

Final Project

Khoi Pham, Quang Tran

CS112 Spring 2019

Decision Memorandum

To: Sebastian Koehler, Thomas König
From: Khoi Pham, Quang Tran
Re: The significance of the effect of Stability and Growth Pact on euro countries

Executive Summary

A study on the impact of Stability and Growth Pact (SGP) on the euro countries by Koehler and König claims that “the SGP effectively constrained overall government debt making” after showing a negative effect of the introduction of euro in these countries on debt/GDP ratio. The authors conducted in-time and in-space placebo studies as supplementary tests for the significance of the effect, and they interpret the results as in support of the effect found. However, our re-implementation of the in-space placebo test and our re-interpretation of the in-time placebo results show that the effect of euro and SGP¹ estimated for the eurozone is **not** statistically significant. We recommend redoing the in-space placebo test with more control countries for a more convincing and conclusive result.

Background

Stability and Growth Pact (SGP) is a fiscal mechanism intended to limit government debt and deficits for all European Union (EU) members. The two most important regulations of SGP took effect in the same year as the introduction of the euro. As there had been some doubt on the effectiveness of SGP for the initial 11 euro members, the paper “Fiscal governance in the eurozone: How effectively does the stability and growth pact limit governmental debt in the euro countries?” estimated the effect of SGP/euro on the euro 11. Based on the in-time and in-space placebo tests, the authors maintained the significance of the found effect.

¹ The authors attributed the effect to both the introduction of the euro and SGP as the two occurred at the same time in the year 1999.

Objective

We investigate the correctness of the procedure of the authors' conducting of the placebo tests. We first replicate figure 3 in the paper that shows the negative effect of SGP on debt/GDP ratio in euro 11 countries and then explore the significance of the effect by:

- 1) Computing the p-value using the correct version of the in-space placebo test;
- 2) Discussing the authors' interpretation of the in-time placebo test.

Synthetic Control Study²

Koehler et al. conducted a synthetic control study where the treated unit is euro 11, and the control group includes ten countries that did not introduce the euro currency. The treatment variable is the introduction of the euro in 1999, and the outcome of interest is the population-weighted debt/GDP ratio of euro 11 countries.³

The control group is used to produce one synthetic entity (called the synthetic control) that matches well the euro 11's debt/GDP ratio in the pre-treatment period (1983-1999). The difference between the euro 11's outcome and the extrapolated outcome for the synthetic control is the treatment effect estimate. Below is our replication of figure 3 in the paper that visualizes this method. As the euro 11 curve consistently lies below the synthetic curve, figure 1 shows that the euro 11's debt/GDP ratio would have been higher had the euro not been introduced.

² **#organization:** organized the memorandum with clear structure, order, and headings.

³ **#variables:** described and distinguished between independent variables, dependent variables, and treatment in a study.

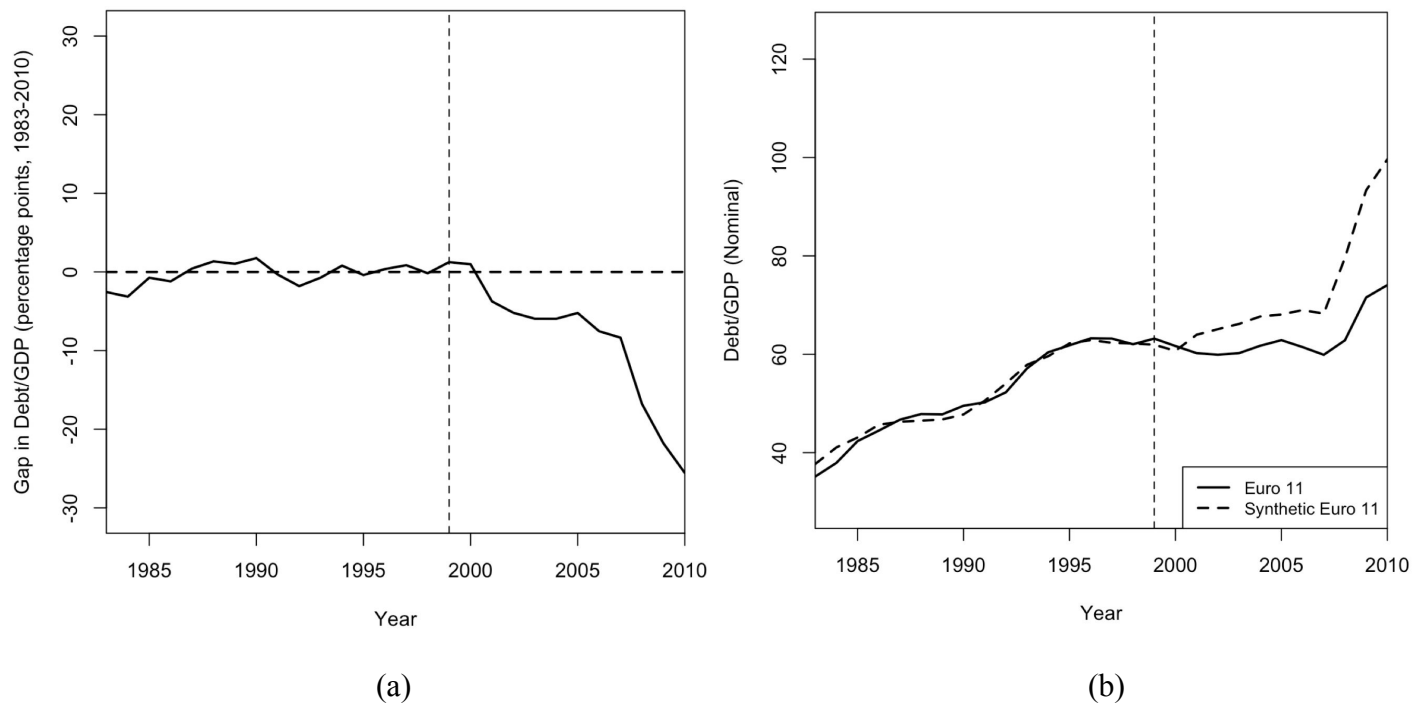


Fig. 1: Figure 3 in Koehler et al. Comparison of the euro 11 group of states and its synthetic counterpart (1983-2011). Data and code for replication found [here](#).

In-time Placebo Test

The paper reports conducting in-time placebo studies to evaluate the credibility of the effect by re-assigning the treatment to alternative times. The results are that for all treatment years from 1988 to 1997, the difference between euro 11 and the synthetic one is positive, and for the year 1998, although negative, the effect is not as obvious as the effect for the real treatment year 1999. The authors concluded that the effect estimated for 1999 is “substantial” because there are “no comparable effect for those placebo studies.” However, **this is not a correct interpretation of the method detailed in Abadie, Diamond, and Hainmueller (2012).** What matters in the technique is not a “comparable” estimate but “large estimate.” Because clear effects were found for years 1988-1997, they actually undermine the estimated effect for 1999 and make it non-substantial.

In-space Placebo Test

Despite the claim, the eurozone paper did not follow Abadie, Diamond, and Hainmueller (2015) in applying the in-space placebo test. The eurozone paper estimates the placebo effects only for five countries, one of which is not even from the control countries considered for creating the synthetic euro 11. Figure 2(a) represents the correct implementation of the test where effects are estimated for *all* ten control countries, using the same synthetic control method.

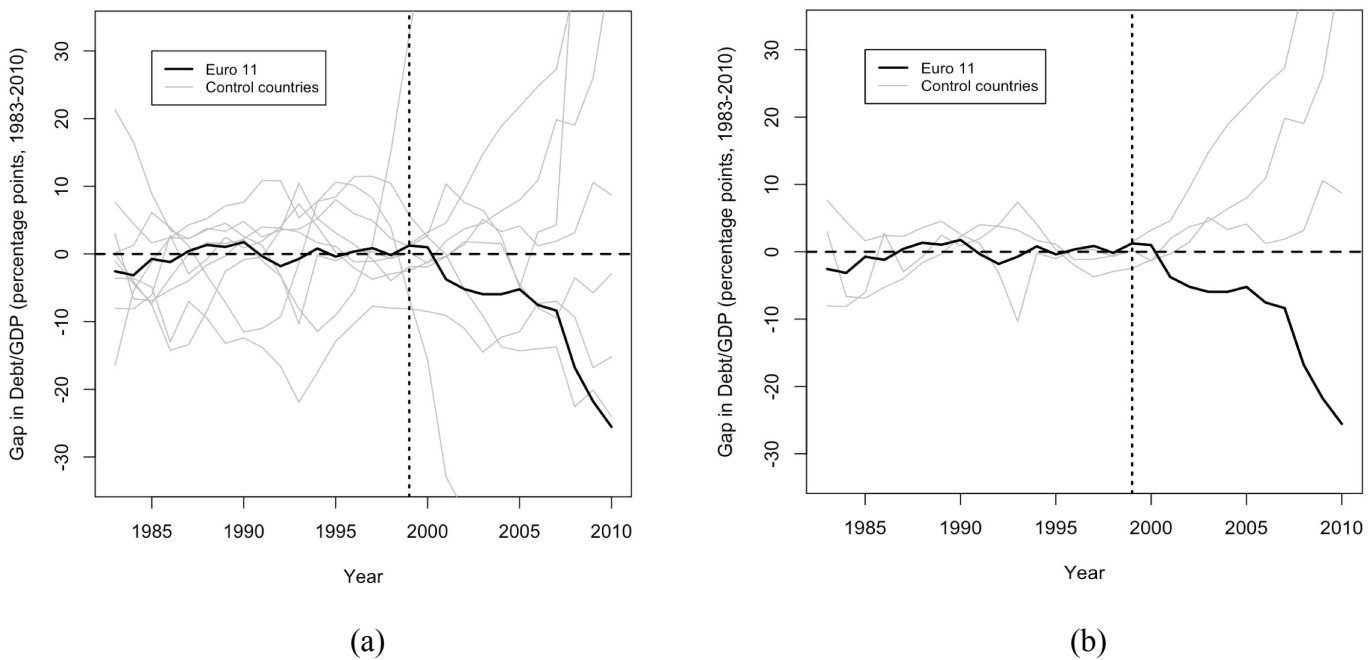


Fig. 2: Debt/GDP (percentage points) gaps in euro 11 group and placebo gaps in (a) all 10 control countries and (b) control countries that have MSPE smaller than 20 times that of euro 11.

Using the in-space placebo test, we can perform a significance test to check which one of the following hypotheses we should reject and which one we should accept:

- H_0 (null hypothesis): Introduction of SGP and euro in 1999 has not had any noticeable effect on the debt-to-GDP ratio of euro 11 countries;

- H_a (alternative hypothesis): Introduction of SGP and euro in 1999 has affected the debt/GDP ratio in a considerable way (specifically, it has decreased the debt/GDP ratio) in euro 11 countries.

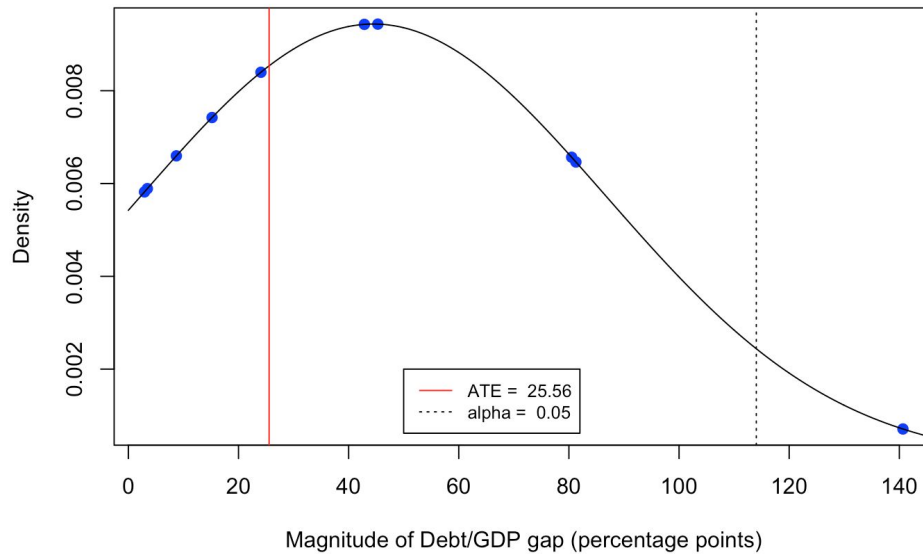
If the magnitude of the euro 11 gap is larger than most of its control counterparts, the negative effect found for euro 11 is significant. However, this is not the case, as the percentage of gaps of magnitude larger than or equal to the euro 11 group's gap in 2010 is $p = 6/11 = 0.545$, far larger than the significance level 0.05, indicating the non-significant effect of SGP estimated for euro 11.

Figure 2(b) is another variation of figure 2(a), where all control countries with mean square prediction errors (MSPE) larger than 20 times that of the euro 11 are excluded. This approach to discarding placebo runs that provide irrelevant information was also introduced in Abadie et al. (2015). The p-value is still really large (0.75) in this case.

Maximum Likelihood Estimate Distribution

Because there are only ten countries in the control group, the lowest significance level the effect for euro 11 can get is $1/11 \approx 0.09$, which is surely not statistically significant. Doing a significance test like above would not be a fair test in case there is indeed a significant result. We address this by additionally fitting the maximum likelihood estimate (MLE) to the distribution of the magnitudes of the gaps and see where the effect for euro 11 lies. MLE allows us to approximate the underlying trend of the control units' differences in outcomes by maximizing the probability of observed magnitudes given a proposed estimate distribution. Using such a distribution, we can approximate how the population of control placebos would look like and get a more representative p-value for the treatment effect for the treated. Figure 3 provides the fitted normal distributions for the two variations shown in figure 2.

(a)



(b)

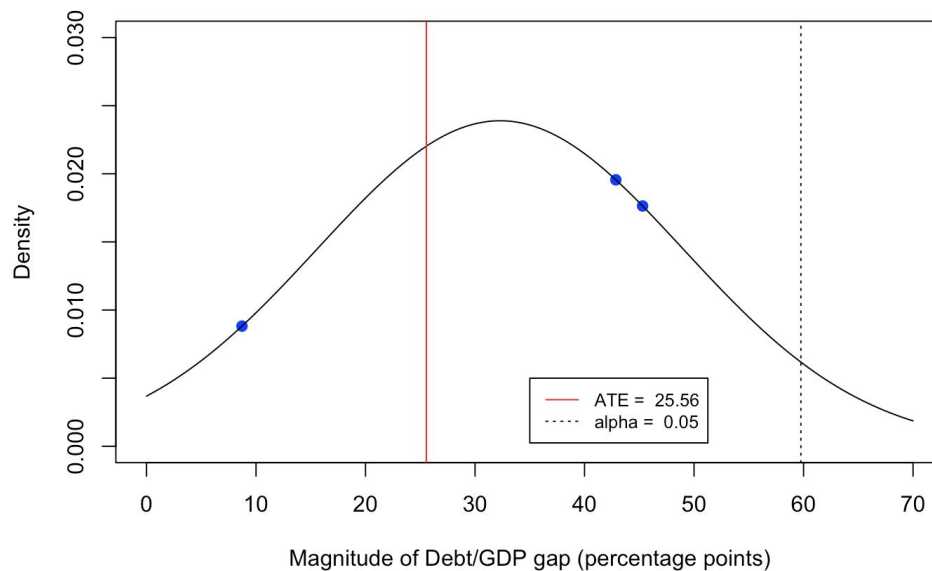


Fig. 3: Distribution of magnitudes of differences in outcomes (blue dots) in 2010 of the control units fitted with maximum likelihood estimation (MLE). Red line shows the magnitude of the estimated average treatment effect (ATE) for the “treated” euro 11 countries. Dotted line indicates the one-tailed threshold for significance test. (a) and (b) is the distribution corresponding to fig. 2(a) and (b), respectively.⁴

⁴ **#dataviz:** visualized distributions with effective use of vertical lines to indicate critical points on the x-axis; clear legend and visualization description are attached.

The p-values obtained from the maximum likelihood estimate (MLE) distribution are $p = 0.673$ and $p = 0.657$, respectively. These are in line with the results from vanilla in-space placebo tests shown in figure 2, indicating that the effect of euro introduction is **not** statistically significant.⁵ From the p-value test, we have to reject the alternative hypothesis and accept the null hypothesis which states that the introduction of SGP and euro in 1999 has not had an impact on the debt-to-GDP ratio.

Link to the R code

<https://github.com/Munchic/euro11-replication/blob/master/euro11.R>

Conclusions

The paper misinterpreted the results from the in-time placebo test and misapplied the in-space placebo study approach introduced in Abadie et al. (2012).⁶ Our re-evaluation shows that the estimate for the effect of the introduction of euro and the SGP is not substantial like what the interpretation in the paper shows. It is worth noting, however, that this insignificance should be treated as a preliminary result because the small number of control countries has necessitated our MLE distributions. Ideally, we would want a reliable p-value right from the regular in-space placebo test. This shows that the in-space placebo test and its MSPE-truncated version would be more informative and its results would be more convincing if the authors could use a larger pool of control countries drawn from a larger, more complete data set.

⁵ **#significance:** performed significance test using Fisher's exact test in the in-space placebo test as well as using maximum likelihood estimation to account for a small sample size.

⁶ **#observationalstudy:** critiqued the synthetic controls method used in the original study on observational data and proposed replication that, given the data, shows the lack of originally proclaimed statistical significance.

References

- Abadie, A., Diamond, A., Hainmueller, J., & National Bureau of Economic Research. (2012). *Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program* (Nber working paper series, no. 12831). Cambridge, Mass.: National Bureau of Economic Research. (2007). Retrieved April 21, 2019, from <https://economics.mit.edu/files/11859>
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and synthetic control method. *American Journal of Political Science*, 59(2), 495-510. Retrieved March 10, 2016, from https://www.researchgate.net/publication/228322749_Comparative_Politics_and_the_Synthetic_Control_Method/download
- Koehler, S., & König, T. (2015). Fiscal governance in the eurozone: How effectively does the stability and growth pact limit governmental debt in the euro countries? *Political Science Research and Methods*, 3(02), 329-351. doi:10.1017/psrm.2014.26
- Koehler, S., & König, T. (2014). Replication data for: Fiscal Governance in the Eurozone: How effectively does the Stability and Growth Pact limit governmental debt in the Euro countries?. <https://doi.org/10.7910/DVN/24640>, Harvard Dataverse, V1