When Does Introducing a Value-Added Tax Increase Economic Efficiency? Evidence from Synthetic Control Methods

Job Market Paper
October 28, 2015

Bibek Adhikari *

Abstract

Theory predicts that unlike sales and turnover taxes, a value-added tax (VAT) does not distort firms’ production decisions, making it an efficient tax instrument. It is this feature of the VAT that has led to its rapid adoption by more than 160 countries in less than 60 years. This prediction, however, is based on the assumption that the VAT is well-designed and well-enforced, which is often violated in practice. Thus, whether the VAT, in practice, is an efficient tax system is ultimately an empirical question. Yet empirical analysis of the efficiency effects of the VAT is surprisingly sparse. This paper provides causal estimates of the efficiency gains of adopting a VAT in a worldwide sample of countries. I identify countries that adopted the VAT between 1967 and 2000 and apply synthetic control methods to estimate their counterfactual trajectory of GDP per worker in the absence of a VAT. The synthetic control is a weighted average of countries without a VAT that closely resembles the economic structure and outcomes of the country with a VAT for several years before the adoption of a VAT. In line with previous studies, I find that the VAT has, on average, positive and economically meaningful impact on economic efficiency. However, once I allow for the impact of the VAT to be heterogeneous across countries, I find that this result is driven by richer countries only. There is no significant impact of the VAT on poorer countries. I find similar results when estimating the impact of the VAT on total factor productivity and capital stock per worker, two important channels through which a VAT affects economic efficiency. In this paper, I provide evidence that a success of VAT almost entirely depends on the initial level of income of a country, which, in result, determines whether a country is able to properly design, implement, and enforce a VAT. The findings are robust across a series of placebo studies and sensitivity checks.

Keywords: Value-Added Tax, Economic Efficiency, Synthetic Control Methods.

JEL Codes: H20, H25, O40, E6.

*Department of Economics, 206 Tilton Hall, Tulane University, New Orleans, LA 70118 (email: badhikar@tulane.edu). I am grateful to James Alm, Stefano Barbieri, Alan Barreca, Romain Duval, Keith Finlay, Sean Higgins, Michael Keen, Jaewoo Lee, Mario Mansour, Ruud de Mooij, Whitney Ruble, Steven Sheffrin, Jay Shimshack, and Dan Teles for helpful comments and discussions. Earlier versions of this paper were presented at the 2014 Annual Meetings of the National Tax Association, 2014 and 2015 Annual Meetings of the Southern Economic Association, Jobs, Growth, and Structural Reforms seminar series at the International Monetary Fund, and economics department seminar series at Tulane University. I am also thankful to the participants of these conferences and seminars for their thoughtful comments. I am especially thankful to Steve Sheffrin for providing funds to purchase industrial statistics from UNIDO, Eric Weding for adding International VAT Monitor to the library collection, and Hideki Fujioka for helping me with parallel processing of the C++ codes.
1 Introduction

The value-added tax (VAT) is a general tax on consumption that is levied on the difference between the value of a firm’s output and its inputs. In its modern form, the VAT was first introduced in Brazil and Denmark in 1967, making it one of the youngest, yet among the most important sources of government revenue. As of 2015, the VAT has been adopted by more than 160 countries and its revenues represent more than seven percent of the world’s GDP. Moreover, these numbers are expected to continue to grow because of increasing reliance on consumption taxes and because of the requirement by many multi-lateral organizations that their member countries use the VAT.

One of the key reasons for the rapid adoption of the VAT is the belief that it is an efficient tax system. Under a VAT, tax is charged at all stages of production with some mechanism to provide full credit or refund to firms of the VAT that has been charged on their input costs. Thus, a VAT leaves production undistorted as it does not affect the prices firms ultimately pay for inputs and it does not create cascading - the tax-on-tax as a good passes from one production stage to another. The theoretical justification for an efficient tax system comes from one of the fundamental results in public economics, the production efficiency theorem (Diamond and Mirrlees, 1971), which states that tax systems should be designed such that they do not distort firms’ production decisions even in second-best environments. Furthermore, the use of efficient taxes can also increase social welfare by reducing the marginal costs of providing public services.

The VAT’s efficiency is based on the assumption that it is well-designed and well-enforced. However, there are many settings under which a VAT can introduce inefficiencies. First, almost all countries exempt small firms from the VAT as the costs of administration do not justify the amount of tax collected from these firms. Exempt firms do not qualify for the refunds on their input costs, which affect the prices they pay for inputs and can introduce production inefficiencies. Second, the production efficiency theorem assumes zero tax evasion, but tax evasion is widespread, especially in developing and transitioning countries (Alm and Embaye, 2013; Schneider, 2005). The presence of tax evasion (Emran and Stiglitz, 2005) and the informal economy (Piggott and Whalley, 2001) also distort behavior, which introduces inefficiencies in the VAT system. Furthermore, there is some concern that a VAT increases tax evasion and informality because the cost of complying with a VAT is considered to be higher than that of a sales or a turnover tax, especially for small and medium enterprises (Hines, 2004; Coolidge, 2012). Thus, whether the VAT is an efficient tax system is ultimately an empirical question.

This paper claims that the question - Is a VAT an efficient tax system - is not only
empirical, but also an incomplete one.\textsuperscript{1} Although research has been done to test whether the VAT is as promising in the real world as it is in theory, these studies have ignored the heterogeneous environments under which the VATs are adopted. These include a country’s stage of development, tax capacity, tax evasion, and informal economy, as well as the design and implementation of the VAT. As a result, when it comes to the question of the VAT’s efficiency, it is difficult to come to a singular ‘yes’ or ‘no’ conclusion when factors determining it are multiple. So perhaps, the real question is - in what conditions should a VAT be adopted for it to be an efficient tax system. This paper explores the importance of one of such conditions by asking if richer countries benefit more from a VAT than poorer countries. In particular, this paper provides one of the first causal estimates of the impact of replacing distortionary taxes such as sales or turnover tax with a relatively non-distortionary VAT on production efficiency using newly available data on output-side GDP per worker. Moreover, this paper also investigates two primary channels through which a VAT affects production efficiency: by increasing total factor productivity and by increasing the stock of capital.

In addition to having limitations, previous studies on empirical analysis of the efficiency effects of a VAT is surprisingly sparse as well, largely due to the unavailability of comparable data across countries on economic efficiency. My primary data source is Penn World Table 8.1 (\textit{Feenstra, Inklaar and Timmer, 2015}), which, for the first time, includes comparable cross-country data on output-side real GDP, total factor productivity, stock of physical capital, and stock of human capital.\textsuperscript{2} I combine PWT 8.1 with various data sources, such as proprietary data on the manufacturing industry from United Nations Industrial Development Organization (UNIDO), economic and demographic variables from the World Bank’s World Development Indicators, and institutional variables from the Polity IV database. Finally, I collect data on the properties of the VAT at time of its introduction from various issues of International VAT Monitor, Tax Notes International, International Monetary Fund’s country documents, and tax authority websites.

Even when data availability is not a problem, estimating the causal effect of VAT adoption on economic efficiency using traditional approaches faces several challenges. First, countries choose whether to adopt a VAT and when to adopt a VAT. Thus, cross-country regressions are subject to numerous biases and are unlikely to reveal the causal effect of VAT on economic efficiency. One way to tackle this problem is to use propensity score matching technique to generate the probability of VAT adoption and compare countries with VAT to those countries without VAT, but with similar propensity to adopt a VAT. However,\textsuperscript{1} Similar claims have been made in \textit{Bird and Gendron (2011), Casanegra de Jantscher (1990)}, but they do not go beyond descriptive analysis or individual cases to substantiate these claims.\textsuperscript{2} Output-side GDP is a better measure of productive capacity of an economy than the standard expenditure-side real GDP, which is better at measuring average living standard of an economy.
matching techniques construct the counterfactual by matching on a single index number and, importantly, they cannot account for the biases caused by unobserved confounders. When the biases by unobservables are of concern, difference-in-differences (DID) technique can be used. This approach allows us to control for both country and year fixed effects so that all time-invariant differences across countries (such as geography) and changes common to all countries in the same year can be controlled for. However, identifying the effects of a VAT using DID requires the paths of the outcome variable between the VAT adopters and non-VAT adopters to be parallel. As Figure 17 indicates, this assumption is not satisfied in our case. Another important limitation of the DID is that it assumes the impact of the VAT on economic efficiency is homogenous across reforming countries. Thus, it does not allow us to explore whether the impacts of the VAT systematically varies across countries according to their capacity to properly design and implement it. Finally, DID cannot account for the biases caused by time-varying unobservables.

This paper makes progress in addressing all these challenges. To account for the endogeneity caused by time-varying unobserved confounders, I use synthetic control methods (SCM), a data-driven research design that constructs the counterfactual trajectory of GDP per worker in the absence of a VAT by taking the weighted average of countries without a VAT such that it closely resembles the economic structure and outcomes of the country with a VAT for several years before the adoption of a VAT. SCM calculates the dynamic effects of VAT reform on a case-by-case basis, which allows for the effect of the reform to be vary across countries and over time. This enables me to aggregate the dynamic treatment effects into four income groups according to the World Bank’s income classification based on reforming country’s level of income at time of VAT’s introduction. In addition, I also contribute to the improvement of synthetic control methods by developing a version of leave-one-out test, which is an intuitive graphical approach of testing whether the average effects are driven by a few outlier countries. This will be useful to researchers applying SCM to multiple treated units and studying the aggregate impact of the treatment (Abadie and Gardeazabal, 2003; Abadie, Diamond and Hainmueller, 2010, 2015).

The results of this research provide a new insight to the topic of ‘VAT and efficiency’. This paper finds a strong evidence of efficiency gains of VAT adoption, but only in richer countries. For instance, five years after the reform GDP per worker of the high-income group was 10.9 percent higher than the synthetic group, which increased to 11.2 percent by the end of the sample period (all significant at the 1 percent level). GDP per worker of the upper-

---

3The World Bank’s country classification starts as early as 1971 for low-income countries and 1975 for lower-middle-income countries, but the classification for upper-middle-income and high-income countries start from 1987 only, therefore I use the information from the classification of low-income countries to impute the country classification for years before 1987. For details about the imputation see Appendix 1.
middle-income group was also positive and significant at the 5 percent or better level, but after some time lag. By five year after the reform GDP per worker was 25.5 percent higher compared to the synthetic group, which increased to 33.1 percent 10 years after the reform. However, this paper finds no significant impact of VAT on poorer countries. The effect of VAT on lower-middle-income countries were slightly positive and the effect on low-income countries were mostly negative, but these estimates were statistically indistinguishable from zero at the 10 percent level. The same differential trend in the impact of VAT is found across income groups for total factor productivity and capital stock per worker, two of the primary channels thorough which VAT affects economic efficiency. These results indicate that the theoretical advantages of VAT do not necessarily translate into practice.

To probe the robustness of these results, I perform a series of additional tests. In some cases VATs were adopted as part of the trade liberalization, World Trade Organization (WTO) membership or in anticipation of joining the European Union (EU). Thus, it is important to test if their impacts are commingled. To that end, I report the average treatment effect calculated by removing the countries that liberalized their economy, joined the WTO or joined the EU in the sample period. My main results remain unchanged. Another concern is that the estimates may be biased due to reverse causation. If VAT reforms are motivated by expectation of future growth prospects, this would bias the estimates obtained from SCM as long as growth expectations are not captured by the unobservable heterogeneity included in the estimation. To mitigate the concern of reverse causality, I supplement the study with data from the manufacturing industry. The rationale being that the decision to adopt VAT takes place at the national level and thus they are not likely to be affected by one particular sector of the economy in any significant way. Again, the results are similar to the main results. Likewise, the results are similar when estimated using conventional difference-in-differences methods. I also conduct various other sensitivity checks and placebo studies and the estimates are robust to these tests.

The rest of the paper is organized as follows. Section 2 assesses key advantages and disadvantages of the VAT, provides context of its adoption, describes the features of the VATs that have been adopted, and provides a brief review of the literature. Section 3 lays out the conceptual framework to evaluate the efficiency gains of VAT adoption and discusses the data used for the estimation. Section 4 discusses the research design used to isolate the impact of VAT from various other confounding factors and discusses the sample selection procedure. Section 5 presents the main results, section 6 presents results for potential mechanisms, section 7 presents various robustness tests and section 8 concludes with some policy implications.
2 Background

2.1 Assessing Key Properties of the VAT

There are various properties of a VAT that makes it an attractive tool for raising revenue. First, by not taxing intermediate transactions, it does not distort firms’ production decisions (unlike sales and turnover taxes). To be more specific, as long as some of the taxed inputs can be substituted by untaxed inputs, any taxes on intermediate transactions will drive a wedge between the buying and selling prices of producers, violating production efficiency. Since a VAT is levied on final consumption and not on intermediate transactions between firms, it does not create a wedge between the prices that producers face in buying and selling from one another, thus, not violating production efficiency (Ebrill, Keen and Bodin, 2001; Keen and Lockwood, 2010; Tait, 1988).

Second, by fully refunding the taxes on inputs, the VAT prevents cascading - a phenomenon where a tax-on-tax occurs as a good passes from one production stage to another (for instance, from manufacturing to wholesale to retail). Cascading creates distortions in real production decisions by providing firms incentive to vertically integrate to reduce the number of stages that are subject to tax, implying efficiency loss. An alternative to the VAT that can avoid cascading is taxing only one stage of production, i.e., levy a tax only at manufacturing stage, wholesale stage, or retail stage. However, cascading is unavoidable even in single stage taxation. For instance, many producers buy inputs from manufacturers and wholesalers, or even from big box retailers. Thus, the lack of mechanism to refund taxes on inputs implies cascading even in single stage sales tax, with the probability (and magnitude) of cascading greater, the further it is from the retail stage. For instance, Ring (1989, 1999), and Mikesell (2014) estimate that about 40 percent of the retail sales tax revenue in the U.S. comes from business inputs and Smart and Bird (2009) estimate that about 43 percent of the retail sales tax revenue in Canada comes from business inputs.4

Third, theory predicts that a consumption tax, which encourages savings and investment, is superior to an income tax in terms of its growth effects (Jones and Manuelli, 1990; Lucas, 1988; Rebelo, 1991; Oakland, 1967a) and many empirical studies support this prediction (Alm and El-Ganainy, 2013; Barro, 1991; Kneller, Bleaney and Gemmell, 1999; Lee and Gordon, 2005). As Metcalf (1995) notes, VATs can be considered as a replacement for income taxes in the sense that most countries that adopt a VAT collect a significantly smaller

---

4A VAT is also superior to cascading taxes from tax incidence perspective since cascading results in a pattern of effective tax rates that depends on arbitrary things such as number of taxable production stages, making effective taxes opaque and calculation of tax incidence difficult (Ebrill, Keen and Bodin, 2001; Oakland, 1967b).
percent of tax revenues from personal and corporate income taxes. Another advantage of the VAT over an income tax is that the VAT does not influence the forms or methods of doing business since tax liability under a VAT remains the same whether the product is made in the corporate, or non-corporate sector or, whether capital-intensive, or labor-intensive technology is used in production (Cnossen, 1998).

Fourth, leakages in revenue are less likely under the VAT than under sales taxes because under a VAT revenue is collected in all stages of production rather than at the final sale. The retail stage is also considered to be the weakest link in the chain to collect revenue. Furthermore, the invoice driven VAT system can provide an adequate paper trail that facilitates enforcement and makes tax evasion more difficult (Ebrill, Keen and Bodin, 2001; Pomeranz, 2015).

A VAT, however, is not without its weaknesses. As discussed earlier, exempting some firms from the VAT or the presence of an informal economy introduces some inefficiency in the VAT system. Further, the credit and refund mechanism of the VAT creates opportunities for fraud that are unique to a VAT. Some of the examples are false claims for credit or refund, credit claimed for a VAT on purchases that are not creditable, and creating fake invoices so that credit can be claimed (Keen and Smith, 2006).

There are two major criticisms of the VAT. The first criticism, ironically, is the perceived efficiency of the VAT. Critics argue that politicians exploit the lower marginal cost of raising revenue through a VAT to raise more revenue than that is economically optimal, leading to an inefficiently large level of government spending (Brennan and Buchanan, 1977). For instance, the 2005 presidential panel on tax reform in the United States could not reach a consensus on whether to recommend introducing VAT in the U.S., primarily because “Some members were also concerned that introducing a VAT would lead to higher total tax collections over time and facilitate the development of a larger federal government - in other words, that the VAT would be a “money machine.” (The President’s Advisory Panel on Federal Tax Reform, 2005).

The second criticism of the VAT is that it is a regressive tax system. Critics argue that any consumption tax will impose a heavier burden on low-income families than on high-income families because the fraction of annual income spent on consumption tends to be higher for low-income than for high-income families. Also, if a higher reliance on the consumption tax leads to less reliance on the income tax, it means that there will be less redistribution via the income tax. Similarly, the consumption tax exempts capital and capital income from the tax base, which reduces the tax burden on high-income families, while increasing the burden on low-income families.


2.2 The Context of VAT Reforms

All tax policies are endogenous to some extent and one of the main reasons to use synthetic control methods (SCM) is its ability to account for unobserved heterogeneity more flexibly than traditional panel regressions or matching techniques. That said, the context of VAT adoptions provide a unique and relatively more exogenous setting to assess the impact of tax reforms on economic efficiency.

First, a VAT take-up decisions are often influenced by external forces that are arguably exogenous to the internal conditions of the reforming countries. Many academic and non-academic sources highlight the role of multi-lateral organizations, especially the European Union (EU) and the International Monetary Fund (IMF), in influencing the take-up decision of a VAT. The EU requires the member states to adopt a VAT upon entry to the EU. So, any country joining the EU or aspiring to join the EU needs to adopt a VAT. Similarly, the IMF is a strong advocate of the VAT and often puts the adoption of a VAT as one of the conditions for providing a loan, or other assistance. Thus, any country that needs IMF’s help has a much higher probability of adopting a VAT. For instance, Keen and Lockwood (2010), Ufier (2014), and Cizek, Lei and Ligthart (2012) find that the probability of a VAT adoption increases by up to 25 percent in the year following country’s participation in IMF program. Ebrill, Keen and Bodin (2001) estimate that more than half of all countries that introduced a VAT during 1980s and 1990s used IMF advice in doing so.\(^5\)

Second, as Case, Rosen and Hines (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, which is often referred to as the “bandwagon” theory of policy adoption. If countries adopt a VAT 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 their internal economic conditions. The map documenting the spread of the VAT is shown in Figure 1 and it clearly suggests that the VAT was adopted in regional waves. For instance, more than 11 other European countries adopted a VAT between 1967 when it was first introduced and 1973. Similarly, 11 other Latin American countries introduced the VAT within a decade of Brazil’s decision to adopt a VAT. The spread of the VAT slowed down in the first half of the 1980s, but accelerated again in the second half of the 1980s. The 1990s

---

\(^5\) Riswold (2004) criticizes the role of IMF in Sub-Saharan Africa’s VAT adoption. James (2015) argues that a significant role is played by other multi-lateral organizations such as the World Bank (e.g., in the cases of Cameroon, Ghana, Hungary, Philippines, Thailand, Tanzania, and Venezuela), the World Trade Organization, Inter-American Development Bank, Asian Development Bank, the Organization for Economic Co-operation and Development, as well as aid and development agencies such as United States Agency for International Development (e.g., in the cases of Serbia, Egypt, El Salvador, Guatemala, and Jamaica), German Aid and Development Agency (e.g., in the cases of Serbia, Croatia, Macedonia, and Pakistan), and the UK Department for International Development (e.g., in the cases of Ghana, and Pakistan).
saw more than 70 countries introduce a VAT. The copycat behavior is particularly strong in the Eastern Europe where 18 countries adopted a VAT within five years of Hungary’s VAT adoption in 1988. A similar pattern is also found in the developing countries of the Sub-Saharan Africa and Asia. This trend has continued in the 2000s, with about 40 more countries introducing a VAT by 2015. Moreover, the copycat behavior in VAT adoption is also demonstrated in Keen and Lockwood (2010), Ufier (2014), and Cizek, Lei and Ligthart (2012).

2.3 Features of the VATs Adopted

In this section, I describe the properties of the VATs adopted and compare them with the ideal VAT. These information are not readily available so I collect them from various sources. I start with Tait (1988) and Cnossen (1998) and supplement them by the data from various issues of International VAT Monitor, Tax Notes International, IMF’s country documents, and tax authority websites. Even so, I do not have the information for some of the countries with a VAT. First, I present the summary statistics for all the countries for which I was able to collect data. I then focus on the sample of 33 countries that are analyzed in this study. I divide these countries into four income groups to highlight how the characteristics of the VAT differ among them.

An ideal VAT will have a broad base, that is, it will cover most of the goods and services consumed. What are the bases of the real-world VATs? Out of 102 countries with a VAT for which I have data, 81 of them indeed taxed both goods and services, while 13 more countries taxed goods and services, but only selected services were included in the tax base. The rest of the 8 countries’ VAT base included goods only and in some cases, the VAT base also included capital goods.

How broad is the coverage of the VATs in practice? Again, an ideal VAT would also have a broad coverage, that is, it would tax all stages of production through the retail stage. Out of 102 countries with data, 92 of them indeed taxed through the retail stage, while 3 of them extended the VAT through the wholesale stage and 7 of them only extended it through the manufacturing stage.

The VAT was designed to replace various forms of sales taxes, turnover taxes, and tariffs that distorted economic activity. Out of 58 countries for which I have data, 21 replaced retail sales tax with a VAT, 19 replaced turnover taxes with a VAT, and 7 replaced both the turnover tax and the sales tax. The 1990s and later saw the adoption of the VAT by transitioning countries and the developing countries from Africa and Asia. The transitioning countries replaced the traditional sources of revenue such as levies on state enterprises with
a VAT and the developing countries from Asia and Africa replaced sales and turnover taxes with a VAT. Another reason for a VAT adoption by developing countries was to offset the loss in revenue from tariffs due to deepening of trade liberalizations.

What was the proposed effect of the VAT on government revenue? Out of 46 countries for which I have data, 27 of them designed VAT to yield equal revenue as the predecessor taxes it replaced, while 15 of them designed it to increase revenue, and 2 of them to decrease it. The remaining 2 countries designed VAT such that the revenue would either remain the same or increase.

The 13 countries from the high-income group analyzed in this study are Austria, Belgium, Canada, Denmark, France, Greece, Italy, Japan, Netherlands, New Zealand, Spain, Sweden, and United Kingdom. All of them had a broad base VAT reform that levied taxes through retail stage. All of the countries introduced the VAT to replace some kind of sales or turnover taxes while some countries also adjusted their income or payroll taxes at the same time. Most of the VATs were designed to be revenue neutral, with the exceptions of Canada, Denmark, and New Zealand that designed their VAT to increase revenue and the United Kingdom that designed its VAT to decrease revenue. The average of the standard VAT rate was 17.2 percent.

The 6 upper-middle-income countries included in this study are Argentina, Chile, Ireland, Mauritius, Panama, and Portugal. All of them had a broad base VAT reform that included both goods and services into the tax base and levied VAT through retail stage, except for Argentina which levied taxes through retail stage, but only taxed selective services. In terms of revenue yield, 4 of them were designed to be revenue neutral, and 2 (Chile and Panama) of them to increase revenue. All of the countries introduced VAT to replace some kind of sales or turnover taxes while some countries also reduced tariffs and eliminated stamp taxes. The average of the standard VAT rate was 15.3 percent.

The 9 countries from lower-middle-income group analyzed in the study are Colombia, Costa Rica, the Dominican Republic, Honduras, Jamaica, Peru, Senegal, Thailand, and Uruguay. Six out of 9 of the countries had a broad base VAT reform that included both goods and services into its base and levied the taxes through retail stage. The 3 exceptions were Costa Rica, the Dominican Republic, and Jamaica. Costa Rica and Jamaica taxed selected services only. The Dominican Republic extended its VAT only through the manufacturing stage, it also taxed selected services and included some capital goods in the tax base. The average of the standard VAT rate was 15.1 percent.

The 5 countries from low-income group included in the study are Bangladesh, Guinea, Kenya, Nepal, and Pakistan and only Nepal had a broad base VAT that taxed both goods and services and that extended the VAT through retail stage. The rest of them only taxed
selective services and extended the VAT through the manufacturing stage only. Pakistan also included some capital goods in the tax base. The average standard VAT rate was 15.2 percent.

To conclude, most of the VATs in practice have broad tax base and broad coverage. Nevertheless, there are some significant exceptions, especially in developing countries. In addition to this, even when the VAT is well-designed, factors related to the development stage of a country, such as tax capacity, informal economy, and tax evasion influence the effectiveness of the VAT. Thus, the question still remains whether the adoption of the VAT results in efficiency gains or not.

2.4 Related Literature

There are many applied general equilibrium models evaluating the economic effects of VAT adoptions or some changes in the structure of the VAT. The main advantage of this methodology is that it provides a clear bridge between theoretical and applied aspect of tax policy analysis. These papers use data from a particular country to calibrate their model. For instance, Boeters et al. (2010) use data from Germany to evaluate the impact of a VAT simplification reform on economic efficiency, Ballard, Scholz and Shoven (1987) use data from U.S. to examine the impact of hypothetical VAT adoption on economic efficiency, Bye, Strom and Avitsland (2012) use data from Norway to estimate efficiency effects of various VAT base-broadening scenarios, and Bovenberg (1987) uses data from Thailand to evaluate the impact of a VAT on economic efficiency. In general, these studies find positive impact of VAT on economic efficiency and other macroeconomic variables. One thing common across all these studies is that they assume there is no informal economy. Piggott and Whalley (2001) use data from Canada where they model the presence of informal economy and assume that a VAT encourages some firms to switch to the informal sector and some individuals to increase home production, both of which are plausibly inefficient compared to the formal sector. Piggott and Whalley (2001) find that VAT base-broadening reduces economic efficiency. Similarly, Bye, Strom and Avitsland (2012) find that including selective services in the VAT base reduces welfare compared to both not including services in the VAT base and including all services in the VAT base. They highlight the role of informal economy and a narrow VAT base in undermining the effectiveness of the VAT, ignoring which can lead to an upward bias of the estimated results.

The applied general equilibrium models of the VAT reforms have important limitations. For instance, they make strong and ad-hoc assumptions about the model’s functional forms, elasticity type, tax treatment, market structure, technology type, and so on that usually do
not hold in the real world. This methodology also suffers from the lack of statistical test to confirm the validity of model specifications. Finally, almost all of these are single-country studies focusing on high-income countries only (Bovenberg, 1987; Shoven and Whalley, 1984; Andre, Cardenete and Romero, 2010).

In contrast, there are relatively few reduced-form studies on the economic effects of the VAT. Nellor (1987) is one of the earliest attempts to empirically study the impact of a VAT on economic efficiency. Nellor (1987) analyzes 11 European countries that introduced a VAT in the 1960s and 1970s using cross-country regression and finds that introduction of a VAT raised revenue ratio in those countries. This increase in revenue ratio provides indirect evidence on the efficiency gains of a VAT adoption. Keen and Lockwood (2010) prove that, under some weak conditions, access to a more efficient tax instrument leads an optimizing government to increase the revenue ratio. In two recent studies, Keen and Lockwood (2006) using a panel of Organization for Economic Co-operation and Development (OECD) countries and Keen and Lockwood (2010) using an unbalanced panel of 143 countries over 26 years, estimate the effect of the VAT on the revenue ratio using fixed effects regression and find that a VAT adoption is associated with an increase in economic efficiency. Similarly, Smart and Bird (2009) and Ferede and Dahlby (2012) use DID research design and province-level data from Canada and find that investment and economic growth respond positively when a VAT replaces a sales tax. More recently, Ufier (2014) looks at the effect of the VAT on various macroeconomic variables such as growth, investment, trade, inflation, and government size using propensity score matching and finds evidence that VAT increases investment and finds weak evidence that it increases growth.

Most of the studies mentioned above use standard cross-country regression methodology that do not attempt to identify the causal link between a VAT adoption and economic efficiency. Smart and Bird (2009) and Ferede and Dahlby (2012) are the only studies that attempt to establish causality while also controlling for the biases caused by time-invariant unobservables, but they only analyze the impact of province-level VAT in Canada. My study is more similar to Ufier (2014) in that both of us account for the selection of a country into a VAT. However, my paper differs from Ufier (2014) both in terms of the methodology used and the topics analyzed. In terms of the methodology, I use SCM which can flexibly control for both observed and unobserved confounders unlike propensity score matching which can only account for observed confounders. Second, SCM constructs the counterfactual by matching the level and the trend of the outcome variable (10 years in this study) unlike propensity score matching that constructs the counterfactual by matching on a single index number. Thus, SCM produces more credible counterfactual. In terms of the topics analyzed, I focus on the impact of a VAT adoption on economic efficiency and I empirically estimate the
mechanisms through which a VAT can affect economic efficiency. Moreover, rather than estimating the average impact of the VAT across all countries, I allow for the impact of the VAT to vary across countries. This allows me to provide new insights: that the impact of a VAT depends on the initial level of income of a country, which can lead to very different policy implications.

3 Conceptual Framework and Data

3.1 Conceptual Framework

Suppose the production function in terms of output per worker is given by:

\[ y_{it} = A_{it} f(k_{it}, h_{it}) \]  (1)

where \( y_{it} \) is the output per worker of country \( i \) at time \( t \), \( A \) is the total factor productivity (TFP), \( k \) is the stock of physical capital per worker, and \( h \) is the average stock of human capital. Thus, higher values of \( k \) and \( h \) means that the same \( A \) leads to more output and a higher value of \( A \) means that the same inputs lead to more output.

I focus on the effect of VAT introduction on \( y \), \( A \), and \( k \). GDP per worker provides a direct empirical identification of efficiency gains associated with the replacement of sales or turnover taxes with a VAT. An increase in production efficiency means the production of some outputs can be increased without decreasing the production of any other outputs. Thus, the adoption of an efficient tax system should lead to higher aggregate output.

The introduction of a VAT can increase output through two channels. First, when a distortionary tax such as a sales tax or a turnover tax is replaced with a relatively non-distortionary tax such as a VAT, a profit maximizing firm will have an incentive to allocate its resources to their most productive uses, which increases \( A \), and thus leads to higher output. Second, as investment goods are not taxed under a VAT, a move from sales tax or a turnover tax to a VAT will decrease the cost of capital and remove the bias against the use of capital-intensive technology in production, which increases \( k \), and thus leads to higher aggregate output.

3.2 Data

To study the impact of the VAT on production efficiency, I create a country-by-year panel for the period between 1950 and 2010 by combining various data sources.
My primary data source is Penn World Table (PWT) version 8.1, which covers 167 countries between 1950 and 2011 (Feenstra, Inklaar and Timmer, 2015). PWT 8.1 has made some major additions to the database such as reintroducing measures of capital stock and introducing for the first time measures of human capital, the share of labor income in GDP and total factor productivity. This implies that for the first time, there now exists a database that is highly comparable across countries and over time, which can be used for comparing production efficiency, TFP, and capital stocks. Prior to PWT 8.1, the data on TFP was only available for a small sample of countries, with most of them being high-income countries because the construction of TFP is very difficult as it requires comparable data on output, labor stock, and capital stock. Even so, TFP data prior to PWT 8.1 came from various different databases, making comparisons of countries from different income groups problematic. Furthermore, PWT 8.1 uses an improved methodology to compute these series, making PWT 8.1 the best database for the study.

The outcome of interest is output-side real GDP per worker, which is calculated by dividing output-side real GDP by number of persons engaged, both of which are from the PWT 8.1. Feenstra, Inklaar and Timmer (2015) argue that output-side real GDP per worker is a more appropriate measure of the production possibility of the country compared to commonly used expenditure-side GDP per worker. Expenditure-side GDP reflects the standard of living of the country, which can differ significantly from the productive capacity of the country due to differences in the terms of trade the country faces. For instance, small open countries such as Singapore and Luxembourg enjoy a much higher standard of living than what their productive capacities allow, due to favorable terms of trade. However, I am interested in measuring relative production efficiency and not the relative living standard of the countries. Hence, output-side GDP per worker is an appropriate measure (Inklaar and Timmer, 2013).

The main explanatory variables are stock of physical capital per worker ($k$), and stock of human capital per worker ($h$) and they are obtained from the PWT. Below is a brief discussion of how these variables are constructed. For further details see Inklaar and Timmer (2013).

Capital stock per worker is obtained by dividing capital stock by number of persons engaged. For the construction of capital stock, the PWT uses the perpetual inventory method. Unlike previous databases that assume constant depreciation rate for all assets, PWT divides assets into six categories (structures, transport equipment, computers, communication equipment, software, and other machinery and assets) and allows the depreciation rate to

---

6PWT 5.6 had data on physical capital stocks up to 1992 and was based on older basic data using a single depreciation rate.
vary across each category. This will adjust for differences in composition of investment across countries and over time, thus producing tighter estimates of capital stock.

Average stock of human capital is calculated as a function of the average years of schooling for the population aged 15 and older from Barro and Lee (2012) and rates of return of schooling based on Psacharopoulos (1994). The return to schooling is given by a piece-wise linear function to account for the fact that early years of education have a higher return (i.e., higher wages) than the later years.

To anchor the paper to the endogenous growth model, I include additional control variables that are meant to capture the impact of institutions (democracy and trade openness), macroeconomic environment (inflation rate), and demography (population growth rate) on the production possibilities (Barro, 1991; Mankiw, Romer and Weil, 1992). Data on trade as a share of GDP (trade openness) and population growth are also obtained from the PWT. Data to construct a measure of democracy are obtained from Polity IV, which contains unbalanced panel data on the types and qualities of government of 167 countries spanning 1800 to 2012 (Marshall and Jaggers, 2013). Following Persson and Tabellini (2007), I classify a country as democratic if the “polity2” in the Polity IV data set is strictly positive. The data on inflation are obtained from the World Bank’s World Development Indicators.

The data on the treatment indicator, that is the year VAT was introduced, is obtained primarily from Ebrill, Keen and Bodin (2001) and supplemented by International Tax Dialogue (2005) and Ufier (2014).

4 Empirical Strategy

In order to isolate the impact of a VAT adoption from other influences, I employ synthetic control methods (SCM), a data-driven way of finding the most appropriate counterfactual in generalized difference-in-differences (DID) estimation. DID estimation consists of identifying a specific treatment, 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. The primary motivation to use synthetic control methods is the belief that the effect of a particular intervention can be empirically assessed only by comparison with the appropriate counterfactual. The SCM was developed by Abadie and Gardeazabal (2003) and expanded by Abadie, Diamond and Hainmueller (2010, 2015). Abadie and Gardeazabal (2003) use the approach to quantify the impact of political violence on economic growth in the Basque region of Spain, Abadie, Diamond and Hainmueller (2010) use it to estimate the impact of a large anti-tobacco initiative on the per capita sales of cigarettes in California, and Abadie, Diamond and Hainmueller (2010) use it to assess the
economic effects of the 1990 German reunification in West Germany. The use of SCM has been growing, and it is now being applied to study a very diverse set of topics.  

There are numerous advantages of using the SCM in my setting. First, SCM allows for the effect of the reform to be heterogeneous across both countries and over time, as opposed to the DID estimator where only average effects can be analyzed. The average effects can mask large differences across countries. This is especially important for my purposes because the broader environment under which the VAT is adopted, such as the stage of development of a country, design of the VAT, and its implementation vary significantly across countries, thus, the assumption that the impacts across countries are homogenous might not be valid. Indeed, I find very heterogeneous impacts of the reforms.

Second, SCM can substantially reduce any potential endogeneity problem caused by the omitted variables. Abadie, Diamond and Hainmueller (2010) prove that if a synthetic country can be found such that it matches the pre-treatment trajectory of the 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. The intuitive explanation is that only countries that are alike in both observed and unobserved predictors of the outcome variable as well as in the effect of those predictors on the outcome variable should produce similar trajectories of the outcome variable over extended periods of time.

Third and finally, in comparative case studies, the researcher is allowed to choose the comparison group that resembles the counterfactual. However, Abadie, Diamond and Hainmueller (2010) argue that this introduces substantive 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. The SCM resolves this problem by using a data-driven method to create a comparison unit that closely resembles the fundamentals of the treated unit in the pre-treatment period, using a weighted average of all other control units. 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. By taking the weighted average of control countries, SCM makes explicit the

---

7These applications include: the effects of trade liberalization on economic growth (Nannicini and Billmeier, 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 (Abdallah 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, 2013), 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 (Smith, 2015).
contribution of each comparison unit to the counterfactual of interest. This provides better transparency. Contrast this with a regression technique like conventional DID where the counterfactual is constructed using the regression-weighted average of all control countries. However, these weights are implicitly calculated and are not reported in practice, which makes them opaque. Moreover, since the weights in SCM are restricted to be positive and sum to one, it safeguards against extrapolation bias that plagues traditional regression models. Similarly, the weighted average of a few similar countries can provide a more credible counterfactual than the regression-weighted average of all control countries.

In the absence of experimental variation, quasi-experimental identification is not possible without identification assumptions. I discuss these assumptions as well as a few limitations of the SCM. First, other sources of endogeneity like reverse causation may still be present even with SCM. If VAT reforms are motivated by expectation of future growth prospects, this would bias the estimates obtained from SCM as long as growth expectations are not captured by the unobservable heterogeneity included in the estimation. To mitigate the concern of reverse causality, I supplement the study with data from the manufacturing industry. The rationale being that tax reforms take place at the national level and thus they are not likely to be affected by one particular sector in any significant way.

Second, if the treated country or control countries with positive weights experience major idiosyncratic shocks or other policy changes in the sample period, then the SCM can confound the impact of these shocks to the impact of VAT reform. This is especially true in the post-reform period. I carefully select the control group by removing countries that were known to have shocks of bigger magnitude (e.g., civil war) or those that have higher propensity to experience idiosyncratic shocks (e.g., small island economies, resource-rich economies, and Eastern European countries). In some cases VAT reforms were adopted as part of the trade liberalization or in anticipation of joining the EU. Thus, their impacts are likely commingled. To test whether the results are commingled, I also report the average treatment effect calculated by removing the countries that liberalized their economy in the sample period using two broad measures of economic liberalization: WTO membership and a binary indicator developed by Sachs and Warner (1995) and updated by Wacziarg and Welch (2008). Similarly, I also report the average treatment effect calculated by removing the countries that joined the EU in the sample period.

Third, I assume that the common support exists. That is, I assume that there exists 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 worker and its predictors) of the treated country. I use pre-treatment fit index to assess the similarity of the treated country to the synthetic country. In some cases, I am unable to
find a synthetic control that closely resembles the treated country. To confine the analysis to the common support, I follow the practice in other matching techniques and exclude such cases from the analysis.

Finally, I also assume no interferences between countries, and no anticipation effects. To meet the assumption of no interference between countries, I exclude those countries that adopted the VAT before or in the sample period from the control group. If some of the countries with positive weights in the synthetic control undergo piecemeal tax reforms that increases economic efficiency in response to the implementation of the VAT by my country of interest, then this will bias the treatment effect towards zero. 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. As explained later, I test for the validity of no anticipation effects by using difference-in-differences with leads and lags. The results are not affected.

4.1 Synthetic Control Methods

Suppose that we have \( J + 1 \) countries, where country 1 is a treated country and the remaining \( J \) countries act as potential controls called the donor pool. 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 for country \( i \) at time \( t \) with reform (R). The sample period is given by \( t = 1, \ldots, T_0, T_0+1, \ldots, T \), where \( T_0 \) denotes the number of pre-treatment periods and \( T_0+1 \) denotes the treatment year. The observed outcome variable can be written as:

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

(2)

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

For the treated country, we can observe \( Y_{it}^{R} \), however, we need to estimate the counterfactual \( Y_{it}^{NR} \), which is the GDP per worker of the country that adopted VAT had the country not adopted it. In order to estimate the counterfactual I use linear factor model of the form:

\[
Y_{it}^{NR} = \alpha_t + \theta_t Z_i + \lambda_t \mu_i + D_i(R_i = R_j) + D_i(I_i = I_j) + \epsilon_{it},
\]  

(3)

where \( \alpha_t \) is an unknown common factor with constant factor loadings across countries, \( Z_i \) is a vector of observed covariates with coefficients \( \theta_t \), \( \mu_i \) is a vector of unknown parameters, \( \lambda_t \) is a vector of unobserved common factors, and \( \epsilon_{it} \) are the idiosyncratic error terms with
zero means. $D_i(R_i = R_j)$ is a vector of indicator functions that become 1 when the treated country and the control country belongs to the same geographic region, and $D_i(I_i = I_j)$ is a vector of indicator functions that become 1 when the treated country and the control country belongs to the same income group. Note that this specification allows country effects to vary with time ($\lambda_t\mu_i$), unlike in difference-in-differences that restricts country effects to be time invariant ($\lambda\mu_i$).

Let us 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, $W = (w_2, \ldots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \ldots, J + 1$ and $w_2 + \cdots + w_{J+1} = 1$, where vector $W$ represents a potential synthetic control. Then the outcome variable for each potential synthetic control unit is given by:

$$J+1 \sum_{j=2}^{J+1} w_j Y_{jt} = \alpha_t + \theta_t J+1 \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_i + \sum_{j=2}^{J+1} w_j \epsilon_{jt} \quad (4)$$

Now suppose that there is a vector of potential control countries, $W^*$, with the elements $(w^*_2, \ldots, w^*_{J+1})$, such that:

$$J+1 \sum_{j=2}^{J+1} w^*_j Y_{jt} = Y_{11}, \ldots, J+1 \sum_{j=2}^{J+1} w^*_j Y_{jT_0} = Y_{1T_0} \text{ and } J \sum_{j=2}^{J} w^*_j Z_j = Z_1 \text{ holds.} \quad (5)$$

Abadie, Diamond and Hainmueller (2010) prove that the size of the bias in the difference between the post-treatment outcome variable for the treated and synthetic country is close to zero if the number of pre-intervention period is large relative to the scale of the transitory shocks. Thus, the dynamic treatment effect for time $t$ ($\tau_{1t}$), where $t \in \{T_0+1, \ldots, T\}$ can be estimated by $\hat{\tau}_{1t}$ where:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w^*_j Y_{jt} \quad (6)$$

Importantly, the authors show that this holds even for the case of a strong but imperfect fit.

To find such $W^*$, let $X_1$ be $(k \times 1)$ vector containing the values of the pre-treatment variables of the treated country that we aim to match as closely as possible. Let $X_0$ be the $k \times J$ matrix where each column of the matrix is a vector of the same pre-treatment variables for each potential donor country. Then, the differences between the pre-treatment characteristics of the treated unit and its synthetic control is given by $\|X_1 - X_0W\|$. 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 semidefinite and diagonal matrix such that the root of the
mean squared prediction error (RMSPE) of the outcome variable is minimized for the pre-intervention periods. By minimizing the RMSPE, the algorithm assigns larger weights to those pre-treatment variables that have the highest predictive power.

I use a cross-validation method to choose the matrix of predictors weight $V$. If all available pre-treatment data is used to both choose matrices $V$ and $W$ and to minimize the RMSPE then the algorithm can make a more accurate prediction on the data it has seen (i.e., pre-treatment data), but it may perform poorly on the data it has not seen (i.e., post-treatment data). This phenomenon is called over-fitting in the statistical literature. It is crucial to insure that the algorithm performs well in the post-treatment data, since I am interested in constructing the counterfactual trajectory of GDP per capita such that it can accurately reflect how the trajectory of GDP per capita would evolve in the absence of the treatment. To avoid over-fitting, I use a cross-validation technique suggested by Abadie, Diamond and Hainmueller (2015) to choose the matrix of weights $V$ that minimizes out-of-sample prediction errors. This is achieved by splitting the pre-treatment data into two halves and using the first half of the data for training and the second half for validation. To elaborate, I use the average values of the predictors from the training period to select the predictor weights $V$ such that the resulting synthetic unit minimizes RMSPE over the validation period. Next, I use $V$ from the previous step and the average values of the predictors from the validation period to estimate a synthetic unit that minimizes the RMSPE for the entire pre-treatment period.

To assess whether the comparison country created using SCM is a good counterfactual, we need some measure of how well it resembles the treated country before the treatment. Abadie, Diamond and Hainmueller (2010) use RMSPE of the outcome variable to measure fit or lack of fit between the paths of the outcome variable for treated country and its synthetic counterpart. Following Adhikari and Alm (2015), I use the pre-treatment fit index to assess the overall quality of the pre-treatment fit. The pre-treatment fit index is given by:

$$\text{Pre-treatment fit index} = \frac{\text{RMSPE}}{\sqrt{\frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{1t})^2}}$$

A pre-treatment fit index of $X$ implies that the fit of the path of the outcome variable of treated country and its synthetic control is equal to that created by a $100X$ percent deviation of outcome variable on each pre-treatment period.

One advantage of using the pre-treatment fit index rather than RMSPE is that it normal-
izes RMSPE, which makes it possible to compare the pre-treatment fit between the synthetic control units across different countries, such as when GDP per worker 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.

Next, I aggregate the country-specific treatment effects into four groups based on income classification of the country during the year of VAT adoption. Aggregating it across countries of group $G$, where $G$ contains countries $\{1, \ldots, g, \ldots, G\}$ gives the average dynamic treatment effect for group $G$ at time $t$, or:

$$DTE_t^G = \frac{\sum_{g=1}^G \hat{\tau}_{gt}}{G} = \frac{\sum_{g=1}^G Y_{gt}^g - \sum_{j=2}^{J+1} w_{jt}^g Y_{jt}^g}{G}$$ (8)

Following Cavallo et al. (2013), I normalize the estimates before aggregating them by setting the GDP per worker of each treated country equal to one in the treatment year.

4.2 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) among others, can be used to evaluate the significance of treatment effects. The essence of placebo experiments is to test whether the estimated impact of the VAT adoption could be driven entirely by chance. Specifically, I 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 assuming that they implemented VAT reform in the same year as our country of interest and running the synthetic control method. 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 can conclude that there is no statistically significant evidence of an effect of a VAT 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 VAT within our sample period, then we conclude that there is statistically significant evidence of an impact of a VAT reform in the treated country. Assuming that we are doing inference about a positive treatment effect, the p-value at each
post-treatment year \( t \in \{T_0 + 2, T_0 + 3, \ldots, T\} \) is given by:

\[
Pr(\hat{\tau}_{1t} < \hat{\tau}_{jt}) = \frac{\sum_{j=2}^{J+1} 1(\hat{\tau}_{1t} < \hat{\tau}_{jt})}{J},
\]

where \( j = 1 \) denotes treated unit and \( j \neq 1 \) denotes placebo units and \( 1(\cdot) \) is the indicator function.

Since I am interested in drawing inference about the significance of the aggregate effects, I need to account for the fact that the average smooths out some noise. To that end, I follow Cavallo et al. (2013) and compute p-values for the average treatment effect for group \( g \) at each post-intervention year \( t \in T_{0+2}, \ldots, T \) according to the following steps. First, for each country \( i \) of group \( g \), I compute the placebo effect for all \( J_i \) placebo units from the placebo pool. Second, at each post-treatment year, I 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 \( N_{PL} \) placebo averages where \( N_{PL} = \prod_{j=1}^{G} J_i \). Third, at each post-treatment year, I calculate the p-values for the dynamic treatment effect obtained in Equation 8 by using the following equation.

\[
Pr(\tau_{gt} < \tau_{PL gt}) = \frac{\sum_{N_{PL}} 1(\tau_{gt} < \tau_{PL gt})}{N_{PL}},
\]

where \( 1(.) \) is the indicator function, \( \tau_{gt} \) is the average treatment effect for group \( g \) at time \( t \) after the treatment and \( \tau_{PL gt} \) is the placebo average treatment effect for group \( g \) at time \( t \).

### 4.3 Sample Selection

The first step in using the SCM to evaluate the impact of the VAT reform is to choose suitable countries for analysis from the group of countries that adopted VAT. Since I want to analyze the impact of the VAT up to 10 years after the treatment, I choose countries that adopted a VAT before 2000. Next, SCM requires at least a few pre-treatment observations of GDP per worker to calibrate the synthetic control. If some treated countries do not have pre-treatment data on GDP per worker, then they are excluded from the study. When data availability is not an issue, I restrict the sample period to 10 pre-treatment years to calibrate the synthetic unit and 10 post-treatment period to evaluate the impact of the treatment. It becomes difficult to isolate the impact of a VAT if the country experiences major shocks to its economy around the time of the VAT adoption. Thus, I also exclude countries that were known to have shocks of bigger magnitude (e.g., civil war) or those that have higher

\[\footnote{Note that this number grows rapidly when there are many countries in the group or when there are many placebo units for each country in the group, and especially when both are true.} \]
propensity to experience idiosyncratic shocks such as, transitional countries, small island countries, and resource-rich countries. In my estimation, I include the following set of predictors while estimating the synthetic unit: capital stock per worker, average human capital, a democracy dummy, openness, inflation, and population growth. However, if the country of interest does not have a single observation before the treatment for the relevant predictor, I drop that predictor.

The second step in using the SCM is to select the donor pool for each treated country. For any country to be included in the donor pool, it needs to meet two requirements. First, there cannot be any missing observations for the outcome variable in the sample period. Second, there must be at least one non-missing observation before the treatment for each of the covariates used in the estimation. Any country not meeting these requirements is dropped from the donor pool. Further, I exclude from the potential donor pool any country that adopted a VAT before or within the sample period because the synthetic unit is meant to reproduce the path of the 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 treated in the time period implies that the synthetic unit is not reproducing the potential outcome in the absence of treatment. I also exclude transitional countries, small island countries, resource-rich countries, and countries that went through civil war in the sample period from the donor pool. The next step is to run the synthetic control algorithm.

In the third and final step, I select treated countries for which SCM was able to produce a synthetic unit with a good pre-treatment fit to aggregate their effects by income group. When the pre-treatment fit is bad, we lose the confidence that the treatment effect is due to the reform, because it could be due to the inability of the synthetic unit to mimic the path of the outcome variable of the treated unit. Note that discarding from the analysis the unmatched treatments is similar to confining the analysis to the common support when using matching estimators. I use pre-treatment fit index to decide the quality of the fit, and exclude any country with a pre-treatment fit index greater than 0.10. 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. Therefore, I only include placebo countries with pre-treatment fit index less than or equal to 0.10 in the placebo experiment. In some cases, SCM is unable to find any placebo countries with a good pre-treatment fit, leaving the placebo pool empty. When that happens, I also exclude those treated countries with the empty placebo pool.

10Almost all of the transitional countries adopted a VAT within few years of transitioning to a market economy, thus the effects would be commingled with the effects of wider structural reforms. Very few small island countries and resource-rich countries have a VAT or any sales or turnover taxes for that matter.
5 Main Results

This section presents the results of the empirical analysis. I first present evidence that the VAT increases overall economic efficiency in the sample of 33 treated countries. I then analyze the heterogeneity in the effects across countries, showing that a VAT is efficiency improving in the high-income and the upper-middle-income countries, but not in the lower-middle-income and the low-income countries.

Figure 2 presents the average causal impact of VAT reforms on GDP per worker for all treated countries. The overall effects are positive and economically meaningful. Five years after the reform, the average GDP per worker of all treated countries is 7.6 percent higher than the synthetic group. Overall, post-reform GDP per worker is 6.1 percent higher in the treated group compared to the control group. However, are these results representative of all countries?

To test whether we find differential effects of the VAT adoption by the development stage at which it was adopted, I divide the countries into four groups according to the World Bank’s income classification: high-income countries (H), upper-middle-income countries (UM), lower-middle-income countries (LM), and low-income countries (L). This gives us 13 countries in the high-income group, 6 countries in the upper-middle-income group, 9 countries in the lower-middle-income group, and 5 countries in the low-income group.\footnote{I exclude 6 treated countries from the analysis because the SCM algorithm could not find any placebo countries with pre-treatment fit less than or equal to 0.10 for these countries. Similarly, I exclude 22 treated countries from the analysis because the SCM algorithm produced the synthetic controls with pre-treatment fit greater than 0.10. The majority of these countries were from the low-income group (18), followed by lower-middle-income group (6), high-income group (3), and upper-middle-income-group (1). When I increase the pre-treatment fit threshold to include countries with the fit index less than or equal to 0.20, the number of countries in the low-income group increases to 11, the number of countries in the lower-middle-income group increases to 10, and the number of countries in the upper-middle-income group increases to 6. Even then, the impact of the VAT on economic efficiency displays same differential trends across income groups.}

As a first step, I graph country-specific average treatment effects obtained from running SCM in Figure 3. The horizontal axis represents GDP per worker during the year of VAT adoption normalized by the GDP per Worker of the U.S. The vertical axis represents the change in GDP per worker after a VAT reform of the reforming country compared to change in GDP of its synthetic control. We can observe heterogeneous treatment effects of the VAT reform. Importantly, the average effects are increasing with the initial level of income, with almost none of the low-income countries having a positive effect, about half of lower-middle-income countries having positive effects and most of the upper-middle-income and high-income countries having positive effect. To further explore this pattern and to provide the significance of the estimates, I calculate dynamic treatment effects and their p-values next.
Figure 4 presents the dynamic treatment effects for each of the income group. I find that VAT adoption has positive and significant impact on high-income and upper-middle-income countries, but mixed and insignificant impacts on lower-middle and low-income countries. For high-income countries, 5 years after the reform GDP per worker was 10.9 percent higher than the synthetic group, and it increased to 11.2 percent higher 10 years after the reform, all significant at the 1 percent level. Similarly, I find that upper-middle-income countries have positive and significant impact on GDP per worker, although after a time lag. The estimates are significant at the 10 percent level from year 3 onwards and at 5 percent level from year 4 onwards. Five years after the reform GDP per worker was 25.5 percent higher than the synthetic group, and it increased to 33.1 percent higher 10 years after the reform. For the lower-middle-income group, I find average effects ranging from 6.7 percent to 9.3 percent higher than the synthetic control, but indistinguishable from zero at the 10 percent level. In the case of low-income countries, I find negative, but statistically insignificant, impact of VAT adoption.

6 Exploring Potential Mechanisms

In this section, I explore two mechanisms through which a VAT affects economic performance: capital stock per worker and total factor productivity. I use synthetic control methods to estimate the causal effect of the VAT reforms on these variables following the same procedure used when estimating the effect on GDP per worker. To make the estimates as comparable as possible, I only run the analysis on the countries that were included in the main results.\footnote{Following the selection criteria of the main results, some country cases were not included if the pre-treatment fit was worse than the threshold of 0.10 or if the donor pool was empty. In some cases data on capital stock or total factor productivity was not available. Thus, I had to exclude 2 countries from the high-income group, 1 from the upper-middle-income group, 3 from the low-income group, and 1 from the low-income group from the capital stock analysis. Similarly, I had to exclude 1 upper-middle-income country, 1 lower-middle-income country and all 5 low-income countries from the total factor productivity analysis.}

The covariates used in the estimation of capital stock per worker and total factor productivity are growth rate of real GDP per capita, average stock of human capital, openness, inflation, and democracy dummy.

6.1 Capital Stock

The results for country-specific effects of the VAT reforms on capital stock per worker are presented in Figure 5. Average treatment effects are at the vertical axis and GDP per capita normalized by the GDP per capita of U.S are on the horizontal axis. The impacts on capital stock are heterogeneous, with none of the low-income countries having a positive
impact, but most of the middle-income and high-income countries having a positive impact. The results for dynamic treatment effects are presented in Figure 6. For the high-income countries, I find positive and significant impact of the reform at the 5 percent level for all years except year 5, 6, and 7 where the estimates are significant at the 10 percent level. Five years after the reform, capital stock per worker was 5 percent higher in the treated group compared to the synthetic group, which increased to 7.1 percent by the end of the sample. However, the average effects are positive, but not significant for the middle-income countries and significantly negative for the low-income countries. By the end of the sample, low-income countries with VAT had 10 percent lower capital stock per worker compared to the synthetic control. To summarize, I find the same differential effect across income groups on capital per stock worker as in GDP per worker.

6.2 Total Factor Productivity

Figure 7 graphs country-specific effects of VAT reforms on total factor productivity at the vertical axis and GDP per capita normalized by the GDP per capita of U.S on the horizontal axis. The impact on total factor productivity is heterogeneous, but mostly positive and it is increasing with the initial level of income. Figure 8 presents the results for the average impact across income-groups. Due to data limitation I could not analyze the impact on low-income countries. For the rest of the three income groups, I find that VAT increased total factor productivity. All dynamic treatment effects for the high-income countries are positive and significant at the 1 percent level. Five years after the reform, total factor productivity was 10.1 percent higher than the synthetic group, which increased to 11.8 percent by the end of the sample. The estimates for upper-middle-income countries are also positive and significant at the 10 percent or better level, but after a time lag. Note that the trend in TFP of the treated group started declining a few years before the reform, but the trend started improving immediately after the reform and by year 4 treated group had higher productivity that the synthetic group. For the lower-middle-income group, we notice an immediate jump in the productivity after the reform which starts declining after a few years, but always remain higher than the control group. Five years after the reform, TFP of the treated group was 9.9 percent higher compared to the synthetic group and at its highest, the TFP of treated group was 11.6 percent higher than the synthetic group. Moreover, 7 of the dynamic effects are significant at the 5 percent level and the remaining 3 dynamic effects are significant at the 10 percent level.

13 Only Kenya had data on TFP and Kenya’s TFP continued its decreasing trend after the VAT reform.
7 Robustness Tests

I conduct a series of robustness checks. First, I test whether the average effect is driven by the inclusion of any one treated country. Second, I test the impact of the VAT using more disaggregated data. Third, I test whether the impact of the VAT reform is commingled with other contemporaneous events like EU membership, WTO membership, or economic liberalization. Fourth, I estimate the impact of the VAT by restricting the donor pool to the same geographic region or the same income group. Fifth, I estimate the impact of the VAT adoption using conventional DID method to test whether my results are dependent on the use of synthetic control methods. Sixth, instead of using the World Bank’s income classification system, I create my own system by dividing the countries into 4 quartiles each year, according to their GDP per Worker. Seventh, to test if my results are sensitive to the use of cross-validation technique, I use the entire pre-treatment data to find the synthetic controls. Seventh, I use various alternative measures of economic efficiency. Eighth, and finally, I use tax as a share of GDP to control for overall tax burden while estimating the SCM.

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

7.1 Leave-One-Out Averages

I develop a graphical test to inspect whether the average effect is driven by any one treated country or not by drawing the leave-one-out averages of the difference in the outcome variable between the treated and synthetic groups. The leave-one-out average is constructed by iteratively removing one treated country and its synthetic counterpart and taking the average of the remaining \( N - 1 \) countries. If the average effect is not driven by any particular country then the leave-one-out averages should track the overall average very closely. Figure 9 presents the result for GDP per worker. We notice that the leave-one-out averages track the average result pretty closely across all four income groups. Thus, I can conclude that our main results are not driven by the inclusion of any treated country in particular. Similarly, Figure 10 and Figure 11 present the results for capital stock and total factor productivity respectively. I again do not find that any one country is influencing the results. Although in a few cases removing one country from the average changes the size of the effect, it never changes the sign. Thus, I conclude that a few large outliers are not driving the central finding.
7.2 Manufacturing Industry

There are two concerns with using aggregate data to analyze the effect of the VAT reforms that the use of manufacturing sector level data can help address. First, if the VAT reforms are motivated by expectation of future growth prospects, this would bias the estimates obtained from SCM as long as growth expectations are not captured by the unobservable heterogeneity included in the estimation. Thus, I cannot rule out the possibility of the bias caused by reverse causality. The use of manufacturing-level data helps mitigate the concern of reverse causality because VAT adoptions take place at the national level and thus are not likely to be affected by the underlying trends in one of the numerous sectors in any significant way.

Second, how a VAT affects different sectors of the economy is different across countries. In some countries, services are taxed broadly while in others they are only taxed selectively. In most cases, the VAT is extended through retail stage, but in others it is only extended through manufacturing stage. Similarly, sectors such as agriculture, financial sector, education sector, and health services sector are all treated differently in each country. Thus, the aggregate impact of the VAT masks this heterogeneity. However, the treatment of manufacturing sector is quite homogenous across countries. For instance, in almost all countries, the VAT was introduced to replace sales and turnover taxes that directly affected manufacturing industry and in all cases the VAT includes the manufacturing industry in the tax base. Thus, by focusing exclusively on the manufacturing industry we are better likely to isolate the impact of a more uniform VAT.

I use proprietary data on industrial statistics (INDSTAT2) from United Nations Industrial Development Organization (UNIDO), which provide industrial statistics for a wide range of countries at manufacturing level going back to early 1960s. It covers value-added, employment, and wage, among other things, allowing me to calculate labor productivity (value-added per employee) and average wage (wage per employee). Average wage provides an additional robustness test on labor productivity because if labor is more productive then it increases the marginal product of labor and as a result, the market-clearing wage. The data are reported in current U.S. dollars. I follow the approach suggested by Levchenko, Rancire and Thoenig (2009) and convert them into constant international dollars using price deflators from the PWT 8.1. To my knowledge, this is the only source of data that goes back to 1960s, allowing me to analyze countries from Western Europe and Latin America that adopted VAT in the 1960s and 1970s. It is also the only source of data for middle-income and low-income countries. However, only a few middle-income and low-income countries have long time series, but to make the results comparable to the main results, we need 21 years of non-missing data. Accordingly, the empirical analysis is based on 8 high-income coun-
tries (Austria, Belgium, Denmark, France, Ireland, Italy, Netherlands, Sweden, and United Kingdom) for labor productivity and 6 high-income countries (Austria, Belgium, Denmark, Italy, Netherlands, and United Kingdom) for average wage that meet the data requirements as well as the requirements of the SCM.

The results for labor productivity, average wage, and their leave-one-out averages are presented in Figure 12. Both labor productivity and average wage increased more in the reforming countries than their respective synthetic groups after the VAT reform. The dynamic treatment effects range from 2.7 percent higher 1 year after the reform to 10.3 percent higher 4 years after the reform, after which it starts decreasing and becomes statistically insignificant at the 10 percent level, but remains positive throughout the sample period. The leave-one-out averages suggests that the results are not driven by the inclusion of any particular treated country in the group. The results are slightly stronger and more precisely estimated for average wage. The estimates range from 3.7 percent higher to 13.6 percent higher in the treated group compared to the synthetic group. All estimates from year 2 onwards are significant at the 10 percent level with estimates from year 4 onwards being significant at the 5 percent or better level. Again, the leave-one-out test rules out the possibility that the results are driven by an outlier country.

7.3 Controlling for Contemporaneous Reforms

In some cases, the VAT was adopted as part of the trade liberalization or EU membership. Thus, the baseline estimates might confound their impact. To test whether the results are confounded, I calculate the average treatment effect by removing the countries that liberalized their economy or joined EU in the entire sample window (i.e., 10 years before or 10 years after VAT adoption). Furthermore, I use two of the most widely used measures of economic liberalization: WTO membership and a binary indicator by constructed by Sachs and Warner (1995) and updated by Wacziarg and Welch (2008).

The aggregate effects controlling for the EU membership are presented in the top row of Figure 13. Six countries joined the EU within 10 years of VAT adoption and they belonged to either high-income or upper-middle-income group. The results are unchanged by the exclusion of these countries. For high-income countries, the impact of the VAT controlling for EU membership ranges from 6.6 percent to 14.6 percent higher compared to the synthetic group, all significant at the 1 percent level. For upper-middle-income countries, the impact is positive and significant, but after two year time lag. The estimate from year 3 is significant at the 10 percent level and the estimates from year 4 onwards are significant at the 1 percent level.

\[\text{Denmark, Greece, Ireland, Portugal, Spain, and United Kingdom joined the EU within the sample period and thus were removed from the analysis.}\]
level. The statistically significant estimates range from 4.2 percent to 29.5 percent higher compared to the synthetic group.

The aggregate effects controlling for the WTO membership (14 countries in total) are presented in the bottom row of Figure 13 and the aggregate effects controlling for Wacziarg and Welch (2008) economic liberalization (13 countries in total) are presented in the top row of Figure 14.\textsuperscript{15} For the high-income countries, I again find the impact on GDP per worker is positive and statistically significant at the 1 percent level, although the size of the effects are slightly smaller (ranging from 1.9 percent to 7.4 percent). Similarly, upper-middle-income countries also have positive and significant impact, after a time lag. 7 out of 10 estimates are significant at the 10 percent level and 6 of those are significant at the 1 percent level. Five years after the reform, GDP per worker was 22.4 percent higher, which increased to 37.1 percent by the end of the sample period. The lower-middle-income countries experienced economically meaningful, but statistically insignificant improvement and since all four low-income countries joined the WTO within 10 years of VAT adoption they could not be analyzed.

The results controlling for Wacziarg and Welch (2008) economic liberalization are presented in the top two rows of Figure 14. I again find that VAT adoption has positive and significant impact on high-income countries, positive and significant impact after a time lag on upper-middle-income countries, slightly positive, but insignificant impact on lower-middle-income countries, and mixed, but statistically insignificant impact on low-income countries. Moreover, this trend holds even when I control for EU membership, WTO membership, and economic liberalization at the same time as illustrated in the bottom row of Figure 14.

To conclude, in every specification, I find an economically and statistically significant increase in economic efficiency on high-income countries and upper-middle-income countries, but mixed and insignificant impact on lower-income and low-income countries. Thus, I conclude that the results are driven by the VAT adoption itself and not by other contemporaneous reforms.

\section*{7.4 Restricted Donor Pools}

In the main analysis, I run the analysis using a worldwide sample of countries. The main advantages of doing so are that it includes more countries in the donor pool so the probability

\textsuperscript{15}The WTO members were Bangladesh, Canada, Greece, Guinea, Jamaica, Japan, Kenya, Mauritius, Nepal, New Zealand, Pakistan, Portugal, Spain, and Thailand. The countries that underwent economic liberalization were Bangladesh, Chile, Denmark, the Dominican Republic, France, Guinea, Ireland, Jamaica, Kenya, Nepal, Netherlands, New Zealand, and Sweden.
of finding a synthetic country that matches the treated country increases and it also increases
the size and the power of the test. However, one disadvantage of using a worldwide sample
of countries is that I cannot rule out the possibility of some far-fetched country comparisons,
although this possibility is minimized in practice by including region dummies and income
group dummies in the estimation and by requiring the synthetic controls to match the initial
level of the outcome variable and its important predictors.

I use two alternative control groups where I restrict potential controls to the same income
group or the same geographic region to ensure the existence of a common support between
treated and control countries. This approach has two main advantages. First, it avoids biases
caused by interpolating across countries with very different characteristics. That is, even if
we are able to find a synthetic unit with good pre-treatment fit, interpolation biases may
be large if the linear factor model used in the estimation of the synthetic control does not
hold over the entire set of regions in any particular sample. This happens if the relationship
between the outcome variables and the predictors is highly non-linear and the combination
of two extreme donor units is used to construct a synthetic unit that has the average value.
In the similar spirit, it also controls for unobservable characteristics associated with the level
of economic development or geography and any other secular changes over time that might
affect countries from different income groups or different geographic regions differently.

The results restricting the control group to be from the same geographic region are pre-
presented in Figure 15. I was able to analyze only 8 countries from the high-income group,
and 1 each from the rest of the income groups because of the restrictions imposed to the
donor pool. Even so, I find the same differential pattern that I find in the main results.
For high-income countries, the impact of VAT adoption is positive, ranging from 2.3 per-
cent to 8.8 percent higher compared to the synthetic group and they become statistically
significant at the 10 percent or better level from year 4 onwards. Although much cannot be
said about other income groups because of small number of countries included, I find that
upper-middle-income country have positive impact, lower-middle-income country has mostly
negative impact and low-income country has a negative impact of VAT adoption.

The results restricting the control group to be from the same income group are pre-
presented in Figure 16. I was able to analyze only 8 countries from the high-income group, 2
from upper-middle-income group, 3 from lower-middle-income group, and 1 from low-income
group because of the restrictions imposed to the donor pool. Even so, I find the same differen-
tial pattern that I find in the main results. For high-income countries, I find an immediate
positive and statistically significant impact at the 5 percent or better level. The estimates
range from 2.4 percent to 8.7 percent higher compared to the synthetic group. Although
much cannot be said about other income groups because of the small number of countries in-
cluded, I find that upper-middle-income group has positive impact after some time lag, while lower-middle-income and low-income group have mostly negative impacts of VAT adoption.

7.5 Conventional Difference-in-Differences

I also estimate the impact of VAT reforms using the conventional difference-in-differences method to test whether the estimates are dependent on the use of SCM rather than DID methods. I follow the event study approach by adding leads and lags of the treatment. Doing so allows me to test whether the parallel trends assumption is violated or not and to assess the dynamic nature of the impact on economic efficiency. If coefficients on the leads are significantly different from zero, then this might indicate the failure of the DID approach to create a comparison group with counterfactual trends parallel to the treatment group. In such a case, use of DID produces biased results. The coefficients on lags describe the transition, capturing the average effect of tax reform in years following adoption relative to the effect before the adoption.

The results using conventional DID are presented in Figure 17. The Figure in left panel includes country fixed effects, and year fixed effects while the figure in right panel also includes country-specific linear time trends. The figure in the top row is obtained by running regression on the sample of 33 treated countries used in the baseline SCM estimates, thus the effect of VAT comes from the variation in the timing of VAT adoption. The figure in the bottom row adds 24 countries that contributed to the construction of synthetic controls in the SCM results. Since 23 out of 24 control countries eventually adopted a VAT it implies that I am still exploiting the variation in the timing of VAT adoption to find its effects. In all specifications, the standard errors are clustered at the country level. In the left panel most of the leads are significant, indicating that just using country and year fixed effects violates the parallel trends assumptions. I can control for the differential trends by using country-specific linear trends in the estimation. The results controlling for differential country trends are presented in the right panel. The results in the right panel indicate that VAT adoption has positive and significant impact on economic efficiency. These effects increase with time for first few years and then start decreasing. For instance, GDP per worker was $407 higher in the treated group compared to the control group in the first year after the reform, which increased up to $840 five years after the reform. The estimates from year 2 onwards are significant at the 5 percent level, but they become insignificant from year 7 onwards.\textsuperscript{16} The results for 57 countries are presented in the bottom row. The differential trends between treated group and control group are more prominent than the estimation with just 33 treated countries.

\textsuperscript{16}I also run the analysis that includes all the covariates used in the main SCM estimation, the results are similar.
countries; however, results controlling for country-specific linear trends are very similar.

7.6 Additional Robustness Tests

I also conduct several additional robustness tests in the Appendix. First, I construct the income classification system using a data-driven approach. That is, I divide the countries into 4 quartiles each year, according to their GDP per Worker, and use the first quartile as low-income, second quartile as lower-middle-income, third quartile as upper-middle-income, and fourth quartile as high-income group. Next, I use this classification system to run the SCM and to aggregate the effects into different groups. I obtain very similar results.

Second, in the main analysis, I use a cross-validation technique to minimize the risk of over-fitting by dividing the pre-treatment data into training period and validation period. I test whether my results are sensitive to the use of a cross-validation technique by following a more traditional approach and using the entire pre-treatment data to find the synthetic controls. The results remain unchanged.

Third, I use four alternative measures of economic efficiency to test whether my main results are dependent to the use of output-side GDP per worker from PWT 8.1 as a measure of economic efficiency. The alternative measures are output-side GDP per capita from PWT 8.1, expenditure-side GDP per worker from PWT 8.1, output-side GDP per worker from PTW 8.0, and output-side GDP per capita from PTW 8.0. I find that the results using GDP per worker are slightly smaller in magnitude and they become significant after a year compared to GDP per worker. Other than that the results are largely unchanged.

Fourth, I control for the initial level of tax burden by adding tax as a share of GDP as one of the control variables while estimating the impact of VAT adoption on GDP per worker. The results for high-income, lower-middle-income and low-income remains very similar. However, the estimates for upper-middle-income countries become statistically insignificant (although the point estimates remain positive and 6 out of 10 estimates have p-values less than 0.2).

To conclude, I find the baseline estimates to be largely robust, and they are discussed in detail in the Appendix.

8 Conclusions

This paper analyzes the impact of replacing sales and turnover taxes with VAT on economic efficiency by combining information from comparative case studies obtained with a synthetic control methodology recently developed in Abadie and Gardeazabal (2003) and expanded
in Abadie, Diamond and Hainmueller (2010, 2015). The procedure involves identifying the causal effects by comparing the trajectory of post-treatment GDP per worker with a carefully constructed counterfactual trajectory of GDP per worker. I find that a VAT adoption has an economically meaningful and statistically significant impact on the trajectory of real GDP per worker. However, the positive impact is conditional on the level of development of the country in question. That is, there is a strong correlation between the positive impact of a VAT adoption and the initial level of development - with high-income countries benefiting the most from the reform and low-income countries benefiting the least. Next, I explore two potential mechanisms (capital accumulation and total factor productivity) through which a VAT can affect production efficiency and find that both the channels are important. I again find the same pattern of differential effect of a VAT adoption along the initial income level for these channels.

This paper has two main policy implications. First, these results indicate that the theoretical advantages of VAT do not necessarily translate into practice. In particular, the impact of VAT on economic efficiency depends on the level of development of the country. The level of the development is highly correlated with factors such as tax capacity, tax evasion, and informal economy, which can severely undermine the effectiveness of a VAT. Second, a VAT is often considered a “silver bullet” that can both replace the distortionary taxes and raise much needed revenue for public spending, especially in the developing countries. My study suggests that the results are more nuanced and it highlights the need to modernize tax administration along with VAT adoption, if not before, to benefit from the efficiency properties of a VAT. This policy implication is not new. Voluminous literature on taxation and development have argued that the real tax system is not that which is passed as legislation, but that which is administered (Bird and Gendron, 2011; Carrillo, Pomeranz and Singhal, 2014; Casanegra de Jantscher, 1990; Gordon and Li, 2009). Thus, it is critical to ensure that the implementation and administration of a VAT receive as much attention as the adoption of a VAT does, especially since the results indicate sizable gains in economic efficiency from adopting a well-designed and well-implemented VAT.
References


Figure 1: The Spread of the Value Added Tax

Notes: The map shows the countries that adopted Value Added Tax (VAT) between 1960-2013.
Figure 2: The Overall Impact of VAT Adoption on Average GDP Per Worker

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. N denotes number of countries analyzed from each income group. The left axis consists of labels for the outcome variable. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 3: The Impact of VAT Adoption on Average GDP Per Worker Across Countries with Different Initial Income

Notes: Country-specific average treatment effects are obtained using synthetic control methods by subtracting average value of the pre-intervention difference in the outcome variable from the average value of the post-intervention difference in the outcome variable for each country i and its synthetic control. The y-axis consists of labels for the outcome variable, which is expressed in 2005 constant international dollar and is adjusted for purchasing power parity. The x-axis consists of the income of treated country at treatment year normalized by the income of U.S at that year.
Figure 4: The Impact of VAT Adoption on Average GDP Per Worker Across Different Income Groups

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 5: The Impact of VAT Adoption on Average Capital Stock Per Worker Across Countries with Different Initial Income

Notes: Country-specific average treatment effects are obtained using synthetic control methods by subtracting average value of the pre-intervention difference in the outcome variable from the average value of the post-intervention difference in the outcome variable for each country $i$ and its synthetic control. The y-axis consists of labels for the outcome variable, which is expressed in 2005 constant international dollar and is adjusted for purchasing power parity. The x-axis consists of the income of treated country at treatment year normalized by the income of U.S at that year.
Figure 6: The Impact of VAT Adoption on Average Capital Stock Per Worker Across Different Income Groups

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 7: The Impact of VAT Adoption on Average Total Factor Productivity Across Countries with Different Initial Income

Notes: Country-specific average treatment effects are obtained using synthetic control methods by subtracting average value of the pre-intervention difference in the outcome variable from the average value of the post-intervention difference in the outcome variable for each country $i$ and its synthetic control. The y-axis consists of labels for the outcome variable, which is normalized using TFP of the United States. The x-axis consists of the income of treated country at treatment year normalized by the income of U.S at that year.
Figure 8: The Impact of VAT Adoption on Average Total Factor Productivity Across Different Income Groups

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 9: Average Effect versus Leave-One-Out Average Effects of VAT Adoption on GDP Per Worker Across Different Income Groups

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. The solid black line denotes the difference in the outcome of the group of treated countries and their synthetic controls, and the dashed black line denotes the difference in the outcome of N-1 treated countries and their synthetic controls. The left axis consists of labels for the difference in the normalized outcome variable.
Figure 10: Average Effect versus Leave-One-Out Average Effects of VAT Adoption on Capital Stock Per Worker Across Different Income Groups

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. The solid black line denotes the difference in the outcome of the group of treated countries and their synthetic controls, and the dashed black line denotes the difference in the outcome of N-1 treated countries and their synthetic controls. The left axis consists of labels for the difference in the normalized outcome variable.
Figure 11: Average Effect versus Leave-One-Out Average Effects of VAT Adoption on Total Factor Productivity Across Different Income Groups

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. The solid black line denotes the difference in the outcome of the group of treated countries and their synthetic controls, and the dashed black line denotes the difference in the outcome of N-1 treated countries and their synthetic controls. The left axis consists of labels for the difference in the normalized outcome variable.
Figure 12: Average Effect and Leave-One-Out Average Effects of VAT Adoption on Manufacturing Industry Value Added Per Worker and Wage per Worker Across Different Income Groups

Notes: The results are for high-income countries (H). The outcome variable for the first row is value adder per worker for manufacturing industry, the outcome variable for second row is for wage per worker in manufacturing industry. The first column present the trends in the outcome variables for the treated group and its synthetic control. The second column presents leave-one-out average where N denotes number of countries analyzed. The solid black line denotes the difference in the outcome of the group of treated countries and their synthetic controls, and the dashed black line denotes the difference in the outcome of N-1 treated countries and their synthetic controls.
Figure 13: The Impact of VAT Adoption on Average GDP Per Worker Controlling for European Union and World Trade Organization Memberships

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 14: The Impact of VAT Adoption on Average GDP Per Worker Controlling for Economic Liberalization and All Three Contemporaneous Reforms

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 15: The Impact of VAT Adoption on Average GDP Per Worker Across Different Income Groups using Control Countries from the Same Region Only

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 16: The Impact of VAT Adoption on Average GDP Per Worker Across Different Income Groups using Control Countries from the Same Income Group Only

Notes: The four income groups are H (high-income), UM (upper-middle-income), LM (lower-middle-income), and L (low-income). N denotes number of countries analyzed from each income group. 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 the area below 0.10 is shaded in gray and a horizontal line is drawn to indicate 0.05. To avoid the size effect caused due to varying levels of outcome variable 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.
Figure 17: The Impact of VAT Adoption on Average GDP Per Worker Using Conventional Difference-in-Differences Approach

Notes: In the top row 33 treated countries used in SCM are included, while in the bottom row 24 countries that contributed to the construction of the synthetic controls are also included. All specification includes country and year fixed effects and the standard errors are clustered at the country level. The figures in the right panel also include country-specific linear time trends. The connected line plots the estimates for leads to the left of the vertical line and the estimates for lags to the right. The variable Year = -1 is omitted.