EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Pasquale Marco Marrazzo, Alessio Terzi

Structural reform waves and economic growth

No 2111 / November 2017

Abstract

At a time of slow growth in several advanced and emerging countries, calls for more structural reforms are multiplying. However, estimations of the short- and medium-term impact of these reforms on GDP growth remain methodologically problematic and still highly controversial. We contribute to this literature by making a novel use of the non-parametric Synthetic Control Method to estimate the impact of 23 wide-reaching structural reform packages (including both real and financial sector measures) rolled out in 22 countries between 1961 and 2000. Our results suggest that, on average, reforms started having a significant positive effect on GDP per capita only after five years. Ten years after the beginning of a reform wave, GDP per capita was roughly 6 percentage points higher than the synthetic counterfactual scenario. However, average point estimates mask a large heterogeneity of outcomes. Benefits tended to materialise earlier, but overall to be more limited, in advanced economies than in emerging markets. These results are confirmed when we use a parametric dynamic panel fixed effect model to control for the rich dynamics of GDP, and are robust to a variety of alternative specifications, placebo and falsification tests, and to different indicators of reform.

JEL Classifications: E65, O11, O43, O47

Key words: structural reforms, economic growth, synthetic control

I. **Non-technical summary**

In this paper, we investigate the relationship between a wide range of structural reforms and economic performance over a ten-year time horizon. Our novel empirical approach is composed of two steps. First, building on a methodology developed by IMF (2015), we identify 23 episodes of wide-reaching structural reform implementation (so-called "reform waves"). These are based on a database first assembled by Ostry et al. (2009) and later expanded by Giuliano et al. (2013), which provides detailed information on both real and financial sector reforms in 156 advanced and developing countries over a 40 year period. Indicators considered specifically cover trade-, product market-, agriculture-, and capital-account liberalisation, together with financial and banking sector reform. Then, we track top-reforming countries over the 10 years following adoption and estimate the dynamic impact of reforms by employing the Synthetic Control Method (SCM): a technique first developed in Abadie and Gardeazabal (2003). In a similar vein to matching estimation strategies, this non-parametric data-driven approach creates a control for each individual reforming country, as a linear combination of other countries displaying similar pre-reform characteristics. The SCM allows us to quantify the individual impact of a country's reform package. Our technique therefore caters for the fact that each country and reform package is different, displaying the heterogeneity thereby associated. However, at the same time, it makes it possible to aggregate results and draw some general conclusions on the relationship between reforms and economic performance.

Our main findings are as follows: on average, reforms had a negative but statistically insignificant impact in the short term. This slowdown seems to be connected to the economic cycle, and the tendency to implement reforms during a downturn, rather than an effect of reforms per se. Reforming countries however experienced a growth acceleration in the medium-term. As a result, ten years after the reform wave started, GDP per capita was roughly 6 percentage points higher than the synthetic counterfactual scenario. To put this into perspective, over the past decade, at a time of so-called "great convergence" (Baldwin 2016), the GDP per capita gap between high- and upper-middle-income countries has shrunk by a comparable amount.

We highlight how average point estimates mask the fact that the impact of reforms is highly heterogeneous, in particular between advanced and emerging markets. Benefits tended to materialise earlier, but to be more limited, in the former than in the latter. We show how this result is unlikely to be dictated by the macroeconomic conditions in which reforms were implemented, but rather suggest that this may be due to a diversified composition of the average reform package and overall quality of institutions, affecting policy credibility. Moreover, having a shorter horizon over which reforms pay out could also be increasing the political return to reforms, hence making their systematic adoption more likely in advanced economies, and therefore contributing to explaining the lack of long-term convergence observed in the past. This interpretation of the results lends new credibility to the channels underlying the relation between quality of institutions and long-term growth.

II. Introduction

Policymakers face a perennial struggle to improve economic outcomes, in both advanced and developing countries. Within this context, orthodox economic theory has made a strong case for structural reforms, identified as measures aimed at removing supply-side constraints in an economy. This in turn would favour efficient factor allocation and contribute to medium- to long-term growth. Such measures include, but are not limited to, product and labour market liberalisations, current and capital account openness, and financial liberalisation. For a long time, a collection of these policies has fallen under the name of *Washington Consensus*, following their listing in Williamson (1994). According to this *policy perspective*, which has long been held by multilateral organisations, the depth and breadth of reform packages will determine subsequent economic performance (Easterly 2005).

While individually these measures are built on solid workhorse economic models, and well-established concepts such as Schumpeterian creative destruction or competitive advantage, the growth literature remains divided. The *diagnostics perspective* holds that deep- or wide- orthodox structural reform packages will generally fail to produce positive growth effects, unless they are carefully tailored to the local context (Rodrik 2004). The *endowment perspective* identifies geography and resource endowment as crucially underlying the development of agriculture, human capital, transport costs, the spread on knowledge and knowhow, and therefore there is little policy change can do to significantly shape medium- and long-term growth (Sachs 2001; Diamond 1997). Finally, the *institutional perspective* eclectically builds on the previous two by placing institutions at the centre. Good economic institutions – generally involving the protection of private property and the rule of law – combined with good political institutions – defined as pluralistic and power-constrained – ensure that the right policies will be identified and successfully implemented (Hall & Jones 1999; Acemoglu et al. 2001; Acemoglu & Robinson 2012). At the same time, geography will have a strong impact on the quality of institutions (Rodrik et al. 2004). According to this view, pushing for structural reforms will have only a temporary and limited impact on economic growth as it will be like treating the symptoms of the problem, rather than addressing the root cause.

Motivated by the apparently conflicting prediction on the benefits of (orthodox) policy reform in these growth theories, we investigate the relationship between a wide range of structural reforms and economic performance over a ten-year time horizon. Our novel empirical approach is composed of two steps. First, building on a methodology developed by IMF (2015), we identify 23 episodes of wide-reaching structural reform implementation (so-called "reform waves"). These are based on a database first assembled by Ostry et al. (2009) and later expanded by Giuliano et al. (2013), which provides detailed information on both real and financial sector reforms in 156 advanced and developing countries over a 40 year period. Indicators considered specifically cover trade-, product market-, agriculture-, and capital-account liberalisation, together with financial and banking sector reform. Then, we track top-reforming countries over the 10 years following adoption and estimate the dynamic impact of reforms by employing the Synthetic Control Method (SCM): a technique first developed in Abadie and Gardeazabal (2003) and later refined in Abadie et al. (2010; 2015). In a similar vein to matching estimation strategies (Rosenbaum & Rubin 1983), this non-parametric data-driven approach creates a control for each individual reforming country, as a linear combination of other countries displaying similar pre-reform characteristics.

Crucially, Abadie et al. (2010) prove that once a good match has been established over observable characteristics prior to the reform, time-varying unobservable confounders become a second-order problem. In a SCM framework, we are therefore somewhat less concerned about the endogeneity of reforms, which is instead a persistent issue in standard panel models such as difference-in-difference or fixed-effect estimations.

The SCM allows us to quantify the individual impact of a country's reform package. Our technique therefore caters for the fact that each country and reform package is different, displaying the heterogeneity thereby associated. However, at the same time, it makes it possible to aggregate results and drawing some general conclusions on the relationship between reforms and economic performance. As discussed by Billmeier and Nannicini (2013), the SCM creates data-driven case studies that can be analysed within a unified statistical framework, hence effectively constituting a middle-ground between a case study approach and standard cross-sectional work.

Our main findings are as follows: on average, reforms had a negative but statistically insignificant impact in the short term. However, reforming countries experienced a growth acceleration in the medium-term, with the result that ten years after the reform wave started, GDP per capita was roughly 6 percentage points higher than the synthetic counterfactual scenario. We highlight how average point estimates mask the fact that the impact of reforms is highly heterogeneous, in particular between advanced and emerging markets. Benefits tended to materialise earlier, but to be more limited, in the former than in the latter. We show how this result is unlikely to be dictated by the macroeconomic conditions in which reforms were implemented, but rather suggest that this may be due to a diversified composition of the average reform package and overall quality of institutions, affecting policy credibility.

Given the novelty of our approach in the reform literature, we show that our findings are not model-specific. In particular, we adapt the dynamic panel regression model used by Acemoglu et al (2014) to provide an alternative estimation strategy for the impact of reforms on growth, and illustrate how our baseline estimates are at most to be treated as conservative.

Moreover, we performed extended robustness checks on our main results. These include a placebo test, where we fictitiously placed a reform episode ten years before the actual reform wave, and show how the impact measured by the SCM is not comparable to that identified in the baseline. We also doubled the time horizon over which we fit the model to twenty years, reducing the likelihood of positive self-selection bias. Moreover, we ran a falsification test on countries that did not implement reforms, showing how our model tracks well performance on average in the absence of a reform shock. Finally, we considered alternative indicators of comparable structural reforms, and hence an alternative selection of wide-reaching reform episodes. In all cases, our baseline results remain broadly confirmed.

The remainder of the paper is organised as follows. Section II provides a detailed account of the existing literature on the impact of reforms on growth. Section III explains the theory underpinning the SCM. In Section IV, we describe the data and explain the methodology behind our case selection. Our implementation of the SCM is presented in Section V. Section VI reports the baseline results. In Section VII, we crosscheck our baseline estimates with an alternative parametric estimation strategy. Section VIII presents a range of robustness checks, while Section IX is devoted to discussing two specific case studies. Section X offers some concluding remarks.

III. Literature on reforms and growth

The literature tracing the impact of individual policy reforms on growth in extremely abundant. We do not attempt to summarise it here exhaustively, but rather report some of the findings related to the policies we will consider later in the paper. While we consider only reform packages including multiple reform dimensions, hence making it impossible to quantify the impact of a single policy measure within our setting, it is interesting to note that the empirical literature remains highly divided even on the benefits of individual policy dimensions.

For what regards trade liberalisation, the evidence suggesting a positive effect on growth dates back to the 1990s (Frankel & Romer 1999; Dollar 1992; Sachs & Warner 1995; Dollar & Kraay 2004; Wacziarg & Welch 2008). However, these results have not gone unchallenged (Rodriguez & Rodrik 2001; DeJong & Ripoll 2006). Interestingly, Billmeier and Nannicini (2013) recently employed the SCM in a worldwide sample of countries to establish that liberalisations yielded a positive effect in all regions, but more recent episodes, especially in Africa, had no significant impact.

Turning to domestic financial liberalisation, Levine (1997; 2005) provides a comprehensive overview of the literature, showing how – although with some nuances – the preponderance of evidence suggests a positive link to growth (Rajan & Zingales 1998; Beck et al. 2000; Aghion, Howitt, et al. 2005; Galindo et al. 2007; Abiad et al. 2008). Dervis and Page (1984) argued in favour of product market reforms and policies promoting competition, for which several authors find positive effects, particularly in developed economies (Nicoletti & Scarpetta 2003; Frischtak et al. 1989; Barro 1991; Easterly 1993; Jalilian et al. 2007). Rodrik (1995) however preserves a degree of scepticism: while the theoretical channels are clear, the empirical aggregate impact of these industrial policies seems rather small. Detailing the complexity of the relationship between competition and innovation, Aghion et al. (2005) cast doubts even over the theoretical channels underpinning such policies.

Gollin (2010) offers an overview of the relationship between agricultural reform and long-term growth, offering a mixed picture (see also Rodrik 1995). Beyond the direct impact of these policies on agricultural productivity, the main indirect theoretical channel relates to the fact that agricultural reforms tend to decrease inequality, which in turn the political economy literature has identified as a drag on growth (Alesina & Rodrik 1994; Besley & Burgess 2000). Aside from Anderson (2010), this empirical literature has generally focussed on specific countries, finding strong positive effects in China (Lin 1992; Gulati et al. 2005) and Vietnam (Pingali & Xuan 1992). The evidence is however much thinner for former Soviet countries, which saw collapsing output and productivity (Rozelle & Swinnen 2004; Macours & Swinnen 2000).

Perhaps the most contentious reform among those considered is capital account liberalisation (Eichengreen 2001). While several authors find positive effects on growth (Bekaert et al. 2005; Kose et al. 2009; Dell'Ariccia et al. 2007; Quinn & Toyoda 2008; Quinn 1997), these results tend to be rather weak (Rodrik 1998). Klein and Olivei (2008) and Edwards (2001) suggest that these unsatisfactory empirical findings, which clash with the strong theoretical case in favour of financial globalisation, are driven by developing countries. A point later echoed by Rodrik and Subramanian (2009) and Prasad et al. (2007).

Turning to the broader relationship between policy reforms and economic performance, the body of academic empirical literature we contribute to is perhaps more limited than the importance of the question would warrant. The overall conclusion that can be drawn is that reforms are generally associated with positive subsequent economic performance, but the data displays a high degree of heterogeneity, depending on countries and policies considered. A finding that we will find confirmed within our own empirical setting.

Post-Soviet countries moving towards a market economy have received considerable attention in this respect. Fischer et al. (1996) looked at 26 transition economies over the period 1989-1994. They conclude that structural reforms played a vital role in reviving economic growth. This finding for transition economies was echoed by de Melo et al. (1996), and more recently by Havrylyshyn and van Rooden (2003) and Eicher and Schreiber (2010).

Focussing more broadly on countries implementing wide reform packages covering domestic finance, trade, and the capital account, Christiansen et al. (2013) find a strong impact of the former two on growth in middle-income countries. Moreover, they show how well-developed property rights are a pre-condition in order to reap fully the benefits of structural reforms. The importance of institutions in explaining cross-country heterogeneity is further remarked by Prati et al. (2013), who illustrate how the positive relationship between structural reforms and growth depends on a country's constraints on the authority of the executive power. Distance from the technological frontier seems also to play a role. In line with the spirit and methodology of our paper, Adhikari et al. (2016) recently applied the SCM to six cases of reform waves in advanced economies. Overall, they find evidence suggesting a positive but heterogeneous effect of labour and product market reforms on GDP per capita.

However, as mentioned in the introduction, there are some dissenting opinions among scholars. In line with the *diagnostic* view, Levine and Renelt (1992) show how policies associated with long-term growth have important interaction effects and, when these are taken into account, the strong predictions on their individual positive impact often become fragile or insignificant. Similarly questioning past methodologies, Rodrik (2012) argues there is little to be learned from regressing policies on growth. Easterly (2001) documents how developing countries experienced a growth slowdown over the period 1980-1998, while their extensive reforms would have predicted a growth acceleration. Linking the argument to the *institutional perspective*, Easterly and Levine (2003) show how a broad set of macroeconomic policies becomes irrelevant in explaining economic development once institutions, and the geographical factors that underpin them, are taken into account. A point echoed by Rodrik et al. (2004). Easterly (2005) subsequently shows how the positive relationship established in the literature between national policies and growth is likely to be driven by extreme observations.

IV. Econometric theory

The empirical study presented in this paper is based on the Synthetic Control Method as introduced by Abadie and Gardeazabal (2003), and later extended by Abadie et al (2010; 2015). The SCM proposes an innovative way to address one of the main issues in comparative case studies, namely finding a suitable counterfactual to a treated unit, or within our setting, to a country. The main idea is to use several countries to synthetize a control that resembles as much as possible the "treated" country. In practice, the synthetic replica country is a linear combination of several possible comparable countries that deliver a good match to the country of interest[3].

In a way, this technique is not far away from the widely accepted difference-in-difference estimation. In that framework, studies look for a country that is "similar enough" to the country experiencing a policy change before the treatment and then look at the post-reform difference under an assumption of parallel trends. The SCM works in the same way, with the difference that it creates a linear combination of "similar" countries using transparent weights to produce, year-by-year, the best possible pre-reform match.

Formalisation of the Synthetic Control Method

Abadie et al (2010) formally show how to identify the effect of an intervention (in this paper, a reform wave) by mean of a panel dataset and a factor model. More precisely, suppose we start with a panel dataset collecting the GDP per capita of $J+1$ countries and T periods, where one country $i=1$ exhibits a reform wave at time T_0, while J countries do not exhibit any reform wave during the T periods. Moreover, assume that the GDP per capita of any country i at any time t is given by a factor model:

$$Y_{it} = \delta_t + \boldsymbol{\lambda}_t \boldsymbol{\mu}_i + \boldsymbol{\theta}_t \boldsymbol{Z}_i + D_{it}\alpha_{it} + \varepsilon_{it} \qquad [1]$$

where:

- δ_t is an unobservable common factor with unitary loading to all countries i
- $\boldsymbol{\lambda}_t$ is a vector of unobservable common factors with country specific loading $\boldsymbol{\mu}_i$
- $\boldsymbol{Z}_i$ is a vector of observed covariates uncorrelated with the reform wave
- $\boldsymbol{\theta}_t$ is a vector of unknown parameters for the covariates
- D_{it} is an indicator that takes a value of one for $t > T_0$ and $i = 1$

[3] In the finance literature, the SCM could be described as a portfolio strategy to replicate one specific asset.

- α_{it} is the effect of the reform wave
- ε_{it} is an unobservable transitionary shocks with zero mean.

Suppose now to take a vector of J non-negative weights $\boldsymbol{W}$ summing to one, which effectively define a specific synthetic control. Moreover assume there exists a vector of weights $\boldsymbol{W}_i^*$ such that:

$$Y_{it}^{treated} = \boldsymbol{W}_i^* \, \boldsymbol{Y}_{it}^{control} \text{ for } t < T_0$$
$$\text{and} \qquad [2]$$
$$\boldsymbol{Z}_i^{treated} = \boldsymbol{W}_i^* \, \boldsymbol{Z}_i^{control}$$

where $Y_{it}^{treated}$ and $\boldsymbol{Z}_i^{treated}$ are GDP per capita at time t and a vector of covariates of the country of interest i, respectively. $\boldsymbol{Y}_{it}^{control}$ and $\boldsymbol{Z}_i^{control}$ represent instead a vector of GDP per capita at time t and a matrix of covariates of the J countries that act as control group for the country i.

In absence of non-transitionary shocks, Abadie et al (2010) shows that if the pre-treatment period is large relative to the scale of the transitory shocks, it is possible to estimate the effect of the reform wave α_{it} at a specific time $t > T_0$ as:

$$\hat{\alpha}_{it} = Y_{it}^{treated} - \boldsymbol{W}_i^* \, \boldsymbol{Y}_{it}^{control} \qquad [3]$$

The estimation of the synthetic control is purely nonparametric because, as described in Abadie et al (2010), it is based on the minimisation of a pseudo-distance between the vector of characteristics of the country that experiences a reform wave and the vector of the potential synthetic control[4]. One of the main benefits of the synthetic control method is that, conversely from standard case studies, the selection of the control country is data driven, while other approaches have a strong arbitrary element. Moreover, also the relative weight of each covariate is data-driven. This makes the SCM overall a very transparent approach.

It is interesting to notice that the SCM can be seen as a specific case of a linear regression. In fact, the SCM uses a weighting linear combination, where the coefficients are non-negative and sum to one, while a regression uses a linear combination that can be normalised to sum to one but that is not restricted to non-negative coefficient. This restriction comes with two main benefits: first, and foremost, it allows

[4] From a practical standpoint, given the dual optimisation process, it appears useful to start the estimation of the parameters from many points to reduce the chance of running into local, rather than global, minima. We hence start the optimisation process from six thousand different points and select the one delivering the best overall pre-reform fit.

an intuitive and transparent interpretation of the synthetic as a weighted average of the countries composing the control group; second, it avoids the risk of running in an extrapolation bias, as proved by Abadie et al (2015).

On top of this, there are several reasons that lead us to prefer the SCM with respect to a standard panel regression analysis within this setting. Aside from issues of transparency, due to its non-parametric nature, the SCM is less subject to issues of misspecification, also allowing us to model more closely the yearly impact of reform. Moreover, it is less prone to endogeneity bias originating from time-varying unobservable confounders. This is an improvement vis-à-vis difference-in-difference approaches and time fixed effect estimations, which only account for time-invariant confounders (Billmeier & Nannicini 2013). Finally, what is usually considered a disadvantage of the SCM is that it does not allow for standard statistical inference. We will show however that within our setting, this problem can be overcome.

Our methodological contribution

The SCM was originally devised as a single-country data-driven case study method. Within such setting, an underlying crucial assumption is that after T_0 there are no significant idiosyncratic non-transitionary shocks affecting GDP per capita to neither the country of interest nor any of the countries composing the synthetic control. As this assumption cannot be tested, it is treated loosely speaking as the exclusion restriction in an Instrumental Variable (IV) setting, meaning that the author goes qualitatively at length explaining why there are reasons to believe it is satisfied[5]. In a single country case study, thanks also to the transparency in the construction of the control, this argument is made easier.
We see obvious shortcomings in this discretionary case-by-case approach and suggest an alternative way to deal with this problem. What we propose is to apply the SCM systematically to multiple reform episodes n, and study the average effect across all n episodes, therefore assuming that idiosyncratic (i.e. country-specific) shocks are broadly symmetric, which would ensure the asymptotic consistency of our estimate. As n increases, the average effect converges towards an unbiased estimate of the average impact of a reform wave. More formally, given that condition [2] only holds approximately in most real application, then the estimate of the individual treatment effect [3] might be affected by a zero mean idiosyncratic shock ε_{it}, turning equation [3] into:

[5] Some basic robustness tests are routinely carried out to help this claim, as for example excluding countries composing the synthetic control one by one and verifying whether baseline results broadly hold (Mideksa 2013; Abadie et al. 2015).

$$\hat{\alpha}_{it} + \varepsilon_{it}^{treated} - \boldsymbol{W}^{*}\boldsymbol{\varepsilon}_{it}^{control} = Y_{it}^{treated} - \boldsymbol{W}_{i}^{*}\,\boldsymbol{Y}_{it}^{control} \qquad [4]$$

However, the impact of $\varepsilon_{it}^{treated}$ and $\boldsymbol{W}^{*}\boldsymbol{\varepsilon}_{it}^{control}$ on the average treatment effect will tend to zero as n increases, resulting into:

$$\hat{\alpha}_{avg,t} = \frac{1}{n}\sum_{i=1}^{n} Y_{it}^{treated} - \boldsymbol{W}_{i}^{*}\,\boldsymbol{Y}_{it}^{control} \qquad [5]$$

While this remains a theoretical assumption, a Monte Carlo Simulation lends empirical support to our approach, even in a relatively small n setting as ours (see Appendix 1). Moreover, we note that this method is similar in spirit to recent applications of micro-econometric techniques to macro settings, specifically targeted at estimating the average impact of fiscal austerity (Jordà & Taylor 2016), or of democratisation (Acemoglu et al. 2014), on growth. Clearly, while we see standard statistical inference as a main advantage of our approach with respect to past applications of the SCM[6], nothing forbids subsequent zooming into specific reform episodes, therefore offering a finely balanced mix between cross-national econometrics and case studies.

V. Data and Sample

In order to use the SCM, we need first to identify episodes of extensive reform. Our starting point is a panel dataset of structural reforms assembled by Giuliano et al (2013) covering 156 countries between 1960 and 2005, and displaying indicators reflecting the level of regulation in six economic areas: domestic financial sector, capital account, product market, agricultural market, trade, and current account (Table 1).

Table 1. Policy indices

	Domestic financial sector
Securities market	This indicator assesses the quality of the securities market framework, including the existence of an independent regulator and the extent of legal restrictions on the development of domestic bond and equity markets
Banking	This indicator captures reductions or removal of interest rate controls (floors or ceilings), credit controls (directed credit and subsidized lending), competition

[6] We note that Abadie et al (2010) develop a quasi-p-value for a single-country SCM application building on placebo tests.

restrictions (limits on branches and entry barriers in the banking market, including licensing requirements or limits on foreign banks), and public ownership of banks. It also captures a measure of the quality of banking supervision and regulation, including the power and independence of bank supervisors, the adoption of Basel capital standards, and the presence of a framework for bank inspections

Capital account

This indicator aims to measure the extent of the external capital account liberalisation. The index contains information on a broad set of restrictions including, for example, controls on external borrowing between residents and non-residents, as well as approval requirements for foreign direct investment

Product market

This indicator covers the degree of liberalisation in the telecommunication and electricity markets, including the extent of competition in the provision of these services, the presence of an independent regulatory authority, and privatisation

Agriculture market

This indicator measures the extent of public intervention in the market going from total monopoly or monopsony in production, transportation or marketing (i.e., the presence of marketing boards), the presence of administered prices, public ownership of relevant producers or concession requirement to free market

Trade

This indicator is based on tariff liberalization and is measured by average tariffs

Current account

This indicator captures the extent to which a government is compliant with its obligations under the IMF’s Article VIII to free from government restriction the proceeds from international trade in goods and services

Source: This table presents a brief description of the variables. For a more comprehensive treatment, including data sources, refer to Giuliano et al (2013).

Intuitively, we ideally would want to identify points in time when reforms were (i) wide-reaching, (ii) deep, and (iii) not immediately reversed. First, for each of the six reform variables, we want to identify large and stable jumps, which reflect a positive break (improvement) in the specific policy field. To do so, we adapt an approach first detailed by IMF (2015, pp.59–60). Formally, we look at the three-year differences, $\Delta I_t^i = I_{t+3}^i - I_t^i$ where i refers to the country and I is a specific policy indicator. We pool together the three-year differences over the entire time and country sample $\mathrm{A} := \Delta I_t^i \in \mathrm{A}, \ \forall\, i, t$. We then identify breaks $\forall\, I \in [1,6]$ based on three criteria:

(i) the three-year difference ΔI_t^i belongs to the top 3 percentile of the distribution of all ΔI_t^i;

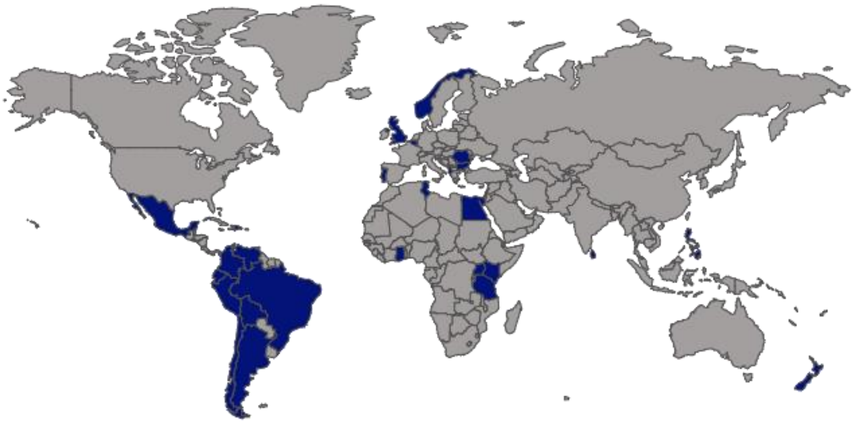
(ii) the three yearly difference that compose ΔI_t^i are all non-negative: $I_{t+1}^i - I_t^i \geq 0$, $I_{t+2}^i - I_{t+1}^i \geq 0$ and $I_{t+3}^i - I_{t+2}^i \geq 0$;

(iii) if condition (i) and (ii) occur more than once consecutively, we consider only the first observation.

We identify reform wave events as years when within a three years interval at least two out of our six variables present a break as defined above. Applying such criteria, we obtain a list of events that consist of 29 episodes distributed over five decades. More precisely, we identify one episode in the 60s, three in the 70s, thirteen in the 80s, eleven in the 90s, and one in the early 2000s. Reforms implemented in these 29 instances cover all six economic areas: Agricultural (8), Product Market (6), Trade (6), Capital account (14), Current Account (17), and Domestic Financial Sector (13). The full list of countries and reforms is displayed in Appendix 2.

The reform waves identified happened in 28 countries, with the only repetition of Argentina in 1974 and 1987. Moreover, as is evident from Figure 1, many of our reform episodes come from Latin America. While this should not surprise, given the region's focus on wide-reaching economic reforms in the 80s and 90s, we note that more than half of the episodes considered in our dataset includes countries located outside Latin America and the Caribbean. As such, we would refrain from considering our results as generated only from reform experiences in this part of the world.

Figure 1. Major reform episodes identified



VI. Methodology in practice: The synthetic control approach

Having identified major reform episodes is only a first step. In order to construct synthetic controls as discussed in section III, we assembled a panel dataset of both real GDP per capita and relevant covariates for 167 countries from all continent and income group over the period 1950-2011. The variables used for the pre-treatment calibration are standard economic growth predictors used in the SCM literature (Abadie et al. 2015; Billmeier & Nannicini 2013). Aside from real per capita GDP in PPP (from here onward referred to as "GDP"), we considered investment rates, the degree of openness of the economy, secondary and tertiary education, population growth, and the value added of industry (full details are reported in Appendix 3).

A necessary condition to be in the donor pool is that for the period of interest the candidate donors should not have experienced a reform wave, both in the pre-treatment and in the post-treatment phase. Within our paper, we will assume the latter to be 10 years following the reform. As such, for example, a country that has experienced a reform wave in 1982 can be a candidate for another country experiencing a reform wave in 2000, but not in 1989.

Extremely small countries are excluded from our analysis because of their limited contribution and high volatility in GDP per capita. For practical purposes, the threshold we apply for the exclusion is 1.5 million people at T_0. Moreover, using the Correlates of War database (Pevehouse et al. 2004), we excluded from the donor pool all countries affected by war in the time period of interest, given that such factor is likely to affect in an extreme (negative) way the donor and be unrelated to the reforming country, hence presenting the risk of distorting our estimates[7].

Not every country's performance can be replicated using the SCM. In some instances, the synthetic control will simply display a limited replication power of the country of interest in the pre-treatment period. In these circumstances, the SCM cannot be used to analyse post-treatment effects, as the parallel trend assumption will be openly violated, as explained in Abadie et al (2010). Therefore, in line with Adhikari et al (2016), we develop a quantitative selection rule to identify whether a reform event should be dropped from our study, based on an arbitrary tolerable error[8]. We standardise the root mean squared prediction error (RMSPE) by GDP per capita at the period before the treatment in the following way:

$$RMSPE_{i,t}^{std} = \frac{RMSPE_{i,t}}{Y_{i,t0-1}} < \gamma$$

where γ is a threshold. Based on the distribution of the $RMSPE_{i,t}^{std}$, we identified a sensible threshold γ = 7% (see Appendix 4)[9]. By standardising over GDP per capita at the year before the reform, we make different event studies comparable in terms of quality of fit. Moreover, this provides an intuitive interpretation to the threshold, i.e. we exclude from the analysis 6 reform episodes in which the average fit error is more than a specific fraction of the level of GDP per capita at the period before the event. In the remainder of the paper, we apply this "best fit" filter, unless mentioned otherwise.

The countries considered in the donor pool of each individual reform episode should be similar to the reforming country, to avoid the risk of running in an interpolation bias, as described in Abadie et al (2010). Intuitively, what this means is that, for example, it would not be desirable to replicate Greece as a weighted mix of Sweden and Angola, which, by averaging, could yield a good pre-reform match, but at the same time are more likely to fail the parallel trend assumption.

A simple way to reduce this risk is to restrict the possible donor pool by income level. An off-the-shelf way to do so would be to use standard income classifications of the World Bank or IMF to construct

[7] In practice, this meant excluding the following countries from selected donor pools: Azerbaijan 1993 – 1994, Cambodia 1977 – 1979, Cyprus 1974, Democratic Republic of the Congo 1975 – 1976, Ethiopia 1998 – 2000, Iraq 1980 – 1988, 1990 – 1991 and 2003, Jordan 1973 and 1991, Rwanda 1994, Sierra Leone 1991 – 2002, South Africa 1975 – 1976, Syrian Arab Republic 1973, Uganda 1978 – 1979, Vietnam 1979.

[8] We note that a similar approach is used also in portfolio theory, where portfolios at times cannot be replicated due to market incompleteness and, as such, an "acceptable" tracking error is used as a threshold for analysis.

[9] Our tolerance index is computed in a slightly different way with respect to Adhikari et al (2016). When we convert our γ=7% to make it comparable, our approach results 30% more stringent than theirs.

donor pools. However, this approach would have the drawback that if the treated country is at the lower (higher) edge of its group, all possible donors will have a higher (lower) income, yielding a synthetic control that is not close to the treated unit. Therefore, we include in the donor pool only countries that have a GDP per capita close to the treated countries, according to the following condition: $Y_t^j \in \left[Y_t^i - \frac{1}{2}Y_t^i, Y_t^i + \frac{1}{2}Y_t^i\right]$ where Y is the GDP per capita, t is the treatment period, i refer to the treated country, j refer to the untreated countries. Table 2 (column 3) offers an empirical backing to this approach, as donor pools constructed in this fashion already start approximating well the economic structure of our countries of interest, although obviously four times less precisely on a yearly basis than our SCM (RMSPE of 13 percent vs 3).

All these transformations and conditions are in line with standard implementations of the SCM in the literature. For example, Abadie et al (2015) builds a synthetic West Germany by restricting the donor pool by income (only member of the OECD in 1990), excludes small-sized countries, members with an income level significantly lower than West Germany, and countries facing large macroeconomic shocks.

Table 2. GDP per capita and covariates means before reform programme implementation

	[t-10,t-1]		
	Top reformers	Synthetic top reformers	Simple average of donor pool
GDP per capita	5932.9	5930.3	5434.7
Investment rate	21.9	24.2	22.8
Industry share	30.5	28.0	30.3
Trade openness	43.6	49.4	56.7
Population growth	1.8	2.1	2.3
Secondary education	10.1	9.0	8.5
Tertiary education	3.6	3.2	3.0
Standardised RMSPE		*3%*	*13%*

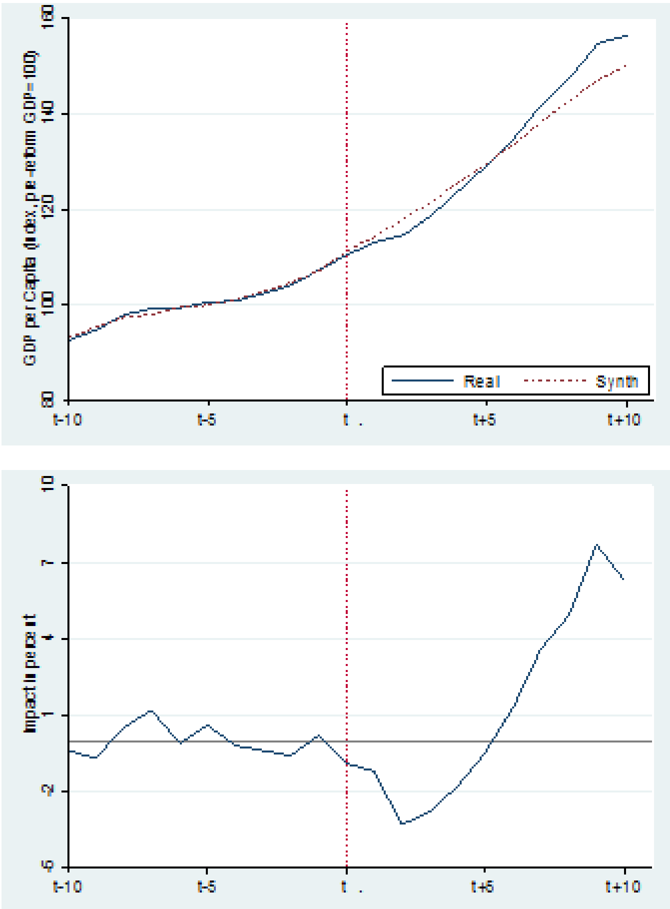
Notes: Covariates are averaged over the 10 years preceding the reform wave [t-10,t-1] and then averaged across reform episodes in the first column. The second column averages across synthetic controls. The third column averages across all countries potentially composing the donor pool. See text for details.

VII. The impact of structural reform packages

In this section, we present and discuss our baseline results. Figure 2 (Panel A) represents graphically the evolution of our interest variable – real GDP per capita – averaged across the 23 reform episodes (real) and their respective synthetic controls (synth)[10]. The close overlap between the lines before *t* suggests we effectively replicated the yearly evolution of GDP in the pre-treatment period, and not only the 10-year average – as already indicated in Table 2.

Our baseline results suggest the average reform country experienced a brief slowdown vis-à-vis counterfactual following the inception of a reform wave. Figure 2 (Panel B) suggests the slowdown lasted for roughly 2 years, after which growth picked up again. Only after 5 years is the average reform country performing better than counterfactual and after 10 years its GDP per capita is 6.3p.p. higher.

Figure 2. (Panel A) Average trend in GDP per capita: reform countries versus synthetic control (Panel B) estimated impact of reforms



One possible source of concern could be that reforms have spillover effects on countries composing the synthetic control, therefore polluting our estimates. To alleviate this concern, we collect bilateral trade

[10] In order to average across reform episodes with a different T_0 without assigning a higher weight to larger countries, we standardise both the real and synthetic, setting GDP over the period [t-10,t-1]=100. This approach sounds reasonable, as both the real and the synth are practically equal over this period, as shown in Table 2.

data from the IMF's Directions of Trade Statistics (DOTS) database and compute the weight of the reformer in the export basket of each respective synthetic control. As displayed in Appendix 5, in 19 out of the 23 episodes considered, the reformer represents less than 1% of the control's export basket. As a conservative check[11], we exclude situations in which the reformer represented more than 3% or 5% of the control's export basket[12]. The impact of reforms at t+10 remains practically unaltered (6.2 p.p. and 5.8 p.p, respectively), displaying similar short-run developments as in the baseline above.

Another source of concern could be that the methodology might be assigning a high weight to individual countries in the synthetic control, hence increasing the risk that post-reform idiosyncratic shocks in the control might be disturbing our estimation. To dispel this concern, we impose a maximum weight restriction of 50% for any donor in each synthetic control. Average results at t+10 stand practically unaltered (6.5p.p. vis-à-vis 6.3 in the baseline case).

A further obvious source of concern could be that reforms are more likely to be implemented at a time of crisis, as suggested *inter alia* by Williamson (1994). If this were predominantly the case in our sample, then our short-term estimates could be biased downwards, dictating the lack of a positive impact of reform over the first 5 years. Moreover, at least part of the subsequent positive effect could be simply imputable to post-crisis recovery or reversion to the mean.

To address this concern, we would want to identify countries experiencing a large financial or macroeconomic crisis that could have later twisted the governments' arm into reform. While these crises could take multiple forms, ranging from sudden stops of capital inflows, to banking crises, or sovereign debt defaults, controlling for all of them would be problematic in a small-*n* setting like ours. However, the financial crisis literature highlights how currency crises have the tendency to manifest themselves in conjunction with all the above (Glick & Hutchison 2011). As such, we take a conservative approach, by conditioning our estimates on the presence of a currency crisis at *t-1*, as defined by Laeven and Valencia (2012)[13]. Crucially, however, we cannot control for a simultaneous crisis at *t*, as this could have been sparked by the (financial) reforms themselves, and would therefore be an integral part of the estimation of the reform impact.

[11] By comparison, the export basket of the synthetic control for West Germany constructed by Abadie et al (2015) depends for 21.0% on (real) West Germany.

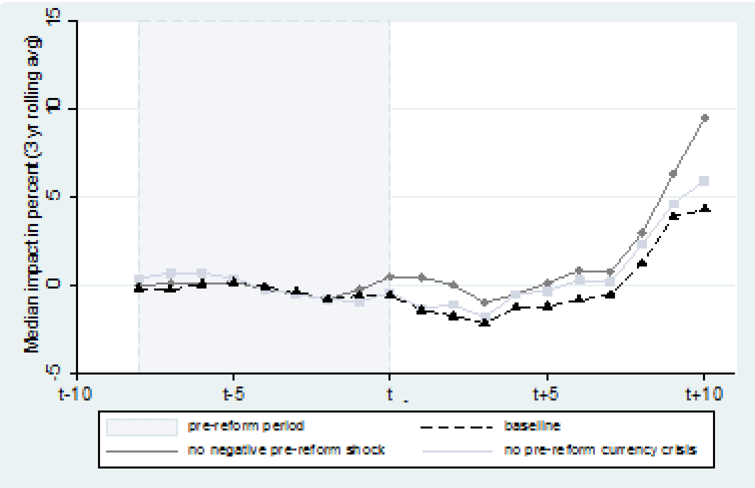
[12] Episodes excluded are Great Britain-1976, Belgium-1988 with a 5% threshold, including also Brazil-1987 in the 3% threshold case.

[13] These included: Argentina – 1987, Bolivia – 1982, Brazil – 1987, Chile – 1973, Egypt – 1990, and Kenya – 1993.

Moreover, we develop our own binary measure of negative idiosyncratic shock by identifying countries that saw a GDP per capita contraction at *t-1*, while their synthetic counterfactual experienced positive GDP growth. This effectively implies that a factor outside our model has negatively affected the reform country's growth.

Figure 3 visualises our results, displaying the median impact of reform waves under alternative specifications[14]. Firstly, the short-term negative effect identified in the average results discussed above is confirmed when looking at the median, suggesting this was not dictated by outliers. Mechanically, this result is clearly less pronounced when countries experiencing a financial crisis or idiosyncratic shock are excluded. However, the overall pattern of the impact of reform seems broadly confirmed, with negative or in any case limited effects of reform to be observed over the short-term. A growth pick up is then observed in the medium term. Looking at our alternative specifications, we can also exclude that the positive effect identified is due entirely to a post-crisis recovery[15].

Figure 3. Median impact of reforms



Statistical testing

While average or median effects are visually informative, we note that point estimates mask a high

[14] Given we are shrinking our sample to an even smaller n, we decided to focus visually on the median impact rather than the mean, to complement Figure 2. This is also due to the median's lower sensitivity to outliers. As the yearly median impact is prone to jumps, we smoothed the data series by taking 3-year averages, hence facilitating a visual comparison across the three specifications.

[15] As an extreme crisis scenario, in an alternative specification we excluded countries implementing reforms at a time of regime change (e.g. Chile in 1973), identified by using the Polity IV database. Long-term positive results remain confirmed, as they do when excluding these instances from the donor pool (results available upon request).

degree of heterogeneity of outcomes: in one out of three reform instances, GDP per capita was still more than 5 percentage points below counterfactual at t+10. As such, we performed basic statistical testing on these results. While the SCM in its original form is non-inferential in nature, the fact that we are applying it serially across reform episodes allows us to construct confidence intervals around the effect, building on the multi-country evidence we have.

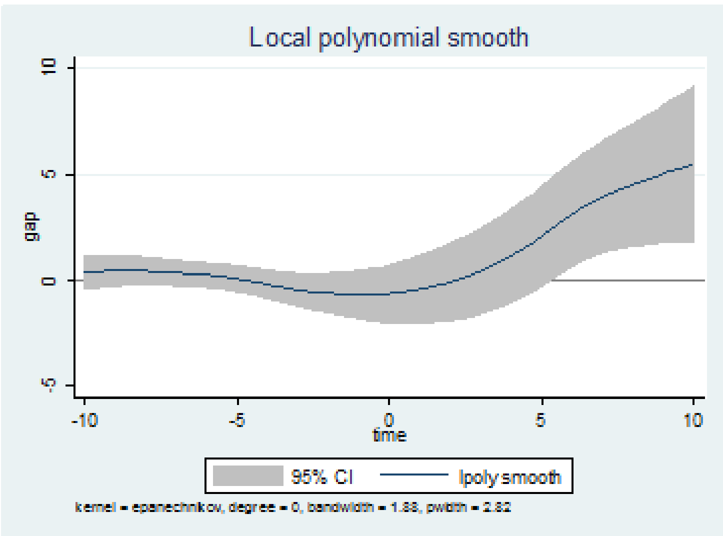
To read common patterns through the cloud of country-specific results, we adopt a non-parametric kernel weighted local polynomial regression model, as proposed by Fan (1992). Within our framework, the model is estimated as follows:

$$\bar{\alpha}_t = \sum\nolimits_i^n \frac{w_{it}\hat{\alpha}_{it}}{\sum_j^n w_{jt}}$$

where $\hat{\alpha}_{it}$ are the standardised gaps between real and synthetic control for the country i at time $t \in [t_0 - 10\,,\ t_0 + 10]$, w_{it} are the kernel weights. This method seems particularly appropriate for our setting, given its good finite sample properties and the fact that it does not force us to over-impose a specific functional form.

As displayed in Figure 4, the difference between real and synthetic is not different from zero in the pre-reform period, suggesting once more a good quality of SCM calibration. Reforms do not appear to have a significant effect in the short term, and hence the average growth slowdown documented above cannot be statistically confirmed, as already suggested by our analysis of the medians. Beyond the 5-year horizon, the difference between the real and the synthetic is positive and significant at the 5% level.

Figure 4. Estimated impact of reforms using a local polynomial regression model



We further checked the sign and significance of the divergence between real and synthetic within a simple panel random effect (RE) model[16], in an SCM variant of what is usually done when running a regression after Propensity Score Matching. More formally, we split our sample in before ($t < 0$) and after ($t \geq 0$) the reform, and estimated β_1 for:

$$\hat{a}_{it} = \beta_0 + \beta_1 t + \varepsilon_{it} \qquad [6]$$

where $\hat{a}_{it}$ is the standardised gap between real and synthetic counterfactual of episode *i* at time *t*, and *t* is a linear time trend[17]. Results (β_1) for the pre- and post-reform period are displayed in Table 4. In line with our earlier findings, there is no statistically significant relationship between gaps before reform implementation. In the 10 years after *t*, the impact of reform is positive and significant at the 1% level. As a further check, we exclude extreme outliers[18], and show that the sign, size, and significance of our measured impact is not substantively affected.

[16] A standard Hausman specification test suggests within our setting a random effect specification is to be preferred to a fixed effect (FE) one. In any case, all results of the paper hold under FE specifications.

[17] An alternative approach would be to run an adapted diff-in-diff panel model, of the sort:

$$\hat{a}_{it} = \beta_0 + \beta_1 t + \delta\ (t\ x\ D_t) + \varepsilon_{it}$$

where D_t is a dummy that takes value 1 for $t \geq 0$. Our coefficient of interest would be δ in this case, effectively testing whether there is a significant change in the slope of the standardised gap between real and synthetic after the reform. This method yields substantially equivalent results to our preferred split sample method, with $\hat{\delta}$ positive and significant (p=0.066) and a β_1 coefficient which is now insignificant, confirming a good model fit.

[18] In the current setting, we define extreme outliers as countries experiencing at any point in time following a reform a gap between real and synthetic that is greater of +/- 45p.p., which would suggest the presence of an idiosyncratic shock of significant proportions.

Table 4. Baseline estimates for a RE model[19]

in p.p.	*pre-reform* [t-10, t-1]	*post-reform* [t, t+10]
23 reform episodes: full sample		
Divergence between reformers and control	-0.013	1.022***
	(0.08)	(0.38)
15 reform episodes: excluding extreme outliers		
Divergence between reformers and control	-0.110	0.812***
	(0.09)	(0.18)

Notes: β_1 coefficients of Model [6] before and after the reform. Positive values indicate a widening gap between reformers and control. Standard errors in parentheses. * p < 0.10, ** p< 0.05, *** p<0.01. See text for additional details.

Reforms in advanced and emerging economies

As discussed in section II, the relevant empirical literature has systematically shown how the impact of reforms tends to be highly heterogeneous across countries, depending on elements such as the composition of reforms or the quality of institutions in a country, as does distance from the technological frontier (Christiansen et al. 2013; Prati et al. 2013). Within our empirical setting, these findings cannot be tested precisely. This is due to the fact that top reformers are identified as implementing multiple reforms at once, but at the same time our small sample setting does not cater for a thorough multivariate regression analysis.

Building on these premises and caveats, we deem worthwhile exploring whether effects are indeed different between advanced[20] and emerging markets. Table 5 displays standard RE regression results, now breaking down the sample in advanced and emerging, and the effect in short-term and long-term[21]. Given our sample contains only few advanced economies, our results in this respect should be treated with caution.

With this caveat in mind, we note how the overall positive effect of reforms is confirmed in both instances. However, advanced economies seem to have reaped fewer benefits from their extensive

[19] The RE model was run using normal standard errors, because of their better small sample properties (Imbens & Kolesar 2016). However, our estimates remain robust at the 5% level to the use of cluster-robust or bootstrapped (200 repetitions) standard errors, at least in the no-outlier specification.
[20] Advanced economy reform episodes are defined here as countries belonging to the OECD in the year the reform was implemented.
[21] We divided the post-reform period in half, defining short-term as [t,t+4] and long-term as [t+5,t+10]. While we acknowledge that this breakdown is somewhat arbitrary, all our results and significance levels are unaffected by changes in the definition of short-term.

reform programmes than emerging markets. Moreover, the time profiling of the payoffs seem somewhat different: while countries closer to the technological frontier, and probably with better institutions, see benefits from reforms reaping already in the first five years, these seem to materialise only in the longer run for emerging markets.

Table 5. Impact of reform estimates for a RE model

in p.p	*post-reform effect*	*short term* [t, t+4]	*long term* [t+5, t+10]
23 reform episodes: full sample			
Divergence between reformers and control	1.022***	-0.336	1.556**
	(0.38)	(0.68)	(0.63)
18 reform episodes: emerging markets			
Divergence between reformers and control	1.123**	-0.651	1.831**
	(0.48)	(0.86)	(0.79)
5 reform episodes: advanced economies			
Divergence between reformers and control	0.656***	0.796**	0.570
	(0.18)	(0.40)	(0.38)

Notes: β_1 coefficients of Model [6] after the reform, subdivided in short- and long term. Positive values indicate a widening gap between reformers and control. Standard errors in parentheses. * p < 0.10, ** p< 0.05, *** p<0.01. See text for additional details.

One argument that could be brought forward to explain this heterogeneous result is that emerging markets have a higher propensity of experiencing a financial or macroeconomic crisis (Eichengreen & Hausmann 2005), leading us to erroneously conclude that reforms only pay off in the longer term in emerging markets. We check for this possibility in Table 6. Independently of the definition of short term adopted, or whether one excludes countries experiencing a financial- or idiosyncratic crisis, emerging markets do not display a similar (positive) effect as the one identified for advanced economies[22].

[22] Advanced economies in our sample had no experience of currency crisis. While one of them was affected by a negative idiosyncratic shock (Norway – 1988), its exclusion does not change the sign and significance of the results displayed in Table 4.

Table 6. RE Model, short-term impact of reforms in emerging markets

	short term		
in p.p	[t, t+4]	[t, t+3]	[t, t+2]
18 reform episodes: emerging markets			
Divergence between reformers and control	-0.651	-1.227	-1.861
	(0.86)	(1.04)	(1.52)
12 reform episodes: emerging markets excluding currency crisis at t-1			
Divergence between reformers and control	-0.797	-1.126	-1.813
	(0.83)	(0.97)	(1.46)
15 reform episodes: emerging market excluding idiosyncratic crisis at t-1			
Divergence between reformers and control	-0.154	-0.748	-1.054
	(0.96)	(1.14)	(1.66)

Notes: β_1 coefficients of Model [6] for emerging markets only under alternative definitions of short term and macroeconomic conditions at t-1. Positive values indicate a widening gap between reformers and control. Standard errors in parentheses. * p < 0.10, ** p< 0.05, *** p<0.01. See text for additional details.

A more plausible explanation for this heterogeneous effect could be that the type of reforms implemented were different in nature. While our framework is not appropriate to test the impact of individual reform categories, we qualitatively observe indeed a different composition in the average reform package between advanced and emerging markets over the analysed period (see Appendix 6). This intuition is further confirmed by our case studies, in section VIII.

In particular, we see that the reform waves of the advanced economies considered: (i) did not contain trade liberalisations, (ii) were relatively more skewed towards the liberalisation of the capital account and of current account transactions, and (iii) were less characterised by agriculture-, domestic finance- and network liberalisation-reforms.

At the same time, the higher quality of institutions in advanced economies could be playing a role not only in the design of country-specific reform packages that are more effective from the onset, but also in determining policy credibility. This in turn is crucial to shift private sector incentives and rapidly affect growth[23]. We will return to this discussion in our concluding remarks.

VIII. Alternative parametric estimation strategy

In the previous section, we presented our baseline estimates of the impact of wide-reaching reforms on growth, based on the non-parametric SCM. Given the novelty of this approach in the literature, in this

[23] Partially lending support to this claim is the fact that following a reform wave, we observe 4 instances of policy reversal over the relevant 10-year horizon, all of which were located in emerging markets. Excluding these cases leads to a slightly larger positive effect of reform at t+10 (10.1 p.p. vs 6.3 in the baseline).

section we show that our findings are not model-specific. In particular, we follow Acemoglu et al (2014) and adapt their dynamic panel regression model to provide an alternative estimate of the impact of reforms on growth.

Our alternative parametric estimation model hence takes the form:

$$y_{ct} = \beta R_{ct} + \sum_{j=1}^{p} \gamma_j y_{ct-j} + \alpha_c + \delta_t + \varepsilon_{ct} \qquad [7]$$

where y_{ct} is the log of real GDP per capita in country c at time t. R_{ct} is a dummy that takes value 1 if country c is a top reformer and $t \in [T_0; T_{10}]$, and zero otherwise; while α_c and δ_t are respectively a full set of country- and time- fixed effects and ε_{ct} is an error term. The specification further includes p lags of log GDP per capita, to control for the dynamics of GDP.

In line with Acemoglu et al (2014), we use the standard fixed effect estimator to estimate equation [7]. Table 7 reports our main results, controlling for different GDP lags. In all specifications, the coefficient of reform ($\hat{\beta}$) is multiplied by 100, to ease reading, and standard errors are clustered and robust to heteroskedasticity.

Table 7. Effect of reforms on GDP per capita

	(1)	(2)	(3)	(4)	(5)
reform	1.133*	1.536**	1.152**	0.930*	0.853*
	(0.585)	(0.656)	(0.560)	(0.515)	(0.510)
log GDP first lag	0.962***	0.962***	1.146***	1.140***	1.122***
	(0.006)	(0.006)	(0.041)	(0.037)	(0.035)
log GDP second lag			-0.189***	-0.131***	-0.126***
			(0.040)	(0.037)	(0.037)
log GDP third lag				-0.049	0.040
				(0.033)	(0.035)
log GDP fourth lag					-0.077***
					(0.019)
Long-run effect	12.5	17.0	12.7	10.3	11.3
p-value	[0.054]	[0.020]	[0.041]	[0.072]	[0.081]
GDP persistence	0.962	0.962	0.957	0.960	0.959
p-value (test<1)	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
SCM baseline sample	no	yes	yes	yes	yes
Reform expectation effect	no	no	no	no	yes
Country FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Observations	5595	5595	5531	5466	5400
Countries in sample	118	118	118	118	118
R-squared	0.996	0.996	0.996	0.996	0.997

Notes: This table presents estimates of the effect of reform on log GDP per capita using a fixed effect dynamic panel regression model. The reported coefficient on reform is multiplied by 100. Long-run effect is the implied aggregate impact of reform at t+10, expressed in percentage points, and p-value for this being different from 0. We report the estimated persistence of the GDP process and the p-value of this being less than 1. Robust standard errors in parentheses. See text for further details.

Column 1 shows the impact of reforms on growth, controlling for one GDP lag. In this specification, we are considering all 29 reform episodes identified in section IV, while from Column 2 onwards we focus only on the 23 "best fit" reform episodes to ensure comparability with our baseline SCM results. We note how GDP persistence is very high, although a standard t-test excludes a unit root in the empirical process of log GDP. Importantly, the impact of reforms is positive and significant, for both specification 1 and 2, suggesting our sample restriction in the baseline came without loss of generality. For the same sample of wide-reaching reforms considered in the baseline, the implied aggregate impact of reforms at t+10 is 17 percentage points of GDP (and the p-value below this estimate suggests this result is significant at the 5% level). In column 3 and 4, we increase the number of GDP lags, accounting for the rich dynamics of GDP. The level of GDP persistence remains comparable to the 1-lag specification in column 2. Moreover, the long-term aggregate effect of reform is reduced to 10.3 p.p. (significant at the 10% level), bringing it strikingly close to our baseline estimates.

Column 5 displays our preferred specification, including four lags of log GDP[24]. Within this setting, we also control for potential expectation effects in the year anticipating the reform[25]. The coefficient of reforms remains positive and significant, pointing towards a long-term increase in GDP per capita of 11.3 p.p. following a reform wave[26].

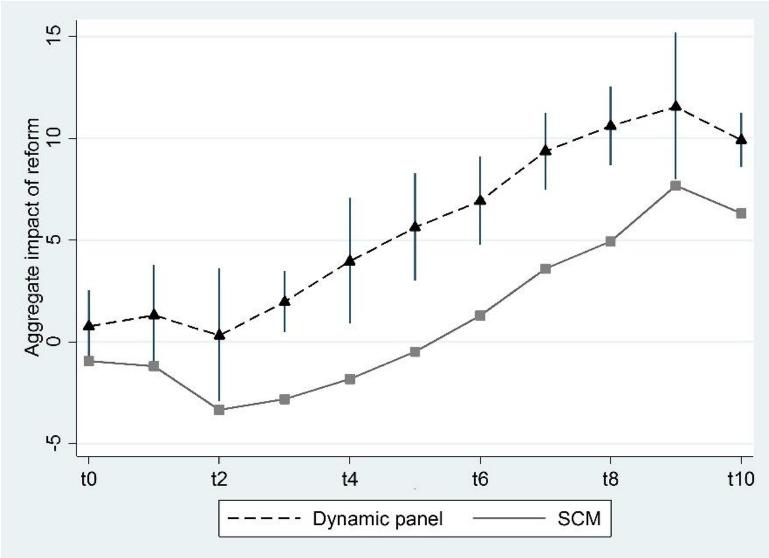
Finally, in order to trace the yearly dynamics of the impact of reform, we slightly modified equation [7] as follows:

$$y_{ct} = \sum_{k=0}^{10} \beta_k R_{k,ct} + \sum_{j=1}^{p} \gamma_j y_{ct-j} + \alpha_c + \delta_t + \varepsilon_{ct} \qquad [8]$$

where $R_{k,ct}$ is a dummy that takes value 1 in the kth-year following a reform wave in country c at time t, and zero otherwise. Estimated coefficients for various specifications are detailed in Appendix 8. Figure 5 displays the imputed impact of reform (for the most sensible *p*=4 specification) and compares it with our baseline SCM estimates. Several key takeaways emerge. First, reassuringly, we note that this alternative estimation strategy points to strikingly similar reform impact dynamics with respect to our baseline, excluding the possibility that our SCM results are entirely a methodological artifact. Second, a panel model better accounting for the GDP cycle confirms our suspicion that the negative short-term dynamics of GDP initially identified by the SCM are not to be imputed to a negative impact of reforms per se. Third, and perhaps most importantly, there are reasons to believe that our SCM estimates of the positive impact of reforms on GDP per capita are to be treated, at most, as conservative.

[24] As illustrated in Appendix 7, four lags seems like the most appropriate specification to map the rich dynamics of GDP. An F-test on further lags (up to 10) does not result in a significant improvement in the specification.
[25] Practically, this implied in the estimation of equation [7] introducing on the right-hand side a dummy taking value 1 for reforming countries in the year preceding a reform. Further reform lags were not significant, suggesting this was not to be interpreted as reforming countries displaying substantially different GDP dynamics vis-à-vis the other countries in the world, which would point in the direction of some sort of self-selection.
[26] Within this specification, we also found that non-robust standard errors are quite similar to the clustered ones (0.508 vs 0.510 for reform), which lends support to the conclusion that we are correctly modelling GDP dynamics.

Figure 5. Estimated impact of reform: dynamic panel and baseline SCM



Note: bars indicate 95% confidence interval. See Appendix 8 for more details.

IX. Robustness checks

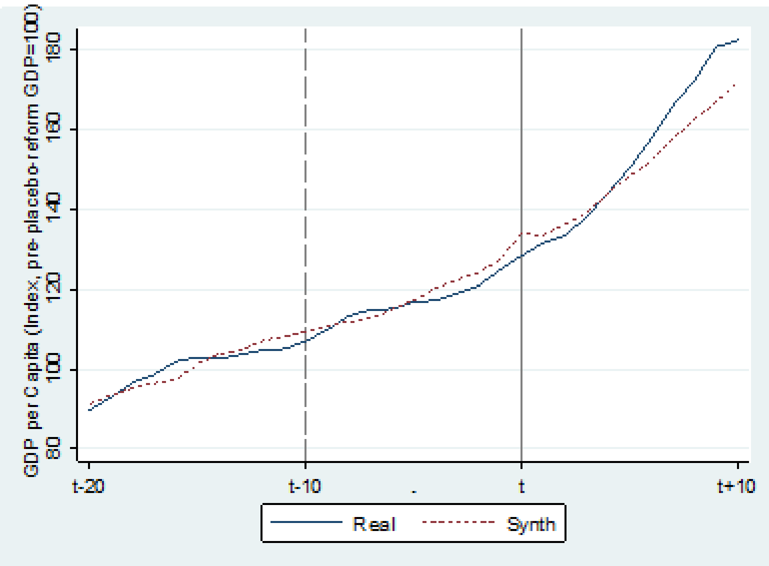
Standard statistical testing and alternative specifications omitting outliers or countries experiencing large macroeconomic shocks ahead of the reform wave show how our baseline results are stable. In this section, we adopt a placebo test, an extended fitting horizon, a falsification test, and alternative indicators of reform to show how our results remain robust under a variety of conditions.

Placebo study

In line with previous SCM studies (Abadie et al. 2010; Mideksa 2013), we carried out an in-time placebo. What this test effectively does is pretend there was a reform at time t-10. The fitting of the synthetic control therefore takes place over the interval [t-20, t-11] for each country. After that, the two are allowed to fluctuate freely. Figure 6 displays the main results.

As can be seen, the quality of our pre-fictitious-reform fit is inferior to that of the baseline specification. However, it must be noted that here we are not applying a “best-fit” filter, but rather displaying results for the 23 countries contained in the baseline, for the sake of comparability. Despite some noise, on average no clear changing pattern can be observed before and after the fictitious reform at t-10. This is particularly true when compared to the effect observed between t and t+10. The divergence between real and synthetic measured at t+10 is more than twice larger than any gap observed over the 20 years preceding the reform. This specification should therefore put to rest potential concerns related to overfitting in the baseline, or else the idea that the reform effect previously identified is just noise that develops as soon as the fit between real and synthetic is no longer imposed by the SCM minimisation. Moreover, the delayed positive effect of reforms is confirmed.

Figure 6. Placebo experiment with fictitious reform wave at t-10



Note: data fit over the 10 years preceding the placebo reform [t-20, t-11], subject to data availability

In line with the statistical testing presented in the baseline section, we employ our standard RE panel regression model to estimate the overall impact of reform in the in-time placebo experiment. Table 9 presents the baseline and placebo results side by side, also breaking the sample between advanced and emerging markets. Following the fictitious reform, we do not observe any statistically significant pattern of divergence between the real and synthetic control. This contrasts with the baseline case. Moreover, we ran a Kolmogorov-Smirnov test, which effectively rejects at the 10% level that the distribution of gaps following reform is the same in the case of the baseline and the placebo for both advanced and emerging markets.

Table 9. RE model, impact of reform comparison between alternative specifications

	post-reform		
in p.p.	baseline [t, t+10]	placebo [t-10, t-1]	*K-S maximum difference*
23 reform episodes: full sample			
Divergence between reformers and control	1.022***	-0.352	0.093
	(0.38)	(0.23)	
18 reform episodes: emerging markets			
Divergence between reformers and control	1.123**	-0.328	*0.131*
	(0.48)	(0.27)	
5 reform episodes: advanced economies			
Divergence between reformers and control	0.656***	-0.439	*0.249*
	(0.18)	(0.40)	

Notes: Columns 1 and 2 display β_1 coefficients of Model [6] after the reform in the baseline and placebo specifications. Positive values indicate a widening gap between reformers and control. Standard errors in parentheses. * p < 0.10, ** p< 0.05, *** p<0.01. Italics denote significant differences in the K-S test at the 10 percent level. See text for additional details.

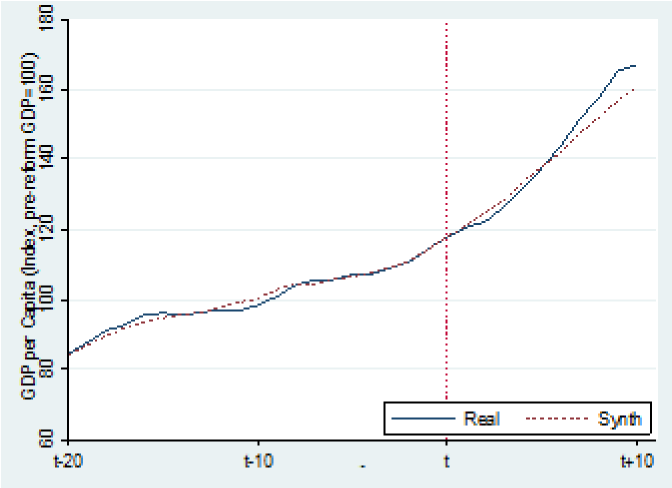
Extended fitting time horizon

As with all macroeconomic studies related to estimating the impact of important policy changes across countries, one of the key concerns could be that of self-selection into the sample. Ultimately, it could be that there are some unobservable characteristics, which make reforms possible and at the same time have an impact on subsequent GDP realisations. Within our empirical setting, while not fully dissipated, there are reasons to believe these concerns are of somewhat minor order with respect to alternative estimation techniques.

Abadie et al. (2010) formally show that once a good match has been established over GDP and observable covariates, the bias originating from time-varying unobservable confounders tends to zero as the fitting horizon tends to infinity. Intuitively, if the synthetic replicates correctly the yearly evolution of GDP of our country of interest, then the likelihood that factors – both observable and unobservable – that have an impact on GDP will be matched by the control increases as the time span of the fitting horizon widens.

To diminish concerns related to potential unobservable characteristics, we double the fitting horizon from 10 to 20 years before the reform wave. As can be seen in Figure 7, our standard results remain unaltered for both the short- and medium term. The final t+10 impact is entirely proportional to what is observed in the baseline scenario (5.9p.p. vs 6.3p.p. in the standard specification). The quality of the pre-reform fit is slightly worse than in the baseline scenario, for the same reasons remarked in the in-time placebo case. Nonetheless, we note that the final reading of the average reform impact at t+10 is over five standard deviations above the pre-reform mean.

Figure 7. Average trend in GDP per capita: reform countries versus synthetic control with extended 20yr pre-reform fitting period



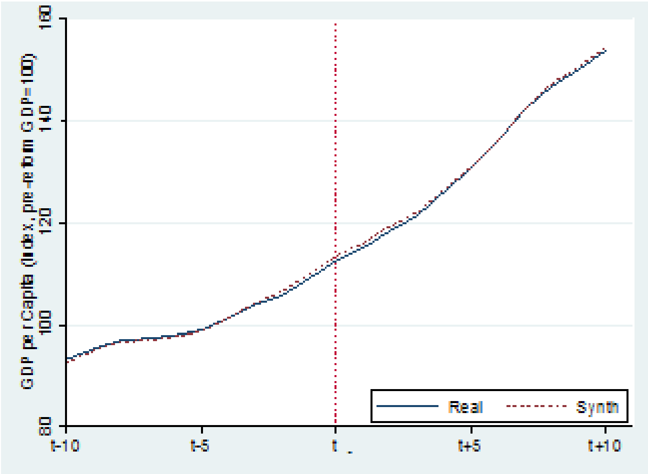
Note: based on baseline countries, subject to data availability

Falsification test

Another form of placebo test consists in applying the SCM to countries that have not had significant reform waves, and observing how their gap evolves vis-à-vis our baseline. Generalising the approach of Abadie et al (2015) to a multi-country setting, we systematically applied the SCM as if a reform wave had happened in every country in each donor pool, using the others to build a counterfactual. After applying our standard "best fit" filter, we were left with 255 fictitious reform episodes and their respective synthetic control, which we then aggregated as in the baseline.

Figure 8 shows how, in the absence of a common reform episode across the countries considered, our model accurately tracks GDP on average. Moreover, this result lends a strong hand to the assumption that lays at the heart of our cross-country application of the SCM, namely that idiosyncratic shocks will be broadly symmetric and therefore that our average impact estimator is asymptotically consistent.

Figure 8. Placebo experiment with fictitious reform wave at t for all controls



Building on these results, we decided to exploit this setting to get a further inferential sense of how significant our baseline results are. We hence computed the average gap at t+10 of a random draw of 23 fictitious reform episodes from the whole control pool[27]. By means of bootstrapping (with over 60,000 repetitions), we obtained a distribution of this average impact of a fictitious reform, which intuitively gives us a sense of the precision of the SCM within our setting. The resulting p-value of our baseline impact of reform was 0.07. The intuition behind this is that there is less than a 7% probability that a result as large as our baseline impact of reform at t+10 was simply random (Appendix 9).

Alternative indicators of structural reforms

A last overarching element of concern with our empirical estimation could be that the database used for our baseline specification does not correctly capture episodes of wide-reaching structural reform or improperly identifies the year of inception of a reform wave. To dissipate at least partially these concerns, we apply our methodology to an alternative, but comparable, list of wide-reaching reform episodes, as identified by IMF (2015). In this setting, the list of policy reform indicators used to identify reform waves is wider, going beyond the Giuliano et al (2013) variables, to include elements such as: (i) legal system and property rights, (ii) hiring and firing regulations, (iii) collective bargaining, (iv) infrastructure, and (v) R&D spending. It is hence unsurprising that IMF (2015) obtains a longer list of reform episodes (as reported in Appendix 10).

[27] The donor pool for each control was built using the standard +/-50% of income rule we used in the baseline. We further imposed the condition that episodes randomly drawn could not be from the same country.

Notwithstanding having some appealing characteristics, including reform breadth and a larger *n*, there are valid reasons that led us to avoid using this database for our baseline analysis. First, the IMF database includes variables that are not commonly referred to as "structural reforms", like expenditure in infrastructure or R&D. Second, while some of the indicators are the same as those used by Giuliano et al (2013), part of the variables are based on opinion surveys (hiring and firing, for example), which are notoriously exposed to subjective biases and tend to be highly correlated with the business cycle more than policy change. This could in turn lead to a fuzzy identification of the reform wave's inception year. Third, as part of the indicators used is based on proprietary data, using this database would have forced us to take the list of reform episodes as given, without allowing in depth analysis of the specific cases and variables. Nonetheless, this alternative list contains information that can prove useful to build a robustness check for our analysis.

After applying our standard filters to the IMF reform wave list, excluding countries experiencing wars, with a population under 1.5 million, and reform episodes that are less than 10 years apart from each other, we remain with 29 episodes. Once we further apply our RMSPE-based "best fit" filter, to ensure the SCM is successfully replicating GDP per capita in the pre-reform period, this number comes down to 22. Finally, in order to ensure a relevant robustness check to our baseline specification, we sift out the countries that implemented reforms in the fields that are covered by the Giuliano et al (2013) database. This leaves 13 episodes for 12 countries.

First of all, it is interesting to note that more than 60% of the reform episodes identified in such a way correspond to those considered for our baseline, though the exact inception years tend to vary somewhat for the reasons discussed above[28].

[28] We considered the episode identified using IMF data to be the same as that using Giuliano et al (2013) data if the inception dates were at most three years apart from each other. The choice of this interval is intuitively based on the methodology we originally adopted to identify reform waves, as discussed in section III.

Table 10. RE model, impact of reform for episodes based on IMF (2015)

in p.p.	*pre-reform* [t-10, t-1]	*post-reform* [t, t+10]	*short term* [t, t+4]	*long term* [t+5, t+10]
13 reform episodes: full comparable sample				
Divergence between reformers and control	-0.113	2.177***	1.245	2.437**
	(0.12)	(0.55)	(0.93)	(1.00)
5 reform episodes: not captured by baseline				
Divergence between reformers and control	-0.227	1.787***	0.344	2.351***
	(0.21)	(0.28)	(0.55)	(0.62)

Notes: β_1 coefficients of Model [6] before and after the reform, further subdivided in short- and long term. Positive values indicate a widening gap between reformers and control. Standard errors in parentheses. * p < 0.10, ** p< 0.05, *** p<0.01. See text for additional details.

Table 10 displays our standard RE model estimations of the impact of reforms, for episodes comparable to our baseline. It also zooms in specifically on the episodes that were not captured by our baseline setting[29]. All main results from our baseline specification are broadly confirmed: the post-reform effect is positive and significant. In particular, reform benefits tend to materialise over the longer term. This is true also when focussing on the reform waves missed by our baseline specification. The fact that point estimates are somewhat higher should come as no surprise given these countries were by IMF definition all simultaneously implementing complementary measures to the ones identified in the baseline. The visual display of these results further confirms that top reformers identified with a slightly modified methodology experienced broadly similar subsequent growth patterns to our baseline cases (see Appendix 11).

X. Economic reform: focus on two case studies

When speaking of the SCM, Billmeier and Nannicini (2013, p.985) refer to it as "*a methodology that builds data-driven comparative case studies within a unified statistical framework*". A great advantage of our approach vis-à-vis standard panel regressions is that it allows us to look in detail at individual countries to explore their reform history and subsequent GDP evolution. In the words of Rodrik (2003, p.10), this can be particularly valuable when speaking about growth processes as "*case studies and cross-national econometrics are not substitutes for each other. […] Any cross-national empirical regularity that cannot be meaningfully verified on the basis of country studies should be regarded as suspect*".

In this section, we therefore look at two specific episodes of deep reform, to put in a specific context the general findings identified above. In particular, we look at the reform episodes starting in the Dominican

[29] These are specifically: Bolivia – 1988, Chile – 1984, Cameron – 1993, Hungary – 1993, and Slovakia – 1999.

Republic in 1989, and in Belgium in 1988. The choice of these two countries is based on several grounds. First, they represent an emerging and advanced economy. Second, they are located in two different continents. Third, the reforms took place broadly at the same time. Finally, while the Dominican Republic undertook broad reforms touching on deregulation, liberalisation, and macroeconomic stabilisation, Belgium focussed on financial reforms of the exchange rate and banking sector.

For both countries, our synthetic control reproduces very closely GDP per capita in the pre-reform period (Appendix 12). As shown in Appendix 13, also the covariates are reproduced in a satisfactory manner, possibly with the exception of Belgium's extraordinary degree of openness to trade. Moreover, the composition of the synthetic sounds intuitively sensible, with neighbouring Panama and Guatemala together composing more than half of "synthetic Dominican Republic", and France having a strong presence in Belgium's control.

In line with our cross-country empirical evidence, Figure 9 and 10 below illustrate how in the Dominican Republic short-term losses were followed by medium term gains. On the other hand, Belgium experienced a positive reform impact from the onset of reform, though the gap stabilised in the medium run. Ten years after the reforms were implemented, the gains in the Dominican Republic were more significant than in Belgium (18p.p. above counterfactual, versus 13p.p.). In the remainder of this section, we detail the historical, economic, and political context characterising the two countries, and detailing what reforms were exactly implemented.

A Latin American story: Dominican Republic in the 1990s[30]

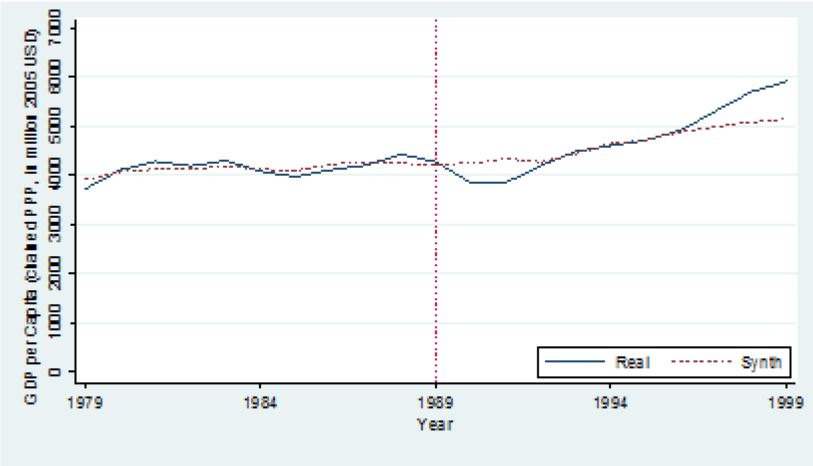
Under the first term of President Joaquin Balaguer, the Dominican Republic of the late 80s was a country characterised by both monetary and fiscal imbalances. At the turn of the decade, the economy deteriorated sharply due to a combination of exogenous shocks: a drop in the price of ferronickel (the country's main export) and a rise in oil prices, in the run up to the Gulf War. This squeezed government revenues. At this point, in an effort to secure the support of the business community ahead of the 1990 elections, Balaguer embarked on a wide-reaching reform programme (Espinal 1995).

The "New Economic Program", as it was known, was later supported by an IMF Stand-By Arrangement from August 1991 onwards. Aside from price and interest rate liberalisations, deregulation, financial sector reform, a tax reform, and an exchange rate devaluation, it rested heavily on fiscal consolidation. The primary balance went from a deficit of about 5 percent of GDP in 1989 to a surplus of almost 2

[30] This section draws heavily from IMF (2001).

percent in 1991-92 (IMF 2001). Moreover, to bring inflation under control, monetary policy was tightened. Perhaps unsurprisingly, Figure 9 shows how this policy mix led to short-term losses. However, reforms did lead to considerable gains in the medium run.

Figure 9. GDP per capita: Dominican Republic (Real) and synthetic control (Synth)



In 1994, the country held once more a presidential election and an interim government was formed. While the new administration lacked the necessary support to push forward on further reforms, it did refrain from reversing the previous ones. As noted by IMF (2001), by the mid-1990s, the Dominican Republic ranked among the world's fastest-growing countries.

Financial reforms in Europe: a case study of Belgium[31]

Towards the end of the 80s, Belgium implemented significant financial sector reforms, which had the appealing empirical characteristic of being exogenously imposed from an EU directive, but at the same time being specifically targeted at Belgium's peculiar financial system arrangement. This is important as it implies that although France has a large weight in the synthetic control, and is equally an EU member, it did not face a similar reform shock because of the directive.

Externally, as reported by Grilli (1989), important steps were being taken in terms of European integration. The Single Market Act was signed in February 1986, establishing 1992 as the deadline for completing the internal market. This is of particular relevance to Belgium as, since 1955, it had been operating a dual exchange rate system together with Luxembourg. Because of the way the system was designed vis-a-vis current account transactions, Wyplosz (1999, p.5) argues that it was a form of

[31] This section draws heavily from Wyplosz (1999).

"market-based capital control" which insulated the monetary authorities from international flows. As such, a 1988 European directive designed to liberalise fully capital movements in the community, explicitly required the abolition of the Belgian-Luxembourgish dual exchange rate system. In compliance with European law, this was abandoned in March 1990.

At the same time, in the aftermath of the Great Depression, Belgian banks had been separated in two categories: deposit- and investment banks. The former were only allowed to collect deposits and invest them in short-term loans, with the result that by 1945 they had specialised in rolling over government debt. In exchange, the government effectively allowed these banks to form a cartel.

Just as in the case of the Dominican Republic, the (exogenous) oil price shock set in motion a process that led to the reform episode defined by Wyplosz (1999, p.6) as the "big bang of the 1989-91" and the end of the banking system cartel. As the debt to GDP ratio hovered above the 100% mark, the government needed to find urgent ways to cut the cost of debt service. To do so, it effectively started borrowing at cheaper rates from foreign investors, rather than making use of direct deals with domestic banks. At this point, the separation of banks in two categories lost its meaning and was effectively scrapped. At the same time, the oligopolistic setting of deposit rates was abolished. Ultimately, in line with the requirements of the Single European Act, the Belgian banking system was fully liberalised and integrated in the common market for financial services.

Figure 10. GDP per capita: Belgium (Real) and synthetic control (Synth)

The impact of these reforms is displayed in Figure 10. Financial reforms had an immediate and progressively growing impact on the Belgian economy over the short run. After this acceleration, growth stabilised in line with the synthetic control. Ten years after the reforms were implemented, the country was doing significantly better than in the counterfactual scenario.

XI. Conclusions

"Structural reforms" are very much the buzzword of the moment, often presented as the silver bullet to reignite growth both in advanced and emerging markets. But have they worked in the past? Our novel empirical approach suggests that, over a period of four decades and across continents, on average, they did. Countries implementing wide-reaching reform programmes have seen their GDP per capita expand over a 10-year horizon, with an average reform effect of 6.3 percentage points. To put this into perspective, over the past decade, at a time of so-called "great convergence" (Baldwin 2016), the GDP per capita gap between high- and upper-middle-income countries[32] has shrunk by a comparable 5.9 percentage points. As such, the *policy perspective* presented in the introduction seems vindicated.

This finding however does not clash necessarily with the *diagnostics perspective*. Average point estimates mask a large degree of heterogeneity in their impact. Within our sample, in one out of three reform instances, GDP per capita was still more than 5 percentage points below counterfactual at t+10. As such, a deep understanding of country-specific factors seems all the more important to increase the chances of designing a successful reform strategy.

Reforms also had a heterogeneous effect across advanced and emerging economies, with the former observing benefits materialise already over the first five years. Reading this result within the context of the relevant institutional literature on the subject, suggests a potential important role for policy credibility. As discussed by Rodrik (2000), the success of reforms in fostering growth crucially hinges on shifting the expectations of the private sector. By protecting more strongly property rights and enjoying the legitimacy originating from participatory and decentralised political systems, high quality institutions as observed in advanced economies are better placed to produce credible policy changes, refrain from policy reversals, and generate changes in incentives more rapidly.

It could be argued, as done by Rodrik (2000), that policymaking is to be seen as an exercise of continuous problem-solving, and therefore that high quality institutions are better placed to identify the "right policies" to respond to ongoing challenges. However, having a shorter horizon over which reforms pay out could also be increasing the political return to reforms, hence making their systematic adoption more likely, and therefore contributing to explaining the lack of long-term convergence observed in the past. This interpretation of the results lends new credibility to the channels behind the *institutional perspective*, while reconciling it with the other perspectives.

[32] World Bank definitions.

Analysing the interaction between policy reform, quality of institutions, political incentives, and long-term growth will surely remain a prosperous avenue for further research.

XII. Bibliography

Abadie, A., Diamond, A. & Hainmueller, J., 2015. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2), pp.495–510.

Abadie, A., Diamond, A. & Hainmueller, J., 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), pp.493–505.

Abadie, A. & Gardeazabal, J., 2003. The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Journal: Economic Policy*, 93(1), pp.114–132.

Abiad, A., Oomes, N. & Ueda, K., 2008. The quality effect: Does financial liberalization improve the allocation of capital? *Journal of Development Economics*, 87(2), pp.270–282.

Acemoglu, D. et al., 2014. Democracy does cause growth. *NBER Working Paper*, (20004), pp.1–64.

Acemoglu, D., Johnson, S. & Robinson, J.A., 2001. The Colonial Origins of Comparative Development : An Empirical Investigation. *The American Economic Review*, 91(5), pp.1369–1401.

Acemoglu, D. & Robinson, J.A., 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty* First., London: Profile Books.

Adhikari, B. et al., 2016. *Can Reform Waves Turn the Tide ? Some Case Studies Using the Synthetic Control Method*,

Aghion, P., Bloom, N., et al., 2005. Competition and Innovation: An Inverted-U Relationship. *The Quarterly Journal of Economics*, 120(2), pp.701–728.

Aghion, P., Howitt, P. & Mayer-Foulkes, D., 2005. The Effect of Financial Development on Convergence: Theory and Evidence. *Quarterly Journal of Economics*, 120(1), pp.173–222.

Alesina, A. & Rodrik, D., 1994. Distributive politics and economic growth. *Quarterly Journal of Economics*, 109(2), pp.465–490.

Anderson, K., 2010. Krueger, Schiff, and Valdes Revisited: Agricultural Price and Trade Policy Reform in Developing Countries since 1960. *Applied Economic Perspectives and Policy*, 1(2), pp.195–231.

Baldwin, R., 2016. *The Great Convergence* First., Cambridge, Massachusetts: Harvard University Press.

Barro, R.J., 1991. Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106(2), pp.407–443.

Barro, R.J. & Lee, J.W., 2013. A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics*, 104, pp.184–198. Available at: http://dx.doi.org/10.1016/j.jdeveco.2012.10.001.

Beck, T., Levine, R. & Loayza, N., 2000. Finance and the sources of growth. *Journal of Financial Economics*, 58(1–2), pp.261–300.

Bekaert, G., Harvey, C.R. & Lundblad, C., 2005. Does financial liberalization spur growth? *Journal of Financial Economics*, 77(1), pp.3–55.

Besley, T. & Burgess, R., 2000. Land Reform, Poverty Reduction, and Growth : Evidence from India. *The Quarterly Journal of Economics*, 115(2), pp.389–430.

Billmeier, A. & Nannicini, T., 2013. Assessing Economic Liberalization Episodes: a Synthetic Control Approach. *The Review of Economics and Statistics*, 95(3), pp.983–1001.

Christiansen, L., Schindler, M. & Tressel, T., 2013. Growth and structural reforms: A new assessment. *Journal of International Economics*, 89(2), pp.347–356.

DeJong, D.N. & Ripoll, M., 2006. Tariffs and growth: An empirical exploration of contingent relationships. *Review of Economics and Statistics*, 88(4), pp.625–640.

Dell'Ariccia, G. et al., 2007. Reaping the Benefits of Financial Globalization. *IMF Occasional Paper Series*, (June), pp.1–37.

Dervis, K. & Page, J.M.J., 1984. Industrial policy in developing countries. *Journal of Comparative Economics*, 8(4), pp.436–451.

Diamond, J., 1997. *Guns, Germs, and Steel* W. W. N. & Company, ed., New York.

Dollar, D., 1992. Outward-Oriented Developing Economies Really Do Grow More Rapidly : Evidence from 95 LDCs. *Economic Development and Cultural Change*, 40(3), pp.523–544.

Dollar, D. & Kraay, A., 2004. Trade, Growth, and Poverty. *The Economic Journal*, 114(493), pp.F22–F49.

Easterly, W., 1993. How much do distortions affect growth? *Journal of Monetary Economics*, 32(2), pp.187–212.

Easterly, W., 2005. National Policies and Economic Growth: A Reappraisal. In Philippe Aghion and Steven N. Durlauf, ed. *Handbook of Economic Growth*. Elsevier B.V., pp. 1015–1059.

Easterly, W., 2001. The Lost Decades : Developing Countries' in Spite of Policy Reform Stagnation. *Journal of Economic Growth*, 6(2), pp.135–157.

Easterly, W. & Levine, R., 2003. Tropics, germs, and crops: How endowments influence economic development. *Journal of Monetary Economics*, 50(1), pp.3–39.

Edwards, S., 2001. Capital Mobility and Economic Performance: Are Emerging Economies Different? *NBER Working Paper 8076*, pp.1–34.

Eichengreen, B., 2001. Capital account liberalization: what do cross-sections studies tell us? *World Bank Economic Review*, 15(3), pp.341–365.

Eichengreen, B. & Hausmann, R., 2005. *Other people's money: Debt Denomination and Financial Instability in Emerging Market Economies*, Chicago and London: The University of Chicago Press.

Eicher, T.S. & Schreiber, T., 2010. Structural policies and growth: Time series evidence from a natural experiment. *Journal of Development Economics*, 91(1), pp.169–179.

Espinal, R., 1995. Economic Restructuring , Social Protest , and Democratization in the Dominican Republic. *Latin American Perspectives*, 22(3), pp.63–79.

Fan, J., 1992. Design-adaptive nonparametric regression. *Journal of the American statistical Association*, 87(420), pp.998–1004.

Fischer, S., Sahay, R. & Carlos, V., 1996. Stabilization and Growth in Transition Economies: The Early Experience. *Journal of Economic Perspectives*, 10(2), pp.45–66.

Frankel, J. & Romer, D., 1999. Does Trade Cause Growth. *American Economic Review*, 89(3), pp.379–399.

Frischtak, C., Hadjimichael, B. & Zachau, U., 1989. *Competition policies for industrializing countries*,

Galindo, A., Schiantarelli, F. & Weiss, A., 2007. Does financial liberalization improve the allocation of investment?. Micro-evidence from developing countries. *Journal of Development Economics*, 83(2), pp.562–587.

Giuliano, P., Mishra, P. & Spilimbergo, A., 2013. Democracy and reforms: Evidence from a new dataset. *American Economic Journal: Macroeconomics*, 5(4), pp.179–204.

Glick, R. & Hutchison, M., 2011. Currency crises. *Federal Reserve Bank of San Francisco - Working Paper Series*, (No. 2011-22), pp.1–30.

Gollin, D., 2010. Agricultural Productivity and Economic Growth. In *Handbook of Agricultural Economics*. Elsevier B.V., pp. 3825–3866.

Grilli, V., 1989. Financial Markets and 1992. *Brookings Papers on Economic Activity*, 2, pp.301–324.

Gulati, A., Fan, S. & Dalafi, S., 2005. *The Dragon and the Elephant: Agricultural and Rural reforms in China and India*,

Hall, R.E. & Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics*, 114(1), pp.83–116.

Hausmann, R. & Rodrik, D., 2003. Economic development as self-discovery. *Journal of Development Economics*, 72(2), pp.603–633.

Havrylyshyn, O. & van Rooden, R., 2003. Institutions Matter in Transition, But So Do Policies. *Comparative Economic Studies*, 45(1), pp.2–24.

Imbens, G.W. & Kolesar, M., 2016. Robust standard errors in small samples: some practical advice. *Review of Economics and Statistics*, 98(4), pp.701–712.

IMF, 2015. *Structural Reforms and Macroeconomic Performance : Initial Considerations for the Fund*,

IMF, 2001. The Dominican Republic: Stabilization, Reform, and Growth. *IMF Occasional Paper Series*, (206), pp.1–80.

Jalilian, H., Kirkpatrick, C. & Parker, D., 2007. The Impact of Regulation on Economic Growth in Developing Countries: A Cross-Country Analysis. *World Development*, 35(1), pp.87–103.

Jordà, Ò. & Taylor, A.M., 2016. The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy. *Economic Journal*, 126(590), pp.219–255.

Klein, M.W. & Olivei, G.P., 2008. Capital account liberalization, financial depth, and economic growth. *Journal of International Money and Finance*, 27(6), pp.861–875.

Kose, M.A. et al., 2009. Financial Globalization: A Reappraisal. *IMF Staff Papers*, 56(1), pp.8–62.

Laeven, L. & Valencia, F., 2012. Systemic Banking Crises Database. *IMF Working Papers*, 12(163), pp.1–32.

Levine, B.R. & Renelt, D., 1992. A Sensitivity Analysis of Cross-Country Growth Regressions. *The American Economic Review*, 82(4), pp.942–963.

Levine, R., 2005. Finance and Growth: Theory and Evidence. In P. Aghion & S. N. Durlauf, eds. *Handbook of Economic Growth*. Elsevier B.V., pp. 865–934.

Levine, R., 1997. Financial development and economic growth: views and agenda. *Journal of Economic Literature*, XXXV(June), pp.688–726.

Lin, J.Y., 1992. Rural Reforms and Agricultural Growth in China. *The American Economic Review*, 82(1), pp.34–51.

Macours, K. & Swinnen, J.F.M., 2000. Causes of output decline in economic transition: the case of Central and Eastern European agriculture. *Journal of Comparative Economics*, 28(1), pp.172–206.

De Melo, M., Denizer, C. & Gelb, A., 1996. From Plan to Market Patterns of Transition. *The World Bank Policy Research Working Paper*, (January 1996), pp.397–424.

Mideksa, T.K., 2013. The economic impact of natural resources. *Journal of Environmental Economics and Management*, 65(2), pp.277–289.

Nicoletti, G. & Scarpetta, S., 2003. Regulation, productivity and growth: OECD evidence. *Policy Research Working Paper Series*, pp.1–26.

Ostry, J.D., Prati, A. & Spilimbergo, A., 2009. Structural Reforms and Economic Performance in Advanced and Developing Countries. *IMF Occasional Paper Series*, 268.

Pevehouse, J., Nordstrom, T. & Warnke, K., 2004. The Correlates of War 2 International Governmental Organizations Data Version 2.0. *Conflict Management and Peace Science*, 21(2), pp.101–119.

Pingali, P.L. & Xuan, V.-T., 1992. Vietnam: Decollectivization and rice productivity growth. *Economic Development & Cultural Change*, 40(4), pp.697–718.

Prasad, E.S. et al., 2007. Financial Globalization, Growth and Volatility in Developing Countries. In

A. Harrison, ed. *Globalization and Poverty*. University of Chicago Press, pp. 457–516.

Prati, A., Onorato, M.G. & Papageorgiou, C., 2013. Which Reforms Work and Under What Institutional Environment ? Evidence From a New Data Set on Structural Reforms. *The Review of Economics and Statistics*, 95(July), pp.946–968.

Quinn, D., 1997. The Correlates of Change in International Financial Regulation. *The American Political Science Review*, 91(3), pp.531–551.

Quinn, D. & Toyoda, A.M., 2008. Does capital account liberalization lead to growth? *Review of Financial Studies*, 21(3), pp.1403–1449.

Rajan, R.G. & Zingales, L., 1998. Financial Dependence and Growth. *The American Economic Review*, 88(3), pp.559–586.

Rodriguez, F. & Rodrik, D., 2001. Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence. In K. Bernanke, Ben S. ; Rogoff, ed. *NBER Macroeconomics Annual 2000, Volume 15*. MIT PRess, pp. 261–338.

Rodrik, D., 2004. Growth strategies. *NBER Working Paper 10050*, pp.1–57.

Rodrik, D., 2000. Institutions for high-quality growth: What they are and how to acquire them. *Studies in Comparative International Development*, 35(3), pp.3–31.

Rodrik, D., 2009. *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*, Princeton, NJ: Princeton Univerity Press.

Rodrik, D., 1995. Trade and industrial policy reform. In J. Behrman & T. N. Srinivasan, eds. *Handbook of Development Economics*. Elvisier, pp. 2925–2982.

Rodrik, D., 2003. What Do We Learn from Country Narratives. In D. Rodrik, ed. *In Search of Prosperity: Analytical Narratives on Economic Growth*. Princeton Univerity Press, pp. 1–20.

Rodrik, D., 1998. Who Needs Capital-Account Convertibility? In P. Kenen, ed. *Should the IMF Pursue Capital Account Convertibility*. Princeton, NJ: Princeton Univerity Press, pp. 55–65.

Rodrik, D., 2012. Why We Learn Nothing from Regressing Economic Growth on Policies. *Seoul Journal of Economics*, 25(2), pp.137–151.

Rodrik, D. & Subramanian, A., 2009. Why Did Financial Globalization Disappoint? *IMF Staff Papers*, 56(1), pp.112–138.

Rodrik, D., Subramanian, A. & Trebbi, F., 2004. Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9(2), pp.131–165.

Rosenbaum, P.R. & Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), pp.41–55.

Rozelle, S. & Swinnen, J.F.M., 2004. Success and Failure of Reform: Insights from the Transition of Agriculture. *Journal of Economic Literature*, 42(2), pp.404–456.

Sachs, J.D., 2001. Tropical underdevelopment. *NBER Working Paper*, 8119.

Sachs, J.D. & Warner, A., 1995. Economic reform and the process of Global integration. *Brookings Papers on Economic Activity*, 26(1), pp.1–118.

Wacziarg, R. & Welch, K.H., 2008. Trade liberalization and growth: New evidence. *World Bank Economic Review*, 22(2), pp.187–231.

Williamson, J., 1994. *The Political Economy of Policy Reform* J. Williamson, ed., Washington D.C.: Institute for International Economics.

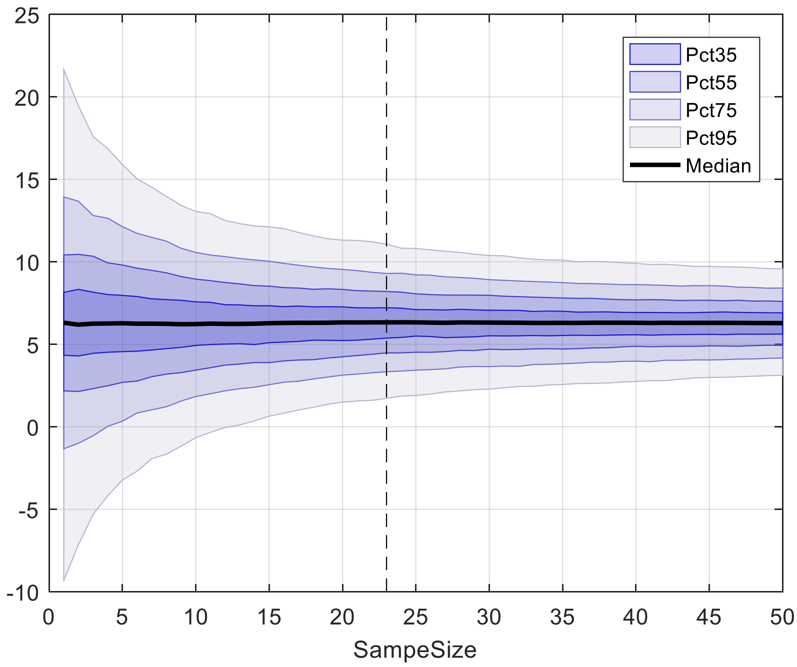
Wyplosz, C., 1999. Financial restraints and liberalization in postwar Europe. *CEPR Discussion paper*, 2253(October), pp.1–58.

Appendix 1. Monte Carlo experiment supporting multi-country application of the SCM

To support our methodological contribution, as described in equations [4] and [5], we designed a Monte Carlo experiment to study the speed at which the average treatment effect becomes less affected by transitionary shock as n increases.

First of all we collected a matrix of effects of reforms $Y_{it}^{treated} - \boldsymbol{W}_i^* \boldsymbol{Y}_{it}^{control}$ for our baseline set of results over the ten year horizon analysed. We subtracted to this matrix the average treatment effect of each period $\left(Y_{it}^{treated} - \boldsymbol{W}_i^* \boldsymbol{Y}_{it}^{control}\right) - \hat{\alpha}_{avg,t}$ obtaining a country specific effect additive to the average treatment effect for each period; such value might be interpreted as an estimate of the difference between the transitionary shocks of the treated and the synthetic country $\left(\varepsilon_{it}^{treated} - \boldsymbol{W}^* \boldsymbol{\varepsilon}_{it}^{control}\right)$. Once we obtained these estimate of the transitionary shocks, we calculated the standard deviation for each treated unit and inputted such standard deviations into a Monte Carlo experiment.

For each one of the 23 cases that compose our baseline results, we generated ten thousands simulation with as mean the average treatment effect $\hat{\alpha}_{avg,t}$ and a standard deviation calculated as described above. Finally, we randomly sampled from the simulation and calculated the average treatment effect increasing the sample size, with n going from 1 to 50. The fan chart of such experiment shows how the uncertainty associated to a transitionary shock shrinks rapidly as n increases. At a 95% confidence level, the average treatment effect would not be different from zero up to $n = 12$, while with our sample size $n = 23$, the 95% percentile interval ranges from 2% to 11%. This suggests that it is very unlikely that the positive average effect of reform at T+10 that we observe in the baseline can be entirely dictated by transitionary shocks.



Appendix 2. Full list of reform wave episodes based on Giuliano (2013) database

Country	Year	Agriculture	Network utilities	Trade liberalisation	Capital account	Current account transactions	Domestic finance
Albania	1991					X	X
Argentina	1974				X	X	X
Argentina	1987				X	X	
Belgium	1988					X	X
Bolivia	1982					X	X
Brazil	1987	X		X			
Bulgaria	1997		X		X		X
Chile	1973			X			X
Colombia	1988				X	X	X
Dominican Republic	1989		X				X
Ecuador	1989			X			X
Egypt	1990	X					X
El Salvador	1995		X		X		
Ghana	1996	X	X				
Jamaica	1988	X			X	X	
Kenya	1993			X	X		
Mexico	1988				X		X
New Zealand	1981	X			X	X	

Norway	1988		X			X	
Peru	1987			X		X	
Philippines	1961			X		X	
Portugal	1990				X	X	
Romania	2000		X		X		
Sri Lanka	1991	X				X	
Tanzania	1991	X				X	X
Tunisia	1992				X		X
Uganda	1990	X				X	
United Kingdom	1976				X	X	
Venezuela	1987				X	X	

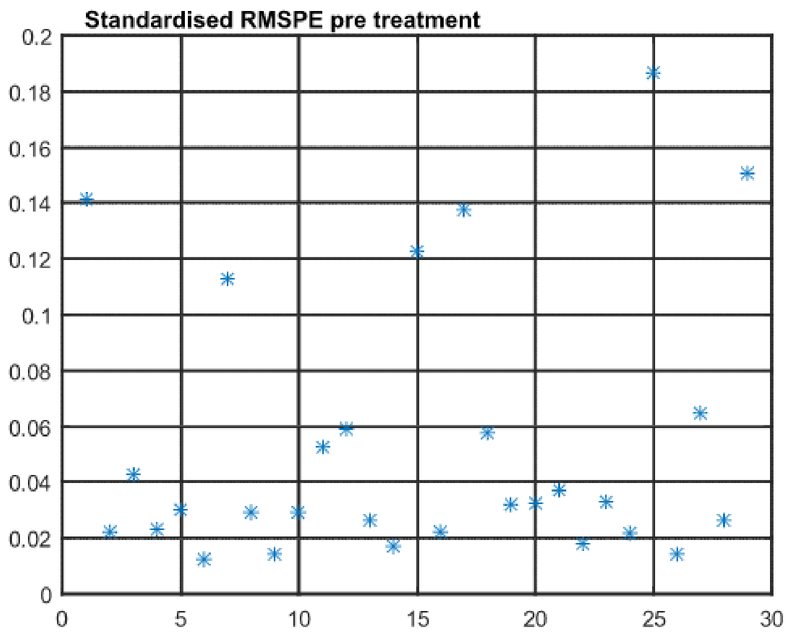
Appendix 3. Data sources

Variables	Description	Unit	Source
GDP per capita	Expenditure-side real GDP at chained PPP (in mil. 2005 US $) divided by Population	2005 International dollar per person	Penn World Table 8.1
Population	Population	Millions	Penn World Table 8.1
Population growth	Population growth	Annual %	Penn World Table 8.1
Investment	Investment Share of PPP Converted GDP Per Capita at 2005 constant prices	% of GDP	Penn World Table 7.0
Openness	Exports plus imports at 2005 constant prices	% of GDP	Penn World Table 7.0
Secondary	Completion ratio for secondary education	% of population aged 25 and over	Barro and Lee (2013)
Tertiary	Completion ratio for tertiary education	% of population aged 25 and over	Barro and Lee (2013)
Industry	Value added by Mining, Manufacturing, Utilities, and Construction	% of GDP	UN Data

Appendix 4. Standardised RMSPE

Figure 1 displays the standardised RMSPE of each reform episode over the period [t-10,t-1]. It is evident that the GDP of some reforming countries is poorly replicated by their respective synthetic control. In particular, we can see that the $RMSPE_{i,t}^{std}$ are already somewhat clustered between good matches ($RMSPE_{i,t}^{std} < 7\%$) and bad matches ($RMSPE_{i,t}^{std} > 11\%$). As such, it looks reasonable to impose a threshold γ = 7%. We note that changes around this level would not alter the selection of countries.

Figure 1. Quality of the fit over [t-10,t-1]



Appendix 5. Composition of synthetic controls

	Argentina 1974	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Peru	0.46	1.6%
Sri Lanka	0.29	0.0%
Romania	0.22	0.0%
Zambia	0.03	0.0%
Trade link between real and synthetic		*0.8%*

	Argentina 1987	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Thailand	0.78	0.0%
Chile	0.17	3.8%
China	0.05	0.0%
Trade link between real and synthetic		*0.6%*

	Begium 1988	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
France	0.9	9.2%
Singapore	0.06	0.6%
Greece	0.04	2.8%
Trade link between real and synthetic		*8.4%*

	Bolivia 1982	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Morocco	0.47	0.0%
Mongolia	0.35	0.0%
Zimbabwe	0.13	0.0%
Ghana	0.04	0.0%
Trade link between real and synthetic		*0.0%*

	Brazil 1987	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Costa Rica	0.25	0.2%
Morocco	0.25	1.0%
Korea, Rep	0.18	0.1%
Uruguay	0.17	27.2%
Zimbabwe	0.16	0.1%
Trade link between real and synthetic		*5.0%*

	Chile 1973	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Mexico	0.47	1.4%
Uruguay	0.31	2.9%
Peru	0.11	1.4%
Hong Kong	0.1	0.0%
Romania	0.01	0.0%
Trade link between real and synthetic		*1.7%*

	Colombia 1988	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Costa Rica	0.41	0.1%
South Africa*	0.27	0.0%
Zimbabwe	0.13	0.0%
Thailand	0.08	0.0%
Uruguay	0.08	0.4%
Turkey	0.03	0.0%
Trade link between real and synthetic		*0.1%*

	Dominican Republic 1989	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Guatemala	0.32	0.8%
Romania	0.19	0.0%
Panama	0.19	0.2%
Philippines	0.18	0.0%
Zimbabwe	0.11	0.0%
Trade link between real and synthetic		*0.3%*

	Ecuador 1989	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Iran	0.43	0.0%
Panama	0.42	0.8%
Thailand	0.15	0.0%
Trade link between real and synthetic		*0.3%*

	Egypt 1990	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Lesotho*	0.32	0.0%
Lao	0.27	0.0%
Mongolia	0.16	0.0%
China	0.13	0.1%
Congo, Rep.	0.08	0.0%
Honduras	0.03	0.0%
Trade link between real and synthetic		*0.0%*

	El Salvador 1995	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Mali	0.58	0.0%
Nepal	0.22	0.0%
Burundi	0.11	0.0%
Central African Republic	0.08	0.0%
Trade link between real and synthetic		*0.0%*

	Ghana 1996	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Pakistan	0.47	0.0%
Honduras	0.17	0.0%
Lesotho*	0.17	0.2%
Zambia	0.1	0.0%
Benin	0.09	0.1%
Trade link between real and synthetic		*0.0%*

	Kenya 1993	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Mauritania	0.44	0.0%
Côte d'Ivoire	0.26	0.0%
Cambodia	0.21	0.0%
Benin	0.1	0.0%
Trade link between real and synthetic		*0.0%*

	New Zeland 1981	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Australia	0.46	4.7%
Sweden	0.22	0.1%
Uruguay	0.21	0.0%
Israel	0.1	0.1%
Trade link between real and synthetic		*2.2%*

	Norway 1988	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Switzerland	0.51	0.7%
Italy	0.3	0.6%
Singapore	0.12	0.2%
Japan	0.06	0.4%
Trade link between real and synthetic		*0.6%*

	Peru 1987	
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Morocco	0.35	0.0%
Philippines	0.35	0.0%
Chile	0.3	1.6%
Trade link between real and synthetic		*0.5%*

Philippines 1961		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Brazil	0.82	0.0%
Thailand	0.13	0.1%
Paraguay	0.05	0.0%
Trade link between real and synthetic		*0.0%*

Portugal 1990		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Spain	0.47	6.2%
Poland	0.33	0.0%
Korea, Rep.	0.14	0.1%
Singapore	0.04	0.1%
Turkey	0.01	0.2%
Trade link between real and synthetic		*3.0%*

Romania 2000		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
South Africa	0.62	0.0%
Peru	0.29	0.1%
China	0.08	0.1%
Trade link between real and synthetic		*0.1%*

Sri Lanka 1991		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Pakistan	0.44	1.2%
Morocco	0.32	0.0%
China	0.21	0.2%
Mongolia	0.03	0.0%
Trade link between real and synthetic		*0.6%*

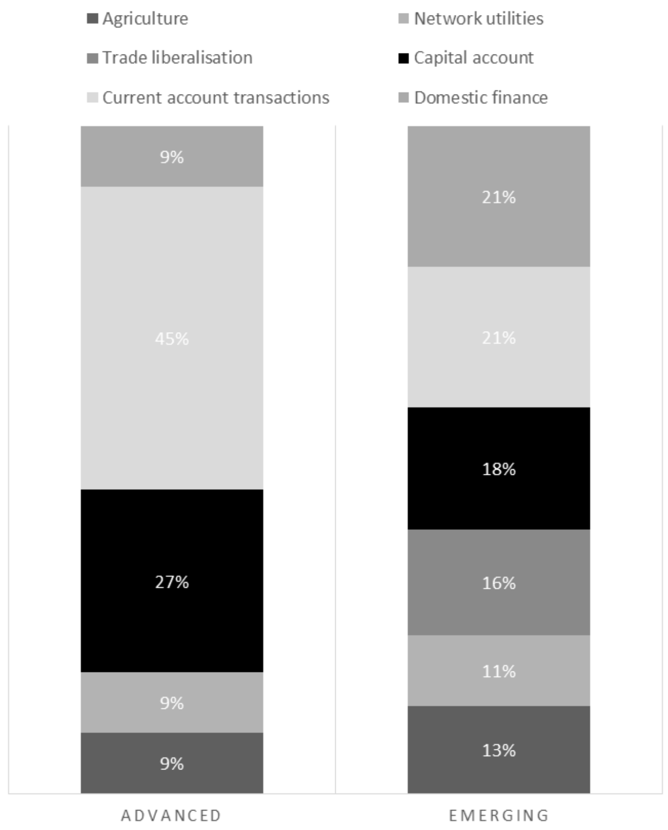
Tunisia 1992		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Thailand	0.54	0.0%
Panama	0.15	0.0%
Indonesia	0.15	0.0%
South Africa*	0.15	0.0%
Trade link between real and synthetic		*0.0%*

Uganda 1990		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Mozambique	0.29	0.0%
Malawi	0.22	0.0%
Nepal	0.19	0.0%
Cambodia	0.15	0.0%
Mali	0.14	0.0%
Trade link between real and synthetic		*0.0%*

Great Britain 1976		
Country	Weight in the synthetic	Weight of reformer in the export basket at t-1
Denmark	0.57	18.8%
Uruguay	0.39	4.7%
Australia	0.04	4.5%
Trade link between real and synthetic		*12.7%*

*indicates data for the overall Southern African Customs Union was used

Appendix 6. Relative composition of reform waves in advanced and emerging markets



Appendix 7. Effect of lags on log GDP per capita

	4 lags	6 lags	8 lags	10 lags
	(1)	(2)	(3)	(4)
log GDP first lag	1.122***	1.117***	1.122***	1.119***
	(0.035)	(0.036)	(0.039)	(0.042)
log GDP second lag	-0.126***	-0.141***	-0.157***	-0.170***
	(0.037)	(0.036)	(0.037)	(0.045)
log GDP third lag	0.040	0.065*	0.081***	0.102***
	(0.035)	(0.033)	(0.026)	(0.032)
log GDP fourth lag	-0.077***	-0.069***	-0.073***	-0.082***
	(0.019)	(0.025)	(0.022)	(0.022)
p-value first four lags	[0.000]	[0.000]	[0.000]	[0.000]
p-value additional lags		[0.330]	[0.198]	[0.486]
Observations	5400	5272	5156	5040
Countries in sample	118	118	118	118

Notes: This table presents estimates of lagged GDP per capita on GDP per capita. In each column we add a different number of lags as specificed in the column table. Only the coefficients of the first four lags are reported. Below each model we report the p-value for a test of joint significance of the first four lags, and the p-value of the additional lags. Robust standard errors in parentheses.

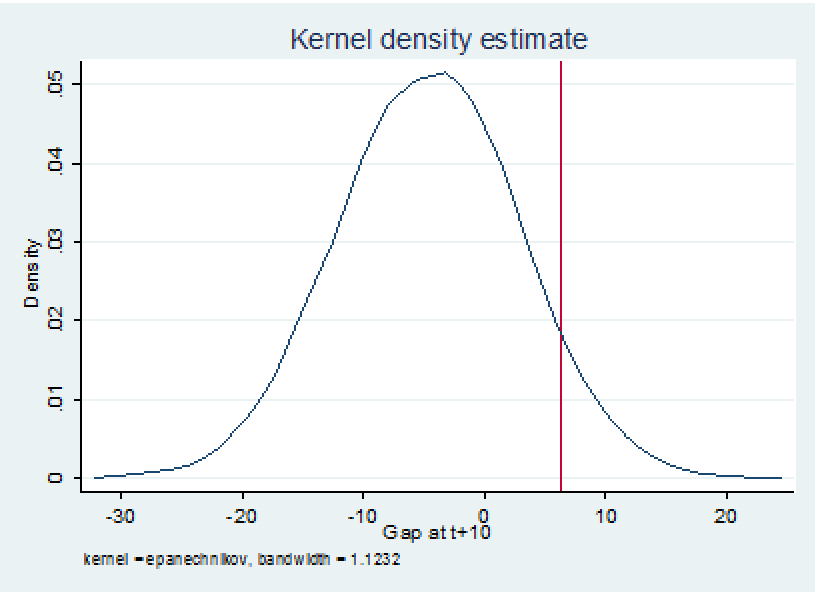
Appendix 8. Estimated yearly impact of reform

	(1)	(2)	(3)	(4)	(5)
reform impact at:					
t0	0.0141	0.00913	0.00747	0.00689	0.00756
	(0.011)	(0.010)	(0.009)	(0.009)	(0.009)
t1	0.012	0.00767	0.00524	0.00476	0.005
	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
t2	0.000264	-0.00296	-0.00825	-0.0105	-0.00985
	(0.019)	(0.018)	(0.016)	(0.016)	(0.016)
t3	0.0235**	0.0223***	0.0187**	0.0158**	0.0163**
	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)
t4	0.0303*	0.0247	0.0205	0.0188	0.0193
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
t5	0.0273**	0.0212	0.0163	0.0153	0.016
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
t6	0.0199	0.0147	0.0143	0.0117	0.0123
	(0.012)	(0.010)	(0.010)	(0.010)	(0.010)
t7	0.0270***	0.0235***	0.0232***	0.0220**	0.0226**
	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
t8	0.016	0.0114	0.0115	0.0107	0.0112
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
t9	0.011	0.0086	0.00816	0.00797	0.008
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)
t10	-0.0114	-0.0128*	-0.0142**	-0.0151**	-0.0147**
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Overall reform effect	17.2	12.9	10.4	8.9	11.4
p-value	[0.021]	[0.040]	[0.070]	[0.099]	[0.078]
Number of GDP lags	1	2	3	4	4
SCM baseline sample	yes	yes	yes	yes	yes
Reform expectation effect	no	no	no	no	yes
Country FE	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes
Observations	5595	5531	5466	5400	5400
Countries in sample	118	118	118	118	118
R-squared	0.996	0.996	0.996	0.997	0.997

Notes: This table presents estimates of the yearly effect of reform on log GDP per capita using a fixed effect dynamic panel regression model. Long-run effect is the implied aggregate impact of reform by t+10, expressed in percentage points, and p-value for this being different from 0. We estimate this for different numbers of GDP lags (coefficients not reported). In all specifications we include a full set of country and year fixed effects. Robust standard errors in parentheses.

Appendix 9. Distribution of placebo impact of fictitious reform at t+10

The aim of this exercise was to generalise the construction of a pseudo-p-value (as in Abadie et al. 2015) to a multi-country setting. To do so, we selected all the possible donors of the 29 case studies and, excluding duplicates, we obtained 345 control cases. We then ran the SCM on all the controls and applied our standard "best fit" filter, bringing the number of control cases to 232. We then constructed a large number of random sub-samples composed of 23 units, sampling (60,000 repetitions) with replacement but excluding the sub-samples where a country was present more than once within a time window of 20 years (for example, the same sub-sample of 23 units could not contain simultaneously Italy-1990 and Italy-1998). Once this rule was applied, this resulted in eight thousand sub-samples of 23 controls, of which we took the average t+10 gap between real and synthetic control. The distribution of gaps is displayed below. The intuition behind this is that there is less than a 7% probability that a result as large as our baseline impact of a reform wave at t+10 (6.3 percentage points) was simply random.

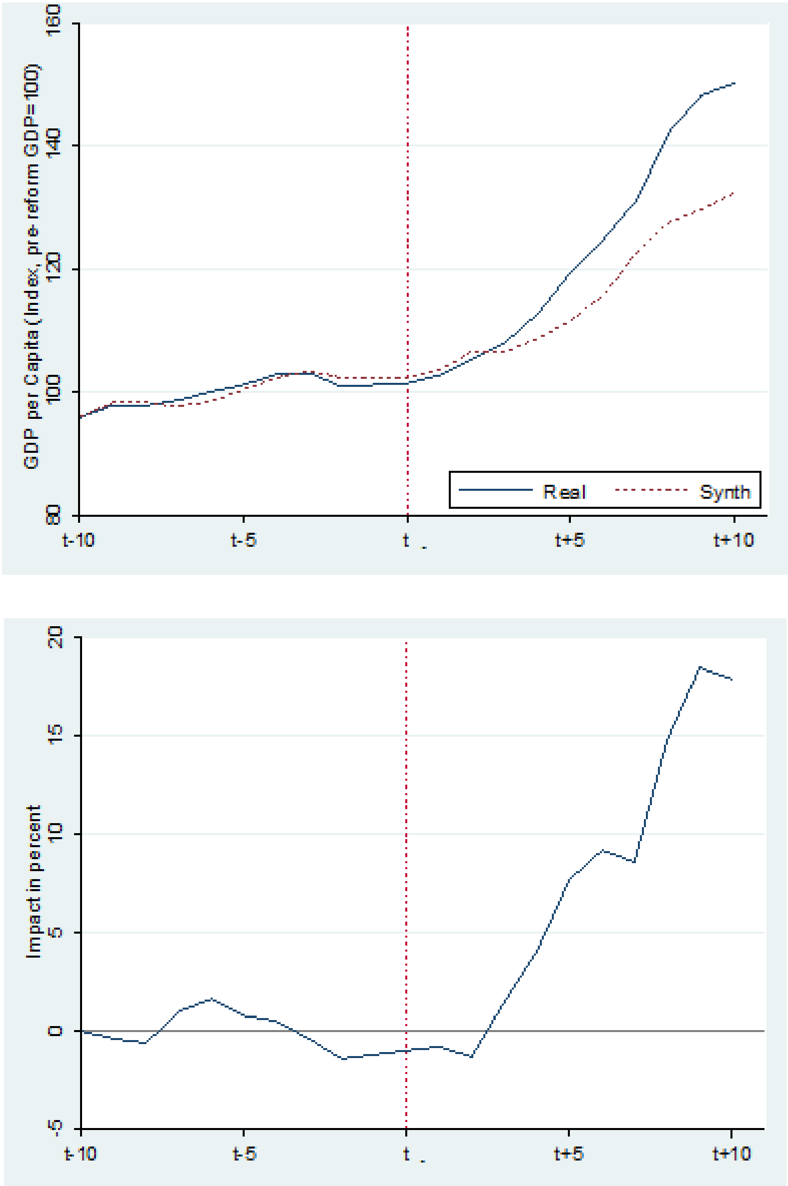


Appendix 10. Full list of reform episodes as reported in IMF (2015)

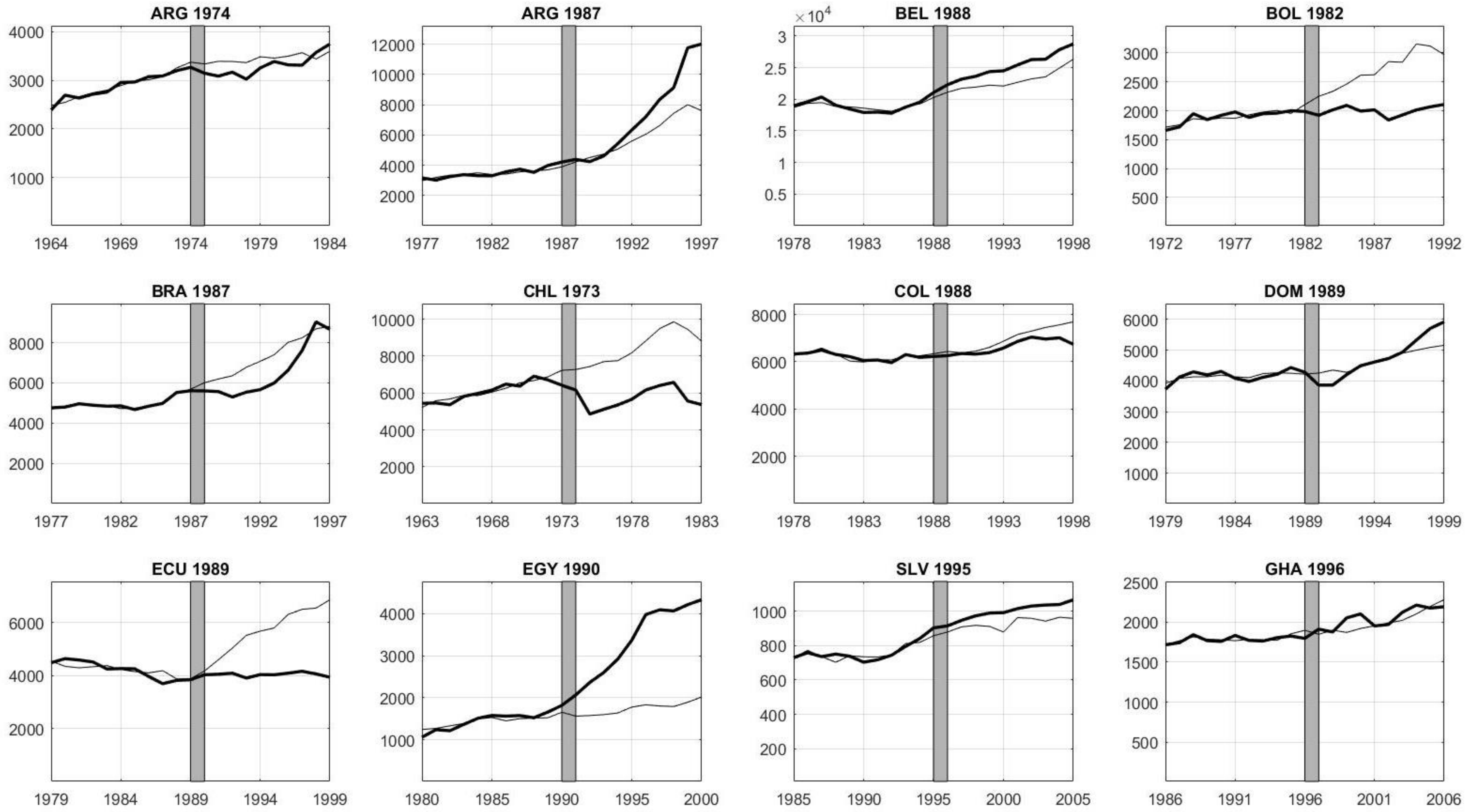
Country	Year	Banking	Capital markets	Tariff	Legal system	Infrastructure	Agriculture	Policy Environment	Promotion of competition	Hiring and firing	Collective bargaining	ETCR	R&D
Argentina	1988	X	X	X									
Bolivia	1988	X	X		X								
Brazil	1985	X	X	X									
Bulgaria	1995	X						X	X				
Cameroon	1993	X	X				X						
Chile	1974	X	X	X									
Chile	1984		X	X								X	
Colombia	1987	X	X	X									
Colombia	2000								X	X	X		
Czech Republic	1995							X	X		X		
Dominican Republic	2002				X					X	X		
Ecuador	1988	X	X	X									
Egypt	1999			X						X	X		
El Salvador	1988	X	X		X								

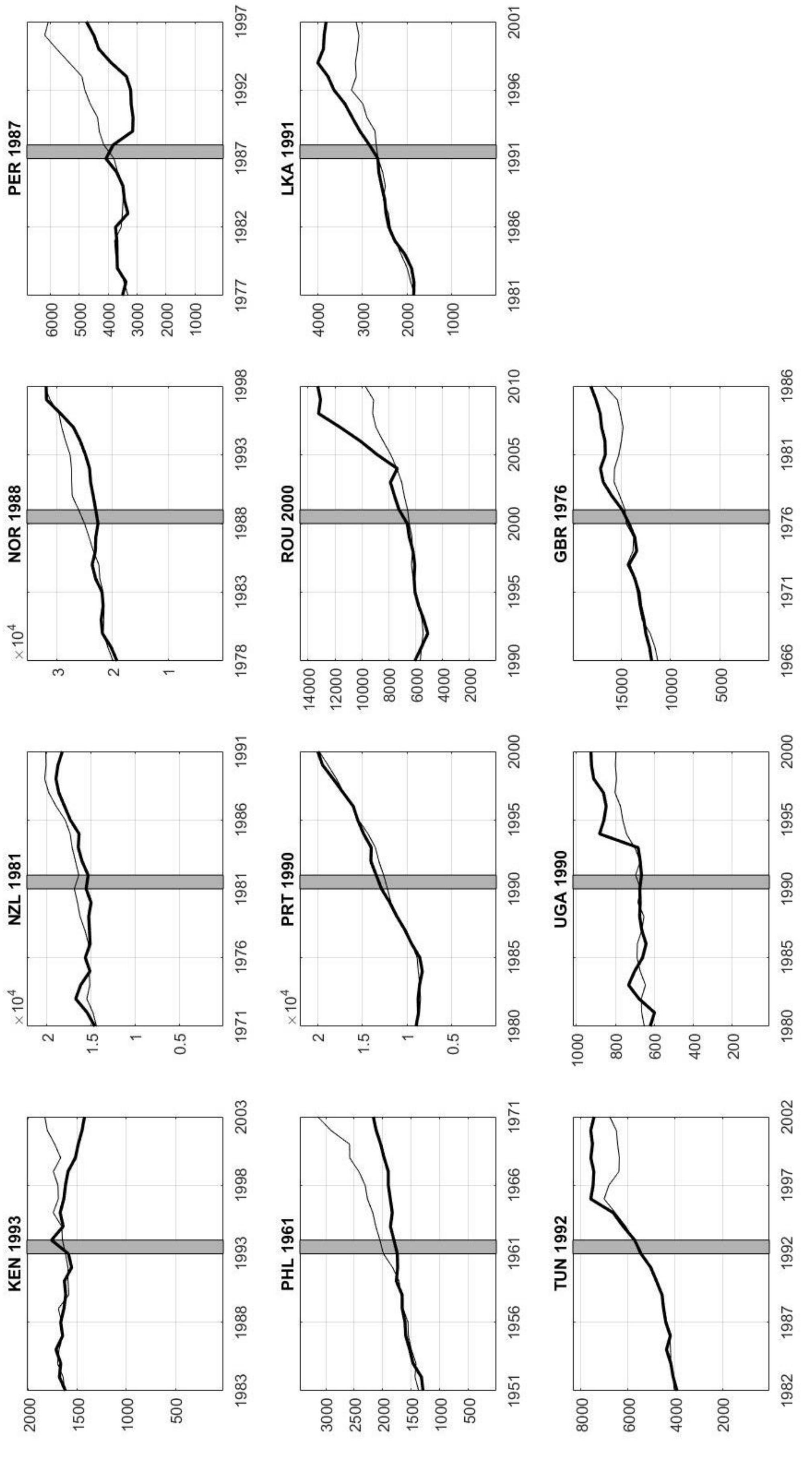
Estonia	1992	X	X									X	
Guatemala	1988	X	X		X								
Guatemala	2000				X					X	X		
Honduras	2001				X					X	X		
Hungary	1993	X	X									X	
Indonesia	1982	X	X	X									
Indonesia	2000				X				X	X	X		
Indonesia	2002								X	X	X		
Israel	2001									X	X	X	
Korea	2001				X					X			X
Madagascar	1986	X	X				X						
Nigeria	2002			X	X						X		
Peru	1986	X		X		X							
Peru	1989		X		X		X						
Poland	1995							X			X	X	
Slovakia	1999								X	X		X	
Sri Lanka	1989		X		X		X						
Uganda	1989	X			X		X						
Venezuela	1974		X		X	X							
Venezuela	1987	X	X	X									

Appendix 11. IMF Robustness check: (Panel A) Average trend in GDP per capita: reform countries versus synthetic control (Panel B) estimated impact of reforms



Appendix 12. Real (bold line) and synthetic (thin line) GDP per capita trends (in USD millions at 2005 prices)




PER 1987
LKA 1991
NOR 1988
ROU 2000
GBR 1976
NZL 1981
PRT 1990
UGA 1990
KEN 1993
PHL 1961
TUN 1992

Appendix 13. Matching table (Table 1) and Synthetic control composition (Table 2) for Dominican Republic (1989) and Belgium (1988)

Table 1. Matching table

	1979-1988		1978-1987	
	Dominican Republic	Synthetic Control	Belgium	Synthetic Control
GDP per capita	4149.3	4145.3	18838.7	18822.7
Investment rate	17.2	20.7	21.4	21.0
Industry share	34.2	34.2	22.4	22.4
Trade openness	60.5	60.5	91.6	46.9
Population growth	2.3	2.3	0.1	0.6
Secondary education	3.8	10.4	14.5	10.1
Tertiary education	1.3	4.3	8.2	4.6

Table 2. Synthetic control composition

	Dominican Republic		Belgium
in %	1989	*in %*	1988
Guatemala	32	France	90
Romania	19	Singapore	6
Panama	19	Greece	4
Philippines	18		
Zimbabwe	11		
Donor pool size	13		23
RMSPE	123		430

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily represent those of the institutions to which they are affiliated. We are especially grateful to Dani Rodrik for extensive discussions on earlier versions of this paper. We would also like to thank Alberto Abadie, Daron Acemoglu, Matteo Alpino, Omar Barbiero, Henrik Enderlein, Ricardo Hausmann, Frank Neffke, Jean Pisani-Ferry, Andrè Sapir, Nicolas Sauter, Julian Schumacher, together with participants of the ECB DGED Seminar, the IMF Jobs, Growth, and Structural Reforms Seminar, the OECD Brown Bag, and the Harvard Growth Lab Seminar for their helpful comments. We retain full responsibility for any remaining errors.

Pasquale Marco Marrazzo
European Central Bank, Frankfurt am Main, Germany; email: marco.marrazzo@ecb.europa.eu

Alessio Terzi (corresponding author)
John F. Kennedy School of Government, Harvard University, Cambridge, United States, Hertie School of Governance, Berlin, Germany; email: alessio_terzi@hks.harvard.edu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

ISSN 1725-2806 (pdf) DOI 10.2866/090439 (pdf)
ISBN 978-92-899-3024-6 (pdf) EU catalogue No QB-AR-17-123-EN-N (pdf)