Counterfactual Analysis with Artificial Controls: Inference, High-Dimensions and Nonstationarity

Marcelo C. Medeiros
PUC-Rio

with

Ricardo Masini
São Paulo School of Economics & ORFE - Princeton University

"Happy the man who has been able to learn the causes of things"

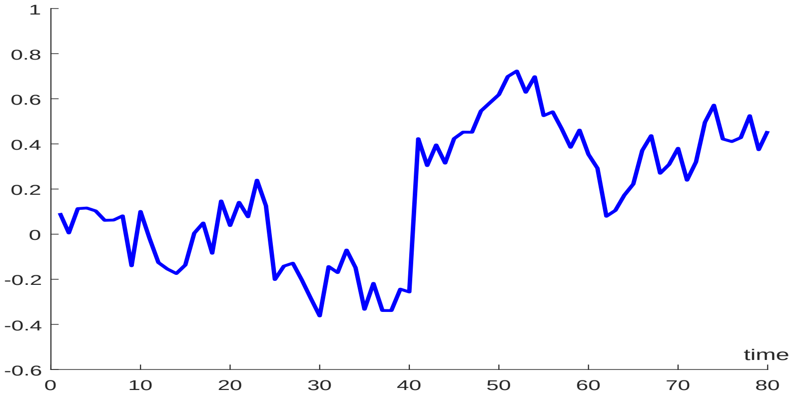
Virgil, Georgics (c. 29 BC), II, 490. in Hoyt's New Cyclopedia of Practical Quotations (1922), p. 91.

COLMEA
July 2021

Overview of the Methods and Motivation

Observe **aggregated time series data**, say y, from $t = 1$ to T.

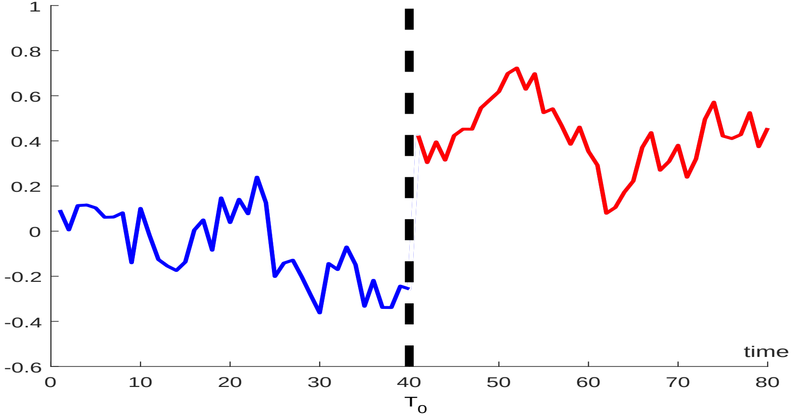
Examples: inflation or output growth of a country or state; returns of a firm.



Overview of the Methods and Motivation

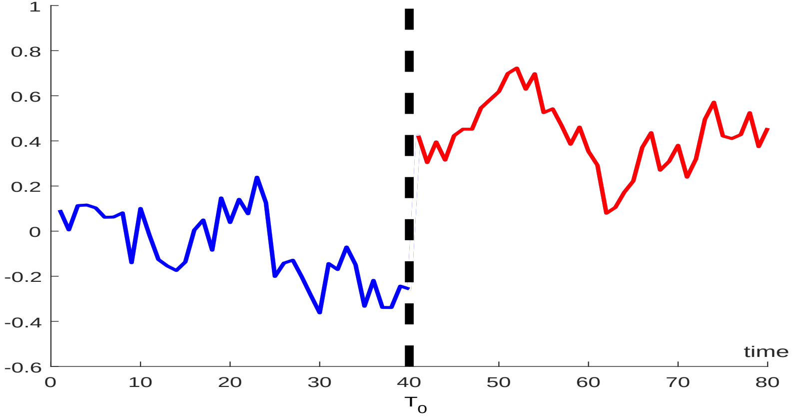
Intervention (treatment, event, ...) occurs at $t = T_0$.

Examples: a new policy/law, outbreak of a war, new government, etc.



Overview of the Methods and Motivation

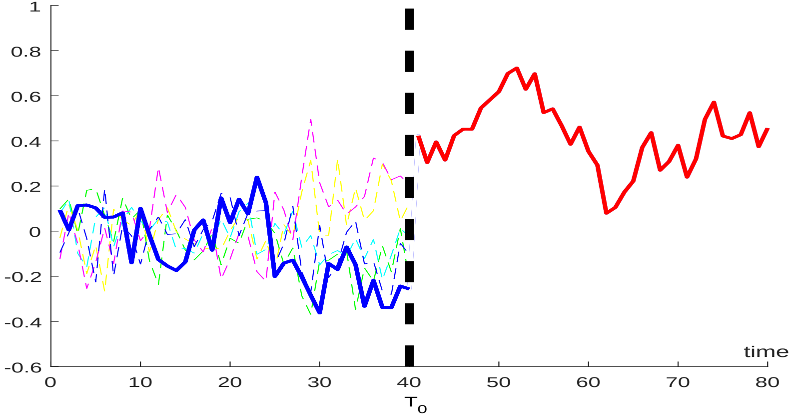
What are the **causal effects** of the intervention on y?



Overview of the Methods and Motivation

Large set of observed variables from **untreated "peers"**, $\boldsymbol{x}$.

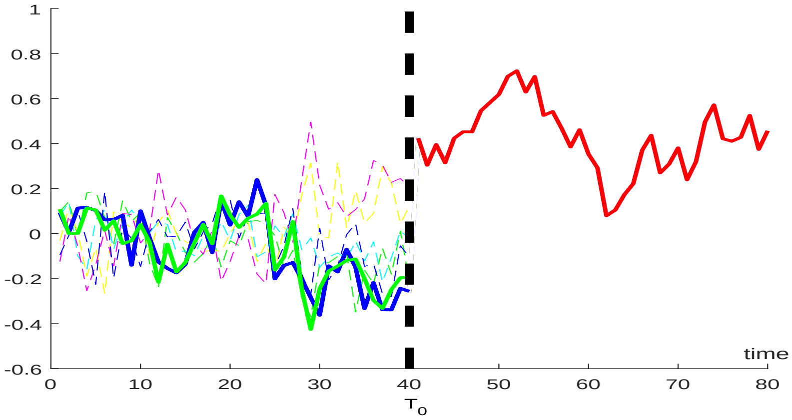
Frequently, the dimension of $\boldsymbol{x}$ is comparable or even larger than T_0.



Overview of the Methods and Motivation

Counterfactual estimation “in-sample” (before intervention).

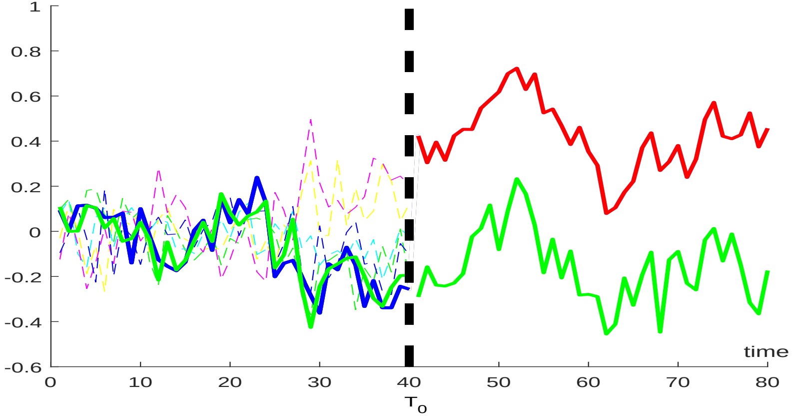
Different methods to estimate the model (synthetic control, ArCo, panel, ...).



Overview of the Methods and Motivation

Counterfactual extrapolation (after the intervention).

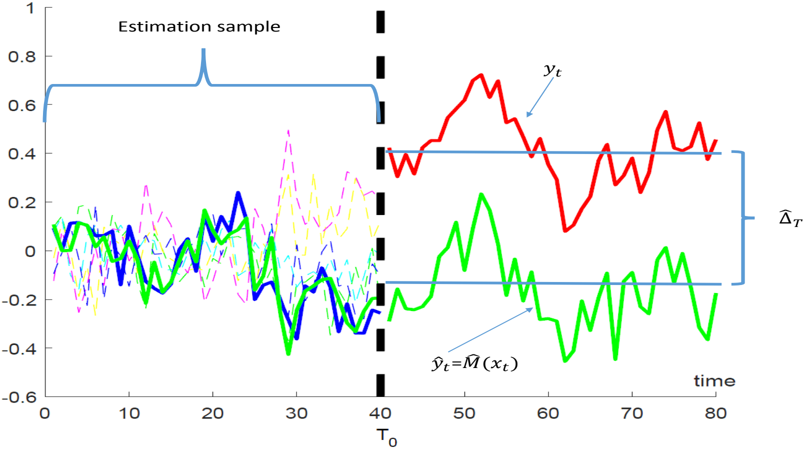
Extrapolation is done based on observables from the peers after T_0.



Overview of the Methods and Motivation

One possible estimator.

Average effect of the intervention.



Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0.

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.
  - Some sort of shrinkage (model restrictions or LASSO) or factor model is needed.

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.
  - Some sort of shrinkage (model restrictions or LASSO) or factor model is needed.
  - Example: In Abadie and Gardeazabal (2003,AER) - 16 parameters and 13 observations.

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.
  - Some sort of shrinkage (model restrictions or LASSO) or factor model is needed.
  - Example: In Abadie and Gardeazabal (2003,AER) - 16 parameters and 13 observations.

- The variables of interest may display trends, either deterministic or stochastic (unit-roots).

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.
  - Some sort of shrinkage (model restrictions or LASSO) or factor model is needed.
  - Example: In Abadie and Gardeazabal (2003,AER) - 16 parameters and 13 observations.

- The variables of interest may display trends, either deterministic or stochastic (unit-roots).
  - With few exceptions, unit-roots have been ignored in this framework.

- Inference on counterfactual dynamics.

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.
  - Some sort of shrinkage (model restrictions or LASSO) or factor model is needed.
  - Example: In Abadie and Gardeazabal (2003,AER) - 16 parameters and 13 observations.
- The variables of interest may display trends, either deterministic or stochastic (unit-roots).
  - With few exceptions, unit-roots have been ignored in this framework.
- Inference on counterfactual dynamics.
  - Either permutation tests or inference on average effects over the post-intervention period (based on $T_0, T \longrightarrow \infty$).

Overview of the Methods and Motivation

Challenges

- The dimension of $\boldsymbol{x}$ is large comparable to T_0. Either because there are a large numbers of covariates from the peers and/or because T_0 is small.
  - Some sort of shrinkage (model restrictions or LASSO) or factor model is needed.
  - Example: In Abadie and Gardeazabal (2003,AER) - 16 parameters and 13 observations.

- The variables of interest may display trends, either deterministic or stochastic (unit-roots).
  - With few exceptions, unit-roots have been ignored in this framework.

- Inference on counterfactual dynamics.
  - Either permutation tests or inference on average effects over the post-intervention period (based on $T_0, T \longrightarrow \infty$).

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)
  2. **Artificial Counterfactual (ArCo)** of Carvalho, Masini and Medeiros (2018, JoE)

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)
  2. **Artificial Counterfactual (ArCo)** of Carvalho, Masini and Medeiros (2018, JoE)
  3. **Panel Approach** of Hsiao, Ching and Wan (2012, JAE).

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)
  2. **Artificial Counterfactual (ArCo)** of Carvalho, Masini and Medeiros (2018, JoE)
  3. **Panel Approach** of Hsiao, Ching and Wan (2012, JAE).

- **Contributions:**

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)
  2. **Artificial Counterfactual (ArCo)** of Carvalho, Masini and Medeiros (2018, JoE)
  3. **Panel Approach** of Hsiao, Ching and Wan (2012, JAE).

- **Contributions:**
  1. Investigate the consequences of estimating counterfactuals when the data are potentially non-stationary, displaying either deterministic or stochastic trends.

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)
  2. **Artificial Counterfactual (ArCo)** of Carvalho, Masini and Medeiros (2018, JoE)
  3. **Panel Approach** of Hsiao, Ching and Wan (2012, JAE).

- **Contributions:**
  1. Investigate the consequences of estimating counterfactuals when the data are potentially non-stationary, displaying either deterministic or stochastic trends.
  2. Theoretical results are derived in a high-dimensional setting: Weighted LASSO which is proved to deliver consistent estimates of the parameters of interest.

One-page Summary

- General framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available, which nests:
  1. **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and Abadie, Diamond and Hainmueller (2010, JASA)
  2. **Artificial Counterfactual (ArCo)** of Carvalho, Masini and Medeiros (2018, JoE)
  3. **Panel Approach** of Hsiao, Ching and Wan (2012, JAE).

- **Contributions:**
  1. Investigate the consequences of estimating counterfactuals when the data are potentially non-stationary, displaying either deterministic or stochastic trends.
  2. Theoretical results are derived in a high-dimensional setting: Weighted LASSO which is proved to deliver consistent estimates of the parameters of interest.
  3. Inferential procedures based on resampling.

The road map

1. The setup
2. The counterfactual estimation
3. Estimator properties and inference
4. Monte Carlo simulation
5. Empirical example: *Demand Estimation and Optimal Pricing*

Setup

- Observe $i = 1, \ldots n$ units for $t = 1, \ldots T$ periods (Panel structure): z_{it}

Setup

- Observe $i = 1, \ldots n$ units for $t = 1, \ldots T$ periods (Panel structure): z_{it}
- **First unit** is treated at **known** T_0.

Setup

- Observe $i = 1, \ldots n$ units for $t = 1, \ldots T$ periods (Panel structure): z_{it}
- **First unit** is treated at **known** T_0.
- The remanning $n - 1$ units $\boldsymbol{z}_{0t} \equiv (z_{2t}, \ldots, z_{nt})'$ are an **untreated potential control group (*donor pool*)**.

Setup

- Observe $i = 1, \ldots n$ units for $t = 1, \ldots T$ periods (Panel structure): z_{it}
- **First unit** is treated at **known** T_0.
- The remanning $n-1$ units $\mathbf{z}_{0t} \equiv (z_{2t}, \ldots, z_{nt})'$ are an **untreated potential control group (*donor pool*)**.
- **Potential Outcome notation**:

$$z_{1t} = d_t z_{1t}^{(1)} + (1-d_t) z_{1t}^{(0)}; \qquad d_t = \begin{cases} 1 & \text{if } t \geq T_0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{z}_{0t} = \mathbf{z}_{0t}^{(0)}$$

where $z_{it}^{(1)}$ is potential outcome under the intervention and $z_{it}^{(0)}$ the potential outcome with no intervention.

Setup

- Simplify notation: $y_t = z_{1t}$.

Setup

- Simplify notation: $y_t = z_{1t}$.
- **Hypotheses of interest**: $y_t^{(1)} = \delta_t + y_t^{(0)}$, $t = T_0 \ldots, T$,

$$\mathcal{H}_0 : \Delta_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \underbrace{\left[y_t^{(1)} - y_t^{(0)} \right]}_{\equiv \delta_t} = 0 \quad \text{or}$$

$$\mathcal{H}_0 : \delta_t = 0, \forall\, t \geq T_0 \quad \text{or}$$

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

Setup

- Simplify notation: $y_t = z_{1t}$.
- **Hypotheses of interest**: $y_t^{(1)} = \delta_t + y_t^{(0)}$, $t = T_0 \ldots, T$,

$$\mathcal{H}_0 : \Delta_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \underbrace{\left[y_t^{(1)} - y_t^{(0)} \right]}_{\equiv \delta_t} = 0 \quad \text{or}$$

$$\mathcal{H}_0 : \delta_t = 0, \forall\, t \geq T_0 \quad \text{or}$$

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

- We **do not** observe the **counterfactual** $y_t^{(0)}$.
  Therefore, we construct an estimate $\widehat{y}_t^{(0)}$ such that:

$$\widehat{\delta}_t \equiv y_t^{(1)} - \widehat{y}_t^{(0)} \quad \text{for } t = T_0, \ldots, T$$

Counterfactual Estimation

- How should we construct $\widehat{y}_t^{(0)}$?

Counterfactual Estimation

- How should we construct $\widehat{y}_t^{(0)}$?

- We choose a (parametric) specification.

Counterfactual Estimation

- How should we construct $\widehat{y}_t^{(0)}$?

- We choose a (parametric) specification.

- Let $\boldsymbol{x}_t = (\boldsymbol{z}'_{0t}, \boldsymbol{z}'_{0t-1}, \ldots, \boldsymbol{z}'_{0t-p})'$ and

$$\boxed{y_t^{(0)} = \mathcal{M}(\boldsymbol{x}_t) + \nu_t,}$$

such that $\mathbb{E}(\nu_t) = 0$ and

$$\boxed{\widehat{y}_t^{(0)} = \widehat{\mathcal{M}}(\boldsymbol{x}_t).}$$

Counterfactual Estimation

- The average estimator is then simply given by

$$\widehat{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \widehat{\delta}_t,$$

where $\widehat{\delta}_t \equiv y_t - \widehat{y}_t^{(0)}$, for $t = T_0, \ldots, T$.

Counterfactual Estimation

- The average estimator is then simply given by

$$\widehat{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \widehat{\delta}_t,$$

where $\widehat{\delta}_t \equiv y_t - \widehat{y}_t^{(0)}$, for $t = T_0, \ldots, T$.

- The estimator is computed in **two-steps**:

  1. **First step**: estimation of $\mathcal{M}$ with the pre-intervention sample;

  2. **Second step**: extrapolate $\mathcal{M}$ with actual data for $\boldsymbol{x}_t$ and compute $\{\delta_t\}_{t \geq T_0}$ and $\widehat{\Delta}_T$.

A Brief Review of the Literature

- Hsiao, Ching and Wan (2012, JAE)
  - Two-step method where $\mathcal{M}(\boldsymbol{x}_t)$ is a linear and scalar function of a **small and stationary** set of variables from the peers.
  - Correct specification ($\mathcal{M}$ is the conditional expectation).
  - Selection of peers by information criteria.

A Brief Review of the Literature

- Hsiao, Ching and Wan (2012, JAE)
  - Two-step method where $\mathcal{M}(\boldsymbol{x}_t)$ is a linear and scalar function of a **small and stationary** set of variables from the peers.
  - Correct specification ($\mathcal{M}$ is the conditional expectation).
  - Selection of peers by information criteria.
- Differences-in-Differences (DiD)
  - Number of treated units must grow.
  - Parallel trends hypothesis.
  - Similar control group.

A Brief Review of the Literature

- Hsiao, Ching and Wan (2012, JAE)
  - Two-step method where $\mathcal{M}(\boldsymbol{x}_t)$ is a linear and scalar function of a **small and stationary** set of variables from the peers.
  - Correct specification ($\mathcal{M}$ is the conditional expectation).
  - Selection of peers by information criteria.
- Differences-in-Differences (DiD)
  - Number of treated units must grow.
  - Parallel trends hypothesis.
  - Similar control group.
- Gobillon and Magnac (2016, REStat)
  - Generalize the above authors by explicitly considering a factor model.
  - Interactive fixed effects with strictly exogenous regressors.
  - Asymptotics both on the cross-section and time dimensions.

A Brief Review of the Literature

- Abadie and Gardeazabal (2003, AER)
  - Convex combination of peers.
  - The weights are estimated using time averages of the observed variables. No time-series dynamics. Stationarity imposed.
  - Generalizations in Doudchenko and Imbens (2016) and Athey and Imbens (2017, JEP):

A Brief Review of the Literature

- Abadie and Gardeazabal (2003, AER)
  - Convex combination of peers.
  - The weights are estimated using time averages of the observed variables. No time-series dynamics. Stationarity imposed.
  - Generalizations in Doudchenko and Imbens (2016) and Athey and Imbens (2017, JEP):
- Bai, Li and Ouyang (2014, JoE):
  - Unit-roots in the framework of Hsiao, Ching and Wan (2012, JAE).
  - Only consistency results in a low-dimensional setting.

A Brief Review of the Literature

- Abadie and Gardeazabal (2003, AER)
  - Convex combination of peers.
  - The weights are estimated using time averages of the observed variables. No time-series dynamics. Stationarity imposed.
  - Generalizations in Doudchenko and Imbens (2016) and Athey and Imbens (2017, JEP):
- Bai, Li and Ouyang (2014, JoE):
  - Unit-roots in the framework of Hsiao, Ching and Wan (2012, JAE).
  - Only consistency results in a low-dimensional setting.
- Chernozhukov, Wuthrich and Zhu (2018a,b):
  - Stationarity imposed.
  - "Different" class of estimators: full-sample to estimate the model.

A Brief Review of the Literature

- Abadie and Gardeazabal (2003, AER)
  - Convex combination of peers.
  - The weights are estimated using time averages of the observed variables. No time-series dynamics. Stationarity imposed.
  - Generalizations in Doudchenko and Imbens (2016) and Athey and Imbens (2017, JEP):
- Bai, Li and Ouyang (2014, JoE):
  - Unit-roots in the framework of Hsiao, Ching and Wan (2012, JAE).
  - Only consistency results in a low-dimensional setting.
- Chernozhukov, Wuthrich and Zhu (2018a,b):
  - Stationarity imposed.
  - "Different" class of estimators: full-sample to estimate the model.
- Carvalho, Masini and Medeiros (2017) and Masini and Medeiros (2019):
  - Low-dimensional nonstationary set-up.

Counterfactual Estimation

Key Assumption

Independence

Let $\boldsymbol{z}_{0t} = (\boldsymbol{z}'_{2t}, \ldots, \boldsymbol{z}'_{nt})'$ denotes the vector of all the observable variables for the **untreated units**. Then, $\boldsymbol{z}_{0t} \perp\!\!\!\perp d_s$, for all t, s.

Counterfactual Estimation

Key Assumption

Independence

Let $\boldsymbol{z}_{0t} = (\boldsymbol{z}'_{2t}, \ldots, \boldsymbol{z}'_{nt})'$ denotes the vector of all the observable variables for the **untreated units**. Then, $\boldsymbol{z}_{0t} \perp\!\!\!\perp d_s$, for all t, s.

- The independence condition $\Rightarrow$ donors are untreated.

Counterfactual Estimation

Key Assumption

Independence

Let $\boldsymbol{z}_{0t} = (\boldsymbol{z}'_{2t}, \ldots, \boldsymbol{z}'_{nt})'$ denotes the vector of all the observable variables for the **untreated units**. Then, $\boldsymbol{z}_{0t} \perp\!\!\!\perp d_s$, for all t, s.

- The independence condition $\Rightarrow$ donors are untreated.
- Examples of interventions (treatments):
  - **Natural disasters**: Belasen and Polachek (2008, AER P&P), Cavallo, Galiani, Noy, and Pantano (2013, ReStat), Fujiki and Hsiao (2015, JoE), ...

Counterfactual Estimation

Key Assumption

Independence

Let $\boldsymbol{z}_{0t} = (\boldsymbol{z}'_{2t}, \ldots, \boldsymbol{z}'_{nt})'$ denotes the vector of all the observable variables for the **untreated units**. Then, $\boldsymbol{z}_{0t} \perp\!\!\!\perp d_s$, for all t, s.

- The independence condition $\Rightarrow$ donors are untreated.
- Examples of interventions (treatments):
  - **Natural disasters**: Belasen and Polachek (2008, AER P&P), Cavallo, Galiani, Noy, and Pantano (2013, ReStat), Fujiki and Hsiao (2015, JoE), ...
  - **Region specific policies (laws)**: Hsiao, Ching, and Wan (2012, JAE), Abadie, Diamond, and Hainmueller (AJPS, 2015), Gobillon and Magnac (ReStat, 2016), Carvalho, Masini and Medeiros (2018, JoE), ...

Counterfactual Estimation

Key Assumption

Independence

Let $\boldsymbol{z}_{0t} = (\boldsymbol{z}'_{2t}, \ldots, \boldsymbol{z}'_{nt})'$ denotes the vector of all the observable variables for the **untreated units**. Then, $\boldsymbol{z}_{0t} \perp\!\!\!\perp d_s$, for all t, s.

- The independence condition $\Rightarrow$ donors are untreated.
- Examples of interventions (treatments):
  - **Natural disasters**: Belasen and Polachek (2008, AER P&P), Cavallo, Galiani, Noy, and Pantano (2013, ReStat), Fujiki and Hsiao (2015, JoE), ...
  - **Region specific policies (laws)**: Hsiao, Ching, and Wan (2012, JAE), Abadie, Diamond, and Hainmueller (AJPS, 2015), Gobillon and Magnac (ReStat, 2016), Carvalho, Masini and Medeiros (2018, JoE), ...
  - **New government or political regime**: Grier and Maynard (2013, JEBO), Masini and Medeiros (2020, JBES), ...

Counterfactual estimation

Weighted Least Absolute Shrinkage and Selection Operator (wLASSO)

- Set $\boldsymbol{z}_{0t} \equiv (z_{2t}, \ldots, z_{nt})'$ and $\boldsymbol{x}_t = (\boldsymbol{z}_{0t}, \ldots, \boldsymbol{z}_{0t-p})'$.

Counterfactual estimation

Weighted Least Absolute Shrinkage and Selection Operator (wLASSO)

- Set $\boldsymbol{z}_{0t} \equiv (z_{2t}, \ldots, z_{nt})'$ and $\boldsymbol{x}_t = (\boldsymbol{z}_{0t}, \ldots, \boldsymbol{z}_{0t-p})'$.
- Set $\boldsymbol{X}_t = (1, \boldsymbol{x}_t)' \in \mathbb{R}^d$:

$$\boxed{\begin{aligned} y_t^{(0)} &= \boldsymbol{\alpha} + \boldsymbol{\beta}'\boldsymbol{x}_t + \nu_t \\ &= \boldsymbol{\theta}'\boldsymbol{X}_t + \nu_t. \end{aligned}}$$

Counterfactual estimation

Weighted Least Absolute Shrinkage and Selection Operator (wLASSO)

- Set $\boldsymbol{z}_{0t} \equiv (z_{2t}, \ldots, z_{nt})'$ and $\boldsymbol{x}_t = (\boldsymbol{z}_{0t}, \ldots, \boldsymbol{z}_{0t-p})'$.
- Set $\boldsymbol{X}_t = (1, \boldsymbol{x}_t)' \in \mathbb{R}^d$:

$$\boxed{\begin{aligned} y_t^{(0)} &= \boldsymbol{\alpha} + \boldsymbol{\beta}'\boldsymbol{x}_t + \nu_t \\ &= \boldsymbol{\theta}'\boldsymbol{X}_t + \nu_t. \end{aligned}}$$

- Estimation:

$$\widehat{\boldsymbol{\theta}} = \arg\min \left[\sum_{t=1}^{T_0-1} \left(y_t^{(0)} - \boldsymbol{\theta}'\boldsymbol{X}_t \right)^2 + \varsigma \sum_{k=1}^{d} \omega_k |\beta_k| \right].$$

- ω_k can be either $|x_{kT_0-1}|$, 1 or $\sqrt{T_0 - 1}$. **The choice will depend on the DGP**.

Where does the model come from?

A factor model example

- Suppose that:

$$z_{it}^{(0)} = c_i + \mu_i f_t + u_{it}^z,$$

where $c_i \in \mathbb{R}$, u_{it}^z is an idiosyncratic shock and $\mu_i \in \mathbb{R}$ is the factor loadings for unit i.

Where does the model come from?

A factor model example

- Suppose that:

$$z_{it}^{(0)} = c_i + \mu_i f_t + u_{it}^z,$$

where $c_i \in \mathbb{R}$, u_{it}^z is an idiosyncratic shock and $\mu_i \in \mathbb{R}$ is the factor loadings for unit i.

- The factor follows either a **unit root process with drift**

$$f_t = \mu_t^f + f_{t-1} + u_t^f, \quad t \geq 1$$

for some initial condition $f_0 = O_P(1)$;

Where does the model come from?

A factor model example

- Suppose that:

$$z_{it}^{(0)} = c_i + \mu_i f_t + u_{it}^z,$$

  where $c_i \in \mathbb{R}$, u_{it}^z is an idiosyncratic shock and $\mu_i \in \mathbb{R}$ is the factor loadings for unit i.
- The factor follows either a **unit root process with drift**

$$f_t = \mu_t^f + f_{t-1} + u_t^f, \quad t \geq 1$$

  for some initial condition $f_0 = O_P(1)$; or a **trend-stationary process**

$$f_t = \mu_t^f + u_t^f,$$

  where in both cases $\{\mu_t^f\}_{t=1}^{\infty}$ is a deterministic sequence.

Where does the model come from?

A factor model example

- Suppose that:

$$z_{it}^{(0)} = c_i + \mu_i f_t + u_{it}^z,$$

  where $c_i \in \mathbb{R}$, u_{it}^z is an idiosyncratic shock and $\mu_i \in \mathbb{R}$ is the factor loadings for unit i.
- The factor follows either a **unit root process with drift**

$$f_t = \mu_t^f + f_{t-1} + u_t^f, \quad t \geq 1$$

  for some initial condition $f_0 = O_P(1)$; or a **trend-stationary process**

$$f_t = \mu_t^f + u_t^f,$$

  where in both cases $\{\mu_t^f\}_{t=1}^{\infty}$ is a deterministic sequence.
- $(u_{1t}^z, \ldots, u_{nt}^z, u_t^f)$ is a zero-mean, independent and identically distributed Gaussian random vector.

Where does the model come from?

A factor model example

- **Common trend** (at least for those units with non-zero loadings, $\mu_i \neq 0$) and a correlation among the stochastic components of the vector $\boldsymbol{z}_t^{(0)}$ due to the presence of u_t^f.

Where does the model come from?

A factor model example

- **Common trend** (at least for those units with non-zero loadings, $\mu_i \neq 0$) and a correlation among the stochastic components of the vector $\boldsymbol{z}_t^{(0)}$ due to the presence of u_t^f.

- The pseudo-true model:

$$y_t = \boldsymbol{\theta}_0' \boldsymbol{X}_t + \nu_t,$$

where $y_t := z_{1t}^{(0)}$ and $\boldsymbol{X}_t := \left[1, \boldsymbol{z}_{0t}^{(0)\prime}\right]'$.

Where does the model come from?

A factor model example

- **Common trend** (at least for those units with non-zero loadings, $\mu_i \neq 0$) and a correlation among the stochastic components of the vector $\boldsymbol{z}_t^{(0)}$ due to the presence of u_t^f.

- The pseudo-true model:

$$y_t = \boldsymbol{\theta}_0' \boldsymbol{X}_t + \nu_t,$$

  where $y_t := z_{1t}^{(0)}$ and $\boldsymbol{X}_t := \left[1, \boldsymbol{z}_{0t}^{(0)\prime}\right]'$.

- Suppose there are $1 < r + 1 \leq n$ units with non-zero loadings $(\mu_i \neq 0)$ including unit 1.

Where does the model come from?

A factor model example

- **Common trend** (at least for those units with non-zero loadings, $\mu_i \neq 0$) and a correlation among the stochastic components of the vector $\boldsymbol{z}_t^{(0)}$ due to the presence of u_t^f.

- The pseudo-true model:

$$y_t = \boldsymbol{\theta}_0' \boldsymbol{X}_t + \nu_t,$$

  where $y_t := z_{1t}^{(0)}$ and $\boldsymbol{X}_t := \left[1, \boldsymbol{z}_{0t}^{(0)\prime}\right]'$.

- Suppose there are $1 < r + 1 \leq n$ units with non-zero loadings $(\mu_i \neq 0)$ including unit 1.

- Without loss of generality, make those the first $r + 1$ units.

Where does the model come from?

A factor model example

- r independent linear relations yielding stationary processes: setting $\widetilde{\mathbf{\Gamma}}'\boldsymbol{z}_t^{(0)}$, where

$$\boxed{\widetilde{\mathbf{\Gamma}}' = \begin{pmatrix} 1 & -\frac{\mu_1}{\mu_2} & 0 & 0 & \\ \vdots & 0 & \ddots & 0 & \mathbf{0}_{r\times(n-r-1)} \\ 1 & 0 & 0 & -\frac{\mu_1}{\mu_{r+1}} & \end{pmatrix}},$$

and $\mathbf{0}_{r\times(n-r-1)}$ is a $r \times (n-r-1)$ matrix of zero elements.

Where does the model come from?

A factor model example

- After normalizing to obtain the representation $\widetilde{\mathbf{\Gamma}}' = (I_r : -\mathbf{\Gamma}')$, we are left with:

$$\mathbf{\Gamma}' = \begin{pmatrix} \widetilde{\mu}_1 & & \\ \vdots & \mathbf{0}_{r\times(n-r-1)} & \\ \widetilde{\mu}_r & & \end{pmatrix},$$

where $\widetilde{\mu}_i := \frac{\mu_i}{\mu_{r+1}}$ for $i \in \{1, \ldots, r\}$.

Where does the model come from?

A factor model example

- After normalizing to obtain the representation $\widetilde{\boldsymbol{\Gamma}}' = (I_r : -\boldsymbol{\Gamma}')$, we are left with:

$$\boldsymbol{\Gamma}' = \begin{pmatrix} \widetilde{\mu}_1 & & \\ \vdots & \mathbf{0}_{r\times(n-r-1)} & \\ \widetilde{\mu}_r & & \end{pmatrix},$$

  where $\widetilde{\mu}_i := \frac{\mu_i}{\mu_{r+1}}$ for $i \in \{1, \dots, r\}$.

- Then, $\boldsymbol{J}_t = \widetilde{\boldsymbol{\Gamma}}' \boldsymbol{z}_t^{(0)}$ is stationary with a typical element given by

$$J_{i,t} = c_i - \widetilde{\mu}_i c_{r+1} + u_{it}^z - \widetilde{\mu}_i u_{r+1,t}^z = \widetilde{c}_i + \widetilde{u}_{it},$$

  where $\widetilde{c}_i := c_i - \widetilde{\mu}_i c_{r+1}$ and $\widetilde{u}_{it} := u_{it}^z - \widetilde{\mu}_i u_{r+1,t}^z$.

Where does the model come from?

A factor model example

- When $r = 1$:

$$\boldsymbol{\theta}_0 = \left(c_1 - \frac{\mu_1}{\mu_2} c_2, \frac{\mu_1}{\mu_2}, 0, \ldots, 0 \right)',$$

Where does the model come from?

A factor model example

- When $r = 1$:

$$\boldsymbol{\theta}_0 = \left(c_1 - \frac{\mu_1}{\mu_2} c_2, \frac{\mu_1}{\mu_2}, 0, \ldots, 0 \right)',$$

- The covariance of $\left(u_t^f, u_{1t}^z, \ldots, u_{nt}^z \right)'$ plays no role in determining the coefficients of the model.

Where does the model come from?

A factor model example

- When $r = 1$:

$$\boldsymbol{\theta}_0 = \left(c_1 - \frac{\mu_1}{\mu_2} c_2, \frac{\mu_1}{\mu_2}, 0, \ldots, 0 \right)',$$

- The covariance of $\left(u_t^f, u_{1t}^z, \ldots, u_{nt}^z \right)'$ plays no role in determining the coefficients of the model.
- When $r \geq 2$:

$$\boldsymbol{\theta}_0 = \left(\widetilde{c}_1 - \boldsymbol{\zeta}' \widetilde{\boldsymbol{c}}_0, \boldsymbol{\zeta}', \widetilde{\mu}_1 - \boldsymbol{\zeta}' \widetilde{\boldsymbol{\mu}}_0, 0, \ldots, 0 \right)',$$

where $\widetilde{\boldsymbol{c}}_0 := (\widetilde{c}_2, \ldots, \widetilde{c}_r)'$, $\widetilde{\boldsymbol{\mu}}_0 := (\widetilde{\mu}_2, \ldots, \widetilde{\mu}_r)'$ and $\boldsymbol{\zeta}$ denote the linear projection of $\widetilde{u}_{1t}$ onto $(\widetilde{u}_{2t}, \ldots, \widetilde{u}_{rt})'$.

Where does the model come from?

A factor model example

- When $r = 1$:

$$\boldsymbol{\theta}_0 = \left(c_1 - \frac{\mu_1}{\mu_2} c_2, \frac{\mu_1}{\mu_2}, 0, \ldots, 0\right)',$$

- The covariance of $\left(u_t^f, u_{1t}^z, \ldots, u_{nt}^z\right)'$ plays no role in determining the coefficients of the model.
- When $r \geq 2$:

$$\boldsymbol{\theta}_0 = \left(\widetilde{c}_1 - \boldsymbol{\zeta}'\widetilde{\boldsymbol{c}}_0, \boldsymbol{\zeta}', \widetilde{\mu}_1 - \boldsymbol{\zeta}'\widetilde{\boldsymbol{\mu}}_0, 0, \ldots, 0\right)',$$

  where $\widetilde{\boldsymbol{c}}_0 := (\widetilde{c}_2, \ldots, \widetilde{c}_r)'$, $\widetilde{\boldsymbol{\mu}}_0 := (\widetilde{\mu}_2, \ldots, \widetilde{\mu}_r)'$ and $\boldsymbol{\zeta}$ denote the linear projection of $\widetilde{u}_{1t}$ onto $(\widetilde{u}_{2t}, \ldots, \widetilde{u}_{rt})'$.
- The covariance of $\left(u_{1t}^z, \ldots, u_{r+1,t}^z\right)'$ affects the coefficients of the model through ζ.

Where does the model come from?

A factor model example

- When $r = 1$:

$$\boldsymbol{\theta}_0 = \left(c_1 - \frac{\mu_1}{\mu_2} c_2, \frac{\mu_1}{\mu_2}, 0, \ldots, 0\right)',$$

- The covariance of $\left(u_t^f, u_{1t}^z, \ldots, u_{nt}^z\right)'$ plays no role in determining the coefficients of the model.
- When $r \geq 2$:

$$\boldsymbol{\theta}_0 = \left(\widetilde{c}_1 - \boldsymbol{\zeta}'\widetilde{\boldsymbol{c}}_0, \boldsymbol{\zeta}', \widetilde{\mu}_1 - \boldsymbol{\zeta}'\widetilde{\boldsymbol{\mu}}_0, 0, \ldots, 0\right)',$$

where $\widetilde{\boldsymbol{c}}_0 := (\widetilde{c}_2, \ldots, \widetilde{c}_r)'$, $\widetilde{\boldsymbol{\mu}}_0 := (\widetilde{\mu}_2, \ldots, \widetilde{\mu}_r)'$ and $\boldsymbol{\zeta}$ denote the linear projection of $\widetilde{u}_{1t}$ onto $(\widetilde{u}_{2t}, \ldots, \widetilde{u}_{rt})'$.
- The covariance of $\left(u_{1t}^z, \ldots, u_{r+1,t}^z\right)'$ affects the coefficients of the model through ζ.
- Finally:

$$\nu_t = u_{1t}^z - \sum_{i=2}^{r+1} \theta_{0,i} u_{it}^z.$$

Data Generating Process (DGP)

Data Generating Process (1/3)

$\{z_{it}^{(0)} : 1 \leq i \leq n, t \geq 1\}$ is either generated by:

1. **Stochastic Trend**, where for a given initial condition $z_{i0}^{(0)} = O_P(1)$, let

$$z_{it}^{(0)} = z_{it-1}^{(0)} + f_{it} + u_{it}, \quad t \geq 1. \tag{1}$$

2. or by a **Deterministic Trend**, where

$$z_{it}^{(0)} = c_i + f_{it} + u_{it}, \quad t \geq 1, \tag{2}$$

In both cases, $\{f_{it}\}_{t \geq 1}$ denotes a deterministic sequence and $\{u_t := (u_{1t}, \ldots, u_{nt})'\}_{t \geq 1} \in \mathcal{U} \subset \mathbb{R}^n$ is a zero mean weakly dependent process.

Data Generating Process

Data Generating Process (2/3)

$\{\boldsymbol{u}_t\}_t$ is a zero mean strong mixing sequence of d_u-dimensional random vectors with mixing coefficients given by $\alpha(m) = \exp(-2cm)$ for some $c > 0$, fulfilling one of the conditions below:

1. For $\xi > 2$, $\sup\{\mathbb{E}|u_{it}|^{\xi+\epsilon} : 1 \leq i \leq d_u, t \in \mathbb{N}\} < \infty$ for some $\epsilon > 0$; or

2. there exist positive constants c_1, c_2, c_3 such that $\sup\{\mathbb{P}(|u_{it}| > x) : 1 \leq i \leq d_u, t \in \mathbb{N}\} \leq c_1 \exp(-c_2 x^{c_3})$ for all $x > 0$,

and, in both cases, the smallest eigenvalue of the matrix $\mathbb{E}(\boldsymbol{u}_t \boldsymbol{u}_t')$ is bounded away from $\mathbf{0}$ uniformly in $t \in \mathbb{N}$.

Data Generating Process

Data Generating Process (3/3)

In the case of a DGP with stochastic trends, there are r independent linear combinations (with non zero coefficient for the treated units) among the units that result in an $I(0)$ process:

$$\boldsymbol{\Gamma}\boldsymbol{y}_t^{(0)} \sim I(0),$$

where $0 < r < n$.

Data Generating Process

Data Generating Process (3/3)

In the case of a DGP with stochastic trends, there are r independent linear combinations (with non zero coefficient for the treated units) among the units that result in an $I(0)$ process:

$$\boldsymbol{\Gamma} \boldsymbol{y}_t^{(0)} \sim I(0),$$

where $0 < r < n$.

- In case of cointegration, the target “true” $\boldsymbol{\theta}$ in the estimation will be a function of the cointegration relation.

Nonstationary case: Main Result

Asymptotic Results

Set $T_1 = T_0 - 1$. Under some regularity assumptions (in the paper) and the additional assumption that for any $c > 0$, the penalty parameter ς is given by either $\varsigma = 4cd^{2/\xi}/\sqrt{T_1}$ or $\varsigma = 4(c + 2\log d)/\sqrt{T_1}$ (depending on the DGP), then:

1. $\|\widehat{\gamma} - \gamma_0\|_1 = O_P[\psi(d)s_0/\sqrt{T_1}] = o_P(1)$, as $T_1 \to \infty$
2. $\widehat{\boldsymbol{\delta}}_t - \boldsymbol{\delta}_t - \boldsymbol{\nu}_t = O_P[\psi(d)^2 s_0/\sqrt{T_1}] = o_P(1)$ for all $T_1 < t \leq T$, as $T_1 \to \infty$
3. $\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T = O_P[\psi(d)s_0/\sqrt{T}] = o_P(1)$, as $T \to \infty$

where $\psi(d) = d^{2/\xi}$ or $\psi(d) = \exp(d)$ (depending on the DGP).

- Not all elements of $\boldsymbol{x}_t$ are of the same order. Therefore, $\gamma = \boldsymbol{A}\boldsymbol{\theta}$, $\boldsymbol{A} = \mathsf{diag}(a_1, \ldots, a_d)$, $a_i = d_{i,T_1}$ or $a_i = \sqrt{T_1}$ (depending on the DGP). d_{it} is the deterministic component of the DGP.

Inference

- Inference procedure based on the sequence of estimators $\{\widehat{\delta}_t\}_{t>T_0}$.

Inference

- Inference procedure based on the sequence of estimators $\{\widehat{\delta}_t\}_{t>T_0}$.

- Continuous mappings $\phi : \mathbb{R}^{T_2} \to \mathbb{R}^b$ whose argument is the T_2-dimensional vector $(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)'$.

Inference

- Inference procedure based on the sequence of estimators $\{\widehat{\delta}_t\}_{t>T_0}$.
- Continuous mappings $\phi : \mathbb{R}^{T_2} \to \mathbb{R}^b$ 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 distribution of $\widehat{\phi} := \phi(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)$ under the null where $\delta_t = 0$ for all $t > T_0$.

Inference

- Inference procedure based on the sequence of estimators $\{\widehat{\delta}_t\}_{t>T_0}$.
- Continuous mappings $\phi : \mathbb{R}^{T_2} \to \mathbb{R}^b$ 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 distribution of $\widehat{\phi} := \phi(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)$ under the null where $\delta_t = 0$ for all $t > T_0$.
- As a direct corollary we have under the asymptotic on the pre-invention period $(T_0 \to \infty)$ that

$$\widehat{\phi} \xrightarrow{p} \phi_0 := \phi(\nu_{T_0+1}, \ldots, \nu_T).$$

Inference

- Consider the construction of $\widehat{\phi}$ using only blocks of size T_2 of consecutive observations from the pre-intervention sample.

Inference

- Consider the construction of $\widehat{\phi}$ 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

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

where $\widehat{\nu}_t := y_t - \widehat{\boldsymbol{\theta}}'_{T_0} \boldsymbol{X}_t$ with the subscript T_0 in $\widehat{\boldsymbol{\theta}}$ is to indicate that the estimator is calculated using the entire pre-intervention sample.

Inference

- Consider the construction of $\widehat{\phi}$ 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

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

  where $\widehat{\nu}_t := y_t - \widehat{\boldsymbol{\theta}}'_{T_0} \boldsymbol{X}_t$ with the subscript T_0 in $\widehat{\boldsymbol{\theta}}$ is to indicate that the estimator is calculated using the entire pre-intervention sample.
- For fixed j, we have that $\widehat{\phi}_j \xrightarrow{p} \phi_j := \phi(\nu_j, \ldots, \nu_{j+T_2-1})$.

Inference

- Consider the construction of $\widehat{\phi}$ 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

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

  where $\widehat{\nu}_t := y_t - \widehat{\boldsymbol{\theta}}'_{T_0} \boldsymbol{X}_t$ with the subscript T_0 in $\widehat{\boldsymbol{\theta}}$ is to indicate that the estimator is calculated using the entire pre-intervention sample.
- For fixed j, we have that $\widehat{\phi}_j \xrightarrow{p} \phi_j := \phi(\nu_j, \ldots, \nu_{j+T_2-1})$.
- Under a strictly stationarity assumption on ν_t we have that ϕ_j is equal in distribution to ϕ_0 for all j.

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

Inference

Resampling

For any continuous $\phi : \mathbb{R}^{T_2} \to \mathbb{R}^b$, let $\widehat{\phi} := \phi(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \ldots, \widehat{\delta}_T - \delta_T)$ and $\phi_0 := \phi(\nu_{T_0+1}, \ldots, \nu_T)$. Under regularity conditions (in the paper) and $s_0 = o\left\{\sqrt{T_0}/\left[\psi(p)\psi(pT_0)\right]\right\}$ then, as $T_0 \to \infty$:

1. $\widehat{\phi} \xrightarrow{p} \phi_0$
2. $\widehat{\mathcal{Q}}_T(x) - \mathcal{Q}_T(x) \xrightarrow{p} 0$ for all $x \in \mathcal{C}_0 :=$ $\{\text{continuity point of } \mathcal{Q}_0(x) := \mathbb{P}(\phi_0 \leq x)\}$
3. If $\mathcal{Q}_0(x)$ is continuous, the result (b) holds uniformly in $x \in \mathbb{R}^b$.
4. If ϕ is real-valued then $\mathcal{Q}_T[\widehat{\mathcal{Q}}_T^{-1}(\tau)] \to \tau$ for all $\tau \in (0, 1)$ such that $\mathcal{Q}_0^{-1}(\tau) \in \mathcal{C}_0$ where $\mathcal{Q}^{-1}(\tau) := \{\inf x : \mathcal{Q}(x) \geq \tau\}$.

Counterfactuals and Spurious Regressions

- What happens when there are unit-roots but no cointegration?
  - In high-dimensions, we do not know but...
  - In Low-dimensions, Carvalho, Masini and Medeiros (WP, 2017) and Masini and Medeiros (2020, JBES) show that:

Under the some regularity conditions, if no cointegration relation exists, then as $T \to \infty$

$$\frac{1}{\sqrt{T}}\left(\widehat{\Delta}_T - \Delta_T\right) \Rightarrow h(\boldsymbol{W}(r), \text{nuisance parameters})$$

where h is a known function of

- $\boldsymbol{W}(r), r \in [0, 1]$ is a standard Wiener process on $[0, 1]^n$;
- intervention fraction $\lambda_0 = T_0/T$; and
- nuisance parameters.

Monte Carlo Simulation

Rejection Rates under the Null (empirical size): Deterministic Trends

	LASSO			**Oracle**			**True**		
	0.01	**0.5**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
				Innovation Distribution					
Normal	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
$\chi^2(1)$	0.0198	0.0602	0.1078	0.0231	0.0703	0.1277	0.0198	0.0591	0.1076
t-stud(3)	0.0187	0.0632	0.1144	0.0275	0.0781	0.1299	0.0208	0.0602	0.1086
Mixed-Normal	0.0205	0.0603	0.1105	0.0300	0.0775	0.1339	0.0186	0.0572	0.1049
				Sample Size					
$T = 50$	0.0270	0.0768	0.1320	0.0494	0.1144	0.1740	0.0262	0.0694	0.1210
100	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
150	0.0194	0.0632	0.1094	0.0220	0.0644	0.1212	0.0152	0.0536	0.1050
200	0.0182	0.0578	0.1042	0.0202	0.0592	0.1116	0.0164	0.0526	0.1018
500	0.0138	0.0530	0.1016	0.0140	0.0544	0.1004	0.0104	0.0514	0.1006
				Number of Total Units					
$n = 200$	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
300	0.0236	0.0671	0.1175	0.0281	0.0743	0.1281	0.0198	0.0579	0.1053
500	0.0268	0.0748	0.1206	0.0289	0.0780	0.1327	0.0224	0.0626	0.1099
1000	0.0325	0.0778	0.1304	0.0273	0.0755	0.1298	0.0193	0.0554	0.1089
				Number of Relevant (non-zero) Covariates					
$s_0 = 2$	0.0201	0.0634	0.1152	0.0210	0.0653	0.1195	0.0174	0.0573	0.1036
5	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
50	0.0223	0.0661	0.1153	0.2480	0.3547	0.4290	0.0196	0.0606	0.1079
97	0.0217	0.0626	0.1088	1.0000	1.0000	1.0000	0.0233	0.0607	0.1091

Monte Carlo Simulation

Rejection Rates under the Null (empirical size): Deterministic Trends

	LASSO			**Oracle**			**True**		
	0.01	**0.5**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
				Deterministic Component					
$f_t^F = \sqrt{t}$	0.0280	0.0809	0.1367	0.0255	0.0745	0.1299	0.0195	0.0572	0.1068
t	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
$t^{3/2}$	0.0317	0.0823	0.1407	0.0314	0.0855	0.1413	0.0224	0.0630	0.1112
t^2	0.0253	0.0685	0.1177	0.0263	0.0742	0.1280	0.0178	0.0508	0.1005
				Serial Correlation					
$\rho = 0$	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
0.5	0.0216	0.0607	0.1134	0.0278	0.0749	0.1281	0.0199	0.0574	0.1037
0.7	0.0246	0.0720	0.1245	0.0308	0.0812	0.1384	0.0191	0.0590	0.1046
0.9	0.0342	0.0889	0.1404	0.0486	0.1111	0.1745	0.0220	0.0635	0.1111
				Post Intervention Periods					
$T_2 = 1$	0.0166	0.0583	0.1061	0.0151	0.0572	0.1099	0.0121	0.0562	0.1027
2	0.0198	0.0631	0.1109	0.0273	0.0685	0.1185	0.0125	0.0566	0.1033
3	0.0205	0.0637	0.1169	0.0297	0.0755	0.1275	0.0207	0.0583	0.1079
4	0.0301	0.0717	0.1247	0.0370	0.0896	0.1467	0.0256	0.0670	0.1151
5	0.0286	0.0686	0.1184	0.0448	0.0933	0.1537	0.0279	0.0650	0.1127

Monte Carlo Simulation

Rejection Rates under the Null (empirical size): Stochastic Trends

	LASSO			**Oracle**			**True**		
	0.01	**0.5**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
				Innovation Distribution					
Normal	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
$\chi^2(1)$	0.0260	0.0765	0.1385	0.0244	0.0727	0.1308	0.0209	0.0598	0.1060
t-stud(3)	0.0282	0.0831	0.1444	0.0261	0.0779	0.1355	0.0194	0.0581	0.1118
Mixed-Normal	0.0357	0.0912	0.1444	0.0330	0.0862	0.1426	0.0208	0.0615	0.1103
				Sample Size					
$T = 50$	0.0566	0.1155	0.1791	0.0512	0.1071	0.1663	0.0247	0.0641	0.1086
100	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
150	0.0226	0.0686	0.1208	0.0216	0.0664	0.1174	0.0156	0.0526	0.0988
200	0.0193	0.0630	0.1145	0.0190	0.0617	0.1143	0.0156	0.0542	0.1022
500	0.0106	0.0546	0.1026	0.0108	0.0544	0.1010	0.0104	0.0520	0.0966
				Number of Total Units					
$n = 200$	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
300	0.0391	0.0875	0.1479	0.0274	0.0748	0.1290	0.0184	0.0581	0.1039
500	0.0471	0.0953	0.1520	0.0281	0.0802	0.1358	0.0198	0.0610	0.1088
1000	0.0583	0.1085	0.1575	0.0293	0.0764	0.1300	0.0224	0.0590	0.1042
				Number of Relevant (non-zero) Covariates					
$s_0 = 2$	0.0256	0.0698	0.1272	0.0225	0.0667	0.1213	0.0188	0.0558	0.1054
5	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
50	0.0497	0.1117	0.1797	0.2541	0.3636	0.4441	0.0174	0.0572	0.1058
97	0.0574	0.1251	0.1950	1.0000	1.0000	1.0000	0.0203	0.0579	0.1060

Monte Carlo Simulation

Rejection Rates under the Null (empirical size): Stochastic Trends

	LASSO			**Oracle**			**True**		
	0.01	**0.5**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
				Deterministic Component					
$f_t^F = 0$	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
1	0.0314	0.0815	0.1373	0.0316	0.0815	0.1393	0.0205	0.0615	0.1122
$\sqrt{t}$	0.0264	0.0693	0.1191	0.0294	0.0814	0.1380	0.0215	0.0605	0.1083
t	0.0265	0.0711	0.1225	0.0292	0.0768	0.1334	0.0184	0.0560	0.1050
				Serial Correlation					
$\rho = 0$	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
0.5	0.0297	0.0785	0.1313	0.0280	0.0761	0.1320	0.0178	0.0572	0.1019
0.7	0.0275	0.0773	0.1335	0.0264	0.0781	0.1342	0.0211	0.0575	0.1064
0.9	0.0299	0.0752	0.1278	0.0323	0.0823	0.1359	0.0222	0.0631	0.1107
				Post Intervention Periods					
$T_2 = 1$	0.0321	0.0753	0.1273	0.0304	0.0714	0.1201	0.0295	0.0690	0.1151
2	0.0289	0.0777	0.1316	0.0271	0.0762	0.1311	0.0219	0.0759	0.1224
3	0.0324	0.0824	0.1384	0.0319	0.0770	0.1348	0.0220	0.0611	0.1095
4	0.0396	0.0930	0.1522	0.0345	0.0879	0.1430	0.0212	0.0608	0.1087
5	0.0516	0.1088	0.1695	0.0464	0.1021	0.1641	0.0293	0.0661	0.1181

Monte Carlo Simulation

Rejection Rates under the alternative (empirical power)

	Deterministic Trends									
	0.01	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**0.1**
	Mean Intervention $\delta_t = c\sigma 1\{t > T_0\}$									
$c = 0.2$	0.10	0.12	0.14	0.16	0.17	0.19	0.20	0.22	0.23	0.25
0.4	0.23	0.27	0.32	0.35	0.37	0.40	0.43	0.46	0.47	0.48
0.6	0.48	0.51	0.56	0.60	0.63	0.65	0.67	0.69	0.70	0.71
0.8	0.76	0.79	0.82	0.86	0.88	0.89	0.91	0.91	0.92	0.93
1.0	0.94	0.95	0.97	0.97	0.97	0.98	0.98	0.98	0.99	0.99
	Variance Intervention $\delta_t = c\sigma Z 1\{t > T_0\}$ where $Z \sim N(0,1)$									
$c = 0.2$	0.09	0.12	0.13	0.15	0.17	0.18	0.20	0.22	0.24	0.25
0.4	0.26	0.29	0.32	0.36	0.38	0.39	0.41	0.44	0.46	0.48
0.6	0.50	0.54	0.58	0.63	0.66	0.69	0.70	0.71	0.73	0.74
0.8	0.78	0.81	0.85	0.88	0.89	0.91	0.92	0.92	0.92	0.93
1.0	0.93	0.95	0.96	0.97	0.97	0.97	0.98	0.98	0.99	0.99

Monte Carlo Simulation

Rejection Rates under the alternative (empirical power)

	Stochastic Trends									
	0.01	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**0.1**
	Mean Intervention $\delta_t = c\sigma 1\{t > T_0\}$									
$c = 0.1$	0.19	0.20	0.24	0.28	0.30	0.32	0.33	0.36	0.38	0.39
0.2	0.63	0.67	0.72	0.73	0.76	0.78	0.80	0.81	0.81	0.83
0.3	0.95	0.96	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98
0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Variance Intervention $\delta_t = c\sigma Z 1\{t > T_0\}$									
$c = 0.1$	0.17	0.20	0.22	0.25	0.27	0.30	0.32	0.33	0.35	0.37
0.2	0.57	0.60	0.65	0.68	0.70	0.72	0.75	0.76	0.78	0.79
0.3	0.91	0.92	0.94	0.96	0.96	0.97	0.97	0.98	0.98	0.98
0.4	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Application: Heterogeneous Price Elasticities

Heterogeneous Price Elasticities

Setup

- Experiments in order to estimate the price elasticity of a specific group of products from a large retailer in Brazil.

Heterogeneous Price Elasticities

Setup

- Experiments in order to estimate the price elasticity of a specific group of products from a large retailer in Brazil.
- More than 1,000 stores distributed over 400 municipalities in Brazil. **Prices are determined at a municipal level**.

Heterogeneous Price Elasticities

Setup

- Experiments in order to estimate the price elasticity of a specific group of products from a large retailer in Brazil.
- More than 1,000 stores distributed over 400 municipalities in Brazil. **Prices are determined at a municipal level**.
- On Average, the company sells more than 29,000 unities per day of this specific group of products, which represents an important share of the company total's revenue.

Heterogeneous Price Elasticities

Setup

- Experiments in order to estimate the price elasticity of a specific group of products from a large retailer in Brazil.
- More than 1,000 stores distributed over 400 municipalities in Brazil. **Prices are determined at a municipal level**.
- On Average, the company sells more than 29,000 unities per day of this specific group of products, which represents an important share of the company total's revenue.
- Optimal pricing policy for this family of products is of utmost importance.

Heterogeneous Price Elasticities

Setup

- Experiments in order to estimate the price elasticity of a specific group of products from a large retailer in Brazil.
- More than 1,000 stores distributed over 400 municipalities in Brazil. **Prices are determined at a municipal level**.
- On Average, the company sells more than 29,000 unities per day of this specific group of products, which represents an important share of the company total's revenue.
- Optimal pricing policy for this family of products is of utmost importance.
- Prices **could** be set at a municipal level. High degree of heterogeneity.

Heterogeneous Price Elasticities

Setup

- 50% of the stores were divided in two groups: Control and Treatment groups.
  - Control group: 126 municipalities
  - Treatment group: 107 municipalities

Heterogeneous Price Elasticities

Setup

- 50% of the stores were divided in two groups: Control and Treatment groups.
  - Control group: 126 municipalities
  - Treatment group: 107 municipalities
- The selection of the treatment and control groups was carried out according to socioeconomic and demographic characteristics of each municipality as well as to the distribution of stores in each city.

Heterogeneous Price Elasticities

Setup

- 50% of the stores were divided in two groups: Control and Treatment groups.
  - Control group: 126 municipalities
  - Treatment group: 107 municipalities
- The selection of the treatment and control groups was carried out according to socioeconomic and demographic characteristics of each municipality as well as to the distribution of stores in each city.
- The randomization process has not used any information about quantities sold of the product in each municipality, which is our output variable.

Heterogeneous Price Elasticities

Setup

- 50% of the stores were divided in two groups: Control and Treatment groups.
  - Control group: 126 municipalities
  - Treatment group: 107 municipalities
- The selection of the treatment and control groups was carried out according to socioeconomic and demographic characteristics of each municipality as well as to the distribution of stores in each city.
- The randomization process has not used any information about quantities sold of the product in each municipality, which is our output variable.
- DiD methods not adequate: no parallel trends and not useful to measure heterogeneous effects (either per state or municipality).

Heterogeneous Price Elasticities

Setup

- 50% of the stores were divided in two groups: Control and Treatment groups.
  - Control group: 126 municipalities
  - Treatment group: 107 municipalities
- The selection of the treatment and control groups was carried out according to socioeconomic and demographic characteristics of each municipality as well as to the distribution of stores in each city.
- The randomization process has not used any information about quantities sold of the product in each municipality, which is our output variable.
- DiD methods not adequate: no parallel trends and not useful to measure heterogeneous effects (either per state or municipality).
- Time-series of sold quantities displays a clear trend and a seasonal pattern.

Heterogeneous Price Elasticities

Setup

- q_{it}: total quantities sold of the group of products on all stores of municipality i.

Heterogeneous Price Elasticities

Setup

- q_{it}: total quantities sold of the group of products on all stores of municipality i.
- Sample runs from June 20, 2016 to October 31, 2016, a total of 134 daily observations.

Heterogeneous Price Elasticities

Setup

- q_{it}: total quantities sold of the group of products on all stores of municipality i.
- Sample runs from June 20, 2016 to October 31, 2016, a total of 134 daily observations.
- The experiment was conducted during the period October 18-31, 14 days.

Heterogeneous Price Elasticities

Setup

- q_{it}: total quantities sold of the group of products on all stores of municipality i.
- Sample runs from June 20, 2016 to October 31, 2016, a total of 134 daily observations.
- The experiment was conducted during the period October 18-31, 14 days.
- During these days, the practiced prices in the municipalities belonging to the treatment group were reduced in Δ_p Brazilian Reais while for the other municipalities where kept fixed.

Heterogeneous Price Elasticities

Setup

- q_{it}: total quantities sold of the group of products on all stores of municipality i.
- Sample runs from June 20, 2016 to October 31, 2016, a total of 134 daily observations.
- The experiment was conducted during the period October 18-31, 14 days.
- During these days, the practiced prices in the municipalities belonging to the treatment group were reduced in Δ_p Brazilian Reais while for the other municipalities where kept fixed.
- The first 126 municipalities are in the control group ($i = 1, \ldots, 126$) whereas the remaining 107 are in the treatment group ($i = 127, \ldots, 233$).

Heterogeneous Price Elasticities

Setup

- In order to determine the optimal price of the product it is necessary to obtain the effects of the price change on the quantities sold.

Heterogeneous Price Elasticities

Setup

- In order to determine the optimal price of the product it is necessary to obtain the effects of the price change on the quantities sold.

- We consider two cases.

  1. Effects are homogeneous across municipalities and our output variable of interest is the total quantity of the product sold in treatment group ($q_{1t} = \frac{1}{107} \sum_{i=126}^{233} q_{it}$).

  2. Effects are heterogeneous.

Heterogeneous Price Elasticities

Setup

- In order to determine the optimal price of the product it is necessary to obtain the effects of the price change on the quantities sold.
- We consider two cases.
  1. Effects are homogeneous across municipalities and our output variable of interest is the total quantity of the product sold in treatment group ($q_{1t} = \frac{1}{107} \sum_{i=126}^{233} q_{it}$).
  2. Effects are heterogeneous.
- Linear Demand

Heterogeneous Price Elasticities

Optimal price computation

- **Linear demand curve**:

$$\widehat{\beta}_i = \frac{\widehat{\Delta}_i}{N_i \Delta p},$$

where:

  - $\widehat{\Delta}_i$ is the estimated treatment effect;
  - N_i: number of treated stores in municipality i;
  - Δp is the price change.

Heterogeneous Price Elasticities

Optimal price computation

- **Linear demand curve**:

$$\widehat{\beta}_i = \frac{\widehat{\Delta}_i}{N_i \Delta p},$$

where:

  - $\widehat{\Delta}_i$ is the estimated treatment effect;
  - N_i: number of treated stores in municipality i;
  - Δp is the price change.

- **Elasticity**:

$$\widehat{\epsilon}_i = \frac{\widehat{\beta}_i p_{i,T_0-1}}{\overline{Q}_i},$$

where:

  - p_{i,T_0-1} is the price before the change;
  - $\overline{Q}_i$ is the counterfactual quantity sold.

Heterogeneous Price Elasticities

Optimal price computation

- **Linear demand curve**:

$$\widehat{\beta}_i = \frac{\widehat{\Delta}_i}{N_i \Delta p},$$

where:

  - $\widehat{\Delta}_i$ is the estimated treatment effect;
  - N_i: number of treated stores in municipality i;
  - Δp is the price change.

- **Elasticity**:

$$\widehat{\epsilon}_i = \frac{\widehat{\beta}_i p_{i,T_0-1}}{\overline{Q}_i},$$

where:

  - p_{i,T_0-1} is the price before the change;
  - $\overline{Q}_i$ is the counterfactual quantity sold.

- **Optimal price**:

$$\boxed{p_i^* = \frac{(1-\mathsf{taxes}_i)(\overline{Q}_i - \widehat{\beta}_i p_{i,T_0-1}) - \widehat{\beta}_i \times \mathsf{Costs}_i}{-2\widehat{\beta}_i(1-\mathsf{Taxes}_i)}.}$$

Heterogeneous Price Elasticities

Homogeneous effects

1. Estimate the parameters of the regression

$$\boxed{\begin{aligned} q_t &= \beta_0 + \sum_{i=1}^{126} \beta_i q_{it} + \pi_1 \mathsf{Mon}_t + \pi_2 \mathsf{Tue}_t + \pi_3 \mathsf{Wed}_t + \pi_4 \mathsf{Thu}_t \\ &\quad + \pi_5 \mathsf{Fri}_t + \pi_6 \mathsf{Sat}_t + V_t, \\ &= \boldsymbol{X}_t \boldsymbol{\beta} + V_t \end{aligned}}$$

by the WLASSO procedure using the 120 observations from June 20, 2016 to October 18, 2016 (pre-treatment sample). The penalty parameter of the WLASSO procedure is selected by the BIC.

Heterogeneous Price Elasticities

Homogeneous effects

1. Estimate the parameters of the regression

$$\boxed{\begin{aligned} q_t &= \beta_0 + \sum_{i=1}^{126} \beta_i q_{it} + \pi_1 \mathsf{Mon}_t + \pi_2 \mathsf{Tue}_t + \pi_3 \mathsf{Wed}_t + \pi_4 \mathsf{Thu}_t \\ &\quad + \pi_5 \mathsf{Fri}_t + \pi_6 \mathsf{Sat}_t + V_t, \\ &= \boldsymbol{X}_t \boldsymbol{\beta} + V_t \end{aligned}}$$

   by the WLASSO procedure using the 120 observations from June 20, 2016 to October 18, 2016 (pre-treatment sample). The penalty parameter of the WLASSO procedure is selected by the BIC.

2. Project the counterfactual for the treatment period as

$$\boxed{\widehat{q}_t = \boldsymbol{X}_t \widehat{\boldsymbol{\beta}}}$$

   and compute

$$\boxed{\delta_t = q_t - \widehat{q}_t.}$$

Heterogeneous Price Elasticities

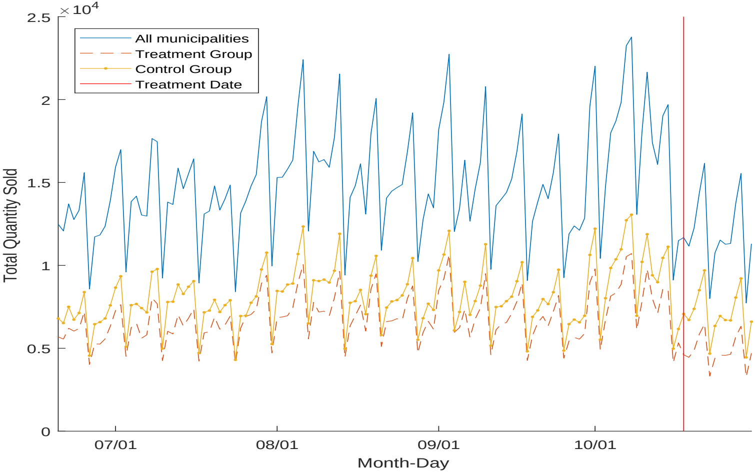
Heterogeneous Effects

- In order to measure the degree of heterogeneity of price elasticities across different municipalities, we estimate the counterfactuals for each one of the municipalities in the treatment group.

$$
\begin{aligned}
q_{jt} = \beta_{k0} + \sum_{i=1}^{126} \beta_{ki} q_{it} + \pi_{k1}\mathsf{Mon}_t + \pi_{k2}\mathsf{Tue}_t + \pi_{k3}\mathsf{Wed}_t + \pi_{k4}\mathsf{Thu}_t \\
+ \pi_{k5}\mathsf{Fri}_t + \pi_{k6}\mathsf{Sat}_t + V_{jt}, \\
= \boldsymbol{X}_{jt}\boldsymbol{\beta}_k + V_{jt}, \quad , \; = 126, \ldots, 233, \; k = j - 126.
\end{aligned}
$$

Heterogeneous Price Elasticities

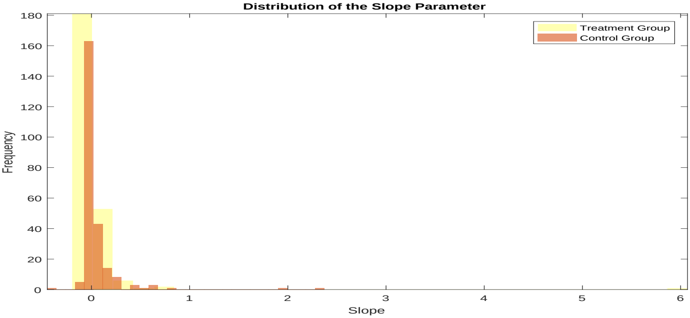
Data: Quantities sold



Heterogeneous Price Elasticities

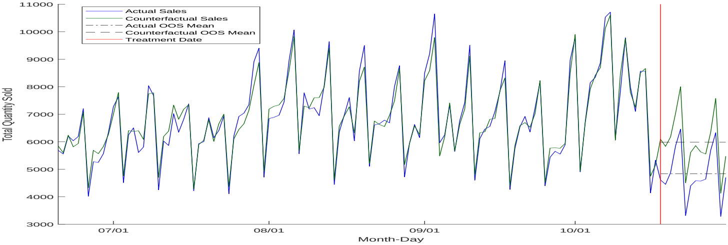
Descriptive Statistics

Aggregated Data: Trend Parameters and ADF Test						
	All	Treatment Group	Control Group	All	Treatment Group	Control Group
Intercept	$13,458.57$	$6,157.138$	$7,301.43$	$8,359.75$	$3,958.18$	$4,401.58$
	(530.68)	(238.33)	(301.35)	(614.94)	(257.12)	(371.18)
Slope	26.08	11.93	14.16	26.55	12.09	14.46
	(10.14)	(4.49)	(5.73)	(10.62)	(4.57)	(6.30)
Days-of-the Week Dummies	No	No	No	Yes	Yes	Yes
ADF (p-value)	0.06	0.07	0.00			

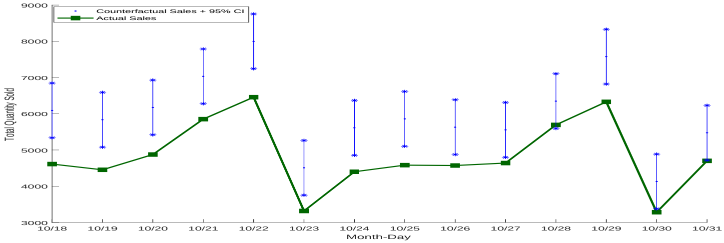


Heterogeneous Price Elasticities

Actual and counterfactual sales



(a) Actual and Counterfactual Sales



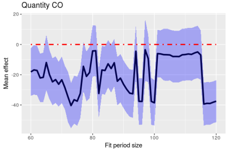
(b) Actual and Counterfactual Sales during Treatment Period

Results

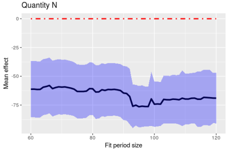
	Panel (a): **Aggregated**	Panel (b) **Disaggregated**			
		Mean	Std. Dev.	Max.	Min.
Δ	-1,147	-12.90	52.08	5.52	-526.70
Δ/#shops	-4.33	-4.21	4.42	5.52	-23.27
p-value (square)	0	0.41	0.29	1	0
p-value (absolute)	0	0.36	0.31	1	0
Proportion (%) of rejection of the null (square)	NA	19	NA	NA	NA
Proportion (%) of rejection of the null (absolute)	NA	31	NA	NA	NA
R-squared	0.96	0.44	0.25	0.95	0
Number of regressors	133	133	NA	NA	NA
Number of relevant regressors	26	9.46	8.06	72	0
Number of pre-treatment observations	120	120	NA	NA	NA
Number of observations during treatment	14	14	NA	NA	NA

Application

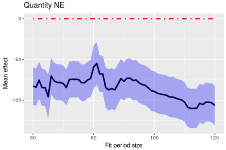
Mean treatment effect using LASSO



(c) CO

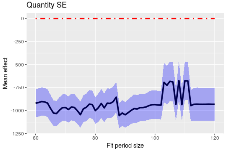


(d) N



(e) NE

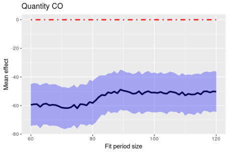
(f) S



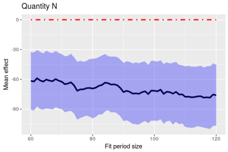
(g) SE

Heterogeneous Price Elasticities

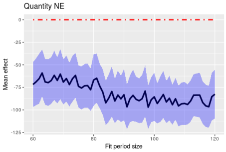
Mean treatment effect using Random Forest



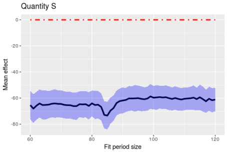
(h) CO



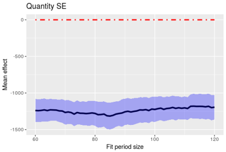
(i) N



(j) NE



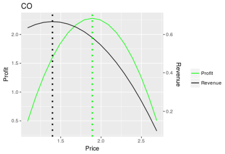
(k) S



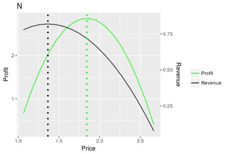
(l) SE

Heterogeneous Price Elasticities

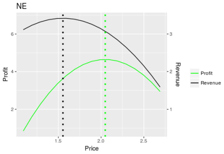
Optimal Pricing with Random Forest



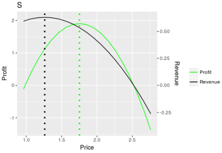
(m) CO



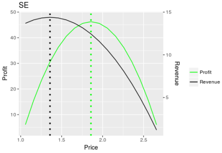
(n) N



(o) NE



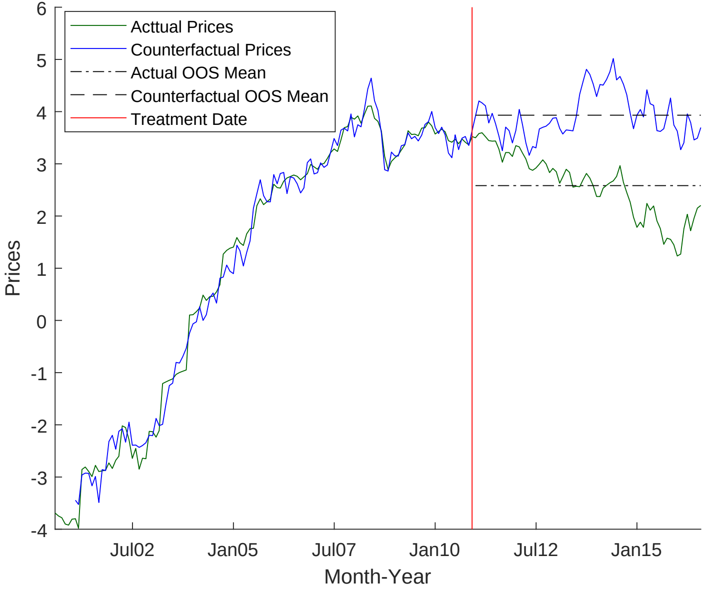
(p) S



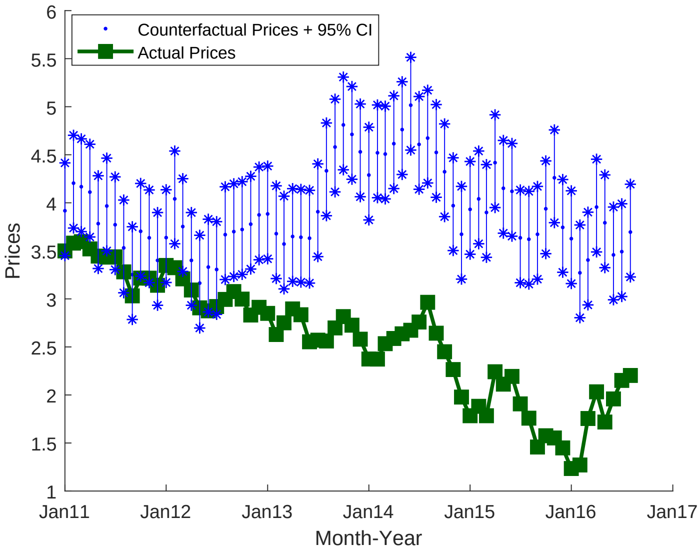
(q) SE

Application: Petrobrás Prices

Additional Example: The Petrobrás Case

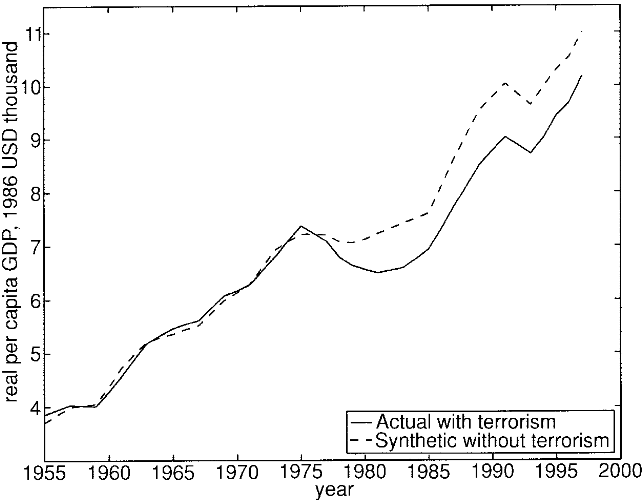


Additional Example: The Petrobrás Case



Conclusions

- General approach for counterfactual analysis with time-series data: stationary and nonstationary.
- By controlling for common shocks that might have occurred in all units after the intervention, it provides a effective methodology to isolate the effect of the intervention of interest.
- R package available at CRAN: Fonseca, Masini, Medeiros and Vasconcelos (2018, R Journal).


real per capita GDP, 1986 USD thousand
11
10
9
8
7
6
5
4
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
year
Actual with terrorism
Synthetic without terrorism

back