Impact Evaluation of a Cluster Program: An Application of Synthetic Control Methods

Diego Aboal*, Gustavo Crespi and Marcelo Perera***

***CINVE**

****IDB**

Roadmap

1. Motivation
2. Objectives
3. The program
4. Data
5. Methodology
6. Results
7. Conclusions

Roadmap

1. Motivation
2. Objectives
3. The program
4. Data
5. Methodology
6. Results
7. Conclusions

Motivation

- **Cluster develpment programs (CDP) are widespread** around the world, including Latin America
- **Clusters** are **agglomeration of firms around specialized productive activities**. Usually they take place at sub-national levels.
- **Cluster policies**: **resolve coordination failures** among firms and between firms and governments in order to **guarantee the provision of club goods** needed for the **competitiveness of the agglomeration**.
- **Only a few impact evaluations available worldwide**: e.g. Figal-Garone et al. (2015), Martin et al. (2011), Nishimura and Okamuro (2011), Falck et el. (2010).
- **Most of them do not account for indirect or "total" effects** of CDPs. A few exceptions: Boneu et al. (2014), Figal-Garone et al. (2015), Castillo et al. (2015).

Roadmap

1. Motivation
2. Objectives
3. The program
4. Data
5. Methodology
6. Results
7. Conclusions

Objective

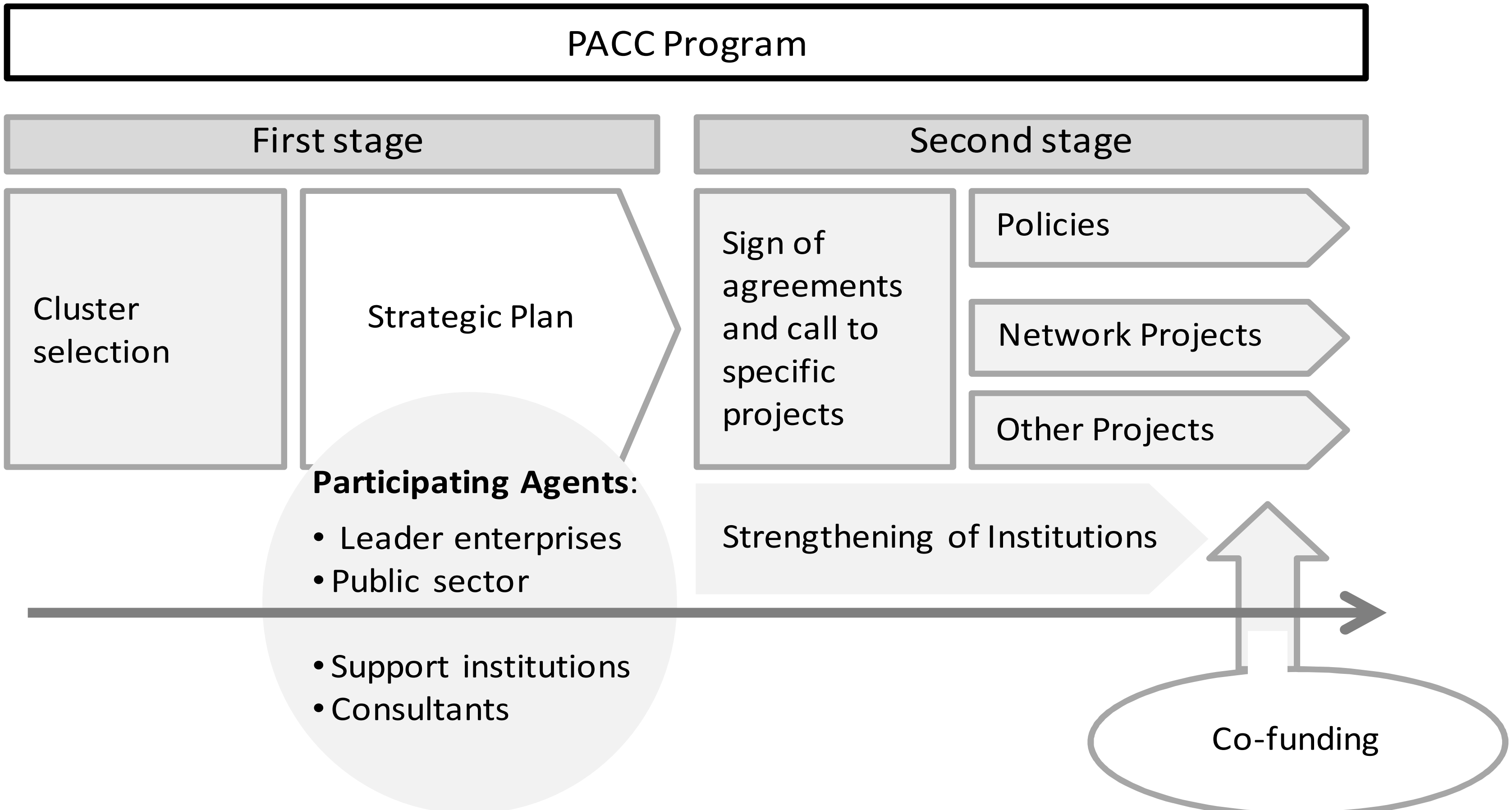
* Evaluate the impact of a Tourism Cluster Program in the Region of Colonia, Uruguay.

* We want to estimate the aggregate effect and not only the one on firms that directly participated in cluster's activities (this is very important given that these programs work through spillovers).

Roadmap

1. Motivation
2. Objectives
3. The program
4. Data
5. Methodology
6. Results
7. Conclusions

The program

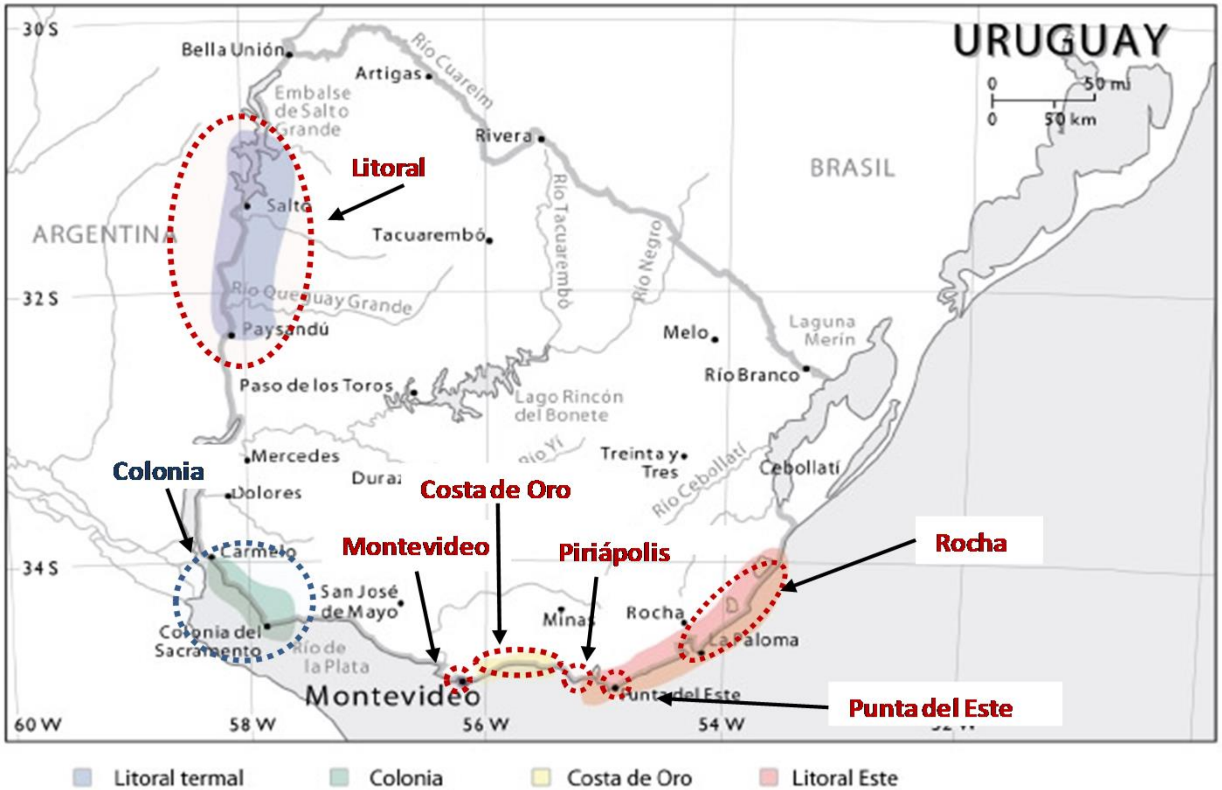


- **IDB supported program. Several initiatives** that required about **US$ 900,000**. Start 2007, most of them implemented in the period **2008-10**.
- Projects: Development of a common trademark, benchmarking exercises with other similar regions around the world, promotion activities, introduction of new marketing technologies, English training for employees, etc..

Roadmap

1. Motivation
2. Objectives
3. The program
4. **Data**
5. Methodology
6. Results
7. Conclussions

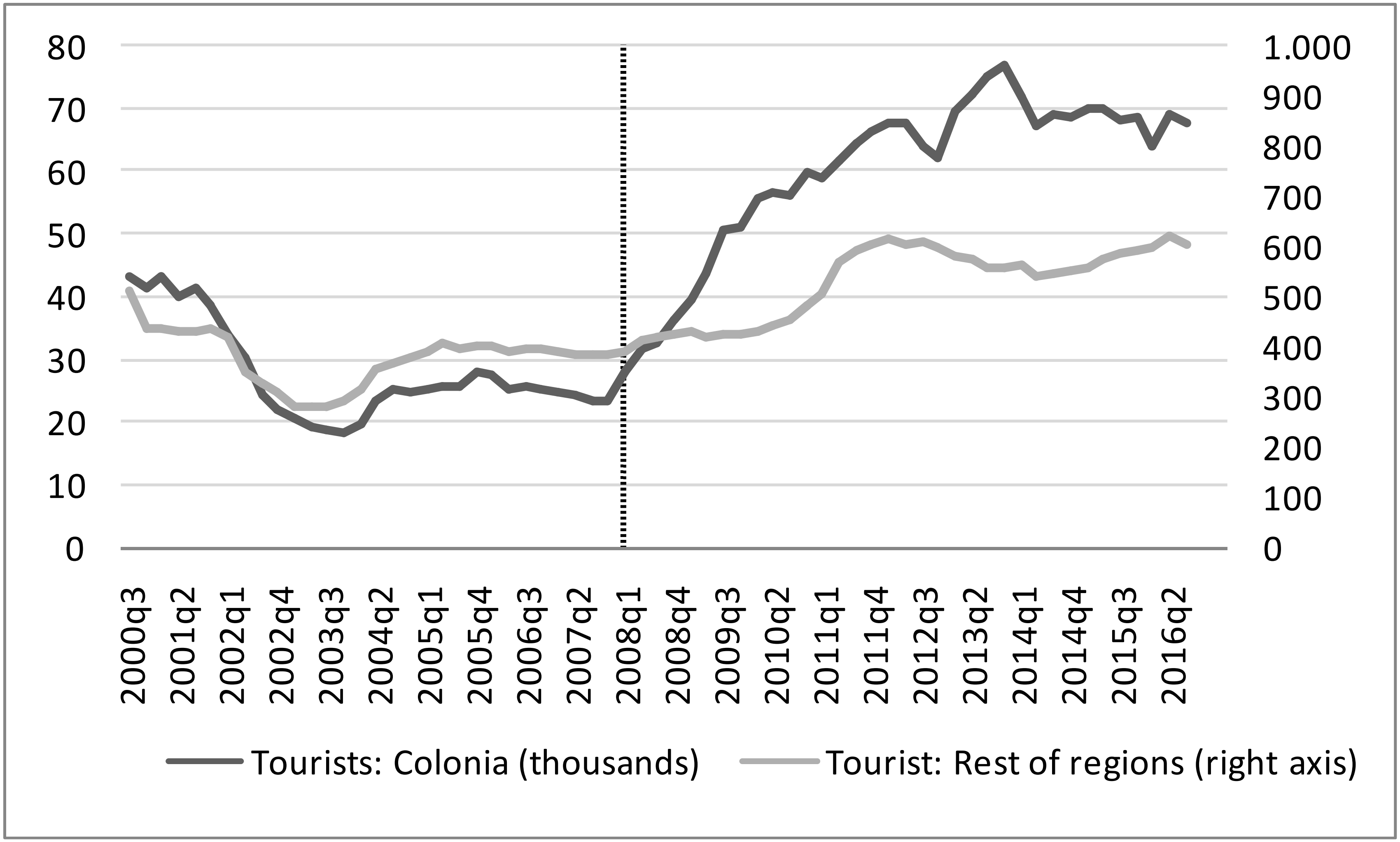
Data



- Main data sources: "Encuesta de Turismo Receptivo", 2000-2016, and Household surveys.
- **Information for Uruguay's seven main touristic destinations**: Colonia, Punta del Este, Montevideo, Costa de Oro, Pirápolis, Rocha and the thermal littoral .
- **Quarterly information** about **number of visitants**, **tourists' expenditures** and **average days of stay of visitants.**

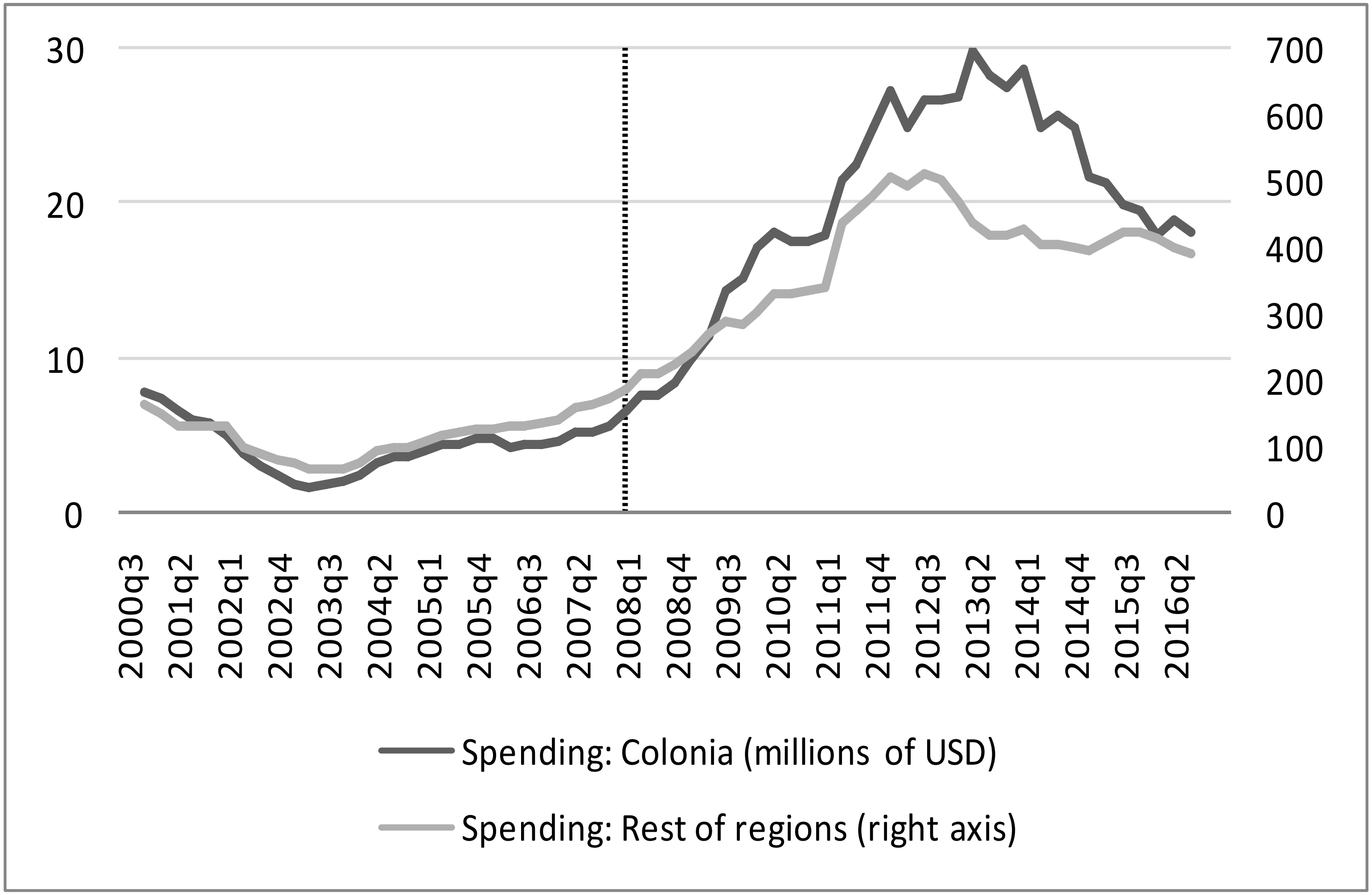
Data

Number of Tourists: Colonia vs. the Other Regions



Data

Total tourists' expenditure: Colonia vs. the Other



Roadmap

1. Motivation
2. Objectives
3. The program
4. Data
5. Methodology
6. Results
7. Conclusions

Empirical Strategy

- We are interested on the **impacts** of a policy intervention that take place **at an aggregate level and affect a geographical area.**

- **The treatment unit and potential controls are aggregated units (regions).**

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) propose a **data-driven procedure to construct suitable comparison groups**: **Synthetic Control Method** (SCM)

Empirical Strategy

- The idea behind the SCM is that **a combination of control units often provides a better comparison for the unit exposed to the intervention than any single unit alone**

- **A Synthetic Control** is a **weighted average of available control units that resembles the treated unit in the pre-treatment period** (makes explicit the relative contribution of each control units)

- **SCM extends the traditional difference-in-differences framework,** allowing that the effects of unobserved variables on the outcome vary with time.

- **And propose a method to perform inferential exercises about the effects of the intervention of interest** (potentially informative regardless of the number of available comparison units).

Synthetic Control Methods (inference)

- The method relies on **placebo o permutation tests for inference**
- **Re-estimate synthetic controls on all *J* donors to get placebo effects**: $\hat{\alpha}_{jt}$
- The two sided **p-values** is then: $= \dfrac{\sum_{j \neq 1} 1(|\hat{\alpha}_{jt}| \geq |\hat{\alpha}_{1t}|)}{J}$
- **Interpretation** is "**what proportion of controls units have estimated effects as large?**"

- **Can take quality of match of pre-treatment period into account**, e.g. an alternative is to divide the effects $\widehat{\alpha}_{jt}$ by the square root of the mean squared error of prediction in the pre-treatment period ($\overleftarrow{s}_j$), and obtain a pseudo t-statistic, $\widehat{\alpha}_{jt}/\overleftarrow{s}_j$.

Roadmap

Results: Number of international tourists

- **Pre-treatment period: 2000q3-2007q4**
- **Post-treatment period: 2008q1-2016q3**

- **Treated Unit: Colonia**
- **Donors**: **6 touristic regions** (Punta del Este, Montevideo, Costa de Oro, Piriápolis, Rocha, Littoral)

- **Outcome variable**: **Number of international tourists**
- **Predictors**: outcome variable for each of the pre-intervention years, expenditure per tourist in 2007 and the average 2005-2007 household income (we have also performed robustness checks including other variables like, informality, employment).

Results: Number of international tourists

Table 1: Syntethic Colonia (regions' weights)

Tourist Region	Weights
Punta del Este	0.00
Montevideo	0.02
Costa de Oro	0.56
Piriapolis	0.00
Rocha	0.20
Litoral	0.22

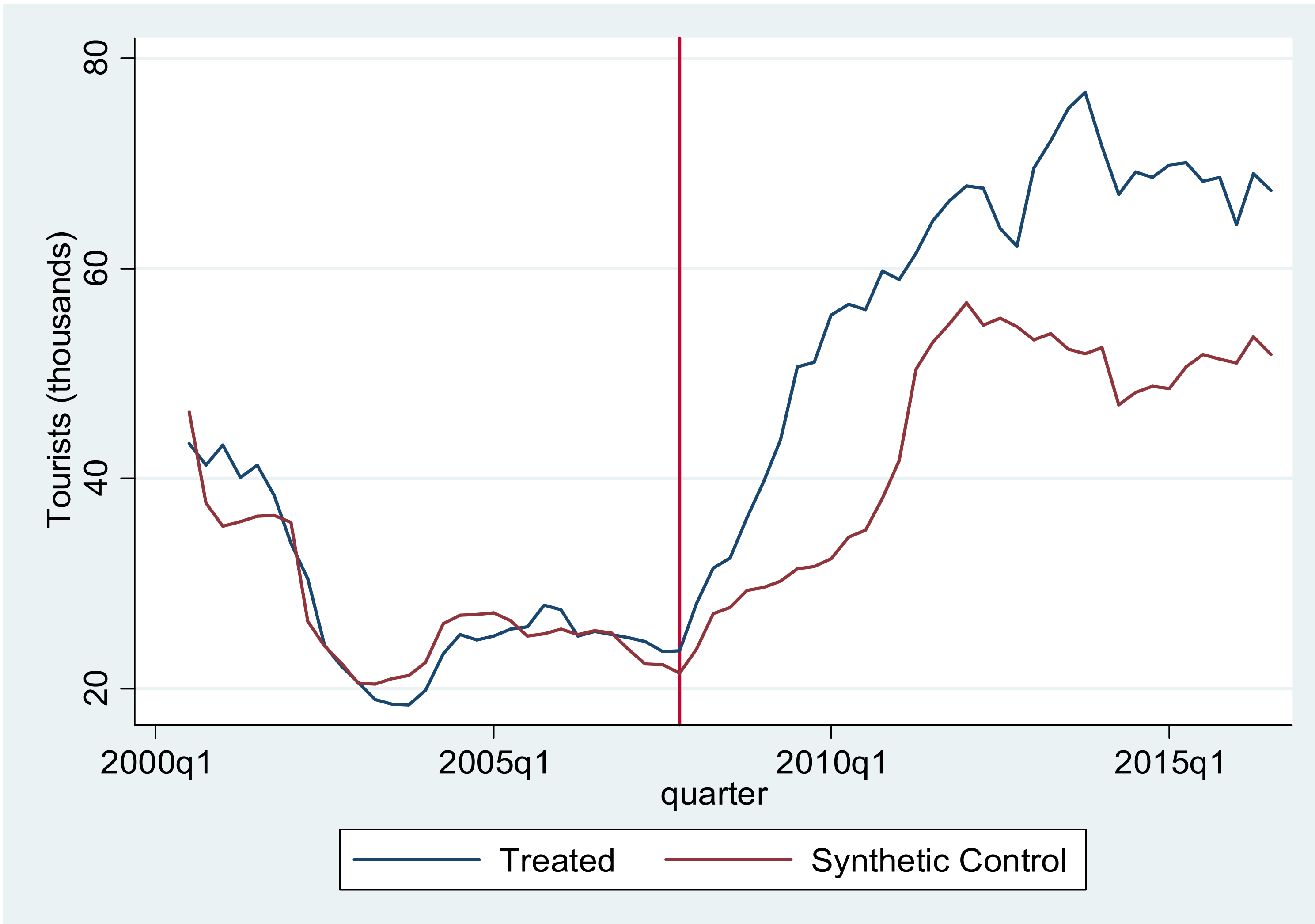
Results: Number of international tourists

Table 2: Predictors' means before treatment

	Colonia	Average of the rest Tourist Regions	Synthetic Colonia
Tourists (thousands)			
2000q3-2000q4	42.3	78.8	44.3
2001q1-2001q4	40.7	72.3	38.2
2002q1-2002q4	27.6	58.7	28.8
2003q1-2003q4	19.1	47.2	21.9
2004q1-2004q4	23.2	58.8	26.9
2005q1-2005q4	26.1	66.4	27.2
2006q1-2006q4	25.8	66.0	26.3
2007q1-2007q4	24.1	64.3	23.2
Spending (millions of USD)			
2000q3-2000q4	8.6	31.3	10.2
2001q1-2001q4	6.5	22.7	7.8
2002q1-2002q4	3.6	16.8	5.2
2003q1-2003q4	1.9	11.4	2.8
2004q1-2004q4	3.3	15.4	3.8
2005q1-2005q4	4.5	19.7	4.3
2006q1-2006q4	4.5	21.9	5
2007q1-2007q4	5.2	26.6	5.4
Spending per tourist (thousands of USD)			
2001q1-2007q4	193.1	344.9	217.7
Per capita household income (USD)			
2005q1-2007q4	725.6	825.8	751.2

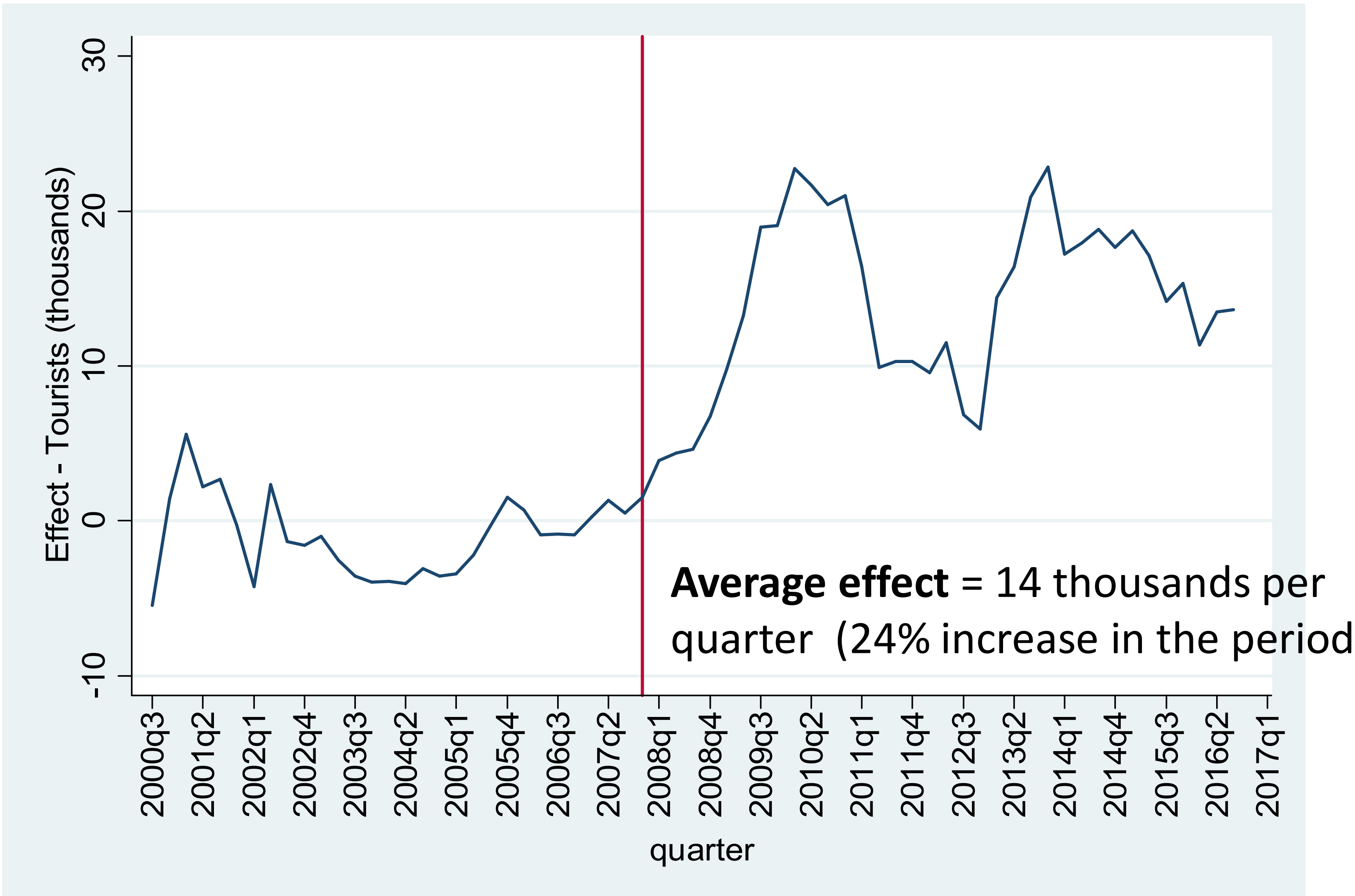
Results: Number of international tourists

Figure 2: Colonia vs Synthetic Colonia 2000q1-2016q3



Results: Number of international tourists

Figure 2: Colonia vs Synthetic Colonia 2000q1-2016q3



Results: Number of international tourists

Table 3: Root Mean Square Error of Prediction (pre and post intervention, and ratio): Colonia vs Placebos

Región	$\overleftarrow{s}_j$	$\overrightarrow{s}_j$	$\overrightarrow{s}_j/\overleftarrow{s}_j$
Colonia	2.7	16.6	6.1
Punta del Este	14.0	17.2	1.2
Montevideo	28.9	48.7	1.7
Costa de Oro	1.6	9.1	5.5
Piriápolis	2.1	8.3	3.9
Rocha	2.9	11.5	3.9
Litoral	15.1	30.4	2.0
p-values:			
$\sum_{j\neq 1} 1\left(\overrightarrow{s}_j/\overleftarrow{s}_j \geq \overrightarrow{s}_1/\overleftarrow{s}_1\right)/J$			0

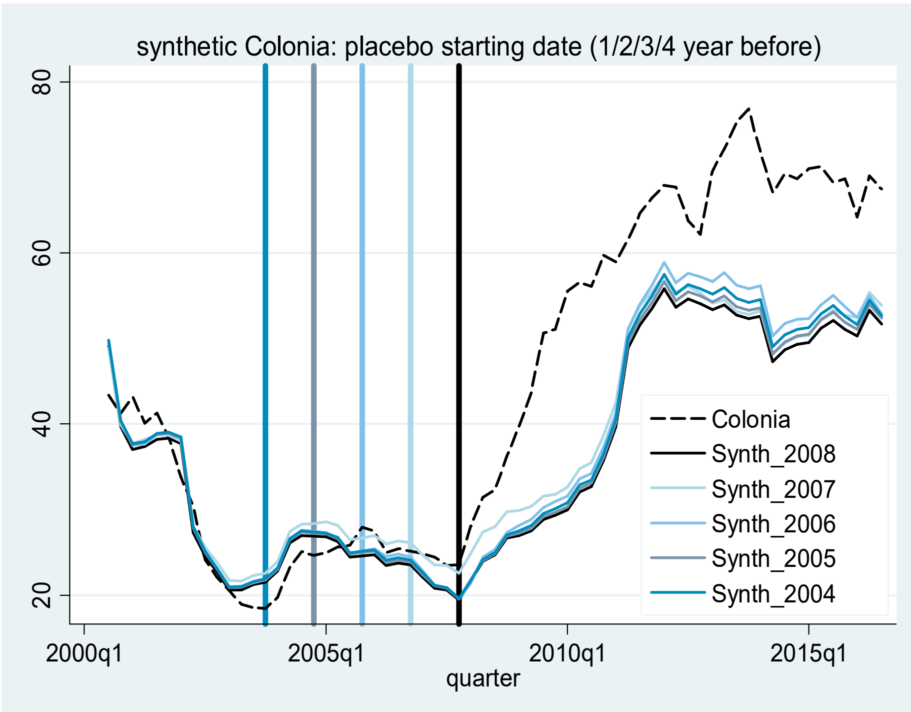
Rubustness

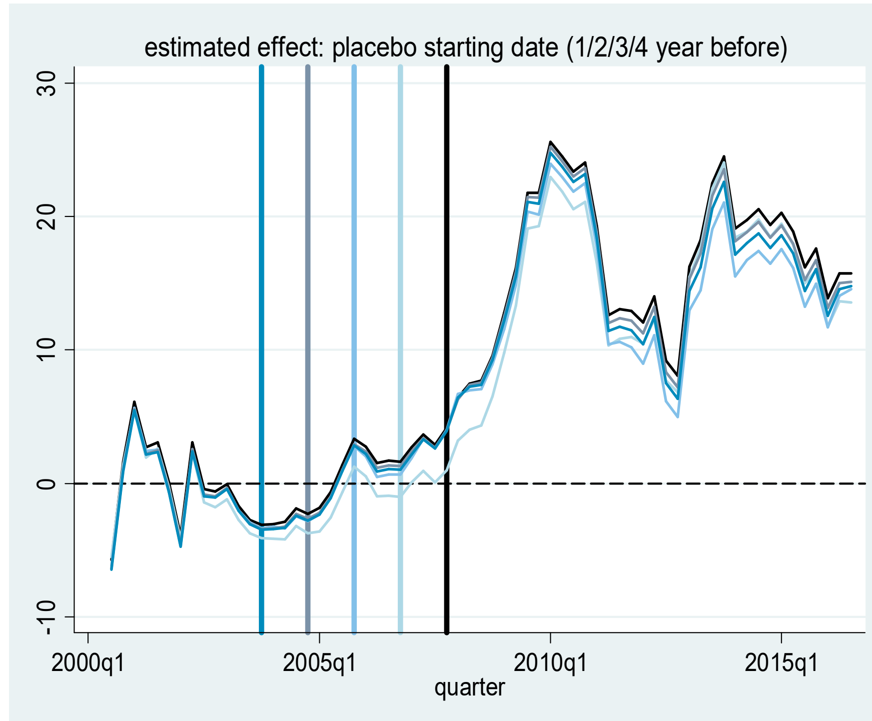
Table 4. Robustness of the significance of the impact to the exclusion of regions from donor group

Excluding from donors:	p-value ($\overrightarrow{s}_j/\overleftarrow{s}_j$)
Costa de Oro	0.2
Rocha	0.0
Litoral	0.0
Costa de Oro, Rocha	0.0
Costa de Oro, Litoral	0.3
Rocha, Litoral	0.0
Costa de Oro, Rocha, Litoral	0.0

Rubustness

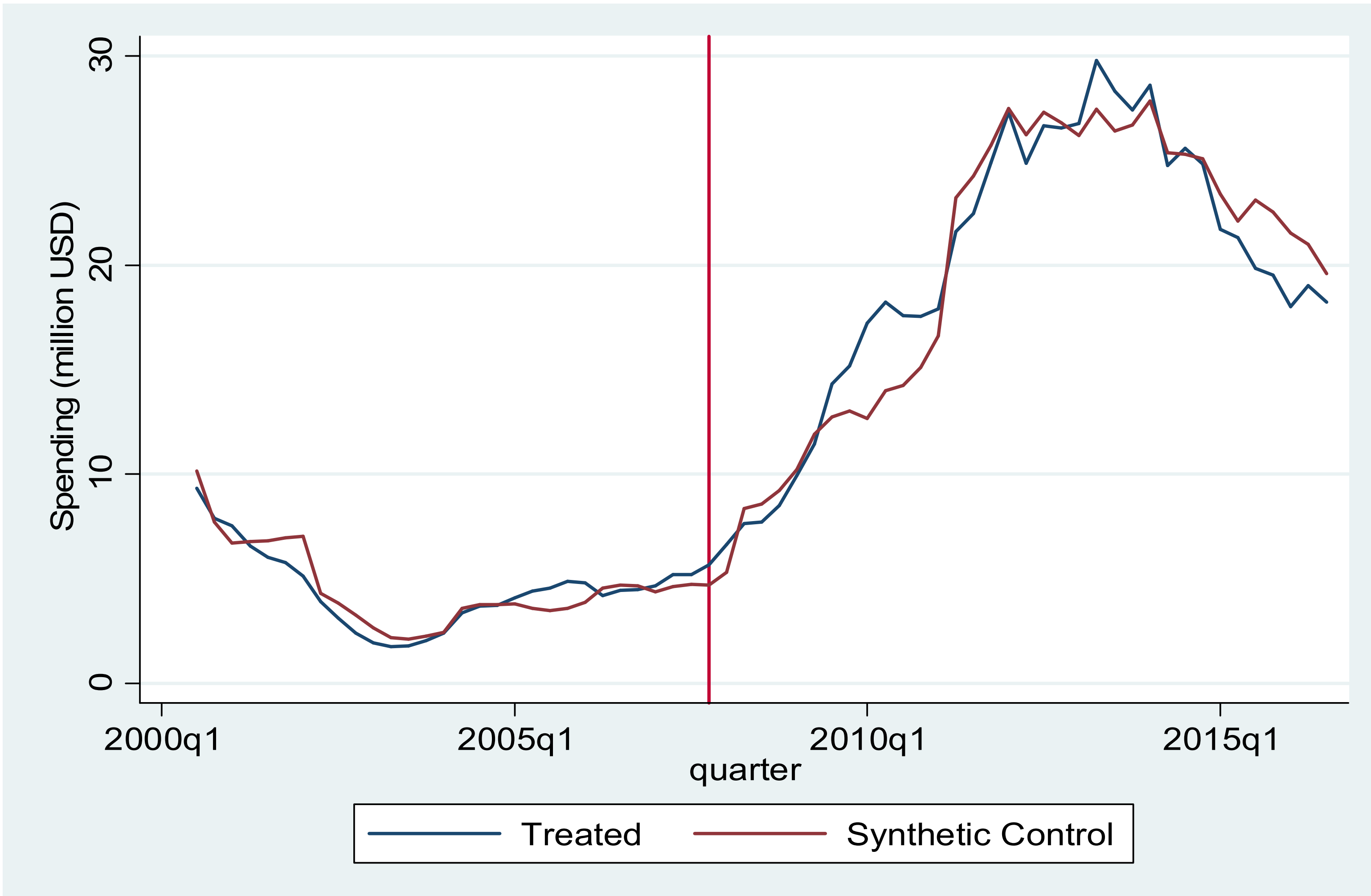
Table 4. Robustness of the impact to the starting date





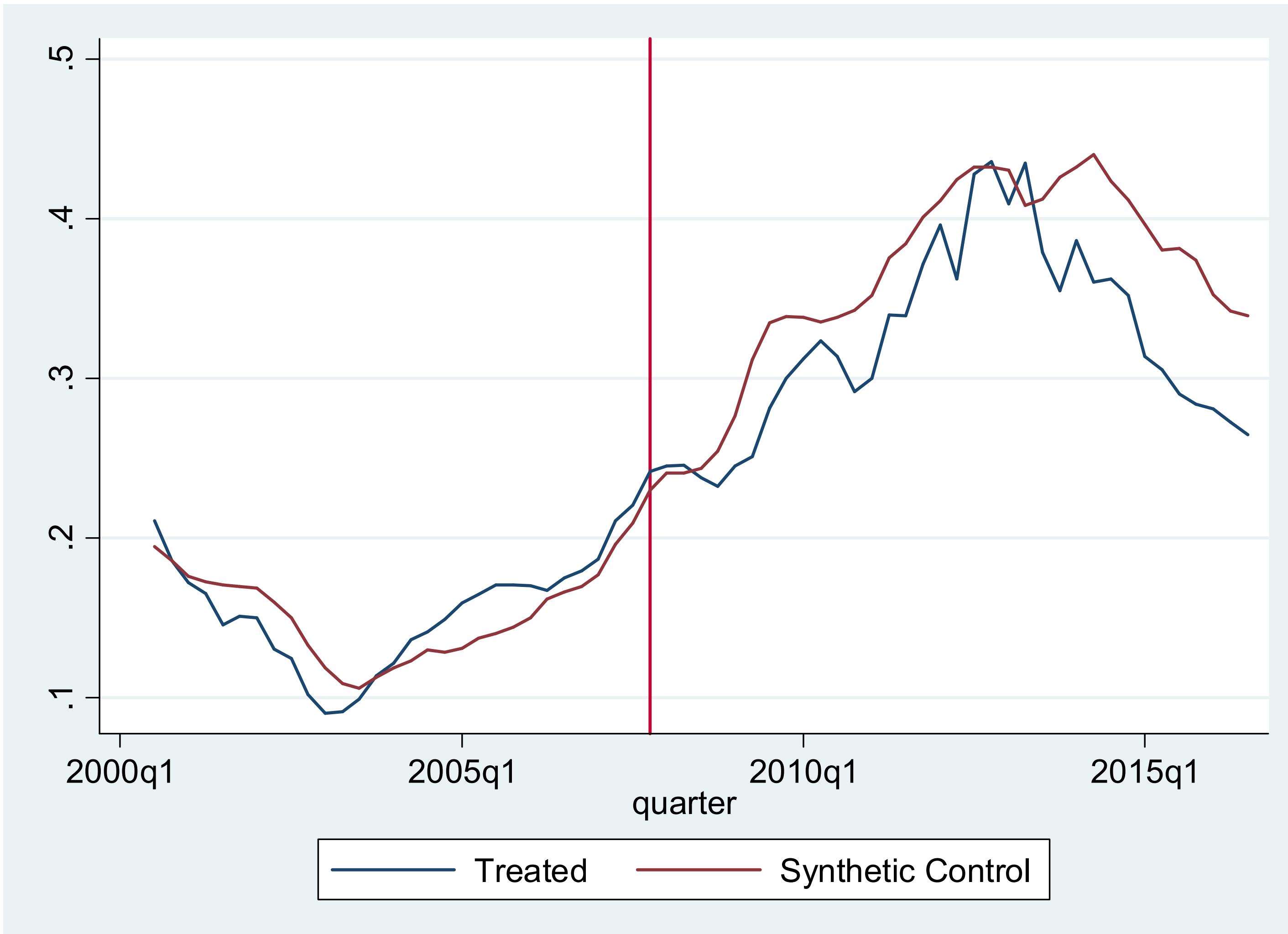
Results: Total expenditure

Table 4. Colonia vs. Synthetic Colonia



Results: expenditure per tourist

Table 4. Colonia vs. Synthetic Colonia



Roadmap

1. Motivation
2. Objectives
3. The program
4. Data
5. Methodology
6. Results
7. Conclusions

Conclusions

- **Limitations: the pool of donors is small.**
- **Positive impact of the cluster program on the inflow of international tourists to Colonia.**
- The estimated impact was of 14 thousands tourists per quarter between 2008 and 2015, which represent a **24% increase in the number of tourists in the period**.
- In addition, **we did not find a significant impact on the total expenditure**.
- This **could be explained by a composition effect in the total number of tourists arriving to Colonia?**
- Probably the incremental number of tourists was concentrated in segments of lower relative income.
- Or alternatively, that due to the border mobility and foreign exchange restrictions in Argentina, there was a negative effect on the expenditure per tourist (less days of stay and/or fewer resources spent).

$\bar{s}_j/\bar{s}_j$

Thank you for your time!

Synthetic Control Methods

- Following Abadie et al. (2010) we define D_{jt} as the indicator of treatment for region j at moment t. The observed outcome variable Y_{jt} equals the sum of the effect of the treatment ($\alpha_{jt}D_{jt}$) and the counterfactual Y^N which is specified as a factor model:

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it}.$$

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it},$$

- Because only the first region (i=1) is exposed to intervention and only after period T_0, we have that:

$$D_{it} = \begin{cases} 1 & \text{if } i = 1 \text{ and } t > T_0, \\ 0 & \text{otherwise.} \end{cases}$$

Synthetic Control Methods

- We want to estimate $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$. For $t > T_0$,

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N.$$

- But we just need to estimate the unobserved counterfactual Y_{1t}^N
- If there are $(w_2^*, \ldots, w_{J+1}^*)$ such that:

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \ \ldots \ , \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* Z_j = Z_1. \quad (2)$$

- Under standard condition $Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ will be close to zero if the number of pre-intervention periods is large relative to the scale of the transitory shocks. Then

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

Synthetic Control Methods (estimation)

- So, choosing a syntethic control which can fit Z_1 and a set of pre-intervention outcomes (Y_{11}, Y_{12},…, Y_{1T0}), we are able to obtain an estimate for the counterfactual whose bias can be bounded by a function that goes to zero as the number of pre-treatment periods increases

- Let "predictors" X comprised of Z and the set of pre-intervention outcomes

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\|_{\mathbf{V}}$$

- W* is chosen to minimize the distance:

- V is a matrix of predictor weights that prioritizes which variables to match better.