TEXTO PARA DISCUSSÃO

No. 678

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

Jianqing Fan
Ricardo Masini
Marcelo C. Medeiros

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

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

Jianqing Fan
Department of Operations Research and Financial Engineering (ORFE)
Princeton University

Ricardo Masini
Department of Economics
Sao Paulo School of Economics (EESP), Getulio Vargas Foundation

Marcelo C. Medeiros
Department of Economics
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)

November 2, 2020

Abstract

The measurement of treatment (intervention) effects on a single (or just a few) treated unit(s) based on counterfactuals constructed from artificial controls has become a popular practice in applied statistics and economics since the proposal of the synthetic control method. In high-dimensional setting, we often use principal component or (weakly) sparse regression to estimate counterfactuals. Do we use enough data information? To better estimate the effects of price changes on the sales in our case study, we propose a general framework on counterfactual analysis for high dimensional dependent data. The framework includes both principal component regression and sparse linear regression as specific cases. It uses both factor and idiosyncratic components as predictors for improved counterfactual analysis, resulting a method called Factor-Adjusted Regularized Method for Treatment (`FarmTreat`) evaluation. We demonstrate convincingly that using either factors or sparse regression is inadequate for counterfactual analysis in many applications and the case for information gain can be made through the use of idiosyncratic components. We also develop theory and methods to formally answer the question if common factors are adequate for estimating counterfactuals. Furthermore, we consider a simple resampling approach to conduct inference on the treatment effect as well as bootstrap test to access the relevance of the idiosyncratic components. We apply the proposed method to evaluate the effects of price changes on the sales of a set of products based on a novel large panel of sale data from a major retail chain in Brazil and demonstrate the benefits of using additional idiosyncratic components in the treatment effect evaluations.
Keywords: counterfactual estimation, synthetic controls, ArCo, treatment effects, factor models, high-dimensional testing, LASSO, FarmTreat.

1 Introduction

The evaluation of treatment (intervention) effects on a single (or just a few) treated unit(s) based on counterfactuals 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 et al. (2010). Usually, these artificial (synthetic) controls are built from a panel of untreated peers observed over time, before and after the intervention; see Doudchenko and Imbens (2016) and Athey and Imbens (2017) for recent discussions.

The great majority of methods based on artificial counterfactuals 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. Therefore, the dimension of the model to be estimated is frequently large compared to the available number of observations and some sort of restrictions must be imposed. In the original method put forward by Abadie and Gardeazabal (2003), the counterfactual model is linear with coefficients restricted to be positive and must add up to one. Li and Bell (2017) and Carvalho et al. (2018) relaxed the original restrictions by considering penalized estimation of the linear model by Tibishirani's (1996) Least Absolute and Shrinkage Operator (LASSO). Carvalho et al. (2018) derived a number of theoretical results, including consistency and asymptotic normality of the average intervention effect.[1] Their results rely on some sort of model sparsity and the analysis is done under the assumption that the number of observations, both before and after the intervention, diverges.

Sparsity is relaxed by some authors as in Chernozhukov, Wuthrich and, Zhu (2018a,b,c) or Masini and Medeiros (2019). In their papers, the authors assume only approximate sparsity. Some others also relaxed the original restrictions but they only considered a low-dimensional setup. See, for example, Ferman and Pinto (2016), Li (2017), or Masini and Medeiros (2020). Nevertheless, low-dimensional settings do not seem to be realistic for most applications. On the other hand, Gobillon and Magnac (2016) estimate counterfactuals based on pure factor models without exploring potential cross-correlations among the idiosyncratic components.

The aim of this paper is to propose a methodology that includes both principal component regression and sparse linear regression for estimating counterfactals as specific examples for better evaluation of the effects on the sales of a set of products after price changes in our case

[1] The average is taken over the post-intervention period and not over the treated units as in most cases there is only one unit suffering the intervention.

study. It does not impose neither sparsity or approximate sparsity in the mapping between the peers and the treated by using the information from hidden but estimable idiosyncratic components. Furthermore, we show that inferential models where the number of post-intervention observations is fixed can be used in the framework considered in the paper. 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 et al. (2020).

The proposed method consists of four steps. In the first step, the effects of exogenous (to the intervention of interest) variables are removed, for example, deterministic 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 crucial when relaxing the sparsity assumption. To explore potential remaining relation among units, 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 last 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 pre-intervention period. Finally, the model is projected for the post-intervention period under the assumption that the peers do not suffer the intervention. Inspired by Fan et al. (2020), we call the methodology developed here `FarmTreat`, the factor-adjusted regularized method for treatment evaluation.

In terms of theoretical results we show that the estimator of the instantaneous treatment (intervention) is consistent which enable the use of straightforward residual resampling procedures to test general hypotheses about the treatment effect without relying on any asymptotic result for the post-intervention period. All our results are uniquely based on pre-intervention asymptotics. We also show that a bootstrap-based inference for cross-section dependence among idiosyncratic components is valid.

We believe our results are of general importance for the following reasons. First and most importantly, the sparsity or approximate sparsity assumptions 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 methods are violated (Fan et al., 2020). Second, first filtering for trends, seasonal effects and/or outliers seems 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.

We conduct a simulation study to evaluate the finite-sample properties of the estimators and inferential procedures discussed in the paper. 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 1,400 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.

The rest of the paper 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 3.1. Inferential procedures are presented in Section 3.2. We present the results of a simulation experiment in Section 4 and discuss the empirical application in detail in Section 5. Section 6 concludes the paper. Finally, the proof of our theoretical result is relegated to the Appendix.

2 Overview of Case Study and the Methodology

This section first briefly describes the problem for our case study and then summarizes the methods that we develop for evaluating the treatment effects. The proposed methods have broader applications than what we applied here.

2.1 Case study

The overarching goal is to optimalize price setting in the retail industry in Brazil via counterfactual analysis. Price changes affect the quantities of sales and the counterfactor analysis is to determine the amounts of changes in sales. Our dataset consists of the daily prices and quantities sold of five different products commercialized by one of the major retail chains in Brazil, aggregated at the municipal level. The company has more than 1,400 stores distributed in more than 400 municipalities over the country.[2] The chosen products differ in terms of magnitude of sales and in importance as a share of the company's total revenue.

Our sample consists of about 50% of the municipalities where there are stores. As the number and size of stores differ across municipalities, we will present the results in terms of total sales per store. To determine the optimal price of each of the products (in terms of profit or revenue maximization), 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 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.

2.2 `FarmTreat`

The dataset is a realization of $\{Z_{it}, \boldsymbol{W}_{it} : 1 \leqslant i \leqslant n, 1 \leqslant t \leqslant T\}$, in which Z_{it} is the quantity of sale in the municipality i at time t and $\boldsymbol{W}_{it}$ describes heterogeneity of municipality i at that time, including seasonal pattern (days of the weeks). Suppose we are interested in estimating the effects on the variable in Z_{1t} of the first unit after an intervention that occurred at $T_0 + 1$. We estimate a counterfactual based on a number peers $\boldsymbol{Z}_{-1t} := (Z_{2t}, \dots, Z_{nt})'$ that are assumed

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

to be unaffected by the intervention. We allow the dimension of $\boldsymbol{Z}_{-1t}$ to grow with the sample size T, i.e. $n := n_T$. We also assume that there are a number of covariates $\boldsymbol{W}_{it}$ which are not affected by the intervention. Our key idea is to use both information in the latent factors and idiosyncratic components, called `FarmTreat`.

The procedure is thus summarized by the following steps:

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

$$Z_{it} = \boldsymbol{\gamma}_i' \boldsymbol{W}_{it} + R_{it}, \quad t = 1, \ldots, T_0,$$

   and compute the residuals $\widehat{R}_{it} := Z_{it} - \widehat{\gamma}_i' \boldsymbol{W}_{it}$. This step removes heterogeneity due to $\boldsymbol{W}_{it}$.

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

$$\boldsymbol{R}_t = \boldsymbol{\Lambda} \boldsymbol{F}_t + \boldsymbol{U}_t,$$

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

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

   where $\widehat{\boldsymbol{U}}_t = \left(\widehat{U}_{1t}, \ldots, \widehat{U}_{nt}\right)'$. There is a large literature on high-dimensional factor analysis; see the book by Fan et al. (2020) for detail.

3. The third estimation 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. If the null is rejected, fit the model in the pre-intervention period

$$\widehat{U}_{1t} = \boldsymbol{\theta}_1' \widehat{\boldsymbol{U}}_{-1t} + V_t, \quad t = 1, \ldots, T_0,$$

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

$$\widehat{\boldsymbol{\theta}}_1 = \mathsf{arg\,min}\left[\sum_{t=1}^{T_0}\left(\widehat{U}_{1t} - \boldsymbol{\theta}_1'\widehat{\boldsymbol{U}}_{-1t}\right)^2 + \xi\|\boldsymbol{\theta}_1\|_1\right]. \tag{2.1}$$

This step uses cross-sectional regression of the idiosyncratic components to estimate that in the treated unit. The model includes sparse linear model on $\boldsymbol{R}_t$ as a specific example (see (2.3) below) and the required model selection conditions are more easily met due to the factor adjustments. It also encompass the principal component regression in which $\widehat{\boldsymbol{\theta}}_1 = 0$, namely, using no cross-sectional prediction.

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

$$\widehat{\delta}_t = Z_{1t} - \left(\widehat{\gamma}_1'\boldsymbol{W}_{1t} + \widehat{\boldsymbol{\lambda}}_1'\widehat{\boldsymbol{F}}_t + \widehat{\boldsymbol{\theta}}_1'\widehat{\boldsymbol{U}}_{-1t}\right). \tag{2.2}$$

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

5. Use the estimator to test for null hypothesis of no intervention effect in the form[3]

$$\mathscr{H}_0 : \delta_t = 0, \quad t \in \{T_0 + 1, \ldots, T\},$$

where δ_t is the (possibly random) intervention effect for periods $t \in \{T_0 + 1, \ldots, T\}$.

The innovations of our approach in estimating counterfactuals are multi-folds. For simplicity, let us suppose that we have no $\boldsymbol{W}_t$ component, so that $\boldsymbol{R}_t = \boldsymbol{Z}_t$. 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 $\boldsymbol{\theta}_1 = 0$. Clearly, the method explores the sparsity of $\boldsymbol{\theta}_1$ to improve the performance and also includes the case of sparse regression on $\boldsymbol{Z}_{-1t}$ to estimate

[3]Clearly, we can also accommodate heterogeneous null hypothesis of the form $\delta_t = c_t$ for given constants c_t.

counterfactuals as in Masini and Medeiros (2019), where counterfactuals are estimated as

$$Z_{1t} = \boldsymbol{\theta}_1' \boldsymbol{Z}_{-1t} + \epsilon_t, \quad t = 1, \cdots, T_0.$$

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

$$Z_{1t} = \boldsymbol{\theta}_1' \boldsymbol{\Lambda}_{-1} \boldsymbol{F}_t + \boldsymbol{\theta}_1' \boldsymbol{U}_{-1t} + \epsilon_t, \tag{2.3}$$

where $\boldsymbol{\Lambda}_{-1}$ and $\boldsymbol{U}_{-1t}$ are defined as $\boldsymbol{\Lambda}$ and $\boldsymbol{U}_t$ without the first row. When we take $\boldsymbol{\lambda}_1 = \boldsymbol{\theta}_1' \boldsymbol{\Lambda}_{-1}$, this reduces to use sparse regression to estimate the counterfactuals, but now use more powerful `FarmSelect` of Fan et al. (2020) to fit the sparse regression. Again, `FarmSelect` imposes the condition $\boldsymbol{\theta}_1' \boldsymbol{\Lambda}_{-1}$ as the regression coefficients of $\boldsymbol{F}_t$. 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 propose 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 pre-intervention period.

2.3 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 bounded deterministic functions of time, i.e, $f(t/T)$, in order to capture trends, an intercept to remove the mean, seasonal dummies or other calendar effects, or any other dummy to remove potential outliers. 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 pre-intervention 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. Note that $\mathbb{E}(\boldsymbol{R}_t) = \mathbf{0}$ by definition. Hence, we can replace the post-intervention observations of R_{1t} by 0 in order to carry the factor analysis. As the number of post-intervention 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.

2. The other alternative 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.

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 3.2. 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 (2019).

3 Assumptions and Theoretical Result

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 $\boldsymbol{Z}_t := (Z_{1t}, \ldots, Z_{nt})'$.[4] 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 was in place.

[4]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.

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 $Z_{it}^{(1)}$ denotes 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. Without loss of generality, we set unit 1 to be the one of interest. The null hypothesis to be tested is:

$$\mathscr{H}_0 : \delta_t := Z_{1t}^{(1)} - Z_{1t}^{(0)} = 0, \quad \forall t > T_0. \tag{3.1}$$

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* – i.e., what would Z_{1t} have been like had there been no intervention (potential outcome).

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

$$Z_{1t}^{(0)} = \mathcal{M}\left(\boldsymbol{Z}_{-1t}^{(0)}; \boldsymbol{\theta}\right) + V_t, \quad t = 1, \ldots, T, \tag{3.2}$$

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).[5], $\mathcal{M} : \mathcal{Z} \times \boldsymbol{\Theta} \to \mathbb{R}$, $\mathcal{Z} \subseteq \mathbb{R}^{n-1}$, 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 pre-intervention 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* $1 \leqslant s, t \leqslant T$.

[5]We could also have included lags of the variables and/or exogenous regressors into $\boldsymbol{Z}_{0t}$, but again, to keep the argument simple, we have considered only contemporaneous variables; see Carvalho et al. (2018) for more general specifications.

The main idea is to estimate (3.2) using just the pre-intervention sample, $t = 1, \ldots, T_0$, since under Assumption 1, $\boldsymbol{Z}^{(0)}_{-1t} = \boldsymbol{Z}_{-1t} := (Z_{2t}, \ldots, Z_{nt})'$ in the pre-intervention period. Consequently, the estimated counterfactual for the post-intervention period, $t = T_0 + 1, \ldots, T$, becomes $\widehat{Z}^{(0)}_{1t} := \mathcal{M}(\boldsymbol{Z}_{0t}; \widehat{\boldsymbol{\theta}}_{T_0})$. Under some sort of stationary assumption on $\boldsymbol{Z}_t$, in the context of a linear model, Hsiao et al. (2012) and Carvalho et al. (2018), show that $\widehat{\delta}_t := Z_{1t} - \widehat{Z}^{(0)}_{1t}$ is an unbiased estimator for δ_t as the pre-intervention 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 $\{Z^{(0)}_{it} : 1 \leqslant i \leqslant n, t \geqslant 1\}$ is generated by*

$$Z^{(0)}_{it} = \boldsymbol{\gamma}'_i \boldsymbol{W}_{it} + \boldsymbol{\lambda}'_i \boldsymbol{F}_t + U_{it} \tag{3.3}$$

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, deterministic trends, seasonal dummies or any other exogenous (possibly random) characteristic of unit i that the practitioner judge to be helpful in the construction of the counterfactual. Our counterfactual model is nothing more than the sample version of the projection of $Z^{(0)}_{1t}$ onto the space spanned by $(\boldsymbol{W}_{1t}, \boldsymbol{F}_t, \boldsymbol{U}_{-1,t})'$. Under Assumption 2 the counterfactual can be taken as

$$Z^{(0)}_{1t} = \boldsymbol{\gamma}'_1 \boldsymbol{W}_{1t} + \boldsymbol{\lambda}'_1 \boldsymbol{F}_t + \boldsymbol{\theta}'_1 \boldsymbol{U}_{-1t} + V_t, \tag{3.4}$$

where $\boldsymbol{\theta}_1$ is the coefficient of the linear regression of $\boldsymbol{U}_{1t}$ onto $\boldsymbol{U}_{-1t}$ and V_t the respective regression error.

3.1 Theoretical Result

In order to state our result in a precise manner we consider the following technical assumption

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

(a) *The covariance matrix of* $\boldsymbol{W}_{1t}$ *is non-singular;*

(b) $\mathbb{E}|W_{it\ell}|^p \leqslant C$ *and* $\mathbb{E}|U_{it}|^{p+\epsilon} \leqslant C$ *for some* $p \geqslant 4$ *and* $\epsilon > 0$ *for* $i \in [n]$, $t \in [T]$ *and* $\ell \in [k]$*;*

(c) *The process* $\{(\boldsymbol{F}'_t, \boldsymbol{U}'_t)', t \in \mathbb{Z}\}$ *is weakly stationary with strong mixing coefficient* α *satisfying* $\alpha(m) \leqslant \exp(-2cm)$ *for some* $c > 0$ *and for all* $m \in \mathbb{Z}$*;*

(d) $\|\boldsymbol{\theta}_1\|_\infty \leqslant C$*;*

(e) $\kappa_0 := \kappa\left(\mathbb{E}\boldsymbol{U}_t\boldsymbol{U}'_t, \mathcal{S}_0, 3\right) \geqslant C^{-1}$ *where* $\kappa()$ *is the compatibility condition defined in* (A.1) *in the Appendix and* $\mathcal{S}_0 := \{i : \theta_{1,i} \neq 0\}$.

A few words on the assumptions above are in order. Condition (a) is necessary for the linear projection parameter $\boldsymbol{\gamma}_1$ to be well defined. Conditions (b) and (c) taken together are sufficient for a law of large number for strong mixing processes that can 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 et al. (2020), we could state a stronger result in terms of the allowed number of non-zero coefficients as a fraction of the same 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 Proposition 1.

Finally, conditions (d) and (e) in Assumption 3 are regularity condition on the high-dimensional linear model to be estimated by LASSO in step 3. Condition (e) ensures the (restricted) strong convexity of the objective function which is necessary for consistently estimate $\boldsymbol{\theta}_1$ when $n > T$. In effect, it lower bounds the minimum restricted ℓ_1-eigenvalue of the covariance matrix of $\boldsymbol{U}_t$ uniformly. For simplicity, the bounds appearing in (d) and (e) were assumed to hold uniformly. However, both conditions could also be somewhat relaxed to allow $\|\boldsymbol{\theta}_1\|_\infty$ 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 condition of Proposition 1.

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

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

(b) $|\mathcal{S}_0| = O\left(\left\{\eta\left[(nT)^{1/p} + \eta\right] + \frac{n^{4/p}}{\sqrt{T}}\right\}^{-1}\right)$.

If the penalty parameter ξ in (2.1) *is set to be at the order of $\frac{n^{2/p}}{\sqrt{T}} + \eta T^{1/p}$ then, as $T_0 \to \infty$, $\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1 = O_P\left(\xi|\mathcal{S}_0|\right)$, and for every $t > T_0$:*

$$\widehat{\delta}_t - \delta_t = V_t + O_P\left\{|\mathcal{S}_0|\left[\eta(nT)^{1/p} + \frac{n^{3/p}}{\sqrt{T}}\right]\right\},$$

where V_t is the stochastic component not explainable by untreated units defined by (3.4)

Remark 1. *Condition (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 the number of non-zero coefficients in $\boldsymbol{\theta}_1$ (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 $\boldsymbol{W}_i$ 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, Corollary 1 in Fan et al. (2020) state conditions under which $\eta = \frac{n^{6/p}}{T^{1/2-3/p}} + \frac{T^{1/p}}{\sqrt{n}}$.*

Proposition 1 is key for our inference procedure discussed in Section 3.2. For instance, it can be used to argue that $\widehat{\delta}_t - \delta_t = V_t + o_p(1)$ provided that $|\mathcal{S}_0|\left[\eta(nT)^{1/p} + \frac{n^{3/p}}{\sqrt{T}}\right] = o(1)$. 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 post-intervention period. Furthermore, as described below, we can estimate the quantiles of V_t using the pre-intervention residuals to conduct a valid inference on δ_t.

3.2 Testing for Intervention Effect

The inference procedure presented in this section is based on the sequence of estimators $\{\widehat{\delta}_t\}_{t>T_0}$ and is grounded on the results of Masini and Medeiros (2019a,b). Let $T_2 := T - T_0$ be the number of observations after the intervention and define a generic continuous mapping $\boldsymbol{\phi} : \mathbb{R}^{T_2} \to \mathbb{R}^b$ whose argument is the T_2-dimensional vector $(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)'$.

We are interested in the distribution of $\widehat{\boldsymbol{\phi}} := \boldsymbol{\phi}(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)$ under the null (3.1), where $\boldsymbol{\phi}$ is a given vector of function such as the average treatment effect, median treatment effect, or maximum treatment effect, among others. The statistic $\widehat{\boldsymbol{\phi}}$ is used to test the presence of the treatment effect. The typical situation is the one where the pre-intervention period is much longer than the post intervention period, $T_0 \gg T_2$. In several cases, 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 and consistency of the treatment effect estimator, it is possible to resample the pre-intervention residuals following the procedure described in Masini and Medeiros (2019a,b) to compute the sample quantile of the statistic of interest.

Under the asymptotic limit taken on the pre-invention 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 $\boldsymbol{\phi}_0$. Under the strict stationary assumption of $\{V_t\}$, we can use the pre-intervention period information to estimate the distribution of $\widehat{\boldsymbol{\phi}}$. Consider the construction of $\widehat{\boldsymbol{\phi}}$ using only blocks of size T_2 of consecutive observations from the pre-intervention sample. There are $T_0 - T_2 - 1$ such blocks denoted by

$$\widehat{\boldsymbol{\phi}}_j := \boldsymbol{\phi}(\widehat{V}_j, \ldots, \widehat{V}_{j+T_2-1}) \quad j = 1, \ldots, T_0 - T_2 + 1,$$

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 pre-intervention period, the same as in (2.2).

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 $\mathcal{Q}_T(\boldsymbol{x}) := \mathbb{P}(\widehat{\boldsymbol{\phi}} \leqslant \boldsymbol{x})$ by its empirical distribution

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

where, for a pair of vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, we say that $\boldsymbol{a} \leqslant \boldsymbol{b} \iff a_i \leqslant b_i, \forall i$. Finally, Theorem 2 in Masini and Medeiros (2019) establish condition under which

$$\sup_{\boldsymbol{x}} |\widehat{\mathcal{Q}}_T(\boldsymbol{x}) - \mathcal{Q}_T(\boldsymbol{x})| = o_p(1) \qquad \text{as } T_0 \to \infty.$$

3.3 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, let us write the DGP 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 $\boldsymbol{\Gamma}$ is such that each block is given by $(\boldsymbol{\gamma}'_1, \ldots \boldsymbol{\gamma}'_n)$. Finally, $\boldsymbol{\Lambda} := (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)'$.

Let $\boldsymbol{\Pi} := (\pi_{ij})_{1 \leqslant i,j \leqslant n}$ denote the $(n \times n)$ covariance matrix of $\boldsymbol{U}_t$. Our proposed 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: $\mathscr{H}_{0,2} : \pi_{1j} = 0,\ \forall\ 2 \leqslant j \leqslant n$.

In order to conduct the test we propose the following test statistic

$$S := \|\boldsymbol{Q}\|_\infty,$$

where $\boldsymbol{Q} := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{D}_t$, $\boldsymbol{D}_t := \widehat{U}_{1t} \widehat{\boldsymbol{U}}_{-1t}$, and $\widehat{U}_{it} := \hat{R}_{it} - \widehat{\boldsymbol{\lambda}}_i{}' \widehat{\boldsymbol{F}}_t$. 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

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

is the estimator of the long-run covariance matrix $\boldsymbol{\Upsilon} := \mathbb{V}\widetilde{\boldsymbol{Q}}$, where $\widetilde{\boldsymbol{Q}} := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{1t} \boldsymbol{U}_{-1t}$. 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 et al. (2020) under a more general setting. In particular, the authors show under some regularity conditions

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

4 Simulations

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

$$
\begin{aligned}
Z_{it} &= \delta_{it} + \boldsymbol{\gamma}_i' \boldsymbol{W}_t + R_{it} \\
R_{it} &= \boldsymbol{\lambda}_i' \boldsymbol{F}_t + U_{it} \\
\boldsymbol{F}_t &= (0.8\boldsymbol{I})\boldsymbol{F}_{t-1} + \boldsymbol{V}_t \\
U_{it} &= \begin{cases} \boldsymbol{\beta}' \boldsymbol{U}_{-1t} + \varepsilon_{it}, & \text{if } i = 1, \\ \varepsilon_{it}, & \text{otherwise}, \end{cases}
\end{aligned}
\tag{4.1}
$$

where $\{\varepsilon_{it}\}$ is a sequence of independent and normally distributed zero-mean random variables with variance equal to 0.25, $\boldsymbol{V}_t$ is a sequence of independent and normally distributed zero-mean random vectors taking values on $\mathbb{R}^2$ such that $\mathbb{E}(\boldsymbol{V}_t \boldsymbol{V}_t') = \boldsymbol{I}$, and $\mathbb{E}(\varepsilon_{it} \boldsymbol{V}_s) = \boldsymbol{0}$, for all i, t, and s. The parameters are set as follows: $\boldsymbol{\gamma}_i$ is p-dimensional vector of ones and, for each replication, the elements of $\boldsymbol{\lambda}_i$, $i > 1$, are drawn independently from a normal distribution with mean two and unit variance and, for $i = 1$, the elements of '$\boldsymbol{\lambda}_i$ are drawn from a normal distribution with mean -6 and variance 0.04. The first two elements of $\boldsymbol{\beta}$ are set to one and the rest is set to zero. We consider the following sample sizes: $T_0 = 50, 100$, and 500; and $T_2 = 5$. For each sample size, n is set as $n = \{T, 2T, 3T\}$. p is kept fixed and equal to five. 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 $t > T_0$.

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 $\widehat{\Delta} = \frac{1}{T-T_0} \sum_{t=T_0}^{T} \widehat{\delta}_t$ 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 et al. (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.

From the inspection of the results in the tables, it is clear that the biases for estimating of the treatment effect are negligible and MSEs are predominately the variance. Furthermore, the ArCo delivers very robust estimates, but the MSE can be substantially reduced by the `FarmTreat` methodology when $T > 50$. Therefore, there is strong evidence supporting methodology derived in this paper, which is consistency with our theoretical results. Furthermore, as expected, the MSE decreases as the sample size increases. Second, as already shown in the simulations in Carvalho et al. (2018), the performance of the pure factor model is poor. 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.

Table 3 presents the empirical size of the resampling test when $T_2 = 5$ and the counterfactual is estimated according to the three methods described above. By inspection of the results it is clear that all methods have negligible size distortions. This demonstrates the validity of our bootstrap methods.

Table 4 presents equivalent statistics to the ones in Table 1 but the DGP has no idiosyncratic contribution, i.e., $\boldsymbol{\beta} = \mathbf{0}$. This case favors to PCR, which is indeed when n and T are sufficiently large. As we can see, `FarmTreat` achieves the best results in terms of MSE reduction. This shows that when n and T are large, the factors are well estimated and `FarmSelect` performs as we as the PCR method which is the best by design. This shows that `FarmSelect` adapts well to this specific case. When n or T are not sufficiently large, latent factors are not well estimated and PCR does not perform as well as expected. In this case, `FarmSelect` augments further the prediction power by using the idiosyncratic components when the latent factors are not well estimated.

5 Application: Price Elasticity of Demand

In this section we report the results of the experiment described in Section 2. 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) 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.

Figures 1–5 show the data considered in the application. For each product, Panel (a) in each figure reports the sales per store aggregated in 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 in the treatment and control groups, respectively. Panels (d) and (e) present fan plots for the evolution of sales per store for each municipality. The black curves there represent the cross-sectional medians over time. Several interesting facts emerge from the plots. First, the dynamics of sales change depending of the product and the sample. Nevertheless, there is a clear weekly seasonal pattern in the data. The big spikes in Panel (a) of Figures 2 and 4 are related to major promotions. We selected this particular product/sample to illustrate that our methodology is robust to outlying observations. One important 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. The experiment involving Product I was a price decrease and we expect, as a consequence, a positive impact on sales. However, eyeballing the graph displayed in Panel (a) of Figure 1, we see a major drop in sales around the date of the experiment. The histograms in Panels (b) and (c) corroborate this fact. However, the fall in sales happened before the beginning of the experiment and happened in both control in treatment groups. We like this experiment as it clearly shows the benefits of our method in comparison, for instance, with the before-and-after estimator. The latter will for sure indicate a negative impact of the price reductions. Finally, observing Panels (d) and (e) in the figures, it is easy to notice a significant heterogeneity across municipalities.

We continue by estimating the models discussed in this paper. For each day t, $q_{it}^{(j)}$ represents the total quantities sold per store of product j in municipality i, where $i = 1, \dots, n$, $t = 1, \dots, T$, and $j = 1, \dots, 5$. For each product and each municipality, we run a first-stage regression of quantities 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 described in Ahn and Horenstein (2013). 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 2.3, the penalty parameter is determined by the BIC. Finally, we compute the counterfactual for each municipality $i = 1, \ldots, n_1$ for $t = T_0 + 1, \ldots, T$: $\widehat{q}_{it}^{(j)}$. We also compute the instantaneous and average intervention impact as $\widehat{\delta}_{it}^{(j)} = q_{it}^{(j)} - \widehat{q}_{it}^{(j)}$ and $\widehat{\Delta}_i^{(j)} = \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \widehat{\delta}_{it}^{(j)}$, respectively.

We consider the null hypothesis of no intervention effect as in (3.1). The results are displayed in Figures 6–10 and in Table 6. For each product, Panel (a) in the figures displays a fan plot of the p-values of the re-sampling test for the null hypothesis $\mathscr{H}_{0,1} : \delta_t = 0$ for each t after the treatment, using the test statistics $\phi_1(\widehat{\delta}_t) = |\widehat{\delta}_t|$, which is the same as using the test statistic $\widehat{\delta}_t^2$. The black curve represents the cross-sectional median across time t. Panels (b) and (c) display the distribution of the p-values of the re-sampling tests for the null

$$\mathscr{H}_0 : \delta_t = 0, \forall t \in \{T_0 + 1, \ldots, T\}$$

using the test statistics $\phi_2(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} \widehat{\delta}_t^2$ and $\phi_3(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} |\widehat{\delta}_t|$, respectively. Panel (d) shows an example for one municipality. The panel shows the actual and counterfactual sales per store for the post-treatment period. 95% confidence intervals for the counterfactual path are also displayed.

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 treated municipalities. In Panel (a) in the table we report the results for the R-squared of the pre-intervention model. Panel (b) displays the p-value results for testing the average intervention effect $\mathscr{H}_{0,1} : \delta_t = 0$ over the experiment period across different treated municipalities. It summarizes the results presented in Panel (a) of Figures 6–10. In particular, the average and the median of the average treatment effects across treated municipalities are also presented there. Panels (c) and (d) depict the results for the p-values of the re-sampling test described in Section 3.2 for the null hypothesis $\mathscr{H}_0 : \delta_t = 0, t = T_0+1, \ldots, T$, using, respectively, the test statistics $\phi_1(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} \widehat{\delta}_t^2$ and $\phi_2(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} |\widehat{\delta}_t|$. Panel (e) presents the results for the p-values of the null hypothesis of no idiosyncratic contribution.

A number of conclusions emerge from the figures and the table. First, apart from Product I, the pre-intervention model in general fits the data quite well as can be attested by the large values of the R-squared. Nevertheless, there is some variation in terms of the goodness-of-fit across municipalities. The low quality of the fit is, in most cases, associated with cities with a very small number of stores and few sales. Second, there is a huge heterogeneity in terms of intervention effects across different municipalities as can be seen from Panels (a)-(c) in the Figures and Panel (b) in the table. For Product I, the price intervention has effects only on a small number of municipalities. More specifically, according to the re-sampling test for $\mathcal{H}_0$, the impacts are statistically relevant (at a 1% level) only on three out of 110 municipalities. As expected, the average effect is positive in all cases. This is not surprising as Product I has very low sales. The maximum value for Δ over the municipalities is less than 2 units per store. This is not surprising as the median sale for this product is zero.

The same pattern of heterogeneity can be found in Product II. However, there are more cases where the price changes had significant effects: 12 out of 100 with 1% significance. This result doubles if we consider 10% significance level. The values for Δ are also much higher.

For Product III the impacts are much more significant: at a 1% significance level there are 15 cities with relevant impacts when the squares statistic is used to test for $\mathcal{H}_0$ and 23 when the absolute value is used. If we set the significance level to 10% the numbers move to 31 and 41, respectively. Products IV and V have a similar behavior as Product III.

Under the hypothesis of linear demand function, price elasticities 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 the intervention and $\overline{Q}_{ij}$ is the average counterfactual quantity sold. Finally, optimal prices for profit maximization can be determined by:

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

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

6 Conclusions

In this paper we considered a new methodology to estimate the effects of interventions when there is potentially only one (or just a very small number) of 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. In our case 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 a specific case. 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. The inadequacy of using only the principal component regression has also been evidenced in our case studies. The formal test is also proposed to prove the case for using the idiosyncratic components.

Acknowledgments: Fan's research was supported by NSF grant DMS-1712591, NIH grant 2R01-GM072611-15, and ONR grant N00014-19-1-2120. Masini's and Medeiros' research was partially supported by CNPq and CAPES.

A Proof of the Main Result

Before proving our main result, we define below the compatibility constant for convenience.

Definition 1. *For a $(n \times n)$ matrix $\boldsymbol{M}$, a set $\mathcal{S} \subseteq [n]$ and a scalar $\zeta \geqslant 0$, the compatibility constant is given by*

$$\kappa(\boldsymbol{M}, \mathcal{S}, \zeta) := \inf \left\{ \frac{\|\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}\|}{\sqrt{|\mathcal{S}|}} \|\boldsymbol{x}_{\mathcal{S}}\|_1 : \boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}_{\mathcal{S}^c}\|_1 \leqslant \xi \|\boldsymbol{x}_{\mathcal{S}}\|_1 \right\}. \tag{A.1}$$

Moreover, we say that $(\boldsymbol{M}, \mathcal{S}, \zeta)$ satisfies the compatibility condition if $\kappa(\boldsymbol{M}, \mathcal{S}, \zeta) > 0$.

Note that the compatibility constant is closely related to ℓ_1-eigenvalue of $\boldsymbol{M}$ restricted to a cone in $\mathbb{R}^n$.

A.1 Proof of Proposition 1

The fact that $\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1 = O_P(\xi|\mathcal{S}_0|)$ follows from Theorem 4 in Fan et al. (2020). We are left to show the second part. By the triangle inequality, for $t > T_0$:

$$\begin{aligned}
|\widehat{\delta}_t - \delta_t - V_t| &= |(\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)'\boldsymbol{W}_{1t} + \widehat{\boldsymbol{\lambda}}_1'\widehat{\boldsymbol{F}}_t - \boldsymbol{\lambda}_1'\boldsymbol{F}_t + \widehat{\boldsymbol{\theta}}_1'\widehat{\boldsymbol{U}}_{-1t} - \boldsymbol{\theta}_1'\boldsymbol{U}_{-1t}| \\
&\leqslant |(\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)'\boldsymbol{W}_{1t}| + |\widehat{U}_{1t} - U_{1t}| + |\widehat{\boldsymbol{\theta}}_1'\widehat{\boldsymbol{U}}_{-1t} - \boldsymbol{\theta}_1'\boldsymbol{U}_{-1t}|.
\end{aligned}$$

Using Hölder's inequality, the third term can be further bounded as

$$\begin{aligned}
|\widehat{\boldsymbol{\theta}}_1'\widehat{\boldsymbol{U}}_{-1t} - \boldsymbol{\theta}_1'\boldsymbol{U}_{-1t}| &\leqslant |\widehat{\boldsymbol{\theta}}_1'(\widehat{\boldsymbol{U}}_{-1t} - \boldsymbol{U}_{-1t})| + |(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)'\boldsymbol{U}_{-1t}| \\
&\leqslant \|\widehat{\boldsymbol{\theta}}_1\|_1\|\widehat{\boldsymbol{U}}_{-1t} - \boldsymbol{U}_{-1t}\|_\infty + \|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1\|\boldsymbol{U}_{-1t}\|_\infty \\
&\leqslant (\|\boldsymbol{\theta}_1\|_1 + \|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1)\|\widehat{\boldsymbol{U}}_{-1t} - \boldsymbol{U}_{-1t}\|_\infty + \|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1\|\boldsymbol{U}_{-1t}\|_\infty \\
&= O_P[(\|\boldsymbol{\theta}_1\|_1 + \upsilon|\mathcal{S}_0|\psi^{-1}(T))\upsilon + \upsilon|\mathcal{S}_0|\psi^{-1}(T)\psi^{-1}(n)].
\end{aligned}$$

Combining the last two expressions we are left with

$$|\widehat{\delta}_t - \delta_t - V_t| \leqslant |(\widehat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1)'\boldsymbol{W}_{1t}| + (1 + \|\boldsymbol{\theta}_1\|_1 + \|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1)\|\widehat{\boldsymbol{U}}_t - \boldsymbol{U}_t\|_\infty + \|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1\|\boldsymbol{U}_t\|_\infty.$$

The first term is $O_P(1/\sqrt{T})$ by Assumption 3(a). The second is $O_P(|\mathcal{S}_0|\eta)$ because by Assumption 3(d) we have that $\|\boldsymbol{\theta}_1\|_1 \leqslant |\mathcal{S}_0|\|\boldsymbol{\theta}_1\|_\infty \leqslant C|\mathcal{S}_0|$ and $\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\|_1 = O_P(1)$ by Assumption 3(f). Finally, the third term is $O_P(\xi|\mathcal{S}_0|n^{1/p})$ by Assumption 3(b) and the maximum inequality. Therefore we conclude that

$$\widehat{\delta}_t - \delta_t - V_t = O_P\left(T^{-1/2} + |\mathcal{S}_0|\eta + \xi|\mathcal{S}_0|n^{1/p}\right) = O_P\left[|\mathcal{S}_0|(\eta + \xi n^{1/p})\right].$$

Table 1: **Average Treatment (Δ) Estimation under the Null.**

The table reports descriptive statistics for the average treatment estimation under the null of no effect. The table reports the mean, median, and mean squared error (MSE) of the estimator $\widehat{\Delta}$ for five post-intervention observations. 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 et al. (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), i.e., 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 Abadie and L'Hour (2019). Finally, Panel (c) displays the results of the `FarmTreat` methodology.

Panel(a): LASSO (ArCo)

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	-0.049	-0.076	0.043	-0.061	-0.150	-0.006	1.988	1.552	1.459
100	-0.057	-0.044	0.057	-0.038	-0.051	0.058	0.862	0.646	0.655
500	-0.001	-0.027	-0.001	0.026	-0.037	-0.003	0.212	0.202	0.186

Panel(b): PCR

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.103	-0.136	0.203	0.214	-0.128	0.112	10.137	5.657	6.913
100	-0.065	-0.031	0.070	-0.026	-0.016	0.033	2.376	1.573	1.402
500	-0.012	-0.047	0.065	-0.023	-0.038	0.046	0.983	0.488	0.459

Panel(c): FarmTreat

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.037	-0.055	0.126	0.019	-0.053	0.083	6.114	2.007	2.785
100	-0.041	-0.019	0.051	-0.030	-0.031	0.035	1.152	0.366	0.340
500	0.018	-0.002	0.011	-0.015	-0.001	0.008	0.516	0.067	0.065

Table 2: **Average Treatment (Δ) Estimation under the Alternative.**

The table reports descriptive statistics for the average treatment estimation under the alternative of an average effect of 2. The table reports the mean, median, and mean squared error (MSE) of the estimator $\widehat{\Delta}$ for five post-intervention observations. 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 et al. (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), i.e., 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 Abadie and L'Hour (2019). Finally, Panel (c) displays the results of the `FarmTreat` methodology.

Panel(a): LASSO (ArCo)

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	1.951	1.924	2.043	1.939	1.850	1.994	1.988	1.552	1.459
100	1.943	1.956	2.057	1.962	1.949	2.058	0.862	0.646	0.655
500	1.999	1.973	1.999	2.026	1.963	1.997	0.212	0.202	0.186

Panel(b): PCR

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	2.103	1.864	2.203	2.214	1.872	2.112	10.137	5.657	6.913
100	1.935	1.969	2.070	1.974	1.984	2.033	2.376	1.573	1.402
500	1.988	1.953	2.065	1.977	1.962	2.046	0.983	0.488	0.459

Panel(c): FarmTreat

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	2.037	1.945	2.126	2.019	1.947	2.083	6.114	2.007	2.785
100	1.959	1.981	2.051	1.970	1.969	2.035	1.152	0.366	0.340
500	2.018	1.998	2.011	1.985	1.999	2.008	0.516	0.067	0.065

Table 3: **Rejection Rates under the Null (empirical size)**

The table reports the rejection rates of the partial ressampling test with five observation after the intervention.

Panel(a): LASSO (ArCo)

	$\alpha=0.01$			$\alpha=0.05$			$\alpha=0.10$		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.2670	0.2490	0.2810	0.3650	0.3560	0.3750	0.4230	0.4060	0.4340
100	0.0550	0.0630	0.0690	0.1510	0.1380	0.1450	0.2150	0.2040	0.2130
500	0.0140	0.0190	0.0190	0.0780	0.0790	0.0680	0.1350	0.1420	0.1330

Panel(b): PCR

	$\alpha=0.01$			$\alpha=0.05$			$\alpha=0.10$		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.3140	0.3000	0.3320	0.3940	0.3860	0.4240	0.4420	0.4460	0.4750
100	0.0220	0.0250	0.0130	0.1000	0.0790	0.0670	0.1690	0.1440	0.1180
500	0.0100	0.0150	0.0090	0.0640	0.0620	0.0560	0.1160	0.1130	0.1030

Panel(c): FarmTreat

	$\alpha=0.01$			$\alpha=0.05$			$\alpha=0.10$		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.3180	0.3450	0.3960	0.4000	0.4350	0.4740	0.4560	0.4820	0.5240
100	0.0170	0.0280	0.0170	0.0870	0.0880	0.0740	0.1580	0.1430	0.1230
500	0.0090	0.0150	0.0090	0.0670	0.0570	0.0550	0.1160	0.1170	0.1080

Table 4: **Average Treatment (Δ) Estimation under the Null and No Idiosyncratic Contribution.**

The table reports descriptive statistics for the average treatment estimation under the null of no effect and $\boldsymbol{\beta} = \mathbf{0}$. The table reports the mean, median, and mean squared error (MSE) of the estimator $\widehat{\Delta}$ for five post-intervention observations. 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 et al. (2018). Panel (b) presents the results when the counterfactual is estimated by principal component regression (PCR), i.e., 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 Abadie and L'Hour (2019). Finally, Panel (c) displays the results of the `FarmTreat` methodology.

Panel (a): LASSO (ArCo)

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	-0.108	-0.087	0.121	-0.133	-0.068	0.181	2.155	1.501	1.172
100	-0.037	-0.041	0.028	-0.028	-0.020	-0.009	0.994	0.796	0.808
500	-0.053	-0.002	-0.002	-0.042	-0.023	0.003	0.466	0.407	0.452

Panel (b): PCR

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.104	-0.122	0.204	0.103	-0.154	0.046	10.854	5.440	6.816
100	-0.017	-0.016	0.027	-0.055	-0.064	0.021	4.458	1.322	1.189
500	-0.067	-0.011	0.001	-0.037	0.016	0.023	0.987	0.238	0.248

Panel (c): FarmTreat

	Mean			Median			MSE		
	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$	$n=T$	$n=2\times T$	$n=3\times T$
$T=50$	0.020	-0.071	0.161	-0.018	-0.088	0.090	7.167	1.937	2.688
100	0.001	-0.017	0.033	-0.014	-0.010	0.049	3.680	0.491	0.469
500	-0.059	-0.006	0.007	-0.035	-0.003	0.025	0.951	0.209	0.222

Table 5: **Experiments.**

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

Product	**Sample**	**Experiment Period**	**Experiment Type**	**Control Group**	**Treatment Group**
I	Aug-14-2016 – May-02-2017	Apr-19-2017 – May-02-2017	Price reduction	328	110
II	May-14-2016 – Jan-23-2017	Jan-17-2017 – Jan-23-2017	Price reduction	321	100
III	Feb-13-2016 – Oct-31-2016	Oct-16-2016 – Oct-31-2016	Price increase	318	97
IV	May-14-2016 – Jan-23-2017	Jan-17-2017 – Jan-23-2017	Price increase	321	102
V	Feb-13-2016 – Oct-31-2016	Oct-16-2016 – Oct-31-2016	Price increase	309	106

Table 6: **Results.**

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 given statistic. 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 p-value results for the average intervention effect over the experiment period $\mathscr{H}_0 : \delta_t = 0$ for a given t. Panels (c) and (d) depict the results for the p-values of the re-sampling test for the null hypothesis $\mathscr{H}_0 : \delta_t = 0, \forall t \in \{T_0 + 1, \ldots, T\}$ using respectively the test statistics $\phi_2(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} \widehat{\delta}_t^2$ and $\phi_3(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} |\widehat{\delta}_t|$. Finally, Panel (e) reports the results for the p-values for the test for idiosyncratic contribution.

Panel (a): **R-squared**

Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	0.0337	0.0514	0.1040	0.1672	0.2705	0.4436	0.6642	0.2002	0.1282
II	0.4028	0.6745	0.8825	0.9323	0.9652	0.9894	0.9988	0.8981	0.1073
III	0.1134	0.1951	0.3610	0.4916	0.6215	0.7566	0.9065	0.4878	0.1764
IV	0.4669	0.7236	0.8744	0.9252	0.9551	0.9848	0.9961	0.8978	0.0916
V	0.1190	0.3092	0.5221	0.6969	0.8254	0.9281	0.9535	0.6691	0.1970

Panel (b): **Average Treatment Effect (over time): Δ**

Product	Min	5%-quantile	25%-quantile	Median	75%-quantile	95% quantile	Max	Average	Std. Dev
I	-1.2630	-0.9181	-0.4980	-0.1862	0.1420	0.6933	1.5493	-0.1672	0.4856
II	-3.0126	-1.8272	-0.4593	0.2748	1.3074	3.7670	6.6975	0.5515	1.6794
III	-19.1670	-16.8416	-7.8397	-3.4310	-1.2491	1.3600	3.5261	-5.1397	5.4411
IV	-45.4717	-28.3762	-14.6982	-7.4852	-3.4748	2.1461	36.6423	-9.4225	11.0010
V	-54.5934	-17.3325	-6.5691	-2.6661	-0.6040	0.8332	7.1110	-5.0361	8.0906

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	0.0638	0.3106	0.6298	0.9319	1.0000	1.0000	0.5970	0.3286
II	0	0	0.1219	0.3657	0.7045	0.9669	1.0000	0.4125	0.3284
III	0	0	0.0638	0.2298	0.5670	0.8438	0.9532	0.3203	0.2954
IV	0	0.0107	0.0826	0.3182	0.6157	0.9306	0.9959	0.3785	0.3068
V	0	0	0.0809	0.2702	0.5830	0.9200	0.9702	0.3525	0.2927

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	0.0596	0.2511	0.6489	0.9191	1.0000	1.0000	0.5967	0.3374
II	0	0	0.1012	0.4029	0.6612	0.9256	1.0000	0.4095	0.3280
III	0	0	0.0426	0.1447	0.5500	0.8787	0.9915	0.2968	0.3091
IV	0	0	0.0537	0.2645	0.6281	0.9091	0.9917	0.3565	0.3149
V	0	0	0.0426	0.2468	0.5957	0.9123	0.9745	0.3320	0.3103

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	0.0110	0.0180	0.2110	0.3445	0.5140	0.7750	0.8810	0.3616	0.2200
II	0.0240	0.0450	0.1030	0.1800	0.3075	0.4420	0.7340	0.2080	0.1375
III	0	0.0010	0.0187	0.0780	0.2240	0.6969	0.7770	0.1617	0.2000
IV	0.0060	0.0242	0.0600	0.1280	0.2600	0.4436	0.6690	0.1810	0.1482
V	0	0	0.0080	0.0705	0.1600	0.3252	0.5330	0.1064	0.1171

Figure 1: Product I data.

Panel (a) reports the sales per store aggregated in 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 medians over time and the vertical green line indicates the date of intervention.

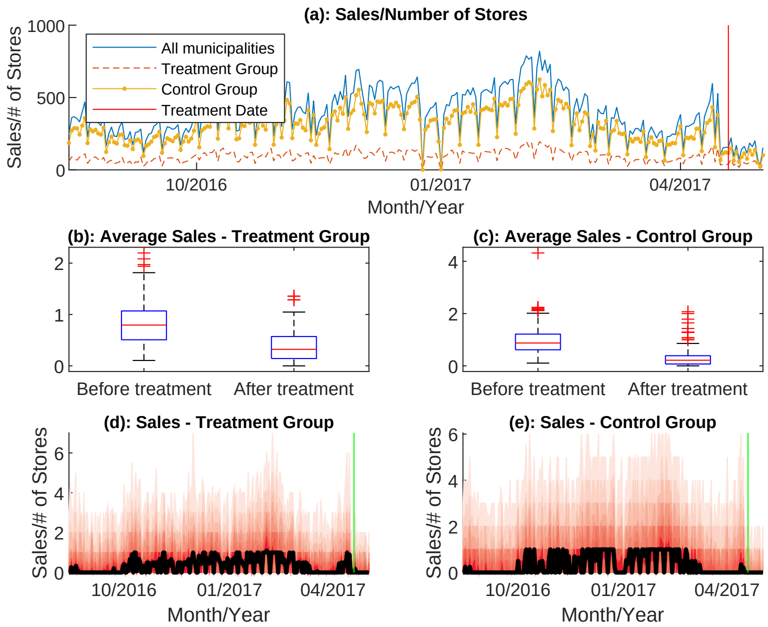


Figure 2: Product II data.

The same caption as Figure 1 is used.

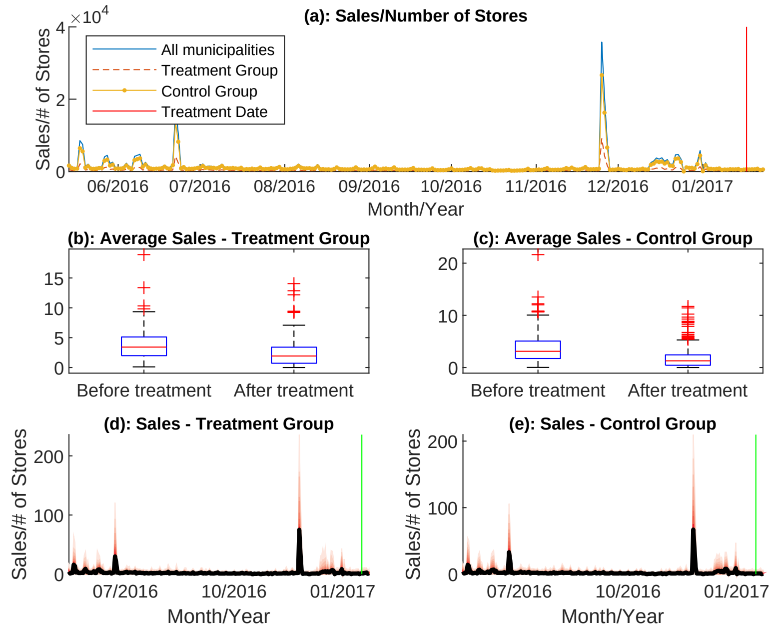


Figure 3: Product III data.

The same caption as Figure 1 is used.

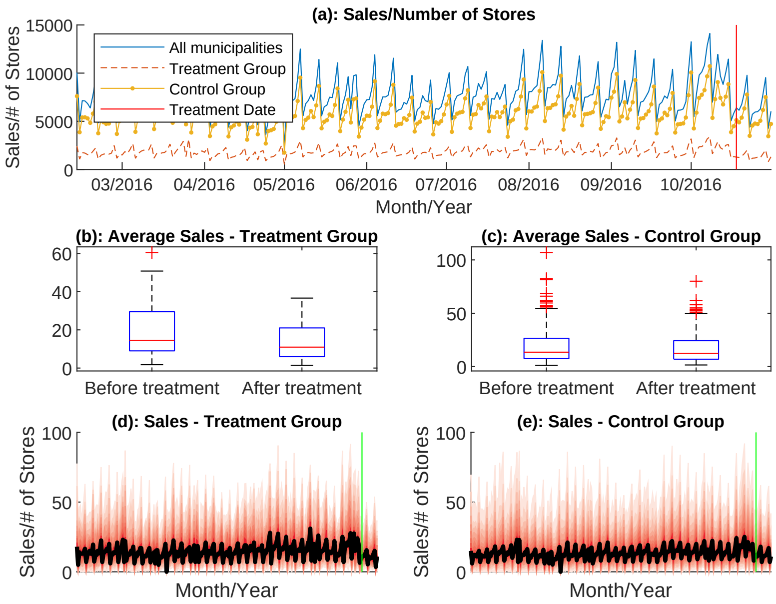


Figure 4: Product IV data.

The same caption as Figure 1 is used.

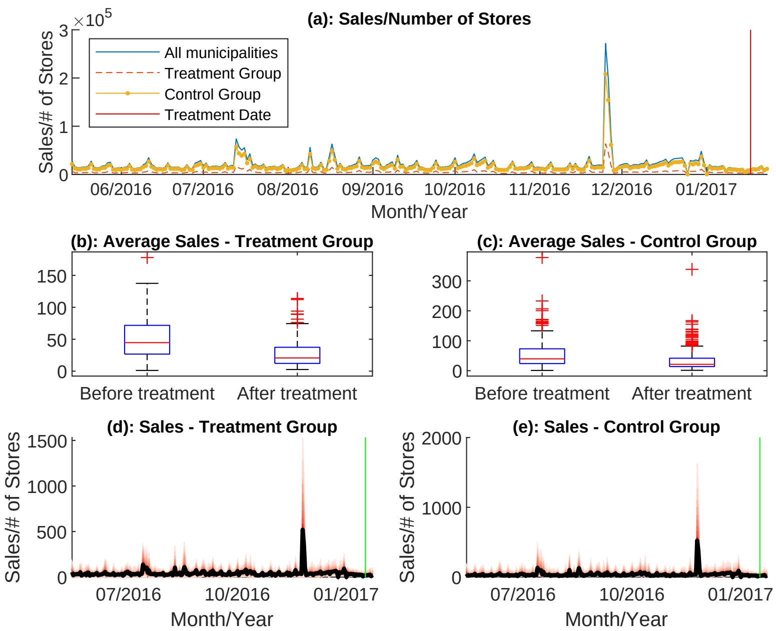


Figure 5: Product V data.

The same caption as Figure 1 is used.

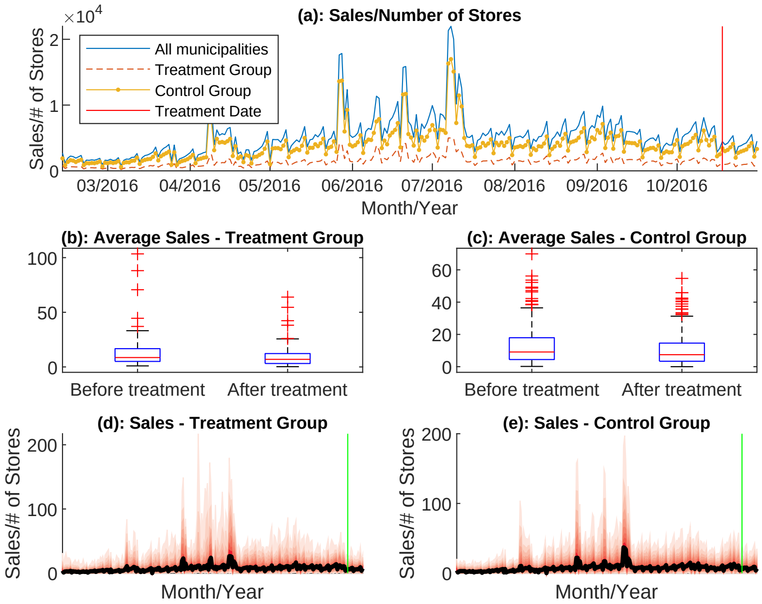


Figure 6: Results for Product I.

Panel (a) displays a fan plot, across n_1 municipalities in the treatment group, of the p-values of the re-sampling test for the null $\mathscr{H}_{0,1} : \delta_t = 0$ at each time t after the treatment. The black curve represents the median p-value across municipalities over t. Panels (b) and (c) display the distribution of the p-values of the re-sampling tests for the null hypothesis $\mathscr{H}_0 : \delta_t = 0, \forall t \in \{T_0 + 1, \ldots, T\}$ using respectively the test statistics $\phi_2(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} \widehat{\delta}_t^2$ and $\phi_3(\widehat{\delta}_{T_0+1}, \ldots, \widehat{\delta}_T) = \sum_{t=T_0+1}^{T} |\widehat{\delta}_t|$. Panel (e) shows an example for one municipality. The panel depicts the actual and counterfactual sales per store for the post-treatment period. 95% confidence intervals for the counterfactual path is also displayed.

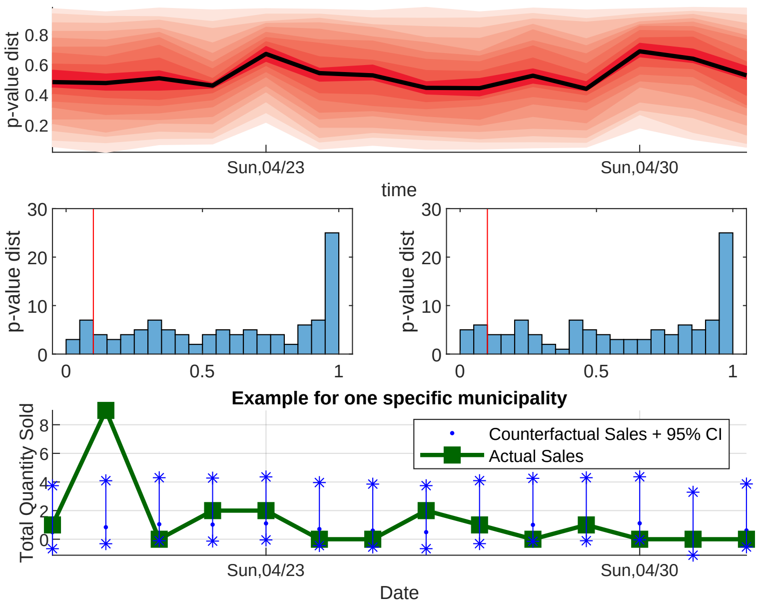


Figure 7: Results for Product II.

The same caption as in Figure 6 is used.

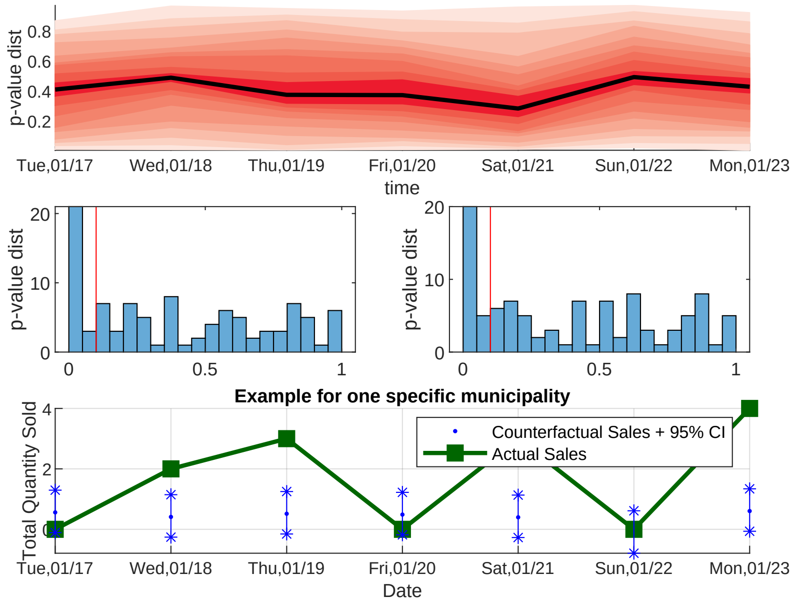


Figure 8: Results for Product III.

The same caption as in Figure 6 is used.

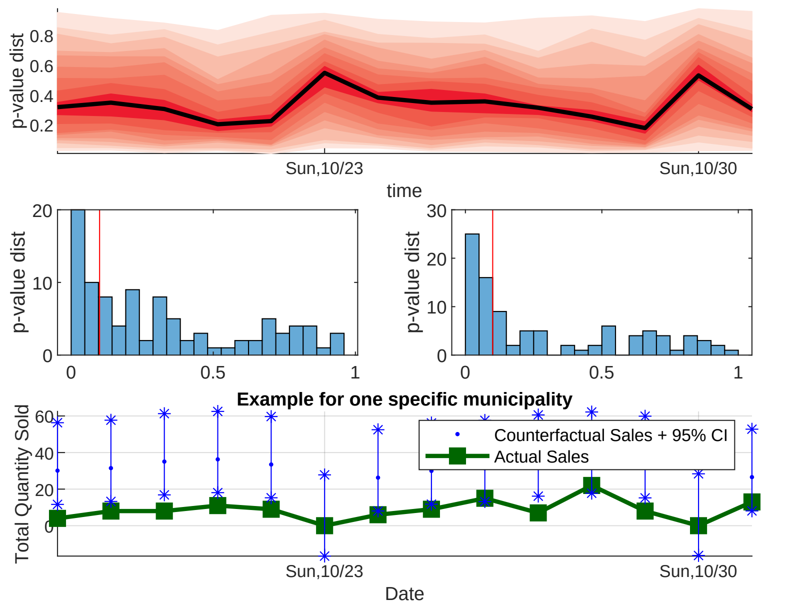


Figure 9: Results for Product IV.

The same caption as in Figure 6 is used.

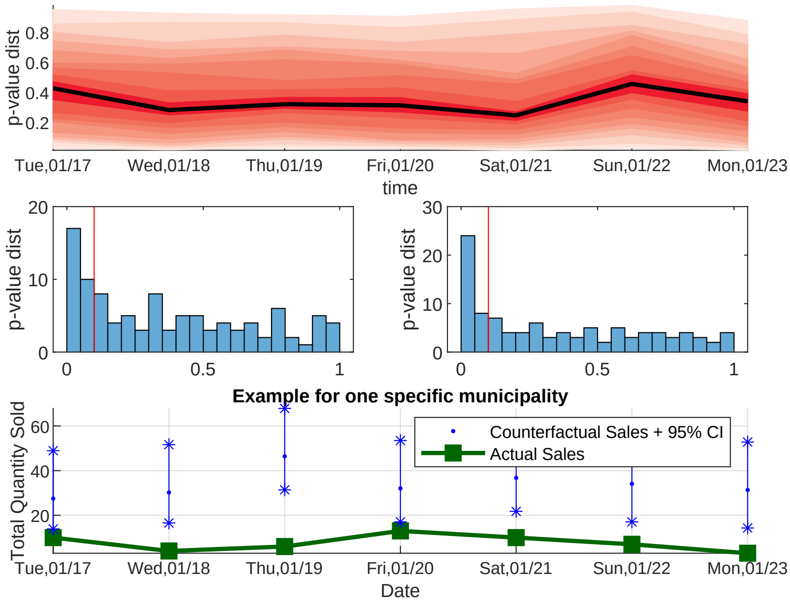
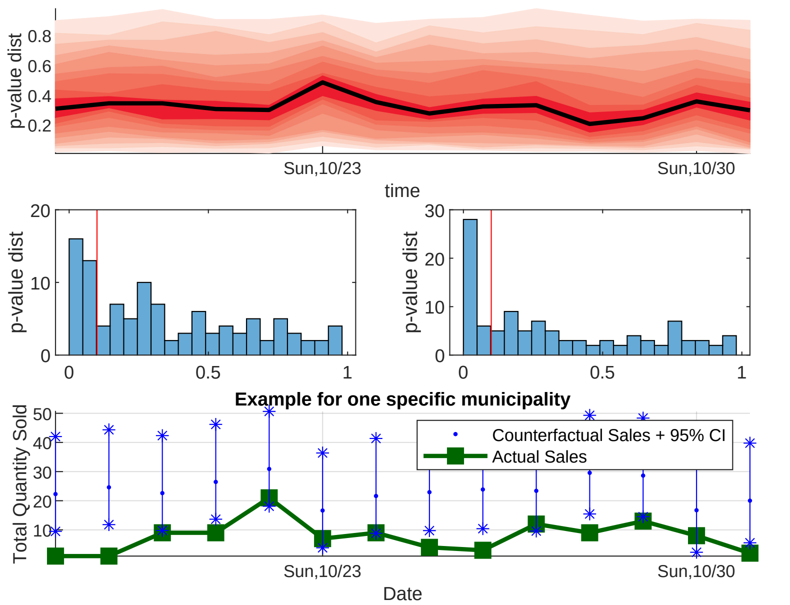


Figure 10: Results for Product V.

The same caption as in Figure 6 is used.



References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association 105*, 493–505.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review 93*, 113–132.

Abadie, A. and J. L'Hour (2019). A penalized synthetic control estimator for disaggregated data. Technical report, CREST.

Ahn, S. and A. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Athey, S. and G. Imbens (2017). The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives 31*, 3–32.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Carvalho, C., R. Masini, and M. Medeiros (2018). Arco: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of Econometrics 207*, 352–380.

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2018a). An exact and robust conformal inference method for counterfactual and synthetic controls. arxiv:1712.09089, arXiv.

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2018b). Exact and robust conformal inference methods for predictive machine learning with dependent data. arxiv:1802.06300, arXiv.

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2018c). Inference on average treatment effects in aggregate panel data settings. arxiv:1812.10820, arXiv.

Doudchenko, N. and G. Imbens (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. 22791, NBER. available at arXiv:1610.07748.

Fan, J., Y. Ke, and K. Wang (2020). Factor-adjusted regularized model selection. *Journal of Econometrics 216*, 71–85.

Fan, J., R. Li, C.-H. Zhang, and H. Zou (2020). *Statistical Foundations of Data Science*. CRC Press.

Fan, J., R. Masini, and M. Medeiros (2020). Do we exploit all information? factor models, idiosyncratic correction and a new test for high-dimensional covariance structure. Working paper, Princeton University.

Ferman, B. and C. Pinto (2016). Synthetic controls with imperfect pre-treatment fit. Working paper, São Paulo School of Economics - FGV.

Gobillon, L. and T. Magnac (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics 98*, 535–551.

Hsiao, C., H. S. Ching, and S. K. Wan (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.

Li, K. (2017). Statistical inference for average treatment effects estimated by synthetic control methods. Working paper, The Wharton School, the University of Pennsylvania.

Li, K. and D. Bell (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics 197*, 65–75.

Masini, R. and M. Medeiros (2019). Counterfactual analysis with artificial controls: Inference, high dimensions and nonstationarity. Working Paper 3303308, SSRN.

Masini, R. and M. Medeiros (2020). Counterfactual analysis and inference with non-stationary data. *Journal of Business and Economic Statistics*. in press.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological) 58*, 267–288.