INTERNATIONAL CENTER FOR PUBLIC POLICY

International Center for Public Policy
Working Paper 23-21
December 2023

Disentangling Tax Capacity and Effort

Andrey Timofeev

Georgia State University | Andrew Young School of Policy Studies

International Center for Public Policy
Working Paper 23-21

Disentangling Tax Capacity and Effort

Andrey Timofeev

December
2023

International Center for Public Policy
Andrew Young School of Policy Studies
Georgia State University
Atlanta, Georgia 30303
United States of America

Phone: (404) 413-0235
Fax: (404) 651-4449
Email: paulbenson@gsu.edu
Website: http://icepp.gsu.edu/

Copyright 2023, the Andrew Young School of Policy Studies, Georgia State University. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means without prior written permission from the copyright owner.

International Center for Public Policy
Andrew Young School of Policy Studies

The Andrew Young School of Policy Studies was established at Georgia State University with the objective of promoting excellence in the design, implementation, and evaluation of public policy. In addition to four academic departments, including economics and public administration, the Andrew Young School houses eight leading research centers and policy programs, including the International Center for Public Policy.

The mission of the International Center for Public Policy (ICePP) at the Andrew Young School of Policy Studies is to provide academic and professional training, applied research, and technical assistance in support of sound public policy and sustainable economic growth in developing and transitional economies.

ICePP is recognized worldwide for its efforts in support of economic and public policy reforms through technical assistance and training around the world. This reputation has been built serving a diverse client base, including the World Bank, the U.S. Agency for International Development (USAID), the United Nations Development Programme (UNDP), finance ministries, government organizations, legislative bodies, and private sector institutions.

The success of ICePP reflects the breadth and depth of its in-house technical expertise. The Andrew Young School's faculty are leading experts in economics and public policy and have authored books, published in major academic and technical journals, and have extensive experience in designing and implementing technical assistance and training programs. Andrew Young School faculty have been active in policy reform in over 40 countries around the world. Our technical assistance strategy is not merely to provide technical prescriptions for policy reform, but to engage in a collaborative effort with host governments and donor agencies to identify and analyze the issues at hand, arrive at policy solutions, and implement reforms.

ICePP specializes in four broad policy areas:

- Fiscal policy (e.g., tax reforms, public expenditure reviews)
- Fiscal decentralization (e.g., reform, intergovernmental transfer systems, urban finance)
- Budgeting and fiscal management (e.g., local, performance-based, capital, and multi-year budgeting)
- Economic analysis and revenue forecasting (e.g., micro-simulation, time series forecasting)

For more information about our technical assistance activities and training programs, please visit our website at icepp.gsu.edu or contact us at paulbenson@gsu.edu.

Disentangling Tax Capacity and Effort

Andrey Timofeev[*]

December 2023

Abstract

For over 50 years, the IMF and other international organizations have compared tax performance across countries. Over this time, both data availability and sophistication of statistical methods have improved considerably. However, these new tools cannot eliminate the fundamental problem of imperfect information, as the observed tax outcome is a joint product of unobservable tax effort and tax capacity. This paper aims to serve as a refresher on the purposes and assumptions underlying this kind of empirical study and to reassess the best technical solutions of currently available modeling approaches, given the improvements in the availability of longitudinal data and advances in dynamic panel analysis made over these 50 years. Using an error-correction model in this paper, I attempt to disentangle the long-run relationship between tax capacity and economic development from short-run adjustments in response to economic cycles and other transient shocks.

Keywords: tax capacity, tax effort, heterogeneous dynamic panels

JEL Codes: C23, C51, H21

[*] tel: (+1) 404-413-0230, fax: (+1) 404-413-0244, e-mail: atimofeev@gsu.edu

Introduction

For over 50 years, the IMF and other international organizations have compared tax performance across countries. Over this time, both data availability and sophistication of statistical methods have improved considerably. However, these new tools cannot eliminate the fundamental problem of imperfect information, as the observed tax outcomes represent a joint product of two distinct groups of mostly unobservable factors: 1) those under government control ("tax effort") and 2) external factors outside government control ("tax capacity").

While the distinction between these two groups of factors is crucial for the correct specification of empirical models and interpretation of their results, from the very early attempts (e.g., Lotz and Morss 1967), this was acknowledged to be a judgment call ultimately that has to be made based on the specific policy application at hand. In particular, multiple iterations of studies conducted by the IMF over the 50 years identified two distinct policy applications: 1) the assessment of tax space available for policy interventions and 2) the monitoring and evaluation of fiscal reform progress. Additionally, these two applications might imply different judgments as to which factors are under government control and thus accounted for as "tax effort," and they might require different empirical strategies.

In particular, early studies identified cross-sectional ("between") estimators as more suitable for the first application (tax space), while the difference-in-difference ("within") estimators were believed to be more appropriate for the second policy application (tracking progress). Although the technical approaches used by these studies evolved into more sophisticated methods, less discussion has been provided in more recent studies for these conceptual issues and underlying assumptions. This might have led to the 'unlearning' of some of the insights gained in the earlier studies.

This paper aims to serve as a refresher on the objectives and assumptions underlying this kind of empirical study and to reassess the most suitable technical solutions from currently available modeling approaches in light of improvements in the availability of longitudinal macro-fiscal data and advances in dynamic panel analysis made over these 50 years. I argue that, by adopting fixed-effects and stochastic frontier models, the evolution of empirical strategies has strayed from the underlying rationale of the policy issues they purportedly address. Rather than the paradigm of the "maximum output attainable given a set of inputs" (i.e., frontier) featured in more recent studies, in my opinion, the notion of tax capacity should be framed in terms of the level of taxation that is *optimal* for the circumstances of a given country so that the actual tax outcomes can be above that optimal level as well as below it. This is certainly the case when the sample includes emerging economies from the former socialist block, which started their transition process with quite high tax-to-GDP ratios that subsequently came down to levels more common for other countries of comparable income levels.

Previous empirical studies found a remarkable persistence of cross-country differences in tax-to-GDP ratios over time, with observable characteristics explaining at best half of the between variation. Furthermore, the bulk of the within (over time) variation appears to be driven by the economic cycle. By employing an error correction model in this paper, I attempt to disentangle the long-run relationship between tax capacity and economic development from short-run policy adjustments in response to economic cycles and other transient shocks.

Motivation and Literature Review

The last decade has witnessed a renewed interest from the international development community in the issue of domestic resource mobilization, due to a variety of reasons (fiscal austerity in developed countries after 2009, costs of climate change adaptation, Millennium Development

Goals, post-COVID-19 recovery, etc.). The tax-to-GDP ratio is commonly used to measure the relative performance of countries in domestic resource mobilization. Initially, IMF produced two complementary measures of tax effort: one being the growth of the tax-to-GDP ratio over time, the other being a cross-country comparison of three- and five-year averages of this ratio.

However, comparisons to the previous-year outcome in the same country (internal benchmarking) can mistake chronically low tax effort for low tax capacity. At the same time, comparisons to other similar countries (external benchmarking) might attribute the impact of unaccounted differences in tax capacity to unequal tax effort.

The notions of "fiscal capacity" and "tax effort" originally emerged in the application to subnational government units in the field of intergovernmental fiscal relations (ACIR 1962). However, that field had a completely different setting and methodology. In particular, the notions of "fiscal capacity" and "tax effort" were proposed as relative rather than absolute indicators. In turn, that framing must have been informed by the previous literature on "taxable capacity," defined as the "capacity to raise revenues without extreme interference with productive activity and the operation of the economy" (Kimmel 1949, p. 154).

In the international development field, multiple iterations of studies conducted by the IMF over the last five decades identified two distinct policy applications. The first is an assessment of tax space available for policy intervention. This, for example, was articulated by Lotz and Morss (1967, p. 478) as the "scope for an increase in the level of taxation in a country as part of a stabilization program, for the mobilization of resources to finance a development program." The second policy application is related to monitoring and evaluation of fiscal reform progress. For example, Chelliah (1971, p. 259) stated that, "it is of interest to those concerned with problems of development—policymakers in developing countries, international organizations, and donors of foreign aid—to know what progress different countries have made in mobilizing resources for development through their tax systems." Not only might these two applications imply different judgments as to what factors can be deemed under government control and thus accounted for under "tax effort," but they might also require different empirical strategies.

One common weakness of such empirical studies is a lack of any normative foundation in the shape of a formal model that could guide the design and interpretation of the econometric analysis. Early on Chelliah (1971, p. 292) acknowledged the role played by "preferences of the people and the leaders as between public and private services, including the institutional arrangements (for the fulfillment of particular needs) arising from them." He further argued that "the needs and preferences for public goods may be said to affect the tax ratio through the willingness to tax" (p. 293).

Similarly, Bahl (1971) pointed out the role played by the "cultural style," in terms of the "nature of the historical arrangements for the division of financial responsibility for certain activities between the public and private sectors" (p. 583), as well as "a country's preference for taxing or not taxing a particular base" (p. 584), and "preferences for public services and for methods of supplying them" (p. 600).

Tait et al. (1979, p. 125–6) made a similar argument:

> *Even when the relative prices of publicly and privately provided goods do not differ across countries, tastes relevant to the public-private goods mix may do so. Two countries seeking to equate the marginal social benefits of public and private provision of goods and services will achieve very different tax ratios, which should not be seen as indices of effort but rather should be seen as the result of conventional maximizing behavior.*

Tait et al. (1979) further argued that the very term “tax effort” might be misleading: instead, they came up with an alternative notion of “international tax comparison” (ITC), which they used to frame their analysis.

Attempts to formalize the notions of tax capacity and tax effort started more recently. Leuthold (1991) offered a very stylized model where the actual tax outcome is a function of both the optimum tax level as well as availability of ‘tax handles.’ In the literature on intergovernmental fiscal relations, tax effort was analyzed in the framework of optimal-tax theory by Dahlby and Wilson (1994). In international settings, Dalamagas et al. (2020) modeled the optimal mix of tax rates as a function of the country-specific households’ preferences over equity/efficiency and then derived tax capacity as the sum of the products of the optimal tax rates and their corresponding tax bases.

While allowing for different weights that might be attached to the same criteria in the decision maker’s optimization function of different countries, in the rest of this section I catalog various factors that have been conjectured in the literature as determinants of either tax capacity or tax effort. These factors can be related to one of the three competing objectives that a social planner is commonly hypothesized to balance against each other in designing an optimal tax system:

1. Equity (or more broadly, political acceptance)
2. Economic costs
3. Administrative convenience/costs

The level of per capita income was identified in the literature as a proxy for the equity dimension. As pointed out by Lotz and Morss (1967, p. 481), with increasing per capita income “a smaller proportion of total income is required for subsistence needs and more ‘surplus’ is available for taxation and other purposes." Other than the ability to pay taxes, per capita income also enters the economic efficiency consideration in the form of optimal allocation between private and public goods. Thus, Chelliah (1971, p. 280) alluded to the "high income elasticity of demand for public goods."

The level of development can also serve as a proxy for many factors affecting administrative constraints on tax collection. As pointed out by Lotz and Morss (1967, p. 481), "Economic development is usually accompanied by a higher rate of literacy, increased monetization, and stricter law enforcement—all of which can be expected to increase taxable capacity." Similarly, tax administration costs are conjectured to be determined by the size of the foreign trade sector.[1] In particular, Lotz and Morss (1967, p. 483) identified two reasons:

> *First, it is administratively easier to tax trade inflows and outflows than domestic transactions. Second, the "degree of openness" in many countries, especially in early stages of development, indicates the relative importance of cash crops and subsistence agriculture, as well as the degree of urbanization and industrialization.*

Similarly, Chelliah (1971, p. 282) hypothesized that "economic growth with its concomitants of urbanization, proliferation of salary incomes, and the development of intangible forms of property would gradually bring about the ascendancy of modern forms of income and property taxation." Tait et al. (1979, p. 125) pointed out that the "importance of large producers,

[1] While not acknowledged in this literature, the flip side of open economies is that their tax capacity might be constrained by the mobility of capital. Even though trade is not the same as capital flows, the two might be correlated due to such business practices as transfer pricing.

employers, and retail establishments is positively correlated with the level of economic development.[2]“

The relative shares of GDP accounted for by mining and agriculture have also been conjectured to affect the administrative (and political) costs, as the former is associated with taxable activities concentrated in a few large (and often foreign-owned) entities while the latter can be associated with a large number of informal businesses. However, it has been pointed out in the literature (e.g., Chelliah 1971, Bahl 1971) that political will to tax agriculture might play as big of a role as administrative feasibility or equity concerns. Thus, some scholars (e.g., Keen 2012) argued that politics aside, it was not hard historically to administer taxes on agriculture due to land holdings remaining highly concentrated in many developing countries after independence. All these studies seem to find the share of agriculture in the economy to be an important determinant of tax outcomes but disagree on whether it should be accounted for under “tax effort” or “tax capacity.”

There has been a lot of cross-fertilization between tax effort studies and positive empirical studies explaining differences in tax systems across countries (e.g., Williamson 1961, Hinrichs 1965, Thorn 1967, Shin 1969, Weiss 1969). However, the latter strand of literature was motivated by a different policy concern and was framed differently, mostly as a “political means of raising the aggregate savings ratio through public saving” (Weiss 1969, p. 348). The main methodological difference is that the tax effort studies have parsed these determinants of tax outcomes into those under and outside government control and excluded the former from their regressions aimed at estimating the tax capacity. Earlier studies of tax effort (e.g., Lotz and

[2] The trends emerging since the 1990s have led some scholars to conjecture that the share of economic activity flowing through the “conduit of the corporation” might peak at some level of development, beyond which it would start shrinking again due to outsourcing, deindustrialization, and so on (e.g., Bird 2002, p. 199).

Morss 1967, Chelliah 1971, Bahl 1971), while acknowledging a number of observable determinants of tax effort, chose not to include those variables as explanatory variables, leaving them in the regression residual. Thus, Chelliah (1971, p. 298) justified this strategy with an assumption that the "factors affecting the willingness to tax are largely independent of the capacity factors whose effects have been estimated."

However, many of those studies acknowledged possible reverse causality or confounding in the relationship between tax effort and tax capacity (e.g., Bahl 1971). Obviously, taxation can affect economic development, which those studies identified among the determinants of tax capacity. Furthermore, under the Stochastic Frontier Analysis framework, employed in some recent studies, a failure to explicitly account for observable determinants of the tax effort can lead to severely biased results (Wang and Schmidt, 2002). Therefore, conceptually, the main methodological difference between these two types of studies (tax effort vs. tax outcomes) should not be as much about the specification of econometric models as about the interpretation of obtained estimates. After performing regression analysis to explain as much variation in tax outcomes as possible using all available proxies for both tax effort and tax capacity, one can use the predicted impact of the latter (determinants of tax capacity) to control for differences in tax capacity and attribute the residual differences to tax effort.

However, when both tax effort and tax capacity are confounded by some unobserved lurking variable, such as the economic cycle, the data might not allow one to distinguish unobserved heterogeneity in tax capacity from unobserved heterogeneity in tax effort, and thus it would become essentially a judgment call. Tait et al. (1979) attempted to use two-stage least squares to address potential endogeneity, but the resulting estimates were similar to those obtained with OLS. Gupta (2007) attempted to address the endogeneity of some explanatory variables, such as foreign debt and indebtedness, with the system-GMM approach. He also did not find qualitative differences compared to the baseline regression. However, the system-GMM approach is known to rely on weak instruments.

Tax-to-GDP ratios are more suitable for static cross-country comparisons, if the economic cycle is accounted for by averaging across years (i.e., the between estimator). In a dynamic setting, tax buoyancy appears as a natural extension of such an approach. Unlike for static (between) estimators, there has been less cross-fertilization between the studies of the dynamic (within) estimators of tax effort and other strands of literature attempting to address similar problems of imperfect information, even if framed differently.

One such relevant strand of literature deals with estimating short- and long-run tax buoyancy (observed tax change) and tax elasticity (counterfactual tax change holding the tax policy constant). Thus, Newlan (1985) suggested revenue elasticity as a benchmark for revenue buoyancy. However, revenue elasticity only takes into account tax policy efforts while neglecting tax administration efforts. Indeed, the relation between tax capacity (unobserved) and tax-to-GDP ratio (observed) is similar to that between revenue elasticity (unobserved) and revenue buoyancy (observed). In a dynamic setting, there is also a distinction between long-term elasticity (equilibrium) and short-term deviations/adjustments.

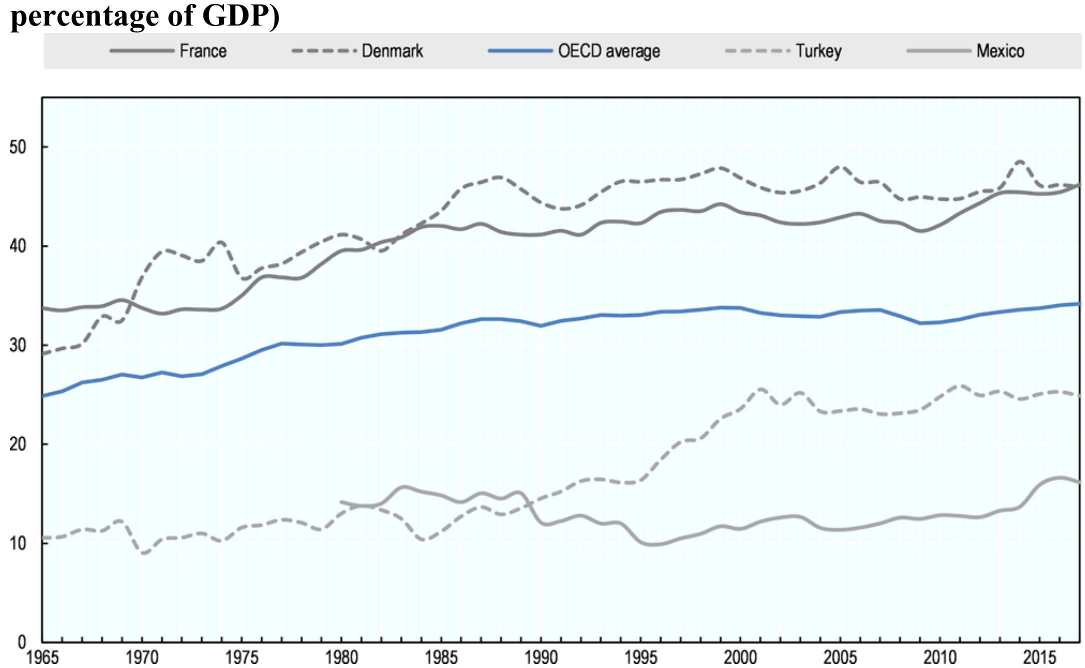
In fact, the relevance of revenue elasticity was identified in the early studies of tax effort, as for example articulated by Chelliah (1971, p. 301):

> *Strictly speaking, tax effort is a process; it takes several forms, including reform of existing taxes, improvement in administration, and introduction of new taxes. All these steps necessarily require time to plan, legislate, and implement. Countries that for historical or other reasons started out with a low tax ratio a decade or so ago might have undertaken considerable effort to raise their respective tax ratios but may not yet have reached even the average*

level of taxes in developing countries. Tax effort, therefore, should be considered also in the dynamic sense of comparing changes in the tax ratio over time. Thus, even if a country has a low tax effort in the static sense, the question may be asked whether it has made efforts over a period of time to increase tax revenues. For this purpose, it seems best (for reasons explained earlier) to compare the income elasticities of total taxes.

Another problem with slowly evolving variables is that past shocks fade out slowly over a number of subsequent periods, which may violate the assumption of independently distributed disturbances and may lead to misleading inferences due to autocorrelation of the disturbances. Thus, empirical studies have found some of the macroeconomic explanatory variables used in tax effort regression, such as GDP per capita, to be non-stationary. When combined with slope heterogeneity, persistence of explanatory variables makes the within estimator inconsistent (Pesaran and Smith 1995).

Figure 1. Trends in Tax-to-GDP Ratios for Selected OECD Countries, 1965-2017 (as a percentage of GDP)



Source: Revenue Statistics (OECD 2018)

Empirical data exhibit considerably more variation in the tax-to-GDP ratio between countries than variation over time for a given country, as revealed by the ANOVA decomposition of longitudinal data. Thus, in the panel dataset on 161 countries over 2003–17, the between variation accounts for 92 percent of the total variation in tax-to-GDP ratios. Furthermore, a large portion of the remaining within variation in the tax-to-GDP ratio appears to be driven by the economic cycle (Figure 1).

Thus, for OECD countries, tax-to-GDP ratios increased through the 1990s, with the OECD average peaking at 33.9 percent in 2000. Tax-to-GDP ratios fell between 2001 and 2004 (to 33 percent), but then rose again between 2005 and 2007 to an average of 33.7 percent before falling back sharply to 32.3 percent by 2009 after the global financial crisis. The average OECD tax-to-GDP ratio of 33.9 percent in 2014 was 0.2 percentage points higher than the pre-crisis level of 33.7 percent in 2007 and returned to the previous high of 33.9 percent, recorded in 2000.

Previous studies found remarkable persistence of cross-country differences in tax-to-GDP ratios over time. This led Bird (2011, p. 438) to argue that "countries appear to achieve a sort of equilibrium position with respect to the size and nature of their fiscal systems that reflects the balance of political forces and institutions and then to stay there, fluctuating around some relatively stable ratio, until 'shocked' into a new equilibrium." Earlier, Peacock and Wiseman (1961, p. xxiv) articulated a similar hypothesis as the "displacement effect:"

> *…in settled times, notions about taxation are likely to be more influential than ideas about desirable increases in expenditure in deciding the size and rate of growth of the public sector. There may thus be a persistent divergence between ideas about desirable public spending and ideas about the limits of taxation. This divergence may be narrowed by large-scale social disturbances, such as major wars. Such disturbances may create a displacement effect, shifting public revenues and expenditures to new levels. After the disturbance is over new ideas of tolerable tax levels emerge, and a new plateau of expenditure*

may be reached, with public expenditures again taking a broadly constant share of gross national product, though a different share from the former one.

In a panel data analysis of tax-to-GDP ratios in eight African countries over nine years, based on the Durbin-Watson test, Leuthold (1991) rejected zero autocorrelation in residuals from the pooled OLS regression. After transforming the data to remove the autocorrelation by using a first-stage estimate of the first-order autoregressive parameters, Leuthold found dramatic changes in the second-stage estimates produced with a generalized least-square (GLS) estimator.

In a panel analysis of 105 developing countries over 25 years, Gupta (2007) found that the Wooldridge (Breusch–Godfrey) Test for Autocorrelation in Panel Data rejected the hypothesis of no autocorrelation in the residuals from the within regression. After removing the autocorrelation with the Prais–Winsten transformation, performed in tandem with panel corrected standard errors (PCSE), Gupta found estimates to be qualitatively similar to those from the within estimator but with larger PCSE standard errors. In the system-GMM framework, by explicitly accounting for the persistence of tax outcomes in the form of a lagged-dependent variable, Gupta estimated the autoregressive parameter to be around 0.8.

In the next section, I propose an empirical strategy to disentangle the long-run relationship between tax capacity and economic development from short-run policy adjustments in response to economic cycles and other transient shocks. Thus, long-term trends in the tax-to-GDP ratios of different countries might be an outcome of welfare optimization aimed at achieving the optimal public-private mix of consumption given country-specific constraints, while short-term fluctuations can represent certain rigidities[3] (e.g., in borrowing or tax administration) that might

[3]Thus, Adolph Wagner reportedly believed that revenue raising constraints were binding only in the short run while in the long run constituents' demands and social progress drive the expansion of the public sector, this hypothesis is known as the Wagner's 'law" (Afxentiou 1980).

cause deviations from this optimum. I argue that such disentanglement of short- and long-run relationships can be achieved with an error correction model.

Empirical Framework

In the most general form, a panel data model of the tax outcome y_{it} in country i, and time t, can be specified as

$$y_{it} = m(x_{it};\beta_i) + c_i + \eta_i + u_{it} + v_{it}$$

where

- $m(x_{it};\beta_i)$ captures the impact of observed country characteristics x_{it},
- c_i captures unobservable time-invariant differences in tax capacity,
- η_i encapsulates unobservable time-invariant (persistent) differences in tax effort,
- u_{it} represents unobservable time-varying (transient) tax effort, and
- v_{it} captures stochastic shocks (impact of all time-varying determinants of tax capacity other than those captured by the measurable country characteristics x_{it}).

Until very recently, there was no methodology available to estimate a model in this general form (i.e., disentangling all four unobservable components and allowing for slope heterogeneity). Older models had to rely on simplifying assumptions to make analysis tractable.

Between estimators

Using this longitudinal data, there are two ways to look at the relationships among variables in this dataset. First, one can examine how cross-country differences in tax-to-GDP ratios are linked to differences in country characteristics, which change very slowly over time, if at all. This is achieved with the help of the "between estimator," which essentially takes the mean of each variable for each country over time and runs a regression on the collapsed dataset of means.

This strategy allows one to examine the impact of the country characteristics discussed above, which, while changing slowly over time, vary more significantly across countries. However, the collapsing of data results in a loss of information, in particular on the evolution of variables over time, which is exactly what tax effort is mostly about.

Chelliah (1971, p. 256) used three-year averages "in order to minimize the influence of fortuitous factors." The same strategy was used by Lotz and Morss (1967) and Bahl (1971). With improved data availability, subsequent studies started to use five-year averages instead (e.g., Weiss 1969, Tait et al. 1979). Essentially, these early studies estimated average relationships conditional on various country characteristics (economic, demographic, and geographic factors, etc.) to filter out the effect of statistical noise. Time averaging of the actual tax outcomes is supposed to eliminate any cyclical fluctuations so that any remaining deviations from the estimated average relationship can be attributed to tax effort. For panels with sufficiently large time dimension (T), this strategy found some validation in Pesaran and Smith (1995).

Therefore, I perform the between estimation as a baseline given that, under certain assumptions, when T and N are large, such a cross-section regression based on the time averages of variables should produce a consistent estimate of the mean of the long-run coefficients (Pesaran and Smith 1995). In particular, the latter result was developed for a general case of heterogenous panels where coefficients, while being constant over time, can differ randomly across countries. However, these random coefficients are assumed to be independent of regressors. In particular, if the initial conditions (country effects) were correlated with regressors, the between estimator would produce biased estimates of the slope coefficients (a heterogeneity bias).

In the case of a dynamic data-generating process with slopes homogenous across individual units, Pirotte (1999) showed that the probability limit of the between estimator converges to the long-run effect as the number of individual units tends to infinity while the number of time periods remaining fixed. Pirotte and Mur (2017) extended this result by showing that, in the presence of spatial dependence, the probability limit of the between estimator also converges to the long-run effect but only as the time dimension approaches infinity. The convergence speed increases with heterogeneity of individual effects but slows with spatial dependence. Again, this result was obtained for strictly exogenous explanatory variables, as in Pesaran and Smith (1995).

In the previous studies of tax effort, cross-sectional regressions, after averaging out cyclical fluctuations, could not explain more than half of the remaining ("between") variation, even for the smaller samples used in the earlier studies, which covered only 50–75 countries due to limited data available at that time. Some determinants of tax capacity could not be accounted for in cross-sectional regressions because they were hard to measure (e.g., culture) or too numerous (e.g., indirect determinants). Repeated observations on the same country allow one to account for the impact of those "unobservable" factors, especially if they change slowly over time (or do not change at all). In addition, using more observations leads to efficiency gains (more precise estimates); however, with panel data comes a range of additional assumptions that one needs to use to model the relationships over time. Different assumptions made in specifying the model can lead to dramatic differences in the estimates of tax efforts.

Static ("within") estimators

In contrast to the between estimator, one can focus on changes of these variables over time while sweeping out the effects of time-invariant factors, in particular those that we cannot observe directly. One implementation of such approach in the tax effort studies was the Prais–Winsten transformation used by Leuthold (1991) and Gupta (2007). Alternatively, Stotsky and Woldemariam (1997) estimated country fixed- and random-effects models on a sample of 43

Sub-Saharan countries over 1990–95. Teera and Hudson (2004) estimated country fixed- and random-effects models on a sample of 116 developed and developing countries over the period 1975–98. Le at al. (2012) estimated a pooled OLS model with fixed effects for world regions and time periods using a panel dataset for 110 developing and developed countries over 1994–2009.

However, two kinds of problems arise when this so called “within” approach is pursued through the common techniques of first-differencing or demeaning achieved with the inclusion of country dummies as the so-called “fixed effects.” First, short-term fluctuations of one’s variables—for example, lower revenues from income taxes in times of recession—would be irrelevant for our study of tax effort, which should be considered over the whole business cycle and, like other policies, fully play out only in the medium term.

It has been long suggested in the literature that the within (fixed-effects) estimator of static equations mostly captures short-run effects (Baltagi and Griffin 1984). Using numerical simulations, Pirotte and Mur (2017) show that the probability limit of the within estimator can be associated with the short-run effects only under a medium memory of explanatory variables. However, for explanatory variables with a short memory, the within estimator tends to underestimate the short-run impact. By contrast, a long memory of explanatory variables would cause the fixed-effects model to overestimate the short-run impact.

The second challenge in examining changes over time is that measurement errors— inevitable as our indicators are imperfect proxies for socio-economic processes—may dominate the changes in slowly evolving institutional variables. Under both first-differencing and fixed-effect techniques, this measurement error problem often leads to econometric results that are “unsatisfactory, with ‘too low’ and insignificant coefficients” (Griliches and Hausman 1986, p. 93). In the empirical growth literature, it was suggested that, when the number of time periods is large or the within variation in explanatory variables is small compared to the between variation, the between estimator might represent the optimal balance in trading-off the extent of measurement error bias against heterogeneity bias (Hauk and Wacziarg 2009). Again, this result was derived for strictly exogenous explanatory variables, as in Pesaran and Smith (1995).

The attempts to account for autocorrelation by including the lagged dependent variable as a regressor in the presence of fixed effects can add a downward bias to the estimates when the number of time periods (T) is small (Nickell 1981). Furthermore, slope heterogeneity can cause a separate kind of inconsistency in models with lagged variables, which persist even for large T (Pesaran and Smith 1995).

When the slopes are the same for all countries, one solution to both the measurement errors[4] and autocorrelation of disturbances is to use the instrumental variable estimation proposed under the GMM approach, which allows for a large number of weak instruments—as is often the case. Additionally, using instrumental variables allows parameters to be estimated consistently even when some of our explanatory variables are endogenous.

Under the difference GMM estimator proposed by Arellano and Bond (1991), lagged levels are used as instruments for subsequent first differences. However, this estimator does not perform very well with a small number of time periods, especially when the persistence of past shocks is high and the variation in country effects is large relative to the variation of the residual white-noise disturbances (Blundell and Bond 1998, Kiviet et al. 2017). Under these circumstances, better results can be obtained using the system GMM estimator developed by Blundell and Bond, who combined equations in differences instrumented with lagged levels and equations in levels

[4] When the measurement errors are autocorrelated rather than being white noise, the difference GMM estimator does not eliminate the measurement error bias (Hauk and Wacziarg 2009).

instrumented with lagged first differences. However, the validity of lagged differences as instruments hinges on the assumption that the country effect is unrelated to the first observable first-difference of the dependent variable. When this assumption does not hold—manifested by a rejection of the test of overidentifying restrictions—one is left to try improving on the difference GMM estimator by finding additional instruments.

Monte Carlo simulations performed by Hauk and Wacziarg (2009) for the augmented Solow growth model suggest that even a small amount of measurement error can exacerbate the small sample bias of the difference GMM estimator. Their data-generating process featured country-specific effects that were highly correlated with explanatory variables as well white-noise residuals also correlated with the regressors other than the lagged dependent variable. These simulations suggest that the between estimator outperforms the fixed-effects and difference/system GMM estimators in terms of the average absolute value of the biases across slope coefficient as well as the bias in the coefficient on the lagged dependent variable (the convergence parameter).[5] This finding reinforces the conjectured superiority of the between estimator previously suggested in cases with strictly exogenous explanatory variables. Intuitively, when the variation in the white-noise residuals is small compared to the variance in regressors (including country effects), even a large covariance between residuals and the regressors translates into a small bias. However, the superiority of the between estimator in the case of the augmented Solow model might be due to the inclusion of the lagged dependent variable capturing country effects, while in the case of static equations, the endogeneity bias could be more severe because the country effects would end up in the error term.

[5] In that study, a time period was defined as five years so that the per capita GDP was measured over eight five-year periods, and the dependent variable was lagged by five years on the left-hand side of the regression equation.

Another critical assumption of the difference/system GMM estimators is that errors are correlated only over time but not across countries. Therefore, when relying on these GMM estimators, this paper includes year dummies to remove common time-related shocks from the errors. In future research, I would like to replicate this analysis using estimators robust to error cross-sectional dependence and slope heterogeneity (e.g., Chudik and Pesaran 2015).

Stochastic frontier models with panel data

The stochastic frontier approach estimates an empirical relationship—between production inputs and outputs in its original application—as a conditional average where the total deviation from the regression curve is decomposed into two terms: statistical noise and inefficiency. While both of these terms are unobserved, under certain distributional assumptions, the inefficiency term can be estimated either for the sample as a whole (e.g., representing a group of countries) or for each individual country. In particular, when applied to tax performance, the stochastic frontier approach assumes that the tax effort is bound so that the tax outcome cannot exceed the tax capacity (hence, the notion of the "frontier"). However, if the tax capacity is framed as some welfare optimum level, there is no reason to expect deviations (tax effort) to be one-sided—as in the stochastic frontier analysis. The only possible justification for one-sided deviations could be when the analysis is performed on a subsample of developing countries that are believed to be well under the optimum level.

If these distributional assumptions of the stochastic frontier approach were true, that alone would not affect the consistency of the OLS estimates of the impact of various observable characteristics of each unit (in this case, each country). However, the OLS estimate of the intercept would be biased downward by the expected level of inefficiency for the entire group of units.

The first application of the stochastic frontier analysis to the estimation of the tax effort can be traced to Alfirman's (2003) study of local government revenues in Indonesia.[6] To analyze his panel data, Alfirman used the estimator from Cornwell et al. (1990) allowing for a country-specific, quadratic specification for the deterministic time trend:

$$y_{it} = m(x_{it};\beta) + c^0_i + c^1_i*t + c^2_i*t^2 + v_{it}.$$

The specification in Cornwell et al. (1990) relies on the standard panel data fixed-effects estimator[7] to obtain estimates of the tax effort in each jurisdiction:

$$\tilde{c}_{it} = c^0_i + c^1_i*t + c^2_i*t^2$$

Then, the estimator of the relative tax effort can be obtained as

$$c_{it} = \max_j (\tilde{c}_{jt}) - \tilde{c}_{it}$$

That is, for each year *t* the best-performing jurisdiction is identified, and the tax effort is calculated relative to that benchmark set in that year.

Since the time trend *t* appears in the tax effort component, it cannot also appear as a determinant of tax capacity x_{it}, which would be required if one were to model a drift in tax capacity, m(x). Thus, the Cornwell et al. (1990) specification cannot separate multi-year tax effort initiatives from persistent trends in tax capacity, which is a fundamental issue in this policy arena—and thus cannot be obfuscated by using sophisticated methods.

[6] Other applications can be found in Fenochietto and Pessino (2013), Cyan et al. (2014), Langford and Ohlenburg (2016), Brun and Diakite (2016), and Garg, Goyal, and Pal (2017).

[7] Han et al. (2005) show that a generalized method of moments (GMM) approach is preferable for estimating the models of Cornwell, Schmidt, and Sickles (1990) because it is asymptotically efficient.

However, if we were only concerned about the relative tax effort, it would matter less where we (mis-)specify the common time trend or common cyclical fluctuation. If the common shift in tax capacity were erroneously captured in the tax effort term, it would wash out after subtracting from the top performer:

$$c_{it} = \hat{c}_t - \tilde{c}_{it} \text{ where } \hat{c}_t = \max_j (\tilde{c}_{jt}) \; \forall t.$$

Even when allowing meaningful point estimates, such misspecification still creates issues for inference due to the violation of the IID assumption (due to common shocks and autocorrelation). Furthermore, if the tax capacity trend is not common but specific to each country, it does not wash out in the computation of the relative tax effort.

Pessino and Fenochietto (2010) applied the stochastic frontier approach to the international setting. However, they used the specification of Battese and Coelli (1992) featuring a monotonic time trend (time decay) in tax effort, and this temporal pattern was assumed to be the same for all countries. More specifically, they relied on a maximum-likelihood estimation of a time-varying stochastic frontier model by specifying the country tax effort as $u_{it} = u_i * exp[\eta * (t-T)]$, where T corresponds to the last time period in each panel, η is the decay parameter estimated from the data, and u_i is the time-invariant (or the average over time) country-specific tax effort assumed to have a normal distribution truncated at zero.

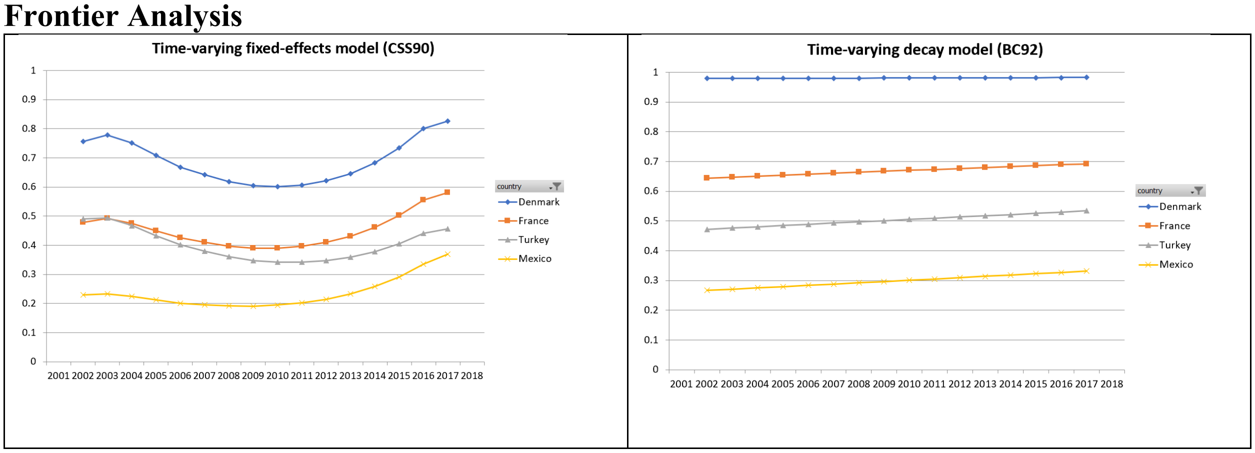
The use of the Battese and Coelli (1992) specification could be justified if the policymakers were more interested in the estimate of the tax effort at the endpoint of the time series (or a country average for the entire time span) rather than the path of tax effort exerted in each previous year. In that case, the assumptions about the effort trajectory would be more of a technical consideration of how to discount the past data points in producing the end-point estimate.

However, despite using the historical data, this estimation approach would not be able to answer two policy-relevant questions: 1) where a country has been in its effort to tap the available tax space and 2) how much further it needs to go.

By contrast, the specification in Cornwell et al. (1990) can be more useful for tracking the tax effort in each individual country over time. It allows tax effort in each country to have one peak (e.g., one tax reform episode) over the years covered by data.

As an illustration, Figure 2 shows the estimated trajectory of the tax effort of selected countries under alternative specifications of stochastic frontier regressions using our dataset. The Battese and Coelli (1992) specification assumes the same monotonic temporal pattern for the tax effort in all countries. The predicted values of the tax effort show exactly that in Figure 2. In particular, this assumption rules out the possibility of countries' relative rankings in terms of their tax effort to switch over time. However, countries' relative rankings do switch over time under the specification of Cornwell et al. (1990), which allows more flexible temporal variation in the tax effort. According to the left panel of Figure 2, Turkey's tax effort was slightly higher than that of France in 2002–03, but it fell significantly behind France in the subsequent years.

Figure 2. Tax Effort Trajectories under Alternatives Specifications of the Stochastic Frontier Analysis



In all of these time-varying stochastic frontier models, the intercept is assumed to be the same across countries, thus generating a misspecification bias in the presence of time-invariant unobservable determinants of tax capacity, which are unrelated to the tax effort but affect the tax outcome. Therefore, the effect of these time-invariant unobservable factors may be captured by the tax effort term, producing biased results. Again, this is a fundamental issue in this policy arena and, unlike for observable country characteristics, it cannot be resolved by making a judgment call about whether a certain factor is under government control.

In the Cornwell et al. (1990) specification, it is more of a philosophical question of whether the time-invariant unobservable factors are erroneously accounted under the country-specific deterministic trend in the tax effort rather than the tax capacity. However, the Battese and Coelli (1992) specification does not allow for a country-specific intercept in either tax capacity or tax effort. Therefore, when the real-world data feature some time-invariant unobservable factors specific to each country, the Battese and Coelli (1992) specification would produce biased estimates of tax effort as a result of this misspecification. Such misspecification is somewhat mitigated in Battese and Coelli's (1995) approach, where the mean of the truncated normal distribution of u_i is expressed as a linear function of some observable characteristics—for example, a corruption perception index in Pessino and Fenochietto (2013).

A handful of more recent studies of tax effort attempted to use four-component stochastic frontier models—called the Generalized True Random-Effects (GTRE) models—allowing for both time-invariant and transient components in the stochastic tax effort in addition to the time-invariant component in the tax capacity specification. Thus, Brun and Diakite (2016) employed the stochastic frontier model of Kumbhakar et al. (2014), which features a random country effect in addition to time-invariant and transient tax efforts. A multi-step method of moments estimator was used first to estimate the standard random-effect panel model and then to decompose its residuals assuming one-sided tax efforts. Canavire-Bacarreza et al. (2021) employed the Bayesian approach to estimate the GTRE model following Tsionas and Kumbhakar (2014). However, even while allowing for transient tax effort, GTRE models assumed all stochastic components to be independent over time periods and thus rule out dynamic adjustment in the tax effort.

In the stochastic frontier analysis, tax capacity is a theoretical concept, which cannot be calculated in practice without knowing the unobserved value of stochastic shocks (v). In theory,

Tax effort = actual/capacity = exp(xb+v+u) / exp(xb+v) = exp(u).

In practice, we do not know the realized value of stochastic shocks (v) that would be required to calculate the denominator. Fortunately, stochastic shocks (v) cancel out from both the numerator and denominator thus allowing one to skip the calculation of the ratio and jump straight to the calculation of the tax effort, *exp(u)*. As a result, the stochastic frontier analysis does not produce an estimate of the tax capacity—only an estimate of tax effort. As the tax effort (u_i) is stochastic, one would not know its actual realized value for any country, only its expected value conditional on the regression residual, which in turn serves as an estimate of the error v+u. Fenochietto and Pessino (2013) attempted to estimate the tax capacity outside the stochastic frontier model by marking up the actual tax-to-GDP ratio to account for the estimated underperformance (u_i).

Dynamic estimators

In a conventional multiplicative form:

Tax Outcome = T/Y = (Tax Capacity) * (Tax Effort),

where T stands for the country's total tax collections and Y stands for its GDP. Then, by taking logs of both sides, we can transform it into an additive form:

$$\log(T/Y) = \log(\text{Tax\_Capacity}) + \log(\text{Tax\_Effort}) =$$

$$= \log(\text{Observed\_Tax\_Capacity}) + \log(\text{Unbserved\_TaxTax\_Capacity}) +$$
$$\log(\text{Observed\_Tax\_Tax\_Effort}) + \log(\text{Unbserved\_Tax\_Effort}) =$$

$$= \log(\text{Observed\_Tax\_Capacity}) + \log(\text{Observed\_Tax\_Tax\_Effort}) + c_i + v_{it} + \eta_i + u_{it}$$

The cross-country relationship between tax capacity and economic development is most likely to be about the potential GDP (i.e., the stage of development), while the within (short-run) coefficient is about the output gap (the denominator effect in the tax-to-GDP ratio and countercyclical policy response). Also, because economic development (*y*) has to be measured per capita in order to capture the level of development, *T* and *Y* in the tax outcome (*T/Y*) should also be expressed per capita to follow the same notation. In our econometric analysis, we also normalize monetary variables by converting them to U.S. dollars.

By separating the level of economic development (*y*) from other variables among the observed country characteristics x_{it}, we can rewrite this additive equation as

$$\log(t/y) = \alpha^*\log(y) + m(x_{it};\beta) + c_i + \eta_i + u_{it} + v_{it} \tag{1}$$

Given the conjectured persistence of the dependent and independent variables used in the tax effort studies (e.g., Gupta 2007), it is reasonable to include the lags of dependent and independent variables among the regressors in the form of an autoregressive distributed lag (ARDL) model. According to Deli et al. (2018), at least in the OECD sample, the Bayesian Information Criterion (BIC) suggests the optimal lag length of 1 for both tax revenues and GDP.[8]

[8] Similarly, the lag length of 1 for both tax and GDP is used in other recent studies of tax buoyancy (e.g., Cornevin et al. 2023; Lagravinese et al. 2020).

Therefore, in this exercise, we assume that autocorrelation can be tackled with the ARDL(1,1) version of equation (1):

$$\log(t/y) = \rho*\log(t_{it-1}/y_{it-1}) + \alpha*\log(y_{it}) + \alpha*\delta*\log(y_{it-1}) + m(x_{it};\beta) + \gamma*m(x_{it-1};\beta) + c_i + \eta_i + u_{it1} + v_{it} \quad (2)$$

Furthermore, a more flexible form can allow for slope heterogeneity:

$$\log(t_{it}/y_{it}) = \rho_i*\log(t_{it-1}/y_{it-1}) + \alpha_i*\log(y_{it}) + \alpha_i*(\delta_i*\log(y_{it-1}) + m(x_{it};\beta_i) + \gamma_i*m(x_{it-1};\beta_i) + c_i + \eta_i + u_{it1} + v_{it}$$

By collecting terms, we can rewrite it in the following Error Correction Model (ECM) form:

$$\Delta\log(t_{it}/y_{it}) = \lambda_i\{\theta_i*\log(y_{it-1}) + \mu_i*m(x_{it-1};\beta_i) - \log(t_{it-1}/y_{it-1})\} + \alpha_i*\Delta\log(y_{it}) + \Delta m(x_{it};\beta_i) + c_i + \eta_i + u_{it1} + v_{it} \quad (3)$$

where

- $\lambda_i = (1-\rho_i)$ is the speed of adjustment,
- $\theta_i = \alpha_i(1+\delta_i)/(1-\rho_i)$ is the long-run relationship between economic development and tax outcome (conjectured to be positive),
- $\beta_i*\mu_i = \beta_i(1+\gamma_i)/(1-\rho_i)$ is the long-run impact of all other observable determinants of tax outcome,
- α_i is the short-run reaction of the tax ratio to the economic shock (positive in the case of countercyclical policy response), and
- β_i is the short-run reaction of the tax ratio to non-economic shocks.

This econometric equation allows one to decompose the observed change in the tax-to-GDP ratio into three distinct driving forces:

- tax gap harvesting: λ_i $\{\theta_i *\log(y_{it-1}) + \mu_i *m(x_{it-1};\beta_i) - \log(t_{it-1}/y_{it-1})\}$,
- countercyclical response: $\alpha_i*\Delta\log(y_{it})$, and
- tax capacity shift: $\Delta m\ (x_{it};\beta)$.

The first component captures the utilization of the available tax space implied by the level of economic development and the existing level of tax effort. It does not appear reasonable to assume that all countries utilize the same fraction λ_i of the available tax space in a given year. Neither would it be reasonable to assume a constant rate of change in the tax effort for any country given the stop-and-go nature of this kind of policy initiatives. It would be more reasonable to assume λ_{it} to vary both over the years and across countries. However, we do not have degrees of freedom to estimate N x T parameters λ_{it}.

When the time dimension T is sufficiently large, additional empirical strategies are possible, in particular those allowing heterogeneity in slopes as opposed to just heterogeneity in intercepts featured in the "within" regressions. When the cross-section dimension (N) is small (e.g., N < 10) and the time series dimension (T) is large, one could treat the set of equations from individual countries as a system of seemingly unrelated regression equations (SURE), thus allowing for general (time-invariant) correlation patterns across the errors in the different cross-section equations. However, if the number of countries (N) were allowed to rise, the application of the SURE approach would involve nuisance parameters whose number would increase at a quadratic rate. This would make SURE not feasible when N were large relative to T as the estimated covariance matrix could not be inverted. Furthermore, the SURE approach is not applicable if the errors are correlated with the regressors.

To deal with these problems, a new approach was proposed by Pesaran (2006) leading to a set of estimators known as common correlated effects (CCE) estimators. This approach aims to eliminate the differential effects of unobserved common factors by filtering the country-specific regressors by means of cross-section (weighted) averages of the dependent variable and the observed country-specific regressors. While Pesaran (2006) initially developed this approach for the case of stationary and exogenous individual-specific regressors and common factors, it was subsequently extended to the case where the unobserved common factors were allowed to follow unit-root processes (Kapetanios et al. 2011).

The CCE approach also offers an empirical framework for estimating an "average" coefficient in the case of heterogeneous panels where each country can have a different slope. This is the essence of the tax capacity concept, which is supposed to capture average or "standard" relationships between countries' characteristics and their tax outcomes (Bahl 1971). The CCE approach offers two alternative estimators for the "average" of the country-specific slope coefficients. One, building on the Mean Group (MG) estimator initially proposed in Pesaran and Smith (1995), is a simple average of the CCE estimators for individual countries, called the "Common Correlated Effects Mean Group" (CCEMG) estimator. The alternative estimator—called the ''Common Correlated Effects Pooled" (CCEP) estimator—is a generalization of the fixed effects estimator, which imposes uniformity of country slopes but now allowing cross-section dependence.

Pesaran (2006) shows that the CCEMG estimator is asymptotically unbiased as $N \rightarrow \infty$ for both T fixed and $T \rightarrow \infty$. While the estimation of country-specific slopes would not be a typical application when the number of countries N is large, it could still be done with a CCE estimator of the country-specific coefficients, which is consistent as N, T $\rightarrow\infty$, jointly, as long as a certain rank condition on the factor loadings is satisfied.

The initial CCE approach allowed for dynamics through autocorrelation both in the common factors (even unit-root processes as shown later in Kapetanios et al. 2011) and in country-specific effects and regressors. However, initially, it did not explicitly model these dynamics by including the lagged values of the dependent variable among the regressors. The assumption of strictly exogenous regressors in Pesaran (2006) not only did not allow the lagged values of the dependent variable among the regressors, but it also ruled out feedback effects from the dependent variable onto the regressors—for example, from the level of taxation to GDP in the context of this paper.

Chudik and Pesaran (2015) extended the CCE approach by introducing lagged dependent variables and allowing regressors to be weakly exogenous. They also considered two bias correction methods to deal with the small T bias of their estimators and found these procedures helpful in mitigating the time series bias but concluded that it could not fully deal with the size distortion unless T was sufficiently large.

Such dynamic CCE estimators open new possibilities for studying tax efforts using data panels with a sufficiently large time dimension (T). For example, for the oldest member countries of OECD, data are available going back to the 1960s. With such data at hand, one could employ the Pooled Mean Group (PMG) estimator, which was initially proposed by Pesaran et al. (1999) for models with lagged dependent variables and heterogeneity of short-run dynamics, but not allowing for error cross-section dependence at that time. With the dynamic CCE extension, it could feature a large degree of heterogeneity, by allowing both the intercepts and the slopes (e.g., tax gap harvesting and countercyclical responses) to vary across countries, both in the short and long run. Thus, one could potentially test whether there is a uniform long-run relationship θ between the tax-to-GDP ratio and the level of economic development.

Table 1. Robustness of Alternative Estimators

Estimator / robust to	**Slope heterogeneity**	**Measurement error**	**Endogeneity**	**Autocorrelation**	**Cross-section dependence**
Between	Yes	Somewhat	No	Yes	Yes
Fixed effects	No	No	No	No	No
Error Components Frontier	No	No	No	No	No
Difference GMM	No	No*	Yes	Yes	No
CCE-MG	Yes	?	No	Yes	Yes
DCCE-MG	Yes	?	Yes	Yes	Yes

Notes: * When the measurement errors are autocorrelated rather than being white noise, the difference GMM estimator does not eliminate the measurement error bias.

Table 1 summarizes the preceding discussion of the robustness of the reviewed econometric methods to various empirical challenges of the tax effort studies. The next section illustrates relative performance of these methods using a large panel of countries.

Estimation Results

For our empirical analyses, I have put together a panel dataset for 155 countries covering the period 2003–17. Due to data availability, the panel is unbalanced with the time dimension varying between 10–16 years. In addition to the tax-to-GDP ratio used as our dependent variable, our dataset also includes three major determinants of tax outcomes:

- GDP per capita in U.S. dollars,
- Imports plus exports as a share of GDP, and
- Value added of the agriculture sector as a share of GDP.

We begin by presenting the results of estimating alternative versions of the static regression in Table 2. The between estimator is supposed to be most appropriate for capturing long-run relationships, although in our case the number of time periods, 10–16 depending on the country, might not be sufficiently long. The estimates suggest a rather high (0.308) impact of economic development (proxied with the GDP per capita) on the tax-to-GDP ratio. Note that all variables are expressed in logarithms. This estimate is also quite close to 0.29 obtained by Williamson (1961) in the between regression without any other covariates for 33 countries over 1951–57.

The between estimator is supposed to be robust to most empirical challenges, in particular slope heterogeneity, but not to the endogeneity of regressors. While we can expect a reverse causality from the level of taxation to economic development, the sign of that relationship should be negative so that, if anything, the estimated impact of economic development on the tax-to-GDP ratio would have a downward bias.

Table 2. Estimation of Static Regressions

	Between Estimator	**Within Estimator**	**Time-invariant Error Components Frontier**	**Time-varying Error Components Frontier (BC92)**	**CCEMG**	**CCEP**
GDP per capita	0.308***	0.135***	0.122***	0.028**	0.236***	0.219
	(0.061)	(0.019)	(0.011)	(0.012)	(0.058)	(0.145)
Imports plus exports	0.113	0.199***	0.191***	0.165***	0.174***	0.161*
as % of GDP	(0.099)	(0.024)	(0.021)	(0.020)	(0.050)	(0.069)
Value added of the agriculture sector as %	0.248***	-0.041*	-0.039*	-0.026	-0.103*	-0.153*
of GDP	(0.077)	(0.022)	(0.022)	(0.070)	(0.046)	(0.065)
Constant	-0.256		2.760***	3.446***		
	(0.674)		(0.111)	(0.122)		
N [countries]	2401	2401	2401	2401	2401	2401
	[155]	[155]	[155]	[155]	[155]	[155]
R^2	0.179	0.045			0.828	0.713
σ_u			0.678	0.705		
σ_v			0.154	0.150		
λ			4.414	4.694		

Notes: Dependent variable is the logarithm of the tax-to-GDP ratio. All independent variables are in logarithms. Time dummies are included in all specifications. Robust standard errors are provided in parentheses. * statistically significant at 10 percent; ** statistically significant at 5 percent; *** statistically significant at 1 percent.

In all other specifications of the static regression, the estimated impact of economic development on the tax-to-GDP ratio was much lower, around 0.15. The impact of trade openness (proxied with the sum of exports and imports as a share of GDP) on the tax-to-GDP ratio appears to be most consistent across various specifications of the static regression. Reliance on agriculture has a positive and statistically significant impact in the between regression but a negative impact in all other specifications of static regressions. Note that imposing uniformity of country slopes with the CCEP estimator produces coefficients similar to the "averages" of the respective country-specific slope coefficients under the unrestricted CCEMG estimator—but with larger standard errors. This suggests that slop coefficients might indeed vary across countries.

Table 3. Estimation of the dynamic model

	Between Estimator	**Within Estimator**	**Difference GMM**
Lag of tax-to-GDP ratio	-0.018***	-0.272***	-0.166**
	(0.003)	(0.028)	(0.073)
Lag of GDP per capita	-0.001	0.023	0.050
	(0.003)	(0.015)	(0.046)
Lag of Imports plus exports as % of GDP	-0.0004	0.102***	0.205*
	(0.004)	(0.025)	(0.116)
Lag of Value added of the agriculture sector as % of GDP	0.002	-0.044*	-0.082
	(0.003)	(0.039)	(0.113)
Δ GDP per capita	0.048	0.141	0.137*
	(0.073)	(0.118)	(0.082)
Δ Imports plus exports as % of GDP	0.178**	0.117***	0.177***
	(0.083)	(0.038)	(0.061)
Δ Value added of the agriculture sector as % of GDP	-0.137**	-0.096*	-0.132*
	(0.083)	(0.055)	(0.073)
Constant	0.062**		
	(0.026)		
R^2	0.298	0.161	
Sargan test (p-value)			0.369

Notes: Dependent variable is the log of the tax-to-GDP ratio. All independent variables in logarithms. Time dummies are included in all specifications. Robust standard errors are provided in parentheses. * statistically significant at 10 percent; ** statistically significant at 5 percent; *** statistically significant at 1 percent.

Next, we report the results of estimating the dynamic ARDL(1,1) model (equation 3), which allows one to disentangle short-run responses of the tax-to-GDP ratio from convergence to the long-run equilibrium. In all three specifications of the dynamic regression, the coefficient at the lagged tax-to-GDP ratio has a quite low absolute value, ranging from 0.02 for the between estimator to 0.27 for the within estimator. Such low values of the error correction rate suggest high persistence of tax-to GDP ratios. Only the difference-GMM specification produces statistically significant estimates for the long-run relationship (e.g., coefficients at lagged values of variables). From the difference-GMM estimates we can calculate the long-run reaction of the tax-to-GDP ratio to changes in GDP per capita as 0.05 / 0.166 = 0.303. This value is quite close to 0.308 produced by the between estimator as reported in Table 1.

Conclusions

This paper attempts to retrace the evolution of the framing and assumptions underlying decades of empirical studies of the tax performance of different countries. I argue that the increased sophistication of empirical methods cannot get around the fundamental problem of imperfect information, as the observed tax outcome is a joint product of unobservable tax effort and tax capacity. Therefore, such assessments inevitably come down to a judgment call about what is under government control. While not uncommon in practice, such judgment calls must be based on the specific policy application at hand. Thus, if the goal is to track the tax effort over time, it would not be helpful to employ an econometric model assuming a time-invariant tax effort (e.g., Schmidt and Sickles 1984) or assuming the same temporal pattern of the tax effort for all countries (e.g., Battese and Coelli 1992).

Furthermore, in the real-life data, the bulk of the within (over time) variation in tax-to-GDP ratios appears to be driven by the economic cycle. We explore the suitability of the Error Correction Models (ECM) for disentangling the long-run relationship between tax capacity and economic development from short-run policy adjustments in response to economic cycles and other transient shocks. While conceptually appealing, a practical implementation of the ECM approach has to overcome many empirical challenges: slowly evolving dependent and independent variables, endogeneity of GDP and other determinants of tax outcomes, cross-section dependences (CSD) due to unobservable shocks common to all countries, heterogeneity of country slopes, measurement errors, etc. While recent advances in dynamic panel data analysis offer empirical strategies that can potentially mitigate some of these challenges, their applicability has to be determined in each case. In particular, this determination is guided by the size of the time (T) and cross section (N) dimensions of the dataset, which in turn is determined by whose tax effort we are measuring and against which comparator countries. This paper illustrates these points by estimating alternative econometric models on a panel dataset for 155 countries covering the period 2003–17.

Based on our review of the literature and our empirical exercise, a number of lessons can be drawn that can guide the evaluation of the tax effort of a country:

- Include all observed determinants of tax outcomes (both tax capacity and effort) to explain as much variation in the tax-to-GDP ratio as possible (long-run relationship).
- If observed, include time-invariant country characteristics among the regressors.
- Make a judgement call to predict tax capacity by disregarding the estimated impact of observable determinants of tax effort.
- Account for common shocks (e.g., time effects) and test for cross-section dependences in the residuals.
- Allow for some slope heterogeneity, at least short run.
- Make adjustments for short-run responses when assessing the tax effort trajectory.

Finally, while I believe these lessons can improve on the current practice of measuring tax effort, I am cognizant of the limitations of such benchmarking studies. As pointed out by Bird (2014), such benchmarking exercises in themselves “neither supply clear explanations of the underlying problems nor insights that are likely to prove helpful in resolving those problems: problems are not solutions and possibilities are not certainties.” Nevertheless, I believe that measuring tax effort can be a useful input to policy making as it can draw one’s attention to an area of possible concern that might require further investigation using more granular data and methods. Thus, one can start such troubleshooting by applying the tax effort methodology to separate components of the tax system, such as direct versus indirect taxes.

References

Advisory Commission on Intergovernmental Relations. (1962). Measures of State and Local Fiscal Capacity and Tax Effort. October 1962, M-16. Available at: library.unt.edu/gpo/acir/Reports/information/M-16.pdf (accessed July 18, 2022).

Afxentiou, P. C. (1980). "Government Revenue and Expenditure and The Developing Countries: A Survey." *The Developing Economies*, 18: 116–139. doi.org/10.1111/j.1746-1049.1980.tb00407.x

Alfirman, L. (2003). "Estimating Stochastic Frontier Tax Potential: Can Indonesian Local Governments Increase Tax Revenues Under Decentralization?" Department of Economics, University of Colorado at Boulder, Colorado. Working Paper No. 03-19.

Bahl, R. W. (1971). "A regression approach to tax effort and tax ratio analysis." *International Monetary Fund Staff Papers*, XVIII (no. 3),: 570–612.

Arellano, M. and S. Bond. (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2): 277.

Baltagi, B. H., and J. M. Griffin. (1984). "Short and Long Run Effects in Pooled Models." *International Economic Review*, 25(3): 631–45. doi.org/10.2307/2526223.

Battese, G.E., Coelli, T.J. (1992). "Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India." *J Prod Anal,* 3: 153–169 doi.org/10.1007/BF00158774

Battese, G.E., Coelli, T.J. (1995). "A model for technical inefficiency effects in a stochastic frontier production function for panel data." *Empirical Economics,* 20: 325–332.

Belinga, V., D. Benedek, R. A. de Mooij, J. Norregaard. (2014). "Tax Buoyancy in OECD Countries." *IMF Working Papers* 2014/110, International Monetary Fund.

Bird, R. (2002). "Why Tax Corporations?" *Bulletin for International Fiscal Documentation,* 56: 194–203.

Bird, R. M. (2011). "Tax System Change and the Impact of Tax Research." Chapter in Emilio Albi & Jorge Martinez-Vazquez (ed.), The Elgar Guide to Tax Systems, chapter 12, Edward Elgar Publishing.

Bird, R. M. (2014). "Foreign Advice and Tax Policy in Developing Countries," in Richard M. Bird and Jorge Martinez-Vazquez, eds., Taxation and Development: The Weakest Link (Essays in Honor of Roy Bahl), chapter 4, Edward Elgar Publishing.

Blundell, R., and Bond, S. (1998). "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115.

Brun, J.-F., & Diakité, M. (2016). "Tax Potential and tax effort: An empirical estimation for non-resource tax revenue and VAT's revenue." *Études et documents*, 10. CERDI. cerdi.org/production/ show/id/1814/type_produ ction_id/1

Chelliah, R. J. (1971). “Trends in Taxation in Developing Countries.” *International Monetary Fund Staff Papers, 18: 254–331*.

Cornwell, C., P. *Schmidt* and R.C. *Sickles.* (*1990*). “Production Frontiers with Time-Series Variation in Efficiency Levels.” *Journal of Econometrics*, 46: 185–200.

Chudik, A., and M. H. Pesaran. (2015). “Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors.” *Journal of Econometrics*, 188: 393–420.

Canavire-Bacarreza, G., M. C. Deza, O. Manzano, and A. Puerta. (2021). “Revisiting tax effort in emerging markets.” *Public Finance Review*, 49(6): 845, 1552–7530.

Cornevin, A., J. S. Corrales, J. P. Angel. (2023). “A Deep Dive into Tax Buoyancy: Comparing Estimation Techniques in a Large Heterogeneous Panel,” IMF Working Paper No. 23/71, International Monetary Fund, Washington, D.C.

Cyan, M., J. Martinez-Vazquez, and V., Violeta. (2014). “New approaches to measuring tax effort.” In Richard M. Bird and Jorge Martinez-Vazquez, eds., *Taxation and Development: The Weakest Link (Essays in Honor of Roy Bahl),* 27–68.

Dahlby, B. and L. S. Wilson. (1994). “Fiscal capacity, tax effort, and optimal equalization grants.” *Canadian Journal of Economics*, 27: 657–672.

Dalamagas, B, Leventides, J, Palaios, P, Tantos, S. (2020). Revising the conventional tax-effort principle. *Scott J Polit Econ*. 2020; 67: 272– 299. doi.org/10.1111/sjpe.12239

Davoodi, H. and Grigorian, D. (2007). “Tax Potential vs. Tax Effort: A Cross-Country Analysis of Armenia’s Stubbornly Low Tax Collection,” WP/07/106 (Washington: International Monetary Fund).

Deli, Yota A.G. Rodriguez, I. Kostarakos, and P. Varthalitis. (2018). “Dynamic tax revenue buoyancy estimates for a panel of OECD countries,” Papers WP592, Economic and Social Research Institute (ESRI).

Dudine, P. and J. Tovar Jalles. (2018). “How Buoyant Is the Tax System? New Evidence from a Large Heterogeneous Panel,” *Journal of International Development*, 30(6): 961–991.

Fenochietto, R. and P. Carola. (2013). Understanding Countries’ Tax Effort. Washington, DC, USA: International Monetary Fund.

Garg, S., Goyal, A., and Pal, R. (2017). “Why tax effort falls short of tax capacity in Indian states.” *Public Finance Review*, *45*(2): 232–259. doi.org/10.1177/10911 42115 623855

Gaspar, V., L. Jaramillo, and P. Wingender. (2016). Tax Capacity and Growth: Is there a Tipping Point? IMF Working Papers.

Greene. (2005). Reconsidering in The Measurement of Productive Efficiency and Productivity Growth.” Chapter in H. O. Fried, C. A. Knox Lovell and S. S. Schmidt (editors), Oxford University Press, Oxford, 92–250.

Griliches, Zvi and Jerry A. Hausman (1986). Errors in variables in panel data. Journal of Econometrics 31(1): 93-118.

Gupta, Abhijit Sen. (2007). Determinants of Tax Revenue Efforts in Developing Countries Working paper. No. 7-184. International Monetary Fund. doi.org/10.5089/9781451867480.001

Gupta, S., Jalles, J.T. & Liu, J. (2022). "Tax Buoyancy in Sub-Saharan Africa and its Determinants." *Int Tax Public Finance* 29: 890–921 doi.org/10.1007/s10797-021-09694-x.

Han, C., L. Orea, and P. Schmidt. (2005). "Estimation of a panel data model with parametric temporal variation in individual effects." *Journal of Econometrics* 126: 241–267.

Hinrichs, Harley H. 1965, "Determinants of Government Revenue Shares Among Less-Developed Countries." *The Economic Journal*, 75(299): 546–56. doi.org/10.2307/2228952.

Hauk, W.R., Wacziarg, R. (2009). "A Monte Carlo study of growth regressions." *J Econ Growth*, 14: 103–147. doi.org/10.1007/s10887-009-9040-3

Kapetanios G, Pesaran MH, Yamagata T. (2011). "Panels with non-stationary multifactor error structures." *Journal of Econometrics*, 160(2): 326–348.

Keen, Michael. (2012). "Taxation and Development; Again" IMF Working Papers 12/220, International Monetary Fund.

Kimmel, Lewis H. (1949). "Our Tax Burdens and Taxable Capacity." *The Annals of the American Academy of Political and Social Science*, 266: 152–60.

Kiviet, Jan; Pleus, Milan; Poldermans, Rutger. (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models.

Econometrics, ISSN 2225–1146, MDPI, Basel, 5(1): 1–54. doi.org/10.3390/econometrics5010014

Kumbhakar S. C., Lien G. and Hardaker J. B. (2014). "Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming", *Journal of Productivity Analysis*, 41(2), 321–37.

Lagravinese, Raffaele, Paolo Liberati, and Agnese Sacchi. (2020). "Tax buoyancy in OECD countries: New empirical evidence," *Journal of Macroeconomics*, Elsevier, vol. 63(c).

Langford, B., & Ohlenburg, T. (2016). Tax revenue potential and effort. International Growth Centre, Working Paper, S-43202-UGA-1.

Leuthold, Jane H. (1991). "Tax shares in developing economies: A panel study," Journal of. Development Economics, 35: 173–85.

Le, Tuan Minh, Blanca-Moreno Dodson and Nihal Bayraktar. (2012). "Tax Capacity and Tax Effort: Extended Cross-Country Analysis 1994–2009," Policy Research Working Paper 6252, The World Bank.

Lotz, Jørgen R. and Elliott R. Morss. (1967). "Measuring 'Tax Effort' in Developing. Countries," Staff Papers (International Monetary Fund), 14(3): 478–499.

Newlyn, W.T. (1985). "Measuring Tax Effort in Developing Countries.' *Journal of Development Studies*, 21(3): 390–405.

Nickell, S. J. (1981). “Biases in dynamic models with fixed effects.” *Econometrica* 49: 1417–1426.

OECD. (2018). Revenue Statistics, OECD Publishing, Paris. doi.org/10.1787/rev_stats-2018-en.

Peacock, A.T. and Wiseman, J. (1961). *The growth of public expenditure in the United Kingdom*. London: National Bureau of Economic Research.

Pirotte, A. (1999). “Convergence of the static estimation toward the long run effects of dynamic panel data models.” *Economics Letters*, 63, 151–158.

Pirotte, A. and J. Mur. (2017). “Neglected dynamics and spatial dependence on panel data: consequences for convergence of the usual static model estimators.” *Spatial Economic Analysis*, 12 (2–3): 202–229. doi: 10.1080/17421772.2016.1232839.

Pesaran, M.H. and R. Smith. (1995). “Estimating Lon-Run Relationship from Dynamic Heterogeneous Panels.” *Journal of Econometrics*, 68: 79–113.

Pesaran, M. H., Y. Shin and R. P. Smith. (1999). “Pooled Mean Group Estimation of Dynamic Heterogeneous Panels.” *Journal of the American Statistical Association,* 94(446): 621–634. doi: 10.1080/01621459.1999.10474156

Pessino, C., and R. Fenochietto. (2010). “Determining Countries’ Tax Effort.” *Hacienda Pública Española/Revista de Economía Pública*, 195: 61–68.

Schmidt, P., and R. C. Sickles. (1984), “Production Frontiers and Panel Data.” *Journal of Business & Economic Statistics*, 2: 367–74.

Stotsky, J. G. and A. Woldemariam. (1997). “Tax effort in Sub-Saharan Africa.” IMF Working Paper No: WP/97/107 (September).

Tait, A. A., W. L. M. Grätz and B. J. Eichengreen. (1979). “International Comparisons of Taxation for Selected Developing Countries, 1972-76,” *Staff Papers (International Monetary Fund)*, 26(1): 123–156.

Teera, J. M. and J. Hudson. (2004). “Tax performance: a comparative study.” *Journal of International Development*, 16(6): 785–802.

Wang, H. J. and P. Schmidt. (2002.) “One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels.” *Journal of Productivity Analysis,* 18: 129–144. doi.org/10.1023/A:1016565719882

Weiss, S. J. (1969). “Factors Affecting the Government Revenue Share in Less Developed Countries.” *Social and Economic Studies*, 18(4): 348–64.

Williamson, J. G. (1961). “Public Expenditure and Revenue: An International Comparison,” The Manchester School of Economic and Social Studies, XXIX: 43-56.