Do We Exploit all Information for Counterfactual Analysis? Benefits of Factor Models and Idiosyncratic Correction

Jianqing Fan, Ricardo Masini & Marcelo C. Medeiros

Taylor & Francis
Taylor & Francis Group

Check for updates

Do We Exploit all Information for Counterfactual Analysis? Benefits of Factor Models and Idiosyncratic Correction

Jianqing Fan[a], Ricardo Masini[b,c], and Marcelo C. Medeiros[d]

[a]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ; [b]Center for Statistics and Machine Learning (CSML), Princeton University, Princeton, NJ; [c]Sao Paulo School of Economics (EESP), Getulio Vargas Foundation, Sao Paulo, Brazil; [d]Department of Economics, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil

ABSTRACT

Optimal pricing, that is determining the price level that maximizes profit or revenue of a given product, is a vital task for the retail industry. To select such a quantity, one needs first to estimate the price elasticity from the product demand. Regression methods usually fail to recover such elasticities due to confounding effects and price endogeneity. Therefore, randomized experiments are typically required. However, elasticities can be highly heterogeneous depending on the location of stores, for example. As the randomization frequently occurs at the municipal level, standard difference-in-differences methods may also fail. Possible solutions are based on methodologies to measure the effects of treatments on a single (or just a few) treated unit(s) based on counterfactuals constructed from artificial controls. For example, for each city in the treatment group, a counterfactual may be constructed from the untreated locations. In this article, we apply a novel high-dimensional statistical method to measure the effects of price changes on daily sales from a major retailer in Brazil. The proposed methodology combines principal components (factors) and sparse regressions, resulting in a method called Factor-Adjusted Regularized Method for Treatment evaluation (`FarmTreat`). The data consist of daily sales and prices of five different products over more than 400 municipalities. The products considered belong to the *sweet and candies* category and experiments have been conducted over the years of 2016 and 2017. Our results confirm the hypothesis of a high degree of heterogeneity yielding very different pricing strategies over distinct municipalities. Supplementary materials for this article are available online.

1. Introduction

The evaluation of treatment effects on a single (or just a few) treated unit(s) based on counterfactuals (i.e., the unobservable outcome had there been no intervention) constructed from artificial controls has become a popular practice in applied statistics since the proposal of the synthetic control (SC) method by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010). Usually, these artificial (synthetic) controls are built from a panel of untreated peers observed over time, before and after the intervention.

The majority of methods based on artificial controls relies on the estimation of a statistical model between the treated unit(s) and a potentially large set of explanatory variables coming from the peers and measured before the intervention. The construction of counterfactuals poses a number of technical and empirical challenges. Usually, the dimension of the counterfactual model to be estimated is large compared to the available number of observations and some sort of restrictions must be imposed. Furthermore, the target variables of interest are nonstationary. Finally, conducting inference on the counterfactual dynamics is not straightforward. Although the original work by Abadie and Gardeazabal (2003) is able to handle some of these challenges,

a number of extensions has been proposed; see Doudchenko and Imbens (2016), Athey and Imbens (2017), or Abadie (2021) for recent discussions. Motivated by an application to the retail industry where the optimal prices of different products have to be determined, we develop a new methodology to construct counterfactuals which nests several other methods and efficiently explores all the available information. Our proposed method is well-suited for both stationary and nonstationary data as well as for high- or low-dimensional settings.

1.1. Heterogeneous Elasticities and Optimal Prices

The determination of the optimal price of products is of great importance in the retail industry. By optimal price we mean the one that either maximizes profit or revenue. To determine such quantity we need first to estimate the price elasticity from the demand side. This is not an effortless task as standard regression methods usually fail to recover the parameter of interest due to confounding effects and the well-known endogeneity of prices.

Our novel dataset consists of daily prices and quantities sold of five different products for a major retailer in Brazil, aggregated at the municipal level. The company has more than 1400 stores distributed in over 400 municipalities, covering all the states of

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

the country.[1] The chosen products differ in terms of magnitude of sales and in importance as a share of the company's total revenue. The overarching goal is to compute optimal prices at a municipal level via counterfactual analysis. Our method determines the effects in sales due to price changes and provides demand elasticities estimates which will be further used to compute the optimal prices.

To determine the optimal price of each of the products, a randomized controlled experiment has been carried out. More specifically, for each product, the price was changed in a group of municipalities (treatment group), while in another group, the prices were kept fixed at the original level (control group). The magnitudes of price changes across products range from 5% up to 20%. Furthermore, for three out of five products, the prices were increased, and for the other two products there was a price decrease. The selection of the treatment and control groups was carried out according to the socioeconomic and demographic characteristics of each municipality as well as to the distribution of stores in each city. Nevertheless, it is important to emphasize three facts. First, we used no information about the quantities sold of the product in each municipality, which is our output variable, in the randomization process. This way, we avoid any selection bias and can maintain valid the assumption that the intervention of interest is independent of the outcomes. Second, although according to municipality characteristics, we keep a homogeneous balance between groups, the parallel trend hypothesis is violated, and there is strong heterogeneity with respect to the quantities sold and consumer behavior in each city, even after controlling for observables. This implies that price elasticities are quite heterogeneous and optimal prices can be remarkably different among municipalities. Finally, there are a clear seasonal pattern in the data as well as common factors affecting the dynamics of sales across different cities.

Our results confirm the heterogeneous patterns in the intervention effects, yielding different elasticities and optimal prices across municipalities. In addition, the impacts also differ across products. Overall, the effects of price changes are statistical significant in more than 20% of the municipalities in the treatment group and the optimal prices in terms of profit maximization are usually below the actual ones. Therefore, we recommend that the optimal policy in terms of profit maximization is to change the prices in the cities where the effects were statistically significant. Further experiments may be necessary to evaluate the effects of price changes in the cities were it was not possible to find statistically significant results.

1.2. Methodological Innovations

Driven by the empirical application discussed in the previous subsection, this paper proposes a methodology that includes both principal component regression (factors) and sparse linear regression for estimating counterfactuals for better evaluation of the effects on the sales of a set of products after price changes. It does not impose either sparsity or approximate sparsity in the mapping between the peers and the treated by using the information from hidden but estimable idiosyncratic compo-

nents. Furthermore, we show that when the number of postintervention observations is fixed, tests like the ones proposed in Masini and Medeiros (2021) or Chernozhukov, Wüthrich, and Zhu (2020), can be applied. Finally, we also consider a high-dimensional test to answer the question whether the use of idiosyncratic component actually leads to better estimation of the treatment effect. Our framework can be applied to much broader context in prediction and estimation and hence we leave more abstract and general theoretical developments to a different paper (Fan, Masini, and Medeiros 2021).

The proposed method consists of four steps, called FarmTreat. In the first one, the effects of exogenous (to the intervention of interest) variables are removed, for example, heterogeneous deterministic (nonlinear) trends, seasonality and other calendar effects, and/or known outliers. In the second step, a factor model is estimated based on the residuals of the first-step model. The idea is to uncover a common component driving the dynamics of the treated unit and the peers. This second step is key when relaxing the sparsity assumption. To explore potential remaining relation among units, in the third step a LASSO regression model is established among the residuals of the factor model, which are called the idiosyncratic components in the factor model. Sparsity is only imposed in this third step and it is less restrictive than the sparsity assumption in the second step. Note that all these three steps are carried out in the preintervention period. Finally, in the fourth step, the model is projected for the postintervention period under the assumption that the peers do not suffer the intervention. Inspired by Fan, Ke, and Wang (2020), we call the methodology developed here FarmTreat, the factor-adjusted regularized method for treatment evaluation.

The procedure described above is well suited either for stationary data or in the case of deterministic nonlinear and heterogeneous trends. In case of unit-roots, the procedure should be carried out in first-differences under the assumption that factors follow an integrated process (with or without drift). In this case, our result follows from sec. 7 in Bai and Ng (2008). After the final step, the levels of both the target variable and the counterfactual can be recovered and the inference conducted.

We show that the estimator of the instantaneous treatment is unbiased. This result enables the use of residual ressampling procedures, as the ones in Masini and Medeiros (2021) or Chernozhukov, Wüthrich, and Zhu (2020), to test hypotheses about the treatment effect without relying on any asymptotic result for the postintervention period. The testing procedures proposed in Masini and Medeiros (2021) or Chernozhukov, Wüthrich, and Zhu (2020) are similar with the crucial difference that the first paper considers models estimated just with the preintervention sample, while the second paper advocates the use of the full sample to estimate the models. As shown by the authors and confirmed in our simulations, using the full data yields much better size properties in small samples.

We believe our results are of general importance for the following reasons. First and most importantly, the sparsity or approximate sparsity assumptions on the regression coefficients do not seem reasonable in applications where the cross-dependence among all units in the panel are high. In addition, due to the cross-dependence, the conditions needed for the consistency of LASSO or other high-dimensional regularization

---

[1] Due to a confidentiality agreement, we are not allowed to disclosure either the name of the products or the name of the retail chain.

methods are violated (Fan, Ke, and Wang 2020). Second, first filtering for trends, seasonal effects and/or outliers seem reasonable in order to highlight the potential intervention effects by removing uninformative terms. Finally, modeling remaining cross-dependence among the treated unit and a sparse set of peers are also important to gather all relevant information about the correlation structure about the units.

Under the hypothesis that the treatment is exogenous which is standard in the synthetic control literature, we have an unbiased estimator for the treatment effect on the treated unit for each period after the intervention. In the case the treatment is exogenous with respect only to the peers, we can identify the effects of a specific intervention on the treated unit, that is, the time of a single intervention is fully known. This might be the quantity of interest in several macroeconomic applications as, for instance, the effects of Brexit on the UK economy fixing the date of the event.

We conduct a simulation study to evaluate the finite-sample properties of the estimators and inferential procedures discussed in the article. We show that the proposed method works reasonably well even in very small samples. Furthermore, as a case study, we estimate the impact of price changes on product sales by using a novel dataset from a major retail chain in Brazil with more than 1400 stores in the country. We show how the methods discussed in the paper can be used to estimate heterogeneous demand price elasticities, which can be further used to determine optimal prices for a wide class of products. In addition, we demonstrate that the idiosyncratic components do provide useful information for better estimation of elasticities.

1.3. Comparison to the Literature

Several papers in the literature extend the original SC method and derive estimators for counterfactuals when only a single unit is treated. We start by comparing with Carvalho, Masini, and Medeiros (2018). Differently from this article, we do neither impose sparsity nor our results are based on preintervention and postintervention asymptotics. We just require the preintervention sample to diverge in order to prove our results. Furthermore, by combing a factor structure with sparse regression we relax the (weak) sparsity assumption on the relation between the treated unit and its peers. In addition, we allow for heterogeneous trends which may not be bounded as in the case of the aforementioned article; for a similar setup to Carvalho, Masini, and Medeiros (2018), see Li and Bell (2017). Masini and Medeiros (2021, 2022) consider a synthetic control extension when the data are nonstationary, with possibly unit-roots. However, the former paper imposes weak-sparsity on the relation between the treated unit and the peers and the later only handles the low-dimensional case. The low-dimensional nonstationary case is discussed in many other papers. See, for example, Hsiao, Ching, and Wan (2012), Ouyang and Peng (2015), Du and Zhang (2015), and Li (2020), among many others.

Compared to Differences-in-Differences (DiD) estimators, the advantages of the many estimators based on the SC method are three folds. First, we do not need the number of treated units to grow. In fact, the workhorse situation is when there is a single treated unit. The second, and most important difference, is that our methodology has been developed for situations where the

$n-1$ untreated units may differ substantially from the treated unit and cannot form a control group, even after conditioning on a set of observables. For instance, in the application in this paper, the dynamics of sales in a specific treated municipality cannot be perfectly matched by any other city exclusively. On the other hand, there may exist a set of cities where the combined sales are close enough to ones of the treated unit in the absence of the treatment. Another typical example in the literature if to explain the gross domestic product (GDP) of a specific region by a linear combination of GDP from several untreated regions; see Abadie and Gardeazabal (2003). Finally, SC methods and their extensions are usually consistent even without the parallel trends hypothesis.

More recently, Gobillon and Magnac (2016) generalize DiD estimators by estimating a correctly specified linear panel model with strictly exogenous regressors and interactive fixed effects represented as a number of common factors with heterogeneous loadings. Their theoretical results rely on double asymptotics when both T (sample size) and n (number of peers) go to infinity. The authors allow the common confounding factors to have nonlinear deterministic trends, which is a generalization of the linear parallel trend hypothesis assumed when DiD estimation is considered. Our method differs from Gobillon and Magnac (2016) in a very important way as we consider cross-dependence among the idiosyncratic units after the common factors have been accounted for.

Finally, we should compare our results with Chernozhukov and Wüthrich and Zhu (2021b). Chernozhukov, Wüthrich, and Zhu (2020) propose a general conformal inference method to test hypotheses on the counterfactuals which can be applied to our model setup as discussed above. When the sample size is small we strongly recommend the use of the approach described in Chernozhukov, Wüthrich, and Zhu (2020) to conduct inference on the intervention effects. Chernozhukov, Wüthrich, and Zhu (2021a) propose a very nice generalization of Carvalho, Masini, and Medeiros (2018) with a new inference method to test hypotheses on intervention effects under high dimensionality and potential nonstationarity. However, their approach differs from ours in three aspects. First, and more importantly, their results are based on both preintervention and postintervention samples diverging. Second, their inferential procedure is designed to test hypothesis only on the average effect. Our procedure can be applied to a wide class of hypothesis tests. Finally, they impose that exactly the same (stochastic) trend is shared among all variables in the model. This is a more restricted framework than the one considered here.

1.4. Organization of the Paper

The rest of the article is organized as follows. We give an overview of the proposed method and the application in Section 2. We present the setup and assumptions in Section 3 and state the key theoretical result in Section 3.2. Inferential procedures are presented in Section 3.3. We present the results of a simulation experiment in Section 5. Section 4 is devoted to provide guidance to practitioners and a discussion of the empirical application can be found in Section 6. Section 7 concludes the article. Finally, the proof of our theoretical

result and additional empirical results are relegated to the supplementary material.

2. Methodology

The dataset is a realization of $\{Z_{it}, W_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$, in which Z_{it} is the variable of interest and W_{it} describes potential covariates, including seasonal terms and/or deterministic (non-linear and heterogeneous) trends, for example. Suppose we are interested in estimating the effects on the variable Z_{1t} of the first unit after an intervention that occurred at $T_0 + 1$. We estimate a counterfactual based on the peers $Z_{-1t} := (Z_{2t}, \ldots, Z_{nt})'$ that are assumed to be unaffected by the intervention. We allow the dimension of Z_{-1t} to grow with the sample size T, that is, $n := n_T$. We also assume that W_{it} are not affected by the intervention. Our key idea is to use both information in the latent factors and idiosyncratic components and we name the methodology as `FarmTreat`.

The procedure is, thus, summarized by the following steps:

1. For each unit $i = 1, \ldots, n$, run the regression:

$$Z_{it} = \gamma_i' W_{it} + R_{it}, \quad t = 1, \ldots, T^*,$$

and compute $\widehat{R}_{it} := Z_{it} - \widehat{\gamma}_i' W_{it}$, where $T^* = T_0$ for $i = 1$ and $T^* = T$, otherwise. This step removes heterogeneity due to W_{it}. As mentioned before, W_{it} may include an intercept, any observable factors, dummies to handle seasonality and outliers, and determinist (polynomial) trends, for example. In the case of our particular application, Z_{it} represents the daily quantity of a product sold per store in a municipality i and W_{it} includes a constant, six dummy variables for the days of the week and a linear trend.

2. Write $R_t := (R_{1t}, \ldots, R_{nt})'$, which is the cross-sectional data $Z_t := (Z_{1t}, \cdots, Z_{nt}')'$ after the heterogeneity adjustments. Fit the factor model

$$R_t = \Lambda F_t + U_t,$$

where F_t is an r-dimensional vector of unobserved factors, and Λ is an unknown $n \times r$ loading matrix and U_t is an n-dimensional idiosyncratic component. The second step consists of using the panel data $\{\widehat{R}_t\}_{t=1}^T$ to learn the common factors F_t and factor loading matrix Λ and compute the estimated idiosyncratic components by

$$\widehat{U}_t = \widehat{R}_t - \widehat{\Lambda}\widehat{F}_t,$$

where $\widehat{U}_t = (\widehat{U}_{1t}, \ldots, \widehat{U}_{nt})'$. There is a large literature on high-dimensional factor analysis; see, chap. 10 of the book by Fan et al. (2020) for details. One important point is that we should not use data after T_0 for the treated unit. There are many possibilities to handle this issue that are discussed in Section 4.

3. The third step is to use the idiosyncratic component to further augment the prediction on the treatment unit. It consists of first testing for the null of no remaining cross-sectional dependence (optional). If the null is rejected, fit the model in the preintervention period

$$\widehat{U}_{1t} = \theta_1' \widehat{U}_{-1t} + V_t, \quad t = 1, \ldots, T_0,$$

by using LASSO, where $\widehat{U}_{-1t} = (\widehat{U}_{2t}, \ldots, \widehat{U}_{nt})'$. Namely, compute

$$\widehat{\theta}_1 = \arg\min \left[\sum_{t=1}^{T_0} \left(\widehat{U}_{1t} - \theta_1' \widehat{U}_{-1t} \right)^2 + \xi \|\theta_1\|_1 \right]. \quad (1)$$

This step uses cross-sectional regression of the idiosyncratic components to estimate the effects in the treated unit. It is approximately the same as using $\widehat{F}_t$ and $\widehat{U}_{-1t}$ to predict $\widehat{R}_{1t}$ with the sparse regression coefficients for $\widehat{U}_{-1t}$, due to the orthogonality between $\{\widehat{F}_t\}_{t=1}^T$ and $\{\widehat{U}_t\}_{t=1}^T$. The model includes sparse linear model on R_t as a specific example (see (4) below) and the required model selection conditions are more easily met due to the factor adjustments. It also encompasses the principal component regression (PCR) in which $\theta_1 = 0$, namely, using no cross-sectional prediction.

4. Finally, the intervention effect δ_t is estimated for $t > T_0$ as

$$\widehat{\delta}_t = Z_{1t} - \left(\widehat{\gamma}_1' W_{1t} + \widehat{\lambda}_1' \widehat{F}_t + \widehat{\theta}_1' \widehat{U}_{-1t} \right). \quad (2)$$

where $\widehat{\lambda}_1$ is the estimated loading of unit 1, the first row of $\widehat{\Lambda}$. During the post treatment period, the realized factors $\widehat{F}$ are learned without using $R_{1,t}$.

5. Use the estimator (2) to test for null hypothesis of no intervention effect in the form described by (5).

The innovations of our approach in estimating counterfactuals are multi-folds. For simplicity, let us suppose that we have no W_{it} component, so that $R_t = Z_t$. First of all, the proposed procedure explores both the common factors and the dependence among idiosyncratic components. This not only makes use of more information, but also makes the newly transformed predictors less correlated. The latter makes the variable selection much easier and prediction more accurate. Note that factor regression (principal component regression) to estimate counterfactuals is a special case when $\theta_1 = 0$. Clearly, the method explores the sparsity of θ_1 to improve the performance and also includes the case of sparse regression on Z_{-1t} to estimate counterfactuals as in Masini and Medeiros (2021), where counterfactuals are estimated as

$$Z_{1t} = \theta_1' Z_{-1t} + \epsilon_t, \quad t = 1, \ldots, T_0. \quad (3)$$

However, the variables Z_{-1t} are highly correlated in high dimensions as they are driven by common factors, which makes variable selection procedures inconsistent and prediction ineffective. Instead, Fan, Ke, and Wang (2020) introduces the idea of lifting, called factor adjustments. Using the factor model in step 2, we can write the linear regression model (3) as

$$Z_{1t} = \theta_1' \Lambda_{-1} F_t + \theta_1' U_{-1t} + \epsilon_t, \quad (4)$$

where Λ_{-1} and U_{-1t} are defined as Λ and U_t without the first row. When we take $\lambda_1 = \theta_1' \Lambda_{-1}$, this reduces to use sparse regression to estimate the counterfactuals, but now use more powerful `FarmSelect` of Fan, Ke, and Wang (2020) to fit the sparse regression. Again, `FarmSelect` imposes the condition $\theta_1' \Lambda_{-1}$ as the regression coefficients of F_t. Our method does not require this constraint. This flexibility allows us to apply our new approach even when the sparse linear model does not hold.

Finally, we also consider a test for the contribution of the idiosyncratic components by testing the null hypothesis that $\boldsymbol{\theta}_1 = \mathbf{0}$. Note that this is a high-dimensional hypothesis test, which is equivalent to testing the uncorrelatedness between the idiosyncratic component U_{1t} for the treated unit and those from the untreated units $\boldsymbol{U}_{-1t}$ in the preintervention period.

3. Assumptions and Theoretical Result

3.1. Assumptions

Suppose we have n units (municipalities, firms, etc.) indexed by $i = 1, \ldots, n$. For every time period $t = 1, \ldots, T$, we observe a realization of a real valued random vector $\boldsymbol{Z}_t := (Z_{1t}, \ldots, Z_{nt})'$.[2] We assume that an intervention took place at $T_0 + 1$, where $1 < T_0 < T$. Let $\mathcal{D}_t \in \{0, 1\}$ be a binary variable flagging the periods where the intervention for unit 1 was in place. Therefore, following Rubin's potential outcome framework, we can express Z_{it} as

$$Z_{it} = \mathcal{D}_t Z_{it}^{(1)} + (1 - \mathcal{D}_t) Z_{it}^{(0)},$$

where $Z_{it}^{(1)}$ denote the potential outcome when the unit i is exposed to the intervention and $Z_{it}^{(0)}$ is the potential outcome of unit i when it is not exposed to the intervention.

We are ultimately concerned with testing the hypothesis on the potential effects of the intervention in the unit of interest, that is, the treatment effect on the treated. Without loss of generality, we set unit 1 to be the one of interest. The null hypothesis to be tested is:

$$\mathcal{H}_0 : \boldsymbol{g}(\delta_{T_0+1}, \ldots, \delta_T) = \mathbf{0}, \tag{5}$$

where $\delta_t := Z_{1t}^{(1)} - Z_{1t}^{(0)}$, $\forall t > T_0$, and $\boldsymbol{g}(\cdot)$ is a vector-valued continuous function. The general null hypothesis (5) can be specialized to many cases of interest, as for example:

$$\mathcal{H}_0 : \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \delta_t = 0 \quad \text{or} \quad \mathcal{H}_0 : \delta_t = 0, \forall t > T_0.$$

It is evident that for each unit $i = 1, \ldots, n$ and at each period $t = 1, \ldots, T$, we observe either $Z_{it}^{(0)}$ or $Z_{it}^{(1)}$. In particular, $Z_{1t}^{(0)}$ is not observed from $t = T_0 + 1$ onwards. For this reason, we henceforth call it the *counterfactual*—that is, what Z_{1t} would have been like had there been no intervention (potential outcome).

The counterfactual is constructed by considering a model in the absence of an intervention:

$$Z_{1t}^{(0)} = \mathcal{M}\left(\boldsymbol{Z}_{-1t}^{(0)}; \boldsymbol{\theta}\right) + V_t, \quad t = 1, \ldots, T, \tag{6}$$

where $\boldsymbol{Z}_{-1t}^{(0)} := (Z_{2t}^{(0)}, \ldots, Z_{nt}^{(0)})'$ be the collection of all control variables (all variables in the untreated units).[3] $\mathcal{M} : \mathcal{Z} \times \boldsymbol{\Theta} \to \mathbb{R}$, $\mathcal{Z} \subseteq \mathbb{R}^{n-1}$, is a known measurable mapping up to

a vector of parameters indexed by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$ is a parameter space. A linear specification (including a constant) for the model $\mathcal{M}(\boldsymbol{Z}_{0t}; \boldsymbol{\theta})$ is the most common choice among counterfactual models for the preintervention period. `FarmTreat` uses a more sophisticated model.

Roughly speaking, in order to recover the effects of the intervention, we need to impose that the peers are unaffected by the intervention in the unit of interest. Otherwise our counterfactual model would be invalid. Specifically we consider the following key assumption

Assumption 1 (Intervention Independence). $\boldsymbol{Z}_t^{(0)}$ is independent of $\mathcal{D}_s$ for all $1 \leq s, t \leq T$.

Remark 1. Assumption 1 identifies the treatment effect on the treated. If only $\boldsymbol{Z}_{-1t}^{(0)}$ is independent of $\mathcal{D}_s$ for all $1 \leq s, t \leq T$, we can recover the effect of the intervention on the treated unit given that T_0 is deterministic and known. This later case is typical in papers on SC.

The main idea is to estimate (6) using just the preintervention sample, $t = 1, \ldots, T_0$, since under Assumption 1, $\boldsymbol{Z}_t^{(0)} = (\boldsymbol{Z}_t^{(0)} | \mathcal{D}_t = 0) = (\boldsymbol{Z}_t | \mathcal{D}_t = 0)$ for all t. Consequently, the estimated counterfactual for the postintervention period, $t = T_0 + 1, \ldots, T$, becomes $\widehat{Z}_{1t}^{(0)} := \mathcal{M}(\boldsymbol{Z}_{-1t}; \widehat{\boldsymbol{\theta}}_{T_0})$. Under some sort of stationary assumption on $\boldsymbol{Z}_t$, in the context of a linear model, Hsiao, Ching, and Wan (2012) and Carvalho, Masini, and Medeiros (2018), show that $\widehat{\delta}_t := Z_{1t} - \widehat{Z}_{1t}^{(0)}$ is an unbiased estimator for δ_t as the preintervention sample size grows to infinity in the low and high dimensional sparse case, respectively.

We model the units in the absence of the intervention as follows.

Assumption 2 (DGP). The process $\{Z_{it}^{(0)} : 1 \leq i \leq n, t \geq 1\}$ is generated by

$$Z_{it}^{(0)} = \boldsymbol{\gamma}_i' \boldsymbol{W}_{it} + \boldsymbol{\lambda}_i' \boldsymbol{F}_t + U_{it} \tag{7}$$

where $\boldsymbol{\gamma}_i \in \mathbb{R}^k$ is the vector of coefficients of the k-dimensional observable random vector $\boldsymbol{W}_{it}$ of attributes of unit i, $\boldsymbol{F}_t$ is a r-dimensional vector of common factors and $\boldsymbol{\lambda}_i$ its respective vector of loads for unit i; and U_{it} is a zero mean idiosyncratic shock. Finally, we assume that $\boldsymbol{W}_{it}$, $\boldsymbol{F}_t$ and U_{it} are mutually uncorrelated.

The reason to include $\boldsymbol{W}_{it}$ is to accommodate an intercept, heterogeneous deterministic trends, seasonal dummies or any other exogenous (possibly random) characteristic of unit i that the practitioner judges to be helpful in the construction of the counterfactual. As mentioned before we include an intercept, dummies to account for the effects of different days of the week and a linear trend. Other possibilities could be dummies of nation-wide promotions and/or holidays, for example.

In case of stochastic heterogeneous trends, we let the factors follow a random walk with (or without) drift: $\boldsymbol{F}_t = \boldsymbol{\mu} + \boldsymbol{F}_{t-1} + \boldsymbol{\eta}_t$, where $\{\boldsymbol{\eta}_t\}$ is a second-order stationary vector process. When this is the case, the methodology must be applied in first-differences and levels should be reconstructed in the end.

---

[2]We consider a scalar variable for each unit for the sake of simplicity, and the results in the paper can be easily extended to the multivariate case.

[3]We could also have included lags of the variables and/or exogenous regressors into $\boldsymbol{Z}_{-1t}$, but again, to keep the argument simple, we have considered only contemporaneous variables; see Carvalho, Masini, and Medeiros (2018) for more general specifications.

Therefore, our approach can be directly applied even if the units have heterogeneous and stochastic trends.

Our counterfactual model is the sample version of the projection of $Z_{1t}^{(0)}$ onto the space spanned by $(W_{1t}, F_t, U_{-1,t})'$. Under Assumption 2 the counterfactual can be taken as

$$Z_{1t}^{(0)} = \gamma_1' W_{1t} + \lambda_1' F_t + \theta_1' U_{-1t} + V_t, \tag{8}$$

where θ_1 is the coefficient of the linear regression of U_{1t} onto U_{-1t} and V_t the respective projection error.

3.2. Theoretical Guarantees

In order to state our result in a precise manner we consider the technical assumption below. First, let $W_{S,it}$ denotes the subvector of W_{it} after the exclusion of all deterministic (nonrandom) components (constant, dummies, trends, etc). We state the assumption for the case where the unobserved factors are stationary. For nonstationary (unit-root) factors, the only difference is to state the conditions for the first-differences of the variables involved: $\Delta W_{S,it}$, ΔF_t, and ΔU_t. In this case, our results can be derived following in sec. 7 Bai and Ng (2008). Note that, as mentioned before, if the interest lies on the intervention effects on the levels of the series, after the final step, the levels of both the target variable and the counterfactual can be recovered and the inferential procedures can be applied unaltered.

Assumption 3 (Regularity Conditions). There is a constant $0 < C < \infty$ such that:

(a) The covariance matrix of $W_{S,it}$ is nonsingular;
(b) $\mathbb{E}|W_{S,it}|^p \leq C$ and $\mathbb{E}|U_{it}|^{p+\epsilon} \leq C$ for some $p \geq 6$ and $\epsilon > 0$ for $i = 1, \ldots, n, t = 1, \ldots, T$;
(c) The process $\{(W_{S,t}', F_t', U_t')', t \in \mathbb{Z}\}$ is weakly stationary with strong mixing coefficient α satisfying $\alpha(m) \leq \exp(-2cm)$ for some $c > 0$ and for all $m \in \mathbb{Z}$;
(d) $||\theta_1||_\infty \leq C$;
(e) $\kappa_0 := \kappa\left[\mathbb{E}(U_t U_t'), S_0, 3\right] \geq C^{-1}$ where $\kappa[\cdot]$ is the compatibility condition defined in (S.1) in the supplementary material and $S_0 := \{i : \theta_{1,i} \neq 0\}$.

Condition (a) is necessary for the parameters $\gamma_i, i = 1, \ldots, n$, to be well defined. Conditions (b) and (c) taken together allow the law of large numbers for strong mixing processes to be applied to appropriately scaled sums. In particular, (b) bounds the p-th plus moment uniformly. However, if U_{it} has exponential tails as contemplated in Assumption 3 in Fan, Masini, and Medeiros (2021), we could state a stronger result in terms of the allowed number of nonzero coefficients as a fraction of the sample size. The mixing rate in condition (c) can be weaken to polynomial rate at the expense of an interplay between (c) and the conditions appearing in Proposition 1.

Finally, conditions (d) and (e) in Assumption 3 are regularity conditions on the high-dimensional linear model to be estimated by the LASSO in step 3. Condition (e) ensures the (restricted) strong convexity of the objective function, which is necessary for consistently estimate θ_1 when $n > T$. In effect, it uniformly lower bounds the minimum restricted ℓ_1-eigenvalue of the covariance matrix of U_t. For simplicity, the bounds appearing in (d) and (e) are assumed to hold uniformly. However, both conditions could be somewhat relaxed

to allow $||\theta_1||_\infty$ to grow slowly and/or κ_0 decreases slowly to 0 as n diverges. Once again, at the expense of having both terms included in the conditions of Proposition 1.

Proposition 1. Under Assumptions 1–3, assume further that:

(a) There is a bounded sequence $\eta := \eta_{n,T}$ such that $||\widehat{U} - U||_{\max} = O_P(\eta)$; and

(b) $|S_0| = O\left(\left\{\eta\left[(nT)^{1/p} + \eta\right] + \frac{n^{4/p}}{\sqrt{T}}\right\}^{-1}\right)$.

If the penalty parameter ξ in (1) is set to be at the order of $\frac{n^{2/p}}{\sqrt{T}} + \eta T^{1/p}$ then, as $T_0 \to \infty$, $||\widehat{\theta}_1 - \theta_1||_1 = O_P(\xi|S_0|)$, and for every $t > T_0$:

$$\widehat{\delta}_t - \delta_t = V_t + O_P\left\{|S_0|\left[\eta(nT)^{1/p} + \frac{n^{3/p}}{\sqrt{T}}\right]\right\},$$

where V_t is the stochastic component not explainable by untreated units defined by (8)

Remark 2. Conditions (a) and (b) are high level assumptions that translate into a restriction on the estimation rate in steps 1 and 2 of the proposed methodology, which in turn puts an upper bound on the number of nonzero coefficients in θ_1 (sparsity) in order for the estimation error to be negligible. The rate η can be explicitly obtained in terms of n and T by imposing conditions on projection matrix of W_{it} and the factor model. For the former, we need uniform consistencies of both the factor and the loadings estimators that take into account the projection error in the previous step. In a more general setup, Fan, Masini, and Medeiros (2021) state conditions under which $\eta = \frac{n^{6/p}}{T^{1/2-6/p}} + \frac{T^{1/p}}{\sqrt{n}}$.

Proposition 1 is key for our inference procedure discussed in Section 3.3. For instance, it can be used to argue that $\widehat{\delta}_t - \delta_t = V_t + o_p(1)$ provided that $|S_0|\left[\eta(nT)^{1/p} + \frac{n^{3/p}}{\sqrt{T}}\right] = o(1)$. Since V_t is zero mean by construction, as $T_0 \to \infty$, $\widehat{\delta}_t$ is an unbiased estimator for δ_t for every postintervention period. Furthermore, as described below, we can estimate the quantiles of V_t using the preintervention residuals to conduct a valid inference on δ_t.

3.3. Testing for Intervention Effect

We test the null of no intervention effects based on estimators $\{\widehat{\delta}_t\}_{t > T_0}$ and the results of Masini and Medeiros (2019, 2022) and Chernozhukov, Wüthrich, and Zhu (2020). Let $T_2 := T - T_0$ be the number of observations after the intervention and define a generic continuous mapping $\phi : \mathbb{R}^{T_2} \to \mathbb{R}^b$ whose argument is the T_2-dimensional vector $(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)'$ with given treatment effects $\delta_{T_0+1}, \cdots, \delta_T$.

We are interested in the distribution of $\widehat{\phi} := \phi(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)$ under the null (5), where ϕ is a given statistic. The typical situation is the one where the preintervention period is much longer than the post intervention period, $T_0 \gg T_2$. Frequently, it could be well the case that $T_2 = 1$. However, V_t does not vanish as in most cases there is a single treated unit. Nevertheless, under strict stationarity of the process $\{V_t\}$ and unbiasedness of the treatment effect estimator, it is possible

to resample the pre-intervention residuals following either the procedure described in Masini and Medeiros (2019, 2022) or the one in Chernozhukov, Wüthrich, and Zhu (2020) to compute the sample quantile of the statistic of interest. As pointed out earlier, the main difference between the two approaches is that the former estimate the counterfactual model using only pretreatment observations while the later considers the estimation using the full sample.

Under the asymptotic limit taken on the preinvention period $(T_0 \to \infty)$, by Proposition 1, we have that $\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 = o_P(1)$, where $\boldsymbol{\phi}_0 := \boldsymbol{\phi}(V_{T_0+1}, \ldots, V_T)$. Thus, the distribution of $\widehat{\boldsymbol{\phi}}$ can be estimated by that of $\boldsymbol{\phi}_0$. Consider the construction of $\widehat{\boldsymbol{\phi}}$ using only blocks of size T_2 of consecutive observations from the preintervention sample. There are $T_0 - T_2 + 1$ such blocks denoted by $\widehat{\boldsymbol{\phi}}_j := \boldsymbol{\phi}(\widehat{V}_j, \ldots, \widehat{V}_{j+T_2-1})$, $j = 1, \ldots, T_0 - T_2 + 1$, where $\widehat{V}_t := Z_{1t} - \left(\widehat{\boldsymbol{\gamma}}_1' \boldsymbol{W}_{1t} + \widehat{\boldsymbol{\lambda}}_1' \widehat{\boldsymbol{F}}_t + \widehat{\boldsymbol{\theta}}_1' \widehat{\boldsymbol{U}}_{-1t} \right)$ for the preintervention period. The estimators $\widehat{\boldsymbol{\gamma}}_1, \widehat{\boldsymbol{\lambda}}_1, \widehat{\boldsymbol{F}}_t, \widehat{\boldsymbol{\theta}}_1'$, and $\widehat{\boldsymbol{U}}_{-1t}$ use either the preintervention or the full sample depending on the inferential approach chosen by the practitioner.

For each j, we have that $\widehat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j = o_P(1)$ where $\boldsymbol{\phi}_j := \boldsymbol{\phi}(V_j, \ldots, V_{j+T_2-1})$ and $\boldsymbol{\phi}_j$ is equal in distribution to $\boldsymbol{\phi}_0$ for all j. Hence, we propose to estimate the distribution $\mathcal{Q}_T(\boldsymbol{x}) := \P(\widehat{\boldsymbol{\phi}} \leq \boldsymbol{x})$ by its empirical distribution

$$\widehat{\mathcal{Q}}_T(\boldsymbol{x}) := \frac{1}{T_0 - T_2 + 1} \sum_{j=1}^{T_0 - T_2 + 1} \mathbb{1}(\widehat{\boldsymbol{\phi}}_j \leq \boldsymbol{x}),$$

where, for a pair of vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, we say that $\boldsymbol{a} \leq \boldsymbol{b} \Leftrightarrow a_i \leq b_i, \forall i$. See Masini and Medeiros (2019, 2022) and Chernozhukov, Wüthrich, and Zhu (2020) for further details.

3.4. Testing for Idiosyncratic Contributions

The question of statistical and practical interest is if the idiosyncratic component contributes the estimation of the treatment effect. To answer this question, write (7) as:

$$\boldsymbol{Z}_t = \boldsymbol{\Gamma} \boldsymbol{W}_t + \boldsymbol{\Lambda} \boldsymbol{F}_t + \boldsymbol{U}_t, \qquad t \in \{1, \ldots, T\},$$

where $\boldsymbol{Z}_t := (Z_{1t}, \ldots, Z_{nt})'$, $\boldsymbol{U}_t := (U_{1t}, \ldots, U_{nt})'$, and $\boldsymbol{W}_t := (\boldsymbol{W}_{1t}', \ldots, \boldsymbol{W}_{nt}')'$. The $(n \times nk)$ block diagonal matrix $\boldsymbol{\Gamma}$ has blocks given by $(\boldsymbol{\gamma}_1', \ldots \boldsymbol{\gamma}_n')$. Finally, $\boldsymbol{\Lambda} := (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)'$.

Let $\boldsymbol{\Pi} := (\pi_{ij})_{1 \leq i,j \leq n}$ denote the $(n \times n)$ covariance matrix of $\boldsymbol{U}_t$. Our method exploits the sparsity of the off-diagonal elements of $\boldsymbol{\Pi}$. In particular, we are interested in testing whether $\boldsymbol{U}_{-1t}$ has linear prediction power on the treated unit U_{1t}. This amounts to the following high-dimensional hypothesis test: $\mathcal{H}_0 : \pi_{1j} = 0, \forall 2 \leq j \leq n$.

In order to conduct the test we propose the following test statistic $S := ||\boldsymbol{Q}||_\infty$, where $\boldsymbol{Q} := \frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} \boldsymbol{D}_t$, $\boldsymbol{D}_t := \widehat{U}_{1t} \widehat{\boldsymbol{U}}_{-1t}$, and $\widehat{U}_{it} := \widehat{R}_{it} - \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{F}}_t$. Also let $c^*(\tau)$ be the τ-quantile of the Gaussian bootstrap $S^* := ||\boldsymbol{Q}^*||_\infty$, where $\boldsymbol{Q}^*|\boldsymbol{Z}, \boldsymbol{W} \sim \mathcal{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Upsilon}})$. For a given symmetric kernel $k(\cdot)$ with $k(0) = 1$ and bandwidth $h > 0$ (determining the number of lags), we have that

$$\widehat{\boldsymbol{\Upsilon}} := \sum_{|\ell| < T_0} k(\ell/h) \widehat{\boldsymbol{M}}_\ell \quad \text{with} \quad \widehat{\boldsymbol{M}}_\ell := \frac{1}{T_0} \sum_{t=\ell+1}^{T_0} \boldsymbol{D}_t \boldsymbol{D}_{t-\ell}'$$

is the estimator of the long-run covariance matrix $\boldsymbol{\Upsilon} := \mathbb{V}\widetilde{\boldsymbol{Q}}$, where $\widetilde{\boldsymbol{Q}} := \frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} U_{1t} \boldsymbol{U}_{-1t}$. Notice that $\widehat{\boldsymbol{\Upsilon}}$ is just the Newey-West estimator if $k(\cdot)$ is chosen to be the triangular kernel. More generally, the choice of kernels can be made in class of kernels described in Andrews (1991). The validity of such a method has been proved in Fan, Masini, and Medeiros (2021) under a more general setting. In particular, the authors show under some regularity conditions

$$\sup_{\tau \in (0,1)} |\P(S \leq c^*(\tau)) - \tau| = o(1) \quad \text{under } \mathcal{H}_0.$$

4. Guide to Practice

In this section we provide practical guidance to the implementation of the `FarmTreat` method.

The first step involves the definition of the variables in $\boldsymbol{W}_{it}$. This is, of course, application dependent. Nevertheless, typical candidates are deterministic functions of time, that is, $f(t)$, in order to capture trends, an intercept to remove the mean, seasonal dummies or other calendar effects, or any other dummies to remove potential outliers. Unit-root tests on the variable of interest may also be important in order to decide whether first-differences of the data should be taken or not.

The second step is the estimation of $\boldsymbol{\Lambda}$ and the sequence of factors $\{\boldsymbol{F}_t, t \in \mathbb{Z}\}$ for the full sample, before and after the intervention. Therefore, we cannot just rely on preintervention period to estimate the factors. On the other hand, if we use all the observations from the treated unit, we will bias our estimation under the alternative of nonzero treatment effects. Therefore, there are two possible ways to estimate the factors and the factor loadings:

1. A simple approach is to estimate the factors and factor loadings without the treated unit. In order to estimate the loadings $\widehat{\boldsymbol{\lambda}}_1$ of the first unit, we then regress R_{1t} on the estimated factors. This is the approach adopted in both simulations and in the empirical application.
2. The imputed approach is to use the imputation $\widehat{\boldsymbol{\lambda}}_1' \widehat{\boldsymbol{F}}_t$ for the post intervention period of the treated unit 1 and then apply the whole data to reestimate the factor and factor loadings. $\widehat{\boldsymbol{\lambda}}_1$ and $\widehat{\boldsymbol{F}}_t$ are estimated with just the preintervention period.
3. Note that $\mathbb{E}(\boldsymbol{R}_t) = \boldsymbol{0}$ by definition. Hence, we can replace the postintervention observations of R_{1t} by 0 in order to carry the factor analysis. As the number of postintervention observations is expected to be quite small, this replacement will have negligible effects. It is important to notice, however, that we do this just to estimate the factors.

To determine the number of factors we advocate the use of the eigenvalue ratio test (Ahn and Horenstein 2013). Other possibility is the use of one of the information criteria discussed in Bai and Ng (2002).

After the estimation of the common factor structure, we can test for remaining cross-dependence using the test described in Section 3.3. In the case of rejection of the null of no remaining dependence, the last step consists of a LASSO regression. This step of testing is optional for evaluating the treatment effect, as the sparsity of LASSO includes no effect as a specific example. Nevertheless, it is an interesting statistical problem whether the

idiosyncratic component contributes to the prediciton power. For selecting the penalty parameter in LASSO, we recommend the use of an information criterion, such as the BIC as in Masini and Medeiros (2021).

The final step is to test the null hypothesis concerning the intervention effects. When the preintervention sample is small, we follow Chernozhukov, Wüthrich, and Zhu (2020) and estimate the models under the null. Note that in this case we should reestimate the model using the full sample.

5. Simulations

In this section we report simulations results to study the finite sample behavior of the method proposed in this article. We consider the following data generating process:

$$
\begin{aligned}
Z_{it} &= \delta_{it} + \gamma_i' W_t + R_{it}, \quad R_{it} = \lambda_i' F_t + U_{it}, \\
F_t &= 0.8 F_{t-1} + V_t, \quad U_{it} = \begin{cases} \beta' U_{-1t} + \varepsilon_{it}, & \text{if } i = 1, \\ \varepsilon_{it}, & \text{otherwise,} \end{cases}
\end{aligned}
\tag{9}
$$

where $\{\varepsilon_{it}\}$ is a sequence of independent and normally distributed zero-mean random variables with variance equal to 0.25 if $i = 1$ and $\beta \neq 0$ or variance equal to 1 if $i > 1$ or $\beta = 0$. V_t is a sequence of independent and normally distributed zero-mean random vectors taking values on $\mathbb{R}^2$ such that $\mathbb{E}(V_t V_t') = 0.25 \times I$, and $\mathbb{E}(\varepsilon_{it} V_s) = 0$, for all i, t, and s. W_{it} consists of a constant, a linear trend, and two independent Gaussian random variables with mean and variance equal to 1. The parameters are set as follows: γ_i is $(p + 2)$-dimensional vector where, for each replication, the first entry is randomly pick from a Gaussian random variable with zero mean and variance 1; the second term is randomly selected from an Uniform distribution between -5 and 5; and the last two elements are Gaussian distributed with mean 0.5 and variance 1. For each replication, the elements of $\lambda_i, i > 1$, are drawn independently from a normal distribution with mean two and unit variance and, for $i = 1$, the elements of λ_i are drawn from a normal distribution with mean -6 and variance 0.04. The first two elements of β are either set to 0.5 and the rest is set to zero or we set all the elements equal to zero. We consider the following sample sizes: $T_0 = 50, 75, 100, 150, 250, 500$ and 1000; and $T_2 = 1$. For each sample size, n is set as $n = \{T, 2T, 3T\}$. The number of factors is set to two. For size simulations, $\delta_{it} = 0$ for all i and t. For power simulations, $\delta_{it} = 2$ for $i = 1$ and $t = T_0 + 1$.

Tables 1 and 2 show descriptive statistics for the counterfactual estimation. The table depicts the mean, the median and the mean squared error (MSE) for δ_{T_0+1} under the null and alternative hypotheses, respectively. Three cases are considered. In the first one, the factor structure is neglected and a sparse LASSO regression of the first unit against the remaining ones is estimated. This is the ArCo methodology put forward by Carvalho, Masini, and Medeiros (2018). The second one is equivalent to the approach of Gobillon and Magnac (2016), where a pure factor model is considered. Finally, the FarmTreat approach is considered, which encompasses the previous two methods as a specific example. We also report, between brackets, the same statistics when the full sample is used to estimate the counterfactual model as advocated by Chernozhukov, Wüthrich, and Zhu (2020).

Tables 1 and 2 show descriptive statistics for the counterfactual estimation. The table depicts the mean, the median and the mean squared error (MSE) for δ_{T_0+1} under the null and alternative hypotheses, respectively. Three cases are considered. In the first one, the factor structure is neglected and a sparse LASSO regression of the first unit against the remaining ones is estimated. This is the ArCo methodology put forward by Carvalho, Masini, and Medeiros (2018). The second one is equivalent to the approach of Gobillon and Magnac (2016), where a pure factor model is considered; we call this method the Principal Component Regression (PCR). Finally, the FarmTreat approach is considered, which encompasses the previous two methods as a specific example. We also report, between brackets, the same statistics when the full sample is used to estimate the counterfactual model as advocated by Chernozhukov, Wüthrich, and Zhu (2020).

From the inspection of the results in the tables, it is clear that the biases for estimating of the treatment effect are small and MSEs decrease as the sample size increase, as expected. Furthermore, the ArCo delivers very robust estimates, but the MSE can be substantially reduced by the FarmTreat methodology. Therefore, there is strong evidence supporting methodology derived in this paper, which is consistency with our theoretical results. Second, as already shown in the simulations in Carvalho, Masini, and Medeiros (2018), the performance of the pure factor model is poor in terms of MSE. This is particularly the case when n or T is small, since the factors are not well estimated. When this happens, the prediction power of the idiosyncratic components comes to rescue (comparing the performance with FarmSelect). This demonstrates convincingly the need of using the idiosyncratic component to augment the prediction. When comparing with the results when the full sample is used to estimate the model, two facts emerge from the tables. First, when the null hypothesis is true, the gains of using the full sample are undebatable. However, when the null is false, using the full sample is a bad idea, specially when T_0 is small. This is somewhat expected as in the later case we are including observations affected by the intervention in the estimation sample.

Table 3 presents the empirical size of the ressampling test when there is a single observation after the intervention and the counterfactual is estimated according to the methods described above. It is clear that size distortions are high when T_0 is small. The size converges to the nominal one as the sample increases. On the other hand, using the full sample to estimate the models correct the distortions and are strongly recommended in the case of small samples.

Table 4 shows the empirical power. The ressampling approach delivers high power, specially when ArCo and FarmTreat methodologies are considered. On the other hand, the test looses a lot of power when the full sample is considered. This is expected as the estimator of the treatment effect will be biased, specially in small samples.

Figure 1 compares the MSEs of PCR and FarmTreat when DGP has no idiosyncratic contribution, that is, $\beta = 0$. This case favors to PCR. As we can see, FarmTreat achieves comparable results to PCR, indicating that the methodology is quite robust.

Table 1. Average treatment (Δ) estimation under the null.

	Panel(a): LASSO (ArCo) - Carvalho, Masini, and Medeiros (2018)								
	Mean			Median			MSE		
	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$	−0.036	0.010	0.023	−0.073	0.023	0.074	1.114	0.975	0.901
	[0.017]	[0.003]	[0.000]	[0.024]	[−0.004]	[0.008]	[0.254]	[0.178]	[0.188]
75	−0.084	0.025	−0.011	−0.080	0.025	0.015	0.832	0.785	0.763
	[−0.027]	[−0.018]	[−0.011]	[−0.035]	[−0.018]	[0.001]	[0.383]	[0.216]	[0.206]
100	−0.016	0.068	0.026	−0.021	0.079	0.028	0.732	0.674	0.632
	[−0.020]	[−0.005]	[−0.022]	[0.014]	[0.014]	[0.009]	[0.317]	[0.311]	[0.264]
150	0.004	0.037	−0.070	0.021	0.038	−0.083	0.608	0.655	0.590
	[−0.000]	[−0.020]	[0.022]	[−0.016]	[−0.037]	[−0.012]	[0.362]	[0.288]	[0.349]
250	−0.013	−0.026	−0.028	0.021	−0.040	−0.039	0.539	0.517	0.566
	[−0.000]	[−0.020]	[0.022]	[−0.016]	[−0.037]	[−0.012]	[0.362]	[0.288]	[0.349]
500	0.018	−0.028	0.052	0.032	−0.062	0.042	0.419	0.382	0.424
	[0.002]	[0.016]	[−0.011]	[0.009]	[0.007]	[0.005]	[0.321]	[0.300]	[0.280]
1,000	0.029	0.033	−0.028	0.047	0.049	0.005	0.323	0.378	0.350
	[−0.026]	[−0.035]	[0.029]	[−0.053]	[−0.026]	[0.054]	[0.274]	[0.303]	[0.275]
	Panel(b): PCR - Gobillon and Magnac (2016)								
	Mean			Median			MSE		
	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$	−0.030	−0.001	0.038	−0.120	0.008	0.015	1.259	0.916	0.842
	[0.032]	[−0.013]	[0.014]	[0.004]	[−0.053]	[0.042]	[0.605]	[0.665]	[0.696]
75	−0.041	0.011	−0.002	0.014	0.023	−0.005	0.957	0.958	0.893
	[−0.037]	[−0.047]	[0.008]	[−0.004]	[−0.057]	[0.018]	[0.727]	[0.767]	[0.670]
100	−0.065	0.083	0.014	−0.087	0.080	−0.011	0.989	0.807	0.863
	[−0.019]	[−0.012]	[−0.012]	[−0.001]	[0.024]	[0.004]	[0.683]	[0.719]	[0.636]
150	−0.045	−0.022	−0.093	−0.005	−0.040	−0.083	1.071	0.860	0.914
	[−0.017]	[−0.031]	[0.036]	[−0.004]	[−0.036]	[0.023]	[0.732]	[0.701]	[0.772]
250	−0.042	−0.045	−0.038	−0.041	−0.057	−0.018	0.982	0.778	0.861
	[−0.008]	[−0.047]	[−0.015]	[−0.001]	[−0.007]	[−0.025]	[0.728]	[0.762]	[0.778]
500	0.006	0.001	0.070	0.066	−0.011	0.033	0.765	0.692	0.758
	[−0.018]	[0.057]	[−0.006]	[−0.009]	[0.015]	[−0.078]	[0.740]	[0.809]	[0.751]
1,000	0.028	0.050	−0.046	0.075	0.049	−0.052	0.720	0.783	0.763
	[−0.024]	[−0.051]	[0.057]	[−0.026]	[−0.053]	[0.076]	[0.739]	[0.801]	[0.790]
	Panel(c): `FarmTreat`								
	Mean			Median			MSE		
	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$	−0.024	−0.036	0.035	−0.080	−0.026	0.026	0.964	0.684	0.630
	[0.017]	[0.003]	[0.000]	[0.024]	[−0.004]	[0.008]	[0.254]	[0.178]	[0.188]
75	−0.048	−0.017	−0.019	−0.030	−0.036	−0.015	0.607	0.529	0.471
	[−0.027]	[−0.018]	[−0.011]	[−0.035]	[−0.018]	[0.001]	[0.383]	[0.216]	[0.206]
100	−0.011	0.048	0.014	−0.036	0.061	0.009	0.548	0.377	0.404
	[−0.020]	[−0.005]	[−0.022]	[0.014]	[0.014]	[0.009]	[0.317]	[0.311]	[0.264]
150	−0.063	0.016	−0.055	−0.035	0.012	−0.063	0.585	0.343	0.343
	[−0.000]	[−0.020]	[0.022]	[−0.016]	[−0.037]	[−0.012]	[0.362]	[0.288]	[0.349]
250	−0.033	0.003	−0.024	−0.048	0.007	−0.025	0.453	0.312	0.301
	[0.002]	[−0.022]	[0.001]	[−0.018]	[−0.003]	[0.025]	[0.311]	[0.313]	[0.317]
500	0.017	−0.021	0.026	0.007	−0.010	0.019	0.301	0.260	0.269
	[0.002]	[0.016]	[−0.011]	[0.009]	[0.007]	[0.005]	[0.321]	[0.300]	[0.280]
1,000	0.031	0.036	−0.034	0.035	0.068	−0.024	0.246	0.291	0.263
	[−0.026]	[−0.035]	[0.029]	[−0.053]	[−0.026]	[0.054]	[0.274]	[0.303]	[0.275]

NOTE: The table reports descriptive statistics for the average treatment estimation under the null of no effect ($\delta_{T_0+1} = 0$). The table reports the mean, median, and mean squared error (MSE) of the estimator $\widehat{\Delta}$ for one post-intervention observation. Panel (a) considers the case where the counterfactual is estimated by a LASSO regression of the treated unit on all the peers. This is the Artificial Counterfactual (ArCo) approach proposed by Carvalho, Masini, and Medeiros (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), that is, an ordinary least squares (OLS) regression of the treated unit on factors computed from the pool of peers. This is equivalent to the method of Gobillon and Magnac (2016). The number of factors is determined by the eigenvalue ratio test of Ahn and Horenstein (2013). Finally, Panel (c) displays the results of the `FarmTreat` methodology. Between brackets we report the same statistics but with the model estimated using the full sample as advocated by Chernozhukov, Wüthrich, and Zhu (2020).

6. Price Elasticity of Demand

6.1. Data Description

As described in Section 1.1 the goal is to determine the optimal price of products for a large retail chain in Brazil. The optimal prices should be computed for each city. Our dataset consists of the daily prices and quantities sold of five different products, aggregated at the municipal level. The company's more than 1400 stores differ substantially across and within municipalities, ranging from small convenience stores with a limited selection of products up to very large ones, selling everything from sweets to home appliances and clothes. The stores can be street stores or can be located in shopping malls. The products sold are divided into several departments. Here, we consider products from the

Table 2. Average treatment (Δ) estimation under the alternative.

Panel(a): LASSO (ArCo) - Carvalho, Masini, and Medeiros (2018)								
Mean			Median			MSE		
$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
T = 50								
1.998	2.048	2.001	1.938	2.077	1.973	0.982	0.972	0.812
[0.945]	[0.830]	[0.826]	[0.811]	[0.680]	[0.694]	[1.597]	[1.763]	[1.711]
75								
2.002	2.025	1.946	1.955	2.012	1.913	0.871	0.821	0.828
[1.132]	[1.013]	[0.993]	[1.028]	[0.882]	[0.866]	[1.216]	[1.397]	[1.372]
100								
2.003	1.998	2.087	2.025	2.024	2.060	0.737	0.691	0.681
[1.316]	[1.250]	[1.177]	[1.236]	[1.169]	[1.093]	[0.899]	[1.011]	[1.098]
150								
2.014	2.015	1.967	2.051	2.025	1.985	0.561	0.617	0.587
[1.486]	[1.435]	[1.323]	[1.458]	[1.376]	[1.262]	[0.670]	[0.722]	[0.853]
250								
2.037	1.989	2.033	2.033	2.061	2.022	0.497	0.550	0.491
[1.581]	[1.532]	[1.526]	[1.593]	[1.509]	[1.476]	[0.535]	[0.564]	[0.631]
500								
2.047	2.022	1.960	2.043	2.036	1.949	0.388	0.392	0.383
[1.717]	[1.639]	[1.696]	[1.719]	[1.638]	[1.694]	[0.392]	[0.394]	[0.399]
1,000								
1.974	2.014	1.969	1.957	2.036	1.987	0.380	0.334	0.378
[1.810]	[1.785]	[1.715]	[1.815]	[1.792]	[1.725]	[0.297]	[0.341]	[0.350]

Panel(b): PCR - Gobillon and Magnac (2016)								
Mean			Median			MSE		
$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
T = 50								
1.941	2.071	2.025	1.981	2.100	1.932	1.150	0.944	0.897
[1.524]	[1.718]	[1.757]	[1.527]	[1.682]	[1.732]	[1.161]	[0.717]	[0.652]
75								
1.998	2.004	1.990	1.998	2.020	1.988	1.237	0.955	0.946
[1.637]	[1.813]	[1.789]	[1.716]	[1.812]	[1.786]	[0.985]	[0.793]	[0.741]
100								
2.019	1.962	2.061	1.977	1.937	2.024	1.050	0.920	0.810
[1.662]	[1.929]	[1.857]	[1.708]	[1.926]	[1.820]	[1.028]	[0.689]	[0.750]
150								
1.995	1.988	1.954	1.997	1.978	1.937	0.941	0.838	0.790
[1.776]	[1.867]	[1.806]	[1.871]	[1.859]	[1.802]	[1.000]	[0.772]	[0.838]
250								
2.032	1.970	2.016	2.009	1.983	1.979	0.843	0.802	0.723
[1.893]	[1.893]	[1.901]	[1.863]	[1.890]	[1.921]	[0.837]	[0.737]	[0.814]
500								
2.013	2.074	1.964	2.025	2.047	1.976	0.777	0.758	0.731
[1.969]	[1.969]	[2.037]	[2.029]	[1.971]	[1.997]	[0.744]	[0.669]	[0.731]
s1,000								
2.026	2.029	1.936	2.055	2.046	1.965	0.786	0.765	0.800
[2.010]	[2.033]	[1.939]	[2.062]	[2.037]	[1.933]	[0.714]	[0.770]	[0.751]

Panel(c): FarmTreat								
Mean			Median			MSE		
$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
T = 50								
1.933	2.052	2.044	1.938	2.062	2.031	0.800	0.712	0.633
[0.986]	[0.865]	[0.796]	[0.982]	[0.603]	[0.509]	[1.595]	[1.713]	[1.874]
75								
1.995	2.010	1.985	1.990	2.007	1.996	0.913	0.503	0.510
[1.423]	[1.508]	[1.439]	[1.515]	[1.533]	[1.459]	[0.828]	[0.689]	[0.745]
100								
2.007	1.982	2.080	2.000	1.965	2.094	0.565	0.418	0.395
[1.597]	[1.765]	[1.752]	[1.639]	[1.772]	[1.769]	[0.660]	[0.366]	[0.363]
150								
1.997	2.003	1.986	2.050	2.029	1.962	0.509	0.320	0.311
[1.701]	[1.843]	[1.789]	[1.770]	[1.833]	[1.785]	[0.567]	[0.303]	[0.322]
250								
2.019	2.000	2.033	2.017	2.000	2.014	0.363	0.322	0.294
[1.875]	[1.908]	[1.887]	[1.865]	[1.919]	[1.892]	[0.332]	[0.288]	[0.279]
500								
2.037	2.005	1.964	2.049	2.023	1.981	0.262	0.247	0.262
[1.967]	[1.933]	[1.980]	[1.976]	[1.938]	[1.978]	[0.290]	[0.251]	[0.256]
1,000								
1.993	2.008	1.985	2.005	2.003	1.968	0.278	0.264	0.271
[2.007]	[2.013]	[1.945]	[2.008]	[2.042]	[1.957]	[0.245]	[0.283]	[0.259]

NOTE: The table reports descriptive statistics for the average treatment estimation under the null of no effect ($\delta_{T_0+1} = 2$). The table reports the mean, median, and mean squared error (MSE) of the estimator $\widehat{\Delta}$ for one post-intervention observation. Panel (a) considers the case where the counterfactual is estimated by a LASSO regression of the treated unit on all the peers. This is the Artificial Counterfactual (ArCo) approach proposed by Carvalho, Masini, and Medeiros (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), that is, an ordinary least squares (OLS) regression of the treated unit on factors computed from the pool of peers. This is equivalent to the method of Gobillon and Magnac (2016). The number of factors is determined by the eigenvalue ratio test of Ahn and Horenstein (2013). Finally, Panel (c) displays the results of the FarmTreat methodology. Between brackets we report the same statistics but with the model estimated using the full sample as advocated by Chernozhukov, Wüthrich, and Zhu (2020).

Sweets and Candies unit. The chosen products differ in terms of magnitude of sales, price range, and in importance as a share of the company's revenue. For example, the median daily sales per store over the available period and across municipalities vary from 0 (Product V) to 35 units (Product II).

Our sample consists of about 50% of the municipalities where there are stores. As the number and size of stores differ across municipalities, we divide the daily sales at each city by the number of stores in that particular location. To determine the optimal price of each of the products (in terms of profit or revenue maximization) and avoid confounding effects, a randomized controlled experiment has been carried out. For each product, the price was changed in a set of municipalities (treatment group), while in another group, the prices were kept fixed at

Table 3. Rejection rates under the null (empirical size).

	Panel(a): LASSO (ArCo) - Carvalho, Masini, and Medeiros (2018)								
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$	0.294	0.398	0.326	0.398	0.490	0.430	0.452	0.562	0.492
	[0.028]	[0.028]	[0.026]	[0.052]	[0.070]	[0.072]	[0.096]	[0.102]	[0.110]
75	0.156	0.244	0.254	0.260	0.408	0.372	0.354	0.476	0.452
	[0.022]	[0.026]	[0.018]	[0.054]	[0.064]	[0.062]	[0.134]	[0.114]	[0.110]
100	0.096	0.160	0.220	0.210	0.282	0.316	0.288	0.366	0.394
	[0.016]	[0.024]	[0.010]	[0.050]	[0.078]	[0.056]	[0.092]	[0.124]	[0.086]
150	0.090	0.114	0.118	0.166	0.228	0.220	0.252	0.304	0.290
	[0.010]	[0.012]	[0.014]	[0.046]	[0.044]	[0.052]	[0.104]	[0.086]	[0.118]
250	0.064	0.050	0.060	0.146	0.146	0.142	0.198	0.218	0.230
	[0.010]	[0.014]	[0.014]	[0.044]	[0.052]	[0.060]	[0.092]	[0.116]	[0.116]
500	0.032	0.024	0.040	0.110	0.102	0.108	0.172	0.150	0.176
	[0.016]	[0.014]	[0.004]	[0.062]	[0.052]	[0.046]	[0.112]	[0.106]	[0.102]
1,000	0.012	0.024	0.026	0.068	0.096	0.082	0.122	0.166	0.160
	[0.010]	[0.022]	[0.014]	[0.048]	[0.048]	[0.054]	[0.088]	[0.110]	[0.106]
	Panel(b): PCR - Gobillon and Magnac (2016)								
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$	0.152	0.042	0.040	0.216	0.104	0.080	0.242	0.162	0.126
	[0.018]	[0.020]	[0.022]	[0.040]	[0.064]	[0.052]	[0.084]	[0.108]	[0.102]
75	0.100	0.032	0.022	0.134	0.122	0.068	0.194	0.184	0.150
	[0.018]	[0.008]	[0.012]	[0.064]	[0.064]	[0.066]	[0.130]	[0.130]	[0.100]
100	0.086	0.012	0.010	0.138	0.060	0.066	0.194	0.100	0.132
	[0.006]	[0.008]	[0.008]	[0.056]	[0.058]	[0.044]	[0.110]	[0.110]	[0.082]
150	0.084	0.020	0.024	0.128	0.078	0.088	0.176	0.118	0.144
	[0.012]	[0.010]	[0.016]	[0.046]	[0.040]	[0.070]	[0.116]	[0.102]	[0.120]
250	0.026	0.014	0.026	0.080	0.052	0.078	0.128	0.112	0.130
	[0.010]	[0.016]	[0.014]	[0.038]	[0.052]	[0.060]	[0.102]	[0.106]	[0.104]
500	0.018	0.010	0.010	0.060	0.046	0.048	0.110	0.084	0.122
	[0.014]	[0.016]	[0.006]	[0.036]	[0.062]	[0.060]	[0.090]	[0.100]	[0.118]
1,000	0.008	0.002	0.010	0.050	0.056	0.052	0.096	0.102	0.104
	[0.012]	[0.018]	[0.012]	[0.058]	[0.064]	[0.054]	[0.084]	[0.114]	[0.110]
	Panel(c): `FarmTreat`								
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$	0.332	0.400	0.362	0.388	0.468	0.460	0.434	0.532	0.496
	[0.018]	[0.020]	[0.028]	[0.052]	[0.056]	[0.058]	[0.080]	[0.084]	[0.092]
75	0.120	0.084	0.110	0.186	0.182	0.190	0.248	0.292	0.272
	[0.020]	[0.010]	[0.018]	[0.066]	[0.064]	[0.048]	[0.146]	[0.112]	[0.088]
100	0.096	0.028	0.028	0.158	0.096	0.098	0.208	0.168	0.160
	[0.004]	[0.024]	[0.016]	[0.046]	[0.058]	[0.052]	[0.120]	[0.118]	[0.078]
150	0.078	0.022	0.026	0.156	0.080	0.084	0.206	0.134	0.140
	[0.014]	[0.012]	[0.008]	[0.054]	[0.052]	[0.070]	[0.094]	[0.104]	[0.118]
250	0.028	0.006	0.022	0.096	0.066	0.070	0.134	0.136	0.132
	[0.014]	[0.012]	[0.014]	[0.046]	[0.044]	[0.056]	[0.102]	[0.098]	[0.094]
500	0.014	0.010	0.022	0.052	0.058	0.044	0.124	0.090	0.092
	[0.010]	[0.012]	[0.004]	[0.058]	[0.052]	[0.044]	[0.124]	[0.118]	[0.094]
1,000	0.008	0.016	0.012	0.058	0.054	0.060	0.092	0.124	0.114
	[0.012]	[0.022]	[0.008]	[0.052]	[0.060]	[0.050]	[0.110]	[0.114]	[0.098]

NOTE: The table reports the rejection rates of the ressampling test under the null. Panel (a) considers the case where the counterfactual is estimated by a LASSO regression of the treated unit on all the peers. This is the Artificial Counterfactual (ArCo) approach proposed by Carvalho, Masini, and Medeiros (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), that is, an ordinary least squares (OLS) regression of the treated unit on factors computed from the pool of peers. This is equivalent to the method of Gobillon and Magnac (2016). The number of factors is determined by the eigenvalue ratio test of Ahn and Horenstein (2013). Finally, Panel (c) displays the results of the `FarmTreat` methodology. Between brackets we report the rejection rates but with the model estimated using the full sample as advocated by Chernozhukov, Wüthrich, and Zhu (2020).

the original level (control group). Note that the randomization is carried out at the city-level not at the store-level. With the application of our methodology optimal prices can be computed for each city in the treatment group as well as other levels of aggregation. In order to determine the prices for the locations in the control group, the experiment can be repeated by inverting the groups in a second batch of experiments. Here, we will report the results concerning the first group of randomized experiments.

The selection of the treatment and control groups was carried out according to the socioeconomic and demographic characteristics as well as to the distribution of stores in each city. The following variables were used: human development index, employment, GDP per capita, population, female population,

Table 4. Rejection rates under the alternative (empirical power).

Panel(a): LASSO (ArCo) - Carvalho, Masini, and Medeiros (2018)								
$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$								
0.764	0.826	0.856	0.850	0.894	0.906	0.894	0.920	0.926
[0.394]	[0.448]	[0.448]	[0.514]	[0.562]	[0.596]	[0.616]	[0.626]	[0.678]
75								
0.728	0.812	0.806	0.830	0.886	0.882	0.878	0.912	0.918
[0.412]	[0.462]	[0.494]	[0.606]	[0.642]	[0.660]	[0.702]	[0.754]	[0.750]
100								
0.744	0.800	0.816	0.858	0.886	0.888	0.906	0.914	0.916
[0.464]	[0.514]	[0.540]	[0.646]	[0.722]	[0.728]	[0.736]	[0.804]	[0.820]
150								
0.778	0.766	0.766	0.892	0.870	0.858	0.908	0.910	0.900
[0.596]	[0.586]	[0.554]	[0.756]	[0.736]	[0.728]	[0.824]	[0.830]	[0.812]
250								
0.812	0.780	0.808	0.912	0.878	0.892	0.946	0.920	0.922
[0.674]	[0.634]	[0.602]	[0.824]	[0.794]	[0.780]	[0.872]	[0.864]	[0.856]
500								
0.856	0.854	0.836	0.944	0.938	0.932	0.960	0.958	0.964
[0.744]	[0.700]	[0.756]	[0.880]	[0.888]	[0.882]	[0.930]	[0.932]	[0.948]
1,000								
0.860	0.878	0.838	0.948	0.944	0.942	0.962	0.974	0.972
[0.808]	[0.774]	[0.772]	[0.922]	[0.916]	[0.896]	[0.966]	[0.954]	[0.954]

Panel(b): PCR - Gobillon and Magnac (2016)								
$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$								
0.488	0.510	0.448	0.654	0.676	0.660	0.728	0.744	0.740
[0.266]	[0.308]	[0.344]	[0.450]	[0.542]	[0.562]	[0.560]	[0.644]	[0.668]
75								
0.448	0.414	0.426	0.658	0.642	0.622	0.754	0.750	0.722
[0.282]	[0.322]	[0.328]	[0.488]	[0.576]	[0.570]	[0.624]	[0.706]	[0.700]
100								
0.400	0.338	0.390	0.624	0.610	0.646	0.752	0.710	0.744
[0.264]	[0.320]	[0.290]	[0.504]	[0.604]	[0.570]	[0.626]	[0.720]	[0.706]
150								
0.464	0.418	0.398	0.672	0.632	0.630	0.764	0.740	0.738
[0.372]	[0.362]	[0.346]	[0.574]	[0.574]	[0.560]	[0.676]	[0.684]	[0.680]
250								
0.412	0.414	0.400	0.654	0.642	0.650	0.752	0.744	0.754
[0.354]	[0.368]	[0.378]	[0.594]	[0.604]	[0.612]	[0.704]	[0.710]	[0.708]
500								
0.392	0.434	0.376	0.650	0.666	0.628	0.766	0.788	0.750
[0.374]	[0.360]	[0.386]	[0.644]	[0.628]	[0.664]	[0.760]	[0.752]	[0.760]
1,000								
0.412	0.434	0.362	0.668	0.640	0.604	0.760	0.754	0.702
[0.418]	[0.436]	[0.344]	[0.640]	[0.630]	[0.628]	[0.734]	[0.746]	[0.744]

Panel(c): FarmTreat								
$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$	$n = T$	$n = 2 \times T$	$n = 3 \times T$
$T = 50$								
0.826	0.854	0.884	0.892	0.920	0.938	0.912	0.940	0.946
[0.458]	[0.586]	[0.626]	[0.602]	[0.712]	[0.768]	[0.680]	[0.776]	[0.838]
75								
0.774	0.812	0.800	0.894	0.906	0.912	0.940	0.954	0.950
[0.618]	[0.680]	[0.676]	[0.752]	[0.834]	[0.858]	[0.852]	[0.902]	[0.920]
100								
0.758	0.800	0.818	0.886	0.916	0.958	0.934	0.950	0.980
[0.662]	[0.752]	[0.762]	[0.816]	[0.910]	[0.894]	[0.868]	[0.956]	[0.940]
150								
0.852	0.862	0.878	0.950	0.960	0.964	0.978	0.972	0.976
[0.764]	[0.832]	[0.834]	[0.860]	[0.946]	[0.928]	[0.902]	[0.972]	[0.952]
250								
0.884	0.872	0.908	0.952	0.962	0.968	0.972	0.974	0.994
[0.872]	[0.872]	[0.874]	[0.940]	[0.956]	[0.948]	[0.968]	[0.980]	[0.976]
500								
0.918	0.910	0.878	0.978	0.976	0.974	0.990	0.986	0.988
[0.878]	[0.912]	[0.908]	[0.964]	[0.968]	[0.976]	[0.984]	[0.986]	[0.984]
1,000								
0.898	0.916	0.924	0.974	0.978	0.970	0.988	0.988	0.990
[0.918]	[0.890]	[0.888]	[0.974]	[0.974]	[0.968]	[0.990]	[0.984]	[0.988]

NOTE: The table reports the rejection rates of the ressampling test under the alternative. Panel (a) considers the case where the counterfactual is estimated by a LASSO regression of the treated unit on all the peers. This is the Artificial Counterfactual (ArCo) approach proposed by Carvalho, Masini, and Medeiros (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), that is, an ordinary least squares (OLS) regression of the treated unit on factors computed from the pool of peers. This is equivalent to the method of Gobillon and Magnac (2016). The number of factors is determined by the eigenvalue ratio test of Ahn and Horenstein (2013). Finally, Panel (c) displays the results of the FarmTreat methodology. Between brackets we report the rejection rates but with the model estimated using the full sample as advocated by Chernozhukov, Wüthrich, and Zhu (2020).

literate population, average household income (total), household income (urban areas), number of stores, and number of convenience stores. Details about the method can be found in the supplementary material. As mentioned in the Introduction, we used no information about the quantities sold of the product in each municipality to create the treatment and control groups. Therefore, we avoid any selection bias, maintaining valid the assumption that the intervention of interest is independent of the outcomes.

It is important to highlight that although the experiment is randomized, traditional differences-in-differences estimators cannot be considered as the goal is to estimate the price elasticities at the municipal level which is exactly the same level of the randomization. Nevertheless, we can rely on differences-in-
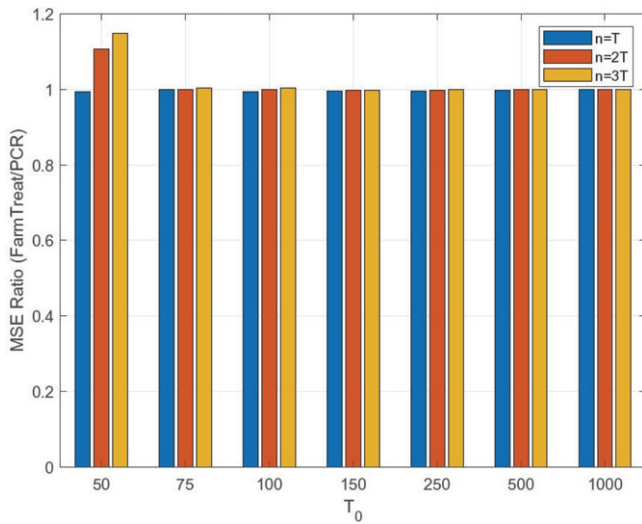
Figure 1. MSE ratio
The figure reports the ratio of the mean squared errors (MSE) of the `FarmTreat`
methodology and Principal Component Regression (PCR) when there is no cross-
dependence among idiosyncratic components.

differences to estimate the intervention effects at the country
level.

6.2. Results

In this section we report the results of the experiment described
in the previous subsection. Table 5 describes each one of the
experiments carried out for each product. The table shows
the sample date, the period of the experiment (usually two
weeks), the type of the experiment (if the price was increased or
decreased), the magnitude of the price change, and the number
of municipalities in the treatment (n_1) and control groups (n_0).
n is the total number of municipalities considered. n, n_0, and
n_1 vary according to the product, but we omit the product
identification to simplify notation.

For each day t, $q_{it}^{(j)}$ represents, for municipality i, the quanti-
ties sold of product j, where $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $j =$
$1, \ldots, 5$. For convenience of notation assume that $i \in \{1, \ldots, n_0\}$
represents cities in the control group and $i \in \{n_0 + 1, \ldots, n\}$
indexes the municipalities in the treatment group. Finally, define
$\widetilde{q}_{it}^{(j)} = q_{it}^{(j)}/n_i^s$, where n_i^s is the number of stores at location i. The
analysis is carried on for $\widetilde{q}_{it}^{(j)}$.

Figure 2 shows the data for the first product considered in the
application. The data for the remaining products are displayed
in Figures S.1–S.4 in the supplementary material. Panel (a) in
the figures reports the daily sales at each group of municipalities
(all, treatment, and control) divided by the number of stores
in each group. More specifically, the plot shows the daily evo-
lution of $q_{\text{all},t\cdot}^{(j)} = \frac{1}{s} \sum_{i=1}^{n} q_{it}^{(j)}$, $q_{\text{control},t\cdot}^{(j)} = \frac{1}{s_0} \sum_{i=1}^{n_0} q_{it}^{(j)}$, and
$q_{\text{treatment},t\cdot}^{(j)} = \frac{1}{s_1} \sum_{i=n_0+1}^{n} q_{it}^{(j)}$. The plot shows the data before
and after price changes and the intervention date is represented
by the horizontal line. Panels (b) and (c) display the distribution
across municipalities of the time averages of $\widetilde{q}_{it}^{(j)}$, before and
after the intervention and for the treatment and control groups,
respectively. Panels (d) and (e) present fan plots for the evolution

of $\widetilde{q}_{it}^{(j)}$. The black curves there represent the cross-sectional
means over time.

Several facts emerge from the plots. First, the dynamics of
sales change depending of the product and the sample con-
sidered. Second, there is a weekly seasonal pattern in the data
which is common to all products. The big spikes for Products II
and IV, observed in Panel (a), are related to major promotions.
We selected these particular products and sample to illustrate
that our methodology is robust to outliers. One point that
deserves attention is that promotions took place in both control
and treatment groups and, therefore, do not have any harmful
implication to our methodology. Eyeballing the graphs in Panel
(a) of Figures 2 and S.4, we observe a substantial drop in sales
before the start of the experiment and happened in both control
in treatment groups. This experiment clearly shows the benefits
of our method in comparison, for instance, with the before-and-
after (BA) estimator. BA estimator does not take into account
common trends or global shocks that affects both treatment and
control groups. Finally, a point to highlight concerning Products
IV and V is the fact the daily sales are quite small as compared
to the other three products. For instance, the average daily sales
per store is less than one unit for Product V. One of the reasons
for the drop in sales for Product V just before the intervention
is a large drop in the number of available units in some of the
municipalities. For example, in about 3% of the municipalities,
both in the treatment and control groups, there were not a single
unit of the product available to be sold. As we are going to see
later, this will have an impact on the results obtained for this
specific product. Finally, by observing Panels (d) and (e), we
notice a significant heterogeneity across municipalities.

The models are estimated at the municipal level. For each
product and each municipality, we run a first-stage regression
of $\widetilde{q}_{it}^{(j)}$ on seven dummies for the days of the week, a linear
deterministic trend and the number of stores that are open at
municipality i on day t. For the municipalities in the control
group the above regression is estimated with the full sample.
For the municipalities in the treatment group we use data only
up to time T_0. The second step consists of estimating factors
for the first-stage residuals. We select the number of factors, k,
by the eigenvalue ratio test. In the third step, we run a LASSO
regression of each idiosyncratic component of treated units on
the idiosyncratic terms of the control group. As described in
Section 4, the penalty parameter is determined by the BIC.
Finally, we compute the counterfactual for each municipality
$i = 1, \ldots, n_1$ for $t = T_0 + 1, \ldots, T$: $\widehat{\widetilde{q}}_{it}^{(j)}$. We also compute
the instantaneous and average intervention impact as $\widehat{\delta}_{it}^{(j)} =$
$\widetilde{q}_{it}^{(j)} - \widehat{\widetilde{q}}_{it}^{(j)}$ and $\widehat{\Delta}_i^{(j)} = \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \widehat{\delta}_{it}^{(j)}$, respectively. We test
the null hypothesis of intervention effect, $\mathcal{H}_0 : \delta_{it}^{(j)} = 0 \,\forall t \geq T_0$,
with the ressampling procedure with either $\phi(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) =$
$\sum_{t=T_0+1}^{T} \widehat{\delta}_t^2$ or $\phi(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} |\widehat{\delta}_t|$. We also test
the for daily effects.

Under the hypothesis of linear demand function, price elas-
ticities ϵ_{ij} for each municipality i and product j can be recovered
as $\widehat{\epsilon}_{ij} = \frac{\widehat{\beta}_{ij} p_{ij,T_0-1}}{\overline{Q}_{ij}}$, where $\widehat{\beta}_{ij} = \frac{\widehat{\Delta}_{ij}}{N_i \Delta_{p_j}}$, $\widehat{\Delta}_{ij}$ is the estimated
average effect for municipality i and product j, N_i is the number
of stores, Δ_{p_j} is the price change, p_{ij,T_0-1} is the price before

Table 5. Experiments.

Product	Sample	Experiment period	Experiment type	Magnitude	Control group	Treatment group
I	Feb-13-2016–Oct-31-2016	Oct-18-2016–Oct-31-2016	Price increase	10.58%	318	97
II	May-14-2016–Jan-23-2017	Jan-17-2017–Jan-23-2017	Price increase	5.01%	321	102
III	Feb-13-2016–Oct-31-2016	Oct-18-2016–Oct-31-2016	Price increase	20.13%	309	106
IV	May-14-2016–Jan-23-2017	Jan-17-2017–Jan-23-2017	Price reduction	18.20%	321	100
V	Aug-14-2016–May-02-2017	Apr-19-2017–May-02-2017	Price reduction	10.05%	328	110

NOTE: The table shows, for each product considered in the article, the sample, the period when the experiment was carried out, the type of the experiment (price increase or decrease), the magnitude of the price change, and the number of cities in the control and treatment groups.
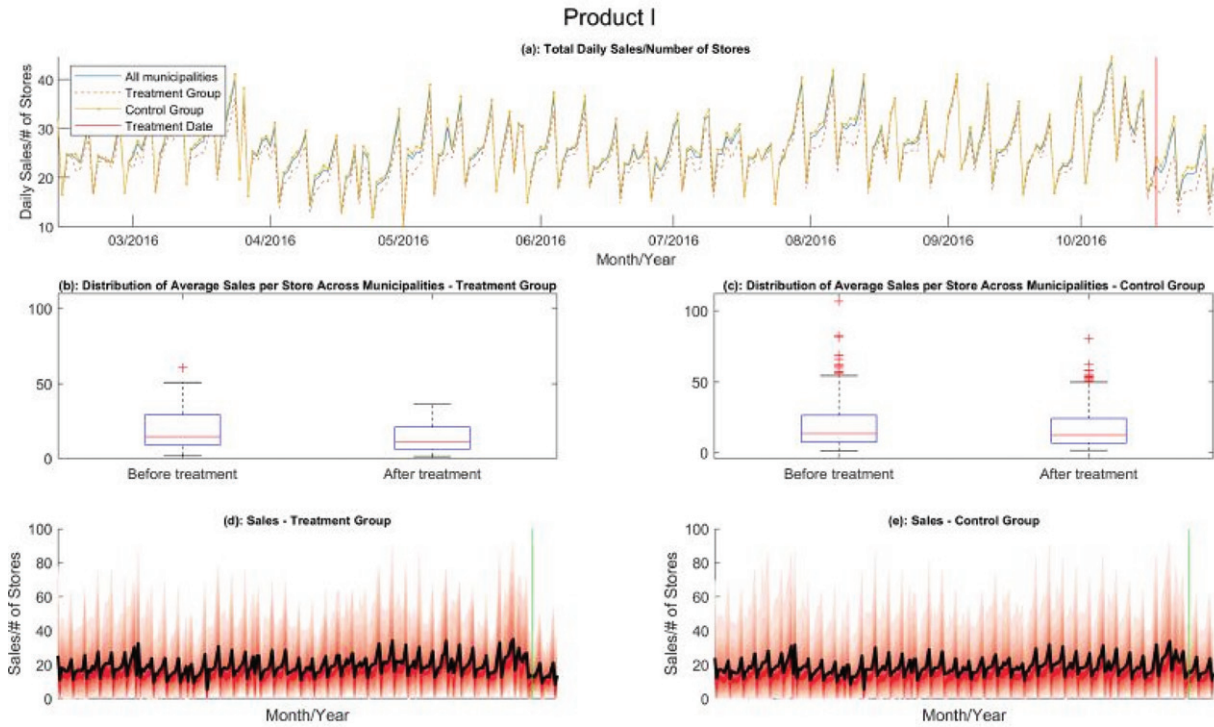

Figure 2. Data for Product I.
Panel (a) reports the daily sales divided by the number of stores aggregated for all cities as well as for the treatment and control groups. The plot also indicates the date of the intervention. Panels (b) and (c) display the distribution of the average sales per store over time across municipalities in the treatment and control groups, respectively. Panels (d) and (e) present fan plots of sales across municipalities in the treatment and control groups for each given time point. The black curves represent the cross-sectional mean over time and the vertical green line indicates the date of intervention.

the intervention and $\overline{Q}_{ij}$ is the average counterfactual quantity sold. Finally, optimal prices for profit maximization can be determined by:

$$p_{ij}^* = \frac{(1 - \text{Taxes}_{ij})(\overline{Q}_{ij} - \widehat{\beta}_{ij} p_{ij,T_0-1}) - \widehat{\beta}_{ij} \times \text{Costs}_{ij}}{-2\widehat{\beta}_{ij}(1 - \text{Taxes}_{ij})},$$

where Taxes_{ij} and Costs_{ij} are the municipality-product-specific tax and costs, respectively.

Table 6 reports, for each product, the minimum, the 5%-, 25%-, 50%-, 75%-, and 95%-quantiles, maximum, average, and standard deviation for several statistics. We consider the distribution over the all treated municipalities. In Panel (a) in the table we report the results for the R-squared of the preintervention model. Panel (b) displays the average intervention effect over the experiment period. Panels (c) and (d) depict the results for the p-values of the ressampling test described in Section 3.3 for the null hypothesis of no intervention effect with the square or absolute value statistic, respectively. Panel (e) presents the results for the p-values of the null hypothesis of no

idiosyncratic contribution. Table 7 presents, for each product, the same descriptive statistics for the estimated elasticities and the percentage difference between the estimated optimal price and the current price. Contrary to what we show in Table 6, in Table 7 we report only results with respect to the municipalities where the estimated average effects have the correct sign (positive when there is a price reduction and negative when there is a price increase) and are statistically significant at the 10% level. The last column in the table shows the fraction of municipalities where the above criterium is satisfied.

Additional results are displayed in Figures 3 and S.5–S.8 in the supplementary material. For each product, Panel (a) in the figures displays a fan plot of the p-values of the ressampling test for the null hypothesis $\mathcal{H}_0 : \delta_t = 0$ for each given t after the treatment, using the test statistic $\phi(\widehat{\delta}_t) = |\widehat{\delta}_t|$, which is the same as using the test statistic $\widehat{\delta}_t^2$. The black curve represents the cross-sectional median across time t. Panel (b) shows an example for one municipality. The panel shows the actual and counterfactual sales per store for the posttreatment period. 95% confidence intervals for the counterfactual path are also displayed.

Table 6. Results: estimation and inference.

				Panel (a): R-squared					
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	0.1134	0.1949	0.3605	0.4913	0.6215	0.7542	0.9077	0.4865	0.1749
II	0.4669	0.7237	0.8744	0.9298	0.9552	0.9863	0.9959	0.8984	0.0926
III	0.1190	0.3094	0.5236	0.7101	0.8341	0.9278	0.9540	0.6750	0.1981
IV	0.4028	0.6744	0.8822	0.9327	0.9633	0.9865	0.9988	0.8972	0.1073
V	0.0366	0.0527	0.1040	0.1670	0.2795	0.4258	0.6517	0.1995	0.1260
				Panel (b): average treatment effect (over time): Δ					
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	−18.9451	−16.3384	−7.7380	−3.2545	−1.1891	1.4580	3.8921	−4.8996	5.3115
II	−44.4114	−26.8298	−14.6979	−7.2601	−3.6173	3.7216	42.2832	−9.0446	11.1052
III	−48.6364	−17.2678	−6.3314	−2.5950	−0.6001	0.7309	9.0886	−4.8875	7.9211
IV	−2.9729	−1.7697	−0.4535	0.2842	1.3057	3.7759	6.5858	0.5660	1.7131
V	−1.5141	−0.9080	−0.4308	−0.1544	0.1455	0.6984	1.6005	−0.1440	0.4949
				Panel (c): p-value of the test on squared values					
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	0	0	0.0755	0.2468	0.5660	0.8623	0.9617	0.3338	0.2961
II	0	0.0091	0.0826	0.3409	0.6116	0.9331	1.0000	0.3774	0.2992
III	0	0	0.0809	0.2872	0.5702	0.9191	0.9702	0.3488	0.2913
IV	0	0	0.1178	0.3781	0.7149	0.9711	1.0000	0.4151	0.3292
V	0	0.0596	0.3021	0.6234	0.9234	1.0000	1.0000	0.5998	0.3285
				Panel (d): p-value of the test on absolute values					
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	0	0	0.0457	0.1660	0.6053	0.8894	0.9957	0.3093	0.3128
II	0	0	0.0620	0.2831	0.6074	0.9190	0.9917	0.3500	0.3059
III	0	0	0.0511	0.2383	0.5957	0.9174	0.9787	0.3333	0.3119
IV	0	0	0.0992	0.3988	0.6860	0.9463	1.0000	0.4113	0.3311
V	0	0.0553	0.2511	0.6553	0.9234	1.0000	1.0000	0.6006	0.3346
				Panel (e): p-value of the test for idiosyncratic contribution					
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	0	0	0.0120	0.0700	0.2415	0.6651	0.7380	0.1570	0.1975
II	0.0120	0.0232	0.0560	0.1400	0.2540	0.5012	0.7000	0.1819	0.1531
III	0	0	0.0140	0.0720	0.1480	0.3296	0.5100	0.1041	0.1155
IV	0.0140	0.0430	0.0970	0.1770	0.3050	0.4400	0.6900	0.2082	0.1365
V	0.0040	0.0120	0.1860	0.3340	0.5120	0.7680	0.8540	0.3524	0.2208

NOTE: The table reports estimation results. In each panel we report, for each product, the minimum, the 5%-, 25%-, 50%-, 75%-, and 95%-quantiles, maximum, average, and standard deviation for a variety of different statistics. We consider the distribution over the treated municipalities. In Panel (a) we report the results for the R-squared of the pre-intervention model. Panel (b) displays the results for the average intervention effect over the experiment period (Δ). Panels (c) and (d) depict the results for the p-values of the ressampling test for the null hypothesis $\mathcal{H}_0 : \delta_t = 0, \forall t \in \{T_0 + 1, \ldots, T\}$ using respectively the test statistic $\phi(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} \widehat{\delta}_t^2$ or $\phi(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} |\widehat{\delta}_t|$. Finally, Panel (e) reports the results for the p-values for the test for the idiosyncratic contribution.

Table 7. Results: elasticities and optimal prices.

				Panel (a): elasticities						
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev	Fraction
I	−6.5063	−4.9033	−3.8701	−3.2394	−2.8788	−1.6952	−0.9111	−3.3737	1.0605	0.2887
II	−17.6172	−15.9339	−12.4079	−9.2212	−5.3922	−2.1923	−1.7181	−9.1675	4.2633	0.2549
III	−3.3421	−3.2692	−2.8714	−2.3356	−1.7182	−1.2854	−1.1852	−2.3187	0.6597	0.2642
IV	−23.7060	−20.9949	−9.8284	−7.2089	−2.9920	−1.8260	−1.7398	−8.4089	6.2820	0.2100
V	−43.7705	−43.7705	−24.8448	−19.7129	−9.8875	−3.6541	−3.6541	−20.3380	13.2366	0.0636
				Panel (b): price discrepancies (% difference)						
Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev	Fraction
I	−15.9781	−13.3710	−10.7336	−8.2254	−6.2937	7.0732	31.2139	−6.8061	8.4973	0.2887
II	−21.4575	−21.1495	−20.2659	−18.8623	−15.0229	−1.1660	4.8060	−16.4801	6.3444	0.2549
III	−10.3946	−10.0601	−7.9305	−3.9473	3.7551	13.5708	16.8302	−1.7678	7.6748	0.2642
IV	−17.6040	−17.2995	−14.5927	−12.7773	−2.5345	7.7192	9.0263	−9.1644	8.3262	0.2100
V	−18.5970	−18.5970	−17.7256	−17.2029	−13.4599	−6.0562	−6.0562	−15.2362	4.5547	0.0636

NOTE: The table reports elasticities estimates as well the percentage difference between the current prices and the optimal price maximizing profit. In each panel we report, for each product, the minimum, the 5%-, 25%-, 50%-, 75%-, and 95%-quantiles, maximum, average, and standard deviation for a given statistic. We consider the distribution over the selected treated municipalities. We only report results concerning the cities where the estimated Δ has the correct sign and the effects are statistical significance at the 10% level. The last column indicates the fraction of cities that satisfy the criterium described above. In Panel (a) we report the results for the estimated elasticities. In Panel (b) we show the results for the difference between the current price and the optimal price.

Several facts emerge from the results. First, the average R-squared are quite high for Products II and IV and moderate for Products I and III. This fact provides some evidence that, on average, the estimated models are able to properly describe the dynamics of the sales. For Products II and IV this finding is even more pronounced as in the worst case, the R-squared are 0.4669 and 0.4028, respectively. On the other hand, the model for Product V yields very low R-squared. A potential reason for
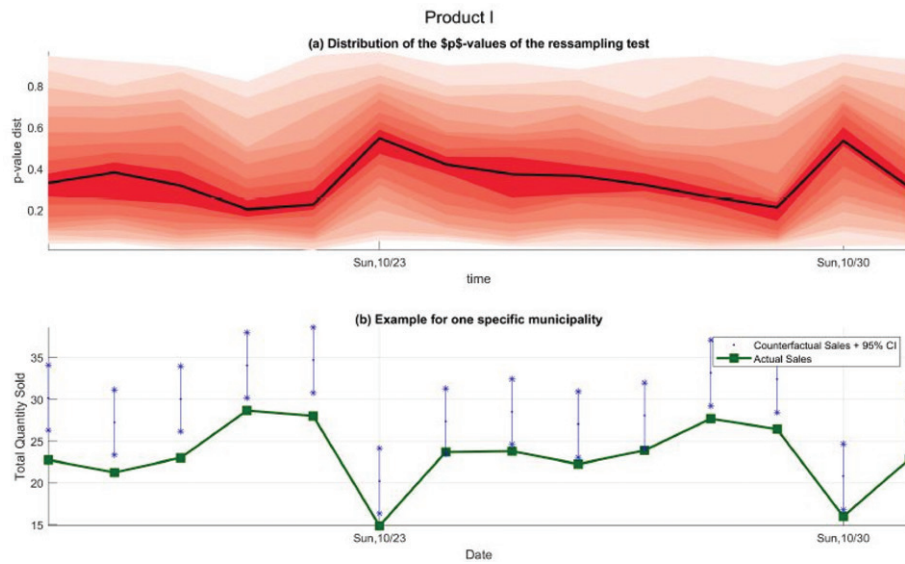
Figure 3. Results for Product I.
Panel (a) displays a fan plot, across n_1 municipalities in the treatment group, of the p-values of the resampling test for the null $\mathcal{H}_0 : \delta_t = 0$ at each time t after the treatment. The black curve represents the median p-value across municipalities over t. Panel (b) shows an example for one municipality. The panel depicts the actual and counterfactual sales per store for the posttreatment period. 95% confidence intervals for the counterfactual path is also displayed.

the poor fit is the fact that sales per store of Product I are very small and there are some municipalities that displays no sales in some days.

The second finding is related to the estimation of the average intervention effect (Δ). As expected, the estimated mean effects ($\widehat{\Delta}$) have the correct sign for Products I–IV, on average. For Product I, Δ has the correct sign (negative) for 90% of the municipalities and is statistically significant at the 10% level in about 37% of all treated cities. Among the cities with $\widehat{\Delta} > 0$, in only one we find statistical significance at the 10% level. For Product II, $\widehat{\Delta}$ has the correct sign for 89% of the treated municipalities and the results are significant in about 33% of all the treated cities. For Product III, the numbers are similar. However, in none of the cities where Δ has been estimated with the opposite sign, the effects are significant. For Product IV, the estimated average treatment effect has the correct sign in 59% of the cities. Fortunately, in the 41 cases where the estimates have the wrong sign, the results are significant in only three of them. For Product V the estimates have the correct sign in only 35% of the cities. However, in only four cities the results with the wrong sign are statistically significant at the 10% level. The reasons for poor results concerning Product V are possibly twofold. First, as mentioned before, the sales per store are quite small and the in-sample fit is poor. Second, there was a stock problem around the time of the experiment. Figure S.9 in the supplementary material displays the evolution of the distribution of available product units across municipalities. From the inspection of Panel (a) in the figure it is clear that the distribution changes around the experiment dates, pointing to large decrease in stocks for Product V.

It is worth comparing the results in Panel (b) of Table 6 with the ones if we use the before-and-after (BA) estimator to compute the average treatment effects. The BA estimator for each municipality is just average sales over the period after the intervention minus the average sales over the days before the intervention. The results are reported in Table S.9 in the sup-

plementary material. As expected from our previous discussion, the BA over estimates the effects of the price changes, specially for Products I and II, and yields estimates with the wrong sign for Product IV. For Product V, the BA estimates are even more negative that the ones from the farmTreat methodology.

A final fact from the inspection of Panel (e) is that the contribution of the idiosyncratic terms to construct the counterfactual is statistically relevant in several cases.

Now we turn attention to Table 7. If we focus only on the cities with estimated average effects that have the correct sign and where the such effects are statistically significant, we estimate very high elasticities on average; see Panel (a) in Table 7. From Panel (b), we note that on average prices must be decreased.

7. Conclusions

In this article we proposed a new method to estimate the effects of interventions when there is potentially only one (or just a few) treated units. The outputs of interest are observed over time for both the treated and untreated units, forming a panel of time series data. The untreated units are called peers and a counterfactual to the output of interest in the absence of intervention is constructed by writing a model relation the unit of interest to the peers. The novelty of this paper concerns how this model is constructed. We combine factor models with sparse regression on the idiosyncratic components. This model includes both the principal component regression and sparse regression on the original measurements as specific cases. The main advantage of our proposal is that we avoid the usual assumption of (approximate) sparsity and make model selection consistency conditions easier to be satisfied. A formal test is also proposed to prove the case for using the idiosyncratic components.

In terms of practical application we show how our methodology can be used to compute optimal prices for products from the retail industry in Brazil. Our results indicate optimal

prices substantially lower than the current prices adopted by the company.

Supplementary Materials

In the Supplementary Material, we report additional empirical findings and the proof of the main result. We also compare the proposed methodology to some alternatives available in the literature.

ORCID

Marcelo C. Medeiros http://orcid.org/0000-0001-8471-4323

References

Abadie, A. (2021), "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 59, 391–425. [574]

Abadie, A., Diamond, A., and Hainmueller, J. (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505. [574]

Abadie, A., and Gardeazabal, J. (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132. [574,576]

Ahn, S., and Horenstein, A. (2013). "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [580,582,583,584,585]

Andrews, D. W. K. (1991). "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [580]

Athey, S., and Imbens, G. (2017). "The State of Applied Econometrics – Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31, 3–32. [574]

Bai, J., and Ng, S. (2002). "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [580]

Bai, J., and Ng, S. (2008). "Large Dimensional Factor Analysis," *Foundations and Trends in Econometrics*, 3, 89–163. [575,579]

Carvalho, C., Masini, R., and Medeiros, M. (2018). "Arco: An Artificial Counterfactual Approach for High-Dimensional Panel Time-Series Data," *Journal of Econometrics*, 207, 352–380. [576,578,581,582,583,584,585]

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2020). "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," arxiv:1712.09089. [575,576,579,580,581,582,583,584,585]

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021a). "A *t*-test for Synthetic Controls," arxiv:1812.10820. [576]

Chernozhukov, V., Wüthrich, K., and Zhu, Y.(2021b). "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*, 116, 1849–1864. [576]

Doudchenko, N., and Imbens, G., (2016). "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis," 22791, NBER. available at arXiv:1610.07748. [574]

Du, Z., and Zhang, L. (2015). "Home-Purchase Restriction, Property Tax and Housing Price in China: A Counterfactual Analysis," *Journal of Econometrics*, 188, 558–568. [576]

Fan, J., Ke, Y., and Wang, K. (2020). "Factor-Adjusted Regularized Model Selection," *Journal of Econometrics*, 216, 71–85. [575,576,577]

Fan, J., Li, R., Zhang, C.-H., and Zou, H. (2020). *Statistical Foundations of Data Science*, Boca Raton, FL: CRC Press. [577]

Fan, J., Masini, R., and Medeiros, M. (2021). "Bridging Factor and Sparse Models," arxiv:2102.11341, Princeton University. [575,579,580]

Gobillon, L., and Magnac, T. (2016). "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98, 535–551. [576,581,582,583,584,585]

Hsiao, C., Ching, H. S., and Wan, S. K. (2012). "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, 27, 705–740. [576,578]

Li, K. (2020). "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," *Journal of the American Statistical Association*, 120, 2068–2083. [576]

Li, K., and Bell, D. (2017). "Estimation of Average Treatment Effects with Panel Data: Asymptotic Theory and Implementation," *Journal of Econometrics*, 197, 65–75. [576]

Masini, R., and Medeiros, M. (2022). "Counterfactual Analysis and Inference with Non-Stationary Data," *Journal of Business and Economic Statistics*, 40, 227–239. [576,579,580]

Masini, R., and Medeiros, M. (2021). "Counterfactual Analysis with Artificial Controls: Inference, High Dimensions and Nonstationarity," *Journal of the American Statistical Association*, 116, 1773–1788. [575,576,577,581]

Ouyang, M., and Peng, Y. (2015). "The Treatment-Effect Estimation: A Case Study of the 2008 Economic Stimulus Package of China," *Journal of Econometrics*, 188, 545–557. [576]