ArCo: An Artificial Counterfactual Approach for High-Dimensional Data

Carlos V. de Carvalho
BCB and PUC-Rio

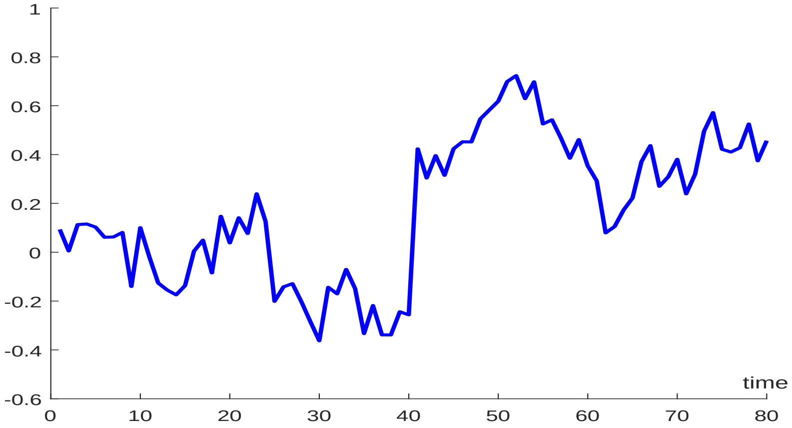
Marcelo C. Medeiros
PUC-Rio

Ricardo Masini
São Paulo School of Economics

BIG DATA, MACHINE LEARNING AND THE MACROECONOMY
Norges Bank, Oslo, 2-3 October 2017

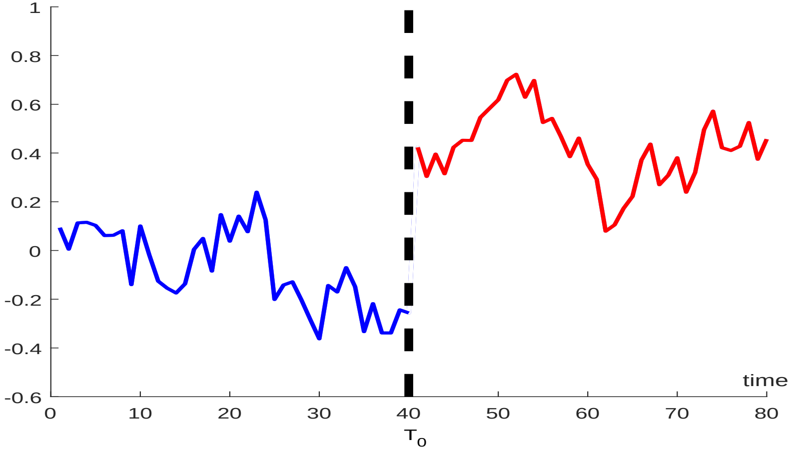
Overview of the Method

Observe aggregated time series data from $t = 1$ to T.



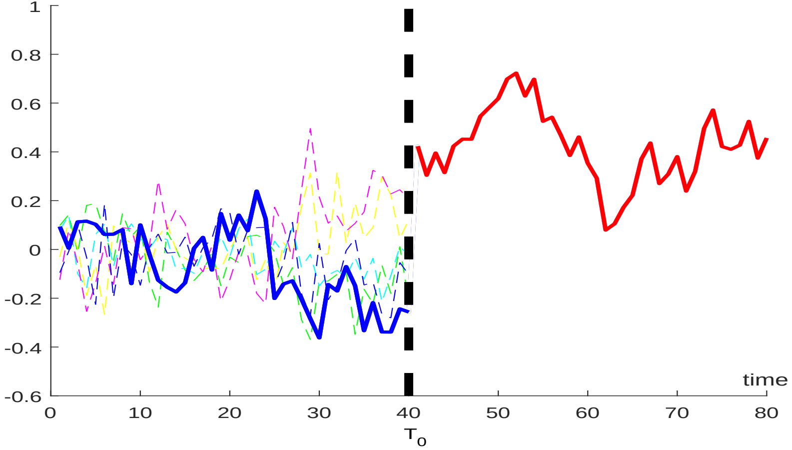
Overview of the Method

Intervention occurs at $t = T_0$.



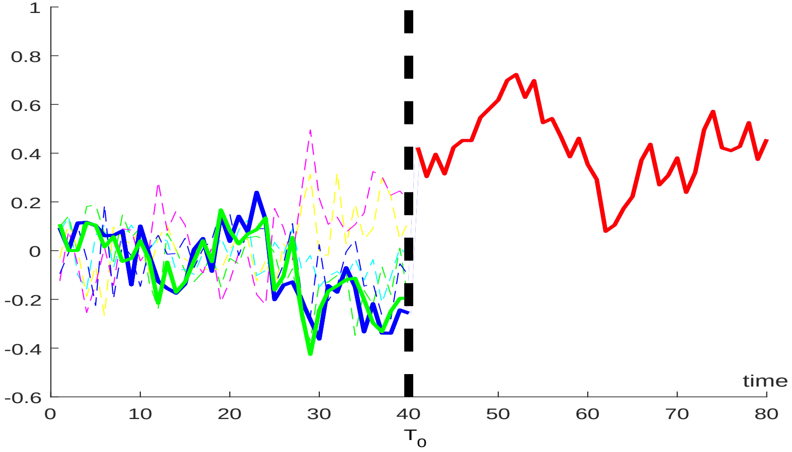
Overview of the Method

No clear controls. Observed variables from untreated “peers”.



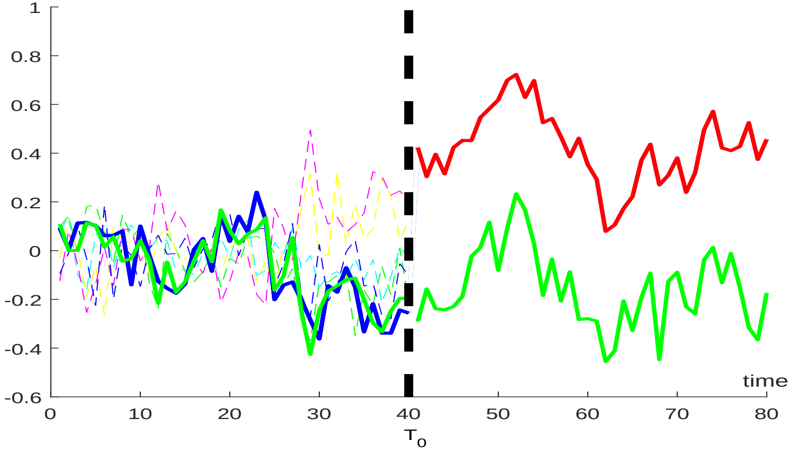
Overview of the Method

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



Overview of the Method

Counterfactual extrapolation (after the intervention).



One-page Summary

- **Artificial Counterfactual (ArCo)**: alternative framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available.

One-page Summary

- **Artificial Counterfactual (ArCo)**: alternative framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available.
  - Generalization of the **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and the **panel approach** of Hsiao, Ching and Wan (2012, JAE).

One-page Summary

- **Artificial Counterfactual (ArCo)**: alternative framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available.
  - Generalization of the **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and the **panel approach** of Hsiao, Ching and Wan (2012, JAE).
- **Our contribution:** estimator with ❶ asymptotic theory, ❷ rigorous inference, ❸ high-dimensionality, and ❹ new tests such as detection of the time of the intervention, contamination effects and effects on multiple moments.

One-page Summary

- **Artificial Counterfactual (ArCo)**: alternative framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available.
  - Generalization of the **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and the **panel approach** of Hsiao, Ching and Wan (2012, JAE).
- **Our contribution:** estimator with ❶ asymptotic theory, ❷ rigorous inference, ❸ high-dimensionality, and ❹ new tests such as detection of the time of the intervention, contamination effects and effects on multiple moments.
- **Key hypothesis**: ❶ Interventions takes place in variables that can be modelled as **trend-stationary processes**. ❷ The intervention affects **only** the unit of interest

One-page Summary

- **Artificial Counterfactual (ArCo)**: alternative framework to measure the impact of an **intervention** in aggregate data when a **control group** is not readily available.
  - Generalization of the **Synthetic Control** (SC) method of Abadie and Gardeazabal (2003, AER) and the **panel approach** of Hsiao, Ching and Wan (2012, JAE).
- **Our contribution:** estimator with ❶ asymptotic theory, ❷ rigorous inference, ❸ high-dimensionality, and ❹ new tests such as detection of the time of the intervention, contamination effects and effects on multiple moments.
- **Key hypothesis**: ❶ Interventions takes place in variables that can be modelled as **trend-stationary processes**. ❷ The intervention affects **only** the unit of interest

The road map

1. The setup
2. The counterfactual estimation
3. Estimator properties
4. Inference
5. Extensions
6. Simulations
7. Empirical example: *Nota Fiscal Paulista*
8. Research agenda
9. Concluding remarks

Setup

- Observe $q_i > 0$ variables for $i = 1, \ldots n$ units for $t = 1, \ldots T$ periods (Panel structure): $\boldsymbol{z}_{it} = (z_{it}^1, \ldots, z_{it}^{q_i})'$.

Setup

- Observe $q_i > 0$ variables for $i = 1, \ldots n$ units for $t = 1, \ldots T$ periods (Panel structure): $\boldsymbol{z}_{it} = (z_{it}^1, \ldots, z_{it}^{q_i})'$.
- **Unit** 1 is exposed to the intervention at **known** T_0.

Setup

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

Setup

- Observe $q_i > 0$ variables for $i = 1, \dots n$ units for $t = 1, \dots T$ periods (Panel structure): $\boldsymbol{z}_{it} = (z_{it}^1, \dots, z_{it}^{q_i})'$.
- **Unit** 1 is exposed to the intervention at **known** T_0.
- The remanning $n - 1$ units $\boldsymbol{z}'_{0t} \equiv (\boldsymbol{z}'_{2t}, \dots, \boldsymbol{z}'_{nt})'$ are an **untreated potential control group (*donor pool*)**.
- **Potential Outcome notation**:

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

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

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

Setup

- Effects on functions of $\boldsymbol{z}_{1t}$: $\boldsymbol{h}\left(\boldsymbol{z}_{1t}\right) : \mathbb{R}^{q_1} \longrightarrow \mathbb{R}^{q}$

$$\boldsymbol{y}_t = \boldsymbol{h}\left(\boldsymbol{z}_{1t}\right) \quad \text{e.g. } h(v_t) = \begin{cases} v_t^p \\ |v_t| \end{cases}$$

$$\boldsymbol{y}_t = d_t \boldsymbol{y}_t^{(1)} + (1 - d_t)\boldsymbol{y}_t^{(0)}$$

Setup

- Effects on functions of $\boldsymbol{z}_{1t}$: $\boldsymbol{h}\left(\boldsymbol{z}_{1t}\right) : \mathbb{R}^{q_1} \longrightarrow \mathbb{R}^{q}$

$$\boldsymbol{y}_t = \boldsymbol{h}\left(\boldsymbol{z}_{1t}\right) \quad \text{e.g. } h(v_t) = \begin{cases} v_t^p \\ |v_t| \end{cases}$$

$$\boldsymbol{y}_t = d_t \boldsymbol{y}_t^{(1)} + (1 - d_t)\boldsymbol{y}_t^{(0)}$$

- Hypothesis of interest: $\boldsymbol{y}_t^{(1)} = \boldsymbol{\delta}_t + \boldsymbol{y}_t^{(0)}, \quad t = T_0 \ldots, T,$

$$\boxed{\mathcal{H}_0 : \boldsymbol{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \underbrace{\left[\boldsymbol{y}_t^{(1)} - \boldsymbol{y}_t^{(0)}\right]}_{\equiv \boldsymbol{\delta}_t} = \mathbf{0}.}$$

Setup

- Effects on functions of $\boldsymbol{z}_{1t}$: $\boldsymbol{h}(\boldsymbol{z}_{1t}) : \mathbb{R}^{q_1} \longrightarrow \mathbb{R}^{q}$

$$\boldsymbol{y}_t = \boldsymbol{h}(\boldsymbol{z}_{1t}) \quad \text{e.g. } h(v_t) = \begin{cases} v_t^p \\ |v_t| \end{cases}$$

$$\boldsymbol{y}_t = d_t \boldsymbol{y}_t^{(1)} + (1 - d_t)\boldsymbol{y}_t^{(0)}$$

- Hypothesis of interest: $\boldsymbol{y}_t^{(1)} = \boldsymbol{\delta}_t + \boldsymbol{y}_t^{(0)}, \quad t = T_0 \ldots, T,$

$$\boxed{\mathcal{H}_0 : \boldsymbol{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \underbrace{\left[\boldsymbol{y}_t^{(1)} - \boldsymbol{y}_t^{(0)}\right]}_{\equiv \boldsymbol{\delta}_t} = \boldsymbol{0}.}$$

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

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

Counterfactual Estimation

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

Counterfactual Estimation

- How should we construct $\widehat{\boldsymbol{y}}_t^{(0)}$?
- Ideally (in the MSE sense) we would set

$$\widehat{\boldsymbol{y}}_t^{(0)} = \mathbb{E}(\boldsymbol{y}_t^{(0)} | \boldsymbol{z}_{0t}, \boldsymbol{z}_{0t-1}, \ldots, \boldsymbol{z}_{0t-p})$$

Counterfactual Estimation

- How should we construct $\widehat{\boldsymbol{y}}_t^{(0)}$?
- Ideally (in the MSE sense) we would set

$$\widehat{\boldsymbol{y}}_t^{(0)} = \mathbb{E}(\boldsymbol{y}_t^{(0)}|\boldsymbol{z}_{0t}, \boldsymbol{z}_{0t-1}, \ldots, \boldsymbol{z}_{0t-p})$$

- In practice, we choose a (parametric) specification (could be linear or not). Let $\boldsymbol{x}_t = (\boldsymbol{z}'_{0t}, \boldsymbol{z}'_{0t-1}, \ldots, \boldsymbol{z}'_{0t-p})'$ and

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

such that $\mathbb{E}(\boldsymbol{\nu}_t) = \mathbf{0}$ and

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

ArCo estimator

- The Artificial Counterfactual (ArCo) estimator is then simply given by

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

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

ArCo estimator

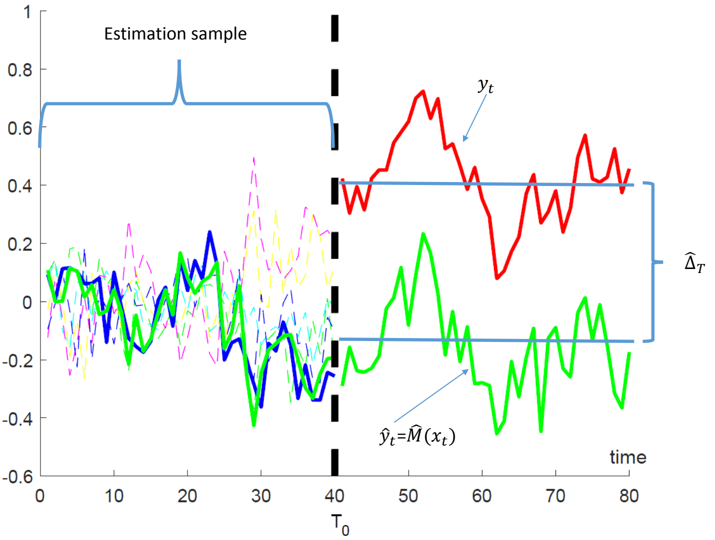
- The Artificial Counterfactual (ArCo) estimator is then simply given by

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

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

- ArCo estimator is a two-step estimator:
  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 $\widehat{\boldsymbol{\Delta}}_T$.

ArCo estimator



ArCo and 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** set of variables from the peers.
  - Correct specification ($\mathcal{M}$ is the conditional expectation).
  - Selection of peers by information criteria.
  - Date of the intervention known.

ArCo and 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** set of variables from the peers.
  - Correct specification ($\mathcal{M}$ is the conditional expectation).
  - Selection of peers by information criteria.
  - Date of the intervention known.
- Differences-in-Differences (DiD)
  - Number of treated units must grow.
  - Parallel trends hypothesis.
  - Similar control group.

ArCo and 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** set of variables from the peers.
  - Correct specification ($\mathcal{M}$ is the conditional expectation).
  - Selection of peers by information criteria.
  - Date of the intervention known.

- 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.

ArCo and 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.

ArCo and 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.
- Pesaran et al. (2004, JBES); Dees et al. (2007, JAE)
  - Counterfactual based on variables that belong to the treated unit. No donor pool.
  - Their key assumption is that a subset of variables of the treated unit is invariant to the intervention.

ArCo and 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.
- Pesaran et al. (2004, JBES); Dees et al. (2007, JAE)
  - Counterfactual based on variables that belong to the treated unit. No donor pool.
  - Their key assumption is that a subset of variables of the treated unit is invariant to the intervention.
- Angrist, Jordà and Kuersteiner (2016, JBES)
  - Information only on the treated unit and no donor pool is available.

ArCo estimator

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.

ArCo estimator

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.

ArCo estimator

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), ...

ArCo estimator

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), ...

ArCo estimator

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), ...
  - **New government or political regime**: Grier and Maynard (2013, JEBO)

ArCo estimator

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.

ArCo estimator

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):
  - Abadie and Diamond (2004, AER) and Abadie, Diamond, and Hainmueller (2010, JASA) have, respectively, 1712 and 1280 Google cites as measured on August 2, 2017.

ArCo estimator

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):
  - Abadie and Diamond (2004, AER) and Abadie, Diamond, and Hainmueller (2010, JASA) have, respectively, 1712 and 1280 Google cites as measured on August 2, 2017.
  - "*Arguably the most important innovation in the evaluation literature in the last fifteen years is the synthetic control method.*"

    Athey and Imbens (2016)

Counterfactual Estimation

- Why do we expect the method to work?

Counterfactual Estimation

- Why do we expect the method to work?
- Possible data generating mechanism (in reduced form):

$$\boxed{\begin{aligned} \boldsymbol{z}_{it}^{(0)} &= \boldsymbol{\mu}_i + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{ij} \boldsymbol{\varepsilon}_{it-j} \\ \boldsymbol{\varepsilon}_{it} &= \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \boldsymbol{\eta}_{it} \end{aligned}}$$

where $\boldsymbol{f}_t (f \times 1) \sim (\mathbf{0}, \boldsymbol{Q})$ is a vector of common unobserved factors . $\boldsymbol{\Lambda}_i (q_i \times f)$ are matrices of factor loadings; $\boldsymbol{\eta}_{it} (q_i \times 1) \sim (\mathbf{0}, \boldsymbol{R}_i)$ is idiosyncratic error term.

Counterfactual Estimation

- Why do we expect the method to work?
- Possible data generating mechanism (in reduced form):

$$\boldsymbol{z}_{it}^{(0)} = \boldsymbol{\mu}_i + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{ij}\boldsymbol{\varepsilon}_{it-j}$$

$$\boldsymbol{\varepsilon}_{it} = \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \boldsymbol{\eta}_{it}$$

  where $\boldsymbol{f}_t(f \times 1) \sim (\mathbf{0}, \boldsymbol{Q})$ is a vector of common unobserved factors . $\boldsymbol{\Lambda}_i(q_i \times f)$ are matrices of factor loadings; $\boldsymbol{\eta}_{it}(q_i \times 1) \sim (\mathbf{0}, \boldsymbol{R}_i)$ is idiosyncratic error term.
- Quality of the pool of donors comes from the factor structure.

Counterfactual estimation

Least Absolute Shrinkage and Selection Operator (LASSO)

- Set $\boldsymbol{X}_t = (1, \boldsymbol{x}_t)' \in \mathbb{R}^d$ and consider $\mathcal{M}(\boldsymbol{x}_t)$ linear:

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

Counterfactual estimation

Least Absolute Shrinkage and Selection Operator (LASSO)

- Set $\boldsymbol{X}_t = (1, \boldsymbol{x}_t)' \in \mathbb{R}^d$ and consider $\mathcal{M}(\boldsymbol{x}_t)$ linear:

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

- Estimation:

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

Counterfactual estimation

Least Absolute Shrinkage and Selection Operator (LASSO)

- Set $\boldsymbol{X}_t = (1, \boldsymbol{x}_t)' \in \mathbb{R}^d$ and consider $\mathcal{M}(\boldsymbol{x}_t)$ linear:

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

- Estimation:

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

- Why LASSO?
  - Avoid overfitting.
  - Large dataset compared to the sample size.
  - "Automatic" model selection.

Counterfactual estimation

LASSO – Catalog of hypotheses

Design

Let $\boldsymbol{\Sigma} \equiv \frac{1}{T_1}\sum_{t=1}^{T_1}\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ and $S_0 = \{i : \theta_{0,1} \neq 0\}$ (set of non-zero parameters). There exists a constant $\psi_0 > 0$ such that

$$\|\boldsymbol{\theta}[S_0]\|_1^2 \leq \frac{\boldsymbol{\theta}\boldsymbol{\Sigma}\boldsymbol{\theta} s_0}{\psi_0^2},$$

for all $\|\boldsymbol{\theta}[S_0^c]\|_1 \leq 3\|\boldsymbol{\theta}[S_0]\|_1$.

Counterfactual estimation

LASSO – Catalog of hypotheses

Design

Let $\boldsymbol{\Sigma} \equiv \frac{1}{T_1}\sum_{t=1}^{T_1}\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ and $S_0 = \{i : \theta_{0,1} \neq 0\}$ (set of non-zero parameters). There exists a constant $\psi_0 > 0$ such that

$$\|\boldsymbol{\theta}[S_0]\|_1^2 \leq \frac{\boldsymbol{\theta}\boldsymbol{\Sigma}\boldsymbol{\theta}s_0}{\psi_0^2},$$

for all $\|\boldsymbol{\theta}[S_0^c]\|_1 \leq 3\|\boldsymbol{\theta}[S_0]\|_1$.

- Compatibility condition of Bülhmann and van der Geer (2011).

Counterfactual estimation

LASSO – Catalog of hypotheses

Design

Let $\boldsymbol{\Sigma} \equiv \frac{1}{T_1}\sum_{t=1}^{T_1}\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ and $S_0 = \{i : \theta_{0,1} \neq 0\}$ (set of non-zero parameters). There exists a constant $\psi_0 > 0$ such that

$$\|\boldsymbol{\theta}[S_0]\|_1^2 \leq \frac{\boldsymbol{\theta}\boldsymbol{\Sigma}\boldsymbol{\theta} s_0}{\psi_0^2},$$

for all $\|\boldsymbol{\theta}[S_0^c]\|_1 \leq 3\|\boldsymbol{\theta}[S_0]\|_1$.

- Compatibility condition of Bülhmann and van der Geer (2011).
- Similar to the restriction of the smallest eigenvalue of $\boldsymbol{\Sigma}$.

Counterfactual estimation

LASSO – Catalog of hypotheses

Design

Let $\boldsymbol{\Sigma} \equiv \frac{1}{T_1}\sum_{t=1}^{T_1}\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ and $S_0 = \{i : \theta_{0,1} \neq 0\}$ (set of non-zero parameters). There exists a constant $\psi_0 > 0$ such that

$$\|\boldsymbol{\theta}[S_0]\|_1^2 \leq \frac{\boldsymbol{\theta}\boldsymbol{\Sigma}\boldsymbol{\theta} s_0}{\psi_0^2},$$

for all $\|\boldsymbol{\theta}[S_0^c]\|_1 \leq 3\|\boldsymbol{\theta}[S_0]\|_1$.

- Compatibility condition of Bülhmann and van der Geer (2011).
- Similar to the restriction of the smallest eigenvalue of $\boldsymbol{\Sigma}$.
- Important for prediction consistency and ℓ_1-consistency of the LASSO.

Counterfactual estimation

LASSO – Catalog of hypotheses

Heterogeneity and dynamics

Let $\boldsymbol{w}_t \equiv (\nu_t, \boldsymbol{x}_t')'$, then:

(a) $\{\boldsymbol{w}_t\}$ is strong mixing with $\alpha(m) = \exp(-cm)$ for some $c \geq \underline{c} > 0$

(b) $\mathbb{E}|w_{it}|^{2\gamma+\delta} \leq c_\gamma$ for some $\gamma > 2$ and $\delta > 0$ for all $1 \leq i \leq d, 1 \leq t \leq T$ and $T \geq 1$,

(c) $\mathbb{E}(\nu_t^2) \geq \epsilon > 0$, for all $1 \leq t \leq T$ and $T \geq 1$.

Counterfactual estimation

LASSO – Catalog of hypotheses

Heterogeneity and dynamics

Let $\boldsymbol{w}_t \equiv (\nu_t, \boldsymbol{x}_t')'$, then:

(a) $\{\boldsymbol{w}_t\}$ is strong mixing with $\alpha(m) = \exp(-cm)$ for some $c \geq \underline{c} > 0$

(b) $\mathbb{E}|w_{it}|^{2\gamma+\delta} \leq c_\gamma$ for some $\gamma > 2$ and $\delta > 0$ for all $1 \leq i \leq d, 1 \leq t \leq T$ and $T \geq 1$,

(c) $\mathbb{E}(\nu_t^2) \geq \epsilon > 0$, for all $1 \leq t \leq T$ and $T \geq 1$.

- $\boldsymbol{w}_t$ is an α-mixing process with exponential decay.

Counterfactual estimation

LASSO – Catalog of hypotheses

Heterogeneity and dynamics

Let $\boldsymbol{w}_t \equiv (\nu_t, \boldsymbol{x}_t')'$, then:

(a) $\{\boldsymbol{w}_t\}$ is strong mixing with $\alpha(m) = \exp(-cm)$ for some $c \geq \underline{c} > 0$

(b) $\mathbb{E}|w_{it}|^{2\gamma+\delta} \leq c_\gamma$ for some $\gamma > 2$ and $\delta > 0$ for all $1 \leq i \leq d, 1 \leq t \leq T$ and $T \geq 1$,

(c) $\mathbb{E}(\nu_t^2) \geq \epsilon > 0$, for all $1 \leq t \leq T$ and $T \geq 1$.

- $\boldsymbol{w}_t$ is an α-mixing process with exponential decay.
- Part (b) bounds uniformly some higher moments $\Rightarrow$ Law of Large Numbers.

Counterfactual estimation

LASSO – Catalog of hypotheses

Heterogeneity and dynamics

Let $\boldsymbol{w}_t \equiv (\nu_t, \boldsymbol{x}_t')'$, then:

(a) $\{\boldsymbol{w}_t\}$ is strong mixing with $\alpha(m) = \exp(-cm)$ for some $c \geq \underline{c} > 0$

(b) $\mathbb{E}|w_{it}|^{2\gamma+\delta} \leq c_\gamma$ for some $\gamma > 2$ and $\delta > 0$ for all $1 \leq i \leq d, 1 \leq t \leq T$ and $T \geq 1$,

(c) $\mathbb{E}(\nu_t^2) \geq \epsilon > 0$, for all $1 \leq t \leq T$ and $T \geq 1$.

- $\boldsymbol{w}_t$ is an α-mixing process with exponential decay.
- Part (b) bounds uniformly some higher moments $\Rightarrow$ Law of Large Numbers.
- Part (c) Sufficient condition for the Central Limit Theorem.

Counterfactual estimation

LASSO – Catalog of hypotheses

Regularity

(a) $\varsigma = O\left(\frac{d^{1/\gamma}}{\sqrt{T}}\right)$

(b) $s_0 \frac{d^{2/\gamma}}{\sqrt{T}} = o(1)$

Counterfactual estimation

LASSO – Catalog of hypotheses

Regularity

(a) $\varsigma = O\left(\frac{d^{1/\gamma}}{\sqrt{T}}\right)$

(b) $s_0 \frac{d^{2/\gamma}}{\sqrt{T}} = o(1)$

- Part (a) puts discipline on the growth rate of the regularization parameter.

Counterfactual estimation

LASSO – Catalog of hypotheses

Regularity

(a) $\varsigma = O\left(\frac{d^{1/\gamma}}{\sqrt{T}}\right)$

(b) $s_0 \frac{d^{2/\gamma}}{\sqrt{T}} = o(1)$

- Part (a) puts discipline on the growth rate of the regularization parameter.
- Part (b) bounds the number of (total/relevant) parameters.

Counterfactual estimation

LASSO – Catalog of hypotheses

Regularity

(a) $\varsigma = O\left(\frac{d^{1/\gamma}}{\sqrt{T}}\right)$

(b) $s_0 \frac{d^{2/\gamma}}{\sqrt{T}} = o(1)$

- Part (a) puts discipline on the growth rate of the regularization parameter.
- Part (b) bounds the number of (total/relevant) parameters.
- Both conditions can be relaxed if normality is assumed.

Counterfactual estimation

LASSO – Results

Consistency and Asymptotic Normality

Let $\mathcal{M}$ be the model defined as before, whose parameters are estimated by LASSO, then under previous assumptions and as $T \to \infty$:

$$\sup_{P \in \mathcal{P}} \sup_{\boldsymbol{a} \in \mathbb{R}^q} \left| \mathbb{P}_P \left[\sqrt{T} \boldsymbol{\Omega}_T^{-1/2} (\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T) \leq \boldsymbol{a} \right] - \Phi(\boldsymbol{a}) \right| \to 0,$$

where $\boldsymbol{\Omega}_T$ is defined in the previous proposition and $\Phi(\cdot)$ is the cumulative distribution function of a zero-mean normal random vector with identity covariance matrix. The inequality is defined element-wise.

Counterfactual estimation

LASSO – Results

Uniform Confidence Interval

- Let $\widehat{\mathbf{\Omega}}_T$ be a consistent estimator for $\mathbf{\Omega}_T$ uniformly in $P \in \mathcal{P}$. Under the same conditions as before:

$$\boxed{\mathcal{I}_\alpha \equiv \left[\widehat{\Delta}_{j,T} \pm \frac{\widehat{\omega}_j}{\sqrt{T}} \Phi^{-1}(1-\alpha/2)\right]}$$

  for each $j = 1, \ldots, q$, where $\widehat{\omega}_j = \sqrt{[\widehat{\mathbf{\Omega}}]_{jj}}$ and $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution.

- $\mathcal{I}_\alpha$ is uniformly valid (honest) in the sense that for a given $\epsilon > 0$, there exists a T_ϵ such that for all $T > T_\epsilon$:

$$\boxed{\sup_{P \in \mathcal{P}} \left|\mathbb{P}_P\left(\Delta_{j,T} \in \mathcal{I}_\alpha\right) - (1-\alpha)\right| < \epsilon.}$$

Counterfactual estimation

LASSO – Results

Uniform Hypothesis Test

Let $\widehat{\boldsymbol{\Omega}}_T$ be a consistent estimator for $\boldsymbol{\Omega}_T$ uniformly in $P \in \mathcal{P}$. Under the same conditions as before, for a given $\epsilon > 0$, there exists a T_ϵ such that for all $T > T_\epsilon$:

$$\boxed{\sup_{P \in \mathcal{P}} \left| \mathbb{P}_P \left(W_T \leq c_\alpha \right) - (1 - \alpha) \right| < \epsilon,}$$

where $W_T \equiv T \widehat{\boldsymbol{\Delta}}'_T \widehat{\boldsymbol{\Omega}}_T^{-1} \widehat{\boldsymbol{\Delta}}_T$, $\mathbb{P}(\chi^2_q \leq c_\alpha) = 1 - \alpha$ and χ^2_q is a chi-square distributed random variable with q degrees of freedom.

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.

2. Multiple interventions.

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.
2. Multiple interventions.
3. Test for the untreated peers and /or treated unit.

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.
2. Multiple interventions.
3. Test for the untreated peers and /or treated unit.
4. The framework is extended to conduct joint test (e.g., affecting the mean and the variance simultaneously)

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.
2. Multiple interventions.
3. Test for the untreated peers and /or treated unit.
4. The framework is extended to conduct joint test (e.g., affecting the mean and the variance simultaneously)
5. R package:
   https://cran.r-project.org/web/packages/ArCo/index.html
   – Fonseca, Masini, Medeiros, and Vasconcelos (2017a)

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.
2. Multiple interventions.
3. Test for the untreated peers and /or treated unit.
4. The framework is extended to conduct joint test (e.g., affecting the mean and the variance simultaneously)
5. R package:
   https://cran.r-project.org/web/packages/ArCo/index.html
   – Fonseca, Masini, Medeiros, and Vasconcelos (2017a)
6. Carvalho, Masini and Medeiros (2017): Integrated data

Additional Results

1. Tests under **unknown** intervention date. Results are not altered when the intervention date is estimated.
   - $\widehat{\lambda}$ is super-consistent.
2. Multiple interventions.
3. Test for the untreated peers and /or treated unit.
4. The framework is extended to conduct joint test (e.g., affecting the mean and the variance simultaneously)
5. R package:
   https://cran.r-project.org/web/packages/ArCo/index.html
   – Fonseca, Masini, Medeiros, and Vasconcelos (2017a)
6. Carvalho, Masini and Medeiros (2017): Integrated data
7. Fonseca, Masini, Medeiros, and Vasconcelos (2017b): Random Forests and other machine learning techniques.

Monte Carlo results

Data Generating Process (DGP)

Consider the following model for $i \in \{1, \ldots, n\}$ and $t \geq 1$:

$$\boldsymbol{z}_{it}^{(0)} = \rho \boldsymbol{A}_i \boldsymbol{z}_{it-1}^{(0)} + \boldsymbol{\varepsilon}_{it},$$

where:

- $\boldsymbol{\varepsilon}_{it} = \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \boldsymbol{\eta}_{it}$,
- $\boldsymbol{f}_t = [1, (t/T)^{\varphi}, v_t]$, $\boldsymbol{z}_{it} \in \mathbb{R}^q$, $\rho \in [0, 1)$, $\varphi > 0$,
- $\boldsymbol{A}_i(q \times q)$ is a diagonal matrix with diagonal elements strictly between -1 and 1,
- $v_t \overset{iid}{\sim} \mathsf{N}(0, 1)$, $\boldsymbol{\eta}_{it} \overset{iid}{\sim} \mathsf{N}(0, r_f^2 \boldsymbol{I}_{nq})$, and $\boldsymbol{\Lambda}_i$ is a $(q \times 3)$ matrix of factor loadings.

Monte Carlo results

Rejection Rates under the Null (1/2)

	Bias	**Var**	$\widehat{s}_0$	$\boldsymbol{\alpha = 0.1}$	**0.05**	**0.01**
	Innovation Distribution $T = 100,\ d = 100,\ s_0 = 5,\ \varphi = 0,\ \rho = 0$					
Normal	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
$\chi^2(1)$	-0.0014	1.1004	5.9287	0.1227	0.0652	0.0154
t-stud(3)	0.0035	1.1026	5.6437	0.1077	0.0543	0.0103
Mixed-Normal	0.0069	1.1267	5.5457	0.1134	0.0607	0.0136
	Sample Size normal dist., $d = 100,\ s_0 = 5,\ \varphi = 0,\ \rho = 0$					
$T = 100$	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
75	-0.0030	1.1449	6.3992	0.1075	0.0546	0.0124
50	0.0021	1.1747	6.1219	0.1092	0.0626	0.0155
25	-0.0050	0.8324	3.2463	0.1330	0.0763	0.0226
	Number of Total Covariates normal dist., $T = 100,\ s_0 = 5,\ \varphi = 0,\ \rho = 0$					
$d = 100$	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
200	-0.0016	1.1655	5.7314	0.1102	0.0565	0.0135
500	-0.0043	1.2112	5.6625	0.1119	0.0556	0.0114
1000	0.0012	1.2477	5.5275	0.1054	0.0566	0.0115

Monte Carlo results

Rejection Rates under the Null (2/2)

	Bias	**Var**	$\widehat{s}_0$	$\boldsymbol{\alpha = 0.1}$	**0.05**	**0.01**
	Number of Relevant (non-zero) Covariates normal dist., $T = 100$, $d = 100$, $\varphi = 0$, $\rho = 0$					
$s_0 = 0$	0.0038	1.0981	0.6105	0.1059	0.0550	0.0136
5	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
10	0.0003	1.0373	9.5813	0.1103	0.0581	0.0120
100	0.0003	-	20.1624	0.1114	0.0574	0.0145
	Determinist Trend $(t/T)^{\varphi}$ normal dist., $T = 100$, $d = 100$, $s_0 = 5$, $\rho = 0$					
$\varphi = 0$	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
0.5	0.0142	1.1245	5.6285	0.1101	0.0598	0.0199
1	0.0183	1.1313	5.5030	0.1188	0.0613	0.0168
2	0.0221	1.1398	5.4259	0.1273	0.0675	0.0261
	Serial Correlation normal dist., $T = 100$, $d = 100$, $s_0 = 5$, $\varphi = 0$					
$\rho = 0.2$	-0.0001	1.4109	5.5246	0.1160	0.0640	0.0158
0.4	0.0002	1.6909	5.9276	0.1223	0.0678	0.0184
0.6	0.0031	1.8895	6.9012	0.1440	0.0871	0.0283
0.8	0.0033	1.9977	7.9464	0.1546	0.0927	0.0329

Monte Carlo results

Rejection rates under the alternative

	$\alpha = 0.1$	0.075	0.05	0.025	0.01
	Step Intervention $\delta_t = c\,\sigma_1 1\{t \geq T_0\}$				
$c = 0.15$	0.2045	0.1695	0.1287	0.0805	0.0436
0.25	0.3783	0.3266	0.2686	0.1890	0.1108
0.35	0.5769	0.5235	0.4545	0.3465	0.2414
0.5	0.8314	0.7945	0.7440	0.6478	0.5227
0.75	0.9876	0.9831	0.9741	0.9520	0.9094
1	0.9998	0.9995	0.9992	0.9983	0.9943
	Linear Increasing $\delta_t = c\,\sigma_1 \frac{t-T_0+1}{T-T_0+1} 1\{t \geq T_0\}$				
$c = 1$	0.8318	0.7938	0.7379	0.6397	0.5121
1.25	0.9877	0.9813	0.9717	0.9459	0.8948
1.5	0.9997	0.9997	0.9990	0.9969	0.9922
	Linear Decreasing $\delta_t = c\,\sigma_1 \frac{T-t+1}{T-T_0+1} 1\{t \geq T_0\}$				
$c = 1$	0.8298	0.7956	0.7434	0.6492	0.5107
1.25	0.9868	0.9818	0.9720	0.9490	0.8985
1.5	0.9995	0.9994	0.9989	0.9968	0.9933

Monte Carlo results

Horse race (1/2)

	BA	SC	DiD*	DiD	GM*	GM	ArCo*	ArCo
	No Time Trend ($\varphi = 0$) and No Serial Correlation ($\rho = 0$)							
Bias	-0.001	-0.678	0.005	0.008	-0.280	-0.273	0.000	0.000
Var	3.151	50.555	17.870	51.444	0.544	0.510	1.001	1.000
MSE	3.152	86.075	17.871	51.449	6.601	6.255	1.001	1.000
	No Time Trend ($\varphi = 0$)							
Bias	-0.003	-0.596	0.000	0.000	-0.353	-0.294	-0.002	-0.002
Var	2.997	12.293	7.215	18.506	3.057	0.705	0.998	1.000
MSE	2.996	27.634	7.214	18.502	8.438	4.427	0.998	1.000
	Common Linear Time Trend ($\varphi = 1$)							
Bias	0.218	-0.579	0.034	0.033	-0.128	-0.195	0.028	0.029
Var	2.900	19.590	6.741	17.720	0.522	0.499	1.007	1.000
MSE	4.677	32.165	6.558	17.159	1.151	1.985	1.004	1.000
	Idiosyncratic Linear Time Trend ($\varphi = 1$)							
Bias	0.744	1.391	0.597	0.577	0.766	0.766	0.161	0.158
Var	0.288	0.564	0.392	1.720	1.499	1.113	0.996	1.000
MSE	2.270	7.544	1.651	2.771	3.493	3.142	0.999	1.000

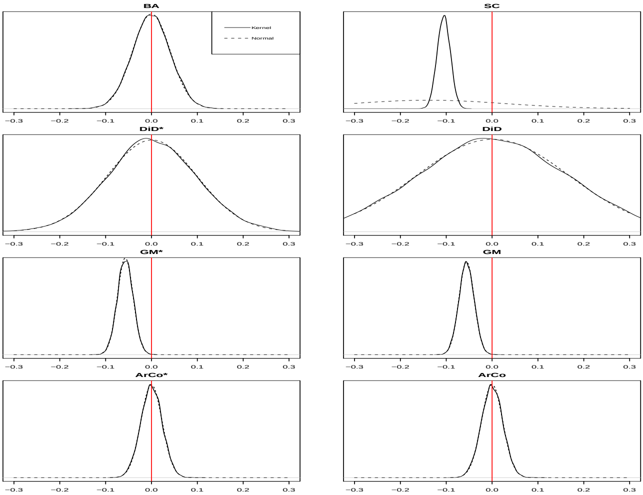
Monte Carlo results

Horse race (2/2)

	BA	SC	DiD*	DiD	GM*	GM	ArCo*	ArCo
	Common Quadratic Time Trend ($\varphi = 2$)							
Bias	0.288	-0.562	0.051	0.053	-0.170	-0.170	0.049	0.048
Var	2.809	18.486	6.571	17.199	0.512	0.488	1.007	1.000
MSE	5.583	28.407	6.105	15.837	1.520	1.498	1.010	1.000
	Idiosyncratic Quadratic Time Trend ($\varphi = 2$)							
Bias	0.994	-0.179	0.780	0.758	0.465	0.465	0.154	0.153
Var	1.443	0.377	3.499	8.878	0.282	0.274	0.992	1.000
MSE	14.786	0.701	10.868	14.002	3.216	3.210	0.998	1.000

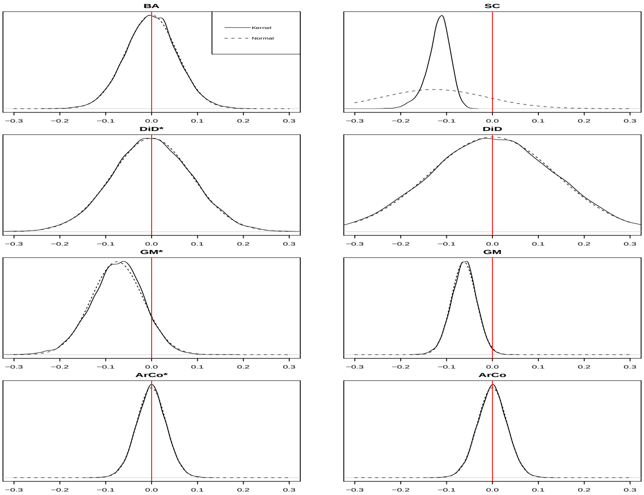
Monte Carlo results

No trend and no serial correlation



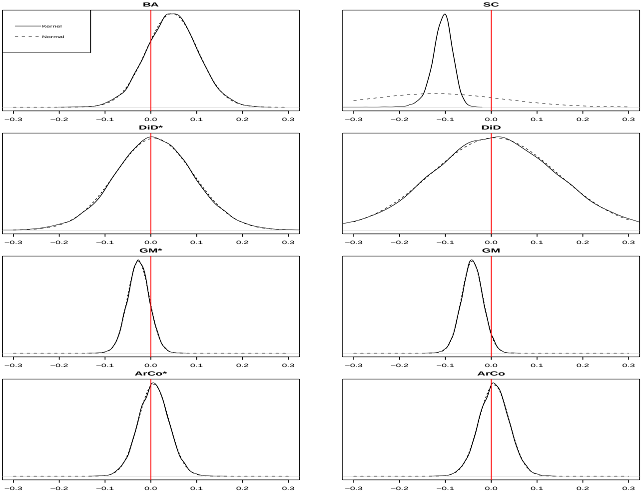
Monte Carlo results

No trend



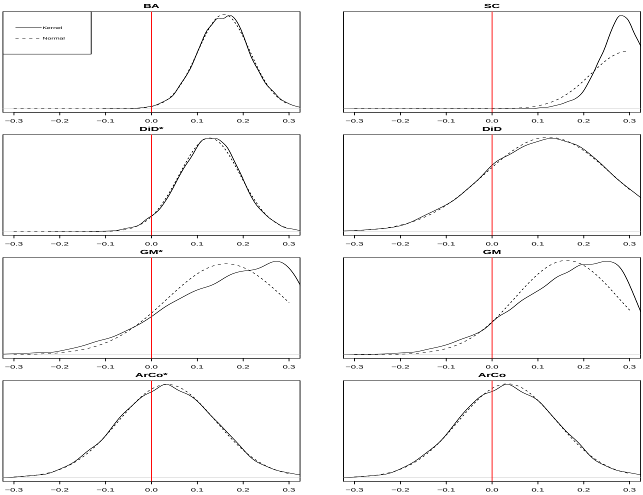
Monte Carlo results

Common linear trend



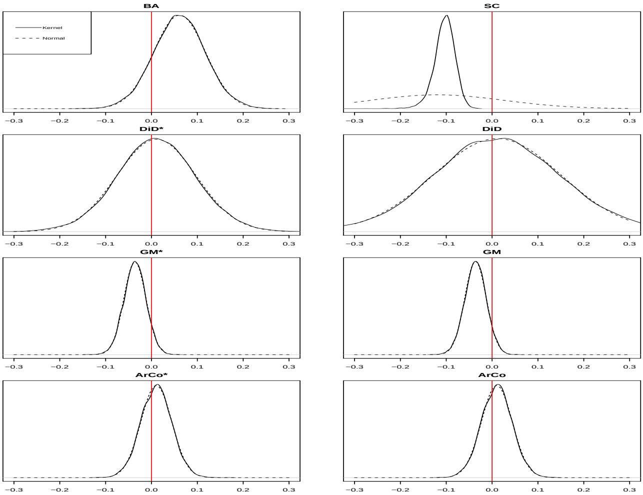
Monte Carlo results

Heterogeneous linear trend



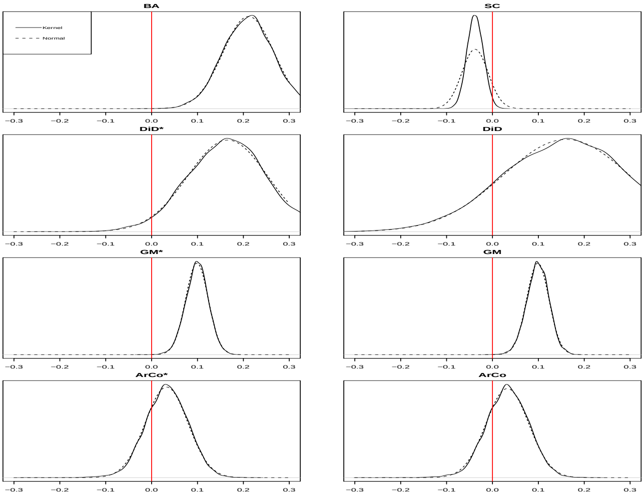
Monte Carlo results

Common quadratic trend



Monte Carlo results

Heterogeneous quadratic trend



Application

- In October 2007, the state government of São Paulo in Brazil implemented a anti tax evasion scheme called *Nota Fiscal Paulista* (NFP).

Application

- In October 2007, the state government of São Paulo in Brazil implemented a anti tax evasion scheme called *Nota Fiscal Paulista* (NFP).
- The NFP consists of a tax rebate from a state tax named ICMS (tax on circulation of products and services).

Application

- In October 2007, the state government of São Paulo in Brazil implemented a anti tax evasion scheme called *Nota Fiscal Paulista* (NFP).
- The NFP consists of a tax rebate from a state tax named ICMS (tax on circulation of products and services).
- Incentive to the consumer to ask for sales receipts.

Application

- In October 2007, the state government of São Paulo in Brazil implemented a anti tax evasion scheme called *Nota Fiscal Paulista* (NFP).
- The NFP consists of a tax rebate from a state tax named ICMS (tax on circulation of products and services).
- Incentive to the consumer to ask for sales receipts.
- Additionally, the registered sales receipts give the consumer the right to participate in monthly lotteries promoted by the government.

Application

- Under the premisses that:
    1. a certain degree of tax evasion was occurring before the intervention,
    2. the sellers has some degree of market power and
    3. the penalty for tax-evasion is large enough to alter the seller behaviour,

  one is expected to see an upwards movements in prices due to an increase in marginal cost.
- Hence, we would like to investigate whether the NFP had an impact on consumer prices in São Paulo.

Application

- The NFP was not implemented throughout the sectors in the economy at once.

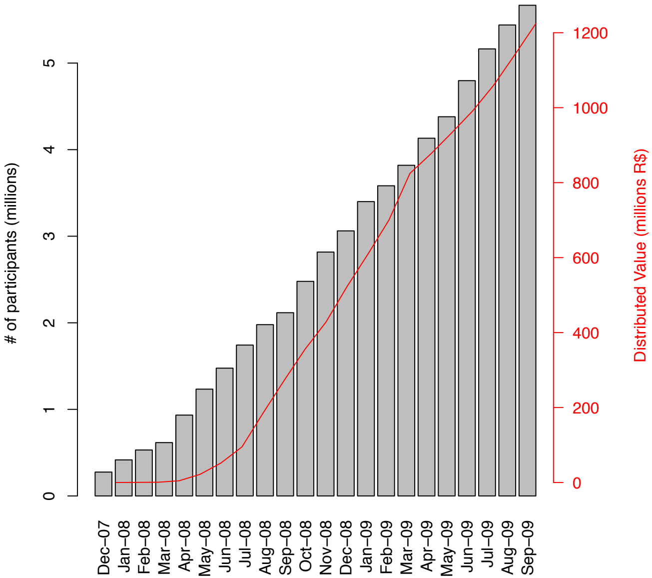
Application

- The NFP was not implemented throughout the sectors in the economy at once.
- The first sector were restaurants, followed by bakeries, bar and other food service retailers.

Application

- The NFP was not implemented throughout the sectors in the economy at once.
- The first sector were restaurants, followed by bakeries, bar and other food service retailers.
- The sample then consists of monthly inflation (food outside home) index for 10 metropolitan areas including São Paulo from Jan-05 to Sep-09.

Application



Application

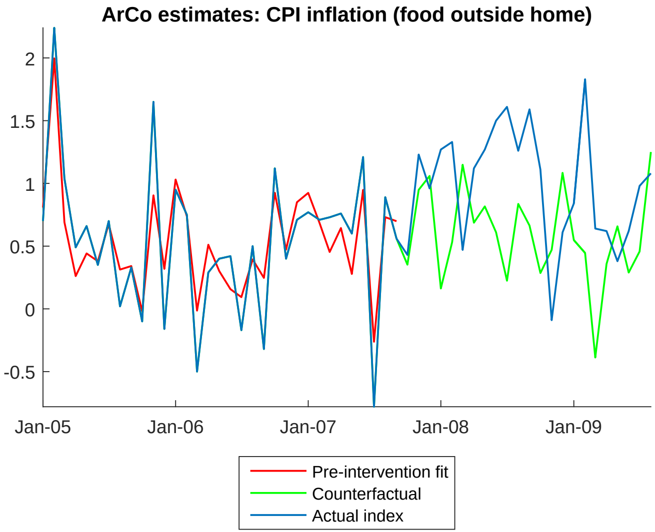
Panel (a): ArCo Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	0.2992 (0.1704)	0.4438 (0.1486)	0.4913 (0.1432)	0.5064 (0.1480)	0.4763 (0.2010)	0.4070 (0.1600)	0.4046 (0.1539)
Inflation	Yes	No	No	No	Yes	Yes	Yes
GDP	No	Yes	No	No	Yes	Yes	Yes
Retail Sales	No	No	Yes	No	No	Yes	Yes
Credit	No	No	No	Yes	No	No	Yes
R-squared	0.6439	0.1213	0.3928	0.1026	0.7960	0.8568	0.8072
d	9	8	9	9	17	26	35
s_0	9	3	7	5	14	17	13
T_1	33	33	33	33	33	33	33
T_2	23	23	23	23	23	23	23

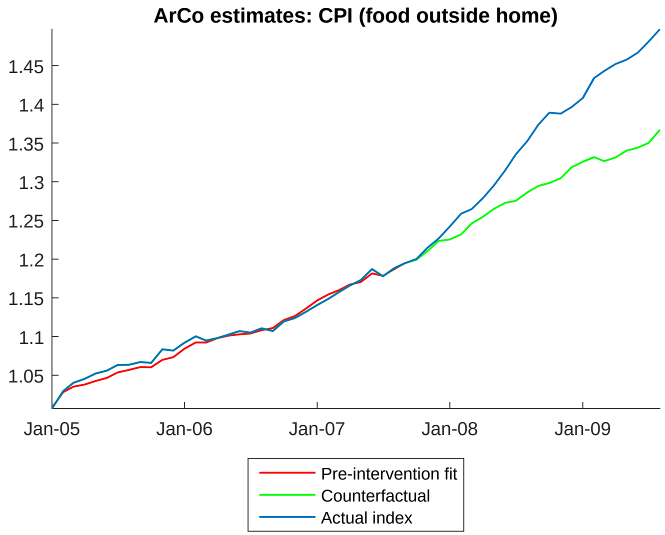
Panel (b): Alternative Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
DiD	0.2524 (0.1466)	0.2407 (0.1456)	0.2494 (0.1467)	0.2412 (0.1556)	0.2387 (0.1457)	0.2520 (0.1466)
GM	0.3694 (0.1234)	0.3788 (0.1243)	0.3595 (0.1246)	0.3775 (0.1227)	0.3660 (0.1228)	–
GDP	Yes	No	No	Yes	Yes	No
Retail Sales	No	Yes	No	Yes	Yes	No
Credit	No	No	Yes	No	Yes	No

Application



Application



Concluding remarks

- Important to note that the ArCo is more than just another structural break estimator.
- 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.
- Providing compelling evidence for the **causality** of the structural break observed in the unit of interest
- It is almost a **model-free** methodology. It implies no structure in the stationary process (no # of lags to be determined nor # MA terms)