Estimating Demand Elasticity with Many Instruments: a Machine Learning Approach

Vadim Gorski Sebastian Schwenen*

January 20, 2022

Abstract

In empirical analyses with large datasets and endogenous regressors, the choice of instruments becomes essential to estimate the effect of interest. We explore the use of machine learning tools to select among potential instruments and to improve inference in two-stage IV estimation. We focus on the canonical two-stage demand estimation problem and rely on high-frequency data from double auctions, where we observe submitted demand curves. Observing "true" demand curves allows us to infer how accurately the proposed regularization-based IV methods approximate true elasticity. We find that regularization in the first stage selects the most relevant instruments and significantly improves inference in the second stage. We derive suggestions for researchers using machine learning tools in IV models with many instruments.

Keywords: Instrumental Variable Estimation, Regularization, Demand Estimation.
JEL-Classification: C13, C36, D12.

*Technical University of Munich, School of Management, vadim.gorski@tum.de, sebastian.schwenen@tum.de.

1 Introduction

Estimating demand elasticity is a classic problem in economics. However, researchers often face difficulties in arriving at good estimates, either because of too much or too little data or due to the necessity to find reasonable instruments. In the worst case, estimates become unreliable and introduce bias in policy-relevant counterfactuals.

Instrumental variable (IV) estimation presents one of the key approaches to estimating demand elasticity. The seminal literature on demand estimation using IV concentrates on cases with scarce data, i.e., only a few available instruments and observations (Angrist and Krueger, 2001). Recently, the increasing availability of large datasets with a high frequency of observations has encouraged novel research in empirical economics and applications of IV, with its own set of obstacles in handling large volumes of data.

In case of scarce data, researchers find it hard to identify the true effect of interest because of potential bias in the local average treatment effect. In case of too much data, researchers struggle to find the right instruments that are best suited to reduce bias. As economies and marketplaces become increasingly computerized, market participants generate vast amounts of data with many possible instrumental variable candidates. Paramount examples are digitized marketplaces such as Amazon or Uber, which have gained growing attention for the estimation of demand and consumer surplus (Bajari et al., 2015; Cohen et al., 2016).

In this article, we study the usefulness of machine learning approaches for demand estimation in high dimensional, high-frequency settings. We make use of the seminal two-stage IV model and estimate demand elasticity from equilibrium prices, quantities, and numerous supply shifters. As well known, bad instruments correlated with the error term can possibly worsen the bias as compared to using ordinary least-squares (OLS). In large datasets, the availability of many, possibly unfit, instruments increases the risk of misidentification and over-fitting, again introducing bias. This risk becomes even more prominent in cases with a large number of instruments but a low amount of observations.

The statistics and machine learning literature has proposed a host of approaches to guide the variable selection and estimation process for the problem at hand.[1] While these methods are originally designed for prediction, a growing literature explores their usefulness for inferring of parameters of underlying statistical and economic models (Belloni et al., 2014; Fessler and Kasy, 2019).

In our study, we apply machine learning techniques to the seminal IV demand estimation problem and explore their usefulness for improving inference. In particular, we investigate how first-stage regularization performs in selecting suitable instruments, and how this selection affects the second-stage estimates. Specifically, we focus on LASSO, RIDGE, and Post-LASSO estimation (Hoerl and Kennard, 1970; Tibshirani, 1996) of the first-stage model. We follow an applied perspective on settings with many instruments, in our case supply shifters, and apply different IV model formulations to the canonical demand elasticity estimation problem (e.g., Angrist et al., 2000).

Our data stem from computerized double auctions in wholesale electricity markets. This dataset provides a rich test environment for our study as it entails revealed demand curves submitted to the double auction. Observing demand curves, i.e., the object of interest, enables us to compute the "true" underlying demand elasticity. This, in turn, allows us to evaluate the different estimates from IV regression based on equilibrium prices, quantities, and instrumental variables. Knowing the true elasticity furthermore allows us to infer how well different IV models with and without machine learning approximate the true elasticity.

Our findings show that regularization in the first-stage significantly reduces bias and mean squared error (MSE) of the estimator. This is, we find that in situations with many

[1] See, e.g., Varian (2014) and Athey (2018) for comprehensive reviews and implications for empirical economics.

instrument candidates for IV estimation, the application of regularization in the first stage dominates standard two-stage models. As our results show, regularization is especially valuable under certain conditions. First, we find that regularization improves inference when instruments are correlated or exhibit outliers. Second, we find that regularization is effective in addressing the weak instrument problem in that it aids the instrument selection process and avoids over-identification. It is commonly known that selecting the seemingly best instrument requires economic reasoning and an understanding of the underlying economic mechanism (Angrist and Krueger, 2001). Yet, using several instruments at once, even if theoretically sensible, can introduce additional bias. In this regard, we find that combining economic reasoning with machine learning-based instrument selection presents a data-driven answer to this trade-off.

Our findings relate to several strands of research. First, we contribute to the emerging literature on the estimation of demand and consumer surplus with large data and machine learning tools (Bajari et al., 2015; Cohen et al., 2016). While Bajari et al. (2015) use machine learning to predict the treatment effect of price promotions on demand, our approach instead aims at improving the traditional IV demand estimation procedure. Although we conduct our analysis within the demand estimation framework, we believe it is widely applicable to a broader set of IV models with large datasets.

Second, we relate to the literature on parameter estimation in sparse settings. While researchers typically apply regularization to settings with a large number of variables but low amount of observations (Belloni et al., 2012; Carrasco, 2012), we show that even in cases with a lower amount of instruments and a higher amount of observations, regularization techniques still provide better estimates.

Third, we relate to the long-lasting literature on demand estimation, which has been applied to a variety of problems and settings, in particular in the field of industrial organization (e.g., Genesove and Mullin, 1998; Berry et al., 1995). More specifically, as our data stems from power markets, we relate to the literature on the estimation of demand elasticity in energy markets (Baumeister and Peersman, 2013; Kilian and Murphy, 2014; Coglianese et al., 2017) and in particular electricity markets (Lijesen, 2007; Bönte et al., 2015; Boogen et al., 2017), where estimating elasticity has become of high interest for policy evaluation (e.g., Aigner et al., 1994; Aubin et al., 1995; Fabra et al., 2021). As we show, adding machine learning tools to the traditional two-stage IV design significantly improves estimates and, as such, can also be widely used for studying policy counterfactuals that require information on the demand side.

Last and more broadly, we add to the growing literature at the intersection between applied economics and econometrics on the use of machine learning for causal inference. In a recent paper, Abadie and Kasy (2019) show that data-driven regularization yields estimates that are close to estimators obtained for optimal levels of regularization. We draw from this idea but apply regularization to select among many instruments within a two-stage IV setting, which has previously been studied in Okui (2011) and Belloni et al. (2012). Similar works develop methods for high-dimensional instrumental variable settings, including Zou (2006), Belloni et al. (2014), and Carrasco and Tchuente (2015). These works mostly concentrate on the derivation of asymptotic properties of the estimators. Our focus is on the application of these methods and the investigation of how these estimators perform “in the field”.

The remainder is organized as follows. Section 2 introduces our theoretical model, notation, and our empirical strategy. Section 3 describes the dataset. Section 4 presents our findings and results. Section 5 concludes.

2 Model and empirical strategy

This section first outlines a brief market model. We then illustrate how we use the model to measure and estimate demand elasticity. Our setup and notation draw from Angrist et al. (2000), whose setting we amend by adding regularization techniques.

Consider a set of n competitive markets in which buyers and sellers exchange a perfectly homogeneous good.[2] Each single market $i = 1, \ldots, n$ is subject to supply shocks from k supply shifters that we denote as z_j with $j = 1, ..., k$. We write the set of all supply shifters as $Z = [z_1, \ldots, z_k]$. The market supply function hence depends on the realization of Z and can be written as $S_i(p, Z)$. In equilibrium, realized supply and demand determine a market-clearing price-quantity pair $\{p_i^*(Z), q_i^*(p_i^*(Z))\}$.

We denote the demand function in market i as $Q_i(p)$. Furthermore, we assume that demand is stochastic and has a zero mean random component so that $Q_i(p) = Q_i(p, \xi)$ with $\mathbb{E}[\xi|p] = 0$. The error term ξ may contain any perturbations due to unobserved consumer behavior in market i.

As the good is traded over n markets, we measure demand elasticity, denoted by ϵ, as the average elasticity over all n markets, formally,

$$\epsilon = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathbb{E}[Q_i(p, \xi)]}{\partial p} \frac{p}{Q_i} = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i, \tag{1}$$

where ϵ_i is the elasticity in market i. The average demand elasticity defined in equation (1) is the object of interest that we seek to estimate.

If the realization of the demand curve in market i is observable, say as submitted demand curve in a double auction, we can estimate (1) from data on Q_i and p, e.g., by fitting a log-log specification around the observed demand curve $Q_i(p)$. Typically, researchers however do

[2] An alternative interpretation to having n markets that trade homogeneous goods is that the same market clears n times. In our empirical application, markets clear hourly, implying that each market $i \in [1, ..., n]$ represents one hour of a particular day.

not observe demand curves. In this case, empirical specifications have to rely on observed equilibrium price-quantity pairs $\{p_i^*(Z), q_i^*(p_i^*(Z))\}$ and instruments (supply shifters), Z, to approximate the true demand elasticity.

2.1 Estimating true demand from bid curves

To estimate "true" underlying elasticities from observed demand curves, we start from equation (1) and assume that demand is iso-elastic. Then, the demand function can be written as

$$Q_i = \alpha_i p^{\epsilon_i}. \tag{2}$$

Note that in our setting, market i refers to a specific market clearing in hour i. The demand elasticity in hour i is then directly estimable by using data points along the revealed aggregate demand curve and applying a logarithmic specification

$$\log(Q_i) = \log(\alpha) + \epsilon_i \log(p) + \xi, \tag{3}$$

where ϵ_i is a point estimate of the elasticity in market i as in equation (1). In essence, the above fits an iso-elastic curve around the observed demand function in market i. We apply equation (3) to all hours $i = 1, ..., n$ in our sample, compute the underlying demand elasticities, and calculate the corresponding average elasticity as shown in equation (1). As this estimate is obtained directly from using demand curve data, we refer to this estimate as "true" elasticity and denote it by ϵ^{true} henceforth. Similarly, we use ϵ_i^{true} to refer to the true elasticity for one market i.

2.2 IV Estimation

In most applied cases, demand curves $Q_i(p)$ are unobservable. The frequently used approach, in this case, is to apply the two-stage least squares (2SLS) estimator and first approximate the observed equilibrium price, p_i^*, given instruments that typically represent exogenous shifts in market supply. Using a log-log specification,

$$log(p_i^*) = log(\delta) + log(Z)\pi + X\gamma + \eta, \tag{4}$$

where π and γ are first-stage coefficients, δ is an intercept, Z are supply shifters that cause exogenous shocks to the equilibrium price, p_i^*, $X = [x_1, \ldots, x_m]$ represents further controls, and η is the first-stage error term. In our application below, the matrix X comprises hourly, daily, weekly and monthly dummies.[3] As standard, instruments are assumed to be relevant and independent, i.e., each supply shifter $z_j \in Z$ is uncorrelated with the second-stage error term, but correlated with the equilibrium price.

Importantly, the fitted values $\tilde{p}_i^*$ depend on the choice of instruments Z and the choice of weights π. Below, we employ machine learning methods to estimate (4) and study how they affect the fitted values $\tilde{p}_i^*$ and, in turn, the elasticity as approximated by the second stage equation

$$q_i^* = \alpha + p_i^*\epsilon^* + X\beta + \xi, \tag{5}$$

where β represents coefficients of all exogenous control variables in X that capture observable heterogeneity other than price. Using this 2SLS approach based on data on equilibrium price-quantity pairs, the estimate for the demand elasticity is then represented by ϵ^*.

[3] Note that for ease of notation we use p^* and q^* to denote both, the theoretical equilibrium price-quantity tupels and the underlying data.

2.3 Regularized IV Estimation

Instead of fitting equilibrium market prices as in equation (4), we apply different machine learning techniques. First, we apply the L1 penalized estimator (LASSO) in the first stage, which we denote as π_L. To compute the latter, we solve

$$\pi_L \in \underset{\pi \in \mathbb{R}^{k+m}}{\arg\min}\, \mathbb{E}[(\log(p_i^*) - (\log(Z), X)\pi)^2] + \frac{\lambda}{n} \left\|\pi\right\|_1, \tag{6}$$

where λ is chosen such that the 10-fold cross-validated mean squared error is minimal. Note that π_L has $k + m$ components, as we regress on both, k instruments and m exogenous variables and controls X in the first stage.

Following the same notation, the L2 penalized first-stage estimator (RIDGE) is defined as

$$\pi_R \in \underset{\pi \in \mathbb{R}^{k+m}}{\arg\min}\, \mathbb{E}[(\log(p_i^*) - (\log(Z), X)\pi)^2] + \frac{\lambda}{n} \left\|\pi\right\|_2^2, \tag{7}$$

where again the regularization intensity λ is chosen to minimize MSE using 10-fold cross-validation.

Last, we construct the Post-LASSO estimator π_{PL} by first performing a LASSO-penalty regularization in the first stage and then taking all non-zero coefficients to estimate the second stage equation using OLS,

$$\pi_{PL} \in \underset{\pi \in \mathbb{R}^{k+m}}{\arg\min}\, \mathbb{E}[(\log(p_i^*) - (\log(Z), X)\pi)^2] : \pi_j = 0\ \forall\ j : \pi_j^L = 0. \tag{8}$$

The idea for this estimator is similar to Belloni et al. (2012), who also apply LASSO as a pre-selection procedure. In all three models above, the regularization parameter λ, if positive, introduces first-stage bias as compared to the standard 2SLS case with $\lambda = 0$. As λ increases, so do the costs of large instrument coefficients when minimizing the sum of squared errors.

The above setup allows for obtaining several elasticity estimators ϵ^* for the second-stage model. First, using fitted values for p_i^* with $\lambda = 0$ yields the standard two-stage least squares estimator, which we denote as ϵ^*_{2ls}. In addition, we obtain three regularized IV estimators when using fitted values from (6), (7), and (8), i.e., using LASSO, RIDGE or Post-LASSO penalties. We denote the second-stage estimator based on a RIDGE first-stage as ϵ^*_R, the second-stage estimator based on LASSO regression as ϵ^*_L, and the Post-LASSO second-stage estimator as ϵ^*_{PL}.

2.4 Comparing estimates

Finally, to make our various two-stage estimates comparable with those derived from averaging the observed demand elasticities, we use samples that cover identical time spans. This is, when we average true observed demand elasticities over a set of n markets, we apply the corresponding n equilibrium prices, quantities, and instruments to our two-stage procedures. To be able to compare our results for a range of different samples, we furthermore partition our data in different sub-samples ω, where each sub-sample ω contains n_ω observations. For each sub-sample ω, we thus obtain different two-stage estimators ϵ^*_ω and one true elasticity ϵ^{true}_ω by averaging observed elasticities.

Hence, when looking at, say, the weekly average elasticity, the sample ω consists of all hours belonging to a given calendar week. For example, we fit iso-elastic demand curves for each hour of every calendar week 20 in the sample and obtain ϵ^{true}_ω for this particular calendar week. Subsequently, we use all hourly price-quantity pairs and instruments from the same calendar week 20 and obtain the corresponding IV estimates. This procedure guarantees that we compare estimators obtained from identical samples, but over a range of different samples. Our main analysis is performed at the weekly level. For additional tests, we also pool markets at the hourly level and, e.g., compute elasticities for all hours from 8 am to 9 am of a given year.

In sum, our empirical approach works as follows. First, we estimate demand elasticity based on observed demand curves as in equation (1), which provides our benchmark estimate for the true underlying elasticity. Next, we estimate elasticity using equilibrium price-quantity pairs following the canonical IV approach as in equations (4) and (5), as well as the regularized approach as in equations (6), (7), and (8). We then pool the observations to construct identical samples and time spans for each estimator as described above and compare estimates. The goal is to provide insights on the usefulness of the employed machine learning techniques for inference in the second stage of the model. We focus on bias, variance, and mean squared error.

3 Data

For our empirical investigation, we exploit data from the European Power Exchange (EPEX). The EPEX is the largest European power exchange and clears, amongst other commodities, day-ahead wholesale markets for electricity in several European countries. More specifically, we employ EPEX data from the German day-ahead electricity market.

We make use of this rich power market data for several reasons. First, electricity is a homogeneous good and allows us to abstract from competing differentiated products. Second, our setup offers a well-defined test environment for demand estimation, because aggregate demand functions, and hence true underlying demand elasticities, are directly observable in the data. We therefore can benchmark estimated elasticities against observed demand curves. Third, we observe the equilibrium price-quantity pairs at hourly granularity, which allows for computing two-stage IV estimates with a large number of observations. Finally, there exist data on several plausible instruments, e.g., we observe weather-dependent generation from wind and solar sources in the German market. Wind and solar energy act as supply shifters and increase overall market supply for high wind speed and high solar radiation levels, while market supply is significantly lower at times of little wind and solar output.

Market clearing in the German day-ahead power market is organized as a sealed bid double auction, that determines a market-clearing price for every delivery hour of the consecutive day. To do so, the exchange matches submitted supply and demand schedules for any hour of the following day, and announces the corresponding hourly equilibrium prices and quantities. Our main data comprise the aggregated hourly market demand curves, including the corresponding equilibrium market prices and quantities. We collect these hourly data for a period of ten years from 2011 to 2020.

In addition to the submitted demand curves, equilibrium prices, and quantities, our setup also requires a range of instruments. The instruments that we use are solar and wind output forecasts which are again available in hourly resolution. Because solar and wind output have marginal costs close to zero, both constitute exogenous supply shifters. Finally, we merge in data on daily input prices for natural gas, coal, and prices for CO_2 allowances. All latter are inputs to producing electricity and can be considered as exogenous. In total, we hence make use of $k = 5$ first-stage instruments.

Because we use solar output as an instrument, we restrict our sample to 12 daytime hours between 8 am and 8 pm .[4] We also exclude weekend hours that typically follow different demand patterns. Table 1 presents summary statistics for our main variables. As shown, the data entail 31,308 hourly observations on clearing prices and quantities.[5] Rows three and four of Table 1 show statistics for solar and wind output. Last, Table 1 shows costs of thermal generation and carbon prices, which are at the daily level and amount to 2,870 observations.

[4]We note that our results remain robust regardless of this choice. In particular, leaving out solar feed-in as an instrument and considering hours during the night yields similar results.

[5]There are no missing values for prices and quantities so that the number of observations covers all peak hours during working days from January 1, 2011, up to December 31, 2020. Further, note that because power is costly to store, power prices can be negative in case of oversupply. We work with logged variables and keep negative prices in the sample by transforming prices with the function $\text{sign}(p)\log(1 + |p|)$. This transformation also significantly improves the fit as compared to leaving negative prices out of the logged sample. Our results on the performance of each estimator remain unchanged when deleting negative prices from the sample instead.

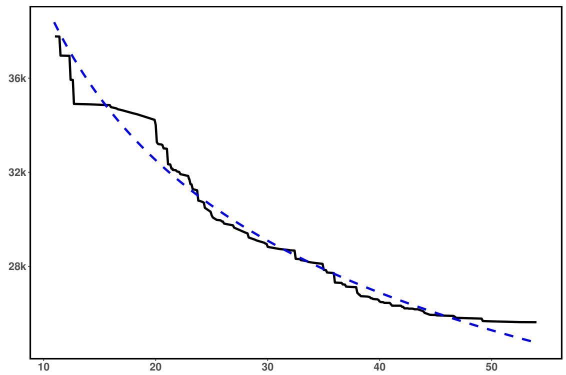
Table 1: Summary statistics

	Mean	St. Dev.	Min	Max	Obs.
Price (EUR/MWh)	45.22	16.19	-83.94	210.00	31,308
Quantity (GW)	30.25	5.09	16.84	51.47	31,308
Wind output (GW)	9.07	8.23	0.22	47.09	31,212
Solar output (GW)	7.47	7.09	0	32.48	30,960
Gas price (EUR/MWh)	19.37	5.69	3.63	29.35	2,870
Coal price (EUR/metric ton)	9.07	2.07	5.08	14.41	2,870
CO_2 price (EUR/metric ton)	11.60	7.82	2.75	33.44	2,870

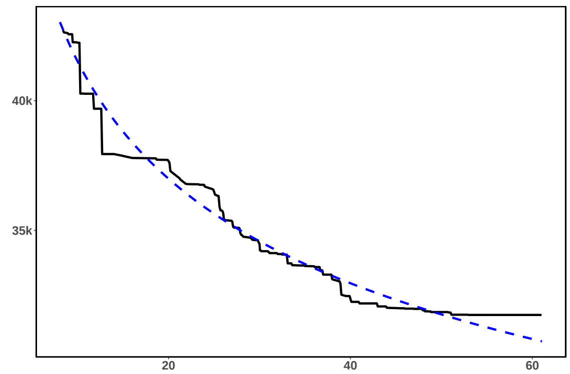
Notes: Prices and quantities are for day-ahead electricity at the German market, traded at EPEX. Gas prices are day-ahead prices for the TTF hub. Coal prices are for imported coal to Europe (API2-CIF), CO_2 prices are from the ICE exchange. We observe gas, coal and CO_2 prices in daily resolution.

While Table 1 summarizes the data that we use for our two-stage IV estimations, Figure 1 depicts representative demand curves that we use to compare our elasticity estimates against. Specifically, Figure 1 shows four hourly demand curves from our sample, along with their corresponding "true" elasticity estimate ϵ_i^{true} as estimated by equation (3). As can be seen, the assumption of iso-elastic demand fits the data well. Across all ten years of our sample, we find an average demand elasticity of -0.198. This estimate is in line with extant literature on demand elasticity in the energy market (e.g., Lijesen, 2007). We also find that the step-wise demand functions are well-approximated by an iso-elastic fit with an average squared approximation error of 2.24% within our sample.[6]

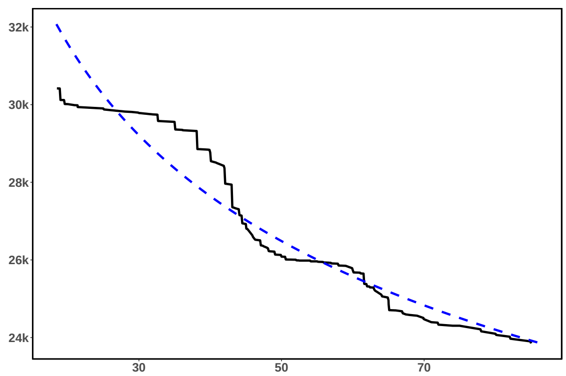
[6]Note that market participants can announce demand for prices between -500 and 3000 EUR/MWh, while Figure 1 plots demand curves for the typical range of market prices, which is far narrower. To avoid fitting demand curves for prices way out of sample, we estimate the true elasticities only within the range of observed clearing prices. Appendix 5 presents this approach in more detail.



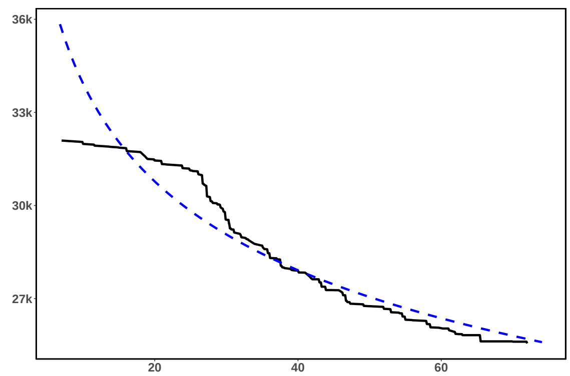
(a) Hour 9, September 1^{st}, 2016



(b) Hour 15, March 16^{th}, 2017



(c) Hour 8, August 10^{th}, 2018



(d) Hour 10, July 11^{th}, 2019

Figure 1: Submitted demand curves (quantity in thousands of MW over price in EUR/MWh) and elasticity ϵ_i^{true}. Part (a) of this figure shows the observed demand curve for the hour 9-10am on September 1^{st}, 2016. The dashed line shows the fitted iso-elastic demand curve with elasticity ϵ_i^{true} that we obtain when estimating the log-log specification in Equation (3). Parts (b), (c), and (d) show demand curves and corresponding log-log fits for three further representative hours.

4 Results

We compare and report the performance of estimators based on the bias, variance, and MSE. Specifically, for each sub-sample ω, we define and compute the sample analogs of bias, MSE, and standard deviation as

$$Bias(\epsilon_\omega^*) = \epsilon_\omega^{true} - \epsilon_\omega^*, \tag{9}$$

$$MSE(\epsilon_\omega^*) = \sum_i [(\epsilon_\omega - \epsilon_\omega^{true})^2], \tag{10}$$

$$\sigma(\epsilon_\omega^*) = \sqrt{\frac{\sum_i(\epsilon_\omega - \epsilon_\omega^{true})^2}{n_\omega}}, \tag{11}$$

where n_ω denotes the total number of market clearing hours i in sub-sample ω. In addition, we investigate how the regularized methods choose instrument weights and which instrument coefficients are set to zero by the LASSO estimator.

We start by presenting results for estimating demand at the weekly level. This is, we pool hourly market results for each of the 52 calendar weeks over the entire time span of the dataset. For each of these weekly samples, we then estimate ϵ_ω^{true} and $\{\epsilon^*_{2ls,R,L,PL}\}$. As we have ten years of data, the estimates are hence based on ten weeks of data for each sample.

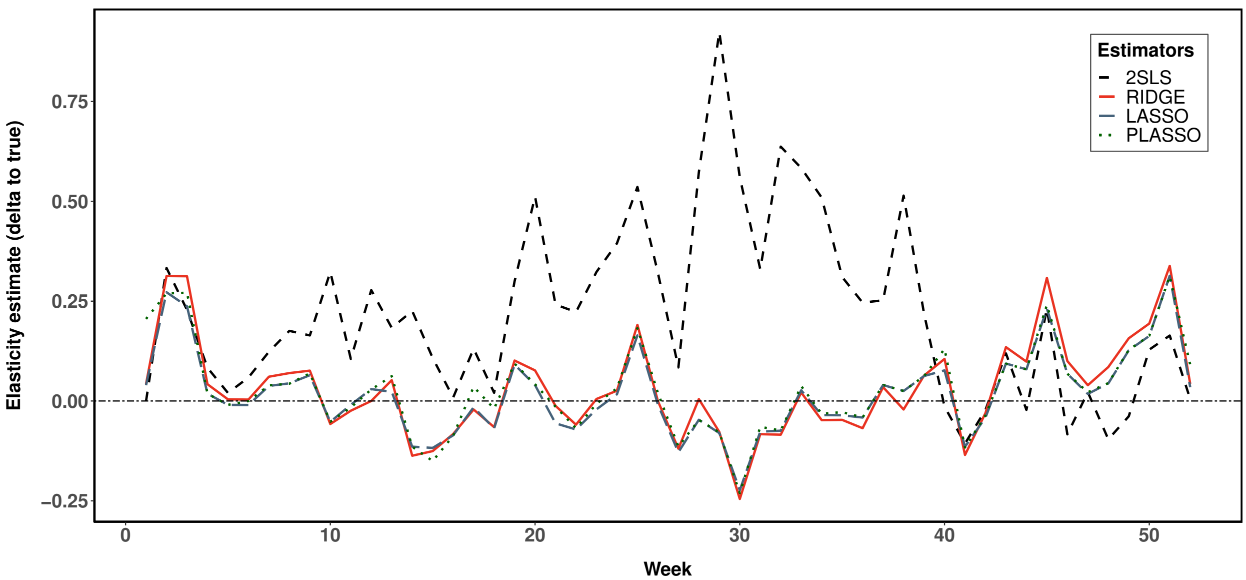


Figure 2: Bias of two-stage estimators (2SLS, RIDGE, LASSO, Post-LASSO). This figure shows the bias of each estimator for each of the 52 calendar weeks. The bias is shown relative to the true demand elasticity for the respective week. Note that we plot the bias as stated in equation (9), but swap the estimated elasticity and the true elasticity. This allows for a more natural interpretation of the plot, as positive values represent an overestimation while negative ones represent an underestimation.

Figure 2 displays the bias over all calendar weeks for all considered estimators. As shown, the bias reduces considerably, and estimates vary much less for the regularized first-stage models. In particular, the relative over-performance of regularized estimators is especially pronounced from around week 10 to week 40. This is due to several reasons. First, the high variance of solar energy during summertime often leads to relatively large outliers. Second, we find that fuel prices exhibit outliers during the first part of the year surrounding week 10. Overall, this leads to poorer estimates for the two-stage least squares procedure, while all regularized estimators are adjusting instrument coefficients accordingly and, in turn, depict more accurate point estimates.[7] As a result, all three of the regularized estimators show lower bias and return more stable results.

Furthermore, we find that the lower bias of regularized IV estimators relates to the collinearity of instruments. Indeed, we compute simple variance inflation factors (VIF) of up to 8 in our instruments data, showing clear signs of high collinearity. In particular, collinearity is relevant for the fuel prices among our set of instruments, which are typically highly correlated. We believe this contributes to the performance of the regularized estimators, as shown in Figure 2, since collinearity is more pronounced in those weeks when the overperformance is higher.

To probe into results for the complete set of metrics for all considered estimators, Table 2 reports the mean bias, mean squared error, and variance of each estimator. As shown, over all metrics, the estimation quality improves considerably for our regularized IV estimators, and two-stage estimates based on LASSO perform best.

To investigate how the observed improvements are related to instrument selection, we look into which instruments are being selected over time by the regularized estimators. In case of LASSO and Post-LASSO, this can be done by screening for instruments with coefficients of zero. We find that in our weekly models, the instruments coal, gas, and CO_2 prices are set to zero in 6%, 20%, and 14% of cases, respectively, while wind and solar output remain relevant instruments almost always. Taken together, we find that instrument selection is relevant even for only five instruments, and that the improved inference especially is due to

[7] When applying Cook's distance on the set of instruments, we find that, on average, around 3% of observations are outliers.

Table 2: Bias, mean squarer error, and variance of two-stage estimators over 52 weeks

Estimator	Bias	MSE	Variance	Observations
ϵ^*_{2ls}	0.220	0.427	0.218	31,094
ϵ^*_{R}	0.029	0.088	0.104	31,094
ϵ^*_{L}	0.018	0.064	0.109	31,094
ϵ^*_{PL}	0.030	0.087	0.108	31,094

Notes: Bias, Variance and MSE are computed as in equations (9), (10), and (11). To compute each metric, we use 31,094 observations on true demand curves and estimated demand elasticities. The difference in observations as compared to the data in Table 1 results from, first, missing demand curves for some hours and, second, estimators that do not yield an elasticity estimate if too much instrument data is missing.

exploiting the high-frequency variation in wind and solar output.

Furthermore, we find that the relevance of the instruments changes over time. During our sample, wind and solar energy show strong investment and market entry. We, therefore, conjecture that the explanatory power of the forecasted renewable energy in-feed should increase over time. Table 3 reports how often each instrument is selected in the first stage over the years. We find that while wind feed-in is kept in all models, the role of solar feed-in in explaining first-stage price variance is increasing over time. Also, the CO_2 price gains more explanatory power over time. In contrast, coal and gas prices become less relevant over time.

In sum, the findings of this section show that using statistical learning in the first stage significantly outperforms traditional two-stage estimators. Importantly, while regularization is known to outperform unregularized estimators, particularly in cases of small numbers of observations compared to the number of explanatory variables (Belloni et al., 2012), our findings imply that these improvements hold even for a relatively small number of potential instruments and high-frequency data with many observations. Furthermore, our results above show that the benefits of selecting instruments in the first stage carry through to

Table 3: Instrument relevance over time: percentage of models which use the corresponding instrument in the first stage after LASSO-regularization.

Year	Solar	Wind	Coal price	Gas price	CO_2 price
2011	0.83	1.00	0.58	0.83	0.75
2012	0.92	1.00	0.83	0.83	0.83
2013	0.92	1.00	0.75	0.83	0.83
2014	0.92	1.00	0.50	0.91	0.58
2015	1.00	1.00	0.50	0.58	0.91
2016	1.00	1.00	0.66	0.41	0.58
2017	1.00	1.00	0.50	0.50	0.91
2018	0.92	1.00	0.50	0.66	0.91
2019	1.00	1.00	0.50	0.58	0.91
2020	1.00	1.00	0.50	0.58	0.91

Notes: For each year, we consider how many estimation models per month are using the corresponding instrument variable after regularizing the first-stage and express this amount as percentage.

the second-stage estimate, which improves considerably in bias, variance, and mean squared error.

4.1 Alternative samples

For robustness, we re-run our analysis by pooling the data by hours instead of weeks, and again compare ϵ_{ω}^{true} and $\{\epsilon_{2ls,R,L,PL}^{*}\}$ where each sample ω now contains all data from each hour of the day, i.e., for each daytime hour from 8 am to 8 pm. Hence, we obtain different estimators for 12 different samples, where each sample collects all data for one particular hour of the day during our 10-year sample.

Figure 3 shows the results and displays the difference between the true elasticity and our estimates based on price-quantity pairs. Again we find that the traditional IV approach performs worst as compared to all regularized estimators. The average bias of the 2SLS estimator is around -0.095, while the average bias for the regularized estimators is between

-0.041 and -0.043. Overall, Figure 3 corroborates our finding that regularized estimators perform significantly better in estimating demand elasticities.

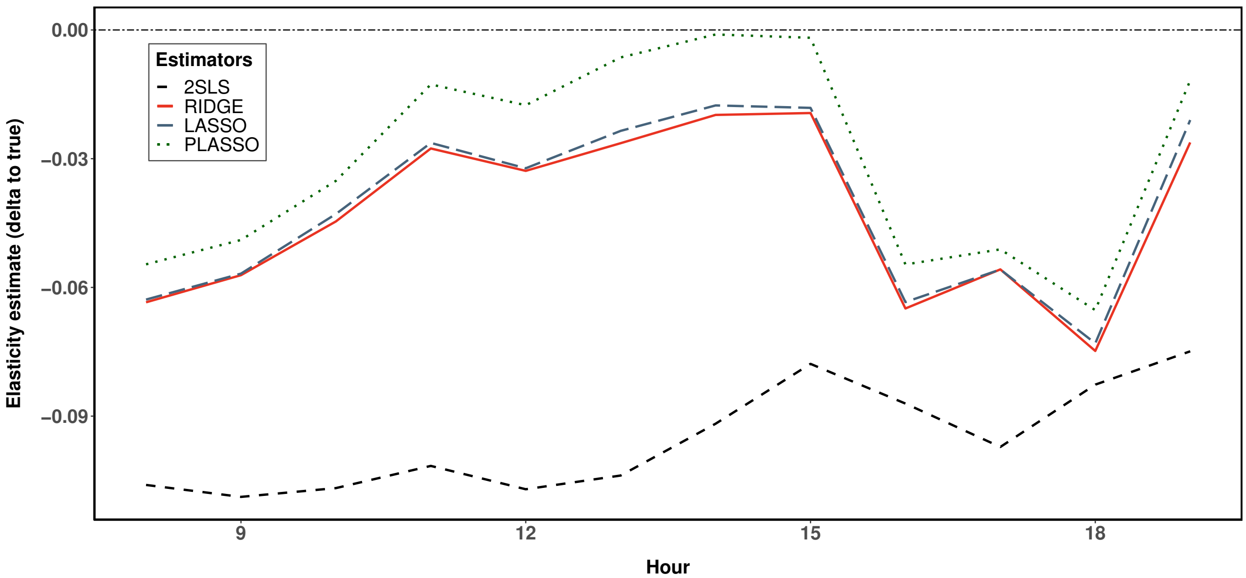


Figure 3: Bias of two-stage estimators (2SLS, RIDGE, LASSO, Post-LASSO). This figure shows the bias of each estimator for each of the 12 peak hours. The bias is shown relative to the true demand elasticity for the respective hour.

4.2 Sensitivity of shrinkage levels

In closing, we discuss the sensitivity of our results with respect to the level of regularization. We find that the performance of regularized estimators can drop significantly when the level of applied shrinkage becomes too high. Without the possibility to benchmark against true elasticities in the second stage, the task of choosing the shrinkage level becomes non-trivial. Indeed, we find that the standard 10-fold cross-validation with respect to the first-stage MSE as implemented in most statistical packages yields too high regularization levels in about 20% of the cases. As a result, in these cases relaxing the level of shrinkage yielded a better second-stage estimate. Our results therefore stress that all regularized estimators outperform the two-stage least squares approach even for below optimal shrinkage.

5 Conclusion

The increasing availability of large datasets has led to the growing prominence of statistical learning methods in empirical economics. While these methods have originally been developed to improve out-of-sample prediction and MSE, they are increasingly used to improve inference.

In this article, we make use of a unique empirical setting to explore the usefulness of statistical learning methods for instrumental variable estimation. We apply regularization techniques, i.e., LASSO, RIDGE, and Post-LASSO to the canonical two-stage demand estimation problem and show that, with many instruments, regularization in the first stage improves inference significantly. Our empirical approach makes use of observed demand curves, allowing us to benchmark obtained estimators against true underlying demand elasticities.

Using this empirical strategy, we derive a set of results on the properties of RIDGE, LASSO, and Post-LASSO estimators. We find that bias, variance, and MSE of the second-stage coefficients can be significantly reduced when employing regularization. We find that the bias reduction property is more prominent, for higher variance, for a higher number of outliers, and for higher collinearity of instruments.

In addition, we investigate how the level of shrinkage affects the results. We find that the estimators are sensitive with regard to the selection of the shrinkage parameter. We argue that "over-shrinking" is a concern, as the quality of estimates can drop significantly when it occurs. Yet, our results show that regularization outperforms traditional IV estimation for a range of shrinkage levels.

In sum, our findings suggest that regularization methods offer measurable advantages for IV estimation with large data and many instrument candidates, with LASSO-type regularization performing best on our data.

References

Abadie, Alberto and Maximilian Kasy (2019) "Choosing among regularized estimators in empirical economics: The risk of machine learning," *Review of Economics and Statistics*, Vol. 101, pp. 743–762.

Aigner, Dennis J, J Newman, and Asher Tishler (1994) "The response of small and medium-size business customers to Time-of-Use (TOU) electricity rates in Israel," *Journal of applied econometrics*, Vol. 9, pp. 283–304.

Angrist, Joshua D, Kathryn Graddy, and Guido W Imbens (2000) "The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish," *The Review of Economic Studies*, Vol. 67, pp. 499–527.

Angrist, Joshua D and Alan B Krueger (2001) "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *The Journal of Economic Perspectives*, Vol. 15, pp. 69–85.

Athey, Susan (2018) "The impact of machine learning on economics," in *The Economics of Artificial Intelligence: An Agenda*: University of Chicago Press.

Aubin, Christophe, Denis Fougere, Emmanuel Husson, and Marc Ivaldi (1995) "Real-time pricing of electricity for residential customers: Econometric analysis of an experiment," *Journal of Applied Econometrics*, Vol. 10, pp. S171–S191.

Bajari, Patrick, Denis Nekipelov, Stephen P Ryan, and Miaoyu Yang (2015) "Machine learning methods for demand estimation," *American Economic Review*, Vol. 105, pp. 481–85.

Baumeister, Christiane and Gert Peersman (2013) "The role of time-varying price elasticities in accounting for volatility changes in the crude oil market," *Journal of Applied Econometrics*, Vol. 28, pp. 1087–1109.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012) "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, Vol. 80, pp. 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014) "High-dimensional methods and inference on structural and treatment effects," *Journal of Economic Perspectives*, Vol. 28, pp. 29–50.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995) "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Bönte, Werner, Sebastian Nielen, Niyaz Valitov, and Torben Engelmeyer (2015) "Price elasticity of demand in the EPEX spot market for electricity—New empirical evidence," *Economics Letters*, Vol. 135, pp. 5–8.

Boogen, Nina, Souvik Datta, and Massimo Filippini (2017) "Dynamic models of residential electricity demand: Evidence from Switzerland," *Energy Strategy Reviews*, Vol. 18, pp. 85–92.

Carrasco, Marine (2012) "A regularization approach to the many instruments problem," *Journal of Econometrics*, Vol. 170, pp. 383–398.

Carrasco, Marine and Guy Tchuente (2015) "Regularized LIML for many instruments," *Journal of Econometrics*, Vol. 186, pp. 427–442.

Coglianese, John, Lucas W Davis, Lutz Kilian, and James H Stock (2017) "Anticipation, tax avoidance, and the price elasticity of gasoline demand," *Journal of Applied Econometrics*, Vol. 32, pp. 1–15.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe (2016) "Using big data to estimate consumer surplus: The case of uber,"Technical report, National Bureau of Economic Research.

Fabra, Natalia, David Rapson, Mar Reguant, and Jingyuan Wang (2021) "Estimating the Elasticity to Real Time Pricing: Evidence from the Spanish Electricity Market," *American Economic Association Papers and Proceedings forthcoming.*

Fessler, Pirmin and Maximilian Kasy (2019) "How to use economic theory to improve estimators: Shrinking toward theoretical restrictions," *Review of Economics and Statistics*, Vol. 101, pp. 681–698.

Genesove, David and Wallace P Mullin (1998) "Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914," *The RAND Journal of Economics*, pp. 355–377.

Hoerl, Arthur E and Robert W Kennard (1970) "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, Vol. 12, pp. 55–67.

Kilian, Lutz and Daniel P Murphy (2014) “The role of inventories and speculative trading in the global market for crude oil,” *Journal of Applied econometrics*, Vol. 29, pp. 454–478.

Lijesen, Mark G (2007) “The real-time price elasticity of electricity,” *Energy economics*, Vol. 29, pp. 249–258.

Okui, Ryo (2011) “Instrumental variable estimation in the presence of many moment conditions,” *Journal of Econometrics*, Vol. 165, pp. 70–86.

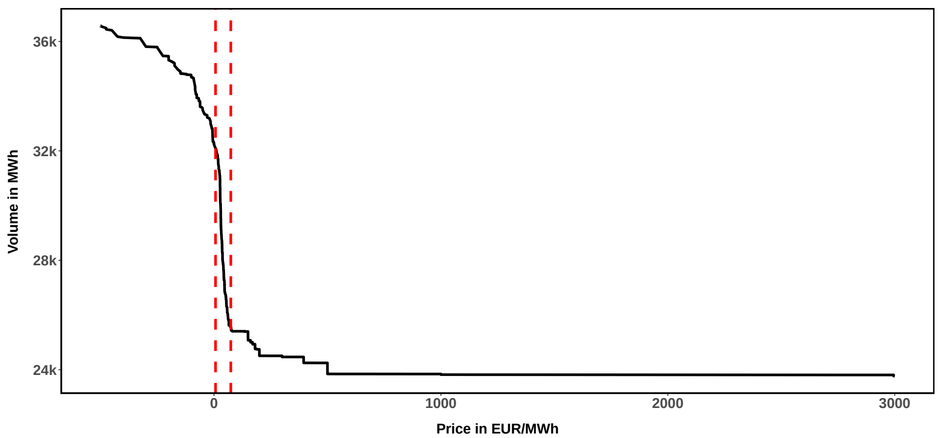
Tibshirani, Robert (1996) “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Varian, Hal R (2014) “Big data: New tricks for econometrics,” *Journal of Economic Perspectives*, Vol. 28, pp. 3–28.

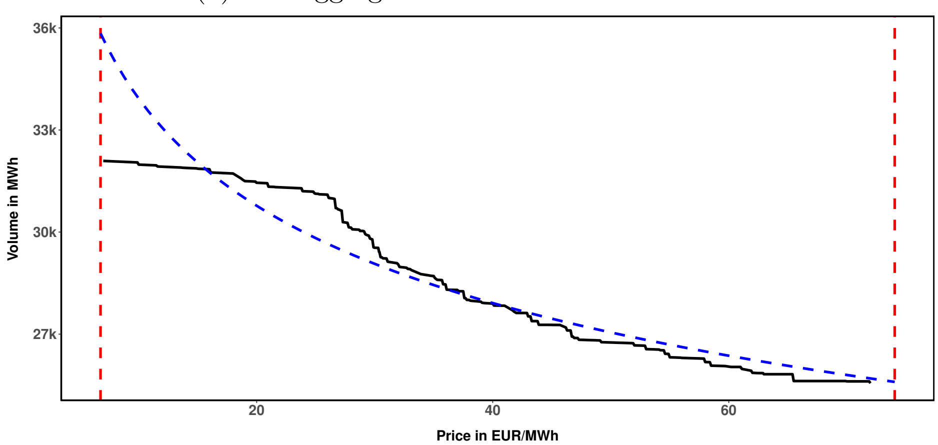
Zou, Hui (2006) “The adaptive lasso and its oracle properties,” *Journal of the American Statistical Association*, Vol. 101, pp. 1418–1429.

Appendix: Demand curve sampling

To avoid fitting demand out of the sample, we apply the log-log specification in equation (3) only to the part of the demand curve between the minimum and the maximum observed clearing price for the corresponding month. If a demand curve falls in, say, July, we only estimate the elasticity for this curve for prices ranging between the minimum and maximum price observed in July. Figure A.1 depicts a selected demand curve for the entire possible price range and the relevant part of the demand that we use for our estimations.



(a) Full aggregated curve of submitted bids



(b) Relevant part of the curve (depicted in panel (a) between the dashed red lines)

Figure A.1: Quantity (y-axis) over price (x-axis) of a complete demand curve for a representative hour (panel a), along with a zoomed-in version of the relevant part of the curve (panel b). The demand curve is for hour 10 of July 11th, 2019. The maximum equilibrium price in July was 74.06 EUR/MWh, while the minimum price was 26.72 EUR/MWh.