathcal{N}} \omega_i \hat{b}_i$. This estimator will be subject to the same high-dimensional aggregation problem as the estimator discussed in the previous subsection.[22]

[21] One can relax this assumption to accommodate variation over time or non-linearities in emission factors. For instance, Archsmith (2022) presents an analysis that accounts for the emissions due to adjustments in electricity generation levels (i.e., power plants ramping up or down). We make an assumption of linearity here to expose clearly our central contribution on the estimation of marginal generation effects. Our approach to estimating marginal generation effects could be paired with any approach to estimating generator-level emission factors.

[22] In fact, if emissions are exactly determined as a linear function of generation for each generator, $w_{it} = \omega_i y_{it}$, then the two estimators will be algebraically equivalent, $\sum_{i\in\mathcal{N}} \hat{\theta}_i = \sum_{i\in\mathcal{N}} \omega_i \hat{b}_i$.

To obtain an improved estimator of marginal emissions, we combine additional restrictions on marginal generation with the regression model (2). First, we note that electricity grids must always satisfy balancing constraints, whereby electricity produced throughout the grid equates electricity consumed. This means that, for any region k, a change in consumption $x_{t,k}$ must be matched by a corresponding change in generation across the grid, i.e., $\sum_{i\in\mathcal{N}} b_{i,k} = 1$ $\forall\, k \in \mathcal{K}$.[23]

Second, we impose an additional assumption of monotonicity. For any region $k \in \mathcal{K}$ and any generator $i \in \mathcal{N}$, it is possible that, on average, increases in consumption $x_{t,k}$ do not affect generation y_{it} ($b_{i,k} = 0$) or lead to increases in generation ($b_{i,k} > 0$), but generator i does not systematically decrease its generation in response to increases in consumption ($b_{i,k} < 0$ is ruled out). Since the regression model (2) captures effects averaged over time for each generator, this restriction of monotonicity allows for some generators to decrease their output in response to an increase in electricity consumption at particular points in time as long as their average response is positive. This captures the intuitive idea that, as electricity consumption increases, more and more factors of production come into use.[24]

We propose using these two restrictions by simply imposing them as constraints on the regression estimator:

$$\{\tilde{b}_i : i \in \mathcal{N}\} = \arg\min_{\{\beta_i \in \mathbb{R}^K : i \in \mathcal{N}\}} \sum_{(i,t)\in\mathcal{N}\times\mathcal{T}} (y_{it} - x_t\beta_i)^2 \text{ s.t. } \beta_i \geq 0\,\forall\, i,\, \sum_{i\in\mathcal{N}} \beta_i = \mathbf{1}_K, \tag{3}$$

where "$\geq$" denotes an element-by-element comparison, and $\mathbf{1}_K$ is a $K \times 1$ vector of ones. We refer to the constraints as positivity and summing-to-one, respectively, and note that this amounts to restricting our estimator to the standard simplex. With known emissions factors ω_i, one could estimate marginal emissions with a plug-in estimator $\sum_{i\in\mathcal{N}} \omega_i \tilde{b}_i$.

In the next section, we show that imposing the constraints in (3) leads to a natural reg-

[23]Consumption, or more precisely load, is measured as a region's electricity generation adjusted for the net electricity interchange with neighboring regions. Because load equals the quantity of electricity consumers use plus local transmission and distribution losses, the marginal generation coefficients $b_{i,k}$ must sum to one. To calculate how emissions change with respect to household electricity use, our estimates of marginal emissions could be scaled to account for local transmission and distribution losses. The EIA estimates that these losses are approximately 5% (EIA (2022)). Accounting for marginal interregional transmission losses is beyond the scope of this paper.

[24]In this paper, we take both monotonicity and summing-to-one constraints as being true. Perlman (1969), Wolak (1987), Wolak (1989) have considered the problem of testing inequality and equality constraints in linear regression models. It is also interesting to note that Wolak (1987) considers a constrained regression estimator similar to ours, albeit in the context of testing inequality and equality constraints rather than for estimating a high-dimensional linear functional as we do.

ularization that can solve the high-dimensional aggregation problem formalized above when a final assumption of sparsity is imposed on the marginal generation b_i. We call the estimator (3) naturally regularized because it does not rely on a penalization parameter that the researcher must choose. Instead, the Lagrange multiplier corresponding to each of the summing-to-one constraints plays the role of a penalization parameter.

2.3.1 Estimation error in emission factors and estimating marginal generation on the standard simplex

There is an additional advantage of restricting the estimation of marginal generation effects to the standard simplex when estimating marginal emissions. Plant-specific emission factors, ω_i, will generally also be estimated by the researcher. Denote the resulting estimator by $\hat{\omega}_i$. When combined with estimated marginal generations $\tilde{b}_i$ that are positive and sum to one, the incidence of the estimation error in $\hat{\omega}_i$ on the estimation error in estimated marginal emissions $\sum_{i\in\mathcal{N}}\hat{\omega}_i\tilde{b}_i$ will vanish asymptotically under mild restrictions. This is because

$$|\sum_{i\in\mathcal{N}}\hat{\omega}_i\tilde{b}_{i,k}-\sum_{i\in\mathcal{N}}\omega_i b_{i,k}|\leq\sum_{i\in\mathcal{N}}|\hat{\omega}_i-\omega_i|\tilde{b}_{i,k}+|\sum_{i\in\mathcal{N}}\omega_i(\tilde{b}_{i,k}-b_{i,k})|,$$

where the inequality is obtained from the triangle inequality and because $\tilde{b}_{i,k}\geq 0$.

The next section is devoted to the more complex problem of establishing that $|\sum_{i\in\mathcal{N}}\omega_i(\tilde{b}_{i,k}-b_{i,k})|$ converges to zero stochastically even in high dimensional settings (i.e., when $n,T\to\infty$ with $\frac{n}{T}\not\to 0$), but here one can immediately note that, since $\tilde{b}_{i,k}\geq 0$ and $\sum_{i\in\mathcal{N}}\tilde{b}_{i,k}=1$, we have $\sum_{i\in\mathcal{N}}|\hat{\omega}_i-\omega_i|\tilde{b}_{i,k}\leq\max_{i\in\mathcal{N}}|\hat{\omega}_i-\omega_i|=O_p(\sqrt{\frac{\log(n)}{T}})$, where the last equality is obtained under typical regularity conditions. Thus, when constraining estimated marginal generation to be positive and sum to one, we drastically reduce the order of estimation uncertainty in marginal emissions due to the estimation of emission factors.[25]

3 Constrained Estimation as a Natural Regularization

We begin our study of asymptotic results with a simplified setting in which a single electricity consumption region is included in the model, so that x_t is a scalar, and there are no control covariates. Motivated by the results above on the estimation error due to estimating emission factors, ω_i, we also treat ω_i as known. Section 3.5 discusses several extensions, with

[25] When using the unconstrained regression estimator $\hat{b}_i$, $\sum_{i\in\mathcal{N}}|\hat{\omega}_i-\omega_i|\hat{b}_{i,k}$ would generally not converge to zero when $\frac{n}{T}\not\to 0$ (e.g., it would be of order $\frac{n}{T}$ under serial and cross-sectional uncorrelation when the estimation errors in b_i and in ω_i are correlated).

the most noteworthy being a multivariate setting where multiple regions $k \in \mathcal{K}$ determine a vector of electricity consumption x_t. Before doing that, we: (i) define our simplified univariate setting, (ii) show that consistent estimation is impossible without further restrictions, (iii) show that our constrained estimator is consistent under an additional assumption of sparsity, and (iv) propose an asymptotically valid approach to inference for marginal emissions.

3.1 Univariate model and estimators

In our simplified setting, we have a constrained univariate regression model where a single covariate x_t determines electricity generation y_{it} across all generators $i \in \mathcal{N}$. For notational simplicity, we will treat x_t as deterministic throughout this section (equivalently, one can interpret all statements below as being conditional on $X = \{x_t : t \in \mathcal{T}\}$). We then consider the model:

$$y_{it} = b_i x_t + u_{it}, \qquad E(u_{it}) = 0, \tag{4}$$

$$b_i \geq 0, \ \forall\, i \in \mathcal{N}, \qquad \sum_{i \in \mathcal{N}} b_i = 1, \tag{5}$$

and our object of interest is the scalar $\Delta = \sum_{i \in \mathcal{N}} \omega_i b_i$, where ω_i is known.

The unconstrained OLS estimator of marginal generation effects is given by $\hat{b}_i = \frac{\sum_{t \in \mathcal{T}} x_t y_{it}}{\sum_{t \in \mathcal{T}} x_t^2}$ and a plug-in estimator is $\hat{\Delta} = \sum_{i \in \mathcal{N}} \omega_i \hat{b}_i$.[26] Our constrained estimator is given by:[27]

$$\{\tilde{b}_i : i \in \mathcal{N}\} = \arg\min_{\{\beta_i,\, i \in \mathcal{N} : \beta_i \geq 0,\, \sum_{i \in \mathcal{N}} \beta_i = 1\}} \sum_{(i,t) \in \mathcal{N} \times \mathcal{T}} (y_{it} - \beta_i x_t)^2. \tag{6}$$

The corresponding plug-in estimator for Δ is $\tilde{\Delta} = \sum_{i \in \mathcal{N}} \omega_i \tilde{b}_i$. The purpose of the results presented below is to understand whether this constrained estimator is able to achieve con-

[26] As discussed in the introduction, this estimator is similar to Correlated Random Coefficient model estimators (see Chamberlain (1992) and subsequent papers), but we consider aggregate marginal effects ($\sum_{i \in \mathcal{N}} \omega_i b_i$ with $\sum_{i \in \mathcal{N}} \omega_i \propto n$) and not average marginal effects ($\sum_{i \in \mathcal{N}} w_i b_i$ with $\sum_{i \in \mathcal{N}} w_i = 1$).

[27] It is interesting to note the similarity of this estimator with the synthetic control (SC) estimator of Abadie, Diamond, and Hainmueller (2010) and a very large subsequent literature. Many versions of SC estimators also base their estimation on a restriction of the estimated parameters to the standard simplex. More generally, the estimator (6) can also be seen as a positive constrained LASSO estimator, see Raskutti, Wainwright, and Yu (2011), Hastie, Tibshirani, and Wainwright (2015), Doudchenko and Imbens (2016), Zerbib et al. (2016), Chernozhukov, Wüthrich, and Zhu (2021), or Cattaneo, Feng, and Titiunik (2021) for related discussions. The differences arise from the fact that, in SC or constrained LASSO, a researcher generally considers a model with a high-dimensional vector of covariates x_t (e.g., has many control units in SC) and a single outcome variable y_t (e.g., a single treated unit in SC), whereas we have outcome variables y_{it} for a large number of units $i \in \mathcal{N}$ but a single covariate x_t (or finitely many covariates in section 3.5 below). In addition, our object of interest is the aggregate effect $\Delta = \sum_{i \in \mathcal{N}} \omega_i b_i$, whereas SC generally considers a per-period or average effect such as $\frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} (y_{it} - x_t b_i)$. Other examples of estimation restricted to the standard simplex in econometrics include Ichimura and Thompson (1998), Fox et al. (2011), and Heiss, Hetzenecker, and Osterhaus (2022) who study mixture estimators, but in a different setting and without the asymptotic results that we present here.

sistency (and asymptotically valid inference) in high-dimensional settings (i.e., when $\frac{n}{T} \not\to 0$). The next subsection begins by establishing that, in fact, no estimator can achieve consistency under the restrictions in (4) and (5) alone. In subsection 3.3 we show that consistency is achieved by our constrained estimator under a final additional assumption of sparsity (while the unconstrained estimator remains inconsistent regardless of whether the assumption of sparsity holds or not).

3.2 Inconsistency without further restrictions

There are two competing phenomena determining the precision of estimated functionals based on the constrained estimator (6). On the one hand, the restriction of the constrained estimator to the standard simplex yields large precision gains.[28] On the other hand, the "signal" represented by marginal generation b_i may be too small compared to the noise embedded in any estimator for b_i to be able to estimate marginal emissions precisely. Intuitively, the possibility that there be very diffuse scenarios where all power plants on the grid have responses to changes in electricity consumption of similar magnitude, $b_i \propto \frac{1}{n}$, will prevent consistency from being achieved without imposing additional restrictions.

To establish this impossibility formally, we consider a special case of the regression model (4), (5), where the noise u_{it} is assumed to be homoscedastic and Gaussian:

$$y_{it} = b_i x_t + u_{it}, \qquad u_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \qquad \mathbf{b}_n \in \mathcal{B}_n, \tag{7}$$

where, to shorten notation, we defined $\mathbf{b}_n = [b_1, ..., b_n]'$ and $\mathcal{B}_n = \{[\beta_1, ..., \beta_n]' \in \mathbb{R}^n : \beta_i \geq 0 \, \forall \, i \in \mathcal{N}, \sum_{i \in \mathcal{N}} \beta_i = 1\}$ to be the standard simplex.

Define $N = (n, T)$ to combine both indices of sample size. We consider the minimax risk with squared error loss:

$$R_N = \inf_{\hat{\Delta}(.)} \sup_{\mathbf{b}_n \in \mathcal{B}_n} E((\hat{\Delta}(\mathbf{y}_N) - \Delta(\mathbf{b}_n))^2),$$

[28] This is established, for instance, by finding that $\tilde{b}_i$ is consistent for b_i in the L_2-norm under the constrained regression model (4), (5) and typical regularity conditions (namely, serial independence, bounded higher moments, and variation in x_t), with $\sum_{i \in \mathcal{N}} (\tilde{b}_i - b_i)^2 = O_p(\sqrt{\frac{\log(n)}{T}})$ from standard derivations (see, e.g., Hansen (2022), or Chernozhukov, Wüthrich, and Zhu (2021), Chernozhukov, Wuthrich, and Zhu (2022), Hastie, Tibshirani, and Wainwright (2015), Vershynin (2015), Plan, Vershynin, and Yudovina (2017), and Zerbib et al. (2016) in the context of synthetic control or constrained LASSO estimation), whereas we would have $\sum_{i \in \mathcal{N}} (\hat{b}_i - b_i)^2$ of order $\frac{n}{T} \not\to 0$ under cross-sectional independence, as discussed before. In our setting, L_2-consistency is not sufficient to consistently estimate our object of interest, as evidenced in Proposition 1 below. Instead, L_1-consistency for estimating $\{b_i : i \in \mathcal{N}\}$ would yield consistency for estimating marginal emissions since $|\sum_{i \in \mathcal{N}} \omega_i \tilde{b}_i - \sum_{i \in \mathcal{N}} \omega_i b_i| \leq \max_{i \in \mathcal{N}} |\omega_i| \sum_{i \in \mathcal{N}} |\tilde{b}_i - b_i|$. This difference in choice of norms is important in settings where one wishes to estimate high-dimensional functionals.

where $\Delta(\mathbf{b}_n) = \boldsymbol{\omega}_n'\mathbf{b}_n$, $\boldsymbol{\omega}_n = [\omega_1, ..., \omega_n]'$, and $\mathbf{y}_N = [y_{it} : (i,t) \in \mathcal{N} \times \mathcal{T}]'$.

Applying the results in Donoho (1994) enables us to establish that consistent point estimation (in terms of convergence in mean square) of Δ is impossible when $\frac{n}{\sum_{t\in\mathcal{T}} x_t^2}$ does not converge to zero (i.e., when n is not small relative to T).[29] Define $\bar{\omega} = \frac{1}{n}\sum_{i\in\mathcal{N}} \omega_i$ to be the average of ω_i.

Proposition 1. Under (7), if $\frac{1}{n}\sum_{i\in\mathcal{N}}(\omega_i - \bar{\omega})^2 \geq c_1$, $\max_{i\in\mathcal{N}}|\omega_i - \bar{\omega}| \leq C \; \forall n$, and $\frac{n}{\sum_{t\in\mathcal{T}} x_t^2} \geq c_2 \; \forall N$ for positive constants $c_1, c_2 > 0$ and $C < \infty$, then $R_N \geq c \; \forall N$ for a positive constant $c > 0$ (i.e., the minimax risk for estimating Δ does not converge to zero).

All proofs are in the online appendix. The first two conditions in Proposition 1, $\frac{1}{n}\sum_{i\in\mathcal{N}}(\omega_i - \bar{\omega})^2 \geq c$ and $\max_{i\in\mathcal{N}}|\omega_i - \bar{\omega}| \leq C$, restrict attention to the most interesting case where there is variation in ω_i of the same order as the number of cross-sectional units n and no single observation $i \in \mathcal{N}$ dominates in its contribution to the total variation in ω_i. It is easy to understand that, if these conditions were not imposed, some extreme cases, for instance when $\omega_i = 0$ for a majority of observations $i \in \mathcal{N}$, could yield consistent estimation since they would effectively be low-dimensional problems. Imposing these two conditions restricts attention to truly high-dimensional problems. Similarly, the condition $\frac{n}{\sum_{t\in\mathcal{T}} x_t^2} \geq c_2$ restricts attention to cases where n is large compared to T.[30]

3.3 Consistency under sparsity

As discussed in the previous section, consistent estimation of marginal emissions in high-dimensional settings will be impossible when there are very diffuse responses to changes in electricity consumption across the grid (e.g., $b_i \propto \frac{1}{n}$). In practice, however, researchers may not be overly concerned about such extreme scenarios. While a researcher may not know the exact set of electricity generators that responds to changes in electricity consumption in,

[29] Related results are found in Ye and Zhang (2010) and Raskutti, Wainwright, and Yu (2011) in the form of lower bounds for minimax L_r-risk for estimation on L_q-balls. As mentioned in footnote 28, L_1-consistency would imply consistency for Δ, but the converse is not a given, so that our result on the impossibility of consistently estimating Δ in our setting is, to our knowledge, novel.

[30] Despite the lack of consistency, it is possible that, in the simplified setting considered here with homoscedastic Gaussian errors, one could still perform valid bias-aware inference as in Armstrong and Kolesár (2018) or Armstrong, Kolesár, and Kwon (2023) since the single variance parameter $\sigma^2 = \mathrm{Var}(u_{it})$ can be estimated consistently. This simplified setting, however, is unlikely to fit the practical needs of researchers since heteroscedasticity both over time and across plants is highly likely (e.g., plants of different sizes, time periods with more volatility in electricity generation), as well as cross-sectional dependence (i.e., correlated shocks in electricity generation across plants).

say, New England, it may be plausible to assume that only a relatively small subset of all electricity generators on the electricity grid respond to these changes.

In the univariate case, this additional restriction is captured by an assumption of approximate sparsity:

$$b_i \leq \exp(-c_1 \cdot n) \; \forall \, i \notin \mathcal{A}_n, \qquad\qquad b_i \geq \frac{c_2}{|\mathcal{A}_n|} \; \forall \, i \in \mathcal{A}_n, \tag{8}$$

where $\mathcal{A}_n$ is a subset of cross-sectional observations, $\mathcal{A}_n \subseteq \mathcal{N}$, and c_1, c_2 are small positive constants, $c_1, c_2 > 0$.

Our assumption of approximate sparsity is composed of two parts. There is a response set $\mathcal{A}_n$ that collects the generators with a significant average response to changes in electricity consumption x_t. Generators not contained in $\mathcal{A}_n$ have marginal generations that are approximately zero. Generators contained in $\mathcal{A}_n$ have marginal generations that are greater than a threshold $\frac{c_2}{|\mathcal{A}_n|}$. This second part of the assumption is a "minimum signal" or "beta-min" condition, which is similar to conditions imposed in the Lasso literature to obtain consistent support recovery or variable selection results, as in, e.g., Wainwright (2009), Bühlmann and van de Geer (2011), or Lahiri (2021). Intuitively, assumption (8) rules out very diffuse cases discussed above where all generators across the electricity grid may have small responses of similar magnitudes to changes in electricity consumption.[31]

The role of approximate sparsity in the asymptotic behavior of $\tilde{\Delta}$ can be understood by noting that our constrained estimator of marginal generations can be expressed as a thresholding estimator:

$$\tilde{b}_i = 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N),$$

where γ_N is the Lagrange multiplier on the summing-to-one constraint and solves $\sum_{i \in \mathcal{N}} 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) = 1$.[32] Therefore, the estimator $\tilde{\Delta} = \sum_{i \in \mathcal{N}} \omega_i \tilde{b}_i$ can be seen as selecting a set of cross-sectional observations that is estimated to respond to the policy, $\hat{\mathcal{A}}_n = \{i \in \mathcal{N} : \hat{b}_i \geq \gamma_N\}$, and setting $\tilde{b}_i$ to zero for observations outside of the selected set $\hat{\mathcal{A}}_n$.

[31] Note that, for notational simplicity, we use the cardinality of the response set $\mathcal{A}_n$ to define the threshold $\frac{c_2}{|\mathcal{A}_n|}$ for the beta-min condition in (8). We could have defined this threshold separately, $b_i \geq r_N$, and imposed an additional rate condition $\sqrt{\frac{\log(n)}{T}} \frac{1}{r_N} \to 0$ to establish Proposition 2 below. As is, the condition $b_i \geq \frac{c_2}{|\mathcal{A}_n|}$ captures the practically relevant case where all electricity generators inside the response set have marginal generation effects of similar magnitudes.

[32] Note that this algebraic result is analogous to simple representations of the Lasso estimators for the orthonormal design case, as discussed in, e.g., Tibshirani (1996).

Under sparsity, it is possible that the resulting estimator is consistent if: (i) the true response set $\mathcal{A}_n$ is not too large compared to sample size T, and (ii) the estimated set $\hat{\mathcal{A}}_n$ approximates the true response set. To establish consistency formally, we impose a few regularity conditions.

Assumption 1. Assume that (i) u_{it} is serially independent, x_t and u_{it} have bounded support, (ii) the range of ω_i is bounded, (iii) $\frac{1}{T}\sum_{t\in\mathcal{T}} x_t^2 \geq c\ \forall T$ for a positive constant $c > 0$, and (iv) $\sigma_{it}^2 \geq c\ \forall i, t$ for a positive constant $c > 0$, where $\sigma_{it}^2 = \mathrm{Var}(u_{it})$.

For simplicity, Assumption 1 imposes relatively strong conditions of bounded support for x_t and u_{it} and serial independence of u_{it}, but these could be relaxed.[33] As discussed in the previous subsection, the case where ω_i is bounded, i.e., no observation dominates in its contribution to the total variation in ω_i, is the case of interest where a high-dimensional aggregation problem occurs. The condition $\frac{1}{T}\sum_{t\in\mathcal{T}} x_t^2 \geq c$ can also be relaxed, as the "effective sample size" is given by the total amount of variation in x_t, $\sum_{t\in\mathcal{T}} x_t^2$, but imposing that T and $\sum_{t\in\mathcal{T}} x_t^2$ be of the same order simplifies notation throughout this section. Finally, assuming that $\mathrm{Var}(u_{it})$ is bounded away from zero is standard and will allow us to apply high-dimensional central limit theorems and use normal asymptotic approximations.

Proposition 2. Under Assumption 1, the constrained regression model (4) and (5), and approximate sparsity (8), if $|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}} \to 0$, $n \cdot \exp(-c_1 \cdot n) = o(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}})$ where c_1 is the constant used in (8), and $\frac{\log^7(n)}{T} \to 0$, then as $n, T \to \infty$ we have: (i) (consistent selection) $P(\mathcal{A}_n \subseteq \hat{\mathcal{A}}_n) \to 1$, (ii) (rate of penalization) $\gamma_N = O_p(\sqrt{\frac{\log(n)}{T}})$, (iii) (consistent point estimation) $\tilde{\Delta} - \Delta = O_p(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}})$.

Proposition 2 establishes our main result regarding point estimation for the aggregate marginal effect Δ under sparsity. It requires that the number of observations that have a significant response to the policy x_t, $|\mathcal{A}_n|$, be small compared to the rate of precision with which each individual coefficient b_i can be estimated by $\hat{b}_i$, $\sqrt{T}$, up to a log term. It also requires that the approximation due to the approximate sparsity in (8) be asymptotically ignorable, $n \cdot \exp(-c_1 \cdot n) = o(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}})$. Finally, the condition $\frac{\log^7(n)}{T} \to 0$ is used to

[33] Namely, one could assume bounded higher moments for u_{it} (and for x_t if it were treated as random) at the expense of additional notation. Our results would also extend directly to cluster dependence with many small clusters, while extending our results to weak serial dependence (e.g., α-mixing) seems possible but would require more cumbersome notation, as discussed in section 3.5 below.

apply the high-dimensional central limit theorem of Chernozhukov, Chetverikov, and Kato (2017). Proposition 2 shows that the selected set $\hat{\mathcal{A}}_n$ contains the true response set with probability approaching one, and that this yields a consistent estimator for Δ.[34] Note that Proposition 2 does not rely on any restrictions on the cross-sectional dependence that may exist in the data (unlike the results on inference provided in the next subsection), so that the rate of convergence obtained in Proposition 2, $|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}}$, only differs from the rate of convergence of an oracle estimator that takes $\mathcal{A}_n$ as known, $\sum_{i\in\mathcal{A}_n}\hat{b}_i$, by a factor of $\log(n)$.[35]

To compare our results with existing results on high-dimensional linear models, note that we can stack our data across all observations to define the $nT \times 1$ vectors $\mathbf{y}_N = [y_{it} : (i,t) \in \mathcal{N} \times \mathcal{T}]'$, $\mathbf{u}_N = [u_{it} : (i,t) \in \mathcal{N} \times \mathcal{T}]'$, the $n \times 1$ vector $\mathbf{b}_n = [b_1, ..., b_n]'$, and the $nT \times n$ block-diagonal matrix $\mathbf{X}_N = I_n \otimes \mathbf{x}_T$, where $\otimes$ is a Kronecker product, I_n is the $n \times n$ identity matrix, and $\mathbf{x}_T = [x_1, ..., x_T]'$. We can write our regression model as:

$$\mathbf{y}_N = \mathbf{X}_N \mathbf{b}_n + \mathbf{u}_N, \qquad\qquad E(\mathbf{u}_N) = 0. \tag{9}$$

Defining $\boldsymbol{\omega}_n = [\omega_1, ..., \omega_n]'$, our object of interest is given by $\Delta = \boldsymbol{\omega}_n \cdot \mathbf{b}_n$. Sample size in our setting is given by $n \cdot T$, but observations are split into T independent clusters of size n. The parameter vector $\mathbf{b}_n$ is of dimension n, but is assumed to only contain $|\mathcal{A}_n|$ elements that are not approximately zero. Because of the panel data structure of our problem, we obtain an important property for the design matrix associated with $\mathbf{X}_N$ in (9) because $\mathbf{X}_N'\mathbf{X}_N = (\sum_{t\in\mathcal{T}} x_t^2) \cdot I_n$. This is why we obtain consistency results without having to specify additional restrictions on the covariates x_t (other than Assumption 1).

Note that an important difference between our proposed estimation method and traditional LASSO estimation is that our penalization parameter is the Lagrange multiplier on the summing-to-one constraint, γ_N, and is not tuned by the researcher. This result leads us to refer to the constrained estimator studied here as naturally regularized, since it leverages balancing constraints and the monotonicity assumption to break the issue of high-dimensional aggregation without requiring that the researcher take a stance on the degree of regulariza-

[34]Since $\tilde{b}_i \geq 0$ and $\sum_{i\in\mathcal{N}}\tilde{b}_i = 1$, we have $|\tilde{\Delta} - \Delta| \leq 2 \cdot \max_{i\in\mathcal{N}}|\omega_i|$, and Proposition 2 also implies convergence in mean square since $E((\tilde{\Delta}-\Delta)^2) \leq r_N^2 P(|\tilde{\Delta}-\Delta| \leq r_N) + 4 \cdot \max_{i\in\mathcal{N}}|\omega_i|^2 P(|\tilde{\Delta}-\Delta| > r_N) \to 0$ when choosing $r_N \to 0$ such that $\frac{1}{r_N}|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}} \to 0$. Therefore the results in Proposition 2 can be compared to Proposition 1 despite the difference in convergence criterion.

[35]See also Cai and Low (2004) (Theorem 7) who provide a lower bound of order $|\mathcal{A}_n|^2\frac{\log(n)}{T}$ for minimax mean squared error in a related setting.

tion. In addition, differences in convergence rates and approaches to inference arise from the high-dimensional linear functional considered, $\boldsymbol{\omega}_n \cdot \mathbf{b}_n$, instead of considering a single element of $\mathbf{b}_n$ or a weighted average. Even if a researcher wished to use a more traditional penalization as in LASSO, much of our results on consistency and inference would be novel because of the nature of our object of interest.

3.4 Inference after debiasing by sample splitting

In this section we discuss inference for Δ under the assumptions imposed in the previous subsection. Basing inference on the point estimator $\tilde{\Delta}$ is challenging because the high-dimensional selection implemented by the constrained estimator (6) leads to a selection bias that prevents the use of a normal approximation for the asymptotic distribution of the estimation error $\tilde{\Delta} - \Delta$. This is similar to the problem of inference with parameters on the boundary or inference post-model selection as in, e.g., Andrews (1999), Leeb and Pötscher (2005), Fang et al. (2022), or Andrews, McCloskey, and Kitagawa (2022), among many.[36]

A significant share of previous work on inference with high-dimensional models has relied on debiasing or doubly robust estimation, see, e.g., Belloni, Chernozhukov, and Hansen (2014), Efron (2014), van de Geer et al. (2014), Zhang and Zhang (2014), Farrell (2015), Belloni et al. (2016), van de Geer (2019), Chernozhukov et al. (2022).[37] In our context, debiasing will take a different form from the studies above because of the particular structure of our problem.[38] We will take advantage of the known sum of the coefficients b_i, $\sum_{i \in \mathcal{N}} b_i = 1$, to express our

[36] Note that Fang et al. (2022) have a setting that is related to ours, but their results cannot be applied here because the asymptotic framework that they consider rules out our high-dimensional setting where $\frac{n}{T} \not\to 0$.

[37] Instead of debiasing, another approach to inference relies on targeting uniform inference, adjusting critical values of tests so that they account for model selection. This was proposed in Berk et al. (2013), Bachoc, Preinerstorfer, and Steinberger (2020), Kuchibhotla et al. (2020), or Andrews, McCloskey, and Kitagawa (2022). See also Pelger and Zou (2022) for an application of this approach in a panel data setting. The difficulty in applying this approach here is in determining how critical values robust to model selection would be obtained to account for the restrictions $\mathbf{b}_n \in \mathcal{B}_n$ satisfying (8), where n is large, and with flexible restrictions on the cross-sectional dependence and heteroscedasticity of u_{it}. We propose an approach based on debiasing and sample-splitting here, but considering alternative approaches would be interesting for future work.

[38] While our work is related to previous work on high-dimensional models, it is interesting to note that the combination of (i) a panel data setting, (ii) a tuning parameter-free estimator, and (iii) considering aggregate effects as object of interest, leads to significant differences compared to existing asymptotic results on LASSO estimation. For instance, in our setting, post-selection OLS estimation would lead to an inconsistent estimator, while in Belloni and Chernozhukov (2013), Belloni et al. (2012), or Manresa (2015), post-selection OLS estimation yields the same rate of convergence as LASSO estimation. We establish this result formally in the online appendix (section B.1).

parameter of interest, Δ, in terms of its deviation from the estimator $\tilde{\Delta}$:

$$\Delta = \tilde{\Delta} + \sum_{i \in \mathcal{N}} (\omega_i - \tilde{\Delta}) b_i. \tag{10}$$

We will then base inference on a new estimator that will combine $\tilde{\Delta}$ with an estimator for the second term in (10) based on sample-splitting.[39] Partition the set of observations $\mathcal{T}$ into subsets $\mathcal{T}_1$ and $\mathcal{T}_2$ of approximately the same size. The debiased estimator is:

$$\breve{\Delta} = \tilde{\Delta} + \sum_{i \in \hat{\mathcal{A}}_{1,n}} (\omega_i - \tilde{\Delta}) \hat{b}_{2,i},$$

where $\hat{\mathcal{A}}_{1,n}$ is the selected set of cross-sectional units obtained from applying the constrained estimator (6) to a subsample of the data determined by the time periods $t \in \mathcal{T}_1$, and $\hat{b}_{2,i} = \frac{\sum_{t \in \mathcal{T}_2} x_t y_{it}}{\sum_{t \in \mathcal{T}_2} x_t^2}$ is the simple OLS regression estimator applied to the rest of the data, i.e. observations with $t \in \mathcal{T}_2$.[40]

Intuitively, sample-splitting yields asymptotically valid inference because it avoids selection bias, since selection and estimation are based on separate, independent, sets of observations. To understand this result more precisely, we can use (10) to write:

$$\breve{\Delta} - \Delta = \sum_{i \in \mathcal{N}} (\omega_i - \tilde{\Delta})(1[i \in \hat{\mathcal{A}}_{1,n}] \hat{b}_{2,i} - b_i). \tag{11}$$

In the proof of our next result, we show that the estimation error defined by (11) is asymptotically equivalent to $\sum_{i \in \mathcal{N}} (\omega_i - \Delta)(1[i \in \hat{\mathcal{A}}_{1,n}] \hat{b}_{2,i} - b_i)$ (i.e., the estimation error in $\tilde{\Delta}$ for Δ is asymptotically ignorable when interacted with the estimation error in $1[i \in \hat{\mathcal{A}}_{1,n}] \hat{b}_{2,i}$ for b_i). For simplicity, consider here the case where we have exact sparsity, so that $b_i = 0$ if $i \notin \mathcal{A}_n$. From Proposition 2 above, we know that model selection performed by $\hat{\mathcal{A}}_{1,n}$ is consistent, so that $P(\mathcal{A}_n \subseteq \hat{\mathcal{A}}_{1,n}) \to 1$ under the conditions imposed in Proposition 2, which are maintained here. This will allow us to write:

$$\sum_{i \in \mathcal{N}} (\omega_i - \Delta)(1[i \in \hat{\mathcal{A}}_{1,n}] \hat{b}_{2,i} - b_i) \simeq \sum_{i \in \mathcal{A}_n} (\omega_i - \Delta)(\hat{b}_{2,i} - b_i) + \sum_{i \notin \mathcal{A}_n} (\omega_i - \Delta) 1[i \in \hat{\mathcal{A}}_{1,n}] \hat{b}_{2,i}, \tag{12}$$

where $\simeq$ denotes asymptotic equivalence.

The decomposition in (12) allows us to understand the gains from using sample-splitting

[39] Sample splitting is a natural approach to inference in the presence of model selection, and dates back to early work by Barnard (1974), Cox (1975), Faraway (1995), Hartigan (1969), Miller (2002), Moran (1973), Mosteller and Tukey (1977), and Picard and Berk (1990), and is recently reviewed in Rinaldo, Wasserman, and G'Sell (2019).

[40] It might be possible to also use the OLS estimates from $\mathcal{T}_1$ and the selected set from $\mathcal{T}_2$ to obtain a more precise bias correction and efficiency gains. We do not pursue this here for simplicity and concision.

for inference. Sample splitting helps with avoiding selection bias originating from the second term on the right-hand side of (12). Because of sample splitting and serial independence, an observation $i \notin \mathcal{A}_n$ being selected by $\hat{\mathcal{A}}_{1,n}$ ($1[i \in \hat{\mathcal{A}}_{1,n}]$) is independent of the OLS estimates obtained using the second subsample, $\hat{b}_{2,i}$. This leads to mean-zero summands, $E(1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i}) = 0 \; \forall\, i \notin \mathcal{A}_n$. When using the same sample for model selection and estimation we would obtain a positive selection bias that would be strong enough to rule out consistency, as discussed in footnote 38 and section B.1 of the online appendix.[41]

Given mean-zero summands, we can apply a cross-sectional central limit theorem if we impose a restriction on the strength of the cross-sectional dependence, which is done in our next assumption.

Assumption 2. Assume that Assumption 1 holds, with the full rank restriction applying to each sub-sample: $\frac{1}{T}\sum_{t\in\mathcal{T}_1} x_t^2, \frac{1}{T}\sum_{t\in\mathcal{T}_2} x_t^2 \geq c$ for some positive constant $c > 0$, and that:

1. (dependence neighborhoods) $u_{it} \perp u_{jt}$ if $j \notin \mathcal{N}_i$, where $\mathcal{N}_i \subseteq \mathcal{N}$ is such that $|\mathcal{N}_i| \leq C$ $\forall\, i \in \mathcal{N}$, $\forall\, n$, for a finite constant $C < \infty$.

Define $\mathcal{D}_{1,N} = \{y_{it} : (i,t) \in \mathcal{N} \times \mathcal{T}_1\}$ to be the first subsample of data, and $\mathcal{D}_1 = \{\mathcal{D}_{1,N} : N \in \mathbb{N}^2\}$ to be its entire sequence. Define $r_N = \sum_{i\notin\mathcal{A}_n} 1[i \in \hat{\mathcal{A}}_{1,n}]$ and $\Sigma_N = \mathrm{Var}(\sum_{i\in\mathcal{A}_n}(\omega_i - \Delta)\sqrt{T}e_{2,iT} + \sum_{i\notin\mathcal{A}_n}(\omega_i - \Delta)1[i \in \hat{\mathcal{A}}_{1,n}]\sqrt{T}e_{2,iT} \,|\, \mathcal{D}_1)$, where $e_{2,iT} = \hat{b}_{2,i} - b_i$. Assume that:

2. (non-degeneracy of sums) $\frac{\Sigma_N}{r_N + |\mathcal{A}_n|} \geq c \; \forall\, n, T \geq C$, for positive constants C and $c > 0$.

Part 1 of Assumption 2 imposes that cross-sectional dependence be limited to dependence neighborhoods of bounded cardinality. Note that the indices in these neighborhoods do not have to be known to the researcher. It is likely that this assumption could be relaxed to other forms of weak cross-sectional dependence (e.g., mixing or near-epoch dependence), but at the expense of more cumbersome technical details and notation. Part 2 of Assumption 2 is standard in results establishing asymptotic normality in the presence of cross-sectional dependence. It requires that the variance of the influence function $\sum_{i\notin\mathcal{A}_n}(\omega_i - \Delta)1[i \in \hat{\mathcal{A}}_{1,n}]\sqrt{T}e_{2,iT} + \sum_{i\in\mathcal{A}_n}(\omega_i - \Delta)\sqrt{T}e_{2,iT}$ not be of a vanishing rate compared to the number of observations that enter this influence function, $r_N + |\mathcal{A}_n|$. It rules out degeneracy scenarios

[41] The summands in the first term on the right-hand side of (12) have mean zero since $E(\hat{b}_{2,i}) = b_i$ under the regression model (4). This result would have been obtained without sample-splitting as well.

such as, for instance, clusters of two observations that cancel each other out, $u_{it} = -u_{i't}$ when i and i' belong to the same cluster.[42]

Proposition 3. Under Assumption 2, the constrained regression model (4), (5), and approximate sparsity (8), if $|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}} \to 0$, $n \cdot \exp(-c_1 \cdot n) = o(\sqrt{\frac{|\mathcal{A}_n|}{T}})$ where c_1 is the constant used in (8), and $\frac{\log^7(n)}{T} \to 0$, then as $n, T \to \infty$, if $r_N + |\mathcal{A}_n| \to \infty$, we have:

$$\frac{1}{\sqrt{\Sigma_N}}\sqrt{T}(\breve{\Delta} - \Delta) \xrightarrow{d} N(0, 1).$$

Given Proposition 3, asymptotically valid inference can proceed if a researcher can construct consistent standard errors. A natural approach to standard errors here are Driscoll-Kraay standard errors (Driscoll and Kraay (1998)) that do not rely on knowledge by the researcher of the structure of the dependence neighborhoods discussed in Assumption 2:

$$\hat{\Sigma}_N = \frac{1}{s^2_{2,X,T}}\frac{1}{T}\sum_{t\in\mathcal{T}_2}\Big(\sum_{i\in\hat{\mathcal{A}}_{1,n}}(\omega_i - \tilde{\Delta})\hat{v}_{it}\Big)^2,$$

where $\hat{v}_{it} = x_t(y_{it} - \hat{b}_{2,i}x_t)$ and $s_{2,X,T} = \frac{1}{T}\sum_{t\in\mathcal{T}_2} x_t^2$.

Proposition 4. Under Assumption 2, the constrained regression model (4), (5), and approximate sparsity (8), if $|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}} \to 0$, $n \cdot \exp(-c_1 \cdot n) = o(\sqrt{\frac{|\mathcal{A}_n|}{T}})$ where c_1 is the constant used in (8), and $\frac{\log^7(n)}{T} \to 0$, then as $n, T \to \infty$, we have:

$$\frac{\hat{\Sigma}_N}{\Sigma_N} \xrightarrow{p} 1.$$

3.5 Extensions

The simplified setting used above allowed us to expose the main ideas in this paper: (i) how constraining a regression estimator to the standard simplex can solve the issue of high-dimensional aggregation, and (ii) how asymptotically valid inference can be conducted. These results can be extended in multiple ways. The most significant extension is the case where multiple consumption regions are included in the model, each with regression coefficients that are restricted to the standard simplex. While this leads to an estimator with an algebraic

[42] Note that the variance of the influence function, Σ_N, is defined conditionally on the first-subsample. Since the two sub-samples are independent, this is not restrictive. We apply this conditioning because it allows to account in a simple way for the dependence in selection outcomes $1[i \in \hat{\mathcal{A}}_{1,n}] = 1[\hat{b}_{1,i} \geq \gamma_{1,N}]$ across $i \in \mathcal{N}$ due to the comparison with a shared value, $\gamma_{1,N}$, where $\hat{b}_{1,i}$ is the OLS regression estimator from the first-subsample and $\gamma_{1,N}$ is the Lagrange multiplier on the summing-to-one constraint in the first sub-sample.

structure that is significantly different from the simple univariate estimator discussed above, we show that our results on consistency for the univariate estimator extend to the multivariate estimator. After proving this result, we briefly discuss several other extensions.

3.5.1 Extension to a multivariate setting

Consider a multivariate model:

$$y_{it} = x_t b_i + u_{it}, \qquad E(u_{it}) = 0, \tag{13}$$

$$b_{i,k} \geq 0, \qquad \sum_{i \in \mathcal{N}} b_{i,k} = 1, \ \forall\, k \in \mathcal{K}, \tag{14}$$

where $x_t = [x_{t,1}, ..., x_{t,K}]$ collects the electricity consumption across all regions $k \in \mathcal{K}$ and is treated as deterministic for notational simplicity, and $b_i = [b_{i,1}, ..., b_{i,K}]'$ collects marginal generation from changes in consumption in each region, with each set of coefficients $\{b_{i,k} : i \in \mathcal{N}\}$ satisfying the positivity and summing-to-one constraints discussed above. The marginal emissions of each region are then given by $\Delta_k = \sum_{i \in \mathcal{N}} \omega_i b_{i,k}$, with ω_i treated as known, as before.

The assumption of approximate sparsity is extended to:

$$b_{i,k} \leq \exp(-c_1 \cdot n) \ \forall\, i \notin \mathcal{A}_{n,k}, \qquad b_{i,k} \geq \frac{c_2}{|\mathcal{A}_{n,k}|} \ \forall\, i \in \mathcal{A}_{n,k}, \ \forall\, k \in \mathcal{K}, \tag{15}$$

where c_1 and c_2 are strictly positive constants, and $\mathcal{A}_{n,k}$ denotes a region-specific response set.

A natural estimator extends the univariate estimator discussed in the previous section by minimizing the least-squares objective function subject to the constraints for each region:

$$\{\tilde{b}_i : i \in \mathcal{N}\} = \arg\min_{\{\beta_i \in \mathbb{R}^K : \beta_i \geq 0, \sum_{i \in \mathcal{N}} \beta_i = \mathbf{1}_K\}} \frac{1}{T} \sum_{(i,t) \in \mathcal{N} \times \mathcal{T}} (y_{it} - x_t \beta_i)^2. \tag{16}$$

Given these estimates of the marginal generation in plant i from changes in consumption across each region, one can estimate marginal emissions with $\tilde{\Delta}_k = \sum_{i \in \mathcal{N}} \omega_i \tilde{b}_{i,k}$.

In this multivariate setting, the estimator $\tilde{b}_i$ does not take the simple thresholding form of the univariate estimator discussed above. Instead, for each power plant $i \in \mathcal{N}$, a subset of loads $x_{t,k}$, $k \in \mathcal{K}$ is selected based on comparing the OLS estimated coefficients of sub-model regressions with linear transformations of the Lagrange multipliers corresponding to each summing-to-one constraint. In practice, as we will find in section 4 below, selecting a subset of covariates may be very helpful for precision when covariates exhibit a high degree of collinearity, but this more complex structure of the estimator complicates the derivation of the asymptotic properties of the estimator. Nevertheless, the same results of consistency can be derived. We first impose an assumption of serial independence and no-perfect-multicollinearity that is a straightforward extension of the assumption imposed in the previous section. Define $S_{x,T} = \frac{1}{T}\sum_{t\in\mathcal{T}} x_t' x_t$.

Assumption 3. Assume that (i) u_{it} is serially independent, x_t and u_{it} have bounded support, (ii) the range of ω_i is bounded, (iii) $\lambda_{\min}(S_{x,T}) \geq c \;\forall\, T$ for a positive constant $c > 0$, (iv) $\sigma_{it}^2 \geq c \;\forall\, i, t$ for a positive constant $c > 0$, where $\sigma_{it}^2 = \mathrm{Var}(u_{it})$.

To state our next result, we introduce straightforward generalizations of the notation developed in the previous section to the multivariate case: (i) $\hat{\mathcal{A}}_{n,k}$ denotes the "selected set" of generators for region $k \in \mathcal{K}$, i.e., the units $i \in \mathcal{N}$ for which the positivity constraint is not binding in problem (16) for region k, (ii) $\gamma_{N,k}$ denotes the Lagrange multiplier corresponding to the summing-to-one constraint in problem (16) for region k.

Proposition 5. Under Assumption 3, the constrained regression model (13), (14), and approximate sparsity (15), if, $\forall\, k \in \mathcal{K}$, $|\mathcal{A}_{n,k}|\sqrt{\frac{\log(n)}{T}} \to 0$, $n \cdot \exp(-c_1 \cdot n) = o(|\mathcal{A}_{n,k}|\sqrt{\frac{\log(n)}{T}})$ where c_1 is the constant used in (15), and $\frac{\log^7(n)}{T} \to 0$, then as $n, T \to \infty$ we have: (i) (consistent selection) $P(\mathcal{A}_{n,k} \subseteq \hat{\mathcal{A}}_{n,k}) \to 1$, (ii) (rate of penalization) $\gamma_{N,k} = O_p(\sqrt{\frac{\log(n)}{T}})$, (iii) (consistent point estimation) $\tilde{\Delta}_k - \Delta_k = O_p(|\mathcal{A}_{n,k}|\sqrt{\frac{\log(n)}{T}})$.

To prove Proposition 5, we study the structure of a particular numerical method for solving the constrained estimation problem (16), namely the gradient projection algorithm of Stoer (1971), who proved its numerical convergence in finitely many steps in a setting that nests ours. The structure of this algorithm, with a specific choice of starting values, allows us to establish consistency of model selection, from which the rest of the results in Proposition 5 follows.

Given the results of consistency provided in Proposition 5, asymptotically valid inference can be conducted by applying the same debiasing based on sample-splitting as in Section 3.4. As in Section 3.3, one novel aspect of Proposition 5 compared to other work on the asymptotic properties of penalized regression is that our estimating procedure is "tuning parameter-free". We show that our method is consistent despite the fact that $\{\gamma_{N,k} : k \in \mathcal{K}\}$

is entirely determined by the data and the positivity and summing-to-one constraints, i.e., without ex-ante restrictions on its rate of convergence.[43]

3.5.2 Other extensions

Our results extend to models with control covariates that have unconstrained regression coefficients, as in our empirical application below, in a straightforward way.

In addition, while the assumption of serial independence maintained throughout this paper allows for the use of high-dimensional central limit theorems for independent data to characterize the asymptotic behavior of our constrained regression estimator, our results can be directly extended to the case of cluster sampling (with many small clusters).[44] While we believe that imposing serial independence is an effective choice for the exposition of our main results, it is likely that our results could also be extended to other forms of weak serial dependence by making use of high-dimensional central limit theorems for dependent data such as in Fang and Koike (2020) or Chang, Chen, and Wu (2023).[45]

As discussed above, our simplification of the analysis by treating the emission factors ω_i as known is motivated by the fact that, with $\tilde{b}_i$ restricted to the standard simplex, estimation error in $\tilde{\Delta}$ originating from estimation error in $\hat{\omega}_i$ vanishes asymptotically. For inference, the estimation error in $\hat{\omega}_i$ will be asymptotically ignorable, under standard regularity conditions on $\hat{\omega}_i$, when $\frac{|\mathcal{A}_n|+r_N \log(n)}{T} = o(\sqrt{\frac{|\mathcal{A}_n|+r_N}{T}})$. In addition, from a practical perspective, because emissions factors ω_i are determined by a univariate relationship (between a plant's emissions and its own generation), they are likely to be estimated much more precisely than marginal generation effects $b_{i,k}$, $k \in \mathcal{K}$, determined by a multivariate relationship (between a plant's generation and electricity consumption across multiple regions plus control covariates). However, given the linear influence function representation of $\breve{\Delta}$ in Propositions 3 and 4, it would also be straightforward to adjust the standard errors to account for estimation error in $\hat{\omega}_i$.

[43] In addition, it is interesting to note that we obtain consistent model selection without imposing an additional assumption of irrepresentability such as the conditions discussed in Zhao and Yu (2006) and subsequent/related papers. As in the univariate setting, this is due to the panel data structure of our problem, which implies sufficient sparsity in our design matrix to obtain consistent model selection without further restrictions.

[44] This scenario would occur, for instance, when observations are initially independent but many fixed-effects corresponding to small non-overlapping groups are used as control covariates, which is the case in our empirical application below. This leads to our adjusting the Driscroll-Kraay estimated variance discussed in section 3.4 to cluster in the time dimension (by weeks) in addition to clustering across cross-sectional observations.

[45] With serial dependence, the Driscroll-Kraay estimated variance discussed in section 3.4 can be modified to account for the component of the variance of our estimator that arises due to serial correlation.

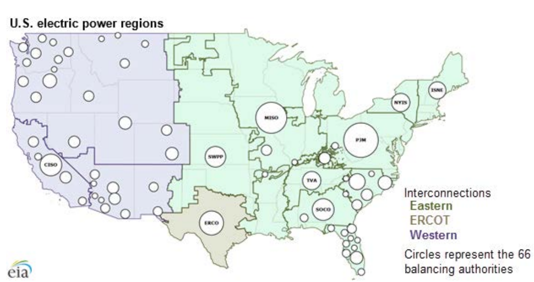
Finally, researchers with ex-ante knowledge on the structure of marginal generation can easily embed this knowledge into our estimation procedure. For instance, if researchers know that electricity generators in, say, Florida, do not respond to changes in electricity consumption in, say, New England, they can impose that $\tilde{b}_{i,\text{New England}} = 0$ for all electricity generators i located in Florida.[46]

4 Marginal Emissions across the U.S. Electricity Grid

In this section we apply our method to the U.S. electricity grid with the dual objective of evaluating the reliability of our method and of providing new results on marginal emissions in the U.S.

4.1 Data Sources

To describe the organization of our data, it is helpful to first describe the organization of the U.S. electricity grid. The grid consists of three interconnections, East, West, and Texas (see Figure 1a). Each interconnection operates as largely independent of the others, and we will treat these interconnections as independent in our analysis. Each interconnection is organized into multiple balancing authorities.[47] Each balancing authority is responsible for maintaining supply and demand balance in their section of the grid at all times (EIA (2016)).



(a) Interconnections and Balancing Authorities.

(b) Regions.

Figure 1: Administrative Organization of the U.S. Electrical Grid.

We use data from 2019-2022. As electricity demand and marginal emissions may be seasonal, we focus on the summer months (June-September). There are two main sources

[46]Such an approach can be thought of as a combination of the approach in Callaway, Fowlie, and McCormick (2018) and subsequent papers, that restrict marginal generation to only occur within the consumption region considered, and the data-driven agnostic approach to sparsity discussed above where any generator on the grid could potentially respond to changes in consumption in any region.

[47]With the exception of Texas, where a single balancing authority — the Electricity Reliability Council of Texas (ERCOT) — oversees the entire interconnection.

of data. First, we obtain data on hourly electricity consumption (load) at the region level from the EIA-930 (regions correspond to either large balancing authorities or the aggregation of multiple smaller balancing authorities, as shown in Figure 1b). Second, we obtain plant-level hourly data on generation and emissions through the Continuous Emissions Monitoring System (CEMS) program (US EPA (2016)). Small fossil fuel plants and non-emitting plants (nuclear, wind, solar, and hydro) are not covered by CEMS. For non-emitting plants, we use balancing authority-level aggregates of electricity generation by fuel type from the EIA-930 data. For small fossil fuel plants, we use balancing authority-level "residuals", the difference between balancing authority-level electricity generation by fuel type (coal, natural gas, other) and the electricity generation obtained from CEMS data when aggregated at the balancing authority level. Our analysis treats these balancing authority-level aggregates as "composite plants". While the CEMS data provides information at the unit-level within each plant, we aggregate this information at the plant-by-fuel level to avoid situations where a particular unit within a plant may be turned off for a substantial amount of time. This leads to over 1,100 plant-by-fuel combinations in the eastern interconnection, 360 in the western interconnection, and 100 in Texas.

We consider three pollutants, carbon dioxide (CO_2), sulfur dioxide (SO_2), and nitrogen oxides (NO_X). CO_2 is a global pollutant, with a homogenous social cost, while SO_2 and NO_X are local pollutants, for which environmental damages depend on local conditions, particularly local population density. To evaluate the social cost of local pollutants, we use the AP3 model of environmental damages (Clay et al. (2019)). A final source of data enables us to convert plant-level gross generation to net generation.[48]

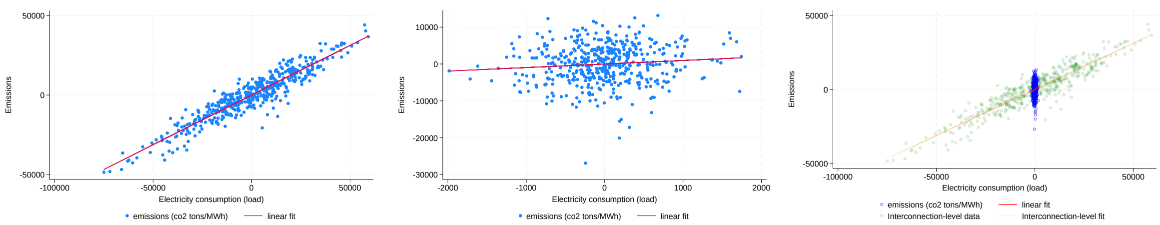
4.2 Marginal emissions with aggregate data

To illustrate the problem of high-dimensional aggregation with this dataset, we first simply regress interconnection-level emissions on electricity consumption, similarly to what is used in many previous studies, such as Graff Zivin, Kotchen, and Mansur (2014) and subsequent work.

[48]CEMS includes hourly information on gross generation, the total amount of electricity generated by each plant, while the relevant quantity for our model below is net generation, the total amount of electricity each plant injects into the electricity grid (i.e., the difference between total electricity generated and the amount of electricity used to run the plant). The EIA reports monthly data on gross and net generation for each plant in its EIA-923 data. We calculate net generation factors that allow us to convert gross generation to net generation by calculating the ratio of total net generation over total gross generation for each plant-by-fuel combination over our entire period of observation.

For concision we concentrate on the eastern interconnection. We obtain interconnection-level marginal emissions by regressing total emissions across the interconnection (we use hourly CO_2 emissions here) on load for this interconnection and year-by-week fixed effects. We obtain region-level marginal emissions by a regression of total emissions across the interconnection on load for all nine regions that partition the interconnection and year-by-week fixed effects.

In Figure 2, we show the scatter plot corresponding to these regressions for UTC hour 23,[49] plotting the partial residual of interconnection-level emissions over (i) the partial residual of interconnection-level load, or (ii) the partial residual of New York load.[50] Estimated marginal emissions are then simply obtained as the slope of each linear fit.



(a) Eastern interconnection. (b) New York. (c) Overlaid for comparison.

Figure 2: Estimating marginal emissions with aggregated data. Scatter plots (UTC hour 23) of partial residuals of emissions across the entire eastern interconnection over partial residuals for the load of (i) the entire eastern interconnection, (ii) New York only. Estimated marginal emissions are given by the slopes of the linear fit.

We see that the results obtained for the eastern interconnection are very precise (0.62 metric tons of CO_2/MWh, with standard errors clustered by week of 0.02). These results, however, do not allow us to uncover heterogeneity in marginal emissions within the interconnection that may be central to policy making (e.g., evaluating a policy with different regional effects, or selecting an optimal location for increases/decreases in net electricity consumption).

Unfortunately, when estimating region-level marginal emissions, the method used in this section yields clearly unreliable results: Estimated marginal emissions for New York are 0.96 metric tons of CO_2/MWh, an implausibly high value, with standard errors of 0.45, denoting that using the data in this way does not provide meaningful information. Intuitively, this is because a region-level regression reduces "effective sample size" (i.e., the identifying variation

[49] UTC hour 23 is 6-7pm in local time across the eastern interconnection, depending on the time zone.

[50] We only show results for New York here for concision, the same issue applies to all regions in the interconnection, as seen below.

in the partial residual of load) compared to an interconnection-level regression, so that the high-dimensional aggregation problem discussed above becomes practically relevant.

As discussed above, our proposed method will help with this issue because it naturally implements a regularizing model selection. For each plant, a subset of load regions will be selected, which will alleviate the high-dimensional aggregation problem because: (i) the marginal emissions corresponding to each load region will be obtained by summing over fewer power plants (section 3.3 above), (ii) for any power plant, model selection reduces the issue of near-perfect multicollinearity if it selects a strict subset of region loads (section 3.5.1 above).

4.3 Marginal emissions with power plant-level data

We estimate marginal emissions by first modeling a plant's generation response to changes in electricity consumption across the grid using a linear regression model. For plant $i \in \mathcal{N}$ at time $t \in \mathcal{T}$, we assume that:

$$y_{it} = x_t b_i + z_{it} d_i + \mu_i(t) + u_{it}, \qquad E(u_{it}|X, Z) = 0, \tag{17}$$

where y_{it} denotes net generation, $x_t = [x_t^1, ..., x_t^K]$ denotes electricity consumption (loads) across all regions in plant i's interconnection (nine regions for the East, three regions for the West, and a single region for Texas), z_{it} collects wind speed and solar radiation across plant i's balancing authority,[51] and $\mu_i(t)$ are year-by-week fixed effects. We apply this model separately to each hour of the day (hour 1-hour 24) to capture variation in marginal emissions across time-of-day.

As discussed above, we will estimate marginal emissions $\Delta = \sum_{i \in \mathcal{N}} \omega_i b_i$ with a constrained regression estimator that restricts the estimated coefficients for each load region to be positive and sum to one. The emission factors ω_i are calculated for each plant as the ratio of total emissions over total generation.[52] We will compare our results with two alternative estimators

[51] The control covariates in z_{it} are defined to be weighted average of county-level wind speed and global horizon irradiation (GHI), obtained from `open-meteo.com`, with averages taken at the balancing authority level, and weights given by wind and solar electricity generation capacity in each county. We include these variables at the balancing authority-level since they are important controls for solar and wind electricity generation, which we only observe at the balancing authority level, as discussed above.

[52] As discussed in footnote 21, our approach to estimating marginal generation could be combined with any alternative approach to estimating emissions factors. Some small fossil fuel-powered generators are included in the EIA data but not in the CEMS data, as discussed above, leading to fuel by balancing authority-level "residuals". We use fuel by balancing authority-level averages in emission factors to impute the emission factor for these residuals. Finally, as discussed in section 3.5.2, estimation error in plant-specific emissions factors will be asymptotically ignorable so that we can treat these coefficients as known, while one could also modify inference to account for estimation error in ω_i if desired.

that are commonly used in the literature using regression-based methods to estimate marginal emissions: (i) an unconstrained regression estimator, (ii) a region-by-region estimator.[53]

The unconstrained OLS regression estimator simply estimates marginal generation coefficients by a regression, but without imposing our constraints that estimated marginal generation be positive and sum to one. The region-by-region estimator is a somewhat simplified version of the estimator of Callaway, Fowlie, and McCormick (2018).[54] It regresses total emissions at the region-by-hour level on total fossil fuel-generation at the region-by-hour level (and weekly fixed effects) to estimate marginal emissions. This approach relies on the assumptions that: (i) there are no imports and exports between regions, (ii) fossil fuel powered generation accounts for all marginal generation.

4.3.1 Evaluating model selection by the constrained estimator

Before presenting our results on marginal emissions, it is important to evaluate the plausibility of the model selection yielded by our constrained estimator. As discussed in Section 3, our constrained estimator can be seen as a naturally regularized estimator that, for each region, yields a selected set of power plants that are estimated to respond to the region's load, while estimated marginal generation is set to zero for the rest of the plants. While Section 3 provides promising theoretical results, in this section we evaluate our method from an empirical standpoint by inspecting two main patterns in model selection: (i) what are the fuel types estimated to respond to loads, (ii) what is the geographical location of marginal generation.

For the first criterion to evaluate model selection, there is a strong prior that fossil fuels, mainly coal and natural gas, account for the wide majority of marginal generation. In this paper, we do not impose any restriction on fuel types for marginal generation. Doing so provides us with an informal way to "test" or "check" the quality of model selection obtained by our method. Figure 3 shows the estimated share of marginal generation using our method by hour of the day and for each region. We see that natural gas and coal account for a

[53]In ongoing work, we also compare our regression-based method to estimating marginal emissions with engineering dispatch models.

[54]Callaway, Fowlie, and McCormick (2018) introduce heterogeneity in marginal emissions across days in an interesting way, with an interactive algorithm that clusters days for each region by their hourly load profile, only selecting groupings that have a significant effect on the estimated marginal emissions. Since there is no formal theory for this approach, we use a regression model that estimates constant marginal emissions for each region here, but it would be interesting to consider such heterogeneity in future work.

wide majority of marginal generation. When we do estimate non-fossil fuel energy sources to contribute to marginal generation, our results are plausible.[55] This is a good sign for the performance of our estimator and the corresponding model selection. If model selection was determined by pure noise, we would not expect the plausible systematic patterns that emerge in our results.

For the second criterion to evaluate model selection, Figure 3 plots the geographic location of marginal generation for each region at UTC hour 23 (which corresponds to 4pm-7pm across the U.S. time zones).[56] We see that there is a systematic pattern where the location of marginal generation follows the location of each region (while not being fully contained within the region). Again, if model selection were determined by unreliable noise, we would not expect such predictable patterns, and this provides us with another good sign about the quality of the model selection obtained by our estimator.

4.3.2 Comparison of estimated marginal emissions

Given the results above on the plausibility of model selection with our estimator, we can proceed with the estimation of marginal emissions. Figure 4 depicts estimated marginal emissions in metric tons of CO_2/MWh for each region and hour, with accompanying pointwise confidence intervals.[57] Figure 4 also shows estimated marginal emissions using an unconstrained regression estimator and using a region-by-region analysis.

Similar to what was found when using a regression with aggregate data in section 4.2 above, we see that the unconstrained regression estimator exhibits excessive noise, leading to uninformative and unreliable results. This further illustrates the high-dimensional aggregation problem discussed in section 2. Figure 5 provides estimated marginal generation by fuel type and geographic location using the unconstrained estimator (only for New York). Compared to Figure 3, we see that the unconstrained estimator assigns (positive or negative)

[55]For instance, hydropower is estimated to be "on the margin" for some regions with large capacities of hydropower generation, and generally during peak times when fossil fuel generation may be expensive and it may be advantageous to respond to changes in electricity consumption with hydropower generation. Wind power is estimated to contribute to marginal generation in some regions that have high capacities of wind power generation (e.g., Central, Southwest and Texas), and mainly at low usage times (nighttime), which may be attributed to the use of curtailment of wind power generation at particular locations and times.

[56]We choose to only display results for UTC hour 23 for concision and because it is a relatively high load period.

[57]Confidence intervals are obtained as described in section 3.4. We use sample-splitting to debias our estimator, which is why our confidence intervals may not be centered around the point estimate. We use standard errors clustered at the weekly level for all estimators.

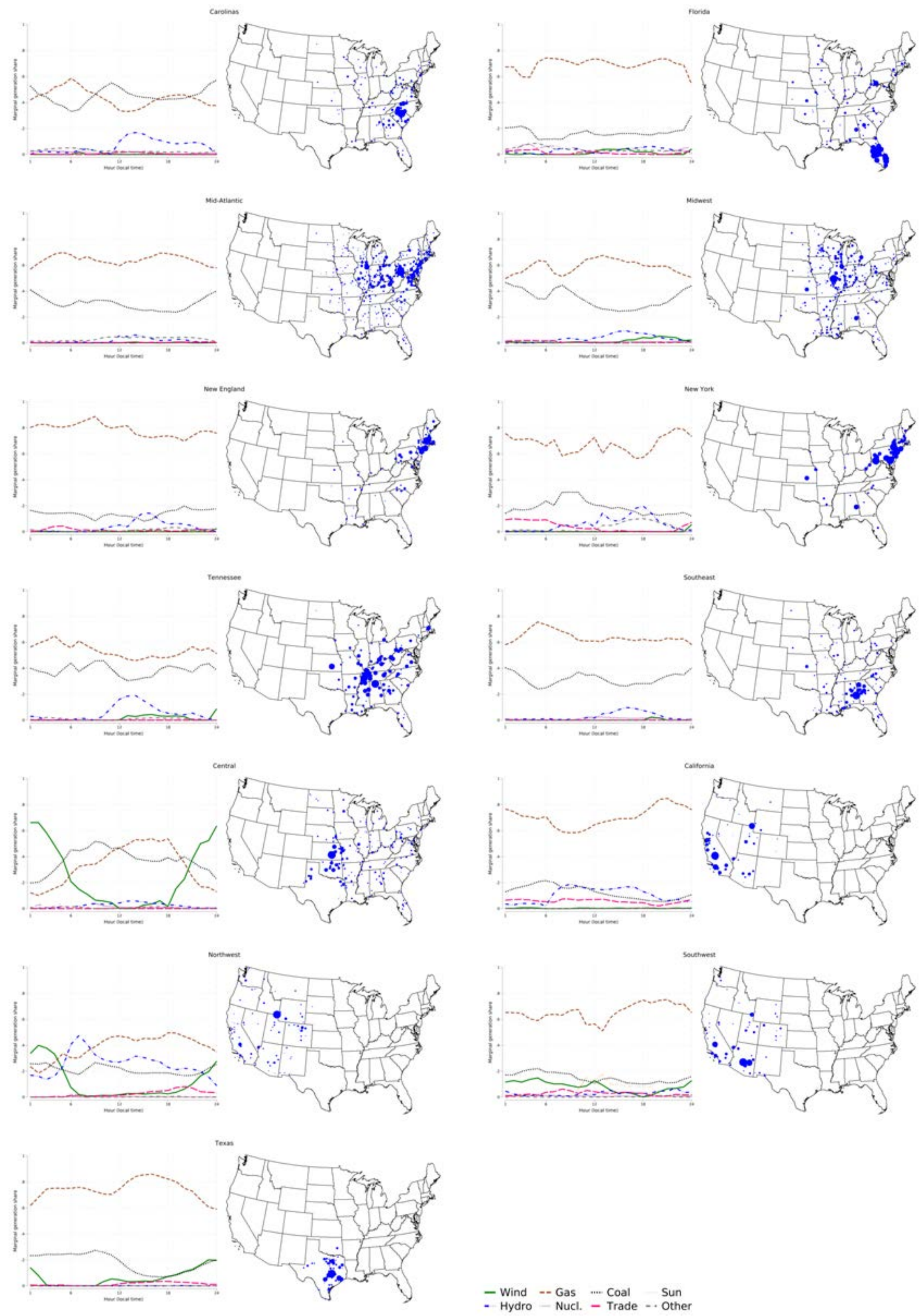


Figure 3: Estimated marginal generation shares by fuel type and, for UTC hour 23, geographic location of marginal generation. Results obtained using our constrained estimator. On the maps, county-level aggregates are represented. For balancing authority-level generation (non-fossil fuel and small fossil fuel generation), centroids are used.

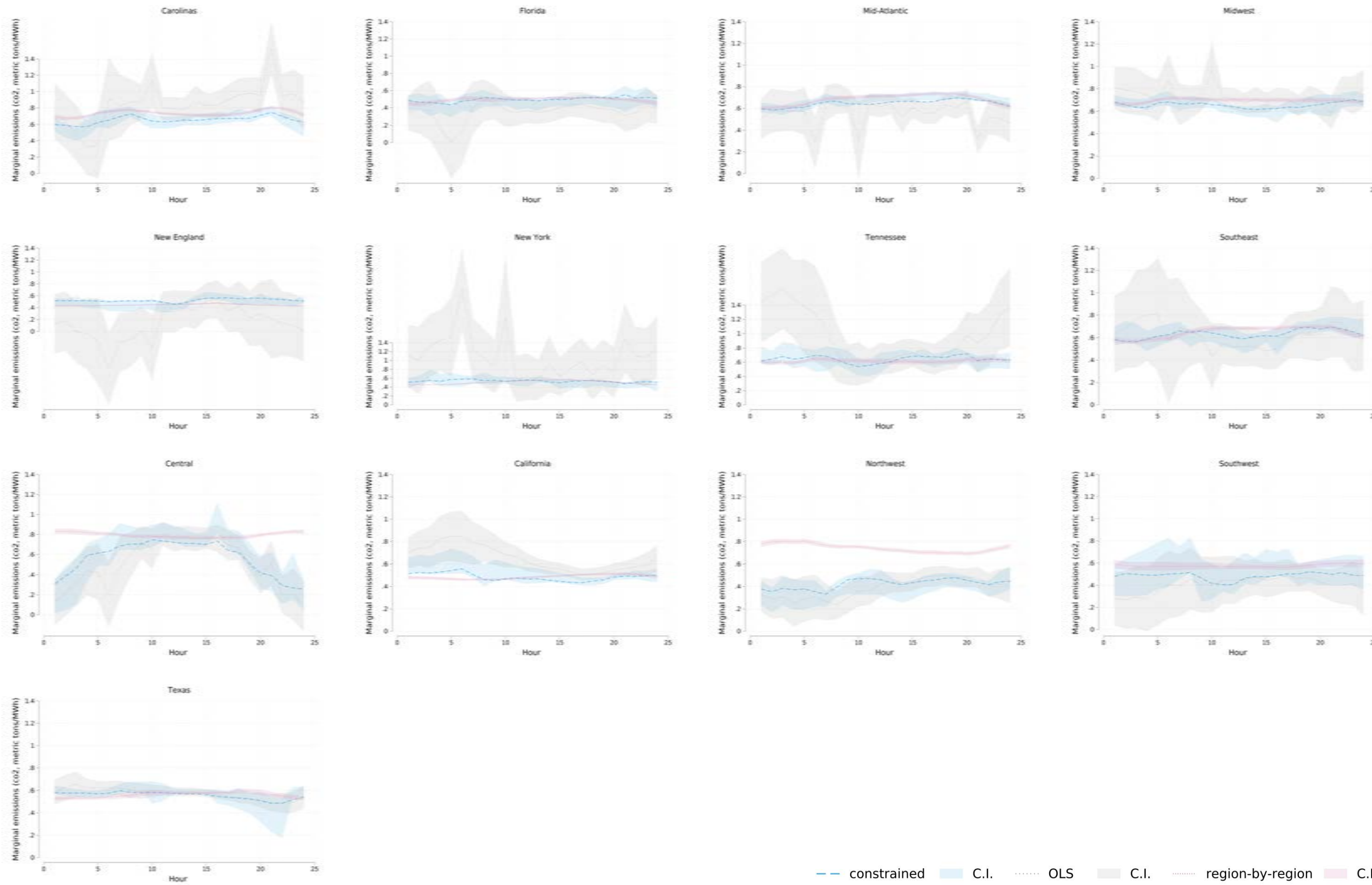


Figure 4: Estimated marginal emissions (metric tons of co2 / MWh) by region and hour (local time) with pointwise confidence intervals for each hour. Constrained is our proposed estimator, OLS aggregates the unconstrained estimates without regularization, and a region-by-region estimator assumes no imports/exports between regions and no non-fossil fuel marginal generation.

estimated marginal generation to a large number of electricity generators, which leads to wide fluctuations in the estimated geographic location of marginal generation or in the estimated fuel types of marginal generation.

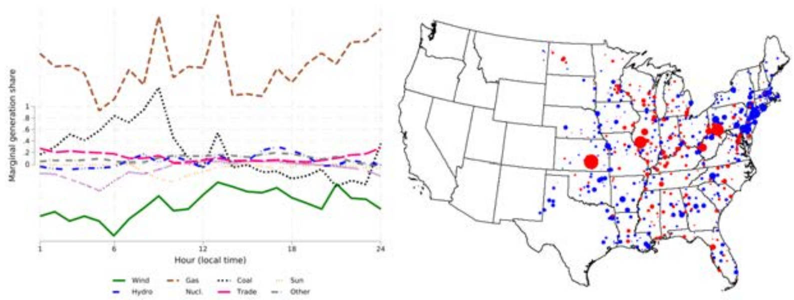
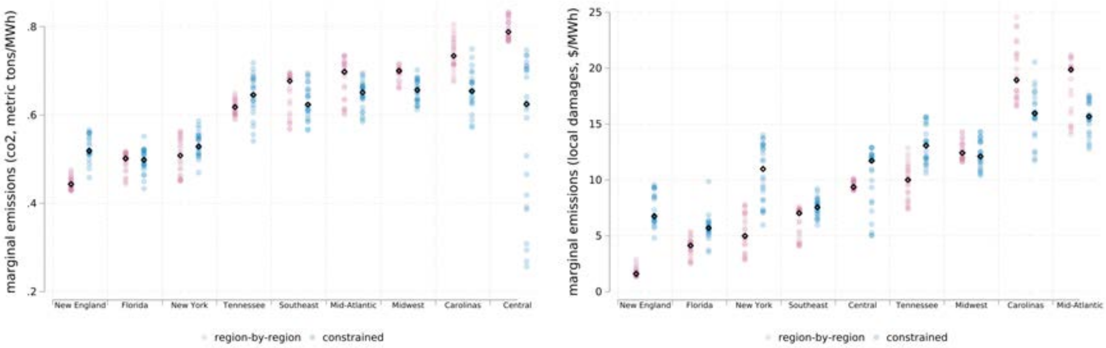


Figure 5: Estimated marginal generation shares by fuel types and geographic location of estimated marginal generation when using OLS. On the map, estimated marginal generation is aggregated at the county-level, negative values are displayed with red markers, and positive values are displayed with blue markers. Displayed for New York only for concision.

Figure 4 is dominated by the the excessive noise and wide fluctuation present in the unconstrained regression estimator. To better compare the results obtained using our estimator with results obtained using a region-by-region analysis, Figure 6 plots estimated marginal emissions in metric tons of CO_2/MWh and in local damages (the monetary damages from emissions of SO_2 and NO_X, calculated using the AP3 model) obtained using both methods. In this figure we see a general pattern whereby a region-by-region analysis tends to attribute too much heterogeneity in marginal emissions to "extreme" regions in terms of emissions, for instance New England on the low emissions side and the Carolinas on the high emissions side. This is because a region-by-region analysis uses the composition of a region's fleet of electricity generators to determine its marginal emissions. New England, for instance, has approximately no coal electricity generators. A region-by-region analysis therefore concludes that New England marginal emissions are very low, since natural gas electricity generators have lower emission factors than coal electricity generators. However, if New England gets a share of its marginal generation from other regions that do have coal electricity generators, its actual marginal emissions may be greater than the emissions coming from electricity generators located in New England only. The converse result would occur for regions that have a greater number of coal electricity generators than other regions, such as the Carolinas.



(a) CO_2.

(b) Local damages.

Figure 6: Estimated marginal emissions across all hours, by region, using our constrained estimator and a region-by-region estimator that assumes no imports/exports between regions and no non-fossil fuel marginal generation. Shown for regions in the eastern interconnection only (Carolinas, Central, Florida, Mid-Atlantic, Midwest, New England, New York, Southeast, Tennessee). Diamonds represent medians for each estimator/region.

To further quantify these differences, consider charging an electric car at 9pm to drive 100 miles. In New England, our proposed method estimates that this would result in 15.3kg of CO_2 emissions and local damages of $0.26 whereas a region-by-region analysis yields substantially lower estimates (12.3kg and $0.05).[58] Conversely, our estimates for charging in Central yield 11.0kg of CO_2 emissions and $0.20 of local damages, whereas a region-by-region analysis yields much larger estimates (22.7kg and $0.27). While the differences for New England can be attributed to the region-by-region analysis ignoring imports of electricity from higher emissions power plants outside of New England, as discussed above, the differences for Central can be attributed to the region-by-region analysis assuming that all marginal power originates from fossil fuel. Our method finds that, during nighttime hours, wind is a marginal source of power for the Central region.

In addition to verifying the reliability of our approach, our results are of independent interest as they are the most up-to-date estimates available that can be used to evaluate the spatially differentiated effects of environmental policies. Our point estimates (with bias correction and standard errors) for CO_2 emissions and local damages are provided for all

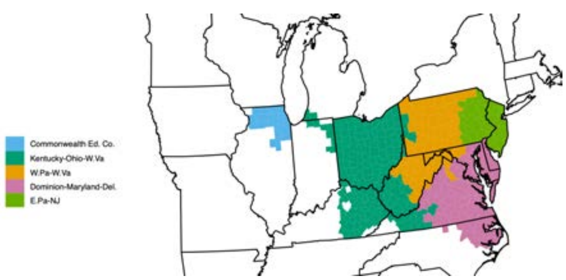
[58]These numbers are for an electric vehicle that requires 0.28kWh of electricity per mile. While the social cost of carbon varies depending on methodology, for comparison with local damages we can use a commonly used value of $50 per metric ton, which would lead to $0.76 damages from carbon emissions using our estimator, $0.61 using a region-by-region estimator. Our estimates of total damages are therefore $1.02, 54% greater than the $0.66 estimate obtained when ignoring imports.

hours and regions in Tables 1 and 2 of the online appendix (section C).

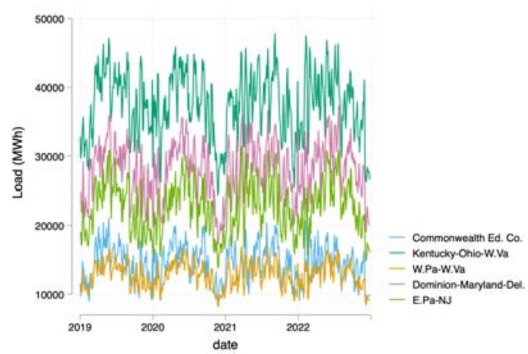
4.3.3 How local are marginal emissions? An analysis of PJM

To further explore the differences between our approach and a region-by-region analysis, we use PJM, a large balancing authority that corresponds to the Mid-Atlantic region used above, as a case study. We split PJM into five subregions depicted in Figure 7a. We then re-run our analysis of the eastern interconnection, keeping the rest of our model unchanged, but obtaining marginal emissions for each of the five PJM subregions instead of treating PJM as a whole.

Because we consider subregion-level loads here, we are drastically reducing "effective sample size", i.e., the amount of identifying variation in our covariates of interest. To compensate this somewhat, we increase sample size by using three hours of data instead of a single hour of data for each regression. For instance, results for UTC hour 23 are obtained by combining data for UTC hours 22-24. Implicitly this relies on an assumption of smoothness in marginal generation across time-of-day.[59]



(a) PJM sub-regions.



(b) Sub-region loads (UTC hour 23 only).

Figure 7: Division of PJM (Mid-Atlantic) into five sub-regions.

Figure 8 displays the fuel type and geographic location for marginal generation for each subregion estimated using our method. We see that, even though we now conduct an analysis at a much finer level of granularity, the estimated patterns of our model selection still seem plausible, with the location of marginal generation following the location of consumption,

[59] We use time of day-by-year-by-week fixed effects in the regression instead of year-by-week fixed effects in the previous results, so that the assumption of smoothness is only on marginal generation, not on the variation in generation captured by the fixed effects.

and natural gas and coal accounting for the wide majority of marginal generation. Figure 9 plots our estimated marginal emissions for each subregion compared to the results obtained with a region-by-region analysis, which treats each sub-region as being independent of the others. We see that our estimator now displays a greater level of estimation noise, particularly for the smaller regions in terms of loads (the Edison Commonwealth Company and West Virginia/Western Pennsylvania subregions, see Figure 7b). This is expected since we are drastically reducing the amount of identifying variation (i.e., effective sample size) in our regression model, and future work could explore additional approaches to further reduce estimation noise within our framework (e.g., impose ex-ante restrictions, as discussed in section 3.5.2).[60]

Figure 8: Estimated marginal generation shares by fuel type and geographic location of marginal generation for PJM subregions. Results obtained using our constrained estimator. Maps are for UTC hour 23, with the magnitude of the markers determined by marginal generation shares aggregated at the county level. For balancing authority-level generation (non-fossil fuel and small fossil fuel generation), centroids are used.

Nevertheless, our method yields surprisingly precise estimates given the disaggregation of loads, and Figure 10 displays estimated marginal emissions for each sub-region compared to the results obtained by a region-by-region analysis. We see that, again, a region-by-region

[60] One could also explore analyses where subregion-level results are obtained across the entire U.S. electricity grid.

analysis applied to each subregion separately would overstate the degree of heterogeneity in marginal emissions that exists between sub-regions. On the other hand, a region-by-region analysis applied to the Mid-Atlantic region as a whole would miss some of the heterogeneity in marginal emissions that we estimate to exist within the region. Both biases could be important for policy making.

5 Conclusion

In this paper we proposed a new method for estimating marginal emissions using a regression model combined with what we call a natural regularization, using economic constraints to implement a penalization that does not need to be directly tuned. Beyond its theoretical promise, we find that our method yields new empirical insights. For instance, we find sizable biases in marginal emissions when using a commonly used region-by-region analysis that rules out marginal generation from electricity imports and exports between regions or from non-fossil fuel generation. These biases can go in either direction depending on the region.

The central idea in our method can be combined with other extensions for estimating marginal emissions. Some are discussed throughout the paper (e.g., using non-linear models for plant-level emissions factors or combining our data-based regularization with ex-ante restrictions), but providing an accurate and flexible new way of identifying marginal generation may allow for new angles of investigation. Finally, our methodological results may also be of interest in other empirical applications (e.g., estimating the aggregate effect of a policy with spatially heterogeneous effects).

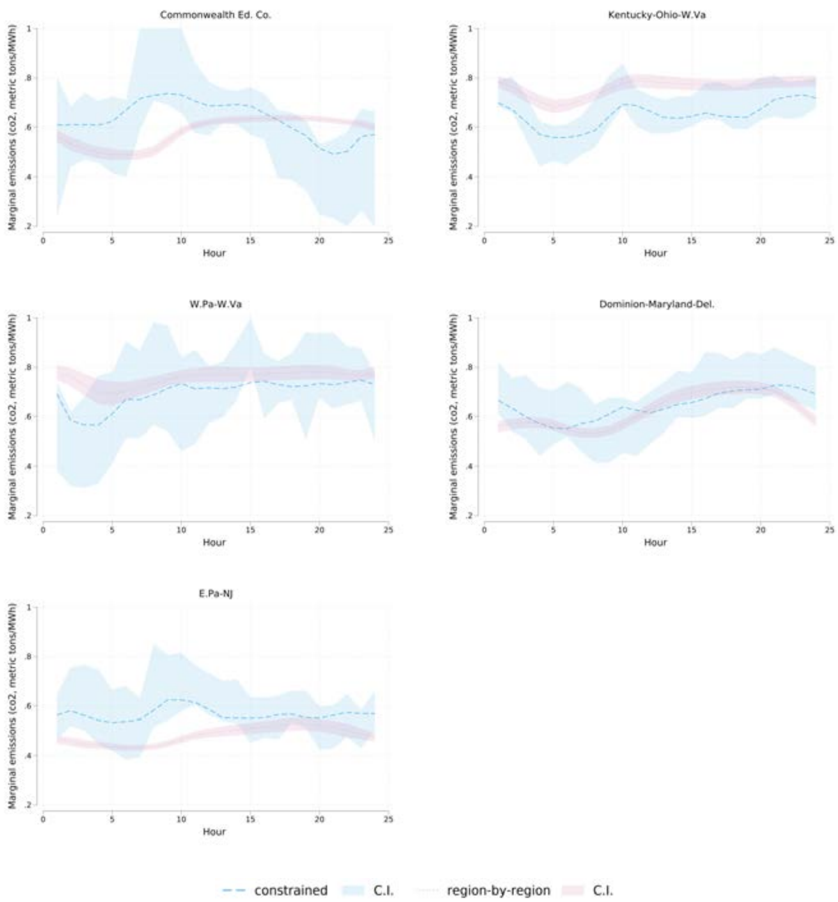
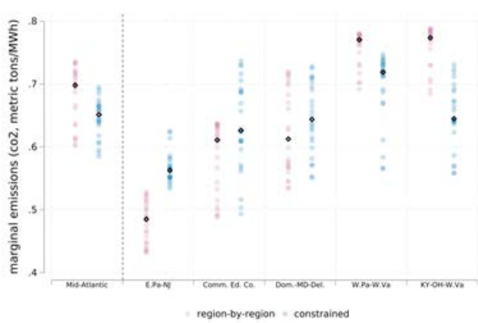
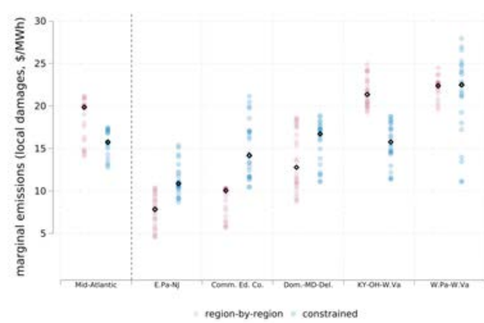


Figure 9: Estimated marginal emissions for PJM subregions, local hour, with our constrained estimator and an estimator that assumes no imports/exports.



(a) CO_2.



(b) Local damages.

Figure 10: Estimated marginal emissions across all hours for subregions of PJM (Mid-Atlantic). Results for Mid-Atlantic included for comparison, previously reported in figure 6.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490):493–505.

Andrews, Donald W. K. 1999. "Estimation When a Parameter Is on a Boundary." *Econometrica* 67 (6):1341–1383.

Andrews, Isaiah, Adam McCloskey, and Toru Kitagawa. 2022. "Inference on Winners." *Working Paper* .

Archsmith, James. 2022. "Adjustment on the Margin: Evaluating Emissions Reduction Policies in the Face of Short-Run Adjustment Costs." *Working Paper* .

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review* 111 (12):4088–4118.

Armstrong, Timothy B. and Michal Kolesár. 2018. "Optimal Inference in a Class of Regression Models." *Econometrica* 86 (2):655–683.

Armstrong, Timothy B., Michal Kolesár, and Soonwoo Kwon. 2023. "Bias-Aware Inference in Regularized Regression Models."

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2021. "Matrix Completion Methods for Causal Panel Data Models." *Journal of the American Statistical Association* 116 (536):1716–1730.

Bachoc, François, David Preinerstorfer, and Lukas Steinberger. 2020. "Uniformly Valid Confidence Intervals Post-Model-Selection." *The Annals of Statistics* 48 (1):440–463.

Barnard, G. 1974. "Discussion of "Cross-Validatory Choice and Assessment of Statistical Predictions" by Stone." *Journal of the Royal Statistical Society. Series B (Methodological)* :133–135.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica* 80 (6):2369–2429.

Belloni, Alexandre and Victor Chernozhukov. 2013. "Least Squares after Model Selection in High-Dimensional Sparse Models." *Bernoulli* 19 (2):521–547.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls†." *The Review of Economic Studies* 81 (2):608–650.

Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur. 2016. "Inference in High-Dimensional Panel Models With an Application to Gun Control." *Journal of Business & Economic Statistics* 34 (4):590–605.

Bergman, Aaron, Brian C. Prest, and Karen Palmer. 2022. “How Can Hydrogen Producers Show That They Are "Clean"?” *Resources* .

Berk, Richard, Lawrence Brown, Andreas Buja, Kai Zhang, and Linda Zhao. 2013. “Valid Post-Selection Inference.” *The Annals of Statistics* 41 (2):802–837.

Bonhomme, Stéphane and Elena Manresa. 2015. “Grouped Patterns of Heterogeneity in Panel Data.” *Econometrica* 83 (3):1147–1184.

Borenstein, Severin and James B. Bushnell. 2022. “Do Two Electricity Pricing Wrongs Make a Right? Cost Recovery, Externalities, and Efficiency.” *American Economic Journal: Economic Policy* 14 (4):80–110.

Buchanan, Will, John Foxon, Daniel Cooke, Iyer Sangeeta, Elizabeth Graham, Bill DeRusha, Christian Binder, Kin Chiu, Laura Corso, Henry Richardson, Vaughan Knight, Asim Hussain, Avi Allison, and Nithin Mathews. 2023. “Carbon-Aware Computing: Measuring and Reducing the Carbon Intensity Associated with Software in Execution.” *White Paper* .

Bühlmann, Peter and Sara van de Geer. 2011. “Variable Selection with the Lasso.” In *Statistics for High-Dimensional Data: Methods, Theory and Applications*, edited by Peter Bühlmann and Sara van de Geer, Springer Series in Statistics. Berlin, Heidelberg: Springer, 183–247.

Butters, R. Andrew, Jackson Dorsey, and Gautam Gowrisankaran. 2023. “Soaking Up the Sun: Battery Investment, Renewable Energy, and Market Equilibrium.” *Working Paper* .

Cai, T. Tony and Mark G. Low. 2004. “Minimax Estimation of Linear Functionals over Nonconvex Parameter Spaces.” *The Annals of Statistics* 32 (2):552–576.

Callaway, Duncan S., Meredith Fowlie, and Gavin McCormick. 2018. “Location, Location, Location: The Variable Value of Renewable Energy and Demand-Side Efficiency Resources.” *Journal of the Association of Environmental and Resource Economists* 5 (1):39–75.

Cattaneo, Matias D., Yingjie Feng, and Rocio Titiunik. 2021. “Prediction Intervals for Synthetic Control Methods.” *Journal of the American Statistical Association* 116 (536):1865–1880.

Cattaneo, Matias D., Michael Jansson, and Whitney K. Newey. 2017. “Inference in Linear Regression Models with Many Covariates and Heteroskedasticity.” *Journal of the American Statistical Association* .

Chamberlain, Gary. 1984. “Panel Data.” In *Handbook of Econometrics*, edited by Zvi Griliches and Michael D. Intrigilator. Elsevier, Amsterdam, 1247–1313.

———. 1992. “Efficiency Bounds for Semiparametric Regression.” *Econometrica* 60 (3):567–596.

Chang, Jinyuan, Xiaohui Chen, and Mingcong Wu. 2023. “Central Limit Theorems for High Dimensional Dependent Data.”

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. 2017. “Central Limit Theorems and Bootstrap in High Dimensions.” *The Annals of Probability* 45 (4):2309–2352.

Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K. Newey, and James M. Robins. 2022. "Locally Robust Semiparametric Estimation." *Econometrica* 90 (4):1501–1535.

Chernozhukov, Victor, Christian Hansen, Yuan Liao, and Yinchu Zhu. 2019. "Inference for Heterogeneous Effects Using Low-Rank Estimation of Factor Slopes."

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2021. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." *Journal of the American Statistical Association* 116 (536):1849–1864.

Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu. 2022. "A T-Test for Synthetic Controls."

Chetverikov, Denis and Elena Manresa. 2022. "Spectral and Post-Spectral Estimators for Grouped Panel Data Models."

Clay, Karen, Akshaya Jha, Nicholas Muller, and Randall Walsh. 2019. "External Costs of Transporting Petroleum Products: Evidence from Shipments of Crude Oil from North Dakota by Pipelines and Rail." *The Energy Journal* 40 (1).

Clinton, Bentley C. and Daniel C. Steinberg. 2019. "Providing the Spark: Impact of Financial Incentives on Battery Electric Vehicle Adoption." *Journal of Environmental Economics and Management* 98:102255.

Cox, D. R. 1975. "A Note on Data-Splitting for the Evaluation of Significance Levels." *Biometrika* 62 (2):441–444.

Dauwalter, Travis E. and Robert I. Harris. 2023. "Distributional Benefits of Rooftop Solar Capacity." *Journal of the Association of Environmental and Resource Economists* 10 (2):487–523.

Davis, Rebecca J., J. Scott Holladay, and Charles Sims. 2022. "Coal-Fired Power Plant Retirements in the United States." *Environmental and Energy Policy and the Economy* 3:4–36.

de Chalendar, Jacques A., John Taggart, and Sally M. Benson. 2019. "Tracking Emissions in the US Electricity System." *Proceedings of the National Academy of Sciences* 116 (51):25497–25502.

Deetjen, Thomas A. and Inês L. Azevedo. 2019. "Reduced-Order Dispatch Model for Simulating Marginal Emissions Factors for the United States Power Sector." *Environmental Science & Technology* 53 (17):10506–10513.

Department of Energy. 2023. "Electric Vehicle Registrations by State." https://afdc.energy.gov/data/10962.

Donoho, David L. 1994. "Statistical Estimation and Optimal Recovery." *The Annals of Statistics* 22 (1):238–270.

Doudchenko, Nikolay and Guido W. Imbens. 2016. “Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis.”

Driscoll, John C. and Aart C. Kraay. 1998. “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data.” *The Review of Economics and Statistics* 80 (4):549–560.

Efron, Bradley. 2011. “Tweedie’s Formula and Selection Bias.” *Journal of the American Statistical Association* 106 (496):1602–1614.

———. 2014. “Estimation and Accuracy After Model Selection.” *Journal of the American Statistical Association* 109 (507):991–1007.

EIA. 2016. “U.S. Electric System Is Made up of Interconnections and Balancing Authorities.” https://www.eia.gov/todayinenergy/detail.php?id=27152.

———. 2022. “State Electricity Profiles.” *Report* .

———. 2023. “Electric Vehicles and Hybrids Make up 16% of U.S. Light-Duty Vehicle Sales.” https://www.eia.gov/todayinenergy/detail.php?id=60321.

Fang, Xiao and Yuta Koike. 2020. “High-Dimensional Central Limit Theorems by Stein’s Method.”

Fang, Zheng, Andres Santos, Azeem M. Shaikh, and Alexander Torgovitsky. 2022. “Inference for Large-Scale Linear Systems with Known Coefficients.” *Econometrica* Forthcoming.

Faraway, Julian J. 1995. “Data Splitting Strategies for Reducing the Effect of Model Selection on Inference.” *Working Paper* .

Farrell, Max H. 2015. “Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations.” *Journal of Econometrics* 189 (1):1–23.

Fell, Harrison and Jeremiah X. Johnson. 2021. “Regional Disparities in Emissions Reduction and Net Trade from Renewables.” *Nature Sustainability* 4 (4):358–365.

Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari. 2011. “A Simple Estimator for the Distribution of Random Coefficients.” *Quantitative Economics* 2 (3):381–418.

Gagnon, Pieter J., John E. T. Bistline, Marcus H. Alexander, and Wesley J. Cole. 2022. “Short-Run Marginal Emission Rates Omit Important Impacts of Electric-Sector Interventions.” *Proceedings of the National Academy of Sciences* 119 (49):e2211624119.

Gillingham, Kenneth, Marten Ovaere, and Stephanie M. Weber. 2021. “Carbon Policy and the Emissions Implications of Electric Vehicles.”

Graff Zivin, Joshua S., Matthew J. Kotchen, and Erin T. Mansur. 2014. “Spatial and Temporal Heterogeneity of Marginal Emissions: Implications for Electric Cars and Other Electricity-Shifting Policies.” *Journal of Economic Behavior & Organization* 107:248–268.

Hansen, Bruce. 2022. *Econometrics.* Princeton: Princeton University Press.

Hartigan, J. A. 1969. "Using Subsample Values as Typical Values." *Journal of the American Statistical Association* 64 (328):1303–1317.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations.* Boca Raton: Routledge, 1st edition ed.

Heiss, Florian, Stephan Hetzenecker, and Maximilian Osterhaus. 2022. "Nonparametric Estimation of the Random Coefficients Model: An Elastic Net Approach." *Journal of Econometrics* 229 (2):299–321.

Holland, Stephen P., Matthew J. Kotchen, Erin T. Mansur, and Andrew J. Yates. 2022a. "Reply to Gagnon et al.: Short-run Estimates vs. Long-Run Conjectures." *Proceedings of the National Academy of Sciences* 119 (49):e2214219119.

———. 2022b. "Why Marginal CO2 Emissions Are Not Decreasing for US Electricity: Estimates and Implications for Climate Policy." *Proceedings of the National Academy of Sciences* 119 (8):e2116632119.

Holland, Stephen P., Erin T. Mansur, Nicholas Z. Muller, and Andrew J. Yates. 2016. "Are There Environmental Benefits from Driving Electric Vehicles? The Importance of Local Factors." *American Economic Review* 106 (12):3700–3729.

Ichimura, Hidehiko and T. Scott Thompson. 1998. "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution." *Journal of Econometrics* 86 (2):269–295.

Jochmans, Koen and Martin Weidner. 2022. "Inference on a Distribution from Noisy Draws." *Econometric Theory* :1–38.

Johnson, Sarah and Chenyu Yang. 2023. "An Empirical Analysis of the US Generator Interconnection Policy." *Working Paper* .

Kuchibhotla, Arun K., Lawrence D. Brown, Andreas Buja, Junhui Cai, Edward I. George, and Linda H. Zhao. 2020. "Valid Post-Selection Inference in Model-Free Linear Regression." *The Annals of Statistics* 48 (5):2953–2981.

Kwon, Soonwoo. 2021. "Optimal Shrinkage Estimation of Fixed Effects in Linear Panel Data Models." *Working Paper* .

Lahiri, Soumendra N. 2021. "Necessary and Sufficient Conditions for Variable Selection Consistency of the LASSO in High Dimensions." *The Annals of Statistics* 49 (2):820–844.

Leeb, Hannes and Benedikt M. Pötscher. 2005. "Model Selection and Inference: Facts and Fiction." *Econometric Theory* 21 (1):21–59.

Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide. 2020. "Forecasting With Dynamic Panel Data Models." *Econometrica* 88 (1):171–201.

Manresa, Elena. 2015. "Estimating the Structure of Social Interactions Using Panel Data." *Working Paper* .

Meinshausen, Nicolai. 2013. “Sign-Constrained Least Squares Estimation for High-Dimensional Regression.” *Electronic Journal of Statistics* 7 (none):1607–1631.

Miller, Alan. 2002. *Subset Selection in Regression*. New York: Chapman and Hall/CRC, 2 ed.

Moon, Hyungsik Roger and Martin Weidner. 2023. “Nuclear Norm Regularized Estimation of Panel Regression Models.”

Moran, P. A. P. 1973. “Dividing a Sample into Two Parts a Statistical Dilemma.” *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 35 (3):329–333.

Mosteller, Frederick and John Tukey. 1977. *Data Analysis and Regression: A Second Course in Statistics*. Reading, Mass: Pearson, 1st edition ed.

Pelger, Markus and Jiacheng Zou. 2022. “Inference for Large Panel Data with Many Covariates.”

Perlman, Michael D. 1969. “One-Sided Testing Problems in Multivariate Analysis.” *The Annals of Mathematical Statistics* 40 (2):549–567.

Picard, Richard R. and Kenneth N. Berk. 1990. “Data Splitting.” *The American Statistician* 44 (2):140–147.

Plan, Y., R. Vershynin, and E. Yudovina. 2017. “High-Dimensional Estimation with Geometric Constraints.” *Information and Inference: A Journal of the IMA* 6 (1):1–40.

Raskutti, Garvesh, Martin J. Wainwright, and Bin Yu. 2011. “Minimax Rates of Estimation for High-Dimensional Linear Regression Over Lq -Balls.” *IEEE Transactions on Information Theory* 57 (10):6976–6994.

Rinaldo, Alessandro, Larry Wasserman, and Max G’Sell. 2019. “Bootstrapping and Sample Splitting for High-Dimensional, Assumption-Lean Inference.” *The Annals of Statistics* 47 (6):3438–3469.

Ross, Nathan. 2011. “Fundamentals of Stein’s Method.” *Probability Surveys* 8:210–293.

Sasaki, Yuya and Takuya Ura. 2022. “Slow Movers in Panel Data.” *Working Paper* .

Sexton, Steven, A. Justin Kirkpatrick, Robert I. Harris, and Nicholas Z. Muller. 2021. “Heterogeneous Solar Capacity Benefits, Appropriability, and the Costs of Suboptimal Siting.” *Journal of the Association of Environmental and Resource Economists* 8 (6):1209–1244.

Siler-Evans, Kyle, Inês Lima Azevedo, and M. Granger Morgan. 2012. “Marginal Emissions Factors for the U.S. Electricity System.” *Environmental Science & Technology* 46 (9):4742–4748.

Slawski, Martin and Matthias Hein. 2013. “Non-Negative Least Squares for High-Dimensional Linear Models: Consistency and Sparse Recovery without Regularization.” *Electronic Journal of Statistics* 7 (none):3004–3056.

Spiess, Jann, Guido Imbens, and Amar Venugopal. 2023. "Double and Single Descent in Causal Inference with an Application to High-Dimensional Synthetic Control."

Stoer, Josef. 1971. "On the Numerical Solution of Constrained Least-Squares Problems." *SIAM Journal on Numerical Analysis* 8 (2):382–411.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1):267–288.

US EPA, OAR. 2016. "EMC: Continuous Emission Monitoring Systems." https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems.

van de Geer, Sara. 2019. "On the Asymptotic Variance of the Debiased Lasso." *Electronic Journal of Statistics* 13 (2):2970–3008.

van de Geer, Sara, Peter Bühlmann, Ya'acov Ritov, and Ruben Dezeure. 2014. "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models." *The Annals of Statistics* 42 (3):1166–1202.

Vershynin, Roman. 2015. "Estimation in High Dimensions: A Geometric Perspective." In *Sampling Theory, a Renaissance: Compressive Sensing and Other Developments*, edited by Götz E. Pfander, Applied and Numerical Harmonic Analysis. Cham: Springer International Publishing, 3–66.

Viviano, Davide and Jelena Bradic. 2023. "Synthetic Learner: Model-free Inference on Treatments over Time." *Journal of Econometrics* 234 (2):691–713.

Wainwright, Martin J. 2009. "Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using L1-Constrained Quadratic Programming (Lasso)." *IEEE Transactions on Information Theory* 55 (5):2183–2202.

Wolak, Frank A. 1987. "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model." *Journal of the American Statistical Association* 82 (399):782–793.

———. 1989. "Testing Inequality Constraints in Linear Econometric Models." *Journal of Econometrics* 41 (2):205–235.

Ye, Fei and Cun-Hui Zhang. 2010. "Rate Minimaxity of the Lasso and Dantzig Selector for the Lq Loss in Lr Balls." *The Journal of Machine Learning Research* 11:3519–3540.

Zerbib, Nissim, Yen-Huan Li, Ya-Ping Hsieh, and Volkan Cevher. 2016. "Estimation Error of the Constrained Lasso." In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 433–438.

Zhang, Cun-Hui and Stephanie S. Zhang. 2014. "Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 76 (1):217–242.

Zhao, Peng and Bin Yu. 2006. "On Model Selection Consistency of Lasso." *Journal of Machine Learning Research* 7 (90):2541–2563.

APPENDIX

(FOR ONLINE PUBLICATION)

A Proofs

A.1 Proof of Proposition 1

This proof applies Theorem 2 and Corollary 1 of Donoho (1994) to our particular setting. It will be useful to define the the minimax risk with squared error loss for affine estimators:

$$R_{A,N} = \inf_{\hat{\Delta}(.)\text{ affine}} \sup_{\mathbf{b}_n \in \mathcal{B}_n} E((\hat{\Delta}(\mathbf{y}_N) - \Delta(\mathbf{b}_n))^2).$$

As in Donoho (1994), define the modulus of continuity of our linear functional of interest with respect to the semi-norm induced by $X_N = I_n \otimes [x_t]_{t\in\mathcal{T}}$:

$$\omega(\epsilon) = \sup\{|\Delta(\mathbf{b}_1) - \Delta(\mathbf{b}_{-1})| : ||\mathbf{b}_1 - \mathbf{b}_{-1}||_X \leq \epsilon \text{ and } \mathbf{b}_1, \mathbf{b}_{-1} \in \mathcal{B}_n\},$$

where $||[v_i]_{i\in\mathcal{N}}||_X = ||X_N[v_i]_{i\in\mathcal{N}}||_2 = \sqrt{\sum_{(i,t)\in\mathcal{N}\times\mathcal{T}} x_t^2 v_i^2} = \sqrt{\sum_{t\in\mathcal{T}} x_t^2}||[v_i]_{i\in\mathcal{N}}||_2$.

First, we establish a lower bound for $\frac{\omega(\epsilon)}{\epsilon}$. Note that:

$$\begin{aligned}
\Delta(\mathbf{b}_1) - \Delta(\mathbf{b}_{-1}) &= \sum_{i\in\mathcal{N}} \omega_i (b_{1,i} - b_{-1,i}) \\
&= \sum_{i\in\mathcal{N}} (\omega_i - \bar{\omega})(b_{1,i} - b_{-1,i}) \text{ when } \mathbf{b}_1, \mathbf{b}_{-1} \in \mathcal{B}_n
\end{aligned}$$

where $\bar{\omega} = \frac{1}{n}\sum_{i\in\mathcal{N}} \omega_i$ and the second equality is obtained from $\sum_{i\in\mathcal{N}} b_{1,i} = \sum_{i\in\mathcal{N}} b_{-1,i} = 1$ because $\mathbf{b}_1, \mathbf{b}_{-1} \in \mathcal{B}_n$.

Consider two points $\mathbf{b}_1, \mathbf{b}_{-1} \in \mathcal{B}_n$ such that

$$b_{1,i} = b_{-1,i} + \frac{\omega_i - \bar{\omega}}{\lambda},$$

where λ is such that

$$||\mathbf{b}_1 - \mathbf{b}_{-1}||_X = \epsilon,$$

so that

$$\lambda = \frac{\sqrt{S_{X,T}}||\ddot{\boldsymbol{\omega}}_n||_2}{\epsilon},$$

where we define $S_{X,T} = \sum_{t\in\mathcal{T}} x_t^2$ and $\ddot{\boldsymbol{\omega}}_n = [\omega_i - \bar{\omega}]_{i\in\mathcal{N}}$ to shorten notation.

These requirements can be achieved, for instance, by setting $b_{-1,i} = \frac{1}{n}$ and setting ϵ small enough that $b_{1,i} \geq 0 \;\forall\, i$, i.e., requiring:

$$\begin{aligned}
\frac{\omega_i - \bar{\omega}}{\lambda} &\geq -\frac{1}{n}, \text{ or, equivalently:} \\
-n(\omega_i - \bar{\omega}) &\leq \lambda,
\end{aligned}$$

which is implied by $\lambda \geq \max_{i\in\mathcal{N}}|\omega_i - \bar{\omega}| \cdot n$, or, equivalently, $\epsilon \leq \frac{\sqrt{S_{X,T}}||\ddot{\boldsymbol{\omega}}_n||_2}{\max_i|\omega_i - \bar{\omega}| \cdot n}$.

Define $\bar{\epsilon}_N = \frac{\sqrt{S_{X,T}}||\ddot{\boldsymbol{\omega}}_n||_2}{\max_i|\omega_i-\bar{\omega}|\cdot n}$ to shorten notation. We then have, for $0 < \epsilon \leq \bar{\epsilon}_N$:

$$\omega(\epsilon) \geq \frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{\lambda} = \epsilon\frac{||\ddot{\boldsymbol{\omega}}_n||_2}{\sqrt{S_{X,T}}},$$

and

$$\frac{\omega(\epsilon)}{\epsilon} \geq \frac{||\ddot{\boldsymbol{\omega}}_n||_2}{\sqrt{S_{X,T}}}.$$

Second, we can verify the conditions of Theorem 2 of Donoho (1994). For any n, T, $\mathcal{B}_n$ is closed, bounded and convex. The operators $\Delta(.)$ and $X(.) : X(\mathbf{b}) = [x_t b_i]_{(i,t)\in\mathcal{N}\times\mathcal{T}}$ are also both well-defined (see Definition p. 244 of Donoho (1994)) since, for any n, $|\Delta(\mathbf{b}_1) - \Delta(\mathbf{b}_{-1})| \leq \max_{i\in\mathcal{N}}|\omega_i - \bar{\omega}|\sqrt{n}||\mathbf{b}_1 - \mathbf{b}_{-1}||_2 \to 0$ if $||\mathbf{b}_1 - \mathbf{b}_{-1}||_2 \to 0$ and $||X(\mathbf{b}_1) - X(\mathbf{b}_{-1})||_2 = \sqrt{\sum_{t\in\mathcal{T}} x_t^2}||\mathbf{b}_1 - \mathbf{b}_{-1}||_2 \to 0$ if $||\mathbf{b}_1 - \mathbf{b}_{-1}||_2 \to 0$.

Therefore we can apply Theorem 2 of Donoho (1994) to obtain:

$$\begin{aligned}
R_{A,N} &\geq \sup_{0\leq\epsilon\leq\bar{\epsilon}_N}(\frac{\omega(\epsilon)}{\epsilon})^2\frac{\sigma^2\frac{\epsilon^2}{4}}{\frac{\epsilon^2}{4}+\sigma^2} \\
&\geq \sigma^2\frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{S_{X,T}}\sup_{0<\epsilon\leq\bar{\epsilon}_N}\frac{\frac{\epsilon^2}{4}}{\frac{\epsilon^2}{4}+\sigma^2} \\
&= \sigma^2\frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{S_{X,T}}\frac{\frac{\bar{\epsilon}_N^2}{4}}{\frac{\bar{\epsilon}_N^2}{4}+\sigma^2} \\
&= \sigma^2\frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{S_{X,T}}\frac{\frac{1}{4}\frac{S_{X,T}||\ddot{\boldsymbol{\omega}}_n||_2^2}{\max_i|\omega_i-\bar{\omega}|^2\cdot n^2}}{\frac{1}{4}\frac{S_{X,T}||\ddot{\boldsymbol{\omega}}_n||_2^2}{\max_i|\omega_i-\bar{\omega}|^2\cdot n^2}+\sigma^2} \\
&= \sigma^2\frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{n}\frac{\frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{n}}{\frac{S_{X,T}}{n}\frac{||\ddot{\boldsymbol{\omega}}_n||_2^2}{n}+4\cdot\sigma^2\cdot\max_i|\omega_i-\bar{\omega}|^2} \\
&\geq c
\end{aligned}$$

for some positive constant $c > 0$ under the conditions of this proposition.

Finally, by Corollary 1 of Donoho (1994):

$$R_N \geq \frac{4}{5}R_{A,N} \geq \frac{4}{5}\cdot c,$$

which concludes the proof.

A.2 Lemma 1: High-dimensional central limit theorem

Here we establish an asymptotic normal approximation for the regression scores $\frac{1}{T}\sum_{t\in\mathcal{T}}[x_t u_{it}]_{i\in\mathcal{N}}$ over hyperrectangles in the form of a lemma as this result will be used repeatedly throughout other proofs.

Define:

$$\rho_n(\mathcal{C}_{\rm re}) = \sup_{C\in\mathcal{C}_{\rm re}} |P(\frac{1}{\sqrt{T}}\sum_{t\in\mathcal{T}}[x_t u_{it}]_{i\in\mathcal{N}} \in C) - P(\frac{1}{\sqrt{T}}\sum_{t\in\mathcal{T}} Z_{n,t} \in C)|,$$

where $Z_{n,t}$ is a normally distributed $n \times 1$ random vector with the same variance-covariance matrix as the regression scores $[x_t u_{it}]_{i\in\mathcal{N}}$, i.e., $Z_{n,t} \sim N(0, \Sigma_{n,t})$ where $\Sigma_{n,t} = \mathrm{Var}([x_t u_{it}]_{i\in\mathcal{N}})$, and $\mathcal{C}_{\rm re}$ is the class of all hyperrectangles in $\mathbb{R}^n$, i.e., $\mathcal{C}_{\rm re}$ consists of all sets of the form:

$$C = \{[v_i]_{i\in\mathcal{N}} \in \mathbb{R}^n : a_i \leq v_i \leq b_i \ \forall\, i \in \mathcal{N}\},$$

for some $a_i, b_i \in [-\infty, \infty]$, $i \in \mathcal{N}$.

Lemma 1. Under the assumptions of Proposition 2, $\rho_n(\mathcal{C}_{\rm re}) \to 0$ as $T \to \infty$.

Proof. The proof of this result relies on the high-dimensional central limit theorem of Chernozhukov, Chetverikov, and Kato (2017). We therefore start by verifying conditions of Proposition 2.1 of Chernozhukov, Chetverikov, and Kato (2017).

In our setting, condition (M.1) on p. 2314 of Chernozhukov, Chetverikov, and Kato (2017) is obtained from:

$$E(\frac{1}{T}\sum_{t\in\mathcal{T}} x_t^2 u_{it}^2) = \frac{1}{T}\sum_{t\in\mathcal{T}} x_t^2 \sigma_{it}^2 \geq c_1 \cdot \frac{1}{T}\sum_{t\in\mathcal{T}} x_t^2 \geq c_2,\ \forall\, i \in \mathcal{N},$$

for generic positive constants $c_1, c_2 > 0$.[1]

Condition (M.2) of Chernozhukov, Chetverikov, and Kato (2017) is obtained from $E(\frac{1}{T}\sum_{t\in\mathcal{T}} |x_t u_{it}|^{2+k}) \leq C$ for a constant $C < \infty$, which is given by the assumption that both x_t and u_{it} have bounded support. Finally we have condition (E.2) of Chernozhukov, Chetverikov, and Kato (2017), $E((\frac{\max_{i\in\mathcal{N}}|x_t u_{it}|}{C})^q) \leq 1 < 2$, where the normalizing sequence in the denominator (B_n Chernozhukov, Chetverikov, and Kato (2017)) is constant in our case, and given by C, where C is chosen to exceed the upper bound of the support of $|x_t u_{it}|$, and we set $q = \frac{1}{4}$ for convenience (we could have chosen any strictly positive q here).

Then, define $D_n = (\frac{\log^7(n)}{T})^{\frac{1}{6}}$, Proposition 2.1 of Chernozhukov, Chetverikov, and Kato (2017) implies:

$$\rho_n(\mathcal{C}_{\rm re}) \leq C \cdot D_n = o(1)$$

for a constant $C < \infty$.

[1] Recall that the covariate x_t is treated as deterministic throughout section 3.

A.3 Proof of Proposition 2

Part 1. Upper bound for γ_N.

We begin the proof of this result by establishing an upper bound for the penalization parameter γ_N. To establish this result, first note that the quantity $S_N(r) = \sum_{i\in\mathcal{N}} 1[\hat{b}_i \geq r](\hat{b}_i - r)$ is strictly decreasing in r when $S_N(r) > 0$. Therefore, for any sequence r_N, we have

$$P(\gamma_N \leq r_N) = P(S_N(\gamma_N) \geq S_N(r_N)) = P(1 \geq S_N(r_N)),$$

where the last equality follows from γ_N being defined by solving $S_N(\gamma_N) = 1$.

Consider the sequence $r_N = \frac{\bar{\sigma}}{s_{X,T}}\sqrt{2(1+\delta)\frac{\log(n)}{T}} + \exp(-c_1 \cdot n)$ for a small positive constant $\delta > 0$, where c_1 is the constant from the assumption of approximate sparsity (8), and with $s_{X,T} = \frac{1}{T}\sum_{t\in\mathcal{T}} x_t^2$, $\bar{\sigma} = \sup_{i\in\mathcal{N},N\in\mathbb{N}\times\mathbb{N}}\sigma_{i,T}$, $\sigma_{i,T} = \sqrt{\text{Var}(\sqrt{T}v_{i,T})}$, $v_{i,T} = \frac{1}{T}\sum_{t\in\mathcal{T}} x_t u_{it}$. Note that Assumption 1 implies that $\bar{\sigma}$ and $s_{X,T}$ are bounded away from infinity and zero and that these definitions imply that $\hat{b}_i - b_i = \frac{v_{i,T}}{s_{X,T}}$.

We have:

$$S_N(r_N) = \sum_{i\notin\mathcal{A}_n} 1[\hat{b}_i \geq r_N](\hat{b}_i - r_N) + \sum_{i\in\mathcal{A}_n} 1[\hat{b}_i \geq r_N](\hat{b}_i - r_N). \tag{A.1}$$

Considering the first term, we have

$$\begin{aligned} P(\sum_{i\notin\mathcal{A}_n} 1[\hat{b}_i \geq r_N](\hat{b}_i - r_N) = 0) &= P(\max_{i\notin\mathcal{A}_n}\hat{b}_i \leq r_N) \\ &\geq P(\max_{i\notin\mathcal{A}_n}\sqrt{T}v_{i,T} \leq \bar{\sigma}\sqrt{2(1+\delta)\log(n)}), \end{aligned}$$

where the inequality follows from the assumption of approximate sparsity (8).

Let $[z_{i,T}]_{i\in\mathcal{N}} \sim N(0, \Upsilon_N)$, where $\Upsilon_N = \text{Var}([\sqrt{T}v_{i,T}]_{i\in\mathcal{N}})$. By Lemma 1, we have:

$$P(\max_{i\notin\mathcal{A}_n}\sqrt{T}v_{i,T} \leq \bar{\sigma}\sqrt{2(1+\delta)\log(n)}) = P(\max_{i\notin\mathcal{A}_n} z_{i,T} \leq \bar{\sigma}\sqrt{2(1+\delta)\log(n)}) + o(1).$$

We then have

$$\begin{aligned} P(\max_{i\notin\mathcal{A}_n} z_{i,T} \leq \bar{\sigma}\sqrt{2(1+\delta)\log(n)}) &\geq 1 - \sum_{i\notin\mathcal{A}_n} P(z_{i,T} > \bar{\sigma}\sqrt{2(1+\delta)\log(n)}) \\ &\geq 1 - \sum_{i\notin\mathcal{A}_n} \exp(-\frac{\bar{\sigma}^2(1+\delta)\log(n)}{\sigma_{i,T}^2}) \\ &\geq 1 - \frac{n}{n^{(1+\delta)}} \to 1, \end{aligned}$$

where the second inequality follows from the Gaussian tail inequality and the third inequality follows from $\bar{\sigma}^2 = \sup_{i\in\mathcal{N},N\in\mathbb{N}\times\mathbb{N}}\sigma_{i,T}$. We then have $P(\sum_{i\notin\mathcal{A}_n} 1[\hat{b}_i \geq r_N](\hat{b}_i - r_N) = 0) \to 1$.

Now taking the second term in (A.1), we have

$$P(\min_{i\in\mathcal{A}_n}\hat{b}_i \geq r_N) \geq P(\min_{i\in\mathcal{A}_n}(\hat{b}_i - b_i) \geq r_N - \frac{c_2}{|\mathcal{A}_n|}),$$
$$= P(\min_{i\in\mathcal{A}_n}\sqrt{T}v_{i,T} \geq s_{X,T}\sqrt{T}(r_N - \frac{c_2}{|\mathcal{A}_n|}))$$

where c_2 is the positive constant defined in the assumption of approximate sparsity (8).

Under the rate condition $|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}} \to 0$, we have $\frac{|\mathcal{A}_n|}{\sqrt{T}}\sqrt{T}r_N \propto |\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}} \to 0$ and $\sqrt{T}r_N - \frac{c_2\sqrt{T}}{|\mathcal{A}_n|} = -\frac{c_2\sqrt{T}}{|\mathcal{A}_n|} + o(\frac{\sqrt{T}}{|\mathcal{A}_n|})$, so that we can use the same steps as above to show that $P(\min_{i\in\mathcal{A}_n}\hat{b}_i \geq r_N) \to 1$ and

$$P(\sum_{i\in\mathcal{A}_n} 1[\hat{b}_i \geq r_N](\hat{b}_i - r_N) = \sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N)) \to 1.$$

We also have

$$\sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N) = \sum_{i\in\mathcal{A}_n} b_i + \frac{1}{s_{X,T}}\sum_{i\in\mathcal{A}_n} v_{i,T} - |\mathcal{A}_n|\cdot r_N,$$
$$= 1 + O(n\cdot\exp(-c_1\cdot n)) + O_p(\frac{|\mathcal{A}_n|}{\sqrt{T}}) - |\mathcal{A}_n|\cdot r_N,$$
$$= 1 - |\mathcal{A}_n|\cdot r_N + o_p(|\mathcal{A}_n|\cdot r_N),$$

where the second equality follows from the assumption of approximate sparsity (8) and Assumption 1, the last equality follows from the rate conditions imposed for this result.

We therefore have

$$P(\sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N) \leq 1) \to 1,$$

and

$$P(S_N(r_N) \leq 1) \geq P(S_N(r_N) = \sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N)\,,\ \sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N) \leq 1)$$
$$\geq 1 - P(S_N(r_N) \neq \sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N)) - P(\sum_{i\in\mathcal{A}_n}(\hat{b}_i - r_N) > 1)$$
$$\to 1,$$

which establishes the desired result and $\gamma_N \leq r_N$ w.p.a. 1, which also establishes $\gamma_N = O_p(\sqrt{\frac{\log(n)}{T}})$.[2]

[2]Note that we have $P(\gamma_N \geq -C\cdot\sqrt{\frac{\log(n)}{T}}) \to 1$ since for $C > 0$ large enough, if $\gamma_N \leq -C\cdot\sqrt{\frac{\log(n)}{T}}$, we would have $P(\mathcal{N} = \hat{\mathcal{A}}_n) \to 1$ and $\sum_{i\in\hat{\mathcal{A}}_n}(\hat{b}_i - \gamma_N) \geq \sum_{i\in\mathcal{N}}\hat{b}_i + n\cdot C\cdot\sqrt{\frac{\log(n)}{T}} = 1 + nC\cdot\sqrt{\frac{\log(n)}{T}} + o_p(n\sqrt{\frac{\log(n)}{T}}) > 1$ w.p.a. 1.

Part 2. Consistent model selection.

Consistent model selection follows directly from the above since

$$P(\min_{i\in\mathcal{A}_n}\hat{b}_i \geq \gamma_N) \geq P(\min_{i\in\mathcal{A}_n}\hat{b}_i \geq r_N, \gamma_N \leq r_N)$$
$$\to 1,$$

with r_N still defined as in part 1.

Part 3. Consistent estimation of Δ.

To establish consistency of $\tilde{\Delta}$ for Δ, we start again from $S_N(\gamma_N) = 1$ but now use it to bound the amount of estimation error originating from cross-sectional units $i \notin \mathcal{A}_n$:

$$\begin{aligned}
\sum_{i\notin\mathcal{A}_n} 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) &= 1 - \sum_{i\in\mathcal{A}_n} 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) \\
&= 1 - \sum_{i\in\mathcal{A}_n} (\hat{b}_i - \gamma_N) \text{ w.p.a. } 1 \\
&= O(n \cdot \exp(-c_1 \cdot n)) - \sum_{i\in\mathcal{A}_n} (\hat{b}_i - b_i - \gamma_N) \\
&= O_p(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}}),
\end{aligned}$$

from the results above.

Then

$$\begin{aligned}
\tilde{\Delta} - \Delta &= \sum_{i\notin\mathcal{A}_n} \omega_i 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) + \sum_{i\in\mathcal{A}_n} \omega_i 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) - \sum_{i\in\mathcal{N}} \omega_i b_i, \\
&= \sum_{i\notin\mathcal{A}_n} \omega_i 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) + \sum_{i\in\mathcal{A}_n} \omega_i (\hat{b}_i - \gamma_N) - \sum_{i\in\mathcal{N}} \omega_i b_i \text{ w.p.a. } 1, \\
&= \sum_{i\notin\mathcal{A}_n} \omega_i 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) + \sum_{i\in\mathcal{A}_n} \omega_i (\hat{b}_i - b_i - \gamma_N) - \sum_{i\notin\mathcal{A}_n} \omega_i b_i
\end{aligned}$$

and we have

$$\begin{aligned}
|\sum_{i\notin\mathcal{A}_n} \omega_i 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N)| &\leq \max_{i\in\mathcal{N}}|\omega_i| \sum_{i\notin\mathcal{A}_n} 1[\hat{b}_i \geq \gamma_N](\hat{b}_i - \gamma_N) \\
&= O_p(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}}),
\end{aligned}$$

from the condition that the range of ω_i is bounded.

Under Assumption 1 we have

$$\sum_{i\in\mathcal{A}_n} \omega_i(\hat{b}_i - b_i) = O_p(\frac{|\mathcal{A}_n|}{\sqrt{T}}).$$

We also have $\gamma_N = O_p(\sqrt{\frac{\log(n)}{T}})$ from the above, and $\sum_{i \notin \mathcal{A}_n} \omega_i b_i = O(n \cdot \exp(-c_1 \cdot n)) = o(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}})$ from the assumption of approximate sparsity (8), so that we obtain the desired result that

$$\tilde{\Delta} - \Delta = O_p(|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}}),$$

which concludes this proof.

A.4 Proof of Proposition 3

We begin with the decomposition:

$$\begin{aligned}\breve{\Delta} - \Delta = &\sum_{i \in \mathcal{N}} (\omega_i - \Delta)(1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i) \\ &+ (\tilde{\Delta} - \Delta) \sum_{i \in \mathcal{N}} (1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i),\end{aligned}$$

using the representation $\Delta = \tilde{\Delta} + \sum_{i \in \mathcal{N}} (\omega_i - \tilde{\Delta}) b_i$ given in the main text to obtain $\breve{\Delta} - \Delta = \sum_{i \in \mathcal{N}} (\omega_i - \tilde{\Delta})(1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i)$, from which the above is obtained.

Part 1. Asymptotic ignorability of $(\tilde{\Delta} - \Delta) \sum_{i \in \mathcal{N}} (1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i)$.

Since the conditions provided for this result to hold nest those of Proposition 2, we have $\tilde{\Delta} - \Delta = o_p(1)$.

To establish asymptotic ignorability of $(\tilde{\Delta} - \Delta) \sum_{i \in \mathcal{N}} (1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i)$, under part 2 of Assumption 2, we therefore need to establish:[3]

$$\sqrt{\frac{T}{r_N + |\mathcal{A}_n|}} \sum_{i \in \mathcal{N}} (1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i) = O_p(1).$$

We can write

$$\begin{aligned}\sum_{i \in \mathcal{N}} (1[i \in \hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} - b_i) = &\sum_{i \notin \mathcal{A}_n} 1[i \in \hat{\mathcal{A}}_{1,n}] e_{2,iT} - \sum_{i \notin \mathcal{A}_n} 1[i \notin \hat{\mathcal{A}}_{1,n}] b_i \\ &+ \sum_{i \in \mathcal{A}_n} 1[i \in \hat{\mathcal{A}}_{1,n}] e_{2,iT} - \sum_{i \in \mathcal{A}_n} 1[i \notin \hat{\mathcal{A}}_{1,n}] b_i.\end{aligned}$$

From the assumption of approximate sparsity (8), we have $\sum_{i \notin \mathcal{A}_n} 1[i \notin \hat{\mathcal{A}}_{1,n}] b_i \leq \sum_{i \notin \mathcal{A}_n} b_i = o(\sqrt{\frac{|\mathcal{A}_n|}{T}})$.

As in Proposition 2, we have $P(\mathcal{A}_n \subseteq \hat{\mathcal{A}}_{n,1}) \to 1$, so that $P(\sum_{i \in \mathcal{A}_n} 1[i \notin \hat{\mathcal{A}}_{1,n}] b_i = 0) \to 1$,

[3]Note that the sequence r_N is defined in the main text as $r_N = \sum_{i \notin \mathcal{A}_n} 1[i \in \hat{\mathcal{A}}_{1,n}]$, it is distinct from the sequence r_N used in the proof of Proposition 2 above.

and $\sum_{i\in\mathcal{A}_n} 1[i \notin \hat{\mathcal{A}}_{1,n}]b_i = o_p(\sqrt{\frac{|\mathcal{A}_n|}{T}})$.

We also have:

$$\begin{aligned}
\mathrm{Var}(\sqrt{\frac{T}{r_N+|\mathcal{A}_n|}}\sum_{i\notin\mathcal{A}_n} 1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT}) &= E(\mathrm{Var}(\sqrt{\frac{T}{r_N+|\mathcal{A}_n|}}\sum_{i\notin\mathcal{A}_n} 1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT}\,|\,\mathcal{D}_1)) \\
&= E(\frac{1}{r_N+|\mathcal{A}_n|}\mathrm{Var}(\sum_{i\notin\mathcal{A}_n} 1[i\in\hat{\mathcal{A}}_{1,n}]\sqrt{T}e_{2,iT}\,|\,\mathcal{D}_1)) \\
&\leq C\cdot E(\frac{1}{r_N+|\mathcal{A}_n|}r_N) \leq C,
\end{aligned}$$

where $0 < C < \infty$ is a positive constant, the first equality follows from the law of iterated variances, since $E(\sum_{i\notin\mathcal{A}_n} 1[i \in \hat{\mathcal{A}}_{1,n}]e_{2,iT}\,|\,\mathcal{D}_1) = 0$, and the inequality follows from Assumption 1 (serial independence, bounded support, $\frac{1}{T}\sum_{t\in\mathcal{T}_2} x_t^2 \geq c > 0$) and Assumption 2 (neighborhood dependence) and the Cauchy-Schwarz inequality. Therefore,

$$\sqrt{\frac{T}{r_N+|\mathcal{A}_n|}}\sum_{i\notin\mathcal{A}_n} 1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT} = O_p(1).$$

Similarly:

$$\mathrm{Var}(\sqrt{\frac{T}{r_N+|\mathcal{A}_n|}}\sum_{i\in\mathcal{A}_n} 1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT}) \leq C,$$

for a finite positive constant C, and $\sqrt{\frac{T}{r_N+|\mathcal{A}_n|}}\sum_{i\in\mathcal{A}_n} 1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT} = O_p(1)$, which establishes the result of asymptotic ignorability of $(\tilde{\Delta}-\Delta)\sum_{i\in\mathcal{N}}(1[i\in\hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i}-b_i)$.

Part 2. Asymptotic normality of $\sum_{i\in\mathcal{N}}(\omega_i-\Delta)(1[i\in\hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i}-b_i)$.

As in Part 1, we can write:

$$\begin{aligned}
\sum_{i\in\mathcal{N}}(\omega_i-\Delta)(1[i\in\hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i}-b_i) = &\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT} + \sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)e_{2,iT} \\
&- \sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\notin\hat{\mathcal{A}}_{1,n}]b_i \\
&- \sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)1[i\notin\hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i},
\end{aligned}$$

and we have

$$\sqrt{\frac{T}{\Sigma_N}}\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\notin\hat{\mathcal{A}}_{1,n}]b_i = o_p(1),$$

$$\sqrt{\frac{T}{\Sigma_N}}\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)1[i\notin\hat{\mathcal{A}}_{1,n}]\hat{b}_{2,i} = o_p(1).$$

It then only remains to establish that $\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]e_{2,iT}+\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)e_{2,iT}$ is asymptotically normal. To establish this asymptotic normality, we use Stein's lemma, and more particularly its application to data with neighborhood dependence in Theorem 3.6 of Ross (2011). As discussed in the main text, we have $E(e_{2,iT}|\mathcal{D}_1)=0$ by the independence of the two samples used for selection and estimation. Neighborhood dependence of $e_{2,iT}$ across $i\in\mathcal{N}$, conditional on $\mathcal{D}_1$, is also given by part 1 of Assumption 2 and serial independence in Assumption 1.

We also have

$$E(|e_{2,iT}|^3|\mathcal{D}_1)\leq C\cdot\frac{1}{T^2},$$

from the independence between the two samples used for selection and estimation, serial independence, $\frac{1}{T}\sum_{t\in\mathcal{T}_2}x_t^2\geq c$, and bounded support of x_t and u_{it}.

Therefore we have:

$$\sum_{i\notin\mathcal{A}_n}|\omega_i-\Delta|^3 1[i\in\hat{\mathcal{A}}_{1,n}]E(|e_{2,iT}|^3|\mathcal{D}_1)\leq C\cdot\frac{r_N}{T^2}.$$

We also have

$$\sum_{i\in\mathcal{A}_n}|\omega_i-\Delta|^3E(|e_{2,iT}|^3|\mathcal{D}_1)\leq C\cdot\frac{|\mathcal{A}_n|}{T^2}.$$

Therefore,

$$\frac{T^{\frac{3}{2}}}{(r_N+|\mathcal{A}_n|)^{\frac{3}{2}}}(\sum_{i\notin\mathcal{A}_n}|\omega_i-\Delta|^3 1[i\in\hat{\mathcal{A}}_{1,n}]E(|e_{2,iT}|^3|\mathcal{D}_1)+\sum_{i\in\mathcal{A}_n}|\omega_i-\Delta|^3E(|e_{2,iT}|^3|\mathcal{D}_1)$$
$$\leq C\cdot\sqrt{\frac{1}{(r_N+|\mathcal{A}_n|)T}}=o(1).$$

Similarly,

$$\sum_{i\notin\mathcal{A}_n}|\omega_i-\Delta|^4 1[i\in\hat{\mathcal{A}}_{1,n}]E(|e_{2,iT}|^4|\mathcal{D}_1)\leq C\cdot\frac{r_N}{T^2},$$

and

$$\sum_{i\in\mathcal{A}_n}|\omega_i-\Delta|^4E(|e_{2,iT}|^4|\mathcal{D}_1)\leq C\cdot\frac{|\mathcal{A}_n|}{T^2},$$

and

$$\frac{T^2}{(r_N+|\mathcal{A}_n|)^2}(\sum_{i\notin\mathcal{A}_n}|\omega_i-\Delta|^4 1[i\in\hat{\mathcal{A}}_{1,n}]E(|e_{2,iT}|^4|\mathcal{D}_1)+\sum_{i\in\mathcal{A}_n}|\omega_i-\bar{\omega}|^4E(|e_{2,iT}|^4|\mathcal{D}_1)$$
$$\leq\frac{1}{r_N+|\mathcal{A}_n|}=o(1).$$

From part 2 of Assumption 2 and Theorem 3.6 of Ross (2011), we therefore have $\frac{1}{\sqrt{\Sigma_N}}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]\sqrt{T}e_{2,iT}+\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)\sqrt{T}e_{2,iT})\xrightarrow{d}N(0,1)$ conditionally on $\mathcal{D}_1$, and we obtain unconditional convergence in distribution by the dominated convergence theorem, which completes the proof of this result.

A.5 Proof of Proposition 4

As in the proof of Propositions 2 and 3, we have, w.p.a. 1:

$$\frac{1}{T}\sum_{t\in\mathcal{T}_2}(\sum_{i\in\hat{\mathcal{A}}_{1,n}}(\omega_i-\tilde{\Delta})\hat{v}_{it})^2=\frac{1}{T}\sum_{t\in\mathcal{T}_2}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\tilde{\Delta})1[i\in\hat{\mathcal{A}}_{1,n}]\hat{v}_{it}+\sum_{i\in\mathcal{A}_n}(\omega_i-\tilde{\Delta})\hat{v}_{it})^2,$$

because $P(\mathcal{A}_n\subseteq\hat{\mathcal{A}}_{1,n})\to 1$.

In addition, we have

$$\begin{aligned}\sum_{i\notin\mathcal{A}_n}(\omega_i-\tilde{\Delta})1[i\in\hat{\mathcal{A}}_{1,n}]\hat{v}_{it}&=\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}](v_{it}-x_t^2(\hat{b}_{2,i}-b_i))\\&\quad-(\tilde{\Delta}-\Delta)\sum_{i\notin\mathcal{A}_n}1[i\in\hat{\mathcal{A}}_{1,n}](v_{it}-x_t^2(\hat{b}_{2,i}-b_i)),\\ \text{and}\ \sum_{i\in\mathcal{A}_n}(\omega_i-\tilde{\Delta})\hat{v}_{it}&=\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)(v_{it}-x_t^2(\hat{b}_{2,i}-b_i))\\&\quad-(\tilde{\Delta}-\Delta)\sum_{i\in\mathcal{A}_n}(v_{it}-x_t^2(\hat{b}_{2,i}-b_i)),\end{aligned}$$

where we define $v_{it}=x_tu_{it}$.

From this decomposition, we can show that

$$\begin{aligned}&\frac{1}{T(r_N+|\mathcal{A}_n|)}\sum_{t\in\mathcal{T}_2}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\tilde{\Delta})1[i\in\hat{\mathcal{A}}_{1,n}]\hat{v}_{it}+\sum_{i\in\mathcal{A}_n}(\omega_i-\tilde{\Delta})\hat{v}_{it})^2\\=&\frac{1}{T(r_N+|\mathcal{A}_n|)}\sum_{t\in\mathcal{T}_2}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}+\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)v_{it})^2\\&+o_p(1).\end{aligned}\tag{A.2}$$

For instance:

$$E(|\frac{1}{T}\sum_{t\in\mathcal{T}_2}\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}x_t^2(\hat{b}_{2,i}-b_i)|\,|\,\mathcal{D}_1)$$

$$\leq \sum_{i\notin\mathcal{A}_n}|\omega_i-\Delta|1[i\in\hat{\mathcal{A}}_{1,n}]\cdot E(|\frac{1}{T}\sum_{t\in\mathcal{T}_2}x_t^2v_{it}|\cdot|\hat{b}_{2,i}-b_i|\,|\,\mathcal{D}_1)$$

$$\leq \sum_{i\notin\mathcal{A}_n}|\omega_i-\Delta|\cdot 1[i\in\hat{\mathcal{A}}_{1,n}]\sqrt{E((\frac{1}{T}\sum_{t\in\mathcal{T}_2}x_t^2v_{it})^2|\mathcal{D}_1)}\sqrt{E((\hat{b}_{2,i}-b_i)^2|\mathcal{D}_1)}$$

$$\leq C\cdot\frac{1}{T}r_N,$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the third inequality follows from serial independence and bounded support.

We then obtain

$$E(\frac{1}{T(r_N+|\mathcal{A}_n|)}|\sum_{t\in\mathcal{T}_2}\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}x_t^2(\hat{b}_{2,i}-b_i)|)$$

$$= E(\frac{1}{T(r_N+|\mathcal{A}_n|)}E(|\sum_{t\in\mathcal{T}_2}\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}x_t^2(\hat{b}_{2,i}-b_i)|\,|\,\mathcal{D}_1))$$

$$\leq E(\frac{r_N}{T(r_N+|\mathcal{A}_n|)})$$

$$\leq C\frac{1}{T}=o(1),$$

where the equality applies the law of iterated expectations, so that

$$\frac{1}{T(r_N+|\mathcal{A}_n|)}|\sum_{t\in\mathcal{T}_2}\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}x_t^2(\hat{b}_{2,i}-b_i)|=o_p(1).$$

Similar derivations can be applied to all of the remaining terms to establish (A.2).

We also have:

$$\frac{1}{T}\sum_{t\in\mathcal{T}_2}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}+\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)v_{it})^2=T_{1,N}+2T_{2,N}+T_{3,N},$$

where

$$T_{1,N}\coloneqq\frac{1}{T}\sum_{t\in\mathcal{T}_2}\sum_{i,i'\notin\mathcal{A}_n}(\omega_i-\Delta)(\omega_i^{'}-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}1[i^{'}\in\hat{\mathcal{A}}_{1,n}]v_{i't},$$

$$T_{2,N}\coloneqq\frac{1}{T}\sum_{t\in\mathcal{T}_2}\sum_{i\notin\mathcal{A}_n,i'\in\mathcal{A}_n}(\omega_i-\Delta)(\omega_i^{'}-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}v_{i't},$$

$$T_{3,N}\coloneqq\frac{1}{T}\sum_{t\in\mathcal{T}_2}\sum_{i,i'\in\mathcal{A}_n}(\omega_i-\Delta)(\omega_i^{'}-\Delta)v_{it}v_{i't}.$$

Each term can be shown to converge in probability to its conditional expected value (at rate $r_N + |\mathcal{A}_n|$). For instance:

$$\begin{aligned}\mathrm{Var}(T_{1,N}|\mathcal{D}_1) &= \frac{1}{T^2}\sum_{t\in\mathcal{T}_2}\mathrm{Var}(\sum_{i,i'\notin\mathcal{A}_n}(\omega_i-\Delta)(\omega_i^{'}-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}1[i^{'}\in\hat{\mathcal{A}}_{1,n}]v_{i^{'}t}|\mathcal{D}_1)\\ &\leq C\cdot\frac{r_N^2}{T},\end{aligned}$$

under Assumptions 1 and 2, so that, by the law of iterated expectations:

$$E(\frac{1}{r_N^2}(T_{1,N}-E(T_{1,N}|\mathcal{D}_1))^2)\leq C\cdot\frac{1}{T},$$

and

$$\frac{1}{r_N}(T_{1,N}-E(T_{1,N}|\mathcal{D}_1)) = o_p(1).$$

Similarly, we have

$$\frac{1}{\sqrt{r_N|\mathcal{A}_n|}}(T_{2,N}-E(T_{2,N}|\mathcal{D}_1)) = o_p(1),$$

and

$$T_{3,N}-E(T_{3,N}|\mathcal{D}_1) = o_p(|\mathcal{A}_n|),$$

so that:

$$\frac{1}{r_N+|\mathcal{A}_n|}(\frac{1}{T}\sum_{t\in\mathcal{T}_2}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}+\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)v_{it})^2-V_N) = o_p(1),$$

where $V_N = \mathrm{Var}(\sum_{i\notin\mathcal{A}_n}(\omega_i-\Delta)1[i\in\hat{\mathcal{A}}_{1,n}]v_{it}+\sum_{i\in\mathcal{A}_n}(\omega_i-\Delta)v_{it}|\mathcal{D}_1)$. Therefore, by part 2 of Assumption 2 and the definition $\Sigma_N = \frac{V_N}{s^2_{2,X,T}}$:

$$\frac{\hat{\Sigma}_N}{\Sigma_N}-1 = o_p(1).$$

A.6 Proof of Proposition 5

Since our constrained regression problem amounts to the minimization of a convex objective function over a convex polytope, its solution is unique. To establish the asymptotic properties of the resulting solution, $\{\tilde{b}_i : i\in\mathcal{N}\}$, this proof makes use of the particular structure of a numerical algorithm that can reach this unique solution from any feasible starting value, i.e., any value in $(\mathbb{R}^K)^n$ that satisfies the positivity and summing-to-one constraints. Specifically, we make use of the algorithm and convergence results discussed in Stoer (1971) for the constrained least-squares problem.

This algorithm is an example of a method of feasible directions, where the algorithm proceeds from one feasible point (a point that satisfies the positivity and summing-to-one constraints) to another. Denote $\beta_i[k]$ to be the k^{th} element of $\beta_i\in\mathbb{R}^K$. In this proof we show that, if the algorithm is started at an initial value $\{\beta_i : i \in \mathcal{N}\}$ such that $\beta_i[k] > 0$ if $i \in \mathcal{A}_{n,k}$, then, with probability approaching one, it will preserve $\beta_i[k] > 0$ if $i \in \mathcal{A}_{n,k}$ along all subsequent steps, so that $P(\mathcal{A}_{n,k} \subseteq \hat{\mathcal{A}}_{n,k}) \to 1$, i.e., our multivariate estimator yields consistent model selection.[4] As a by-product of establishing this result, we will also show that the Lagrangian multipliers corresponding to the summing-to-one constraints will have a convergence rate of $\sqrt{\frac{\log(n)}{T}}$. Finally, given these two results we can derive a convergence rate for $\tilde{\Delta}_{n,k} - \Delta_k$ in a similar way as for the univariate estimator. We begin this proof by describing the algorithm of Stoer (1971) in our setting.

A.6.1 Description of the Gradient Projection Algorithm

In our setting, the algorithm of Stoer (1971) proceeds as follows:

For any current value β of the algorithm, define $\mathcal{I}(\beta) = \{(i,k) \in \mathcal{N} \times \mathcal{K} : \beta_i[k] = 0\}$ to be the set of positivity constraints that are binding for that value β. Each step of the algorithm requires a value β as well as a set of "active" constraints $\mathcal{J} \subseteq \mathcal{N} \times \mathcal{K}$ that is contained in $\mathcal{I}(\beta)$, $\mathcal{J} \subseteq \mathcal{I}(\beta)$. The algorithm is initialized with $\mathcal{J} = \mathcal{I}(\beta)$.

Step 1: Direction and length of update.

Define the direction of the update, $s = \{s_i : i \in \mathcal{N}\}$, $s_i = [s_{i,k}]_{k\in\mathcal{K}}$, to solve the least-squares problem with binding constraints given by the set of positivity constraints indexed by $\mathcal{J}$ and the summing-to-one constraints:

$$s = \operatorname{argmin}_{\sigma_i \in \mathbb{R}^K : i \in \mathcal{N}} \sum_{(i,t)\in\mathcal{N}\times\mathcal{T}} (y_{it} - x_t(\beta_i + \sigma_i))^2 \text{ s.t. } \sigma_{i,k} = 0 \,\forall\, (i,k) \in \mathcal{J}, \sum_{i\in\mathcal{N}} \sigma_{i,k} = 0 \,\forall\, k \in \mathcal{K}, \tag{A.3}$$

and the length of the update to satisfy feasibility:

$$\lambda = \max\{\tau \mid \tau \in [0,1],\ \beta_i[k] + \tau s_{i,k} \geq 0\ \forall\, i,k\}.$$

The new value of the algorithm, $\bar{\beta}$, is then defined as $\bar{\beta} = \beta + \lambda s$.

Define $D = \mathcal{I}(\bar{\beta}) \backslash \mathcal{J}$ to be the set of positivity constraints that are binding for $\bar{\beta}$ but not indexed by $\mathcal{J}$. If $D = \emptyset$, go to Step 3, otherwise go to Step 2.

[4] Since this algorithm is only used as a device in our proof, finding such a starting value does not have to be implemented, but we can note that a univariate estimator applied to the multivariate regression problem, treating $x_{k',t}$ as control covariates for all $k' \neq k$, and repeating the estimation for each $k \in \mathcal{K}$, would satisfy this condition with probability approaching one.

Step 2: Addition step.

Define $\mathcal{J} = \mathcal{I}(\bar{\beta})$.

Step 3: Dropping step.

Either $\bar{\beta}$ is found to be optimal, if all Lagrangian multipliers associated with the binding positivity constraints are found to be positive, so that $\bar{\beta}$ solves the KKT conditions, or a positivity constraint corresponding to a negative Lagrangian multiplier is dropped from the active set $\mathcal{J}$. This step of the algorithm is described in Stoer (1971), and the details of the dropping step are not necessary for this proof since our proof concentrates on the addition step only.

Theorem 2.33 in Stoer (1971) shows that this algorithm converges to the solution of the constrained regression problem (16) in a finite number of steps.

A.6.2 Consistent model selection and rate of convergence of the Lagrangian multipliers

Representation of the constrained minimization problem. We begin by providing a useful representation of the solution to the constrained minimization problem (A.3).

Solving for $\grave{\beta}_i = \beta_i + s_i$ from the constrained minimization problem (A.3), we obtain the following first-order conditions for the Lagrangian objective function:

$$S[m_i]\grave{\beta}_i[m_i] - v_i[m_i] + \gamma[m_i] = 0 \; \forall \, i \in \mathcal{N},$$

where, to shorten notation, we write $S = \frac{1}{T}\sum_{t\in\mathcal{T}} x_t' x_t$ instead of $S_{x,T}$ as it was defined for Assumption 3, as further subscripts will have to be used for this matrix, $m_i = \{k \in \mathcal{K} : (i,k) \notin \mathcal{J}\}$, $S[m_i] = [S[k,k']]_{k\in m_i}^{k'\in m_i}$, $v_i = \frac{1}{T}\sum_{t\in\mathcal{T}} x_t' y_{it}$, $v_i[m_i] = [v_i[k]]_{k\in m_i}$, and γ is the $K\times 1$ vector of Lagrangian multipliers corresponding to the summing-to-one constraints, where again we dropped the subscript used in stating this result, writing γ instead of $\gamma_N = [\gamma_{N,k}]_{k\in\mathcal{K}}$, and $\gamma[m_i] = [\gamma[k]]_{k\in m_i}$.

Therefore

$$\begin{aligned}\grave{\beta}_i[m_i] &= S[m_i]^{-1}v_i[m_i] - S[m_i]^{-1}\gamma[m_i] \\ &= b_i[m_i] + S[m_i]^{-1}\frac{1}{T}\sum_{t\in\mathcal{T}} x_t'[m_i]u_{it} - S[m_i]^{-1}\gamma[m_i], \qquad \text{(A.4)}\end{aligned}$$

where $x_t^{'}[m_i] = [x_t[k]]_{k \in m_i}$ and the second equality follows from the regression model (13).

In addition, this solution must satisfy the summing-to-one constraints:

$$\sum_{i:(i,k)\notin \mathcal{J}} \grave{\beta}_i[k] = 1 \;\forall\, k \in \mathcal{K}. \tag{A.5}$$

Consistency of Model Selection. Now we wish to establish that, if the initial value of the algorithm, β, is chosen such that the true response set $\mathcal{A}_{n,k}$ does not overlap with the active constraints indexed by $\mathcal{J}$, i.e., $\{(i,k) : k \in \mathcal{K}, i \in \mathcal{A}_{n,k}\} \cap \mathcal{J} = \emptyset$, then the subsequent active sets $\bar{\mathcal{J}}$ will satisfy the same requirement with probability approaching one, leading to the conclusion that $\mathcal{A}_{n,k} \subseteq \hat{\mathcal{A}}_{n,k}$ with probability approaching one since the algorithm is guaranteed to converge to the solution of our constrained minimization problem.

Clearly, if a constraint is dropped from the active set (Step 3 above), the new active set will not overlap with $\mathcal{A}_{n,k}$ if the previous active set did not. Therefore we can concentrate on cases where constraints are added to the active set (Step 2 above). To prove consistent model selection we show that, w.p.a. 1, combinations (i,k) s.t. $i \in \mathcal{A}_{n,k}$ correspond to update directions that are uniformly bounded away from being binding.

Consider a current active set $\mathcal{J}$ such that $\{(i,k) : k \in \mathcal{K}, i \in \mathcal{A}_{n,k}\} \cap \mathcal{J} = \emptyset$. We then have $|\sum_{i:(i,k)\notin\mathcal{J}} b_i[k] - 1| \leq n \cdot \exp(-c_1 \cdot n)$ from the summing-to-one constraints of our regression model (14) and the assumption of approximate sparsity (15). In addition, Lemma 1 implies, as in Part 1 of the proof of Proposition 2: $\max_{i\in\mathcal{N}}\max_{m\subseteq\mathcal{K}}||\frac{1}{T}\sum_{t\in\mathcal{T}} x_t^{'}[m]u_{it}||_1 \leq \max_{i\in\mathcal{N}}||\frac{1}{T}\sum_{t\in\mathcal{T}} x_t^{'}u_{it}||_1 = O_p(\sqrt{\frac{\log(n)}{T}})$.

Therefore, (A.4), (A.5), and Assumption 3 can be combined to obtain:

$$|\,\frac{1}{n_k}\sum_{i:(i,k)\notin\mathcal{J}} S_{m_i}^{-1}[k]\gamma[m_i]\,| \leq \chi_N, \tag{A.6}$$

where $S_{m_i}^{-1}[k]$ is the row of $S[m_i]^{-1}$ corresponding to $k \in m_i$, $n_k = |\{i \in \mathcal{N} : (i,k) \notin \mathcal{J}\}|$, and χ_N is a $O_p(\sqrt{\frac{\log(n)}{T}})$ stochastic sequence.

Define $B = \sum_{i\in\mathcal{N}}[\,1[\frac{1}{n_k}S_{m_i}^{-1}[k,k^{'}]\,\text{if}\,(i,k),(i,k^{'}) \notin \mathcal{J}]\,]_{k\in\mathcal{K}}^{k^{'}\in\mathcal{K}}$. We obtain $\gamma \leq \zeta_N$ from (A.6), where ζ_N is a $O_p(\sqrt{\frac{\log(n)}{T}})$ sequence, if the determinant of B is uniformly bounded away from zero.

Define $D = \text{diag}([n_k]_{k\in\mathcal{K}})$ and $B_i = [[S_{m_i}^{-1}[k,k^{'}]\,\text{if}\,(i,k),(i,k^{'}) \notin \mathcal{J},\, 0\,\text{otherwise}]_{k\in\mathcal{K}}^{k^{'}\in\mathcal{K}}$, so that $B = D^{-1}\sum_{i\in\mathcal{N}} B_i$. Note that, from Assumption 3, we have $x^{'}S_{m_i}^{-1}x \geq c \cdot x^{'}x$ for some strictly positive constant $c > 0$ and any vector $x \in \mathbb{R}^{|m_i|}$. This implies that $y^{'} B_i y \geq c \cdot \sum_{k \in m_i} y[k]^2$, for any vector $y \in \mathbb{R}^K$. This further implies that $y^{'} D^{-\frac{1}{2}} B_i D^{-\frac{1}{2}} y \geq c \cdot \sum_{k \in m_i} \frac{1}{n_k} y[k]^2$, and

$$\begin{aligned} y^{'} D^{-\frac{1}{2}} \sum_{i \in \mathcal{N}} B_i D^{-\frac{1}{2}} y &\geq c \cdot \sum_{i \in \mathcal{N}, k \in m_i} \frac{1}{n_k} y[k]^2 \\ &= c \cdot \sum_{k \in \mathcal{K}} \sum_{i:(i,k) \notin \mathcal{J}} y[k]^2 \frac{1}{n_k} \\ &= c \cdot \sum_{k \in \mathcal{K}} y[k]^2 \end{aligned}$$

Therefore the matrix $C = D^{\frac{1}{2}} B D^{-\frac{1}{2}}$ is positive definite with $\det(C) \geq c^{'}$ for some positive constant $c^{'} > 0$. Therefore $\det(B) = \det(D^{-\frac{1}{2}} C D^{\frac{1}{2}}) = \det(C) \geq c^{'}$, so that γ is uniformly bounded above by a $O_p(\sqrt{\frac{\log(n)}{T}})$ sequence.

Therefore, by the first part of the assumption of approximate sparsity (15) and the rate condition $|\mathcal{A}_{n,k}| \sqrt{\frac{\log(n)}{T}} \to 0$, $\min_{i \in \mathcal{A}_{n,k}} \grave{\beta}_i[k]$ is bounded away from zero. This implies that $\beta_i[k] + \lambda(\grave{\beta}_i[k] - \beta_i[k]) > 0 \; \forall \, i \in \mathcal{A}_{n,k}$ since we also had $\beta_i[k] > 0$ for these observations, and $\lambda \in [0, 1]$. Therefore $(i, k) \notin \bar{\mathcal{J}}$ if $i \in \mathcal{A}_{n,k}$, implying that $P(\mathcal{A}_{n,k} \subseteq \hat{\mathcal{A}}_{n,k}) \to 1$.

In addition, the results on γ above imply that $\gamma_N = [\gamma_{N,k}]_{k \in \mathcal{K}} = O_p(\sqrt{\frac{\log(n)}{T}})$ since the Lagrangian multipliers corresponding to the summing-to-one constraint can be obtained from (A.3), taking the positivity constraints that are binding in the final iteration of the algorithm (i.e., such that $\tilde{b}_i[k] = 0$) as given.

A.6.3 Consistent estimation of marginal emissions

It therefore only remains to show that $\tilde{\Delta}_k - \Delta_k = O_p(|\mathcal{A}_{n,k}| \sqrt{\frac{\log(n)}{T}})$. Given the results above, this can be obtained in a similar way as for the univariate estimator in the proof of Proposition 2.

Use the same notation as above for the final iteration of the algorithm, so that $\tilde{b}_i = \beta_i$. Define $\delta_{i,k} = S_{m_i}^{-1}[k] \frac{1}{T} \sum_{t \in \mathcal{T}} x_t^{'}[m_i] u_{it} \; \forall \, i \in \hat{\mathcal{A}}_{n,k}$. Note that, as above, by Lemma 1 we have $\max_{i \in \hat{\mathcal{A}}_n, k \in \mathcal{K}} |\delta_{i,k}| = O_p(\sqrt{\frac{\log(n)}{T}})$. Define $\gamma_{i,k} = S_{m_i}^{-1}[k] \gamma_N$, from Assumption 3 and the result above we have $\max_{i \in \mathcal{N}, k \in \mathcal{K}} |\gamma_{i,k}| = O_p(\sqrt{\frac{\log(n)}{T}})$.

Since $\mathcal{A}_{n,k} \subseteq \hat{\mathcal{A}}_{n,k}$ with probability approaching one, from (A.4), we have w.p.a. 1:

$$\sum_{i \in \mathcal{N}} \tilde{b}_i[k] = \sum_{i \in \mathcal{A}_{n,k}} (b_i[k] + \delta_{i,k} - \gamma_{i,k}) + \sum_{i \in \mathcal{N} \setminus \mathcal{A}_{n,k}} \tilde{b}_i[k].$$

We also have $\sum_{i\in\mathcal{N}} \tilde{b}_i[k] = 1$ and $|\sum_{i\in\mathcal{A}_{n,k}} b_i[k]-1| = O(n\cdot\exp(-c_1\cdot n))$ under approximate sparsity (15), so that

$$\sum_{i\in\mathcal{N}\setminus\mathcal{A}_{n,k}} \tilde{b}_i[k] = O_p(|\mathcal{A}_{n,k}|\sqrt{\frac{\log(n)}{T}}),$$

since $\max_{i\in\mathcal{N},k\in\mathcal{K}}|\delta_{i,k}| = O_p(\sqrt{\frac{\log(n)}{T}})$, $\max_{i\in\mathcal{N},k\in\mathcal{K}}|\gamma_{i,k}| = O_p(\sqrt{\frac{\log(n)}{T}})$, and $n\cdot\exp(-c_1\cdot n) = o(|\mathcal{A}_{n,k}|\sqrt{\frac{\log(n)}{T}})$ under the rate conditions of this result.

Therefore, since $\tilde{b}_i[k] \geq 0$:

$$\begin{aligned}
|\tilde{\Delta}_k - \Delta_k| &= |\sum_{i\in\mathcal{A}_{n,k}} \omega_i(\delta_{i,k} - \gamma_{i,k}) + \sum_{i\in\mathcal{N}\setminus\mathcal{A}_{n,k}} \omega_i\tilde{b}_i[k] - \sum_{i\in\mathcal{N}\setminus\mathcal{A}_{n,k}} \omega_i b_i[k]| \\
&\leq \max_{i\in\mathcal{N}}|\omega_i| \cdot (\sum_{i\in\mathcal{A}_{n,k}} |\delta_{i,k} - \gamma_{i,k}| + \sum_{i\in\mathcal{N}\setminus\mathcal{A}_{n,k}} \tilde{b}_i[k] + \sum_{i\in\mathcal{N}\setminus\mathcal{A}_{n,k}} b_i[k]) \\
&= O_p(|\mathcal{A}_{n,k}|\sqrt{\frac{\log(n)}{T}}),
\end{aligned}$$

where the last equality follows from the results above and approximate sparsity (15). This establishes the desired result.

B Additional Results

B.1 Inconsistency of post-selection OLS

Here we show that, unlike in classical applications of LASSO estimation, our constrained OLS estimator does not yield on consistent post-selection OLS estimation. For post-selection OLS estimation, one would estimate Δ with the estimator $\check{\Delta} = \sum_{i\in\hat{\mathcal{A}}_n} \omega_i\hat{b}_i$, where the selected set $\hat{\mathcal{A}}_n$ was determined by the constrained estimator (6), and $\hat{b}_i$ are the simple plant-specific OLS estimators.

Result. Under the assumptions of Proposition 2 in the main text, it is possible that $\check{\Delta} - \Delta \neq o_p(1)$.

Proof.

Here it suffices to find one data generating process that complies with the conditions of Proposition 2 but yields $\check{\Delta} - \Delta \overset{p}{\not\to} 0$.

Let $u_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$, $x_t = 1\ \forall t$, and $b_i = 0$ if $i \notin \mathcal{A}_n$. Note that $u_{it} \sim N(0,1)$ technically does not satisfy the assumption of bounded support in Assumption 1. However this assumption is only imposed in this form for notational simplicity and could easily be relaxed to an assumption of bounded higher moments that the normal distribution would satisfy.

Part 1. Establish $\gamma_N \leq \sqrt{2\frac{\log(n)-\log(|\mathcal{A}_n|)-\log(\log(n))}{T}}$ with probability approaching one.

To do so, we use the same argument as in the proof of Proposition 2. To shorten notation, define $r_N = \sqrt{2\frac{\log(n)-\log(|\mathcal{A}_n|)-\log(\log(n))}{T}}$. We have:

$$P(\gamma_N \leq r_N) = P(\sum_{i \notin \mathcal{A}_n} 1[e_{i,T} \geq r_N](e_{i,T} - r_N) + \sum_{i \in \mathcal{A}_n}(e_{i,T} - r_N) \leq 0) + o(1).$$

Here we have $\sum_{i \in \mathcal{A}_n}(e_{i,T} - r_N) = O_p(\sqrt{\frac{|\mathcal{A}_n|}{T}}) - |\mathcal{A}_n| r_N$.

Define $T_N = \sum_{i \notin \mathcal{A}_n} 1[e_{i,T} \geq r_N](e_{i,T} - r_N)$. We have $T_N \geq 0$ and

$$\begin{aligned} E(T_N) &= (n - |\mathcal{A}_n|)(\frac{1}{\sqrt{T}}\phi(\sqrt{T}r_N) - \Phi(-\sqrt{T}r_N)r_N), \\ &= \frac{1}{\sqrt{T}}(n - |\mathcal{A}_n|)\frac{1}{\sqrt{2\pi}}\exp(-\frac{(\sqrt{T}r_N)^2}{2})(\frac{1}{(\sqrt{T}r_N)^2} + o(\frac{1}{(\sqrt{T}r_N)^2})), \\ &= O(\frac{1}{\sqrt{T}}\frac{n - |\mathcal{A}_n|}{n}|\mathcal{A}_n|), \end{aligned}$$

where the second equality follows from an expansion of $\phi(x) - \Phi(x)x$ around $x = \infty$.

It follows that $T_N = O_p(\frac{|\mathcal{A}_n|}{\sqrt{T}})$, and:

$$P(\frac{1}{|\mathcal{A}_n| r_N}(T_N + \sum_{i \in \mathcal{A}_n}(e_{i,n} - r_N)) \leq 0) \to 1$$

since $\frac{1}{|\mathcal{A}_n| r_N}(T_N + \sum_{i \in \mathcal{A}_n} e_{i,T}) = o_p(1)$.

Part 2. Establish $\gamma_N \geq 0$ with probability approaching one.

We have

$$P(\gamma_N \geq 0) = P(\sum_{i \notin \mathcal{A}_n} 1[e_{i,T} \geq 0]e_{i,T} + \sum_{i \in \mathcal{A}_n} e_{i,T} \geq 0).$$

For the second term in this decomposition, we have $\sum_{i \in \mathcal{A}_n} e_{i,T} = O_p(\sqrt{\frac{|\mathcal{A}_n|}{T}})$.

The first term is positive and

$$\begin{aligned} \sum_{i \notin \mathcal{A}_n} 1[e_{i,T} \geq 0]e_{i,T} &\geq \sum_{i \notin \mathcal{A}_n} 1[e_{i,T} \geq \frac{c}{\sqrt{T}}]e_{i,T} \\ &\geq \frac{c}{\sqrt{T}} \sum_{i \notin \mathcal{A}_n} 1[e_{i,T} \geq \frac{c}{\sqrt{T}}], \end{aligned}$$

for any positive constant $c > 0$.

We have

$$E(\sum_{i\notin\mathcal{A}_n} 1[e_{i,T} \geq \frac{c}{\sqrt{T}}]) = (n - |\mathcal{A}_n|)\Phi(-c),$$

and

$$Var(\sum_{i\notin\mathcal{A}_n} 1[e_{i,T} \geq \frac{c}{\sqrt{T}}]) \leq (n - |\mathcal{A}_n|)\Phi(-c),$$

so that

$$\frac{c}{\sqrt{T}}\sum_{i\notin\mathcal{A}_n} 1[e_{i,T} \geq \frac{c}{\sqrt{T}}] + \sum_{i\in\mathcal{A}_n} e_{i,T} = c\cdot\frac{n-|\mathcal{A}_n|}{\sqrt{T}}\Phi(-c) + o_p(\frac{n}{\sqrt{T}})$$

and

$$P(\gamma_N \geq 0) \to 1.$$

Part 3. Establish that $\check{\Delta} - \Delta \overset{p}{\not\to} 0$.

Choose a sequence $\omega_i : i = 1, ..., n$ such that $\omega_i \geq c\;\forall\, i$ for a positive constant $c > 0$. With probability approaching one, we have $\check{\Delta} - \Delta = \sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T} \geq \gamma_N]e_{i,T} + \sum_{i\in\mathcal{A}_n}\omega_i e_{i,T}$, where $\sum_{i\in\mathcal{A}_n}\omega_i e_{i,T} = O_p(\sqrt{\frac{|\mathcal{A}_n|}{T}})$.

In addition, $\sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T} \geq \gamma_N]e_{i,T} \geq \sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T} \geq r_N]e_{i,T}$ if $0 \leq \gamma_N \leq r_N$. From parts 1 and 2, we can take $0 \leq \gamma_N \leq r_N$ w.p.a. 1 for $r_N = \sqrt{2\frac{\log(n)-\log(|\mathcal{A}_n|)-\log(\log(n))}{T}}$.

We have:

$$\begin{aligned} E(\sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T} \geq r_N]e_{i,T}) &= \sum_{i\notin\mathcal{A}_n}\omega_i\frac{1}{\sqrt{T}}\phi(\sqrt{T}r_N) \\ &\geq c\cdot\frac{|\mathcal{A}_n|\log(n)}{\sqrt{T}} + o(\frac{|\mathcal{A}_n|}{\sqrt{T}}), \end{aligned}$$

for some positive constant $c > 0$.

We also have

$$\begin{aligned} \mathrm{Var}(\sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T} \geq r_N]e_{i,T}) &= \frac{1}{T}\sum_{i\notin\mathcal{A}_n}\omega_i^2(\Phi(-\sqrt{T}r_N) + \sqrt{T}r_N\phi(\sqrt{T}r_N) - \phi(\sqrt{T}r_N)^2) \\ &\leq C\frac{1}{T}\sum_{i\notin\mathcal{A}_n}\sqrt{T}r_N\phi(\sqrt{T}r_N)\,\text{for}\;\sqrt{T}r_N\;\text{large enough}, \\ &= \frac{C}{\sqrt{2\pi}}\cdot\frac{n-|\mathcal{A}_n|}{T}\sqrt{T}r_N\frac{|\mathcal{A}_n|\log(n)}{n}, \end{aligned}$$

for some finite positive constant $C > 0$, where the inequality follows from $\Phi(-x) \leq x\phi(x)$ for large values of x.

Therefore

$$\sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T}\geq r_N]e_{i,T}-E(\sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T}\geq r_N]e_{i,T})=O_p(\sqrt{\frac{|\mathcal{A}_n|\log(n)^{\frac{3}{2}}}{T}}),$$

but

$$E(\sum_{i\notin\mathcal{A}_n}\omega_i 1[e_{i,T}\geq r_N]e_{i,T})\propto\frac{|\mathcal{A}_n|\log(n)}{\sqrt{T}}.$$

By choosing a sequence $(n,T,|\mathcal{A}_n|)$ such that $\frac{|\mathcal{A}_n|\log(n)^{\frac{3}{2}}}{T}\to 0$ (to obtain convergence to the mean), $|\mathcal{A}_n|\sqrt{\frac{\log(n)}{T}}\to 0$ (to satisfy the conditions of Proposition 2), but $\frac{|\mathcal{A}_n|\log(n)}{\sqrt{T}}\not\to 0$ (so that the mean does not converge to zero), we obtain the desired result that $\tilde{\Delta}-\Delta\overset{p}{\not\to}0$.

C Estimated Marginal Emissions Across all Regions and Hours

In the next two tables we report estimated marginal emissions using our constrained estimator, with bias correction and standard errors, for all regions and hours. We report CO_2 emissions in metric tons/MWh, and local damages in dollar values, obtained using the AP3 model as in the main text. Our period of observation was 2019-2022, summer months only (June-September).

Table 1: Estimated Marginal Emissions (metric tons of CO_2/MWh).

Region	hours (UTC)																	
	1			2			3			4			5			6		
CAL	0.448	-0.004	0.019	0.460	-0.010	0.014	0.484	-0.002	0.013	0.494	-0.001	0.016	0.490	0.015	0.019	0.497	-0.002	0.024
CAR	0.674	0.025	0.036	0.675	-0.013	0.031	0.713	-0.037	0.028	0.750	0.004	0.020	0.700	-0.029	0.034	0.654	-0.009	0.044
CENT	0.620	0.050	0.057	0.508	0.048	0.063	0.419	0.024	0.069	0.391	0.068	0.074	0.294	0.110	0.062	0.269	0.104	0.063
FLA	0.519	-0.055	0.023	0.522	-0.126	0.025	0.509	-0.123	0.029	0.552	-0.109	0.033	0.514	-0.089	0.036	0.523	-0.012	0.031
MIDA	0.685	0.026	0.018	0.695	0.013	0.017	0.691	0.014	0.020	0.675	0.024	0.017	0.670	0.025	0.018	0.643	0.027	0.018
MIDW	0.632	0.003	0.027	0.648	-0.003	0.035	0.659	-0.017	0.040	0.678	-0.031	0.044	0.685	-0.024	0.025	0.702	0.024	0.027
NE	0.550	-0.029	0.040	0.559	-0.106	0.026	0.559	-0.025	0.044	0.546	-0.055	0.034	0.537	-0.007	0.033	0.519	-0.059	0.026
NW	0.475	-0.017	0.032	0.479	-0.026	0.032	0.456	-0.021	0.031	0.437	-0.030	0.028	0.417	-0.034	0.036	0.442	-0.010	0.051
NY	0.548	-0.030	0.041	0.527	0.031	0.037	0.513	-0.071	0.029	0.470	-0.035	0.028	0.507	-0.057	0.038	0.522	-0.027	0.040
SE	0.686	-0.011	0.028	0.691	-0.044	0.028	0.674	-0.009	0.030	0.695	-0.031	0.028	0.676	0.002	0.029	0.652	0.046	0.038
SW	0.502	-0.012	0.051	0.523	-0.004	0.049	0.514	0.002	0.060	0.499	0.026	0.062	0.516	0.019	0.068	0.491	0.048	0.071
TEN	0.667	-0.001	0.045	0.705	-0.067	0.058	0.718	-0.077	0.040	0.623	-0.101	0.050	0.648	-0.085	0.041	0.636	-0.078	0.044
TEX	0.533	-0.006	0.048	0.524	-0.036	0.047	0.510	-0.060	0.068	0.486	-0.102	0.080	0.487	-0.135	0.090	0.521	0.065	0.022
	7			8			9			10			11			12		
CAL	0.503	0.000	0.031	0.516	0.073	0.033	0.524	0.079	0.040	0.522	0.066	0.043	0.530	0.077	0.047	0.544	0.080	0.057
CAR	0.628	-0.045	0.054	0.600	0.002	0.040	0.589	-0.087	0.042	0.576	-0.058	0.036	0.572	0.006	0.053	0.625	0.039	0.043
CENT	0.257	0.062	0.077	0.309	-0.142	0.066	0.386	-0.166	0.099	0.466	-0.115	0.107	0.594	-0.069	0.069	0.613	0.005	0.058
FLA	0.511	-0.011	0.033	0.487	-0.004	0.043	0.464	-0.003	0.039	0.464	-0.058	0.042	0.450	-0.027	0.045	0.433	-0.050	0.035
MIDA	0.618	0.035	0.021	0.595	-0.006	0.022	0.585	0.014	0.019	0.591	-0.018	0.019	0.605	0.014	0.022	0.607	0.031	0.024
MIDW	0.686	0.067	0.028	0.681	0.074	0.020	0.655	0.045	0.021	0.633	0.056	0.019	0.632	0.047	0.025	0.674	0.025	0.024
NE	0.513	0.016	0.030	0.518	-0.023	0.030	0.517	-0.026	0.032	0.517	-0.003	0.038	0.520	-0.074	0.032	0.511	-0.051	0.039
NW	0.448	0.031	0.052	0.378	-0.033	0.057	0.354	-0.076	0.057	0.383	-0.026	0.066	0.370	-0.038	0.071	0.379	-0.056	0.066
NY	0.511	-0.109	0.049	0.507	-0.008	0.057	0.525	0.027	0.049	0.552	0.106	0.060	0.528	0.052	0.061	0.571	-0.064	0.065
SE	0.618	0.044	0.037	0.584	0.048	0.039	0.568	0.083	0.046	0.566	0.076	0.040	0.588	-0.020	0.046	0.612	-0.001	0.041
SW	0.488	0.022	0.075	0.481	-0.038	0.069	0.507	-0.037	0.087	0.499	-0.004	0.097	0.493	0.027	0.107	0.493	0.083	0.108
TEN	0.632	0.011	0.055	0.616	0.145	0.056	0.644	0.020	0.050	0.681	0.015	0.064	0.644	0.021	0.086	0.662	-0.015	0.059
TEX	0.542	0.047	0.020	0.579	0.028	0.019	0.578	0.011	0.019	0.578	-0.002	0.017	0.577	-0.010	0.015	0.570	-0.003	0.016
	13			14			15			16			17			18		
CAL	0.559	0.056	0.050	0.507	0.070	0.048	0.462	0.037	0.050	0.454	0.087	0.049	0.469	0.068	0.039	0.477	0.038	0.038
CAR	0.653	-0.002	0.043	0.696	0.029	0.032	0.730	0.027	0.027	0.682	-0.034	0.030	0.642	-0.038	0.045	0.632	-0.036	0.040
CENT	0.629	0.049	0.075	0.686	0.103	0.062	0.704	0.048	0.052	0.703	0.000	0.047	0.747	0.014	0.050	0.735	0.041	0.043
FLA	0.477	-0.013	0.030	0.490	-0.021	0.038	0.521	-0.023	0.035	0.512	-0.038	0.029	0.499	-0.019	0.022	0.488	-0.042	0.036
MIDA	0.639	0.010	0.021	0.659	0.034	0.020	0.667	-0.006	0.023	0.641	0.013	0.018	0.639	0.032	0.018	0.634	0.060	0.018
MIDW	0.680	0.030	0.029	0.665	-0.010	0.027	0.663	0.002	0.028	0.674	-0.005	0.025	0.658	-0.025	0.017	0.651	-0.042	0.021
NE	0.499	-0.068	0.035	0.511	-0.090	0.041	0.515	-0.092	0.044	0.510	-0.045	0.031	0.520	-0.101	0.052	0.490	-0.079	0.042
NW	0.354	-0.013	0.061	0.332	-0.041	0.037	0.393	-0.013	0.037	0.461	-0.046	0.042	0.471	-0.047	0.050	0.471	-0.030	0.058
NY	0.577	-0.002	0.054	0.586	-0.042	0.037	0.542	0.063	0.059	0.555	0.048	0.037	0.528	0.078	0.038	0.547	0.008	0.038
SE	0.624	0.029	0.037	0.659	0.013	0.032	0.645	-0.033	0.032	0.659	-0.060	0.025	0.640	-0.051	0.037	0.623	-0.089	0.049
SW	0.502	0.105	0.112	0.506	0.098	0.075	0.516	0.155	0.081	0.467	-0.023	0.099	0.417	0.014	0.082	0.405	0.044	0.055
TEN	0.695	-0.148	0.064	0.687	-0.135	0.081	0.632	-0.015	0.056	0.574	-0.036	0.063	0.541	-0.049	0.052	0.556	-0.045	0.052
TEX	0.578	0.002	0.019	0.598	0.018	0.021	0.584	0.047	0.024	0.582	0.052	0.022	0.584	-0.001	0.049	0.582	0.002	0.040
	19			20			21			22			23			24		
CAL	0.468	0.031	0.038	0.468	0.029	0.034	0.456	0.035	0.027	0.448	0.029	0.023	0.437	0.026	0.022	0.434	0.009	0.018
CAR	0.637	-0.030	0.037	0.654	-0.016	0.030	0.650	0.001	0.033	0.659	0.011	0.037	0.671	0.015	0.038	0.674	0.000	0.029
CENT	0.721	0.001	0.044	0.711	0.004	0.042	0.711	0.025	0.030	0.700	0.074	0.037	0.736	0.099	0.075	0.641	0.107	0.044
FLA	0.489	-0.043	0.028	0.483	-0.027	0.033	0.496	-0.017	0.028	0.497	-0.012	0.028	0.499	-0.039	0.027	0.520	-0.047	0.026
MIDA	0.645	0.050	0.020	0.657	0.034	0.016	0.664	0.032	0.016	0.664	0.042	0.018	0.657	0.045	0.018	0.664	0.049	0.020
MIDW	0.638	-0.021	0.022	0.621	-0.015	0.023	0.613	-0.012	0.024	0.620	-0.019	0.030	0.622	-0.021	0.038	0.636	-0.013	0.026
NE	0.458	-0.040	0.030	0.479	-0.044	0.028	0.532	-0.019	0.025	0.558	0.032	0.028	0.562	0.043	0.026	0.566	-0.010	0.033
NW	0.464	-0.023	0.058	0.435	-0.050	0.050	0.417	-0.074	0.046	0.434	-0.063	0.042	0.451	-0.038	0.041	0.457	-0.022	0.040
NY	0.560	-0.037	0.046	0.552	0.039	0.036	0.514	0.050	0.033	0.499	-0.014	0.050	0.529	-0.046	0.044	0.543	-0.073	0.041
SE	0.601	-0.042	0.043	0.592	-0.061	0.040	0.609	-0.055	0.036	0.615	-0.015	0.028	0.610	-0.026	0.052	0.643	-0.033	0.027
SW	0.406	0.063	0.075	0.464	0.090	0.072	0.479	0.049	0.064	0.477	0.115	0.069	0.488	0.107	0.054	0.504	0.095	0.067
TEN	0.583	-0.036	0.069	0.606	0.040	0.047	0.660	0.022	0.041	0.686	-0.028	0.041	0.681	-0.008	0.051	0.676	-0.020	0.045
TEX	0.575	0.018	0.010	0.571	0.018	0.010	0.571	0.018	0.009	0.562	0.028	0.009	0.549	-0.001	0.032	0.539	-0.011	0.032

Point estimates, estimated bias correction, and standard error. Hours are in Coordinated Universal Time (UTC). Period of observation: 2019-2022, summer months only (June-September).

Table 2: Estimated Marginal Emissions (local damages, U.S. $/MWh).

Region	hours (UTC)																	
		1			2			3			4			5			6	
CAL	3.65	0.16	0.28	3.58	0.08	0.26	3.65	-0.03	0.24	3.68	0.19	0.26	3.70	0.27	0.34	3.78	-0.26	0.52
CAR	15.56	1.01	1.91	15.96	-1.73	1.64	17.90	-1.83	1.42	20.54	-1.51	1.15	18.37	0.91	1.64	16.75	2.41	1.80
CENT	10.94	0.98	1.39	9.32	0.13	1.61	7.94	-0.55	1.49	7.21	0.31	1.73	5.27	1.35	1.46	5.04	0.64	1.40
FLA	6.23	-1.36	0.99	6.86	-2.25	0.91	5.60	-4.65	0.91	9.85	-6.19	1.30	5.72	-3.15	1.68	5.66	-1.13	0.97
MIDA	16.84	1.54	1.15	17.29	1.40	1.11	17.47	1.62	1.23	17.09	1.94	1.05	17.35	1.79	1.31	15.77	1.67	1.14
MIDW	11.72	0.24	1.10	11.89	-0.33	1.18	12.35	-1.57	1.19	13.36	-2.68	1.34	13.89	-1.43	1.05	14.27	0.91	0.99
NE	8.46	-1.30	2.35	8.27	-6.57	1.83	9.20	-3.47	2.74	9.30	-3.81	2.10	8.48	-1.42	1.62	7.31	-3.61	1.50
NW	5.28	-0.43	0.62	5.30	-0.37	0.66	4.98	-0.11	0.59	4.76	-0.37	0.46	4.50	-0.08	0.55	4.99	0.19	0.73
NY	11.79	-0.50	2.85	8.98	4.53	2.94	7.21	-2.06	2.65	5.96	-2.00	1.52	7.07	-4.38	3.58	8.20	-1.48	3.04
SE	9.20	-1.38	1.35	8.94	-2.99	1.14	7.79	-1.90	1.19	7.60	-2.08	1.03	7.47	-1.74	1.25	7.29	-1.45	1.50
SW	4.40	-0.13	0.95	4.70	0.08	0.86	4.41	0.09	0.88	4.35	0.34	0.88	4.39	0.60	0.86	4.40	0.85	0.86
TEN	14.10	-0.22	1.97	15.60	-3.05	2.29	15.31	-4.11	1.85	13.47	-3.73	1.62	13.55	-3.10	1.66	13.04	-0.42	1.94
TEX	5.30	0.23	0.80	5.18	0.11	0.80	5.31	-0.18	1.09	5.34	-0.48	1.27	5.42	-0.86	1.36	5.86	1.00	0.70
		7			8			9			10			11			12	
CAL	4.04	-0.19	0.56	4.32	0.83	0.58	4.71	0.96	0.73	4.94	0.93	0.70	5.17	1.10	0.86	5.59	1.15	1.04
CAR	13.93	-0.25	2.30	12.54	2.58	1.81	12.39	-1.04	1.73	11.91	-0.89	1.89	11.71	-1.40	2.76	14.17	2.55	2.32
CENT	5.01	0.45	1.31	6.03	-6.03	1.60	8.08	-6.80	2.07	9.95	-4.18	2.53	12.15	-3.63	2.31	12.33	-2.13	1.85
FLA	6.12	-1.23	0.88	5.46	-1.32	0.90	3.73	-1.08	1.00	5.04	-2.83	1.88	4.77	-1.99	2.03	3.52	-3.08	1.32
MIDA	14.29	1.95	1.28	13.05	0.37	1.27	12.79	1.18	1.16	13.28	0.36	0.99	14.06	0.71	1.13	13.74	1.73	1.48
MIDW	13.67	2.81	0.94	13.42	3.14	0.89	12.51	1.65	1.06	11.63	1.70	0.97	11.50	1.70	1.34	13.85	0.72	1.30
NE	6.68	0.32	1.83	6.79	-2.49	1.92	6.67	-4.07	1.72	6.72	-3.78	1.97	6.56	-4.82	1.44	5.69	-4.04	1.87
NW	5.12	0.77	0.83	4.64	-0.33	0.79	4.39	-1.04	0.76	4.64	-0.45	0.94	4.20	-0.60	1.09	4.10	-0.72	1.03
NY	7.37	-5.97	3.50	7.14	-1.03	3.86	8.17	2.28	3.43	10.25	4.61	3.56	9.27	4.72	3.36	13.20	-2.88	4.32
SE	6.71	-0.45	1.34	5.95	-1.56	2.36	6.51	-0.60	2.64	6.93	0.17	2.51	6.48	0.11	1.84	6.34	-1.61	1.67
SW	4.41	0.22	1.01	4.13	-0.92	0.92	4.54	-0.25	1.09	4.79	0.35	1.18	4.91	0.72	1.36	4.73	1.49	1.52
TEN	11.95	1.20	2.43	10.61	5.32	2.11	10.95	2.63	2.09	12.14	0.80	2.85	11.31	0.82	3.54	11.52	2.91	2.39
TEX	5.92	0.59	0.71	6.15	0.63	0.79	6.05	0.07	0.78	6.01	-0.53	0.68	5.97	-0.77	0.59	5.88	-0.50	0.66
		13			14			15			16			17			18	
CAL	5.63	0.90	1.01	5.39	1.06	0.84	4.83	0.37	0.75	4.78	0.87	0.74	5.03	0.35	0.66	5.02	0.04	0.69
CAR	15.42	1.30	2.11	17.41	2.35	1.59	18.72	1.17	1.54	17.85	-1.88	1.57	16.92	-3.17	2.00	16.15	-1.95	1.99
CENT	11.94	-0.98	2.05	12.87	1.03	2.07	12.87	-1.06	2.59	11.95	-2.30	2.01	12.88	-2.56	1.86	12.29	-1.82	1.52
FLA	5.16	-1.59	1.53	5.73	-1.71	1.52	6.39	-1.07	1.38	6.26	-2.11	1.31	5.73	-2.24	0.85	5.28	-2.44	0.90
MIDA	15.00	0.54	1.54	17.07	1.64	1.15	17.58	0.74	1.25	16.80	2.29	0.89	15.41	4.00	0.86	15.83	4.18	0.83
MIDW	14.31	1.39	1.48	13.40	-1.10	1.52	12.52	-1.24	1.52	12.14	-0.92	1.64	12.08	-1.86	1.25	10.96	-1.31	0.95
NE	4.80	-5.27	2.21	6.11	-7.62	2.68	6.41	-6.76	2.99	6.27	-4.77	2.19	6.28	-4.13	2.01	6.40	-5.04	1.84
NW	4.07	-0.50	1.05	3.71	-1.19	0.76	4.37	-0.65	0.62	5.35	-0.95	0.76	5.32	-0.39	0.95	5.10	-0.55	1.07
NY	13.18	1.72	4.02	12.74	-2.45	2.57	10.01	1.40	3.67	9.48	1.32	2.25	12.78	-4.18	1.91	11.69	-2.27	1.77
SE	6.98	-0.80	1.58	8.23	-1.00	1.24	8.13	-2.58	1.29	8.35	-1.71	1.17	7.76	-0.83	1.16	7.50	-2.05	1.56
SW	4.81	1.58	1.48	4.48	1.33	1.14	4.42	1.81	1.26	3.76	-0.21	1.34	3.49	0.29	1.23	3.47	0.73	1.00
TEN	13.33	-5.21	2.73	13.33	-4.84	3.65	12.01	1.40	2.35	11.87	-4.87	3.40	11.36	-4.17	2.91	11.84	-3.21	2.94
TEX	6.15	-0.29	0.73	6.70	0.36	0.81	6.55	1.12	0.93	6.56	1.07	0.88	6.13	0.10	0.92	5.71	0.00	0.77
		19			20			21			22			23			24	
CAL	4.57	0.05	0.65	4.45	0.17	0.56	4.09	0.27	0.51	3.83	0.32	0.43	3.61	0.64	0.35	3.55	0.28	0.27
CAR	15.98	-1.39	1.97	15.94	-0.63	1.59	15.56	0.06	1.65	15.55	1.34	1.89	15.96	1.54	2.08	15.87	0.34	2.05
CENT	11.69	-0.99	1.44	11.72	-0.23	1.44	11.86	0.20	1.32	11.83	1.00	1.13	12.19	1.59	1.75	10.80	1.66	1.18
FLA	5.41	-2.42	0.71	5.60	-1.89	0.90	5.92	-1.14	0.97	5.76	-0.87	1.01	5.56	-1.72	0.98	5.99	-1.09	1.08
MIDA	15.66	4.06	1.09	15.75	2.26	0.91	15.39	2.23	0.96	15.42	2.46	1.02	15.33	2.26	1.02	15.65	3.56	1.28
MIDW	10.86	-0.91	0.92	10.53	-0.87	0.78	10.42	-0.60	0.85	10.69	-0.45	1.11	11.25	-0.08	1.41	11.92	-0.23	1.48
NE	5.84	-2.98	1.87	6.77	-2.18	1.77	8.65	-1.65	1.50	9.37	-0.20	1.70	9.29	-0.38	1.34	9.57	-2.03	1.46
NW	4.98	-0.41	1.03	4.54	-0.86	0.91	4.49	-1.09	0.88	4.74	-0.85	0.83	5.12	-0.85	0.70	5.28	-0.37	0.71
NY	13.00	-4.95	2.80	14.01	1.52	2.63	13.72	1.11	2.44	13.18	-1.52	2.73	13.71	-1.56	2.50	12.37	-5.27	2.46
SE	7.22	-1.40	1.32	7.24	-2.42	1.16	7.86	-2.50	1.15	7.94	-2.21	1.15	7.94	-3.14	1.37	8.46	-2.88	1.09
SW	3.79	0.59	1.02	4.23	0.66	1.48	4.38	0.45	1.09	4.16	1.37	1.27	4.13	0.60	0.88	4.33	1.00	0.99
TEN	12.57	-3.04	3.12	13.08	2.12	2.65	15.08	0.65	2.39	15.64	-0.42	2.58	15.58	-1.07	2.68	14.94	-2.14	2.28
TEX	5.45	0.10	0.47	5.37	0.10	0.45	5.39	0.06	0.38	5.31	0.15	0.34	5.24	-0.16	0.40	5.22	-0.24	0.41

Point estimates, estimated bias correction, and standard error. Hours are in Coordinated Universal Time (UTC). Period of observation: 2019-2022, summer months only (June-September).