Time Series in Tax Performance[*]

Anders Jensen[†]

May 2022

Abstract

The literature on tax capacity has shed light on the descriptive, long-run trends
in tax performance and the micro-designs which incrementally impact tax
collection of individual firms and households. There is a 'missing middle'
of evidence on tax performance at the intermediary level: in the aggregate,
over medium-term horizons. This paper uses methods from the literature
on economic growth to investigate the time-series of tax collection at this
intermediary level. In a global sample of 156 countries between 1965 and
2020, and consistently across methods, we find numerous breaking points in
tax-to-GDP trends that are sustained. Countries' time-series of tax-GDP feature
both upturns and downturns that are economically large in magnitude.

1 Introduction

The literature on tax capacity seeks to understand the determinants which impact a state's ability to collect taxes in an equitable, efficient and progressive manner. Based on an important initial theoretical contribution by Besley and Persson (2014), two research agendas have emerged on tax capacity. The first literature seeks to establish the patterns of tax capacity across the development path – considering the full range of development across countries today and the long-run evolution within countries over time. This literature focuses on understanding which are the approximate, structural determinants of tax capacity, including in labor markets (Jensen (2022)), financial markets (Gordon and Li (2009)) and cross-border integration (Bachas et al. (2022)). The second literature takes a micro-economic approach and leverages administrative data; the papers in this literature evaluate individual tax reforms, and shed light on the tax policy designs which can incrementally improve the efficiency and equity of tax collection. Recent papers have focused on the role of third-party information coverage (Pomeranz (2011); Naritomi (2019)) and enforcement constraints (Best et al. (2015)). The evidence built across these two research strands thus sheds light on either the descriptive, long-run trends in tax capacity or the micro-designs which incrementally impact tax collection amongst individual firms or households.

There is, however, a missing gap of evidence, which lies between the long-run macro analysis and the within-country micro-analysis. Indeed, the literature on tax capacity effectively focuses on the (very) long-run, often comparing tax/GDP levels that differ by more than 40 years or more in time. The implicit assumption is that the long-run change is the outcome of steady, gradual yearly growth – which in turn can be related to the long-run determinants proposed in the theory, which are thought to have steady, gradual impacts on tax systems. In contrast, as our analysis reveals, most countries' tax performances are unstable over the (very) long-run, and are instead characterized by phases of strong growth, stagnation, and decline. Arguably, this is the level at which policy-design may aspire to have the strongest effects. Policy-makers are likely to be strongly pre-occupied by questions such as: how likely is it that the tax system undergoes an acceleration for a sustained period of time? Is it possible to limit the tax system's exposure to deceleration? Policy-makers may correctly perceive that, even if long-run analyses do ultimately predict the

deeper drivers of capacity over sufficiently long periods, these determinants may not be the most relevant for turning points in tax collection.

In this paper, we explore the relevance of several distinct methods to detect breaking points in individual countries' time-series of tax-GDP. The methods, called the statistical method and the filter method, are flexible and allow us to identify both accelerations and deccelerations in the time-series of the tax-to-GDP ratio of individual countries. The methods build on recent developments in the macro-growth literature, where break estimation procedures have been implemented to identify turning points in the GDP growth.

Our data comes from the publicly available data-set on tax revenues which was first developed by Bachas et al. (2022)[1]. For the purposes of the current project, this data-set has two advantageous features. First, it is the most comprehensive data-set on tax revenues, covering the 150 most populous countries from 1965 to 2018, with exceptions only for pre-independence, civil war and command economy areas. The long time-series within country improves the statistical power of our methodology; the comprehensive coverage across countries allows us to draw broad, globally representative conclusions about trends in tax performance. Second, the database is that it includes all taxes – personal income taxes, corporate income taxes, Social Security payroll taxes, property taxes, wealth taxes, estate and inheritance taxes, consumption and other indirect taxes – at all levels of government.

Our analysis uncovers the richness of the time-series variation that exists within-countries in the global sample. Using the statistical method, there are 464 break points in countries' time series; repeated with the filter method, there are an estimated 284 break points. In the main text, we discuss the reasons which could account for the difference in results. We can convert the frequency of turning points into unconditional probabilities – the likelihood that a country in a given years experiences a turning point is 10.15 percent (8.85 percent) according to the statistical (filter) method. To gauge this magnitude, the estimated likelihood that a turning point will occur in GDP growth is 2.8% (Hausmann, Pritchett, and Rodrik (2005)). Turning points can either lead to upturns in the form of sustained increases in the tax-GDP ratio, which we call accelerations, or downturns in the form of sustained decreases (deccelerations). When we study the likelihood of turning points separately by type, we find that the probability of an acceleration

---

[1]The database is publicly available online at https://globaltaxation.world

(6.12 percent using the statistical method) is only slightly higher than the probability of a decceleration (4.97 percent). In other words, far from being summarized by a steady and small positive long-run growth rate, countries' tax-GDP performance is characterized by frequent upturns and downturns which occur with a comparable likelihood.

Apart from the sheer number and frequency of episodes, the magnitude of the typical turning point is also impressive. The average gain in tax/GDP in the case of an acceleration is 5 percentage points using the statistical method. Relative to the average tax/GDP ratio in the full sample of country-years (22.06%), this represents a 22.6% increase. In other words, in the typical episode, tax/GDP stood 22.06% higher in the post-period than it would have without any acceleration. The average loss in tax/GDP in the case of a deceleration episode is 4.8 percentage points – representing a staggering 21.7% loss in tax collection (relative to the average tax/GDP ratio). The magnitudes uncovered with the filter approach are slightly larger. In summary, due both to their frequency and to the magnitude of the induced change in taxation performance, turning points in tax/GDP are meaningful macro-economic events.

The richness of the data-set allows us to study the contributions of sub-components of total taxation to growth during turning points. We focus on the tax-GDP ratios of labor taxes, capital taxes and consumption taxes, following the classification of Bachas et al. (2022). We define a broad-based turning point as an event during which all three sub-components have growth that is the same sign as overall taxation. Despite the large changes in overall taxation induced during a breaking point, only 57% of accelerations and 43% of deccelerations are broad-based.

We finish our analysis by investigating heterogeneity patterns in turning across development. Are break points more prevalent at certain levels of development? We provide a conceptual discussion, which highlights that the relationship between a country's development and the likelihood and magnitude of turning points is ambiguous. Results using the statistical method suggest that the likelihood of accelerations and deccelerations are very comparable across low-income, middle-income and high-income groups. The filter method suggests that the likelihood of accelerations (deccelerations) is slightly higher (smaller) in higher-income countries. However, both methods reveal that the magnitude of changes in tax-GDP are similar (in percentage point terms) across development levels.

The rest of the paper is organized as follows. In Section 2, we relate our findings to both the literature on tax capacity and the literature on turning points in economic growth. In Section 3, we define our methodology and data. In Section 4, we present our results on estimated turning points in the full sample, while in Section 5 we investigate heterogeneity across development. Section 6 concludes.

2 Literature Review

2.1 Trends in Tax Capacity

There is a growing literature on the long-run determinants of tax capacity. Most papers in this literature use the ratio of taxes to GDP as the main proxy for tax capacity. This literature is motivated by the observation that tax/GDP increases over the development path: this fact holds both within countries, based on the time-series of currently developed countries over the long-run, and across countries, based on the cross-section of countries today which differ in GDP per capita. In the case of within-country trends, the time-frame studied is often very long-run, spanning over 100 years of data. Beyond the increase in tax/GDP, the literature on tax capacity has also established facts regarding the composition of taxes. In particular, Besley and Persson (2014) show that, over the development path, the share of income taxes in total taxes increases and the share of trade taxes decreases. Relatedly, Kleven, Kreiner, and Saez (2016) show that the long-run growth of government in 14 advanced countries is entirely explained by what the authors call modern taxes (primarily income taxes and value-added tax). Remarkably, the share of traditional taxes in GDP (excise and trade taxes) has remained constant or has mildly declined.

Several theories have been proposed to explain these long-run trends. First, the famous Wagner's law focuses on the demand for tax revenues and posits that public goods have an income elasticity above 1. Second, the 'bellicist' theory of state building argues that politicians and rulers seek to extract resources from citizens to sustain the funding of protection against external forces. Per example, historical studies have argued that the second world war, by creating a large increase in the demand for military spending, ushered in the modern income tax in the United States and Europe. Third, the 'Leviathan' theory posits that governments

are controlled by self-interested bureaucrats that seek to maximize the revenue that can be extracted, subject to electoral and fiscal constraints.

More recent studies have focused on the role of information. Gordon and Li (2009) assume that firms decide whether to make use of the financial sector or not. When firms interact with financial institutions, the government gains access to the firms' bank records and can use this information to enforce tax laws. Under the assumption that firms derive smaller benefits from access finance in less-developed countries, this leads to a larger share of transactions that are being conducted in cash – that is, outside the scope of enforcement. Related, Kleven, Kreiner, and Saez (2016) develop an agency model, in which firms act as third party reporters on individuals' income. They show that the long-run growth in firm size and complexity leads to the creation of information trails on firms' activities and curbs the ability to mis-report information to the tax authority. Jensen (2022) focuses on individual income, and shows how, over the long-run of development, the growth in enforceable income (proxied by the employee-share of the workforce) occurs gradually further down in the country's income distribution. This allows countries to expand the income tax base by lowering the income tax exemption threshold and gradually collect more taxes as a share of GDP.

A striking finding from these studies is that developing countries are arguably not so different – in terms of tax systems, and the approximate determinants of tax capacity – from modern high-income countries at a similar stage of development (Besley and Persson (2014); Jensen (2022)). From this perspective, the recent literature on tax capacity suggests that the current low levels of tax/GDP in developing countries today reflect factors that also help explain why these countries are less developed. The challenge from this perspective is it implies that more important policies can be taken to encourage development more broadly, rather than special policy measures which focus on improving tax systems.

In parallel to these macro-economic, long-run studies, a series of recent papers have provided rigorous, micro-evidence on specific policies which alleviate constraints on tax capacity. Using administrative data and experimental variation, Pomeranz (2011) shows that the existence of information trails deters mis-reporting of taxable activities under a VAT. Naritomi (2019) shows that financial incentives provided to consumers can curb mis-reporting at the final stage of the production chain. Kleven, Knudsen, et al. (2011) study randomized audits in Denmark, and

find that evasion at the individual level is strongly related to the the extent of third-party coverage. Best et al. (2015) show that implementing turnover taxes can improve enforcement on small firms and ultimately increase tax revenue collection, even if they create distortions (relative to profit taxes). Bergeron, Tourek, and Weigel (2021) show that increasing enforcement capacity allows local governments to simultaneously levy higher tax rates.

In summary, the literature on tax capacity focuses either on the (very) long-run aggregate tax performance, or on the (very) short-run tax collection from individual firms and households. In this sense, there is a 'missing middle' of evidence on tax performance at the intermediary level that policy-makers arguably care about the most: in the aggregate, over medium-term horizons. At this level of observation, the relevant object becomes to identify turning points in tax performance, which is our focus. The use of empirical methods to uncover turning points in tax capacity is very limited in the literature. Akitoby et al. (2018) use a data-set covering 136 countries between 2000 and 2015; they consider episodes defined as 3-year windows of average increase in tax/GDP by 0.5 percentage points per year. Oppel and Chachu (2022) expand the results in Akitoby et al. (2018), using data on 196 countries between 1985 and 2019. They consider 5-year windows, and define the intensity of tax/GDP episodes on the basis of how large the (absolute value of the) change is in percentage point terms. In turn, they create a cumulative score for each country across all periods.[2] Borrowing terminology from the growth diagnostics literature (summarized below), both of these studies use 'filter approaches' – which use subjectively defined filter-rules to measure trends. To the best of our knowledge, Dincecco (2009) is the only study which uses statistical approaches to identify breaking points in tax collection, but the author focuses on historical time-series in European countries between 1650 and 1913. Drawing on developments in the growth literature, our contribution is to systematically investigate the results using both the filter method and the statistical method. Importantly, we develop this framework while drawing on the largest tax revenue database to-date, covering 150 countries since 1965. The database is available online at https://globaltaxation.world.

---

[2]Gaspar, Jaramillo, and Wingender (2016) use a historical unbalanced panel of tax-GDP ratios for 139 countries from 1965 to 2011. They show the distribution of tax/GDP ratios in the full sample, but then go on to focus on whether there are tipping points beyond which tax/GDP matter differentially for long-run growth.

2.2 Methods to Measure Trends in Economic Growth

Our data-framework draws on the recent developments in the empirical literature on economic growth. Pritchett (2000) noted that economic growth is often characterized by frequent 'regime-shifts'. The author showed that a single growth rate fitted over a long time period produced very poor statistical fits for the time-series of economic growth, particularly for developing countries.

Based on this observation, a subsequent literature focused on developing techniques to identify the timing of breaks in economic growth. Two distinct approaches have emerged. In the first method, called the 'filter approach', the researcher identifies growth breaks on the basis of subjectively defined rules. Using this method, Hausmann, Pritchett, and Rodrik (2005) studies breaks in growth accelerations. The second method is based on statistical tests, which uses estimation and testing procedures to identify structural breaks. These approaches are based on the Bai-Perron (BP) methodology (Bai and Perron (1998)) which develops a general framework to estimate multiple structural changes occurring at unknown dates. Jones and Olken (2008) and Kerekes (2011) are among the earliest studies to use the BP methodology to study breaks in economic growth. Berg, Ostry, and Zettelmeyer (2012) extend the methodology to be able to investigate both the timing and duration of growth spells.

Kar et al. (2013) discuss the shortcomings of each approach. In the filter approach, each study defines a set of subjective criteria. Per example, in Hausmann, Pritchett, and Rodrik (2005), growth accelerations are defined as (i) increases in per capita growth rates by 2 percentage points or more; (ii) sustained growth for at least 8 years; and, (iii) post-acceleration growth which has to be at least 3.5 percent per year. Aizenman and Spiegel (2010) focus on 'takeoffs', which they define as transition from stagnation (5-year periods with average per capita growth below 1%) to high growth (exceeding 3% over a minimum of 5 years). The lack of a common framework in the filter-approach is well illustrated by these two studies, which both focus on upticks in economic growth, but define such events in very different ways. Thus, filter-approaches suffer from the lack of a common framework which renders results less comparable across studies.

The statistical approach is, by contrast, based on a common technique which can be applied uniformly across settings – thus overcoming the main shortcoming of the filter approach. However, the statistical approach is limited by the low

7

statistical power of the BP framework. This issue is exacerbated in settings where the time-series of the outcome is particularly volatile, where the BP test may not be able to identify genuine breaks with statistical significance. This 'true negative' problem is established in simulation results in Bai and Perron (1998), and implies that the set of breaks identified using the BP method will be conservative (Jones and Olken (2008)). The issue, however, is that the subset of 'true' breaks captured with the BP method is unlikely to be representative of the full set; in particular, the events not captured by BP in countries with more volatile time-series may in fact be of strong interest to researchers and policy-makers. Finally, both approaches suffer from a common limitation - which is that neither method considers the dependency of any break on the history of previous breaks. Kar et al. (2013) propose a framework which draws on the strengths from each method, and which attempts to solve the dependency issue. Focusing on economic growth for 125 countries between 1950 and 2010, the method in Kar et al. (2013) uncovers a much larger number of possible genuine breaks (314) than the number uncovered (174) using the BP method in Berg, Ostry, and Zettelmeyer (2012).

Our two methods, described in the following section, draws on the methods developed in the economic growth literature. Other methods developed for time-series macro-economics are summarized in Hamilton (2016). First, we explore a filters based approach adapting the subjectively defined rules used in the growth literature by Hausmann, Pritchett, and Rodrik (2005). Next, we adapt the Bai-Perron (BP) procedure (Bai and Perron (1998)) by following the more recent extensions proposed by Berg, Ostry, and Zettelmeyer (2012) to test for multiple structural breaks when both the total number and the location of breaks are unknown. We use sample-specific critical values that take into account heteroskedasticity and sample size as opposed to asymptotic critical values. Following Antoshin and Souto (2008), we use a sequential testing algorithm of structural breaks which improve both the power and size properties of the test. We refer to this as the "statistical" approach.

3   Methodology

In this section, we discuss in turn the two methods that we employ to detect breaking points in countries' tax-GDP time-series.

3.1 Estimation of turning points

3.1.1 Filter approach

Following Hausmann, Pritchett, and Rodrik (2005), we create filters that are a combination of a large and a rapid change in tax to GDP ratio. Our tax acceleration episodes are defined for a time horizon of eight years. We employ the following filters to identify the meaningful breaks:

- absolute change in the 8-year moving average of the tax-to-GDP ratio before and after the break is 5 percentage points,

- the average annual growth rate in the tax-to-GDP ratio is larger than 5 percent during this 8-year period,

- post-acceleration tax-to-GDP ratio exceeds the pre-episode peak level of tax-to-GDP ratio.

The third criteria allows us to rule out cases of temporary dips in tax revenues or catch-up after a growth episode. For deceleration episodes, we use the same time horizon and (absolute) magnitude of changes.

3.1.2 Statistical approach

We apply a variant of a procedure proposed by Bai and Perron (1998) and adapted by Berg, Ostry, and Zettelmeyer (2012) for the estimation of breaks in the tax-to-GDP ratio. The timing of the initiation of the acceleration or deceleration period is chosen by finding the year that minimizes the sum of squared residuals of a regression based on estimated number of breaks. Since for some countries there are a number of consecutive years for which these criteria of a growth episode are met, the "best" starting date is chosen by looking for the best fit among all contiguous eligible dates.

Countries can have more than one instance of tax acceleration. The null hypothesis of no additional structural breaks is tested against the alternative of one more structural break. The location of additional breaks is decided by minimizing the sum of squared residuals between the actual data and the average tax to GDP ratio before and after the break. Critical values are generated through Monte Carlo

simulations using bootstrapped residuals that take into account the properties of the actual time series (that is, sample size and variance). Confidence intervals for each estimated break date are estimated using heteroskedastic and auto-correlated critical values following Berg, Ostry, and Zettelmeyer (2012).

3.2 Data

Our data comes from the publicly available data-set on tax revenues, factor shares and national income components, which can be retrieved at https://globaltaxation.world. The data-set was developed by Bachas et al. (2022) in a study on the impact of globalization on the effective taxation of capital and labor. For the purposes of the current project, this data-set has several advantageous features. First, it is, to the best of our knowledge, the most comprehensive data-set on tax revenues, covering the 150 most populous countries from 1965 to 2018, with exceptions only for pre-independence, civil war and command economy areas. The long time-series within country improves the statistical power of our methodology; the comprehensive coverage across countries allows us to draw broad, globally representative conclusions about trends in tax performance.

The second advantage of this database is that it includes all taxes – personal income taxes, corporate income taxes, Social Security payroll taxes, property taxes, wealth taxes, estate and inheritance taxes, consumption and other indirect taxes – at all levels of government. When available, OECD Revenue Statistics is the preferred data-source, because it covers and classifies all types of tax revenues, usually back to 1965 for OECD countries. OECD data accounts for 41% of the country-year observations in the sample. Its main drawback is its limited coverage of non-OECD countries: in total it covers 93 countries, and only over the past two decades. To increase coverage, the OECD data is augmented with tax revenue data from the ICTD/UNU-WIDER (17% of observations). This dataset achieves near worldwide coverage but only starts in the 1980s.To address these shortcomings, the data-set also draws on historical public finance data from government reports, primarily

from the Harvard Library archives (30% of country-year observations) and from the 2005 version of the IMF GFS (offline historical database; 10% of observations).[3]

Throughout this paper, our main focus is on overall tax collection, which excludes non-tax sources of revenue (including from natural resources). Following Bachas et al. (2022), we relate total taxes collected to net domestic product (as opposed to national income). This relates tax collection to the appropriate measure of the domestic output base; results using national income are very similar in practice (available upon request). At the end of Section 4, we also study the contribution of labor, capital and consumption taxes to turning points. This classification of taxes is based on Bachas et al. (2022).

3.3 Conceptual discussion

The two methods will capture breaks in trends of aggregate tax collection. While our aim in the current paper is to establish the validity of these methods, we here offer a discussion of the types of events which could a priori cause such breaks.

It is perhaps useful to start by clarifying which types of events are *not* likely to cause such breaks in trends. First, both methodologies intend to capture breaks which are large and sustained over a meaningful number of years. As such, yearly jumps or declines or cyclical trends (4-5 year changes) in tax collection are unlikely to be captured in the breaks. Second, if taxes collected were a constant % of total GDP and our variable of interest was the volume of taxes, then our methodology would mechanically pick up breaks in GDP growth. But since we express taxes relative to GDP, there is no immediately mechanical relationship between economic growth breaks and tax collection breaks.

At a high level, we can consider three categories of factors which may plausibly lead to a transition in tax collection regime. The first set is tax reforms, which can include changes to the statutory parameters (e.g. rates, thresholds), the enforcement environment or the tax authority. While individual statutory changes are perhaps unlikely to cause a transition in the overall tax/GDP ratio, broad statutory reform packages may have such an effect. Per example, the introduction of the VAT is commonly thought to have had first-order aggregate effects on tax collection

---

[3]We note that the ICTD/UNU-WIDER data-set is, in turn, largely drawn from the online version of the IMF Government Finance Statistics. The additional use of IMF data is restricted to the offline historical data-set, which covers 1972-89 and fills gaps from the OECD and historical archives data.

in developing countries. Moreover, statutory reforms are often combined with administrative reforms Akitoby et al. (2018). Some administrative reforms overhaul the entire organization of tax collection – including the actual creation of a tax authority and the transfer of the collection mandate away from specific ministerial units. Some elements of the enforcement environment may have such strong effects on collection that they translate into aggregate changes. Per example, Brockmeyer and Hernandez (2019) investigates an increase in the withholding rate in Costa Rica. This reform led to a visually striking and immediate increase of 8.1% in *aggregate* sales tax revenue, which was sustained over several years. Since sales tax revenue is a large component of overall tax collection in the country, this enforcement reform may have contributed to a shift in the overall tax collection regime.

The second set is institutional factors. Political structures which govern the organization of administrations may have first order impacts on collection performance. Dincecco (2009) studies the long-run historical experiences of several European countries, and finds that both administrative centralization and parliamentary reforms which curb the executive's discretionary power were associated with dramatic increases in aggregate per capita revenues. Over the time-frame that our data-set covers, countries around the world have undergone similarly drastic changes in governance structure. More generally, political events which cause sharp changes may trigger a transition in the tax collection regime – so long as the impact on tax collection is disproportionate to the impact on GDP. Per example, the civic and political changes occurring in the Democratic Republic of Congo in the early 1990's led to a significant and sustained decrease in GDP, but an even more pronounced collapse in tax collection.

The third set can broadly be considered as economic factors. Many of the studies cited in the literature review (Section 2) discuss economic factors which vary over development and which cause changes in tax/GDP. These factors could cause transitions in tax collection regime, for two reasons. First, if the change in the economic force is sufficiently sharp, it could induce a break in trend of tax/GDP. Per example, Bachas et al. (2022) use event-study analyses to investigate the impact of large trade liberalization events. These events led to sharp increases in the collection of both labor and capital taxes (expressed relative to their respective national income base), with breaks from trends. The authors provide suggestive evidence that these transitions in tax collection performance were due to sharp growth in the

enforceable (corporate) sectors of the economy. Thus, if the economic changes are sufficiently large and sharp, they may drive the regime changes captured in our methodology. The second reason is that, even if the underlying economic force evolves smoothly, the government's adaptation to these changes may create breaks in the tax/GDP series. Kleven, Kreiner, and Saez (2016) embed an optimal tax model into a macroeconomic growth model where the size and complexity of firms grow with exogenous technological progress. In their model, the first-best effective tax rate is constant along the development path. In earlier stages of development, technological progress causes some firms to be large enough to be taxable if the tax rate is not too high. In these stages, the enforcement constraint binds and the government tax rate is below the first-best but grows over time. In later development stages, firms have become so large that the enforcement constraint no longer binds – the government implements its first best effective tax rate (as a share of GDP), which stays constant over time. Thus, even if the underlying growth in enforcement is smooth over the full development path, tax/GDP breaks from an upward-growing trend to stable collection whenever the enforcement constraint becomes slack.

In summary, there are several factors which could cause a transition in the tax collection regime. Some factors are within the scope of, or directly related to, the tool-kits that are available to tax practitioners (comprehensive statutory reforms, administrative overhauls, enforcement investments). Other factors, such as economic forces and governance shocks, may be larger in scope and impact, in disproportionate ways, both the numerator (tax revenue) and the denominator (overall GDP). To carefully establish the role of these potential factors is beyond the scope of the current paper. Doing so requires not only a comprehensive literature review to establish the full list of potential factors but also the creation of a methodological framework which relates, with statistical confidence, the timing of these factors to the timing of the tax performance transitions. These tasks are best left for future research; in this paper, we are focused with the initial task of establishing a statistical framework for capturing the transitions in tax performance.

4 Results

In this section, we describe the results from applying the two methodologies to identify turning points in countries' tax/GDP ratios.

13

4.1 Prevalence of turning points

We begin by describing the prevalence of turning points. Each method uncovers a meaningfully large number of events. Using the statistical method, there are 464 breaks in countries' time-series. The number of turning points by year is displayed in Panel A of Figure 1. There are spikes in some individual years, and a slight increase in frequency in the 1980s and the 1990s – but, on the whole, there appears to be no sharp trend in the frequency of breaks. In Panel B of Figure 1, we measure the unconditional likelihood of a break occurring in a given country in a given year. This likelihood is not only determined by the number of breaks in a given year, but also by the changing number of countries that are present in each year (given the unbalanced sample). The likelihood of a breaking point calculated with the statistical method is very large – approximately 10.15 percent. Panel B reveals that, similar to the frequency count, there is no striking time-trend for the likelihood of break points in the global sample.

We repeat the exercise but while using the filter approach. One important result is that the filter approach yields a significantly smaller total number of events – 284 (compared to the 464 in the statistical method). Naturally, this difference could be driven by any of the three criteria that are required to qualify as an event in the filter method. Per example, in the case of an acceleration, the statistical method may pick up a turning point which does not satisfy the filter requirement that the post-event tax-GDP ratio must exceed the peak level in the pre-event period. Similarly, imposing the 8-year time-horizon as a filter may mean that the filter approach misses out on consecutive breaks in tax-GDP which the statistical method would otherwise pick up. Conversely, the smaller number of events identified with the filter method may be interpreted as evidence that the statistical method is capturing dynamic changes (such as temporary dips or spikes) in tax-GDP which do not fit the conceptual definition of an economically meaningful turning point. However, as we will see below, the statistical approach uncovers changes in tax-GDP which are large in magnitude and almost comparable to the filter approach, which weakens this argument.

With this important difference in mind, Panels A and B of Figure 2 display the trends over time in the frequency and likelihood of turning points using the filter approach. Similarly to the statistical method, there appears to be no strong time-trend in the prevalence of turning points. The unconditional likelihood of a

14

turning point based on the filter approach is 8.85 percent – economically large, but smaller in magnitude than the statistical approach.[4]

In Figure 3, we split the turning points into accelerations and decelerations. The two top panels show the likelihood of each type of turning point using the statistical method, while the two bottom panels show the likelihoods based on the filter approach. Several interesting results stand out. First, accelerations are more likely than decelerations, but the likelihood of a decelerating event is significant. Per example, using the statistical approach, the likelihood in a given year that a country will experience an acceleration is 6.12 percent, and the likelihood it will experience a deceleration is 4.97 percent. The respective likelihoods using the filter approach are smaller in levels, but there is similarly only a small difference in unconditional probability between acceleration events and deceleration events. In other words, characterizing countries' tax-GDP time-series as a steady but small positive growth in the long-run is a poor summary: countries experience both large, sustained upturns and downturns and they occur with almost comparable frequency.

When we focus on the trends in occurrences in the global sample, there appears to be no meaningful patterns that emerge. With the statistical method, the likelihood of an upturn or downturn across years is similar and the distributions are comparable to a uniform distribution. The distribution of likelihood using the filter approach is more volatile, but the distributions are somewhat comparable across upturns and downturns.

4.2 Magnitude of changes in tax-GDP ratio

Apart from the sheer number of turning points, the magnitude of the change in tax-GDP induced under a turning point is also impressive. In Figure 4, we calculate the average percentage point change in tax/GDP, by type of turning point, over 8 years of the post-event. Panel A shows the results using the statistical method. The

---

[4]It is interesting to note that, using this methodology, turning points in tax/GDP are more frequent than turning points in GDP. Indeed, in the well-known study on turning points in GDP growth, Hausmann, Pritchett, and Rodrik (2005) finds that the average probability of a growth transition is about 2.8 percent per year. The overlap between growth transitions and tax/GDP transitions merits further investigation – including a comparison within the exact same sample of countries and years, and using criteria which are common yet also account for differences in time-series properties of GDP vs tax/GDP.

average changes in tax/GDP are large - approximately 5 percentage points increase for accelerations, and 4.8 percentage points decrease for decceleration. Relative to the average tax/GDP ratio in the full sample of country-years (22.06%), this represents a 22.6% increase for an acceleration (21.7% decrease for a deceleration). In other words, in the typical episode estimated using the statistical method, tax/GDP stood 22.6% higher in the post-period than it would have without an acceleration.

It is important to note that these magnitudes are striking, given that the statistical method imposes no criteria on how large the changes in tax/GDP have to be to qualify as an event. The results in Panel A imply that, by focusing only on the statistical properties of tax performance to detect breaking points, the changes incurred under the detected turning points are economically large. This result lends credence to the interpretation that the statistical method, without any need to combine it with filters, is able to detect meaningful events in the tax/GDP time-series.

In Panel B of Figure 4, we repeat the exercise with the filter approach. In this case, note that the magnitude changes are by construction bounded from below by the filters imposed. At the same time, the magnitude changes are not bounded from above. Finding average magnitude changes which deviate from the lower bound suggest that the filtered events are not constrained by that specific criteria. The results are somewhat consistent with this interpretation. Both in the case of an acceleration and a decceleration, the average change in tax/GDP is, in absolute magnitude, approximately 7.5 percentage points.

Two results can be taken away from this figure which are common to each method. First, the magnitude results, when combined with the earlier result that turning points occur with a high likelihood (Figure 3), imply that breaks in tax-GDP are are economically significant macro-economic events. Second, the magnitude changes are, in absolute terms, close to symmetric across acceleration and decceleration events. This mirrors the results from the previous sub-section, where the likelihood of an acceleration event and of a decceleration event were also comparable.

In Figure 5, we document the average changes in tax-GDP induced by each type of event and using both methods. Similar to the global trends for the frequency of turning points, there appear to be first-order patterns that emerge. It is important to note that, in the global sample, the tax-to-GDP ratio of the representative country has increased between 1965 and 2020 (Bachas et al., 2022). Thus, while the magnitude

16

induced under a turning point has not necessarily shifted in percentage point terms, it has globally decreased when expressed as a fraction of the tax-GDP ratio.

In summary, rather than being characterized by a sustained, small and positive yearly growth rate, as most theories of tax capacity implicitly assume, the results thus far highlight that tax performance is marked by quantitatively large and sustained upturns and downturns. These turning points occur frequently and constitute large macro-economic events. As discussed in Section 3.3, these turning points may be caused by a multitude of factors, which is important work for future research. Regardless of their determinants, the existence of these turning points highlights that a significant amount of potentially valuable insights about tax performance can be drawn from analyzing countries' medium-run trends.

4.3   Contributions of tax components

In the final sub-section, we investigate the contribution of each main sub-component of the tax-GDP ratio to the turning points. Specifically, we divide the total taxes into its capital, labor and consumption components. We assign corporate income taxes, wealth taxes, and property taxes to capital. We assign payroll taxes and social security payments to labor. (3) Personal income taxes are allocated partly to labor and partly to capital, reflecting the fact that personal income is composed of salaries, capital income, and mixed income (see Bachas et al. (2022) for details). International trade taxes, as well as general and excise taxes on domestic goods, are classified as consumption taxes. We express each type of tax as a percentage of GDP – the three tax-GDP ratios add up to 100% of the total tax-GDP ratio in every country and year.

We first attempt to establish if turning points in tax-GDP are based on broad-based changes that occur in all sub-components of the tax-to-GDP ratio. For each sub-component, and each type of turning point based on overall tax-GDP, we code a dummy equal to 1 if the sub-component's tax-GDP ratio has the same sign of growth as the overall tax-GDP ratio. Per example, if a country experiences an acceleration event and the capital tax-GDP ratio also grows over the post-event episode, we code this occurrence with a value of 1; if the capital-tax GDP ratio decreases over the post-acceleration episode, then we code it with a value of 0. We construct these variables for all three sub-components of total taxation, and separately for acceleration and decceleration episodes.

17

The results are based on the pooled set of events identified with the statistical method and with the filter method. We have verified that results are similar within each method. For purposes of interpretation, it is important to note that the likelihood a particular sub-component's growth can have the opposite sign to the overall tax growth is related to that sub-component's share in overall taxation. Per example, consumption taxes constitute a large share of total taxes in developing countries, making it harder for total taxes to grow substantially without growth in consumption taxes. Similarly, labor taxes constitute a large share of total taxes in developed countries, thereby decreasing the likelihood that growth in total taxes and labor taxes are decoupled from one another.

Starting with upturns, in 21% of acceleration in total taxation, capital taxes see a decrease rather than an increase. This number is similar for consumption taxes (18%), but it is only 11% for labor taxes. In other words, upward turning points seem, in the global sample, to be more strongly linked to underlying growth in labor tax performance. We can define a 'broad-based' acceleration event as one in which all three sub-components also have positive growth. This occurs in 57% of acceleration episodes. In other words, upward turning points are not systematically based on strong growth in all tax-components. It is also noticeable that the likelihood of not witnessing a broad-based upturn is not too far from the sum of the likelihoods of witnessing opposite-sign growth for each sub-component.

Interestingly, in the case of downturns, 18% of decelerations are not accompanied by a corresponding fall in capital taxes. The number is similar for consumption taxes (17%), but a staggering 36% of decceleration cases see a decrease in total taxes but an increase in labor taxes. This result suggests that labor taxes may be a relatively robust source of systematically positive tax performance (compared to the other sub-components) – disproportionately contributing to upturns and often retaining positive growth when overall taxation incurs a downturn. Relatedly, the likelihood of a broad-based downturn is only 43%.

Our second investigation is to quantify the relative contribution of each sub-component to the magnitude of overall tax-GDP changes in each event. For the average acceleration event, we simply calculate the % of total growth that is accounted for by each sub-component. Relative to the average increase in tax-GDP during an acceleration (5.8 percentage points), the largest contribution comes from consumption taxes (which grows by 2.6 percentage points, or 45% of the total

increase); the remaining growth is almost evenly split between labor (1.7 percentage points, or 29%) and capital (1.5 percentage points, or 26%). In the case of the average downturn, which sees a 5.7 percentage point drop in total taxes, the main contribution is still from consumption taxes (2.5 percentage point reduction, or 44% of the drop). The contribution of capital taxes, 2.1 percentage drop or 37%, is now twice as large as the contribution of labor taxes (1.1 percentage drop or 19%).

As noted above, the fact that consumption taxes contribute meaningfully to both accelerations and deccelerations is perhaps not too surprising, given that consumption taxes is the most prevalent source of taxation in many countries around the world. What is perhaps more insightful is the asymmetric role that labor taxes and capital taxes appear to play – where strong growth in labor taxes feature prominently in accelerations, but capital taxes play a larger role in deccelerations.

5 Heterogeneity across development

In this section, we investigate if there are important differences in the time-series of tax-GDP across countries which differ in their level of development. We proxy for development by using the World Bank income per capita classification in the most recent sample-year, and consider the three groups of low-income countries; middle-income countries; and, high-income countries.

Conceptual discussion   Are turning points more prevalent at certain levels of development? From a theoretical point of view, the prediction which relates the likelihood of turning points to development is ambiguous. On the one hand, turning points may be more likely in less-developed countries. As constraints on collection are larger in these countries, it is possible that any improvements which alleviate constraints have higher impacts on tax collection than in developed countries. In other words, if the tax capacity returns to alleviating capacity constraints are concave, then reforms or events which improve collection are more likely to trigger turning points at lower levels of development. Moreover, more developed countries have higher tax/GDP, on average. If the level of tax/GDP in those countries lies close to the first-best effective tax rate (Kleven, Kreiner, and Saez (2016)), this also mechanically implies lower likelihood of observing acceleration turning points at higher levels of development. In short, these elements produce convergence in

tax/GDP along the (long-run) development path, and lead to the prediction that turning points are declining along the development path.

On the other hand, more developed countries have stronger tax administrations. Strong administrative capacity may be complementary to tax policies – that is, the same statutory reform may produce larger tax collection impacts whenever the supporting administrative and enforcement environment are stronger. Moreover, countries with levels of GDP per capita may exhibit more volatile time-series in tax/GDP. All else equal, this volatility attenuates our methodology's ability to detect turning points. Thus, there are both theoretical and statistical arguments which would predict that turning points are increasing in the country's level of development.

Results We first investigate whether the prevalence of turning points differs across development groups. In Figure 6, we calculate the likelihood of both accelerations and decelerations, separately across low-middle-high income countries, and separately by method (statistical approach in Panel A, filter approach in Panel B).

The results differ slightly by method. Focusing first on the statistical method, two main results stand out. First, the likelihood of a turning point occurrence is marginally higher in middle income countries than in both high income and low-income countries. Second, the symmetry in likelihood of occurrence between upturns and downturns is remarkably strong in all three groups. In other words, least-developed countries are as likely to witness a acceleration versus a decceleration (conditional on the occurrence of a turning point) as most-developed countries. In Panel B, the filter method suggests a stronger asymmetry between upturns and downturns. In particular, accelerations are more likely than decceleration in all development groups, and the gap gradually grows from the low-income to the high-income countries. In other words, according to the filter method, high-income countries are better shielded against downturns and more inclined to witness upturns.

Our second investigation concerns the magnitude of changes in tax-GDP, conditional on a turning point. In Figure 7, we calculate the post-event percentage point change in tax-GDP, separately by type of event and by development group, and separately using the statistical method (Panel A) and the filter method (Panel B). These results suggest limited heterogeneity across development levels. Similarly

to the global sample, the magnitudes are larger using the filter method than the statistical method. But conditional on a method, the magnitude of changes, both for accelerations and deccelerations, are approximately similar in all three development groups. Naturally, given the lower overall tax-GDP ratio at lower levels of development, the constant magnitudes in percentage point terms imply that both accelerations and deccelerations have larger impacts as a percent of tax-GDP in less-developed countries.

In summary, the investigation of differences across development did not yield a particularly strong and consistent result. Results differ somewhat across methods. Using the statistical method, both the frequency of upturns and downturns as well as the magnitude-change in tax performances are remarkably uniform across levels of development. Relative to the conceptual discussion, this may be due to the presence of several, countervailing forces that simultaneously impact turning points.

6 Conclusion

The literature on tax capacity seeks to understand the determinants which impact a state's ability to collect taxes in an equitable, efficient and progressive manner. Based on an important initial theoretical contribution by Besley and Persson (2014), two research agendas have emerged on tax capacity. The first literature seeks to establish the patterns of tax capacity across the development path – considering the full range of development across countries today and the long-run evolution within countries over time. The second literature takes a micro-economic approach and leverages administrative data. The evidence built across these two research strands thus sheds light on either the descriptive, long-run trends in tax capacity or the micro-designs which incrementally impact tax collection amongst individual firms or households.

In this sense, there is a 'missing middle' of evidence on tax performance at the intermediary level: in the aggregate, over medium-term horizons. This is arguably the level that policy-makers may be the most pre-occupied by, since it relates to important questions such as – how likely is it that the tax system undergoes a sustained acceleration? What can be done to avoid sustained deceleration and the erosion of effective tax collection? At this intermediary level of analysis, the relevant

object becomes to identify turning points in tax performance, which is the focus in our paper.

We deploy two methods, called the statistical method and the filter method, which draw on work in the economic growth literature, to estimate break points in the time-series of the tax-to-GDP ratio of individual countries. The methods are flexible, allowing us both to identify acceleration episodes and deceleration episodes. The methods reveal an impressively large number of turning points – 464 with the statistical method, 284 with the filter method. Importantly, for a country and a year picked at random, the likelihood of an acceleration occurrence is only slightly larger than the likelihood of a decceleration. In other words, far from being summarized by a small and steady long-run positive growth, countries' tax-GDP time-series are characterized by frequent downturns and upturns which occur with comparable likelihood.

Apart from the sheer number of and frequency of turning points, the magnitude of the typical turning point is also remarkably large. The average gain in tax/GDP in the case of an acceleration is 5 percentage points using the statistical method. Relative to the average tax/GDP ratio in the full sample of country-years (22.06%), this represents a 22.6% increase. In other words, in the typical episode, tax/GDP stood 22.06% higher in the post-period than it would have without any acceleration. The average loss in tax/GDP in the case of a deceleration episode is 4.8 percentage points – representing a staggering 21.7% loss in tax collection. Turning points are thus economically meaningful macro-economic events, owing both to the frequency of their occurrence and to the magnitude of tax-GDP changes incurred during these episodes.

This paper has focused on exploring methods to detect with confidence and precision the the existence and timing of regime shifts in countries' tax collection performance. The results in this paper may serve as a starting point for future research which investigates the determinants of these turning points.

References

Aizenman, Joshua and Mark Spiegel (2010). "Takeoffs". In: *Review of Development Economics* 14(2).

Akitoby, Bernardin et al. (2018). "Tax Revenue Mobilization Episodes in Emerging Markets and Low-Income Countries: Lessons from a New Dataset". In: *IMF Working Paper Series* 18(234).

Antoshin Sergei, Andrew Berg and Marcos Souto (2008). "Testing for Structural Breaks in Small Samples". In: *International Monetary Fund*(75).

Bachas, Pierre et al. (2022). "Globalization and Factor Income Taxation". In: *NBER Working Paper Series*(29819).

Bai, Jushan and Pierre Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes". In: *Econometrica* 66(1).

Berg, Andrew, Jonathan Ostry, and Jeromin Zettelmeyer (2012). "What Makes Growth Sustained?" In: *Journal of Development Economics* 98(2).

Bergeron, Augustin, Gabriel Tourek, and Jonathan Weigel (2021). "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC". In: *CEPR Discussion Paper Series*.

Besley, Timothy and Torsten Persson (2014). "Why Do Developing Countries Tax so Little?" In: *Journal of Economic Perspectives* 28(4).

Best, Michael et al. (2015). "Production vs Revenue Efficiency: Theory and Evidence from Pakistan". In: *Journal of Political Economy* 123(6).

Brockmeyer, Anne and Marco Hernandez (2019). "Taxation, Information and Withholding: Evidence from Costa Rica". In: *Unpublished Working Paper*.

Dincecco, Mark (2009). "Fiscal Centralization, Limited Government, and Public Revenues in Europe, 1650-1913". In: *The Journal of Economic History* 69(1).

Gaspar, Vitor, Laura Jaramillo, and Philippe Wingender (2016). "Tax Capacity and Growth: Is there a Tipping Point?" In: *IMF Working Paper Series* 16(234).

Gordon, Roger and Wei Li (2009). "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation". In: *Journal of Public Economics* 93(7).

Hamilton, James (2016). "Macroeconomic Regimes and Regime Shifts". In: *NBER Working Paper Series*(21863).

Hausmann, Ricardo, Lant Pritchett, and Dani Rodrik (2005). "Growth Accelerations". In: *Journal of Economic Growth* 10.

Jensen, Anders (2022). "Employment Structure and the Rise of the Modern Tax System". In: *American Economic Review* 112(1).

Jones, Benjamin and Benjamin Olken (2008). "The Anatomy of Start-Stop Growth". In: *Review of Economics and Statistics* 90(3).

Kar, Sabyaschi et al. (2013). "Looking for a Break: Identifying Transitions in Growth Regimes". In: *Journal of Macroeconomics* 38.

Kerekes, Monica (2011). "Analyzing Patterns of Economic Growth: A Production Frontier Approach". In: *Unpublished Working Paper*.

Kleven, Henrik, Martin Knudsen, et al. (2011). "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark". In: *Econometrica* 79.

Kleven, Henrik, Claus Kreiner, and Emmanuel Saez (2016). "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries". In: *Economica* 83(1).

Naritomi, Joana (2019). "Consumers as Tax Auditors". In: *American Economic Review* 109(9).

Oppel Annalena, Kyle McNabb and Daniel Chachu (2022). "The Dynamics of Domestic Revenue Mobilization Across Four Decades". In: *UNU Wider Working Paper Series* 22(1).

Pomeranz, Dina (2011). "No Taxation without Information: Deterrence and Self-Enforcemen in the Value-Added Tax". In: *American Economic Review* 105(8).

Pritchett, Lant (2000). "Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains and Plains". In: *World Bank Economic Review* 14(2).

Figure 1: Prevalence of turning points using statistical method

### (a) Number of turning points by year
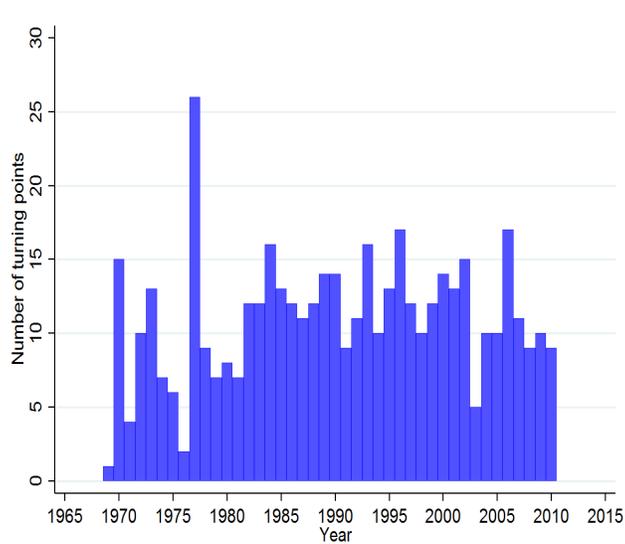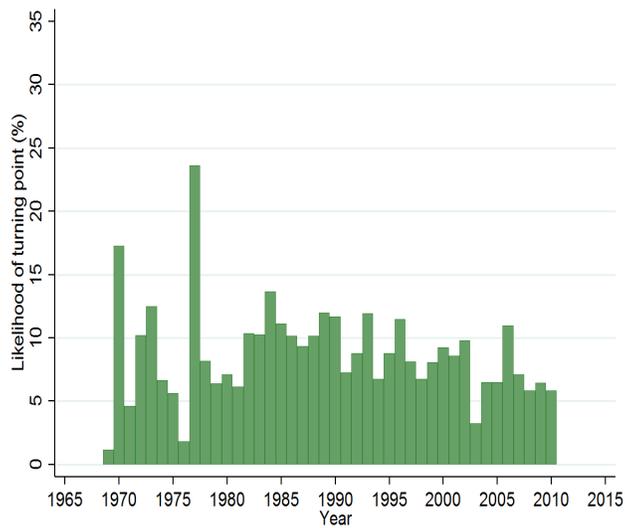
### (b) Likelihood of turning point occurrence by year

Notes: These graphs show the prevalence of turning points estimated using the statistical method. Panel A shows the number of turning points by year in the full sample. Panel B calculates the likelihood of a turning point occurring in a country, separately for each year. For more details on the method, see Section 3.

Figure 2: Prevalence of turning points using filter method

(a) Number of turning points by year
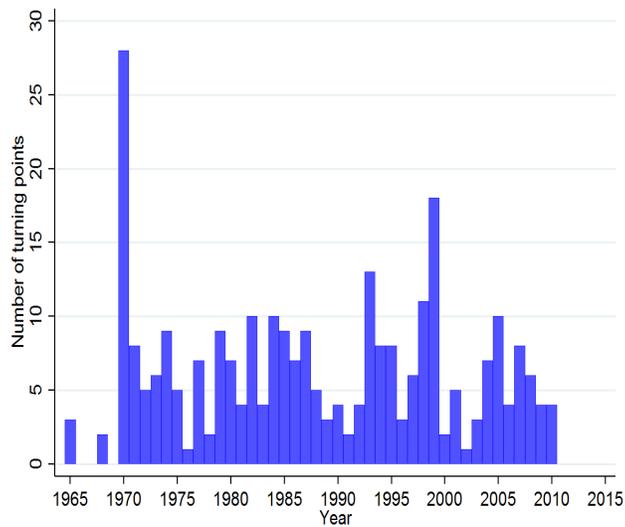
(b) Likelihood of turning point occurrence by year
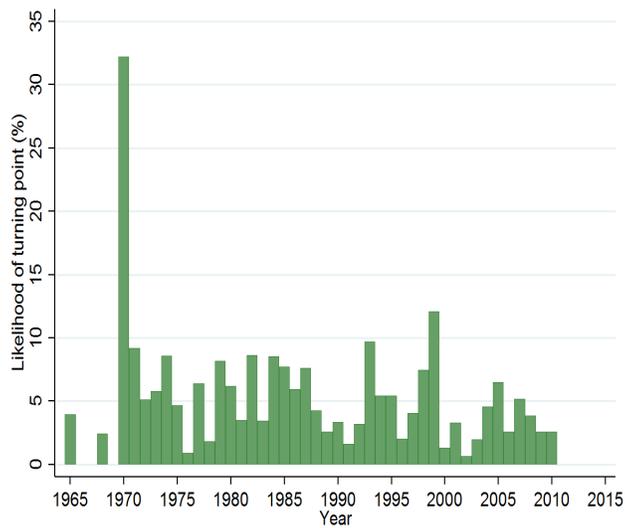
Notes: These graphs show the prevalence of turning points estimated using the filter method. Panel A shows the number of turning points by year in the full sample. Panel B calculates the likelihood of a turning point occurring in a country, separately for each year. For more details on the method, see Section 3.

Figure 3: Prevalence of acceleration and decceleration events

(a) Statistical method: Accelerations
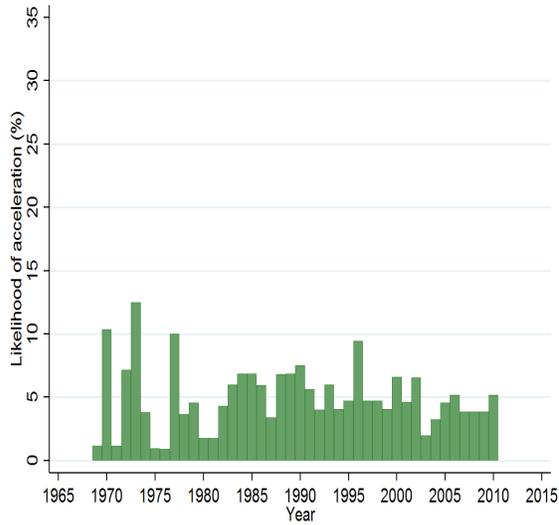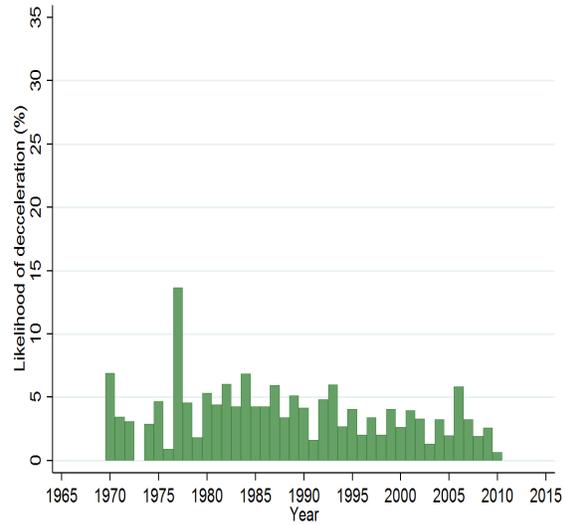
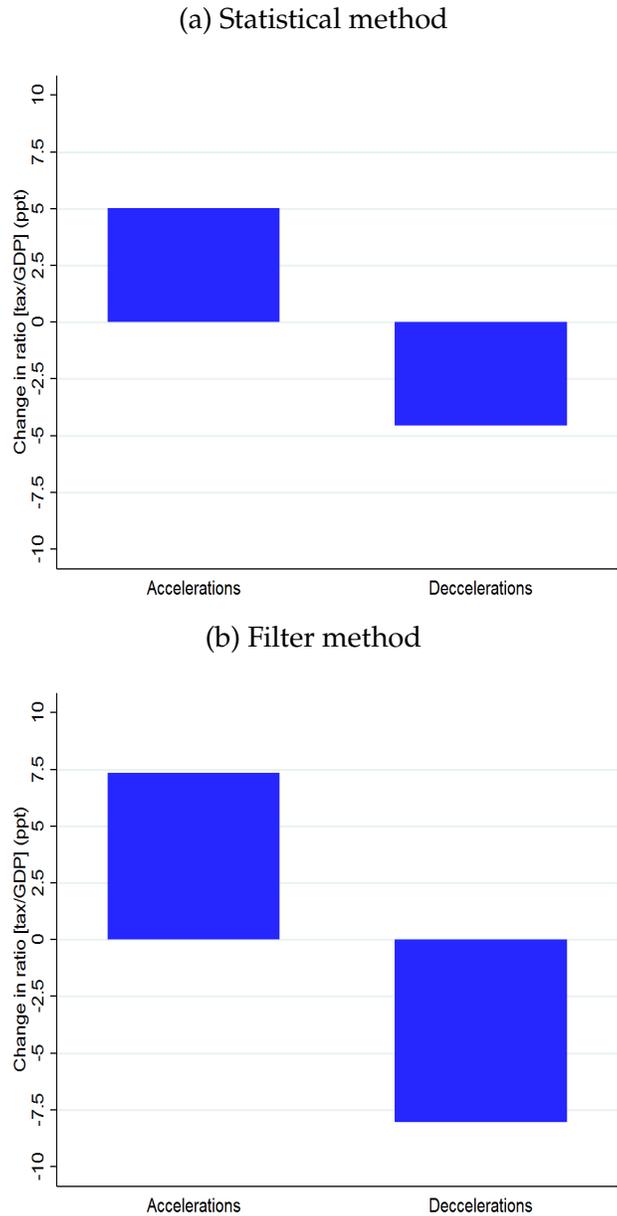(b) Statistical method: Deccelerations

(c) Filter method: Accelerations
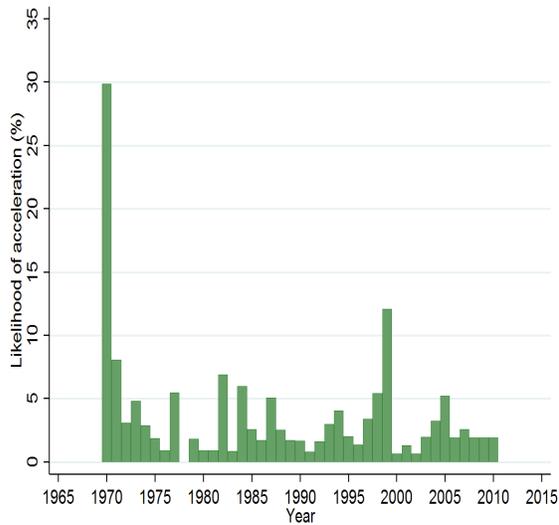
(d) Filter method: Deccelerations
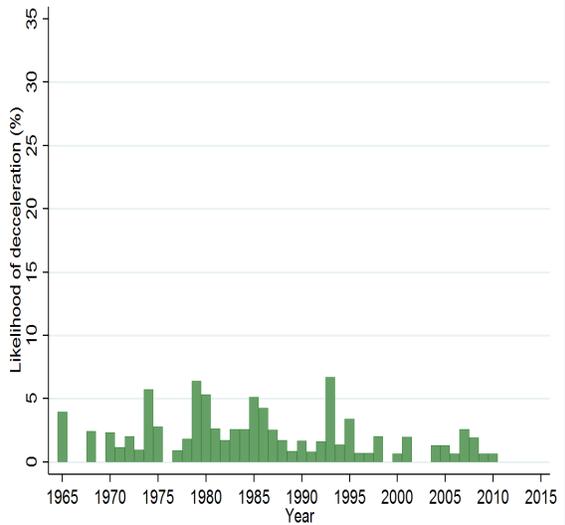
Notes: These graphs show the likelihood of turning points, separately by type and by method. The left-hand panels show the likelihood of accelerations, while the right-hand panels show the likelihood of deccelerations. The top panels use the statistical method, while the bottom panels use the filter method. For more details on the method, see Section 3.

Figure 4: Magnitude of changes in tax-to-GDP ratio

(a) Statistical method

(b) Filter method

Notes: These graphs show the magnitude of changes in tax-to-GDP induced by turning points, separately by type and method. The change is calculated as the percentage point change in the 8-year period following a detected event. In each panel, the bars report the average change, separately for accelerations and for decelerations. Panel A uses the statistical method, while Panel B uses the filter method. For more details on the method, see Section 3.

Figure 5: Magnitude of changes in tax-to-GDP over time
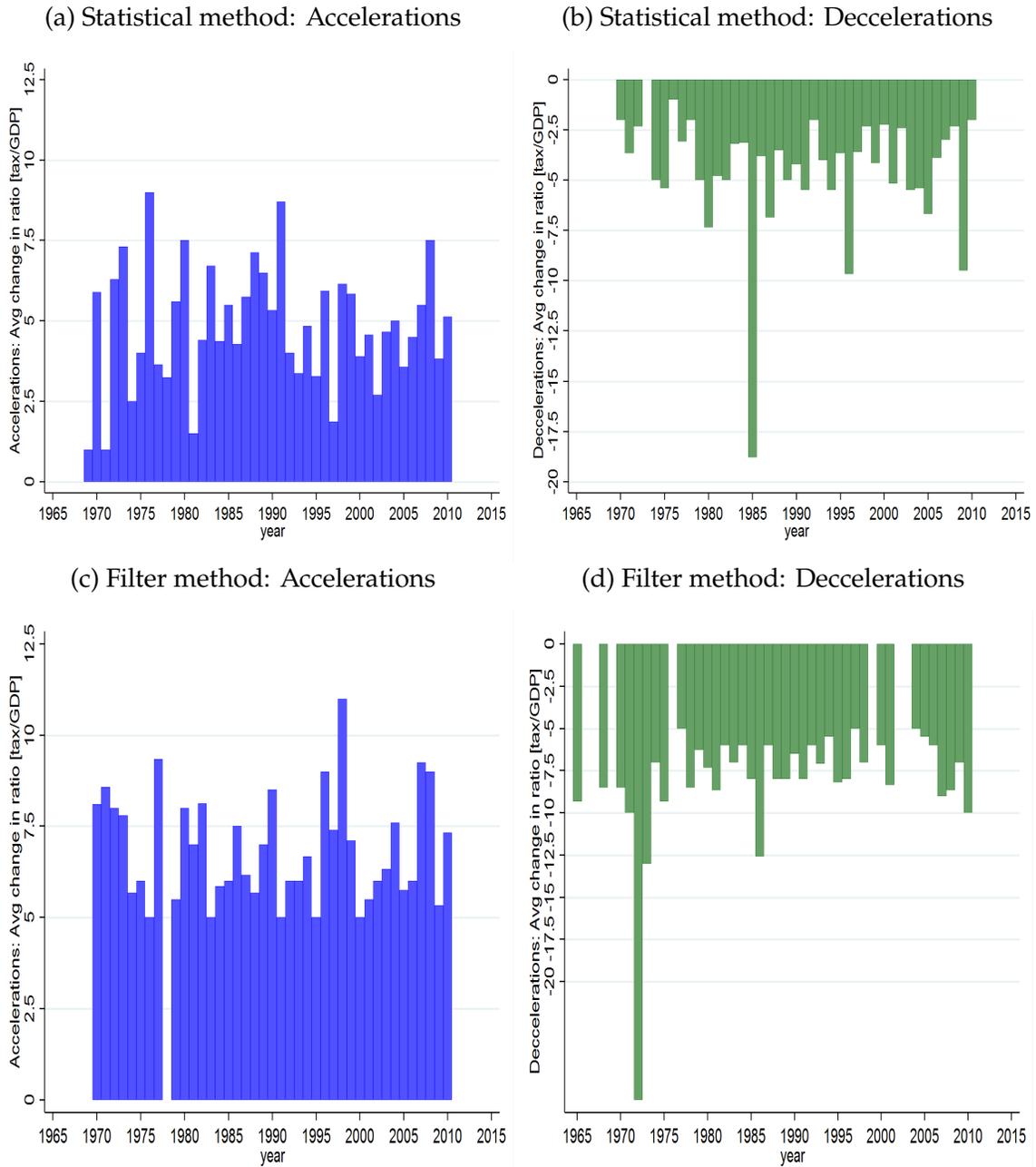
(a) Statistical method: Accelerations

(b) Statistical method: Deccelerations

(c) Filter method: Accelerations
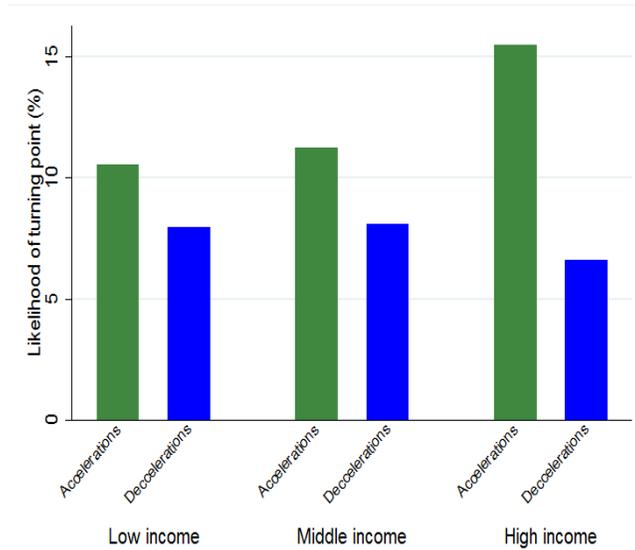

(d) Filter method: Deccelerations

Notes: These graphs show the magnitude of changes in tax-to-GDP induced by turning points, separately by type, method and year. The change is calculated as the percentage point change in the 8-year period following a detected event. In each panel, the bars report the average change in a given year. The left panels report the changes induced by accelerations, while the right-panels show the changes induced by deccelerations. The top panels use the statistical method, while the bottom panels use the filter method. For more details on the methods, see Section 3.

29

Figure 6: Prevalence of turning points by development level
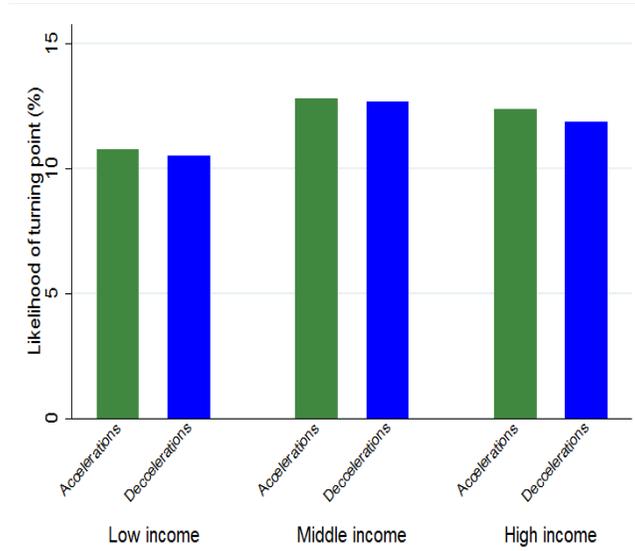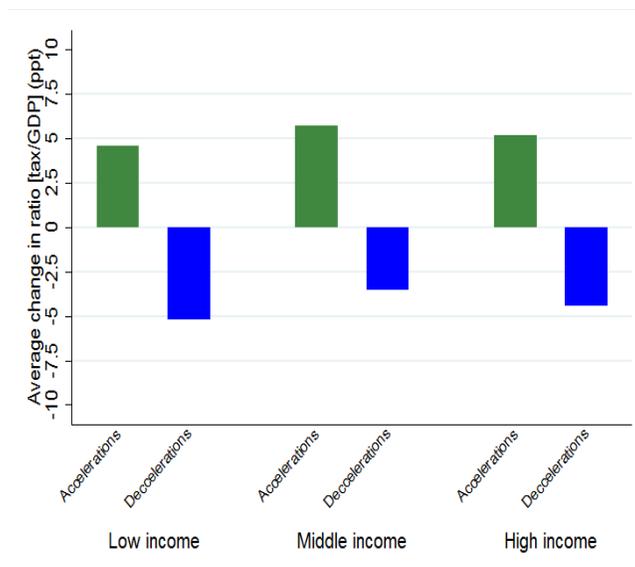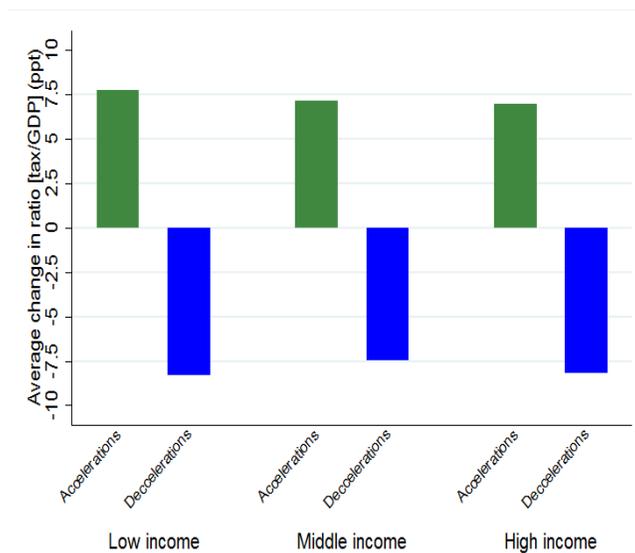
(a) Statistical method

(b) Filter method

Notes: These graphs show the likelihood of turning points, separately by type, method and development group. The likelihood is calculated as the number of events divided by the sample of country-year observations. In each panel, the bars report the probability, separately by type (acceleration and decceleration) and separately by the country's development group (low income, middle income, high income). The top panel uses the statistical method and the bottom panel uses the filter method. The development categories are based on the World Bank's income classification. For more details on the methods, see Section 3.

Figure 7: Magnitude of changes in tax-to-GDP by development level

(a) Statistical method

(b) Filter method

Notes: These graphs show the magnitude of changes in tax-to-GDP induced by turning points, separately by type, method and development group. The change is calculated as the percentage point change in the 8-year period following a detected event. In each panel, the bars report the average changes, separately by type (acceleration and decceleration) and separately by the country's development group (low income, middle income, high income). The top panel uses the statistical method and the bottom panel uses the filter method. The development categories are based on the World Bank's income classification. For more details on the methods, see Section 3.