



Evaluating the Economic Effects of Flat Tax Reforms Using Synthetic Control Methods

Bibek Adhikari and James Alm

WORKING PAPER 01/2017
January 2017

Working Papers in Public Finance



Chair in Public Finance
Victoria Business School

The Working Papers in Public Finance series is published by the Victoria Business School to disseminate initial research on public finance topics, from economists, accountants, finance, law and tax specialists, to a wider audience. Any opinions and views expressed in these papers are those of the author(s). They should not be attributed to Victoria University of Wellington or the sponsors of the Chair in Public Finance.

Further enquiries to:
The Administrator
Chair in Public Finance
Victoria University of Wellington
PO Box 600
Wellington 6041
New Zealand

Phone: +64-4-463-9656
Email: cpf-info@vuw.ac.nz

Papers in the series can be downloaded from the following website:
<http://www.victoria.ac.nz/cpf/working-papers>

Working Papers in
Public Finance



Tulane Economics Working Paper Series

Evaluating the Economic Effects of Flat Tax Reforms Using Synthetic Control Methods

| | |
|-------------------------|-------------------------|
| Bibek Adhikari | James Alm |
| Department of Economics | Department of Economics |
| Tulane University | Tulane University |
| badhikar@tulane.edu | jalm@tulane.edu |

Working Paper 1615
December 2016

Abstract

Tax reforms are often motivated by their potential to improve economic performance. However, their actual impacts are difficult to quantify. We analyze the impact of flat tax reform on incomes using “synthetic control” methods. We identify the 8 Eastern and Central European countries that adopted flat tax systems between 1994 and 2005, and then compare post-reform GDP per capita of “treated” countries with a convex combination of similar but “untreated” countries, while accounting for the time-varying impact of unobservable heterogeneity. We find positive impacts in all 8 countries, with 7 out of 8 cases significant at the conventional level.

Keywords: Flat tax, tax reform, synthetic control methods.
JEL codes: H20, H25, H31

Evaluating the Economic Effects of Flat Tax Reforms Using Synthetic Control Methods

Bibek Adhikari and James Alm*

Abstract

Tax reforms are often motivated by their potential to improve economic performance. However, their actual impacts are difficult to quantify. We analyze the impact of flat tax reform on incomes using “synthetic control” methods. We identify the 8 Eastern and Central European countries that adopted flat tax systems between 1994 and 2005, and then compare post-reform GDP per capita of “treated” countries with a convex combination of similar but “untreated” countries, while accounting for the time-varying impact of unobservable heterogeneity. We find positive impacts in all 8 countries, with 7 out of 8 cases significant at the conventional level.

Keywords Flat tax, tax reform, synthetic control methods.

JEL Classification H20, H25, H31.

* Bibek Adhikari, Department of Economics, 206 Tilton Hall, Tulane University, New Orleans, LA 70118 (email badhikar@tulane.edu). James Alm, Department of Economics, 208 Tilton Hall, Tulane University, New Orleans, LA 70118 (phone +1 504 862 8344; fax +1 504 865 5869; email jalm@tulane.edu). We are grateful to Stefano Barbieri, Raj Chetty, Keith Finlay, Michael Gallagher, Seth Giertz, Michael Keen, Nora Lustig, Ruud de Mooij, Victoria Perry, Steven Sheffrin, and Jay Shimshack for helpful comments and discussions and to Juan Pantano for sharing some codes. We are especially grateful to Alan Barreca, Michael Darden, and Dan Teles for very detailed and insightful suggestions. Earlier versions of this paper were presented at the November 2013 Annual Meetings of the National Tax Association in Tampa, FL, the January 2014 Allied Social Science Association Meetings in Philadelphia, PA, and the August 2015 seminar at the Fiscal Affairs Department of the International Monetary Fund in Washington, D.C.

1. Introduction

The adoption of a flat income tax system by Estonia in 1994 marked the beginning of what has become in two decades a global phenomenon. Before then, only a few countries like Hong Kong, Jamaica, Jersey, and Guernsey had a flat tax (Evans and Aligica, 2008). Within a decade, several other countries in Eastern and Central Europe had followed suit (Latvia, Lithuania, Russia, Serbia, Slovak Republic, and Ukraine). As of 2013, there are more than 30 jurisdictions with some form of a flat income tax, including countries in Latin America and the Caribbean (Grenada, Belize, Paraguay, and Trinidad and Tobago) and in Asia (Iraq, Mongolia, and Timore Leste), and more than 30 countries around the world have recently been considering the adoption of a flat rate income tax (Ji, 2013). Figure 1 documents the spread of the flat tax reforms around the world.

A major argument in the debate over flat tax reform is the effect that such a reform would have on economic performance. Hall and Rabushka (2007), the architects of one variant of the flat tax, argue that the growth effects of flat taxes are one of its major selling points:

“Improved incentives for work, entrepreneurial activity, and capital formation will substantially raise national output and the standard of living.”

However, demonstrating these effects from actual country experience has proved elusive.

This paper analyzes the impact of flat tax reform on economic growth using a novel data-driven empirical method, “synthetic control”. We identify the 8 Eastern and Central European countries that adopted flat tax systems between 1994 and 2005: Estonia, Latvia, Russia, Slovak Republic, Ukraine, Georgia, Romania, and Turkmenistan. We then compare the difference in GDP per capita before and after the reform for the “treated” (e.g., tax reform) country to a convex combination of similar but “untreated” (e.g., no tax reform) countries from Eastern and Central Europe. Overall, we find positive impacts of tax reform on the level of

income in all 8 countries, with these impacts significant at conventional levels in 7 out of 8 cases.

Use of synthetic control methods (SCM) to evaluate the impact of tax reform allows us to make several contributions. First, the standard “Difference-in-Differences” (DID) approach has often been applied to estimate the impact of tax changes on such variables as labor supply and savings where the units of analysis are households or firms, but it has not typically been used in evaluating the overall impact of comprehensive tax reforms because the units of analysis in comprehensive reforms are usually aggregate entities, like countries or states, for which suitable single comparisons often do not exist; also the choice of control countries to include in the comparison group are often ad hoc, creating an ambiguity on whether the comparison group really is an effective counterfactual. Synthetic control methods provide a systematic and transparent methodology to construct the appropriate counterfactual using a weighted average of potential comparisons units that are similar but untreated. Second, most studies on the impact of comprehensive tax reforms have been either theoretical or simulation-based, largely due to the estimation difficulties caused by the endogeneity of tax reforms (Jones, Manuelli, and Rossi, 1993; Pecorino, 1994; Stokey and Rebelo, 1995). The use of SCM helps control for endogeneity problems caused by omitted variables by accounting for the presence of time varying unobserved confounders.¹ Third, the extant literature on flat tax reforms is mostly a single country descriptive study.² We are able to present quantitative results for eight countries using the same methodology, which allows us to make cross country

¹ However, other sources of endogeneity like reverse causation may still be present even with SCM.

² For example, see the website of Alvin Rabushka (“Flat Tax – Essays on the Adoption and Results of the Flat Tax Around the Globe”), at <http://flattaxes.blogspot.com/>. An exception is Russia, where Gorodnichenko, Martinez-Vazquez, and Peter (2009) and Ivanova, Keen, and Klemm (2005) provide rigorous analyses of its flat tax reform. Kopczuk (2012) studies the impact of an optional flat tax reform in Poland using individual tax return data, and finds that the reform substantially improves revenue collection, most likely through decreases in tax avoidance and in underground activity. These results highlight the potential for an increase in economic performance by reducing distortions in the economy.

comparisons while also focusing on the unique circumstances of each country's tax reform. Furthermore, we aggregate country-specific treatment effects to obtain the average treatment effect of flat tax reforms, and we are also able to compare our SCM estimates with DID results. Finally, we contribute to the improvement of synthetic control methods by creating an index to normalize the measure that assesses the quality of fit produced by synthetic control method, so that we can compare the pre-treatment fit across countries of different income levels.

We find that all 8 countries exhibit positive impacts on GDP per capita, in comparison to their respective synthetic controls, with economically meaningful and statistically significant effects in most all cases. Our estimates range from \$96 in Georgia to \$2,558 in Turkmenistan. We also calculate dynamic average treatment effects for the region and for the group of countries that adopted the flat tax in the 1990s (the "1st wave" of reform) and the 2000s (the "2nd wave"), and we find that the average effects are positive and significant in both waves of tax reform. Five years after the treatment, the average GDP per capita of all treated countries, 1st wave countries, and 2nd wave countries are all about 18 percent higher than the respective synthetic group. These results are robust to a variety of alternative specifications and tests. Some possible mechanisms through which flat tax reforms affect economic performance are analyzed in our conclusions.

2. Defining and Assessing a "Flat Tax"

2.1. What Is a "Flat Tax"?

The flat taxes that have become widespread in Eastern and Central Europe typically feature a single tax rate, a broader tax base, and a simpler tax code (Keen, Kim, and Varsano, 2008). Accordingly, we study the overall impact of the flat tax reforms, and we do not

undertake the task of identifying the effects of these separate rate, base, and simplicity reforms.³ Note that the types of all comprehensive flat tax reforms that we examine include flattening of the tax rates, broadening of the tax base, and simplifying of the tax code. Thus, studying the overall impact of the tax reforms is a necessary first step.

These flat tax reforms have generally moved the tax systems closer to being a tax on comprehensive income. However, there is still great diversity in what is meant when talking about a “flat tax”. In this section we discuss the many types of tax systems that are often characterized as a flat tax. There are two main issues that are commingled in nearly all forms of a flat tax, relating to the choice of the tax rate structure and the choice of the tax base. All variants of the flat tax impose the tax at a single proportional marginal tax rate. However, the base upon which the tax rate is imposed often varies significantly.

The flat tax system as originally proposed by Hall and Rabushka in 1985 imposes a tax on businesses and individuals at the same single rate.⁴ The tax is essentially a two-part tax that taxes compensation of individuals and separately taxes businesses through a value added approach, and is more accurately called a “flat rate consumption tax” given that its tax base is consumption.

However, there are also flat rate taxes that are imposed at a single tax rate on an income base. It is this type of tax that has typically been imposed in Eastern and Central Europe. Such a

³ The calculation of the actual tax base in a country requires extensive data, and the separate identification is challenging even then. See Kopczuk (2005), Kawano and Slemrod (2012) for discussion of these difficulties even when examining Western European countries that have much better coverage of data at both the micro- and macro-levels than Eastern European countries.

⁴ Hall and Rabushka were not the first to suggest replacing the income tax with a consumption tax. Among those to whom flat tax ideas can be attributed are: Kaldor (1955), Friedman (1962), the Meade Commission (Institute of Fiscal Studies, 1973), Andrews (1974), and Bradford (1986). Several other versions of the consumption tax have been discussed. See Mieszkowski (1977) for a discussion of a specific cash-flow expenditure tax, and McLure and Zodrow (2007) point out that combining the treatment of expenditures on real assets with the tax prepaid treatment of financial transactions yields what the Meade Commission called the real business cash flow tax base, or “R-base,” and combining it with the tax postpaid approach yields the “real plus financial” business case flow tax base, or “R+F” base”. The “X-tax” proposal is attributed to Bradford (1986). Under certain circumstances (e.g., a uniform tax rate on all transactions), the present value of the R and R+F tax bases are equal.

tax is more accurately referred to as a “flat rate income tax”. Regardless, the many different variations that are somewhat loosely called the “flat tax” are seldom a pure “flat rate consumption tax” or a pure “flat rate income tax”. For example, Russia imposes a constant tax rate of 13 percent on an income base, but it also withdraws an allowance discontinuously as income rises. Georgia imposes a constant tax rate of 12 percent on labor income with no basic allowance. Further differences arise from how countries treat social security and other contributory pensions.

For the purpose of this study, we define a “flat tax” as a “flat rate income tax”. Specifically, we focus on any tax system under which the individual income tax is of the form $T(Y) = \max [t(Y-A), 0]$, where $T(Y)$ denotes tax liability on income Y , t is the single tax rate, and A is an allowance or exemption.

2.2. Assessing the “Flat Tax”

Advocates of a flat tax generally make two main arguments for its adoption. First, advocates argue that a flat tax will increase economic growth, for several reasons. One reason is that the consumption variant of a flat tax would increase saving and investment, especially if capital purchases by businesses are expensed in the year of their purchase. Now there is widespread agreement that taxing consumption will discourage consumption, encourage saving, and thereby increase economic growth. However, the magnitude of the saving increase, and therefore the magnitude of the capital formation increase, is unknown. The emerging view is that the growth effects would be positive and small. Regardless, these growth effects would arise only with a flat rate consumption tax, not a flat rate income tax. Another reason that economic growth might increase under a flat tax, whether income- or consumption-based, is that a flat tax might eliminate the huge array of tax preferences that distort individual and firm

choices. Again, the magnitude of this response is unknown. One last source of an increase in economic growth from a flat tax of any form is from the reduction in marginal tax rates. There is almost universal agreement that, at least in theory, individual and business behavior should respond positively to a reduction in marginal tax rates. However, it is in the area of the magnitude of these responses where there remains enormous – and unresolved – disagreement. (Auerbach and Slemrod, 1997).

The other major advantage of a flat rate consumption tax is that it would lower, perhaps substantially, the administrative and compliance costs of taxation. Most of the complexity of the current income tax stems from efforts to define, measure, and tax capital income. Because a consumption tax does not tax capital income, these difficulties are eliminated or at least reduced. Estimates of the administrative and compliance cost savings vary widely, but there is little doubt that some simplification would occur (Slemrod, 2005), although these gains can easily be overstated and they would not occur under a flat rate income tax.

As for critics of the flat tax, the main concern with any flat tax, whether consumption- or income-based, is its impact on the distribution of income. Critics of a flat tax argue that any flat tax will impose a heavier burden on low-income families than on high-income ones because there can be no redistribution via rising and progressive marginal tax rates and also because the fraction of annual income spent on consumption tends to be higher for low-income than for high-income families, and because any consumption tax effectively exempts from taxation capital and capital income. A related point is that a consumption tax will also redistribute tax burdens across generations, at least when enacted. Those who are or will soon be retired (whose income is almost entirely consumed) will experience a significant increase in tax liabilities.

The relevant trade-off is therefore simply stated: are the benefits of improved growth (as well as the efficiency gains from greater simplicity) worth the equity costs of the reallocation of tax burdens? Central to this discussion is the impact of a flat tax on economic growth. As emphasized by Keen, Kim, and Varsano (2008), much of the current debate about the flat tax is marked by rhetoric and assertion rather than by analysis and evidence. The next sections present our framework for disentangling the impact of flat tax reform on economic growth.

3. Synthetic Control Methods

3.1. Methods and Assumptions

In order to isolate the impact of flat tax reform from other influences, we employ synthetic control methods (SCM), a data-driven way of finding the “best” counterfactual in generalized Difference-in-Differences (DID) estimation. DID estimation consists of identifying a specific treatment (in our case, flat tax reform), and then comparing the difference in outcomes before and after the treatment for the treated country to the difference in outcomes before and after the treatment for the untreated countries. At the heart of the synthetic control method is the idea that the effect of a particular intervention can be empirically assessed only by comparison with the appropriate counterfactual: how would the trajectory of outcome variables evolve had the intervention not happened? The SCM was developed by Abadie and Gardeazabal (2003) and expanded by Abadie, Diamond, and Hainmueller (2010, 2014). Abadie and Gardeazabal (2003) use the approach to assess the impact on economic growth of the violent conflict in the Spanish Basque Country, Abadie, Diamond, and Hainmueller (2010) use it to estimate the impact of Proposition 99 (a large anti-tobacco initiative) on the per capita sales of cigarettes in California, and Abadie, Diamond, and Hainmueller (2014) use it to study

the economic impact of the 1990 German reunification in West Germany. The use of SCM has been growing, and it is now being applied to a very diverse set of topics.⁵

In DID studies, the researcher is allowed to choose the comparison group that mimics the counterfactual. However, Abadie, Diamond, and Hainmueller (2010) argue that this introduces substantial ambiguity about how comparison groups are chosen since researchers select comparison groups based on subjective measures of affinity between the treated country and the untreated countries.⁶ The main advantage of SCM is that it solves this problem by creating a “synthetic” control unit that resembles the “treated” unit in the pre-treatment period, using a weighted average of all other units in the donor pool. The weights are chosen so that the pre-treatment outcomes and the covariates of the synthetic control unit match on average the outcomes and covariates of the treated unit. After the synthetic control unit is chosen, the SCM then looks at the performance of the treated unit against the performance of the synthetic control unit after the intervention, where the difference in the outcomes of the treated unit and the synthetic control unit gives the “dynamic treatment effect”.

The other advantages of SCM are that it safeguards against extrapolation that plagues traditional regression models and that it can account for time variant country characteristics as long as there are enough pre-treatment periods and the pre-intervention match is strong, thus correcting for any potential endogeneity problem from omitted variable bias.⁷ Finally, Bertrand, Duflo, and Mullainathan (2004) show that standard errors in the papers that use DID estimation

⁵ These applications include: the effects of trade liberalization on growth (Billmeier and Nannicini, 2011; Billmeier and Nannicini, 2013); the impact of the terrorist attacks on electoral outcome (Montalvo, 2011); the effects of relaxing restrictions on home equity lending on retail spending by households (Abdalah and Lastrapes, 2012); the impact of natural disasters on economic growth (Cavallo et al., 2013); the effect of civil conflict on economic growth (Dorsett, 2013); the impact of nutrition policies on dietary behavior and childhood obesity (Bauhoff, 2014); the effect of immigration laws on demographic composition (Bohn, Lofstrom, and Raphael, 2014); the impact of decrease in police enforcement on traffic fatalities and injuries (DeAngelo and Hansen, 2014); and the impact of major natural resource discoveries on economic growth (Brock, 2015).

⁶ A similar argument is made by Angrist and Pischke (2009) and Montalvo (2011).

⁷ See Abadie, Diamond, and Hainmueller (2014) for details.

with many years of data severely understate the standard deviation of the estimators due to serially correlated errors. Use of SCM allows us to use placebo experiments to draw valid inference in the presence of correlated errors.

In the absence of experimental variation, quasi-experimental identification is not possible without identification assumptions. We make several assumptions that are common in program evaluation literature: the Conditional Independence Assumption (CIA), the common support assumption, no exogenous shocks, no interference between countries, no spillovers across countries, and no anticipation effects.

The CIA assumes that adoption of the flat tax reform and the potential outcomes of the reform are independent, conditional on a set of observed characteristics. This assumption is satisfied (partially) if there are countries in Eastern Europe and Central Asia that had similar GDP per capita (and other relevant characteristics), and did not adopt flat tax. As most of the tax reforms are motivated by expectation of future growth prospects, this assumption is not completely satisfied, thus leading to potential endogeneity from reverse causation. This would bias the estimates obtained from SCM as long as growth expectations are not captured by the unobservable heterogeneity included in the estimation.

The common support assumption requires that there exist untreated countries with similar characteristics to the treated country such that the weighted combination of untreated countries can reproduce the pre-treatment characteristics (GDP per capita and its predictors) of the treated country.

As for other assumptions, the absence of significant exogenous shocks assumes that the country did not experience any shocks of bigger magnitude than the tax reform. In order to satisfy this assumption, we remove countries that were known to have shocks of bigger

magnitude (e.g., civil violence); we also test for the validity of this assumption using “leave-one-out” tests. However, we still cannot rule out the possibility that our results are affected by other major events that occur simultaneously or after the tax reform, especially in the environment that characterized Eastern and Central European countries during the period of transition. For example, flat tax reform was accompanied by additional reforms of social insurance contributions, benefits, indirect taxation and tax administration, and their impacts are likely commingled. Therefore, we interpret the results as the overall impact of comprehensive tax reforms. Similar comments apply to many other reforms that were taking place, such as trade liberalization, exchange rate changes, and financial liberalization. To meet the assumption of no interference between countries, we exclude from the control group those countries that adopted the flat tax. If some of the countries with positive weights in the synthetic control undergo piecemeal tax reforms in response to the implementation of flat tax by our country of interest, then this will in fact tend to weaken our estimated treatment effect. Finally, the assumption of no anticipation effects implies that variables do not change in anticipation of the future tax reform, before the reform actually takes place.

Finally, the identifying assumption underlying our research design is not random assignment of flat tax reforms. Rather, the assumption is that the outcomes of countries with tax reforms would have trended similarly to the countries without tax reforms in the absence of the tax reforms. This identifying assumption is supported by the data because, as discussed later, these data show no systematic differences in the GDP per capita trends (after accounting for country and year fixed effects) between the treated and the control countries in the several years prior to the implementation of the tax reforms.

3.2. The Institutional Setting

The implementation of flat tax reforms by transition economies provides a unique setting to assess the impact of the flat tax reforms, given their institutions, for several reasons.

First, as argued by Romer and Romer (2010), Favero and Giavazzi (2012), and Cloyne (2013), tax reforms motivated by political or ideological reasons are more exogenous to the economic performance of the country than reforms motivated by the economic condition itself. In the specific case of flat tax reforms, the flat tax was enacted mainly through the efforts of right-wing parties and politicians in coalition governments.⁸

Second, as Case, Hines, and Rosen (1993) and Besley and Case (1995) demonstrate, the design of the tax structure in one jurisdiction influences the design of tax structure in neighboring areas. Thus, if countries in the Eastern and Central Europe adopt flat tax reform due to the influence of their neighbors, then such tax reform will again be more exogenous to the economic condition of the adopting country than a reform motivated by economic conditions. The rapid spread of the flat tax reforms in the Eastern and Central Europe clearly suggests such copycat behavior in the region.

Third, most of the flat tax adoptions went from enactment to implementation in a short period of time, which is useful for our identification strategy as it suggests that it is unlikely that the anticipation of the reform biases our estimates. For example, Estonia adopted its flat

⁸ For the five countries with a clear left-right ideological spectrum (Estonia, Latvia, Romania, Georgia, and Slovak Republic), it was the party on the right that introduced and drove the adoption of the flat tax. For example, in Estonia it was the Pro Patria Party, in Latvia the Latvian Way party, in Romania the Justice and Truth Alliance, in Georgia the National Movement-Democrats, and in Slovak Republic the SDKÚ-DS. In Russia and Ukraine, the usual left-right ideological spectrum is complicated by the practice of forming parties around a key personality. Even there, ideology appeared to be one of the more important factors. In Russia, the flat tax was part of the overhaul of the tax system by the Putin administration with a decidedly liberal prime minister and team of economic advisers; in Ukraine, the strong performance of the right-wing party Our Ukraine Party in the 2002 elections and its control of the Budget Committee's agenda suggests that the flat-tax legislation resulted from the efforts of economically liberal politicians who held powerful positions in the legislature due to the strong electoral performance of economically right-wing parties in the preceding election. See Castanheira, Nicodeme, and Profeta (2010), Ellis (2010), and Appel (2011) for detailed discussions. See also Baturo and Gray (2009) and Ji (2013) for empirical evidence that right-leaning governments were significantly more likely to implement a flat tax than left-leaning governments.

rate tax in less than two months, passing the law at the end of 1993 and implementing it in January 1994. In Romania, the new government led by President Trajan Basecu took power on 28 December 2004, and the flat tax was introduced within 2 days so that it could be implemented in January 2005. The Georgia flat tax legislation was approved in December 2004 and implemented on January 2005, and the Slovak Republic legislation was approved in October 2003 to take effect in January 2004. Russia and Ukraine had a relatively longer time between adoption and implementation, but even there the lag was less than 7 months.⁹

3.3. A Motivating Model¹⁰

Suppose that we have $J+1$ countries, where country 1 adopts flat tax reform at time $[T_0 + 1]$ (a “treated” country) and the remaining J countries act as potential controls called the “donor pool”. Let T_0 be the number of pre-intervention periods, with $1 \leq T_0 < T$. Also, let Y_{it}^{NR} be the outcome variable observed for country i at time t with no reform (NR), and Y_{it}^R be the outcome variable with reform (R). The observed outcome variable can be written as:

$$Y_{it} = \begin{cases} Y_{it}^{NR} & \text{in the absence of tax reform} \\ Y_{it}^R \equiv Y_{it}^{NR} + \tau_{it} D_{it} & \text{in the presence of tax reform,} \end{cases}$$

where $\tau_{it} = (Y_{it}^R - Y_{it}^{NR})$ is the effect of the reform for country i at time t and $D_{it} = 1$ if $t > T_0$ and $i = 1$ and $D_{it} = 0$ otherwise.

For any treated country, we can observe Y_{it}^R . However, we need to estimate the counterfactual Y_{it}^{NR} , which is the level and/or the growth rate of GDP per capita of the country that adopted flat tax reform had the country not adopted it. In order to estimate the counterfactual, we use the linear factor model of the form:

⁹ As discussed later, we empirically test for the presence of anticipation effects, and find no evidence for them.

¹⁰ This discussion is based on Abadie, Diamond, and Hainmueller (2010, 2014).

$$Y_{it}^{NR} = \alpha_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it},$$

where α_t is an unknown common factor with constant factor loadings across countries, Z_i is a vector of observed covariates with coefficients θ_t , μ_i is a $(1 \times F)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, and ϵ_{it} are idiosyncratic error terms with zero mean. Note that this specification allows the effects of confounding unobserved characteristics to vary with time ($\lambda_t \mu_i$), unlike in conventional difference-in-differences that allows for the presence of unobserved confounders but restricts the effects of those confounders to be time invariant ($\lambda \mu_i$).

Define a synthetic control unit as a weighted average of the units in the donor pool. That is, a synthetic control can be represented by a $J \times 1$ vector of weights, $\mathbf{W} = (w_2, \dots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \dots, J + 1$ and $w_2 + \dots + w_{J+1} = 1$, where vector \mathbf{W} represents a potential synthetic control. Then the outcome variable for each potential synthetic control unit is given by:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \alpha_t + \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_i + \sum_{j=2}^{J+1} w_j \epsilon_{jt}$$

Now suppose that there are $(w^*_2, \dots, w^*_{J+1})'$, such that:

$$\sum_{j=2}^{J+1} w^*_j Y_{j1} = Y_{11}, \dots, \sum_{j=2}^{J+1} w^*_j Y_{jT_0} = Y_{1T_0} \text{ and } \sum_{j=2}^J w^*_j Z_j = Z_1 \text{ holds.}$$

Abadie, Diamond, and Hainmueller (2010) prove that if there are weights such the weighted outcome variable of the donor pool countries (e.g., the “synthetic control”) in pre-treatment matches the pre-treatment outcome variable of the treated country, then the size of the bias caused by time varying unobserved confounders in the difference between the post-treatment outcome variable for the treated and the synthetic control countries goes towards zero

as the pre-intervention period increases. Importantly, they show that this holds even for the case of an imperfect but strong fit. Thus, the treatment effect at time $t \in \{T_0+1, \dots, T\}$ can be estimated by:

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

Note that we can aggregate dynamic treatment effects in various ways to illustrate the impact of tax reform. We present two such averages below. First, aggregating across countries of group G (for example, the 1st wave countries) gives the average treatment effect for group G at time $t \in \{T_0+1, \dots, T\}$, or:

$$ATE_t^G = \frac{\sum_{g=1}^G \hat{\tau}_{1t}^g}{G}$$

Second, aggregating across post-treatment years and subtracting from it the amount obtained by aggregating across pre-treatment years for country g gives the average treatment effect for country g , or:

$$ATE = \frac{\sum_{t=T_0+2}^T \hat{\tau}_{1t}}{T - T_0 - 1} - \frac{\sum_{t=1}^{T_0} \hat{\tau}_{1t}}{T_0}$$

To find the optimal weights, let the $(T_0 \times 1)$ vector $K = (k_1, \dots, k_{T_0})'$ define a linear combination of pre-intervention outcomes, $\bar{Y}_j^K = \sum_{s=1}^{T_0} k_s Y_{js}$, where j belongs to $\{1, \dots, J+1\}$, $j=1$ denotes treated country and $j \neq 1$ denote donor countries. Consider M of such linear combinations defined by the vectors (K_1, \dots, K_M) . Let X_1 be $(Z_1', \bar{Y}_1^{K_1}, \dots, \bar{Y}_1^{K_M})'$, the vector of pre-treatment variables that we aim to match as closely as possible for the treated country. Let X_0 be the matrix where each column of the matrix is a vector of same pre-treatment variables

for each potential donor country. The synthetic control algorithm chooses W^* to minimize the distance $\|X_1 - X_0W\|_V = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)}$, where V is a symmetric, positive semi-definite and diagonal matrix such that the mean square prediction error (RMSPE) of the outcome variable is minimized for the pre-intervention periods. Note that the inferential procedure is valid for any choice of V ; however, by minimizing the RMSPE, the algorithm assigns larger weights to those pre-treatment variables that have the highest predictive power.¹¹

3.4. Pre-treatment Fit Index

To assess whether the comparison country created using SCM is a good counterfactual, we need some measure of how well it mimics the treated country before the treatment. Abadie, Diamond, and Hainmueller (2010) use root mean square prediction error (RMSPE) of the outcome variable to measure fit or lack of fit between the path of the outcome variable for treated country and its synthetic counterpart, defined as:

$$\text{RMSPE} = \sqrt{\frac{1}{T_0} \sum_1^{T_0} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2}$$

However, we develop a new measure called the “pre-treatment fit index” to assess the overall quality of the pre-treatment fit. One advantage of using the pre-treatment fit index rather than RMSPE is that it normalizes RMSPE, which makes it possible to compare the fit between the synthetic control method across different outcome variables and different countries, such as when GDP per capita varies quite significantly across the sample countries. A second advantage is that this approach provides an index number that makes assessing the quality of fit very intuitive.

Define the benchmark RMSPE as the RMSPE obtained from the zero fit model, or:

¹¹ See Appendix V in Abadie and Gardeazabal (2003) for details.

$$\text{benchmark RMSPE} = \sqrt{\frac{1}{T_0} \sum_1^{T_0} (Y_{1t})^2}$$

Then the pre-treatment fit index is defined as the ratio of the RMSPE and the benchmark RMSPE:

$$\text{Fit Index} = \frac{\text{RMSPE}}{\text{benchmark RMSPE}}$$

Thus, the range of the index is $[0,U]$, where U is the finite upper bound. The index is defined as long as outcome variable is not 0 for all the pre-treatment period. A fit index of $[0,X]$ means that the RMSPE is equivalent to the RMSPE obtained when the difference between the treated and the synthetic unit is X percent on each pre-treatment year. If the RMSPE is 0, then the fit index will be 0, indicating a perfect fit; if the RMSPE is equal to the benchmark RMSPE, then the fit index will be 1, indicating that the fit is equal to that created by a zero fit model.¹² The fit index will be greater than 1 in cases where the outcome variable of the treated unit is bigger (or smaller) by a magnitude of two or more (or half or less) than that of the synthetic control.¹³

We illustrate the advantage of using the pre-treatment fit index rather than the RMSPE using a numerical example. In Table 1, the first pair of columns is the standard GDP per capita growth rate, the second pair of columns is typical GDP per capita of a low income country, and the third pair of columns is typical GDP per capita of a high income country. We vary the synthetic counterpart's outcome variable by 4 percent in all three cases. Note that the RMSPE

¹² In this sense, the calculated value of $[1 - \text{Fit Index}]$ provides similar information to the information provided by the R^2 statistic in regression analysis.

¹³ Thus, although our fit index has a finite upper bound that in theory cannot be calculated, it can be empirically calculated by the ratio of the outcome variable between the country with the highest GDP per capita trend and the lowest GDP per capita trend in the sample. In any case, a fit index greater than 1 indicates a poor fit, such that this synthetic unit should be discarded, especially when the outcome variable is GDP per capita where it is safe to say that a synthetic country with twice (or half) the GDP per capita of the treated country cannot reasonably be a good counterfactual. We consider the pre-treatment fit as good if the index is less than or equal to 0.10.

varies widely making it impossible to compare the pre-treatment fit across variables or countries. However, the fit index is 0.04 in all three cases, indicating that in all three cases the fit is equivalent to a 4 percent difference between the treated unit and the synthetic unit in each pre-treatment period.

3.5. Inference

In a comparative case study where identification of the treatment effect arises from the change in policy by a small group of countries and where data are usually of small sample size, standard large-sample approximations that are typically used for inference are not appropriate. However, placebo experiments as used in Abadie and Gardeazabal (2003), Bertrand, Duflo, and Mullainathan (2004), Abadie, Diamond, and Hainmueller (2010), and Chetty, Looney, and Kroft (2009) can be used to evaluate the significance of treatment effects.

The essence of placebo experiments is to test whether the estimated impact of the flat tax reform could be driven entirely by chance. Specifically, we conduct a series of placebo experiments by iteratively estimating the “placebo” treatment effect for each country in the donor pool (i.e., untreated countries) by first assuming that these countries implemented a flat tax reform in the same year as our country of interest and by then running the synthetic control method. In selecting a synthetic control group for the countries in the placebo pool, we omit the treated countries.

This iterative procedure provides a distribution of estimated placebo treatment effects for the countries where no intervention took place. If the placebo experiments create enough placebo treatment effects of magnitude greater than the one estimated for the treated country, then we conclude that there is no statistically significant evidence of an effect of flat tax reform in the treated country. If the placebo experiments show that the treatment effect estimated for

the treated country is unusually large relative to placebo treatment effects for countries that did not implement a flat tax system within our sample period, then we conclude that there is a statistically significant evidence of an impact of flat tax reform in the treated country.

Since we are making inferences about positive treatment effects, the p-value for country g for each post-intervention year $t \in \{T_0+2, \dots, T\}$ is given by:

$$\text{p-value}_t = Pr(\hat{\tau}_{1t} < \hat{\tau}_{jt}) = \frac{\sum_{j=2}^{J+1} 1(\hat{\tau}_{1t} < \hat{\tau}_{jt})}{J},$$

where $1(\cdot)$ is the indicator function, $\hat{\tau}_{jt}$ is the difference in the outcome variable between unit j and its synthetic control at time t and $j \in \{1, \dots, J+1\}$, $j = 1$ denotes the treated unit, and $j \neq 1$ denote placebo units. However, when making inferences about the average treatment effect for group G at time t , we need to account for the fact that the average reduces some noise. To do so, we follow Cavallo et al. (2013) and compute p-values for the average treatment effect for group G at each post-intervention year t according to the following steps. First, for each country i of group G , we compute placebo effect for all J_i+1 placebo units from the placebo pool. Second, at each post-treatment year, we compute every possible placebo average effect by picking a single placebo estimate corresponding to each country i and then taking the average across the G placebos. This results in $N_{\overline{PL}}$ placebo averages where $N_{\overline{PL}} = \prod_{i=1}^G J_i$. We index these placebo averages by $np = 1, \dots, N_{\overline{PL}}$.¹⁴ Third, at each post-treatment year, we calculate the p-values using:

$$Pr(\bar{\tau}_{gt} < \bar{\tau}_{gt}^{PL}) = \frac{\sum_{np=1}^{N_{\overline{PL}}} 1(\bar{\tau}_{gt} < \bar{\tau}_{gt}^{PL})}{N_{\overline{PL}}},$$

¹⁴ Note that this number grows rapidly when there are many countries in the group or when there are many placebo units for each country in the group, and especially when both are true.

where $1(\cdot)$ is the indicator function, $\bar{\tau}_{gt}$ is the average treatment effect for group G at time t after the treatment, and $\bar{\tau}_{gt}^{PL}$ is the placebo average treatment effect for group G at time t after the treatment.

Finally, we define p-value to assess the significance of the country-specific average treatment effect, which is the ratios of countries that produced an average “placebo treatment effect” greater than the average treatment effect, and we interpret each as the probability of obtaining an estimate larger than the one obtained for the treated unit. The resulting p-value for country-specific average treatment effect is given by:

$$Pr(\hat{\tau}_1^{Post} - \hat{\tau}_1^{Pre} < \hat{\tau}_j^{Post} - \hat{\tau}_j^{Pre}) = \frac{\sum_{j=2}^{J+1} 1(\hat{\tau}_1^{Post} - \hat{\tau}_1^{Pre} < \hat{\tau}_j^{Post} - \hat{\tau}_j^{Pre})}{J},$$

where $1(\cdot)$ is the indicator function, $\hat{\tau}_j^{Post}$ and $\hat{\tau}_j^{Pre}$ are the average difference in the outcome variable between unit j and its synthetic control after and before the treatment, respectively, and j belongs to $\{1, \dots, J+1\}$, where j = 1 denotes treated unit and $j \neq 1$ denote placebo units. Note that, if $\hat{\tau}_1^{Post} - \hat{\tau}_1^{Pre}$ is negative, then the inequality sign is reversed.

For some countries, GDP per capita during the pre-treatment period cannot be well reproduced by a convex combination of per capita GDP in donor countries.¹⁵ If the synthetic control does not fit GDP per capita before the treatment, then one would be misguided in interpreting the gap between the real and the synthetic unit as the effect of flat tax reform because the gap is likely created simply by the lack of fit than by the tax reform. Similarly, placebo countries that do not have a good pre-treatment fit cannot provide information to measure the relative rarity of estimating a large treatment effect (Abadie, Diamond, and Hainmueller, 2010, 2014). Therefore, we exclude placebo countries with a “bad” pre-treatment

¹⁵ This could happen if one of the countries in the donor pool has either the highest outcome variable or the lowest. If this is the case, then no combination of countries in the sample can reproduce the time series of per capita GDP in the specific country before the treatment.

fit when graphing them in placebo experiment graphs and when calculating p-values. For our baseline results, any placebo countries with fit index greater than the treated country's fit index is considered to have a "bad" pre-treatment fit.¹⁶

4. Data and Sample

4.1. Data

We use panel data obtained from several sources. Our main dependent variable is annual data on PPP converted GDP per capita (Laspeyres method).¹⁷ The set of covariates we use are based on the analyses of endogenous growth models (Barro, 1991; Mankiw, Romer, and Weil, 1992), and are meant to capture the impact of human capital, physical capital, and institutions on economic growth. Data on GDP per capita are obtained from the Penn World Tables (PWT) Version 7.1 (Heston, Summers, and Aten, 2012). We also use information on investment as a share of GDP and trade as a share of GDP (openness) from the Penn World Tables. Information on average years of schooling for individuals aged 15 and above is obtained from the Barro and Lee (2012) dataset on educational attainment for the period 1950-2010. Data on inflation and population growth are obtained from World Development Indicators (World Bank 2013). Following Persson and Tabellini (2007), we classify a country as democratic if the "polity2" variable in Polity IV dataset (Marshall and Jaggers, 2013) is strictly positive. Polity IV contains unbalanced panel data on the types and qualities of government of 167 countries spanning the years 1800 to 2012. Data on the average income tax rates are obtained from the World Tax

¹⁶ As discussed later, we test whether our inference is sensitive to the use of threshold for a "bad fit", and we find that our inference is robust to the choice of the threshold.

¹⁷ Note that we also use as a dependent variable PPP converted investment per capita, the stock of foreign direct investment (FDI) as a share of GDP, and the ratio of employment to population when exploring the mechanisms through which flat tax reforms affect economic performance. Data sources for these variables are described in section 6.

Indicator Database from the International Center for Public Policy at Georgia State University (Peter, Buttrick, and Duncan, 2010).

Our main explanatory variable is the introduction of a flat tax reform. We obtained this information from Keen, Kim, and Varsano (2008) and from the Rabushka website.¹⁸ We refer to the adoption of a flat tax system as the “Treatment”, so the treatment effect is intended to capture the causal impact of tax reform on GDP per capita.

We define all countries from Eastern and Central Europe that adopted flat tax systems between 1994 and 2005 in our data set as the treated countries (other than Lithuania and Serbia, due to data limitations and unrest).¹⁹ These countries (and the year of their tax reform) include: Estonia (1994), Latvia (1997), Russia (2001), Slovak Republic (2004), Ukraine (2004), Georgia (2005), Romania (2005), and Turkmenistan (2005). Note that we limit our sample to countries that implemented a flat tax reform on or before 2005 because, in order to isolate the impact of tax reform on growth, we need to observe the outcome variable for several years after the tax reform. We include data through 2010 for our estimation.

Figure 2 plots the trends in GDP per capita for our 8 treated countries, the rest of the Eastern Europe and Central Africa, and the rest of the world. The rest of the world (or even the rest of the region) may not provide a suitable comparison group for the treated countries because the treated countries and the averages of untreated countries follow different trends in per capita GDP prior to the reform. In the next section we construct synthetic controls that can accurately reproduce the pre-treatment GDP per capita trends.

4.2. Constructing Synthetic Controls

¹⁸ Again, see <http://flattaxes.blogspot.com/>.

¹⁹ We excluded Lithuania (1994) from the study because there was only 1 year of pre-treatment observation of GDP per capita available for Lithuania. We excluded Serbia (2003) because of the Kosovo violence of 1998 and the division of Serbia into Serbia and Montenegro in 2006. Both events are arguably exogenous shocks of a larger magnitude than any comprehensive tax reform.

To implement synthetic control methods, the data must be set up such that they meet several requirements. First, the data must be a balanced panel with all the countries having the same number of years of data available. Second, there cannot be any missing outcome variables for any country. Third, there must be at least one non-missing observation before the treatment for each of the covariates.

These requirements largely determine the pool of feasible control groups and the range of years used in the analysis. The range of years is determined by the availability of outcome variable for the country of interest. When data availability is not an issue, we restrict the first year for the donor pool to 1993 and the post-treatment period to 10 years to evaluate the impact of treatment. The choice to start the donor pool from 1993 is motivated by the dissolution of the Union of Soviet Socialist Republics and the Socialist Federal Republic of Yugoslavia in 1991-1992. This allows us to include countries from these unions in the potential donor pool since most of them start reporting GDP per capita from 1993 onwards. The choice for 10 post-treatment years is motivated by the tradeoff between having a large number of countries in the donor pool and having a long enough post-intervention period to analyze the impact of tax reform. In order for a country to be in the potential donor pool, the country must have at least one observation before the treatment for all the covariates used in the estimation. Any country not having such data is dropped.²⁰ Further, we exclude from the potential donor pool any country that adopted a flat tax reform before or within the sample period because the synthetic unit is meant to reproduce the trend in outcome variable that would have been observed for the treated unit in the absence of treatment. Including any country in the donor pool that was

²⁰ An alternative would be to drop the covariate itself. To make cross-country comparisons comparable, we only drop the covariate if the country of interest does not have a single observation before the treatment for the relevant covariate. For example, we had to drop the variable average schooling when running SCM on Georgia and Turkmenistan because of the unavailability of schooling data for these two countries.

treated in the time period implies that the synthetic unit is not reproducing the potential outcome in the absence of treatment. We also exclude countries from the donor pool if they do not have an income tax system in the pre-treatment year.

When calculating the average treatment effect, we restrict the donor pool to the countries of Eastern Europe and Central Asia only. The advantages of restricting potential controls within a region are that it avoids biases caused by interpolating across regions with very different characteristics and also that it controls for unobservable regional characteristics associated with, say, geography, the level of economic and political development, culture, and any other secular changes over time. It is particularly important to restrict the donor pool to Eastern Europe and Central Asia, where most countries were making their transitions at the same time, so that we can remove any impact of the transition *per se* from our estimation. However, running the placebo experiment only on a regional donor pool leaves us with a small number of countries, which means that the size and the power of the test will be small.²¹ Therefore, we run the placebo experiment using a worldwide sample of countries to make our results more robust. However, each placebo treatment effect is calculated controlling for unobservable regional characteristics, achieved by restricting the donor pool of each placebo units to their respective region.²² Put differently, we calculate regional placebo treatment effects for the global pool of countries, and we then compare them with the regional treatment effect of the treated country.²³

²¹ As discussed later, we present results obtained using a regional donor pool. These results are very similar to the results obtained using the hybrid approach.

²² Countries are divided into following regions: Eastern Europe and Central Asia, Latin America and Caribbean, Middle East and North Africa, Small Islands, South Asia and Pacific Island, Sub Saharan Africa, and Western Europe and offshoots.

²³ A similar strategy is used by Billmeier and Nannicini (2013) when conducting a placebo experiment. They run two iterations of the experiment, one with the regional donor pool and the other with the global donor pool. However, when calculating the placebo treatment effect using their global pool, they do not restrict the donor pool of each placebo unit to their respective region. Billmeier and Nannicini (2013) face the trade-off between the

Following Abadie, Diamond, and Hainmueller (2010), each observation of the relevant outcome variable before the treatment is used as a separate control variable in order to improve the pre-treatment fit.²⁴ For the remaining covariates, only the averages before the treatment are matched, while ignoring the missing observations when taking the average.²⁵

4.3. An Illustrative Example

We illustrate the application of SCM using Estonia as an example. Estonia began the modern wave of flat tax reform in 1994, introducing a flat rate income tax at a rate of 26 percent. The sample period for Estonia is 1990 to 2004, or 4 pre-treatment years and 10 post-treatment years. Lithuania is not included in the donor pool because it implemented a flat tax in the same year, and Latvia, Russia, Slovak Republic, and Ukraine are excluded because they implemented the flat tax within 10 years from Estonia. Nine other countries from the region are not in the donor pool because they had at least one missing outcome variable before 1994. Similarly, six more countries are excluded from the donor pool because they did not have single observation on inflation before the treatment. This leaves us with seven countries (Albania, Bulgaria, Croatia, Hungary, Poland, Romania, and Slovenia) in the donor pool.

The SCM algorithm finds the synthetic country as a linear combination of countries in the donor pool, such that two conditions are fulfilled: the chosen synthetic country matches the values of the variables with highest predictive power as closely as possible with the values of the same variables of the treated country, and the RMSPE of the outcome variable is minimized. The weights to each of the covariates can range from 0 to 100 percent but must add

common support attained by using the regional pool versus the power of the test attained by using global pool. We believe that our method eliminates this trade-off.

²⁴ As discussed later, we present results where we match on the pre-treatment average of GDP per capita as opposed to all realizations of the GDP per capita. The results are very similar to the ones obtained here.

²⁵ This maximizes the sample of countries in the donor pool and also improves the pre-treatment fit by relaxing the constraint so that only averages (as opposed to all realizations) of the covariates before the treatment are matched.

up to 100 percent. The most predictive covariates (with weights shown in parenthesis) are GDP per capita in 1991 (34.8 percent), GDP per capita in 1992 (28.4 percent), GDP per capita in 1990 (21 percent), openness (14.3 percent), population growth rate (1.07 percent), and average schooling (0.03 percent). Other realizations of pre-treatment GDP per capita, investment share, democracy dummy, and inflation have negligible predictive power, and thus are assigned 0 percent weights. Individual countries in the synthetic unit and their contribution (which ranges from 0 to 100 percent and adds up to 100 percent) are Bulgaria (52.6 percent), Croatia (32.4 percent), Hungary (13.1 percent) and Romania (1.9 percent). The remaining countries in the donor pool are assigned a 0 percent weight by the algorithm. Put differently, the weights reported above indicate that GDP per capita trends in Estonia prior to the adoption of flat tax reform is best reproduced by a combination of Bulgaria, Croatia, Hungary, and Romania. The resulting RMSPE is 0.5286 (or a 0.059 fit index), a very good fit.

The differences (shown in parentheses) in the pre-treatment average of covariates between the actual Estonia and the synthetic Estonia are close to zero for GDP per capita (-0.14 percent) and population growth rate (0.33 percent). Another important predictor, trade as a share of GDP (openness), is only 2.3 percent lower (96.6 percent versus 98.9 percent). However, average inflation is 210 percent lower (90 percent versus 299.7 percent) in actual Estonia than in its synthetic unit. Note that the extremely high inflation rate and its bad pre-treatment match appear in most of our eight countries; the pre-treatment match for the rest of those covariates that had negligible predictive power (democracy dummy, inflation, and average schooling in the case of Estonia) is also good. However, the quality of their match is not necessary to the estimation, and the match is not always good for other treated countries.

These impacts are illustrated in Figure 3. In the left panel, the vertical dashed line indicates the treatment year. The difference between the solid line (GDP per capita of Estonia) and dashed line (GDP per capita of the synthetic unit) before the treatment year indicates the quality of the fit; the fit index is also reported in the Figure 3. The same difference after the treatment year indicates dynamic treatment effects. The probability of observing a placebo treatment effect of greater magnitude than the treatment effect at time t is also graphed in the left panel of Figure 3. This estimate is almost significant at the 10 percent level 3 years after the treatment and is significant for all 7 years after that, which indicate a substantial and positive impact of flat tax reform on GDP per capita for Estonia compared to its synthetic counterpart.

The time series graph of the placebo experiments is presented in the right panel of Figure 3. The dashed vertical line denotes the treatment year, the black solid line indicates the regional treatment effect for the treated country, and dashed gray lines indicate regional placebo treatment effects of global sample. Out of 68 placebo units, 63 are consistently below the solid black line, indicating that most of the placebo treatment effects are smaller in magnitude than the true treatment effect. The p-value for average treatment effect is 0.06, implying that the average treatment effect is significant at the 10 percent significance level.

5. Results

In this section we first present results for each of the 8 countries that we examined, discussing each country separately. We then present aggregate results for all countries, including countries that adopted the flat tax in the 1990s and countries that adopted flat tax in 2000s.²⁶ Table 2 presents personal income tax rates and corporate income tax rates before and after the tax reform, as well as the change in allowances accompanying the reform. In the

²⁶ The discussion on the properties of flat tax reforms is mainly based on Keen, Kim, and Varsano (2008).

discussion for each country, Table 3 presents the details of the country-specific data selection, the fit index, the average treatment effects, and their p-values; Figure 4 presents dynamic treatment effects and their p-values; and Figure 5 presents a time series graph of the placebo experiments for all 8 countries. We finish with some robustness tests.

5.1. Country Results

Estonia (1994). Estonia began the modern (or 1st) wave of flat tax reform in 1994, introducing a flat rate income tax at a rate of 26 percent. Before the reform, Estonia had a progressive income tax with rates that ranged from 16 percent to 33 percent. Estonia reduced the corporate tax rate from 35 percent to 26 percent and also removed withholding tax on dividends. The tax rate has been lowered since 1994, and was 21 percent in 2011. The personal allowance was set at EEK 3600.

There was a substantial positive impact of reform on GDP per capita (\$1845 higher on average, significant at the 10 percent significance level). See Table 3 and Figures 4 and 5.

Latvia (1997). Latvia introduced a flat income tax in 1997 at the rate of 25 percent, and as of January 2011 the rate was still 25 percent.²⁷ The corporate income tax (CIT) rate at the time of reform was 25 percent, later reduced to 15 percent. The value of personal allowances was approximately 19 percent of per capita income. Dividend and interest income are exempt. Latvia had a progressive income tax with rates that ranged from 15 percent to 35 percent before the reform. From 1995 to 1997 Latvia had a regressive tax schedule, with marginal tax rates starting at 25 percent and falling to 10 percent for the highest income bracket.

²⁷ There is some confusion regarding the exact date of the flat tax implementation in Latvia, either 1995 or 1997. However, according to Keen, Kim, and Varsano (2008), Latvia reformed its income tax in 1994 with the basic rate at 25 percent and the top marginal rate of 35 percent, which later was reduced to 10 percent. It was only in 1997 that Latvia moved to a single tax rate from the preceding two tax rate system.

We observe positive impacts on Latvian GDP per capita compared to the synthetic country immediately after the flat tax reform. Two out of ten dynamic treatment effects are significant at the 5 percent level, and five more are significant at the 10 percent significance level. The pre-treatment fit is extremely good (0.010), and there is steady growth of GDP per capita in Latvia after the treatment. Latvia's GDP per capita was \$1966 higher on average than the synthetic unit (significant at the 10 percent level).

Russia (2001). The flat tax in Russia was introduced in 2001 at a rate of 13 percent on income, a significant reduction from pre-reform marginal tax rates that ranged from 12 percent to 30 percent. The introduction of the flat tax in Russia started a second wave of flat tax reforms, reforms that are still continuing. Note that the amount of allowances allowed decreases discontinuously as income increases, meaning that the effective marginal tax rate exceeds 13 percent and rises with income. Even so, the marginal rate of 13 percent above the allowance level makes it a flat tax at least above the allowance's threshold.

There is a good pre-treatment fit between Russia and its synthetic control (0.059). The dynamic treatment effects are positive, becoming significant at the 10 percent level from 2004 onwards and at the 5 percent level from 2006 onwards. On average, Russia's GDP per capita was \$2010 higher (significant at the 5 percent level) than the synthetic unit.

Slovak Republic (2004). The Slovak Republic adopted its flat tax in 2004, setting the rates for the individual income tax, the corporate income tax, and value-added tax all at the rate of 19 percent, and also removing the withholding tax on dividends. In January 2013, the Slovak Republic's newly re-elected government replaced the flat tax with a progressive income tax as part of an austerity package.

The pre-treatment fit is very good (0.020), and there is a positive and statistically significant impact of tax reform (after some lag). Average GDP per capita increased by \$1183, significant at the 10 percent level.

Ukraine (2004). Ukraine introduced a flat tax in 2004 at a rate of 13 percent. Before the reform, the marginal tax rates on income ranged from 10 percent to 40 percent. Ukraine later increased the flat rate to 15 percent in 2007 (Keen, Kim, and Varsano, 2008), and in January 2011 it added an additional 17 percent tax rate on high incomes.

The pre-treatment fit is good at 0.090, and the effect of tax reform on GDP per capita is positive (with 4 out of 6 lead specific p-values significant at the 10 percent level). The average treatment effect was \$1363, significant at the 10 percent level.

Georgia (2005). Georgia implemented its flat tax in 2005 at the rate of 12 percent, and also eliminated allowances, making a “true” flat rate income tax system. Tax rates before the reform ranged from 12 percent to 20 percent. The CIT rate remained at 20 percent, social security contributions rate decreased from 33 to 20 percent, and the standard VAT rate decreased from 20 to 18 percent.

The pre-treatment fit is good at 0.069. The impact of tax reform on GDP per capita is small (\$96 higher on average than the synthetic unit) and insignificant for all dynamic treatment effects and for all average treatment effects.

Romania (2005). Romania introduced a flat tax in 2005 at the rate of 16 percent. Tax rates before the reform were progressive, ranging from 18 percent to 40 percent. Romania also reduced its profit tax from 25 to 16 percent, and dividends, interest income and capital gains were taxed at a lower rate than 16 percent.

The pre-treatment fit is good (0.040), with a substantial and significant impact of tax reform on GDP per capita (\$1022 higher on average than the synthetic unit and significant at the 10 percent level). Also, all lead specific treatment effects except for the last lead are significant at the 10 percent level or better.

Turkmenistan (2005). Turkmenistan introduced a flat tax in 2005 at the rate of 10 percent. In 2005, the CIT rate varied from 8 to 20 percent, and the VAT standard rate was 20 percent. There is a substantially positive and significant impact of tax reform on GDP per capita, or \$2558 higher on average than the synthetic unit and significant at the 5 percent level. Similarly, 3 out of 5 lead specific treatment effects are significant at the 5 percent level, and 1 more is significant at the 10 percent level.

Overall Effects. Figure 6 presents the average causal impact of flat tax reforms on GDP per capita for all the treated countries, countries that adopted flat tax in the 1st wave (Estonia and Latvia) and countries that adopted flat tax in the 2nd wave (Russia, Slovak Republic, Ukraine, Georgia, Romania and Turkmenistan). Following Cavallo et al. (2013), we normalize the estimates before aggregating them by setting the GDP per capita of each treated country equal to 1 in the treatment year. We also present lead specific p-values to assess the significance of these estimates.

The overall effects are positive, economically meaningful, and statistically significant in both waves of tax reform. Five years after the treatment, the average GDP per capita of all treated countries is 18.2 percent higher than the synthetic group. Similarly, GDP per capita for 1st wave countries is 17.8 percent higher and GDP per capita for 2nd wave countries is 18.3 percent higher than the respective synthetic group.²⁸ Of some note, the nearly identical impacts of reform in 1st versus 2nd wave countries provides additional evidence that we are capturing the

²⁸ We chose 5 years because all the treated countries in our sample have 5 years of data after the treatment.

effect of the flat tax reforms and not simply the impact of transition. The estimates are significant at the 10 percent level for all countries two years after the reform and at the 5 percent or better level three years after the reform. Also, the estimates are significant at the 10 percent level for 1st wave countries four years after the reform and at the 5 percent level eight years after the reform. The estimates for 2nd wave countries are significant at the 10 percent level two years after the reform and at the 5 percent or better level three years after the reform.

5.2. Robustness Tests

We also conduct a series of robustness checks. First, we test the effect of flat tax reform on the growth rate of GDP per capita. Second, we test the validity of the “no anticipation effects” assumption by assuming that the intervention took place two years prior to the actual implementation date and then running our baseline model. Third, we implement a “leave-one-out test” to check the sensitivity of our baseline estimates to the inclusion of specific donor countries in the construction of synthetic unit. Fourth, we conduct the placebo experiment using countries from Eastern and Central Europe only. Fifth, we estimate the treatment effect for the 1st wave countries by limiting the donor pool to the countries that will adopt the flat tax reform in the future. Sixth, we estimate the impact of flat tax reforms using alternative datasets to test whether our estimates are dependent on the use of Penn World Tables 7.1. Seventh, we estimate the impact of flat tax reforms using conventional Difference-in-Differences (DID) methods to test whether our estimates are dependent on the use of synthetic control methods. We also expand the DID analysis to include an additional 22 treated countries to test whether our main results are sensitive to choice of the original 8 treated countries. Eighth, we draw inference using alternative thresholds for the selection of placebo units to test whether our inference is sensitive to the use of placebo units with pre-treatment fit as good as the treated country. Ninth,

we run the synthetic control algorithm using alternative covariates and alternative donor and placebo pools to check the sensitivity of our estimates to the inclusion of different covariates and control units.

We find our baseline estimates to be largely robust, and they are discussed in detail in the Appendix.

GDP Per Capita Growth Rate. When we use SCM to examine the impact of tax reform on the growth rate of GDP per capita, we find that 4 (Estonia, Latvia, Russia, and Slovakia) out of 7 countries with positive and significant impacts on the level of GDP per capita had positive, economically large, and statistically significant growth effects at the 5 percent level. These estimates range from 3.8 percentage points in the case of Latvia to 6.6 percentage points in the case of Estonia. Three other countries with positive effects on the level also had positive effects on the growth rate, but these effects were statistically insignificant (although economically meaningful at 2.8 percentage points for Ukraine (p-value=0.17), 1.3 percentage points for Romania, and 0.7 percentage points for Turkmenistan). Georgia had insignificant impacts of tax reform on both the level and the growth rate of GDP per capita.

No Anticipation Effects. It is well-known that policies of these magnitudes are often implemented with some time lag. This time lag can potentially create an “anticipation effect” that could bias our estimates. We run the synthetic control algorithm assuming that the treated countries in our sample were treated two years prior to the implementation date. In doing so, we are constructing the synthetic unit with data before the anticipation might reasonably be expected to have any impact; if these estimates are similar to our baseline estimates, then we can be confident that there were no substantial anticipation effects. The estimates are largely similar to our baseline estimates, implying negligible anticipation effects, with only Estonia and

Turkmenistan as exceptions. In the case of Estonia, we believe the explanation is because testing for the anticipation effect leaves only two pre-treatment years for calibration. Furthermore, Estonia was experiencing recession due to transition in the early 1990s, giving a short “J” shaped trend in GDP per capita before the treatment, which is completely missed by assuming that the intervention took place two years earlier. Even so, the treatment effects are qualitatively similar for Estonia. In the case of Turkmenistan, the pre-treatment trends are similar to the trends in the baseline model. However, the new synthetic unit has a higher level of GDP per capita prior to the treatment, which makes the dynamic treatment effects insignificant. Even so, the synthetic unit's trajectory of GDP per capita after the treatment is similar to the baseline model. Overall, we conclude that there is no significant anticipation effect biasing our estimates.

Leave-one-out Tests. We implement “leave-one-out: tests to check the sensitivity of our baseline estimates to the inclusion of specific donor countries in the construction of synthetic unit. This allows us to assess to what degree our results are driven by any particular donor country, and also to test whether there are any significant exogenous shocks after the treatment assumption for the donor units with positive weights.

To apply these tests, we iteratively re-estimate the baseline model excluding in each iteration one of the countries that received positive weights in the baseline model. For all 8 countries, the leave-one-out synthetic units closely track the baseline synthetic units (especially for those with a good pre-treatment fit), giving qualitatively similar treatment effects to the baseline model. Therefore, our baseline estimates are not sensitive to the inclusion of any particular country.

Placebo Experiment Using Countries from Eastern and Central Europe Only. In our main analysis, we run the placebo experiment using a worldwide sample of countries, where each placebo treatment effect is calculated by restricting the donor pool of each placebo units to their respective region. The advantage of doing so is that the size and the power of the test are large. However, one criticism of this approach is that the placebo countries may not be similar to the treated country because they do not belong to the same macro-region. Accordingly, we conduct the placebo experiment using countries from Eastern and Central Europe only. We find that the true treatment effect is similar to but consistently above the placebo effects.

1st Wave Treated Countries with Future Flat Tax Adopters in the Donor Pool. We find the treatment effect for the 1st wave countries by limiting the donor pool to the countries that will adopt the flat tax reform in the future, which allows us to test whether the growth effects of flat tax reform are due to some unobservable characteristics of the adopters and not the reform itself. We find a positive impact of flat tax reform, effects that are larger than the placebo effects.

Alternative Datasets. We use GDP per capita data from Penn World Tables (PWT) for our baseline specification because PWT is the most widely used database for cross-country empirical work. According to Johnson et al. (2013), about two-thirds of all cross-country empirical work is based on PWT. Given some concerns with national accounts data and particularly with different versions of PWT, we consider various alternative measures of GDP per capita for robustness checks (Johnson et al. 2013; Ciccone and Jarocinski 2010). Specifically, we use a chained series of GDP per capita from PTW 7.1, GDP per capita data from World Development Indicators, and finally expenditure-side GDP per capita from PWT 8.0, which is most comparable to the GDP per capita from PWT 7.1 (Feenstra et al. 2013). All

of the measures of GDP per capita are adjusted for purchasing power parity, and all are expressed in 2005 constant prices. Overall, our results are largely consistent across the different data sets, although there are sometimes small differences in specific cases.

Difference-in-Differences Analysis. We estimate the impact of flat tax reforms using standard DID methods to test whether our estimates are dependent on our use of SCM. The dependent variable is GDP per capita. Both SCM and DID require a different sample selection process, and each method constructs the counterfactuals differently. Therefore, we do not expect DID estimates to replicate exactly the estimates from SCM. However, we do expect the results to be similar. When constructing the counterfactual, SCM puts more weight on donor countries that closely resemble the treated country, whereas the DID approach assigns equal weight to each donor country (conditional on covariates).

In our preferred specifications, DID estimates indicate that GDP per capita of countries with a flat rate income tax is from \$1575 to \$1745 higher on average after the treatment than the worldwide sample of countries without a flat rate income tax (Figure 7). Although these estimates cannot be directly compared to the estimates obtained from SCM, they are of similar magnitude. For instance, the average of country-specific average treatment effects of eight countries (column 1 of Table 4) is \$1505.

Figure 7 also presents results for the event study that includes leads and lags of the treatment. All the leads are insignificantly different from zero, and all the estimates from year 0 and forward are positive and statistically significant, with estimates from year 1 onwards being significant at the 1 percent level. The permanent and statistically significant effect of tax reform after a time lag is consistent with the SCM estimates. The use of alternative datasets does not change these results.

We also use DID to test whether our results are sensitive to the choice of the treated countries. To do so, we expand the treated countries to include all countries (30 in total) that adopted a flat rate income tax in the original sample period. The results are presented in the second row of Figure 7. The DID estimates indicate that GDP per capita of countries with a flat rate income tax is from \$1482 to \$1503 higher on average after the treatment than the worldwide sample of countries without a flat rate income tax, which is very similar to both the SCM estimates and the DID estimates using 8 countries. Similarly, all the leads are insignificantly different from zero at the 5 percent level, and all the estimates from year 0 and forward are positive and most are statistically significant, with estimates from year 2 and year 4 onwards being significant at the 5 percent or better level. The permanent and statistically significant effect of tax reform after a time lag is consistent with both the SCM estimates and the DID estimates using 8 countries.

Alternative Thresholds. We use alternative thresholds to select the placebo units for inference. Overall, the significance of the estimates is insensitive to the choice of the threshold, especially for treated countries with a very good pre-treatment fit.

Alternative Covariates and Alternative Donor Pools. Finally, we examine the impact of several alternative covariates and donor pools. For example, we match on the pre-treatment average of GDP per capita as opposed to all realizations of the GDP per capita. We remove inflation from the baseline estimate and re-run the synthetic control algorithm. We also run the synthetic control algorithm using no covariates, i.e., using only the pre-treatment trajectory of GDP per capita as the covariates. Further, we use a global pool of countries instead of the hybrid pool in the main specification. We also test whether the significance of the estimates changes if we keep the treated country in the pool of placebo units when drawing inferences.

Finally, our estimates for the two oil-exporting countries in the sample (Russia and Turkmenistan) might be affected by the increase in the energy prices in the 2000s, so we run the SCM algorithm restricting the donor pools to oil-exporting countries only. Across all of these variants, we find that our results are unaffected by the choices of variables or donor pools.

6. Conclusions

Using synthetic control methods, we find positive impacts of comprehensive tax reform on GDP per capita for all countries in Eastern and Central Europe, in comparison to their respective synthetic controls. We conduct placebo experiments to assess the significance of the estimates, and find that two positive effects (Russia, and Turkmenistan) are significant at the 5 percent significance level, five (Estonia, Latvia, Slovak Republic, Ukraine, and Romania) at the 10 percent significance level, and one (Georgia) is insignificant at conventional significance levels. We also conduct several robustness tests, and we find that our results are largely unaffected.

Our main result demonstrates a strong positive impact of comprehensive flat tax reform. What are the possible mechanisms through which flat tax reforms affect economic performance? Standard theory suggests that tax reform will affect economic performance by affecting such variables as investment per capita, the stock of FDI as a share of GDP, and the employment to population ratio.²⁹ To explore these channels, we use synthetic control methods to estimate the causal effect of flat tax reforms on all of these variables.³⁰

²⁹ Another possible mechanism through which flat tax reforms could improve economic growth is through any “fiscal stimulus” provided by a possible decrease in tax burden after the implementation of the reform. If this is the case, then the growth effects through fiscal stimulus would be commingled with the growth effects through revenue-neutral flat tax reform, the latter of which is the focus of our study. In theory it is easy to test for this fiscal stimulus mechanism. We could use some measure of overall tax burden such as tax revenue as a share of GDP as an outcome variable in the synthetic control methods framework and check whether the tax revenue decreased or not after the adoption of flat tax reforms. However, it is difficult to do so in practice due to the lack of data. The

Overall, these estimations suggest that flat tax reform has largely positive impacts on these variables. These results are shown in Figure 8 and in Table 4. In Figure 8, we see that there are positive and significant impacts on investment per capita for the group of all countries, at least after some time lag. This result is driven mainly by the 1st wave of countries, as the impacts there are positive in all cases and significant in 8 out of 10 cases. Similarly, we see in Table 4 that flat tax reforms affected investment per capita and GDP per capita in the same fashion, with a positive impact of reform on investment per capita for all of countries except Turkmenistan. For the most part, the trends in FDI as a share of GDP are also similar to the GDP per capita trends, although the estimates are not very robust. These effects seem driven largely by the 1st wave countries, where the effects on FDI become statistically significant from year 5 onwards. The effects for 2nd wave countries are mixed and statistically insignificant for 8 out of 9 post-treatment years. The impact of tax reform on employment is positive and

authoritative cross-country data on tax revenue is *Government Finance Statistics* from the International Monetary Fund. Out of 8 treated countries in our sample, 1 country has no revenue data (Turkmenistan), 2 countries have revenue data only for the post-treatment years (Estonia, Russia), and 3 countries have less than or equal to 3 years of pre-treatment data (Latvia, Romania, Slovak Republic), which leaves only 2 countries (Georgia, Ukraine) with more than 3 years of pre-treatment data. Even so, we use these limited data to compare averages of 5 pre-reform years to averages of 5 post-reform years for 5 countries with pre- and post-data. We find that in all cases except Slovak Republic revenue as a share of GDP actually increased after the reform. Further, when we compare the revenue share 1 year before the reform to the share in the year of the reform, we find that revenue increased in Georgia and Latvia, decreased by less than 0.5 percentage points in Romania and Ukraine, and decreased by less than 1.5 percentage points in Slovak Republic. Note that Keen, Kim, and Varsano (2008) compare revenue as a share of GDP 1 year before the reform to the year of the reform for Estonia and Russia, and find that it increased in both countries; they also find that for other flat tax reform countries total tax revenues as a share of GDP were on average only slightly below pre-reform levels. Also, Saavedra (2007) finds that on average there was no significant impact of the flat tax reforms on the revenue, and argues that this was because policy makers adjust tax provisions in order to avoid a significant decline in revenues. Thus, these studies suggest that the fiscal stimulus mechanism cannot explain any economically and statistically meaningful growth effects.

³⁰ Investment per capita data are obtained from the Penn World Tables, FDI data are obtained from the United Nations Conference on Trade and Development, and the employment to population ratio data are obtained from the World Development Indicators. The covariates for investment per capita and FDI as a share of GDP are GDP per capita growth rate, real interest rate, M2 as a share of GDP, average schooling, openness, and a democracy dummy. The covariates for the employment ratio are GDP per capita, GDP per capita growth rate, population growth rate, percent of population age 16 to 64, and average schooling. Data for the real interest rate, M2 as a share of GDP, and percent of population age 16 to 64 are obtained from the World Development Indicators. New firm density data are obtained from the World Bank's Entrepreneurship Survey. Data sources for the other variables were described earlier. Also, the same sample selection procedure is used for these outcome variables and GDP per capita.

statistically significant from year two onwards (except for year 10) for the group of all countries. Again, the impacts are quite different for the 1st and 2nd wave countries.³¹

In sum, our results indicate that flat tax reform generally had significant, positive, and meaningful impacts on income. Even so, these effects seem somewhat heterogeneous across countries, with differential impacts on 1st wave versus 2nd wave countries. Whether future implementation of flat tax reform will continue to have these positive effects is uncertain.

References

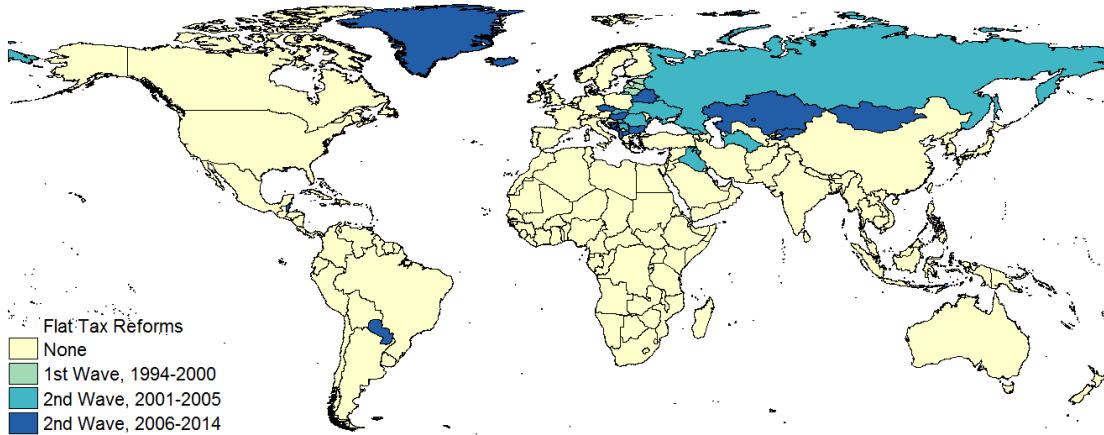
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105 (490): 493-505.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2014. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, forthcoming.
- Abadie, Alberto, and Javier Gardeazabal. 2003. The Economic Costs of Conflict: A Case Study of the Basque Country. *The American Economic Review*, 93 (1): 113-132.
- Abdallah, Chandi S., and William D. Lastrapes. 2012. Home Equity Lending and Retail Spending: Evidence from a Natural Experiment in Texas. *American Economic Journal: Macroeconomics*, 4(4): 94-125.
- Andrews, William D. 1974. A Consumption-type or Cash Flow Personal Income Tax. *Harvard Law Review*, 87 (6), 1113-1188.
- Angrist, Joshua D., Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press.
- Appel, Hilary. 2011. *Tax Politics in Eastern Europe*. Ann Arbor, MI: The University of Michigan Press.
- Auerbach, Alan J. and Joel Slemrod. 1997. The Economic Effects of the Tax Reform Act of 1986. *The Journal of Economic Literature*, 35 (2): 589-632.
- Barro, Robert J. 1991. Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, 106 (2): 407-443.
- Barro, Robert J., and Jong-Wha Lee. 2012. A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104 (1): 184-198.
- Baturo, Alexander and Julia Gray. 2009. Flatliners: Ideology and Rational Learning in the Adoption of the Flat Tax. *European Journal of Political Research*, 48 (2): 130-159.
- Bauhoff, Sebastian. 2014. The Effect of School District Nutrition Policies on Dietary Intake and Overweight: A Synthetic Control Approach. *Economics and Human Biology*, 12 (1), 45-55.

³¹ Note that tax reform is also thought to improve entrepreneurial activity. However, data on entrepreneurship are only available from 2004 onward. Even so, we find strong correlations between flat tax reforms and entrepreneurial activity, using the number of newly registered limited liability companies per 1000 working-age people as a rough measure of entrepreneurship. For instance, for countries without a flat tax reform, the regional yearly average of newly registered companies is 1.402, and the average across the world is 2.973. In contrast, the post-reform average for treated countries (excluding Turkmenistan because of data unavailability) is 4.135. Also, the post-reform average for 1st wave countries is 6.644, and for 2nd wave countries it is 3.090.

- Bertrand, Marianne, Esther Duo, and Sendhil Mullainathan. 2004. How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119 (1): 249-275.
- Besley, Timothy J. and Anne Case (1995). Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits. *The Quarterly Journal of Economics*. 110 (3): 769-798.
- Billmeier, Andreas, and Tommaso Nannicini. 2011. Economies in Transition: How Important Is Trade Openness for Growth? *Oxford Bulletin of Economics and Statistics*, 73 (3): 287-314.
- Billmeier, Andreas, and Tommaso Nannicini. 2013. Assessing Economic Liberalization Episodes: A Synthetic Control Approach. *The Review of Economics and Statistics*, 95 (4): 983-1001.
- Bohn, Sarah, Magnus Lofstrom, and Steven Raphael. 2014. Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population? *The Review of Economics and Statistics*, forthcoming.
- Bradford, David. 1986. *Untangling the Income Tax*. Cambridge, MA: Harvard University Press.
- Case, Anne, James R. Hines, Jr., and Harvey S. Rosen. 1993. Budget Spillover and Fiscal Policy Interdependence: Evidence from the States. *Journal of Public Economics*, 52 (3): 285-307.
- Brock, Smith. 2015. The Resource Curse Exorcised: Evidence From a Panel of Countries. *Journal of Development Economics*, 116: 57-73.
- Castanheira, Micael, Gaetan Nicodeme, and Paola Profeta. 2012. On the Political Economics of Taxation: Survey and Empirical Measurement. *International Tax and Public Finance*, 19 (4): 598-624.
- Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, 2013. Catastrophic Natural Disasters and Economic Growth. *The Review of Economics and Statistics*, 95 (5): 1549-1561.
- Ciccone, Antonio, and Marke Jarocinski. 2010. Determinants of Economic Growth: Will Data Tell? *American Economic Journal: Macroeconomics*, 2(4): 222-246.
- Chetty, Raj, Adam Looney, and Kory Kroft. 2009. Salience and Taxation: Theory and Evidence. *The American Economic Review*, 99 (4): 1145-1177.
- Cloyne, James. 2013. Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom. *The American Economic Review*, 103 (4): 1507-1528.
- DeAngelo, Gregory and Benjamin Hansen. 2014. Life and Death in the Fast Lane: Police Enforcement and Traffic Fatalities. *American Economic Journal: Economic Policy*, forthcoming.
- Dorsett, Richard. 2013. The Effect of the Troubles on GDP in Northern Ireland. *European Journal of Political Economy*, 29: 119-133.
- Evans, Anthony John, and Paul Dragos Aligica. 2008. The Spread of the Flat Tax in Eastern Europe: A Comparative Study. *Eastern European Economics*, 46 (3): 49-67.
- Favero, Carlo, and Francesco Giavazzi. 2012. Measuring Tax Multipliers: The Narrative Method in Fiscal VARs. *American Economic Journal: Economic Policy*, 4 (2): 69-94
- Feenstra, Robert C., Robert Inklaar, Marcel Timmer. 2013. PWT 8.0 – A User Guide. Available online at <http://dx.doi.org/10.15141/S5159X>.
- Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Gorodnichenko, Yuriy, Jorge Martinez-Vazquez, and Klara Sabirianova Peter. 2009. Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare Effects in Russia. *The Journal of Political Economy*, 117 (3): 504-554.
- Hall, Robert E., and Alvin Rabushka. 1985. The Route to a Progressive Flat Tax. *Cato Journal*, 5 (2): 465-480.
- Hall, Robert E., and Alvin Rabushka. 2007. *The Flat Tax* (2nd Edition). Palo Alto, CA: Hoover Institution Press.
- Heston, Alan, Robert Summers and Bettina Aten. 2012. Penn World Table Version 7.1, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
- Institute for Fiscal Studies. 1973. *The Structure and Reform of Direct Taxation: Report of a Committee Chaired by Professor James E. Meade*. London, UK: George Allen and Unwin.

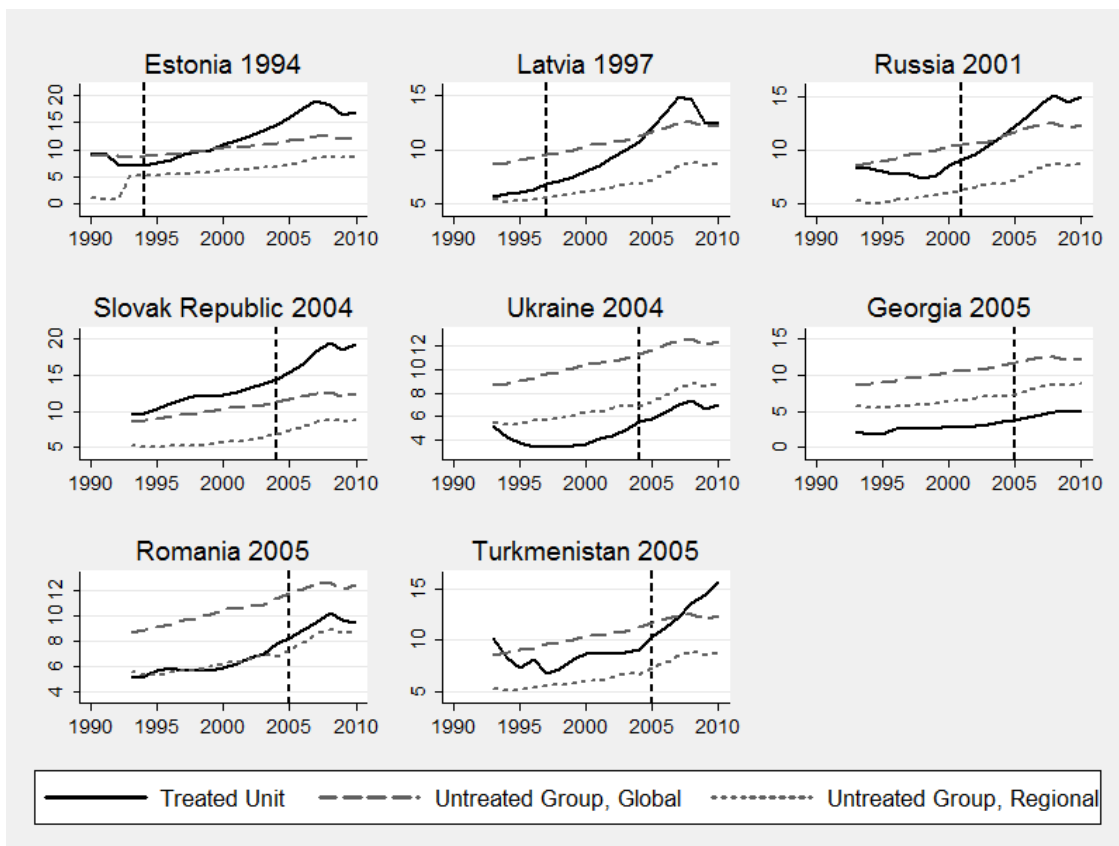
- Ivanova, Anna, Michael Keen, Alexander Klemm. 2005. The Russian “Flat Tax” Reform. *Economic Policy*, 20 (43): 397-444.
- Ji, Kan. 2013. Causes and Consequences of the Flat Income Tax. In *Essays on Tax Policy, Institutions, and Output*, unpublished dissertation, Chapter 3, Tilburg University. Tilburg, The Netherlands: Prisma Print.
- Johnson, Simon, Willian Larson, Chris Papageorgiou, and Arvind Subramanian. Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates. *Journal of Monetary Economics*, 60: 255-274.
- Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi. 1993. Optimal Taxation in a Model of Endogenous Growth. *The Journal of Political Economy*, 101 (3): 485-517.
- Kaldor, Nicholas. 1955. *An Expenditure Tax*. London, UK: George Allen and Unwin.
- Kawano, Laura, and Joel Slemrod. 2012. The Effect of Tax Rates and Tax Bases on Corporate Tax Revenues: Estimates With New Measures of the Corporate Tax Base. NBER Research Working Paper 18440. Cambridge, MA: National Bureau of Economic Research.
- Keen, Michael, Yitae Kim, and Ricardo Varsano. 2008. The Flat Tax (es): Principles and Experience. *International Tax and Public Finance*, 15 (6): 712-751.
- Kopczuk, Wojciech. 2005. Tax Bases, Tax Rates, and the Elasticity of Reported Income. *Journal of Public Economics*, 89 (11): 2093-2119.
- Kopczuk, Wojciech. 2012. The Polish Business “Flat” Tax and its Effect on Reported Incomes: A Pareto Improving Tax Reform? Department of Economics Working Paper. New York, NY: Columbia University.
- Mankiw, Gregory N., David Romer, and David N Weil. 1992. A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107 (2): 407-437.
- Marshall, Monty G., and Keith Jagers. 2013. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2012. College Park, MD: Center for International Development and Conflict Management, University of Maryland.
- McLure, Charles E., Jr., and George R. Zodrow. 2007. Consumption-based Direct Taxes: A Guided Tour of the Amusement Park. *FinanzArchiv*, 63 (2): 285-307.
- Mieszkowski, Peter. 1977. A Cash Flow Version of an Expenditure Tax. Washington, D.C.: U.S. Department of Treasury, Office of Tax Analysis Working Paper No. 26.
- Montalvo, Jose G. 2011. Voting After Bombings: A Natural Experiment on the Effect of Terrorist Attacks on Democratic Elections. *Review of Economics and Statistics*, 2011, 93 (4): 1146-1154.
- Pecorino, Paul. 1994. The Growth Rate Effects of Flat Tax Reform. *Oxford Economic Papers*, 46 (3): 492-501.
- Persson, Torsten, and Guido Tabellini. 2007. The Growth Effect of Democracy: Is It Heterogenous and How Can It Be Estimated? NBER Research Working Paper 13150. Cambridge, MA: National Bureau of Economic Research.
- Peter, Klara Sabirianova, Steve Buttrick, and Denvil Duncan. 2010. Global Reform of Personal Income Taxation, 1981-2005: Evidence from 189 Countries. *National Tax Journal*, 63 (3): 447-478.
- Romer, Christina D., and David H. Romer. 2010. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *The American Economic Review*, 100 (3): 763-801.
- Slemrod, Joel. 2005. The Costs of Tax Complexity. Presentation to the President’s Advisory Panel on Tax Reform, available online at http://govinfo.library.unt.edu/taxreformpanel/meetings/docs/slemrod_03032005.ppt .
- Stokey, Nancy L. and Sergio Rebelo. 1995. Growth Effects of Flat-Rate Taxes. *The Journal of Political Economy*, 103 (2): 419-450.
- United Nations. 2013. United Nations Conference on Trade and Development. Online Edition. <http://unctadstat.unctad.org/> (accessed November 2014).
- World Bank. 2013. 2013 World Development Indicators. Online Edition, available at <http://data.worldbank.org/data-catalog/world-development-indicators/>.

Figure 1: The Spread of the Flat Tax Reforms



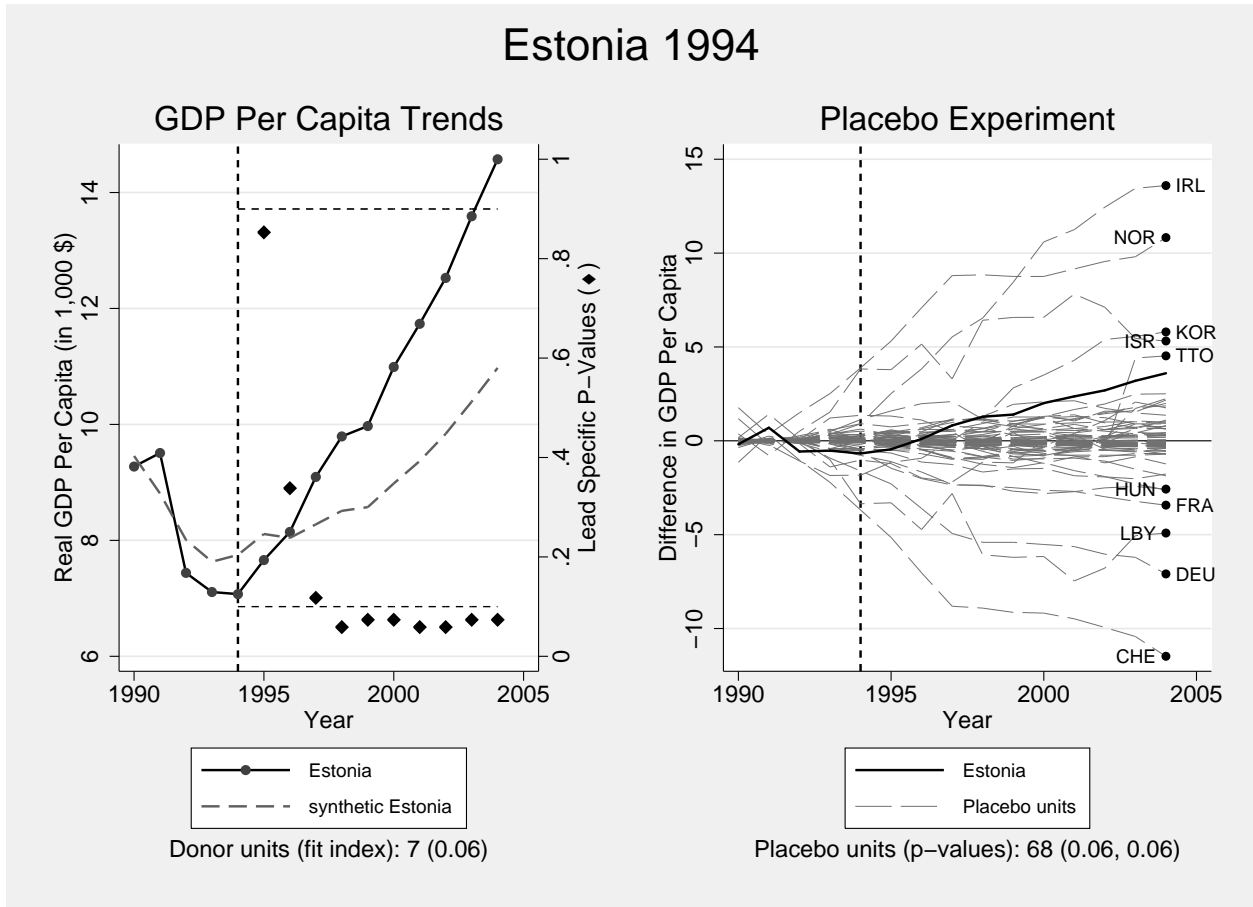
Notes: The map shows the countries that adopted flat rate income tax between 1990-2014. As of 2014, some of these countries have switched back to progressive income tax. Some of the small island countries with a flat rate income tax are not visible in the map due to the small geographical size of those countries.

Figure 2: GDP Per Capita Trends - Treated Countries, Mean of Untreated Countries (Global), and Mean of Untreated Countries (Regional)



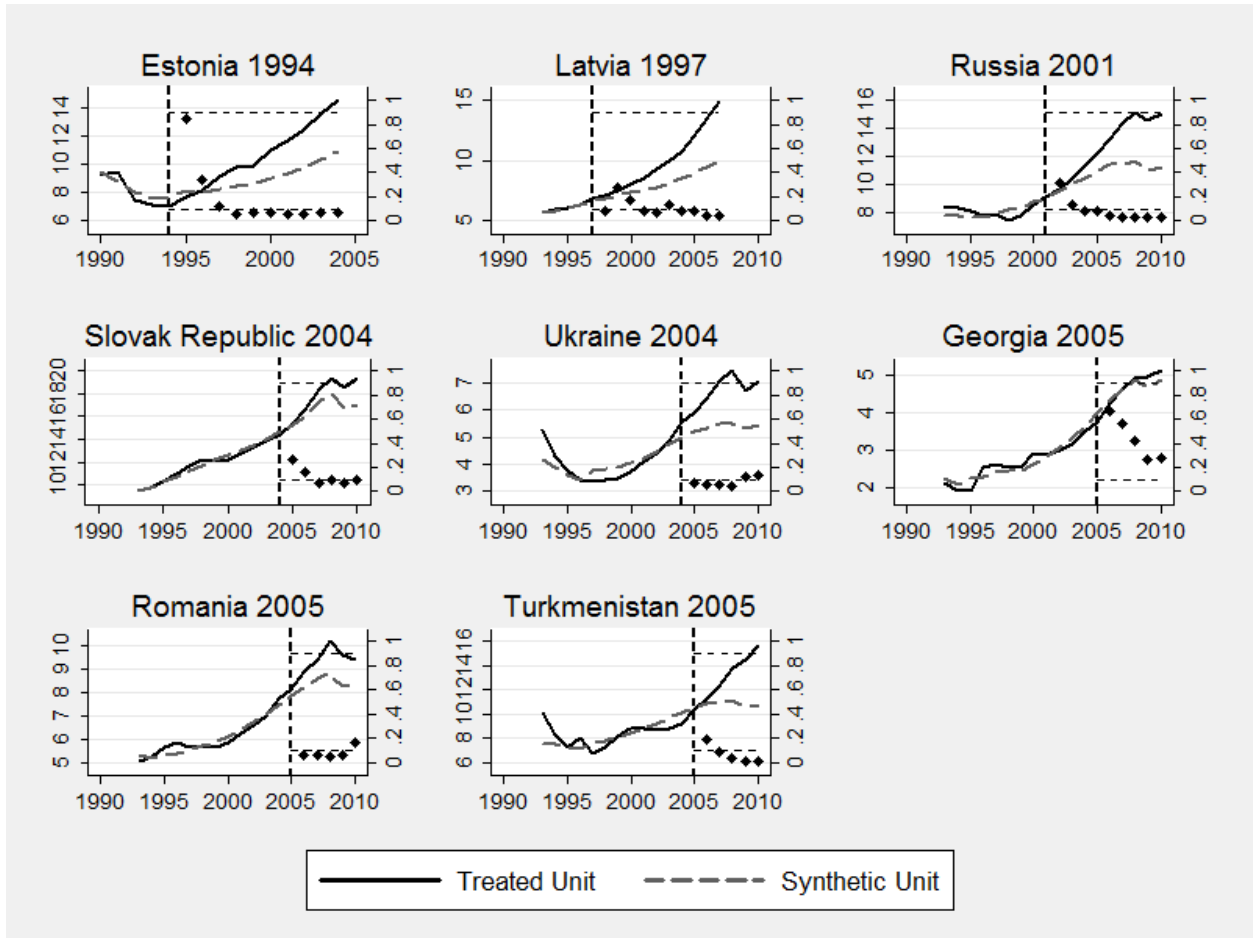
Notes: The solid black line denotes the real GDP per capita trend of treated countries, the dashed gray line denotes the mean real GDP per capita of all untreated countries from the world, and the dotted gray line denotes the mean real GDP per capita of untreated countries from the Eastern and Central Europe.

Figure 3: GDP Per Capita Trends and Placebo Experiments - Estonia versus Synthetic Estonia



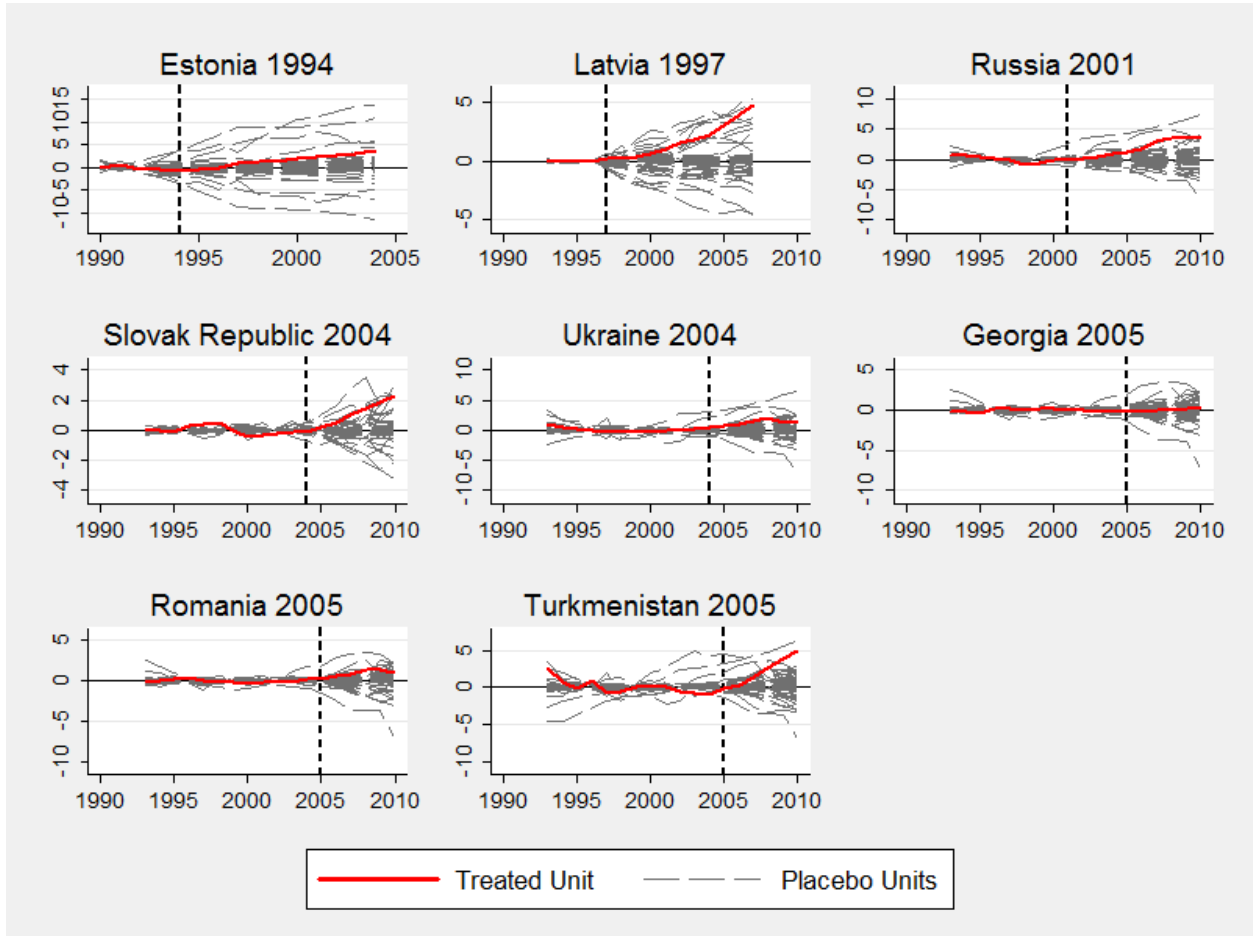
Notes: In the left panel, the solid black line denotes the real GDP per capita trend of Estonia, the dashed black line denotes the real GDP per capita trend of its synthetic control. P-values (the probability that the gap in outcomes would occur by chance for each year after the treatment) are shown with the scatter plot, where horizontal dashed lines are the 0.9 line and the 0.1 line, respectively. In the right panel, the solid black line denotes the difference in real GDP per capita between Estonia and its synthetic control. The dashed gray lines denote the difference in real GDP per capita between each of the Estonia's placebo units and their synthetic control in placebo experiments. The outliers in placebo experiments are labeled using country codes.

Figure 4: GDP Per Capita Trends (1000s of 2005 International Dollar, Left Axis) and P-Values from Placebo Experiments (Right Axis)



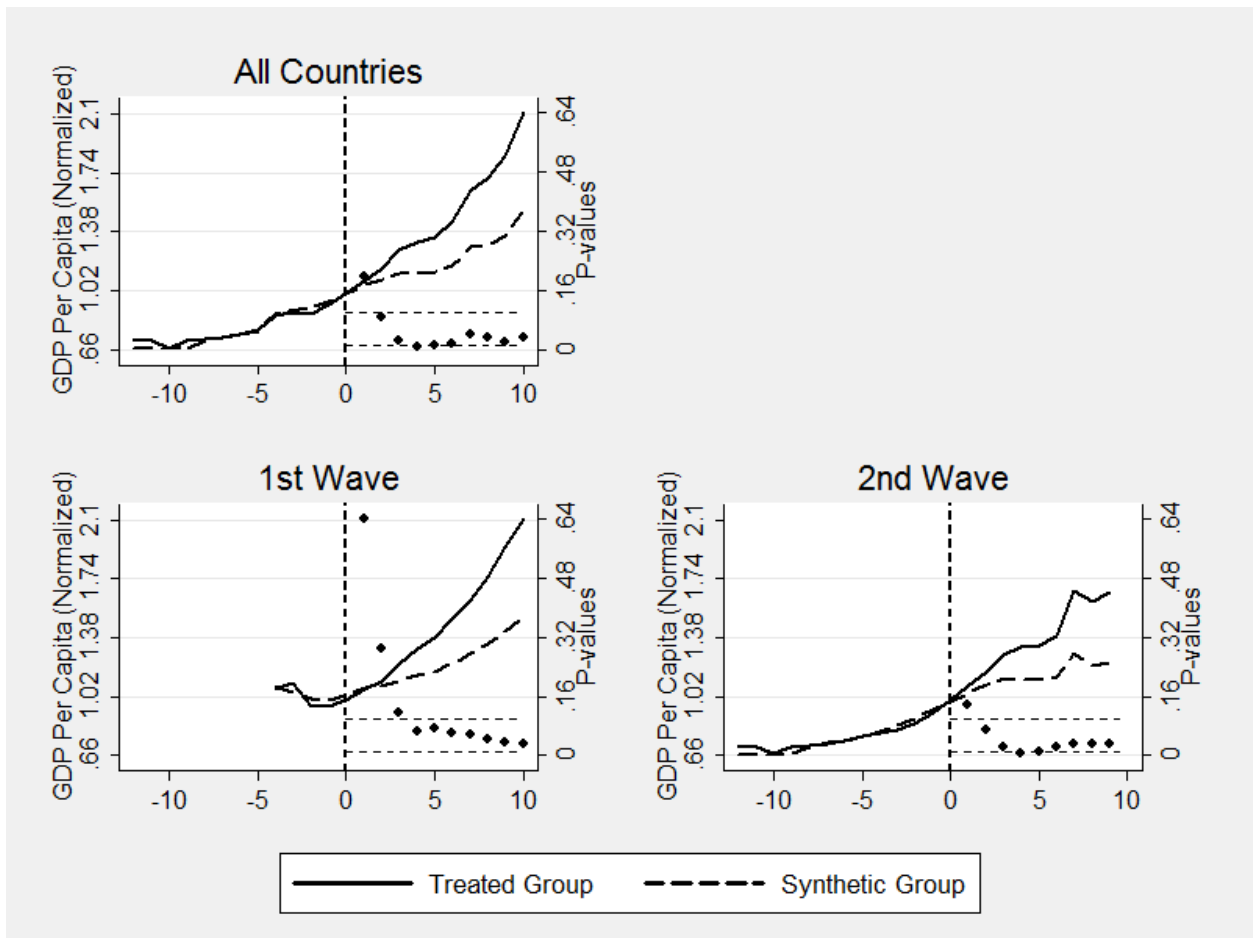
Notes: The solid black line denotes the real GDP per capita trend of treated country, and the dashed black line denotes the real GDP per capita trend of its synthetic control. P-values (the probability that the gap in outcomes would occur by chance for each year after the treatment) are shown with the scatter plot, where horizontal dashed lines are the 0.9 line and the 0.1 line, respectively. The left axis consists of labels for GDP per capita trends, and the right axis consists of labels for P-values.

Figure 5: Time Series Graph of the Placebo Experiments



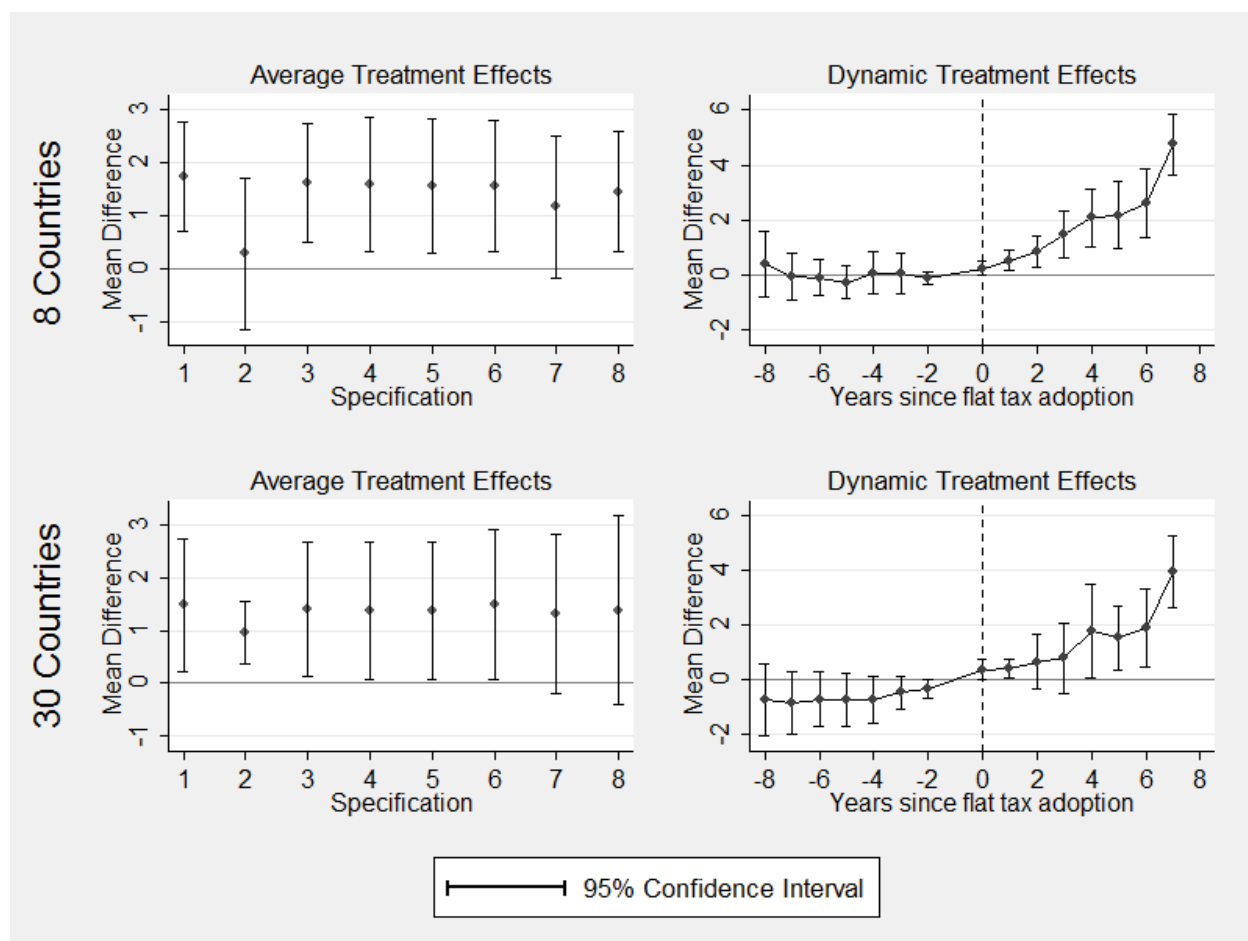
Notes: The solid red line denotes the difference in real GDP per capita between each treated country and its synthetic control. Dashed gray lines denote the difference in real GDP per capita between each of the treated country's placebo units and their synthetic control in placebo experiments.

Figure 6: GDP Per Capita Trends (Normalized, Left Axis) and P-Values from Placebo Experiments (Right Axis) for All Countries: 1st Wave Countries versus 2nd Wave Countries



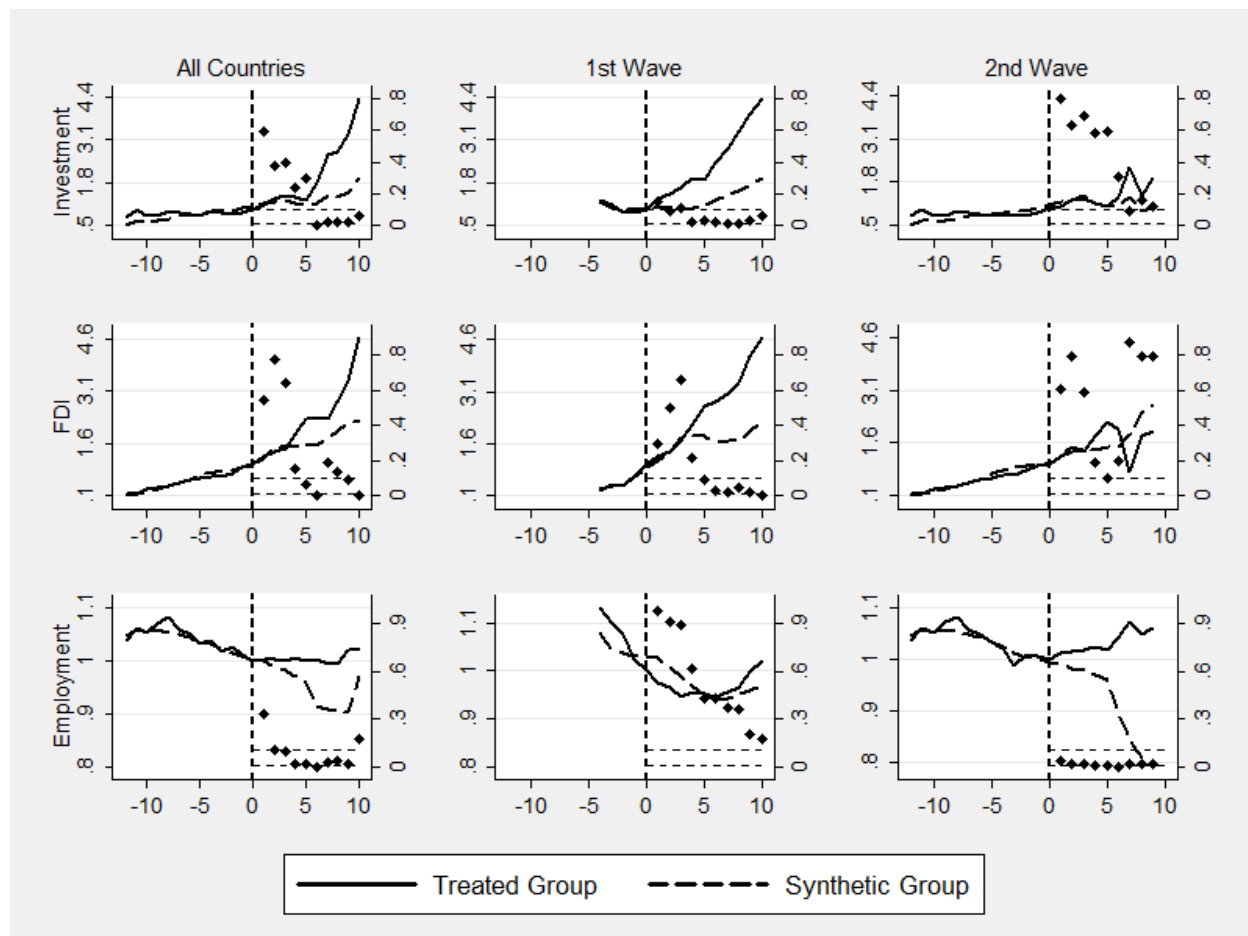
Notes: The solid black line denotes the outcome of the group of treated countries, and the dashed black line denotes the outcome of its synthetic control. The left axis consists of labels for the outcome variable, and the right axis consists of labels for P-values (the probability that the gap in outcomes would occur by chance for each year after the treatment), which are shown with the scatter plot, where horizontal dashed lines are the 0.10 line and the 0.01 line, respectively. To avoid the size effect caused due to varying levels of GDP per capita across countries, the estimates are normalized before aggregating them by setting the GDP per capita of each treated country equal to 1 in the treatment year.

Figure 7: Average Treatment Effects and Dynamic Treatment Effects Using Conventional Difference-in-Differences Approach



Notes: In the top row only 8 treated countries used in SCM are included, while in the bottom row all countries that implemented flat rate income tax by 2010 with non-missing data (30 in total) are included in the estimation. In the left panel, the average treatment effects and their 95 percent confidence intervals are plotted. The estimates come from Table A5. All specifications include year and country fixed effects. Specification 2 adds country specific time trends to specification 1. To get specifications 3-8, we take specification 1 and in each round add one additional control variable in this order: population growth, openness, investment share, democracy dummy, inflation, and average schooling. In the right panel, the connected line plots the estimates for leads to the left of the vertical line and the estimates for lags to the right. The estimates come from specification 1 in Table A6. The variable Year = -1 is omitted. The left-most lead captures the difference in the outcome variable between treated countries and untreated countries 8 years and earlier, and the right-most lag captures the same difference 7 years and beyond the treatment year. See the Appendix 3.5 for details and additional robustness checks.

Figure 8: Investment, FDI and Employment Trends (Normalized, Left Axis) and P-Values from Placebo Experiments (Right Axis) for All Countries: 1st Wave Countries versus 2nd Wave Countries



Notes: The outcome variable for the first row is real investment per capita, the outcome variable for the second row is the stock of FDI as a share of GDP, and the outcome variable for the third row is employment ratio. The solid black line denotes the outcome of the group of treated countries, and the dashed black line denotes the outcome of its synthetic control. The left axis consists of labels for outcome variable and the right axis consists of labels for P-values (the probability that the gap in outcomes would occur by chance for each year after the treatment), which are shown with the scatter plot, where horizontal dashed lines are the 0.9 line and the 0.1 line, respectively. To avoid the size effect caused due to varying levels of outcome variables across countries, the estimates are normalized before aggregating them by setting the outcome variable of each treated country equal to 1 in the treatment year.

Table 1: Root Mean Square Prediction Error (RMSPE) versus Fit Index

| Year | Growth Rate of GDP Per Capita | | GDP Per Capita in Level | | | |
|-----------|-------------------------------|-------------|-------------------------|-------------|-------------|-------------|
| | Y | Synthetic Y | Low Income | | High Income | |
| | Y | Synthetic Y | Y | Synthetic Y | Y | Synthetic Y |
| 1 | 3 | 2.88 | 500 | 480 | 15000 | 14400 |
| 2 | 3.5 | 3.36 | 510 | 489.6 | 16000 | 15360 |
| 3 | 5 | 4.8 | 520 | 499.2 | 15000 | 14400 |
| RMSPE | | 0.16 | | 20.4 | | 613.62 |
| Fit Index | | 0.04 | | 0.04 | | 0.04 |

Notes: The RMSPE measures lack of fit between the path of the outcome variable for any particular country and its synthetic counterpart. The pre-treatment fit index normalizes RMSPE. To illustrate the advantage of using fit index, we vary synthetic counterpart's outcome variable by 4 percent in all three cases. Note that the RMSPE varies widely making it impossible to compare the pre-treatment fit across variables or countries. However, the fit index is 0.04 in all three cases, indicating that in all three cases the fit is equivalent to a 4 percent difference between the treated unit and the synthetic unit in each pre-treatment period. See section 3.2 for details.

Table 2: Tax Systems Before and After the Flat Tax Reform

| | Year | PIT | | CIT | | Basic |
|------------------|---------|--------|-------|--------|-------|----------------------|
| | Adopted | Before | After | Before | After | Allowance |
| Estonia* | 1994 | 16-33 | 26 | 35 | 26 | Modest increase |
| Latvia*‡ | 1997 | 15-35 | 25 | 25 | 25 | Slight reduction |
| Russia | 2001 | 12-30 | 13 | 30 | 26† | Modest increase |
| Ukraine | 2004 | 10-40 | 13 | 30 | 25 | Increase |
| Slovak Republic* | 2004 | 10-38 | 19 | 25 | 19 | Substantial increase |
| Georgia | 2005 | 12-20 | 12 | 20 | 20 | Eliminated |
| Romania* | 2005 | 18-40 | 16 | 25 | 16 | Increase |
| Turkmenistan | 2005 | 10-25 | 10 | 25 | 20 | Increase |

Notes: * The post-reform personal income tax rate is equal to the post-reform corporate income tax rate. † Russia decreased its corporate income tax rate from 35 percent to 26 percent in 2002. ‡ Latvia reformed its income tax in 1994 with basic rate at 25 percent and top marginal rate of 35 percent, which later was reduced to 10 percent. Sources: Keen, Kim, and Varsano (2008), and Ji (2013).

Table 3: Summary of Country Specific Data Selection, Average Treatment Effects and their P-values

| | EST | LVA | RUS | SVK | UKR | GEO | ROM | TKM |
|-------------------|---------|---------|----------|---------|---------|--------|---------|----------|
| First Year | 1990 | 1993 | 1993 | 1993 | 1993 | 1993 | 1993 | 1993 |
| Treatment Year | 1994 | 1997 | 2001 | 2004 | 2004 | 2005 | 2005 | 2005 |
| Last Year | 2004 | 2007 | 2010 | 2010 | 2010 | 2010 | 2010 | 2010 |
| RMSPE | 0.5286 | 0.0782 | 0.5076 | 0.2885 | 0.4246 | 0.1774 | 0.2387 | 0.9558 |
| Fit Index | 0.06 | 0.01 | 0.06 | 0.02 | 0.1 | 0.07 | 0.04 | 0.11 |
| Donor Pool | 7 | 8 | 6 | 7 | 7 | 8 | 7 | 10 |
| Actual Donors | 4 | 3 | 2 | 4 | 2 | 3 | 4 | 2 |
| Included Placebos | 68 | 47 | 63 | 42 | 68 | 74 | 62 | 91 |
| Excluded Placebos | 20 | 39 | 21 | 48 | 22 | 31 | 28 | 23 |
| Covariates | All | All | All | All | All | -S | All | - S, -I |
| Treatment Effect | 1.8453* | 1.9659* | 2.0099** | 1.1834* | 1.3630* | 0.0955 | 1.0220* | 2.5577** |
| P-value | (0.06) | (0.09) | (0.05) | (0.10) | (0.07) | (0.41) | (0.06) | (0.02) |

Notes: Donor Pool denotes the number of potential donor pools, Actual Donor denotes the number of countries with positive weights in the construction of the synthetic unit, Included Placebos denotes the number of placebo units used to draw inference, Excluded Placebo denotes the number of placebo units that were excluded when drawing inference due to their poor pre-treatment fit, and Covariates indicates the covariates that were used. If Covariates is All, pre-treatment realizations of the outcome variable and the pre-treatment average of the following covariates are included: investment share of GDP, population growth rate, openness, average schooling, inflation rate, democracy dummy. If Covariates is -I, inflation is excluded. If Covariates is -S, average schooling is excluded, due to unavailability of the data.

Table 4: Average Treatment Effects and P-values for GDP Per Capita and its Potential Drivers Using Synthetic Control Methods

| | GDP | Investment | FDI | Employment |
|-----------------|--------------------|---------------------|----------------------|---------------------|
| Estonia | 1.8453* (0.06) | 1.2502** (0.04) | 29.6933*** (0.00) | -7.3833 (0.12) |
| Latvia | 1.9659* (0.09) | 1.1593* (0.06) | -12.8943** (0.05) | 2.3581 (0.19) |
| Russia | 2.0099** (0.05) | 0.3633 (0.17) | -2.3884 (0.23) | 9.1506** (0.02) |
| Slovak Republic | 1.1834* (0.10) | 0.2325 (0.31) | 13.7939 (0.16) | 3.0423** (0.05) |
| Ukraine | 1.3630* (0.07) | 0.0809 (0.37) | 9.3829 (0.22) | 5.8395*** (0.01) |
| Georgia | 0.0955 (0.41) | -0.6081 (0.15) | 13.4210 (0.25) | -3.2107** (0.02) |
| Romania | 1.0220* (0.06) | 0.8913 (0.12) | 3.9138 (0.32) | -0.4084 (0.57) |
| Turkmenistan | 2.5577** (0.02) | -1.5317** (0.05) | 4.5669 (0.29) | 0.6058 (0.22) |
| Average | 1.505 | 0.23 | 7.44 | 1.25 |

Notes: Average treatment effects are reported at the top and their p-values are reported in a parenthesis below them. Average treatment effects are obtained using synthetic control methods. P-values are obtained from the placebo experiments. * $p < .1$, ** $p < .05$, *** $p < .01$.

About the Authors

James Alm is Professor and Chair of Economics at Tulane University, New Orleans. He is editor of *Public Finance Review*, a prolific researcher and publisher on the economics of tax compliance.

Email: jalm@tulane.edu

