Effort fiscal en Afrique subsaharienne : les résultats d'une nouvelle base de données

Emilie Caldeira, Ali Compaore, Alou Adessé Dama, Mario Mansour, Grégoire Rota-Graziosi

Tax effort in Sub-Saharan African countries: Evidence from a new dataset

Emilie Caldeira
Associate Professor in Economics, School of Economics, Université Clermont Auvergne, CNRS, CERDI, F-63000 Clermont-Ferrand, France.

Ali Compaore
PhD Student in Economics, School of Economics, Université Clermont Auvergne, CNRS, CERDI, F-63000 Clermont-Ferrand, France.

Alou Adessé Dama
PhD Student in Economics, School of Economics, Université Clermont Auvergne, CNRS, CERDI, F-63000 Clermont-Ferrand, France.

Mario Mansour
International Monetary Fund, Washington, DC, USA.

Grégoire Rota-Graziosi
Professor in Economics, School of Economics, Université Clermont Auvergne, CNRS, CERDI, F-63000 Clermont-Ferrand, France.

Corresponding author: Grégoire Rota-Graziosi

This work was supported by the LABEX IDGM+ (ANR-10-LABX-14-01) within the program "Investissements d'Avenir" operated by the French National Research Agency (ANR).

Abstract

This paper proposes (i) a new database of tax revenue for 42 Sub-Saharan African countries (SSA) over the period 1980-2015, (ii) an estimate of tax effort for these countries, and (iii) some replication analyses of previous tax effort estimations. The database results from statistical information of the African Department of the International Monetary Fund (IMF). In particular, it allows distinguishing tax revenue from the natural resource sector from the other economic sectors. SSA countries collected on average 13.2 percent of GDP in non-resource tax revenue over the studied period and their average estimated tax effort is 0.57. In other words, SSA countries could raise 23.2 percent of GDP in non-resource taxes if they fully used their potential. In line with previous analyses, we find that countries' stage of development measured by per-capita income, financial development, and trade openness are important factors improving tax revenue in the region, while natural resource endowment and the importance of the agriculture sector reduce unambiguously the non-resource tax-to-GDP ratio. Finally, beyond the originality of the database itself and the empirical results, this work participates explicitly to the replication principle given its online development with R software (https://data.cerdi.uca.fr/taxeffort/).

Keywords
Tax effort; Sub-Saharan Africa; Stochastic frontier analysis.

JEL Codes : H20 ; O11 ; O23.

I. Introduction

Since the Addis Ababa Conference in July 2015, Domestic Revenue Mobilization (DRM) became one of the main tools of financing Sustainable Development Goals (SDGs). DRM is now a well-discussed topic to address the issue of economic development (see e.g. Besley and Persson, 2014) and is a privileged tool for donors and international and regional institutions (African Development Bank, International Monetary Fund, World Bank, European Union Commission).

In this paper, we propose (i) an update and complete version of the tax revenue dataset published in Mansour (2014), (ii) an estimate of tax effort for these countries, and (iii) some replication analyses of previous tax effort estimations by Gupta (2007) and Fenechietto and Pessino (2013). The database covers 42 Sub-Saharan African (SSA) countries over the period 1980-2015. It results from statistical information collected in the African Department of the International Monetary Fund (IMF)—most of which is included in public IMF documents. We distinguish tax revenue from the natural resource extractive industry, from those from other economic sectors.

Tax revenue excluding natural resources is on average 13.2 percent of GDP. The average estimated total tax effort is 0.57. In other words, SSA countries could raise on average 23 percent of GDP of non-resource taxes if they fully utilized their potential. We decompose the total tax effort score into time-varying and persistent tax effort and conclude that the total tax effort score is mainly driven by time-varying factors. Moreover, consistent with previous literature, we find that countries' stage of development measured by per-capita income, financial development, and trade openness are important factors improving tax revenue in the region, while natural resource endowment and the importance of the agriculture sector reduce unambiguously the non-resource tax-to-GDP ratio. Regarding the replication exercise, the estimations broadly confirm previous analyses such as Fenechietto and Pessino (2013). However, our verification test failed to replicate the exact results of Gupta (2007) in terms of robust coefficients and significance of the variables, which might be caused from the use of less detailed data than we provide here.

Our analysis contributes to the existing literature by providing a new estimation of SSA countries' total tax efforts and their composition. We decompose tax effort in terms of direct taxation (Corporate Income Tax, CIT, and Personal Income Tax, PIT). In addition, beyond the originality of the database itself and the empirical results, our works participates explicitly to the replication principle given its online application developed with R-Shiny. The need of replicability appears highly relevant for tax effort analysis given the primacy of DRM in the agenda of developing countries, donors, and international organizations. The database is dynamic and is hosted on a webpage that allows users to interact with the data and generate new analytical results, including quick descriptive statistics and running alternative specifications of regressions.

The rest of the paper is organized as follows: Section 2 presents the dataset; Section 3 briefly reviews the literature on the determining factors of tax effort in developing countries; Section 4 describes the empirical methodology and variables. Section 5 presents and discusses the results; section 6 proposes a replication analysis of Gupta (2007) and Fenechietto and Pessino (2013), and section 7 concludes.

II. Tax revenue dataset for Sub-Saharan Africa over 1980-2015

The study of tax policy in developing countries has long been constrained by the availability and the quality of detailed relevant data. Moreover, extractive industries have played and still play a crucial role in the economic development of SSA countries. More than half of these countries are resource dependent, that is natural resources represent 25 percent or more of total country's exports. Tax revenues from this sector are usually large and at high risk of being taken out of the source country through various licit or illicit channels, including: generous tax incentives provided in mining or petroleum codes and other laws; aggressive tax planning such as the use of thin capitalization, trade mispricing, or plain tax evasion; and double taxation agreements that do not always protect appropriately source countries' taxation rights.

We provide here an updated version of the tax revenue dataset published in Mansour (2014), which covered the period 1980-2010 for 41 countries (see https://data.cerdi.uca.fr/taxeffort/).[1] It participates to recent efforts to better apprehend tax revenues in Africa, in particular the revenue statistics in Africa from the OECD, which cover 26 countries in its last release[2] and the Government Revenue Dataset initiated by the International Centre for Tax and Development (ICTD) and updated by UNU-WIDER.[3] There are three advantages that our dataset provides relative to these two alternatives. First coverage for SSA countries is generally broader, and deeper for each of the tax series. Second, the definition of variable is consistent across all countries,[4] Finally, the isolation of resource revenue from non-resource (tax) revenue allows for a better understanding of the interaction of these two fundamentally different (economically) sources. [5]

Distinguishing resource from non-resource revenue is highly relevant to understand countries' tax effort. For instance, Bornhorst, Gupta, and Thornton (2009), Crivelli and Gupta (2014), and James (2015) emphasize a crowding-out effect between resource revenue and non-resource tax revenue: an increase of the former reduces the latter. McGuirk (2013) explains this effect through the strategy of the government to remain in power by reducing its accountability or equivalently the tax pressure. Caldeira et al. (2020) provides an alternative explanation of the negative relationship between resource and non-resource tax revenue in terms of an inter-ministerial tax competition: the Minister in charge of Mining and Petroleum can tax partly the same base than the Minister of Finance. The inter-ministerial tax competition reduces total tax revenue and deteriorates the economic development of these countries by favoring a concentration of the economic activity in the extractive industry.

The dataset covers 9 tax series and 42 SSA countries over the period 1980-2015. The series are: 1. Total Taxes; 2. Trade Taxes; 3. Indirect Taxes; 4. VAT with a decomposition for some countries between domestic VAT, VAT collected at the border, and VAT refunds; 5. Excises; 6. Direct Taxes; 7. Personal Income Tax; 8. Corporate Income Tax with additional information for some countries concerning CIT from extractive industry; 9. Other tax revenues.

In order to isolate the impact of resource revenue on the tax effort, the database reports revenue from extractive activities separately and irrespective of the policy tool used to raise them. As such, resource revenues include royalties and other fees, dividends, and bonuses from extractive activities, the government share of production sharing agreements, and (importantly) corporate income taxes. The latter is included because it is similar in design to production sharing, and hence subject to the same

---

[1] See https://ferdi.fr/publications/a-tax-revenue-dataset-for-sub-saharan-africa-1980-2010

[2] See https://www.oecd.org/ctp/revenue-statistics-in-africa-2617653x.htm

[3] See https://www.wider.unu.edu/project/government-revenue-dataset

[4] The ICTD database, now Government Revenue Dataset (https://www.wider.unu.edu/project/government-revenue-dataset), combines revenue data primarily from OECD revenue statistics, IMF staff reports' statistical tables, and IMF GFS. This produces asymmetries in the definition of resource revenues. For instance, ICTD reports no resource tax revenue for Australia and Canada, and only aggregated corporate income tax (CIT), which include profit taxes from the resource sector if the source is OECD of GFS. These asymmetries are less important in SSA countries since the primary source for ICTD for these countries is IMF staff reports, and Keen and Mansour (2009)—and both report a different concept of resource revenue. For instance, the average resource revenue-to-GDP ratio during 1980-2015 in ICTD is 8.16 percent, which is close to the 8.6 percent in our database. However, the average CIT ratio in ICTD is 1.82 percent over the same period, slightly higher than the 1.7 percent in our database—possibly due to the fact that the CIT revenue for SSA taken from OECD revenue statistics for Africa includes some resource revenue.

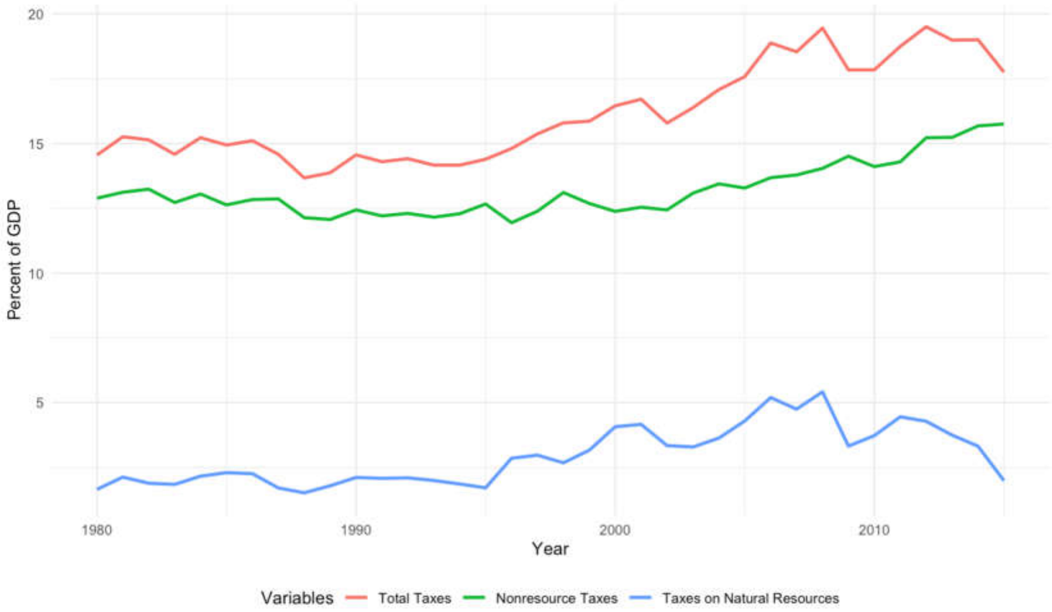
[5] The OECD statistics do not report resource revenues unless they are accounted for as corporate taxes. This may not be an issue in OECD countries, where oil revenue is derived primarily through the tax system. However, in SSA countries, production sharing agreements and turnover-based royalties are prominent.

extent to volatility in commodity prices. However, resource revenues do not include non-refundable VAT on inputs, which we were not able to identify separately—presumably, this is not very important in aggregate given that extractive companies often seek and obtain an exemption from VAT on input, knowing that VAT refund mechanisms in SSA countries are not very effective.

Figure 1 displays the evolution of the average tax revenue in Africa, in percent of GDP.[6] Note that the volatility of commodity prices explains a large share of revenue variations over the period. However, an improvement of non- resource tax revenue is perceptible since 2000: This revenue stagnated around 12.5 percent of GDP from 1980 to 2000 and reached 15 percent by 2015.

Figure 2 highlights the tremendous change in the structure of non-resource tax revenue over the studied period. The reduction of tariff duties (equal to trade tax revenue in the dataset) was offset by an increase in the revenue of taxes on goods and services, which results from the introduction of the Value Added Tax (VAT).

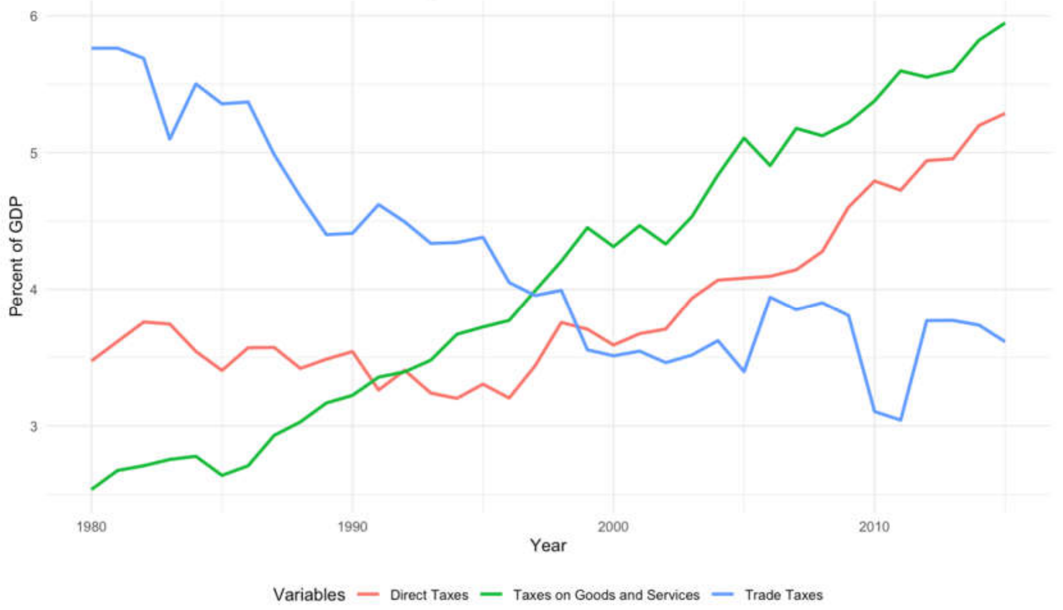
Figure 1. Evolution of revenue in SSA: 1980-2015 (percent of GDP)



Sources: Tax revenue Dataset for SSA and authors' own calculations.

[6] Figures 1, 2 and many others are created from the dedicated website on the page entitled Graphics.

Figure 2. Variation of average direct, indirect, and trade tax revenue over the period 1980-2015 (percent of GDP).

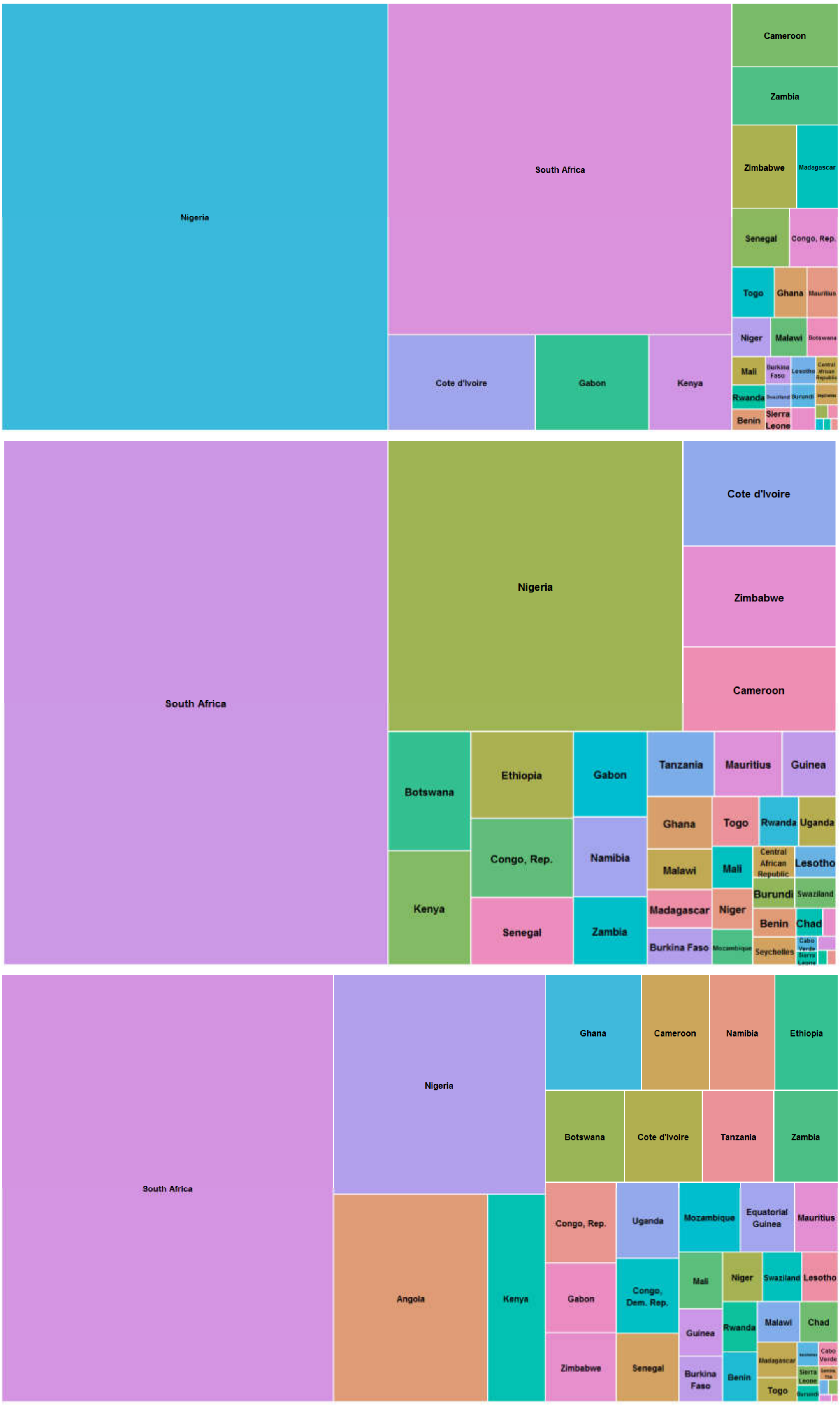


Sources: Tax revenue Dataset for SSA and authors' own calculations.

Figure 3 shows total tax revenue collected across SSA countries in 1980, 1990 and 2015. Total tax revenue amounted to 91 billion USD (constant 2010 USD) in 1980[7], 117 billion USD in 1990, and 259 billion USD in 2015. South Africa and Nigeria are the main contributors with respectively 32 percent of total revenue (including resource revenue) in 1980, 39 percent in 1990, and 40 percent in 2015 for South Africa, and 46 percent in 1980, 36.2 percent in 1990 and only 13 percent in 2015 for Nigeria. This highlights the main role of the natural resource sector in aggregate for SSA and the sharp decrease of total tax revenue collected in Nigeria. The variation of total tax revenue over the period 2000-2015 displays contrasting results: while tax revenue decreases by 49 percent in Nigeria, they increase significantly in Mozambique by 615 percent reaching 3.1 billon USD in 2015, in Rwanda by 403 percent (1,2 billon USD), in Chad by 376 percent (1.1 billon USD), and in Ghana by 341 percent (8 billon USD).

[7] Some countries are missing in 1980.

Figure 3. Total tax revenue in 1980, 1990 and 2015 (constant 2010 USD).



Sources: Tax revenue Dataset for SSA and authors' own calculations.

III. Brief literature review on tax effort

DRM would be a more reliable and sustainable source of financing than its domestic alternative (debt, seignioriage) or international financial flows (i.e. remittances, official development assistance and foreign direct investment). Hence, a non-negligible literature has investigated how countries, specifically developing countries, which face important financial constraints, can levy more domestic resource to finance development and wean themselves from aid.

Several empirical analyses study the macroeconomic and institutional driving factors of countries tax-to-GDP ratio. A first generation of empirical works establish that agriculture, mining (resource rent), and share of external debt in total debt are significant determinants of countries tax ratios. Agriculture share, which is still important in least developed economies, is negatively associated with the level of tax revenue (Chelliah et al., 1975; Leuthold, 1991; Tanzi, 1992; Stotsky and WoldeMariam, 1997), while mining and external debt are positively associated with tax revenues (Cheliiah et al., 1975; Tait et al., 1979 and Tanzi, 1992).

However, the relationship between natural resource sector and tax revenue remains controversial. Indeed, in line with the resource curse debate, recent studies point out a negative association between resource rent and government tax revenue. For instance, Belinga et al. (2017) highlight a crowding-out effect of resource revenue on non-resource revenue for a panel of 30 resource-rich countries over the period 1992-2012. Natural resource boom is associated with less incentive in tax collection. Brun et al. (2014) consider tax revenue from the natural resource sector as an explanatory variable of the non-resource tax effort. They establish a negative effect of the former on tax revenue potential.

A second generation of empirical works outlined the pivotal role of inflation, institutional quality, education, political stability, external aid, and financial development in addition to the previous economic factors (Tanzi and Davoodi, 1997, Grigorian and Davoodi, 2007, Gupta, 2007, Gordon and Li, 2009, Clist and Morrissey, 2011, Fenochietto and Pessino, 2013, Feger and Asafu-Adjaye, 2014).

We provide here a new dataset and focuses on the effort of countries to raise tax revenue. We define tax effort as the extent to which the actual tax revenue collected is from the maximum level of tax revenue given their characteristics. These characteristics correspond to the determinant variables previously studied in the literature, which are mainly: the level of development, trade openness, the size of the agricultural sector, natural resource rent, and financial development. Given these characteristics, we compute for each country potential tax revenue. Tax effort results then from the comparison between potential tax revenue and actual collected tax revenue. Closer they are, the greater is the tax effort. We do not study the details of countries' tax code, nor the organization of their revenue administration or authorities.

Our approach is then purely economic, since it does not rely on countries' tax system[8] but only on economic characteristics. It allows some international comparisons among countries, which share similar economic features. This analysis could be then complemented by some tax policy and revenue administration diagnostics. Indeed, differences in tax effort across countries may result from some distinctions in existing taxes, their statutory rates, their respective tax bases, tax expenditures, the efficiency of revenue administration (organization, IT technology, the number of tax inspectors, and even their remunerations' modalities).[9] Other determinants such as tax morale, the ethno-linguistic

[8] Gillitzer and Slemrod (2015) define tax system as the combination of tax policy (tax laws) and revenue administration.

[9] See for instance Caldeira and Rota-Graziosi (2019) for a detailed analysis of relative tax revenue performance between Benin and Togo, which begins with a tax effort analysis of these two countries and is completed by a review of countries' tax systems.

fragmentation of the countries, political regimes (presidential or parliamentary), inflation rate… are variables, which may be added in our empirical assessment of tax effort.[10]

Tax effort is complementary to tax gap analysis,[11] which measure the difference between expected revenue and collected one. The tax gap approach is legal and microeconomic, while the tax effort analysis relies only on macroeconomic data. Indeed, the computation of expected revenue differs from potential revenue in the tax effort analysis, since the former requires the use of statutory tax rates, tax base's definition, and eventually some assumptions on the behavior of consumers and producers. The tax gap has usually two components: the policy gap and the compliance or administrative gap. The former, roughly equivalent to the cost of tax expenditures, results directly from a political decision to reduce the tax burden of the investor or the consumer. This policy aims at stimulating investments or at protecting the poorest. For instance, investment or sectorial (Petroleum or Mining) codes may provide tax exemptions or reduce tax rates, which would reduce tax revenues. Similarly, one of the main justifications of VAT exemptions or reduced rates is to protect the poorest consumer. For instance, developing countries use to exempt completely the agricultural sector and some SSA countries exempt even from VAT the importations of some foodstuff such as rice and wheat. The rationale is the assumption of a tax incidence close to one, i.e. such exemptions would reduce the price for households.[12] The second element of the gap is the administrative or compliance gap. This gap corresponds to the capacity of tax and customs administration or tax authorities to enforce current tax laws and to the compliance behavior of firms and individuals to pay their taxes.

IV. Empirical methodology: The Stochastic Frontier Analysis

The literature proposes several approaches to capture countries' tax effort. The usual indicator to compare countries' tax effort is the tax-to-GDP ratio. However, Stotsky and WoldeMariam (1997) point out that this simple approach is inappropriate to measure the taxable capacity since not all taxes are explicitly linked to income and to its distribution. Using panel data on 42 Sub-Saharan African countries, during the period 1990-1995, they propose another measure of tax effort consisting of the ratio of the actual to the predicted tax share in GDP. They find that countries with high tax shares tend to have a relatively high tax effort index, even though some disparities remain across countries. Following a similar approach, Gupta (2007) computes the tax effort for 105 developing countries over 25 years but clearly recognize some shortcomings related to this approach.

Cyan et al. (2013) propose a method of estimating tax effort that closely relies on the revenue adequacy approach (This method consists in looking at the deviations between what a country would like to raise in tax revenues – as revealed by the structural choice of the level of public expenditures – and its actual tax revenue level. This approach corroborates the empirical evidence that changes in expenditures induce changes in tax levels (see Baicker and Skinner, 2011). Recently, Yohou and Goujon (2017) proposed a Vulnerability-Adjusted Tax Effort Index (VATEI) for a sample of 120 developing countries over 1990-2012. Their approach consists in building the tax effort as the residual of a standard panel regression model (random effects model) of non-resource tax ratio on the economic vulnerabilities and human asset indices, in addition to the usual determinants of tax revenue. This adjusted tax effort index is assumed to measure the willingness and capacity of governments to collect tax beyond the structural factors.

An alternative and increasingly used approach to capture countries' tax effort is the Stochastic Frontier Analysis (SFA) method, which has been followed by: Alfirman (2003), Fenochietto and Pessino (2013),

[10] The devoted website to this paper allows adding any variable and running new estimates of tax efforts (see https://data.cerdi.uca.fr/taxeffort/).

[11] Several countries provide some tax gap analyses. One of the most exhaustive exercise is the VAT Gap work done by the EU commission. The VAT gap amounts to 137 billion Euro in 2017, or equivalently to 11.2 percent of total VAT revenue.

[12] We do not discuss here the efficiency or the equity of these tax expenditures.

Langford and Ohlenburg (2015), Brun and Diakite (2016). Aigner, Lowell, and Schmidt (1977 and Meeusen and van Den Broeck (1977) developed SFA approach to model firms' production behavior. The rationale is that economic agents cannot exceed an "ideal frontier" of production, which is the optimal level of output given the limited endowment of inputs. In our context, the tax frontier refers to the tax capacity, which is the maximum potential tax revenue, given a country's institutional, demographic, and economic features, while the tax effort is the actual revenue in relation to the frontier. Hence, the closer a country is to that frontier, the greater is its tax effort.

The stochastic frontier approach encompasses parametric and non-parametric models. Data Envelopment Analysis (Charnes *et al*, 1997) and the Free Disposal Hull (Deprins, Simar, and Tulkens, 1984) are the two main and increasingly popular methods used for non-parametric stochastic frontier models. These methods use linear optimization programs to construct the efficiency curve. They display the advantage that no restrictive assumptions on the production function are necessary (except the standard convexity assumption). However, they remain sensitive to random variations in data and measurement errors. Any variation between production units is therefore likely to be interpreted as inefficiency. Furthermore, the inefficiencies estimated by these models are very sensitive to variations in the sample, to the heterogeneity between the units and to the presence of outliers (S

Regarding parametric models in panel data analysis, they are single output-based and categorized into five groups: (i) time-invariant technical inefficiency models (Pitt and Lee, 1981; Schmidt and Sickles, 1984); (ii) time-varying technical inefficiency models Cornwell, Schmidt, Sickles (1990), Kumbhakar (1990), Battese and Coelli (1992), Lee and Schmidt (1993), and Kumbhakar and Wang (2005); (iii) models that separate firm heterogeneity from inefficiency (Greene, 2005; Wang and Ho, 2010); (iv) models distinguishing persistent and time-varying inefficiency (Kumbhakar and Heshmati, 1995); (v) and finally models separating firm effects, persistent inefficiency, and time-varying inefficiency (Colombi *et al*., 2014;. Kumbhakar *et al*., 2014; Filippini and Greene, 2016). In panel data, such models offer the possibility for richer specifications, deal with stochastic noise, and allow testing hypotheses (Hjalmarsson *et al*. 1996; Odeck, 2007).

We follow a parametric approach to estimate the tax effort since we focus on a single output: the total non-resource tax-to-GDP ratio.[13] More precisely, we use the model that separates the error into four components: Generalized True Random Effects model (GTRE). This model was introduced by Colombi *et al*. (2014), Kumbahkar *et al*. (2014), and Tsionas and Kumbahkar (2014). It presents several advantages: (i) it takes into account random shocks; (ii) it is robust to the presence of heterogeneity within the panel; (iii) it allows distinguishing country heterogeneity, and persistent and time-varying factors affecting countries' tax effort. Persistent (i.e. structural) factors are for instance colonial history, culture, geography, the economic structure of the country, which have long-lasting influence on the tax effort. The time-varying factors are both country- and time-specific. They include tax policy, tax administration performance, natural resources discoveries, and commodity price cycles. For example, countries might improve their tax administration performance by clamping down on tax evasion, training their tax officers, or using more sophisticated tax tools. Also, countries' tax effort might change following discovery of natural resources or a boom in commodity prices.

Different methods are proposed in the literature to estimate the parameters of the GTRE model: Colombi *et al*. (2014) use a maximum likelihood estimator (MLE); Kumbahakar *et al*. (2014) propose a multi-step procedure; Tsionas and Kumbahakar (2014) develop a Bayesian approach; Filipini and Greene (2016) use a simulated maximum likelihood approach.

We estimate the following model:

$$\tau_{i,t} = \alpha + f\left(X_{i,t}, \beta\right) + \mu_i + \nu_{i,t} - \eta_i - \varphi_{i,t} \qquad \text{(eq. 1)}$$

[13] We exclude natural resource revenue; which variations are mainly driven by commodities' prices.

The dependent variable, $\tau_{i,t}$, represents the logarithm of the total non-resource tax-to-GDP, the subscripts i and t denote respectively country and time dimensions and $X_{i,t}$ is a vector of covariates explaining countries tax ratio. $\eta_i > 0$ and $\varphi_{i,t} > 0$ are the persistent and time-varying inefficiencies respectively, while μ_i and $\nu_{i,t}$ represent the random effects and the stochastic noise respectively.[14]

Starting from hypotheses on the distribution of the four errors, the MLE approach of Colombi *et al.* (2014) makes it possible to obtain a form of the log-likelihood. Indeed, assuming that $\nu_{i,t}$ is independent and identically distributed (iid) with a normal distribution and $\varphi_{i,t}$ is iid with a half-normal distribution, the error $\varepsilon_{i,t} = \nu_{i,t} - \varphi_{i,t}$ has an asymmetric normal distribution with parameters $\lambda = \sigma_\nu/\sigma_\varphi$ and $\sigma = \sigma_\nu{}^2 + \sigma_\varphi{}^2$. Similarly, the error $\psi_i = \mu_i - \eta_i$ has an asymmetric normal distribution with parameters $\lambda = \sigma_\mu/\sigma_\eta$ and $\sigma = \sigma_\mu{}^2 + \sigma_\eta{}^2$. The two-term error $\varepsilon_i = \varepsilon_{i,t} + \psi_i$ is the sum of two asymmetric normal distributions and then admits a known density. It is therefore possible to define the function of the log-likelihood and to deduce from it the MLE of the parameters. However, the complexity of the likelihood function, which, in his form, involves T_i integrations, makes very hard the implementation in practice.[15] Hence, Filippini and Greene (2016) proposes a computation method based on Butler and Moffitt (1982) formulation to simplify the log-likelihood function and subsequently estimates the MLE using a simulation-based optimization. With the same assumptions on the parameters as before, the idea is to obtain the conditional density $f(\varepsilon_i/\psi_i)$, which is defined on ψ_i. Unlike to the previous case, the manipulation, then, involves only one integration. In order to simplify the implementation of the estimation, we use the multi-step procedure of Kumbhakar *et al.* (2014).[16] The model based on equation 1 is estimated in three stages:

In stage 1, a standard random-effect based regression is used to estimate $\hat{\beta}$ and predicts the values of $\varepsilon_{i,t}$ and ψ_i. We estimate the following equation:

$$\tau_{i,t} = \alpha^* + f\left(X_{i,t}, \beta\right) + \varepsilon_{i,t} + \psi_i \qquad \text{(eq. 2)}$$

Where $\alpha^* = \alpha - E(\eta_i) - E(\varphi_{it})$, $\psi_i = \mu_i - \eta_i + E(\eta_i)$, $\varepsilon_{i,t} = \nu_{i,t} - \varphi_{i,t} + E(\varphi_{i,t})$, $E(\eta_i) = \sqrt{\frac{2}{\pi\sigma_\eta}}$, and $E(\varphi_{i,t}) = \sqrt{\frac{2}{\pi\sigma_\varphi}}$.

In stage 2, by performing a standard stochastic frontier technique, the time-varying tax inefficiency $\varphi_{i,t}$ is estimated using the predicted values of $\varepsilon_{i,t}$ from the first stage. Following Battese and Coelli (1988), this procedure gives the prediction of the time-varying tax effort $\exp(-\varphi_{i,t}|\varepsilon_{i,t})$.

In stage 3, we estimate the persistent tax inefficiency component, denoted by η_i by performing a stochastic frontier model as in the previous stage. The persistent tax effort is then defined by $\exp(-\eta_i)$.

Finally, the overall tax effort is obtained by the product of the time-varying tax effort and the persistent tax effort.

Considering the relevant literature on the determinants of domestic resource mobilization, we identify the following driving factors of government tax revenue and consider them as inputs $X_{i,t}$:

---

[14] We use the logarithm of non-resource tax-to-GDP as dependent. The predictor variables, except the real GDP per capita, are scaled to unit *i.e.* in percent of GDP and not in logarithm. By doing so, we do not assume implicitly the functional form linking the output to the inputs. The log-log form is the most used in the stochastic frontier literature. Note that our results remain robust to the use of the log-log form.

[15] For these authors a direct optimization of the log-likelihood of the model appears complex.

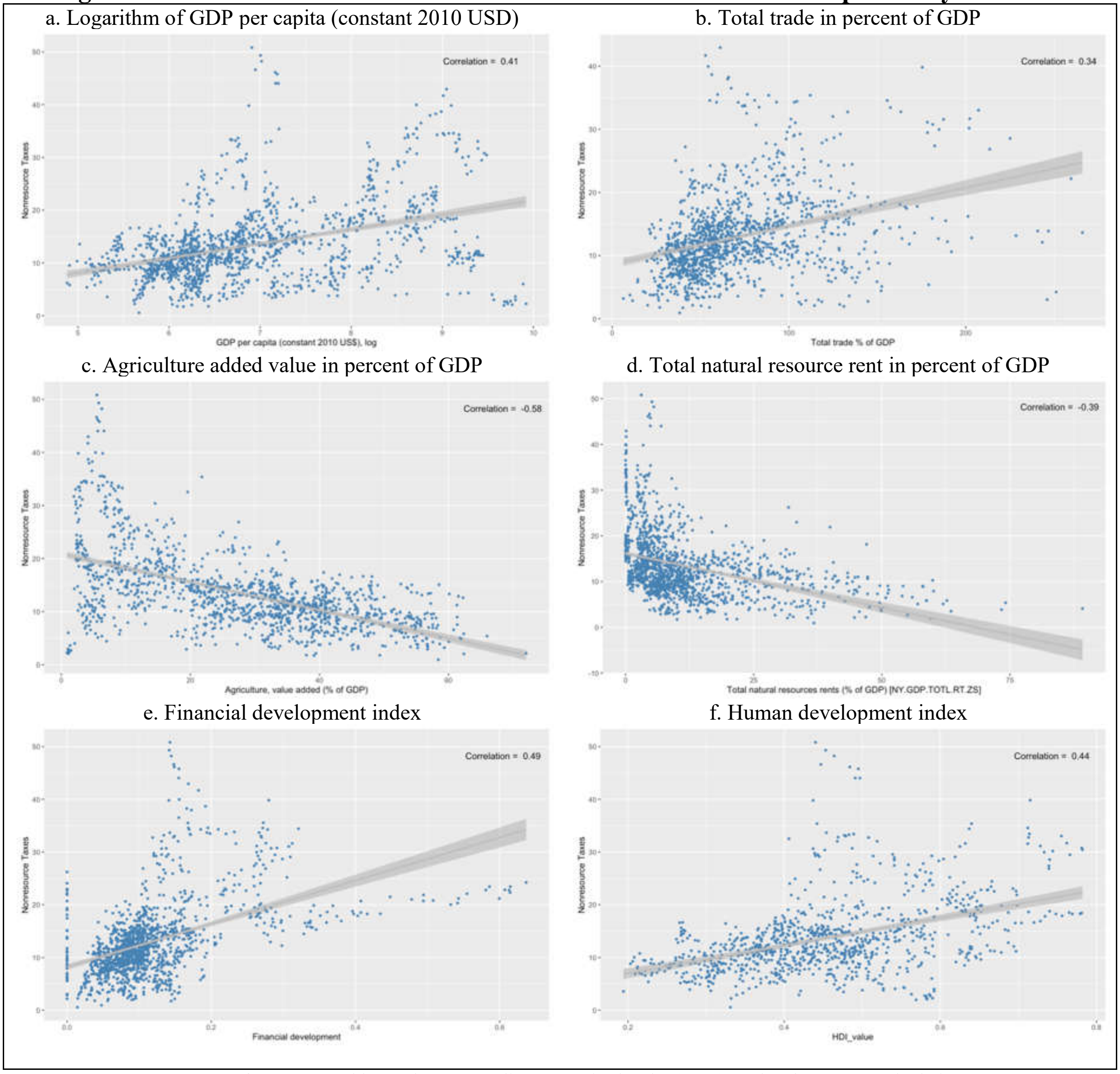
[16] Kumbhakar *et al.* (2015)

1. The level of development: Countries' tax capacity depends on their level of economic development assessed through the level of real GDP per capita. High income countries should raise more tax revenue than developing countries, since they have more efficient tax administration, better institutions, and a higher demand for public goods and services (see Lotz and Morss, 1967; Pessino and Fenechietto, 2010; Crivelli and Gupta, 2014).
2. Trade openness: Trade liberalization policies implemented in most developing countries starting in the early 1970s and stretching well into the 1990s have substantially increased trade volume in these countries. Therefore, trade openness expressed as total trade (value of imports and exports) as a share of GDP is expected to positively influence tax revenue through households' consumption and domestic corporate profits. This reinforces the role of customs administration in collecting taxes, both the external tariff and domestic taxes, on imported goods (see Stotsky and WoldeMariam, 1997; Pessino and Fenechietto, 2010; Gnangnon and Brun, 2018 among others).
3. Agriculture value-added as a percent of GDP: Agriculture is often largely tax exempt in SSA countries from income taxes and other production taxes, and is frequently either tax exempt or out-of-scope of VAT. A first argument in favor of this treatment is that the sector is dominated by small and medium-size farmers that are scattered across geography, and hence hard to tax; and even if tax can be effectively levied, such farmers cannot be significant contributors to tax revenues.[17] A second explanation is the political will to reduce foodstuff prices in order to limit the risk of malnutrition or even famine. This assumes a tax incidence close to one, which means an almost perfect competitive market. However, the tax (VAT mainly) exemptions of foodstuffs involves the inability of farmers to deduce the VAT paid on their inputs. The collected VAT on the intermediary consumptions of farmer can generate some revenue and may have then an ambiguous effect of total tax revenue.
4. Natural resource rent as a percent of GDP: The negative effect of natural resource rent on tax revenue is widely evidenced in the literature. Natural resource endowment is associated with lower tax revenue (Sachs and Warner, 2001; Eltony, 2002; Belinga *et al*., 2007). During commodity prices upswings, governments in resources-rich countries have less incentive to mobilize other tax revenues; resource rent crowds-out tax revenue (Bornhorst et al., 2009, Crivelli and Gupta, 2014, James, 2015) or an inter-ministerial tax competition occurs (Caldeira et al., 2020).
5. Financial development: Financial development may favor higher tax collection (see Gordon and Li, 2009). Combined with improved access to credit, it allows individuals and companies to finance profitable projects and improve the national information system on economic activities—hence, favor tax collection. On the other hand, in a presence of ineffective financial system, firms could successfully evade tax payment by conducting business in cash, which is harder for tax administrations to monitor.
6. Aid received: There is no consensus on the effect of aid on taxes. In a review of several papers studying the relationship between taxation and aid, Morrissey (2015, p.102) stated "the safest conclusion is that we do not have a robust evidence". The study of the relationship raises questions of accountability of governments, good governance and most importantly the repayment contract that comes with aid.

Tables A2 and A3 provide descriptive statistics and more details on variables definition and source. Figure 4 displays scatter plot of total non-resource tax revenue for each of the explanatory variables. These graphs tend to confirm the expected relationships.

[17] The improvement in technologies for farming, including in SSA, and the increase of large farming firms over the past two decades, weaken such arguments. Nevertheless, countries have been very slow in rethinking the taxation of agriculture.

Figure 4. Correlation between total non-resource tax revenue and explanatory variables



Sources: Tax revenue Dataset for SSA, World Development Indicators, Svirydzenka (2016), and authors' own calculations.

V. Results

Table 1 displays the three-stage estimation results.[18] Dependent variables in column [1] to column [4] are respectively total non-resource tax revenue, total income taxes, corporate income tax, and personal income tax. In line with previous studies (Stotsky and WoldeMariam, 1997; Gordon and Li, 2009; Pessino and Fenochietto, 2010; Crivelli and Gupta, 2014) all the variables have the expected sign and are statistically significant. The coefficients associated with the level of development (Log of GDP per capita) and trade openness are positive and significant at the one-percent level. More precisely, a one percentage increase in per-capita GDP is associated with an increase in the total tax revenue by 0.3 percentage points. Similarly, an increase of one percent in total trade-to-GDP ratio is associated with a rise in DRM of 0.2 percent. Agriculture and natural resources sectors harm tax revenue collection. The coefficients associated with these variables are all negative and statistically significant and are consistent with previous analyses (Stotsky and WoldeMariam, 1997; Sachs and Warner, 2001; Eltony, 2002; Brun *et al.*, 2014; Belinga *et al.*, 2017). These results also hold mostly for tax revenue subcomponents (columns [2]-[4]).

For the rest of the analysis (stages 2 and 3), we consider the total non-resource tax revenue as the dependent variable (column [1]). Panel A and Panel B of Table 1 report the second and third stages. We then deduce the time-varying and persistent tax effort scores, and compute the total tax effort (Panel C). The higher is the tax effort score, the closer is the country to the "frontier". Table A4 presents the summary of tax effort for columns [2]-[4].

[18] We exclude Zimbabwe from the sample in all regressions.

Table 1: The three-stage estimation results

	[1]	[2]	[3]	[4]
Variables	NRTAX	Direct	CIT	PIT
Log GDP per cap. (Cst. 2010 USD) $_{(-1)}$	0.265***	0.407***	0.331***	0.410***
	(0.0301)	(0.0450)	(0.0596)	(0.0695)
Total trade (%GDP) $_{(-1)}$	0.002***	0.005***	0.008***	0.004***
	(0.0004)	(0.0006)	(0.0010)	(0.0011)
Agriculture value added (%GDP) $_{(-1)}$	-0.003**	-0.007***	0.002	-0.022***
	(0.0014)	(0.0021)	(0.0029)	(0.0033)
Total natural resource rent (%GDP) $_{(-1)}$	-0.002*	-0.002	-0.001	-0.008**
	(0.0014)	(0.0020)	(0.0029)	(0.0035)
Constant	0.579**	-1.832***	-2.622***	-2.265***
	(0.2299)	(0.3435)	(0.4537)	(0.5285)
Observations	1165	1,165	1,086	1,081
R-squared	0.163	0.240	0.142	0.183
Number of countries	39	39	38	38

Panel A: Stage 2 - estimation of the time-varying tax inefficiency (stochastic frontier)

				Number of obs.		1165
				Wald chi2(1)		317.84
Log likelihood = 104.87				Prob > chi2		0.0000
	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
frontier (one)	0.215	0.012	17.830	0.000	0.191	0.239
usigmas (_cons)	-2.597	0.106	-24.520	0.000	-2.804	-2.389
vsigma (_cons)	-3.759	0.100	-37.690	0.000	-3.954	-3.563

Panel B: Stage 3 - estimation of the persistent tax inefficiency (stochastic frontier)

				Number of obs.		1165
				Wald chi2(1)		1400.00
Log likelihood = -371.66				Prob > chi2		0.0000
	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
frontier (one)	0.446	0.012	37.420	0.000	0.423	0.470
usigmas (_cons)	-1.275	0.060	-21.340	0.000	-1.392	-1.158
vsigma (_cons)	-3.790	0.116	-32.630	0.000	-4.018	-3.563

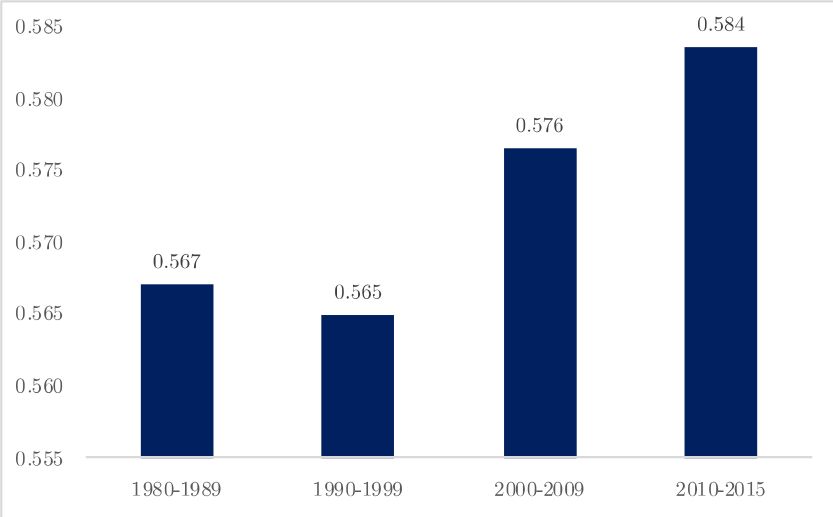
Panel C: Summary of tax effort estimation results

	Obs.	Mean	Std. Dev.	Min	Max
Time-varying tax effort	1165	0.817	0.092	0.318	0.967
Persistent tax effort	1165	0.698	0.167	0.101	0.942
Total tax effort	1165	0.572	0.153	0.057	0.847

Notes: Significance: *** p<0.01, ** p<0.05, * p<0.1; Standard errors in parentheses. NRTAX: Total non-resource tax revenue; Direct: Total income tax revenue; CIT: Corporate income tax revenue; PIT: Personal income tax revenue.

The full sample average stands at 0.817 and 0.698 for time-varying and persistent tax effort, respectively over 1980-2015. The average total tax effort score is equal to 0.572,[19] suggesting that SSA countries mobilize 57 percent of their tax potential. In other words, given their economic features, SSA countries would raise on average 23.16 percent of GDP of non-resource taxes if they fully used their tax potential, rather than the actual 13.22 percent.[20] Furthermore, it is worth underscoring that time-varying factors account for only 36 percent of the total tax effort.[21] Thus, SSA countries would gain significant additional tax revenue by addressing issues related to time-varying factors. In the sample, the minimum tax effort score is 0.057 (Equatorial Guinea in 2011) and the maximum is 0.847 (Burundi in 1998). Note that the tax effort has improved slightly over the period (Figure 5)—from 0.57 during 1980-1989 to 0.58 during the most recent period. An important result is that the number of countries that have improved their tax effort over time is significantly higher than those for which the tax effort has declined (Figure 6).

Figure 5. Evolution of tax effort over time



Sources: Tax revenue Dataset for SSA and authors' own calculations.

Figure 6 shows the evolution of countries tax effort. Most of SSA countries[22] have improved their tax effort over time, particularly in Uganda, Sierra Leone, and Senegal.

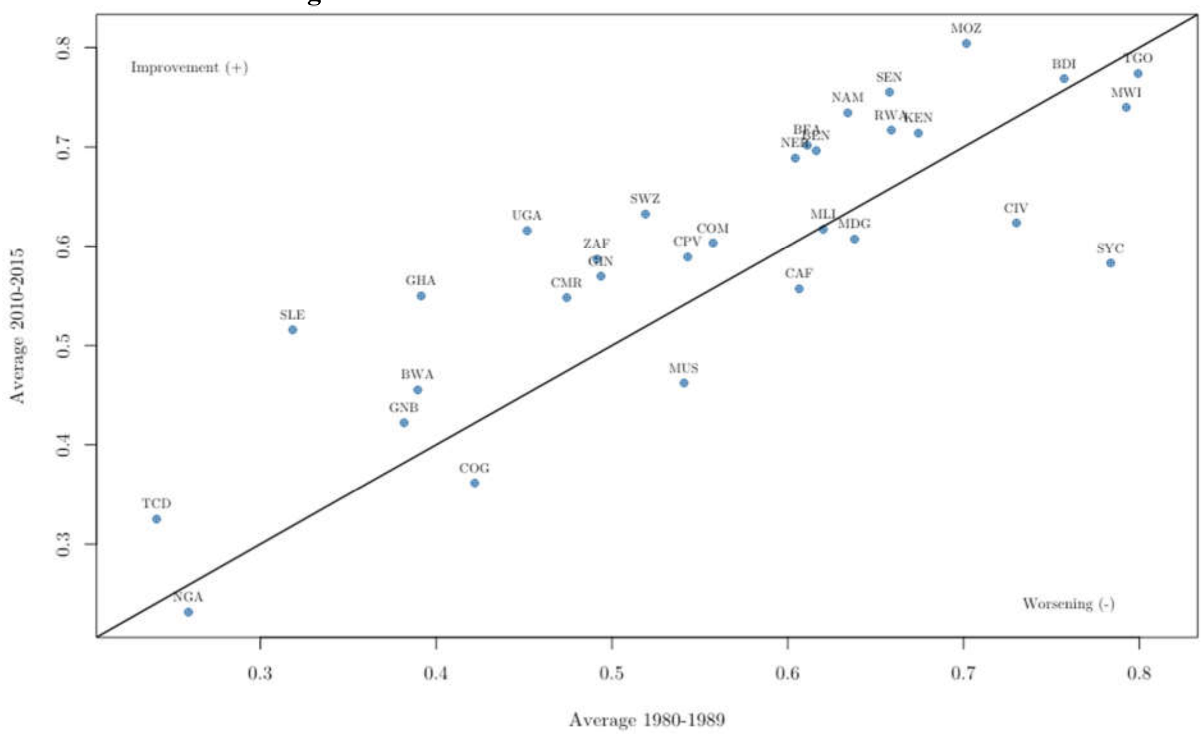
[19] Recall that the total tax effort is obtained by the product of the time-varying tax effort and the persistent tax effort.

[20] On average, total non-resource taxes stand at 13.2% in SSA countries over the period. See table A2.

[21] Since the total tax effort is obtained by a product of the time-varying tax effort and the persistent tax effort components, taking the natural logarithm (Log-linearization) allows us to compute the percentage of each component in the total score.

[22] We display the analysis for 29 countries given the availability of the data for the whole period in particular in 1980.

Figure 6. Evolution of countries tax effort over time



Sources: Tax revenue Dataset for SSA and authors' own calculations.

Table 2 provides a country ranking over different sub-periods (1980-1989, 1990-1999, 2000-2009, and 2010-2015) based on total tax effort scores. Focusing on the last sub-period (i.e. 2010-2015), it emerges that Mozambique, Togo, Burundi, Senegal, and Gambia are top performers with a tax effort score of 0.804, 0.774, 0.769, and 0.755, and 0.742 respectively, while the five poor performers over the sub-period are Congo Republic (0.361), Chad (0.325), Gabon (0.319), Nigeria (0.232) and Equatorial Guinea (0.071). The non-resource tax to GDP has a decreasing trend in Equatorial Guinea, Congo Republic and Gabon. These countries rely a lot and increasingly on revenue from natural resources. As for Nigeria and Chad, they have generally a non-resource tax below 7 percent of GDP over the period 1980-2015.

The poverty level in Mozambique and Burundi explains paradoxically their performance in terms of tax effort. Table 1 highlights the crucial role of GDP per capita to determine the tax revenue to GDP ratio. The coefficient of this variable is not only highly significant, but its level is more important. Togo (as Benin) has a substantial transit activity with Nigeria. Given the Nigerian trade policy to foster domestic production especially (for instance, rice and wax fabrics) or for other considerations (such as environmental and security reason for second-hand cars), some goods subject to high tariff rates are imported in Togo and then smuggled to Nigeria. These importations raise revenue in particular in terms of tariffs and VAT (even if these goods are not consumed in Togo).

Natural resource endowment, especially oil, reduces significantly the computed tax effort (see Figure and 7a). The worst performers are resource-rich SSA countries (Nigeria, Equatorial Guinea) given the crowding out effect (Bornhorst et al., 2009, Crivelli and Gupta, 2014, James, 2015) or the inter-ministerial tax competition weakening the institution in charge of tax policy (Caldeira et al., 2020).

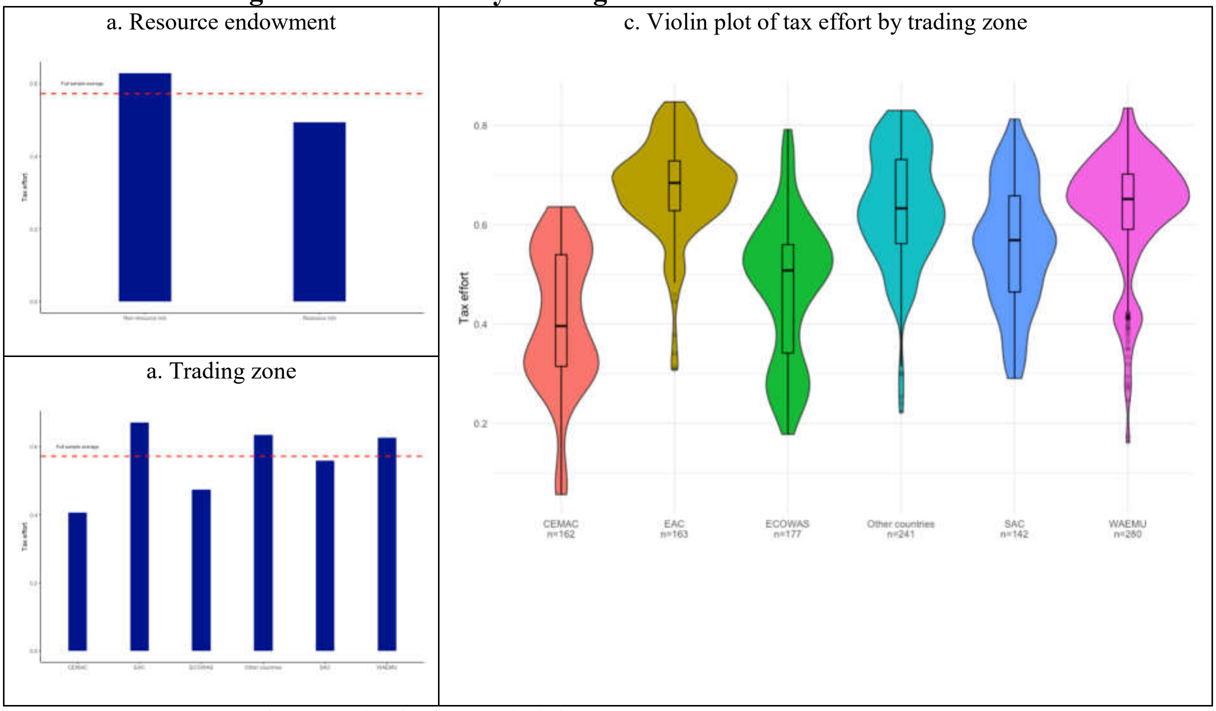
Table 2: Full sample tax effort-based ranking (baseline specification)

Country	1980-1989		1990-1999		2000-2009		2010-2015		
	TE	Rank	TE	Rank	TE	Rank	TE	Rank	Average tax ratio
Angola	-	-	0.265	34	0.279	36	-	-	8.20
Benin	0.616	14	0.627	12	0.691	8	0.696	12	14.67
Botswana	0.390	28	0.361	30	0.411	30	0.455	31	20.96
Burkina Faso	0.611	15	0.660	10	0.689	9	0.702	11	13.15
Burundi	0.757	5	0.806	1	0.792	1	0.769	3	13.40
Cabo Verde	0.543	19	0.605	16	0.619	19	0.589	22	18.71
Cameroon	0.474	24	0.482	25	0.532	26	0.548	28	12.74
Central African Republic	0.607	16	0.560	21	0.569	23	0.557	26	7.79
Chad	0.241	32	0.312	31	0.314	32	0.325	34	5.99
Comoros	0.557	18	0.605	15	0.593	21	0.603	21	11.47
Congo, Dem. Rep.	-	-	-	-	-	-	-	-	9.14
Congo, Rep.	0.422	26	0.375	29	0.307	34	0.361	33	10.87
Cote d'Ivoire	0.730	6	0.682	9	0.645	13	0.624	15	14.02
Equatorial Guinea	-	-	-	-	0.082	37	0.071	37	2.79
Ethiopia	-	-	-	-	-	-	0.724	8	12.01
Gabon	-	-	-	-	0.310	33	0.319	35	11.58
Gambia, The	-	-	-	-	0.735	3	0.742	5	14.93
Ghana	0.391	27	0.480	26	0.534	25	0.550	27	14.82
Guinea	0.494	22	0.422	28	0.521	27	0.570	25	13.38
Guinea-Bissau	0.382	29	0.303	32	0.380	31	0.422	32	8.15
Kenya	0.674	8	0.704	6	0.702	7	0.714	10	15.61
Lesotho	0.808	1	-	-	-	-	-	-	40.97
Madagascar	0.638	11	0.574	19	0.626	18	0.607	19	9.80
Malawi	0.793	3	0.752	3	0.686	10	0.740	6	14.02
Mali	0.620	13	0.585	17	0.634	17	0.617	17	10.85
Mauritius	0.541	20	0.492	24	0.461	29	0.462	30	18.28
Mozambique	0.702	7	0.685	8	0.663	12	0.804	1	19.95
Namibia	0.634	12	0.717	5	0.708	5	0.734	7	31.13
Niger	0.604	17	0.546	22	0.640	14	0.689	13	12.59
Nigeria	0.259	31	0.269	33	0.298	35	0.232	36	4.87
Rwanda	0.659	9	0.623	14	0.703	6	0.717	9	13.70
Sao Tome and Principe	-	-	-	-	-	-	-	-	15.40
Senegal	0.658	10	0.689	7	0.739	2	0.755	4	19.04
Seychelles	0.784	4	0.767	2	0.635	16	0.583	24	30.93
Sierra Leone	0.318	30	0.448	27	0.517	28	0.516	29	9.33
South Africa	0.491	23	0.573	20	0.585	22	0.587	23	22.83
Swaziland	0.519	21	0.542	23	0.606	20	0.633	14	25.53
Tanzania	-	-	0.649	11	0.558	24	0.605	20	10.82
Togo	0.799	2	0.625	13	0.729	4	0.774	2	15.90
Uganda	0.452	25	0.578	18	0.638	15	0.616	18	11.28
Zambia	-	-	0.719	4	0.671	11	0.623	16	14.58

Sources: Tax revenue Dataset for SSA and authors' own calculations.

As an illustration of the potentiality of the devoted website, we explore also the tax effort behavior across African trading zone, on average (figure 7.b) and looking at the heterogeneity within zones (figure 7.c). We find that Eastern African Community (EAC) is the top performing bloc while Central African States Community (rich in oil) is the poorest performer.

Figure 7. Tax effort by trading zone and resource endowment



Sources: Tax revenue Dataset for SSA and authors' own calculations.

For robustness purpose, specifically to deal with the omitted variables bias, we control successively for financial development and development aid and assistance. The three-stage estimation results are respectively reported in tables A.5, and A.7. For each variable, the summary of tax effort for tax subcomponents – columns [2]-[4] – is reported in Tables A.6 and A.8, respectively.

Focusing on non-resource tax, the average total effort does not change when controlling for financial development: the average tax effort remains 0.572. The top performers in the sub-period 2010-2015 remain Togo, Senegal,[23] Burundi, and The Gambia. These countries are resources poor and their tax potential is particularly low (except for Senegal) given their level of poverty and the share of their agricultural sector in their respective GDP. Moreover, Togo and The Gambia raise significant revenue from the transit activities towards landlocked countries or regions. The average tax effort score for subcomponents are slightly higher (+0.01) than those from the baseline results.

Controlling for revenue coming from development aid and assistance results in a slight change in the average tax effort: the average tax effort varies from 0.572 to 0.570. The top performers over the sub-period 2010-2015 are Burundi, Togo, Senegal, and Namibia.[24] The average effort to collect direct tax increases driven by an increase in effort to collect personal income tax and a decrease in effort to collect corporate income tax.

To test the sensibility of our results to the change of production technology between the non-resource tax and its determinants, we use the log-log functional form. The average tax effort varies from 0.572 to 0.571 for the baseline estimation (c.f. tables 1 and A.9). The average tax efforts are 0.581 and 0.566

[23] Note that Senegal comes one place higher.

[24] Namibia replaces The Gambia as the fifth.

when using the log-log specification and controlling for financial development and official assistance and aid respectively.

To further test the robustness of our results, we compute tax effort using some second general stochastic frontier models.[25] We first estimate tax effort using the model of Battese and Coelli (1992). The average tax effort is 0.573 for when using the same variables as in the baseline.

The average tax effort is 0.626 and 0.576 when controlling respectively for financial development and development assistance and aid. Using the model of Kumbhakar (1990), the average tax effort becomes respectively 0.493, 0.626, and 0.587 respectively for the baseline, with financial development and with official assistance and aid.

VI. Replications

Our analysis participates explicitly to the replication effort, which ensures the reliability of produced works. Over recent years, there has been a growing interest in replication particularly in economic research.[26] Following Hamermesch (2007) and Clemens (2017) approach, we undertake a replication of the papers of Gupta (2007) and Fenechietto and Pessino (2013). This replication approach consists in three stages: verification, reproduction, and robustness. Verification means the use of the same sample, population, and empirical specification,[27] while reproduction uses the same econometric model on different samples from the same population.[28] Robustness[29] consists either in running the same specification on different samples and populations, either in applying different econometric specifications on the same sample and population. In our replication exercise,[30] we estimate the same specification for the same sample (same countries and period) as in the original paper for the verification. The reproduction consists in running the same specification on the same sample of countries but including all the available observations for the variables used in the author(s)'s specification(s). Finally, for the robustness, we expand the sample and the time period by using all the countries and years on which data are available to test the author(s)'s specification(s).It is worth mentioning that some differences with respect to the original paper on variables characteristics (i.e. mean, standard deviation, minimum and maximum) emerge even though the sources are the same. This could be due to changes and adjustments in dataset over time. These differences could be also related to some minor treatments by the authors during the dataset compilation, which are not reported in the paper. Furthermore, in the case we do not find a variable from the same source as the author, we take one from another source, if applicable. Otherwise, if the variable is not used in the baseline specification, we do not run the regression for that given specification.

---

[25] Second generation models have their limitation in measuring inefficiency, hence our choice of Generalized True Random Effects model. While they allow for time varying inefficiency, hence improvement over time, the intercept is the same across all countries. Unfortunately, in the presence of time-invariant unobservable factors, they are subject to misspecification bias. Moreover, they are not fitted to deal with heterogeneity between countries. For more discussion, see Wang (2002), Green (2005), and Belotti and Ilardi (2018).

[26] For instance, the top five Reviews in Economics, for a paper to be accepted and published, request the inputs including dataset and program of the paper for replication purpose (Sukhtankar, 2017) and the American Economic Review particularly has dedicated a whole volume to replication. In addition, Anderson and Kichkha (2017)], after a discussion of the three main methods of research synthesis (i.e. traditional literature surveys, meta-analysis and replication), *argue that only pure replication does not contain substantial judgement*.

[27] Hamermesch (2007) calls this procedure

[28] For Hamermesch (2007), this is called a statistical replication.

[29] Hamermesch (2007).

[30] Our replication process is applied based on the following conditions: First, the paper must be an empirical investigation of countries' tax effort (i.e. employing econometric specification) with an actual computation of tax effort. Second, it must be an international comparison of tax effort among countries. In addition, we choose not to replicate number of seminal papers on tax effort prior to the 1990s such as Bahl, (1971, 1972), Chelliah (1971), Chelliah, Baas, and Kelly (1975), Leuthold (1991), Lotz and Morss (1970), Tait, Grätz, and Eichengreen (1979).

Replication of Gupta (2007)

The author estimates countries' revenue potential for a panel of 105 developing countries for the period 1980-2004 using central government revenue dataset. The estimates explained the ratio of central government revenue (excluding grants) to GDP as a function of a set of structural variables (i.e. the log of per capita GDP, the share of agriculture in GDP, the ratio of imports to GDP, share of aid and debt in GDP) and institutional and policy variables (corruption, law and order, government stability, political stability and economic stability). An important limit of this paper is the inclusion of natural resources revenue into tax revenue. Gupta (2007) uses various methods of panel data estimation including Fixed Effect (FE), Random Effects (RE), Common Correlation Coefficient (CCC), Panel Specific Correlation (PSC), and Dynamic Panel Specification (DPS). The results show that the per capita GDP, trade openness, and the share of agriculture in GDP are statistically significant and strong determinants of countries revenue performance. In addition, certain forms of foreign aid improve revenue performance while external debt does not. Regarding politico-institutional factors, the author found that political and economic stability affect positively revenue performance, and corruption significantly and negatively affects revenue performance. The author also emphasizes that countries' tax revenue performance depends on their tax structure. More precisely, countries that depend on indirect taxation as their main source of tax revenue, tend to perform less than countries raising more from direct taxation.

We replicate the key specifications despite a few issues with some variables (economic stability, political stability, and average tariff). The verification test failed to replicate the exact results as in the paper in terms of coefficient and significance of the variables. We have more significant variables than in the original paper (see Tables A.6 and A.7).[31] Moreover, the robustness exercise yields smaller coefficients for all the variables than in the paper, suggesting a smaller effect when non-resources tax is used instead of central government revenue, and when the sample is expanded to all available countries and years (c.f. Tables A.8 and A.9).

Replication of Fenochietto and Pessino (2013)

Fenochietto and Pessino (2013) estimated countries tax capacity and tax effort using Stochastic Frontier Analysis for 113 countries. They first estimated the tax capacity for 96 non-natural resources dependent countries and then on the whole sample using tax and pension contributions revenue collected by central and sub national governments as percent of GDP from the IMF WEO. The authors considered a set of structural and institutional variables (level of development, inflation, education, trade, income inequality, corruption, and the ease of tax collection) explaining countries' tax capacity and estimated tax effort using Battese and Coelli (1992) Half Normal (HN) and Truncated Normal (TN) models incorporating heterogeneity. They also relied on Mundlak (1978) Random Effects Model (REM) to deal with the 'unobserved' heterogeneity.

The verification test for this paper produced almost the same results: the sign and the magnitude are close. As in the paper, the coefficients for non-resource countries are slightly lower than those for all countries (First two columns of Tables A.10 and A.11).

For the robustness analysis, in addition to broadening the sample to all available countries and year while replacing the dependent variable with the ICTD non-resource tax, we also estimate the parameters of Stochastic Frontier (SF) tax function for ICTD non-resource tax while limiting the analysis to the non-resource dependent countries defined in the paper. The robustness results show stable coefficients for all the variables. Nevertheless, the logarithm of the GDP per capita and the logarithm of the GDP per capita square do not have the expected sign or are not significant. We went further in robustness analysis, by relaxing the non-linear relationship assumption between tax revenue and GDP per capita. Thus, we dropped the logarithm of the GDP per capita squared from the specification. Results in last

[31] Although we replicated all the forms of panel data estimations, we present the results for the common correlation coefficient and the panel specific correlation estimations. The reason is that the author expressed his preference for these results in the paper (see Gupta, 2007 p.26).

two columns of Tables A.10 and A.11 show that once we relax the nonlinear relationship assumption, all the variables get the expected sign.

VII. Conclusion

This analysis offers a new dataset of tax revenue, which updates and completes the dataset published in Mansour (2014). We collect statistical information from the African Department of the International Monetary Fund (IMF), most of which is publicly available. We cover 42 SSA countries over the period 1980-2015 distinguishing resource revenue from non-resource (tax) revenue. This work participates to recent efforts to better apprehend tax revenues in Sub-Saharan Africa, in particular the revenue statistics in Africa from the OECD, which cover 26 countries in its last release and the Government Revenue Dataset initiated by the International Centre for Tax and Development (ICTD) and updated by UNU-WIDER.

We provide an estimate of tax effort adopting the Stochastic Frontier Analysis approach. First, we confirm the impact of previously studied factors on countries' DRM capacity such as level of development, financial development, trade openness, natural resource rent, and the size of the agriculture sector. The two latter factors have a negative effect on the domestic revenue mobilization capacity. We estimate on average the total tax effort to be 0.58. Given that non-resource tax revenue amounts to 13.2 percent of GDP, potential tax revenue would be on average 22.75 percent of GDP. Mozambique, Burundi, Togo, Senegal, and Gambia are top performers with a tax effort score above 0.75, while the five lowest performers are resource-rich countries such as Congo Republic (0.366), Chad (0.333), Gabon (0.327), Nigeria (0.243) and Equatorial Guinea (0.073). The poverty level in Mozambique and Burundi explains paradoxically their performance in terms of tax effort. Finally, we did some replication analyses of previous works on tax effort, in particular Gupta (2007) and Fenechietto and Pessino (2013). We fail to replicate the results of Gupta (2007) in terms of robust coefficients and significance of the variables. Some explanatory variables are missing. However, we confirm broadly the analysis of Fenechietto and Pessino (2013).

We acknowledge that our results in terms of tax effort are subject to some caveats. First, additional explanatory variables, in particular regarding political regimes, may be taken into account in the estimation, which could then modify the ranking of countries. Secondly, empirical tools evolve regularly especially the SFA methodology. Third, the estimate of GDP in Africa is particularly weak and heterogeneous across countries. SSA countries regularly update their base year necessary to the computation of GDP with significant changes.[32] That is the reasons why we build a devoted website under R-Shiny linked to this paper. This website allows the reader not only to download our dataset, but also to replicate our empirical analysis and run their own regressions with additional variables.

[32] For instance, Ghana revised upward its total GDP by more than 60% in November 2010.

References

Aigner, Dennis, CA Knox Lovell, and Peter Schmidt (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (1): 21–37.

Alfirman, Luky. (2003). "Estimating Stochastic Frontier Tax Potential: Can Indonesian Local Governments Increase Tax Revenues under Decentralization?" Center for Economic Analysis, Department of Economics, University of Colorado.

Anderson, Richard G. and Areerat Kichkha (2017). "Replication, Meta-Analysis, and Research Synthesis in Economics." *American Economic Review* 107 (5): 56–59.

Bahl, Roy W. (1971). "A Regression Approach to Tax Effort and Tax Ratio Analysis." *Staff Papers* 18 (3): 570–612.

———. (1972). "A Representative Tax System Approach to Measuring Tax Effort in Developing Countries." *Staff Papers* 19 (1): 87–124.

Baicker, Katherine and Jonathan Skinner. (2011). "Health Care Spending Growth and the Future of US Tax Rates." *Tax Policy and the Economy* 25 (1): 39–68.

Battese, George E. and Tim J Coelli. (1988). "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38 (3): 387–99.

———. (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India." *Journal of Productivity Analysis* 3 (1): 153–69.

Belinga, Vincent, Maximillien Kaffo Melou, and Jean-Pascal Nganou. (2017). "Does Oil Revenue Crowd Out Other Tax Revenues? Policy Lessons for Uganda." The World Bank.

Belloti, Frederico and Giuseppe Ilardi (2018). "Consistent inference in fixed-effects stochas- tic frontier models". In: *Journal of Econometrics* 202 (2): 161–177.

Besley, Timothy and Torsten Persson. (2014). "Why Do Developing Countries Tax so Little?" *Journal of Economic Perspectives* 28 (4): 99–120.

Bornhorst, Fabian, Sanjeev. Gupta, and John Thornton. (2009). "Natural Resource Endowments and the Domestic Revenue Effort," *European Journal of Political Economy*, 25(4): 439-446.

Brun J. F., Chambas G. and Mansour M., (2014), "Tax Effort of Developing Countries: An Alternative Measure", in *Financing sustainable development by addressing vulnerabilities*, in M. Boussichas and P. Guillaumont edited by Economica, Paris.

Brun, Jean-François and Maïmouna Diakite. (2016). "Tax Potential and Tax Effort: An Empirical Estimation for Non-Resource Tax Revenue and VAT's Revenue." *Études et Documents 2016/10*, CERDI, France.

Butler, John S. and Robert Moffitt. (1982). "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model." *Econometrica: Journal of the Econometric Society*, 761–764.

Caldeira, Emilie and Grégoire Rota-Graziosi, (2019), "Tax effort in Benin: How can tax gaps be reduced?" with Emilie Caldeira, in F. Bourgignon, R. Houssa, J.-P. Platteau and P. Reding, (eds) *Benin Institutional Diagnostic*, Economic Development & Institutions, Chapter 6.

Caldeira, Emilie, Ali Compaore, Alou A. Dama, and Grégoire Rota-Graziosi, (2020), "Inter-ministerial Tax Competition: The Case of Resource-Rich Developing Countries," Working Paper, CERDI, forthcoming.

Charnes, Abraham, William Cooper, Arie Y Lewin, and Lawrence M. Seiford. (1997). "Data Envelopment Analysis Theory, Methodology and Applications." *Journal of the Operational Research Society* 48(3):332–333.

Chelliah, Raja J. (1971). "Trends in Taxation in Developing Countries." *Staff Papers* 18 (2): 254–331.

Chelliah, Raja J., Hessel J. Baas, and Margaret R. Kelly. (1975). "Tax Ratios and Tax Effort in Developing Countries, 1969-71." *Staff Papers* 22 (1): 187–205.

Clemens, Michael A. (2017). "The Meaning of Failed Replications: A Review and Proposal." *Journal of Economic Surveys* 31 (1): 326–342.

Clist, Paul and Oliver Morrissey. (2011). "Aid and Tax Revenue: Signs of a Positive Effect since the 1980s." *Journal of International Development* 23 (2): 165–180.

Colombi, Roberto, Subal C. Kumbhakar, Gianmaria Martini, and Giorgio Vittadini. (2014). "Closed-Skew Normality in Stochastic Frontiers with Individual Effects and Long/Short-Run Efficiency." *Journal of Productivity Analysis* 42 (2): 123–136.

Cornwell, Christopher, Peter Schmidt, and Robin C Sickles. (1990). "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels." *Journal of Econometrics* 46 (1–2): 185–200.

Crivelli, Ernesto and Sanjeev Gupta. (2014). "Resource Blessing, Revenue Curse? Domestic Revenue Effort in Resource-Rich Countries." *European Journal of Political Economy* 35: 88–101.

Cyan, Musharraf, Jorge Martinez-Vazquez, and Violeta Vulovic. (2013). "Measuring Tax Effort: Does the Estimation Approach Matter and Should Effort Be Linked to Expenditure Goals?" International Center for Public Policy Working Paper 13-08, Andrew Young School of Policy Studies, Georgia State University.

Deprins, Dominique, Leopold Simar, and Henry Tulkens. (1984). "Measuring Labor-Efficiency in Post Offices.", in *The Performance of Public Enterprises: Concepts and Measurement*, Amsterdam, 1984, 243-267.

Eltony, Nagy. (2001). "The Determinants of Tax Effort in Arab Countries." API-Working Paper Series 0207, Arab Planning Institute - Kuwait, Information Center.

Feger, Thuto and John Asafu-Adjaye. (2014). "Tax Effort Performance in Sub-Sahara Africa and the Role of Colonialism." *Economic Modelling* 38: 163–174.

Fenochietto, Ricardo and Carola Pessino. (2013). "Understanding Countries' Tax Effort." 13–244. International Monetary Fund.

Filippini, Massimo, and William Greene. (2016). "Persistent and Transient Productive Inefficiency: A Maximum Simulated Likelihood Approach." *Journal of Productivity Analysis* 45 (2): 187–196.

Gordon, Roger and Wei Li. (2009). "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics* 93 (7–8): 855–866.

Greene, William. (2005). "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model." *Journal of Econometrics* 126 (2): 269–303.

Grigorian, David A. and Hamid R. Davoodi. (2007). "Tax Potential vs. Tax Effort: A Cross-Country Analysis of Armenia's Stubbornly Low Tax Collection." 7–106. International Monetary Fund.

Gupta, Abhijit Sen. (2007). "Determinants of Tax Revenue Efforts in Developing Countries." 7–184. International Monetary Fund.

Hamermesh, Daniel S. (2007). "Replication in Economics." *Canadian Journal of Economics/Revue Canadienne d'économique* 40 (3): 715–733.

Hjalmarsson, Lennart, Subal C.Kumbhakar, and Almas Heshmati. (1996). "DEA, DFA and SFA: A Comparison." *Journal of Productivity Analysis* 7 (2–3): 303–327.

Kumbhakar, Subal C. (1990). "Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency." *Journal of Econometrics* 46 (1–2): 201–211.

Kumbhakar, Subal C. and Almas Heshmati. (1995). "Efficiency Measurement in Swedish Dairy Farms: An Application of Rotating Panel Data, 1976–88." *American Journal of Agricultural Economics* 77 (3): 660–674.

Kumbhakar, Subal C., Gudbrand Lien, and J. Brian Hardaker. (2014). "Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming." *Journal of Productivity Analysis* 41 (2): 321–337.

Kumbhakar, Subal C. and Hung-Jen Wang. (2005). "Estimation of Growth Convergence Using a Stochastic Production Frontier Approach." *Economics Letters* 88 (3): 300–305.

Kumbhakar, Subal C., Hung-Jen Wang, and Alan P. Horncastle. (2015). "A Practitioner's Guide to Stochastic Frontier Analysis Using Stata." Cambridge University Press.

Langford, Ben and Tim Ohlenburg. (2015). "Tax Revenue Potential and Effort." International Growth Centre Working Paper. 2015.

Leuthold, Jane H. (1991). "Tax Shares in Developing Economies a Panel Study." *Journal of Development Economics* 35 (1): 173–185.

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

———. (1970). "A Theory of Tax Level Determinants for Developing Countries." *Economic Development and Cultural Change* 18 (3): 328–341.

Mansour, Mario. (2014). "A Tax Revenue Dataset for Sub-Saharan Africa: 1980-2010." *Revue D'Économie Du Développement, Forthcoming*.

McGuirk, Eoin, (2013), "The illusory Leader: Natural Resources, Taxation, and Accountability," *Public Choice*, 154, 285-313.

Meeusen, Wim and Julien van Den Broeck. (1977). "Efficiency Estimation from Cobb-Douglas

Production Functions with Composed Error." *International Economic Review*, 435–444.
Morrissey, Oliver. (2015). Aid and Government Fiscal Behavior: Assessing Recent Evidence. *World Development* 69, 98–105.
Mundlak, Yair. (1978). "On the Pooling of Time Series and Cross Section Data." *Econometrica: Journal of the Econometric Society*, 69–85.
Odeck, James. (2007). "Measuring Technical Efficiency and Productivity Growth: A Comparison of SFA and DEA on Norwegian Grain Production Data." *Applied Economics* 39 (20): 2617–2630.
Pessino, Carola and Ricardo Fenochietto. (2010). "Determining Countries' Tax Effort." *Hacienda Pública Española/Revista de Economía Pública*, 65–87.
Pitt, Mark M. and Lung-Fei Lee. (1981). "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry." *Journal of Development Economics* 9 (1): 43–64.
Sachs, Jeffrey D. and Andrew M. Warner. (2001). "The Curse of Natural Resources." *European Economic Review* 45 (4–6): 827–838.
Schmidt, Peter and Robin C Sickles. (1984). "Production Frontiers and Panel Data." *Journal of Business & Economic Statistics* 2 (4): 367–374.
Stotsky, Janet Gale and Asegedech WoldeMariam. (1997). "Tax Effort in Sub-Saharan Africa." 97–107. International Monetary Fund.
Sukhtankar, Sandip. (2017). "Replications in Development Economics." *American Economic Review* 107 (5): 32–36.
Svirydzenka, Katsiaryna. (2016). "Introducing a New Broad-based Index of Financial Development." IMF Working Paper, WP/16/05.
Tait, Alan A., Wilfrid L.M Grätz, and Barry J. Eichengreen. (1979). "International Comparisons of Taxation for Selected Developing Countries, 1972-76." *Staff Papers* 26 (1): 123–156.
Tanzi, Vito. 1992. "Structural Factors and Tax Revenue in Developing Countries: A Decade of Evidence," In Open economies: Structural Adjustment and Agriculture, Ian Goldin, and L. Alan Winters (Eds.), Cambridge: Cambridge University Press, 267-281.
Tanzi, Vito and Hamid R. Davoodi. (1998). "Corruption, Public Investment, and Growth." In *The Welfare State, Public Investment, and Growth*, 41–60. Springer.
Tsionas, Efthymios G. and Subal C. Kumbhakar. (2014). "Firm Heterogeneity, Persistent and Transient Technical Inefficiency: A Generalized True Random-Effects Model." *Journal of Applied Econometrics* 29 (1): 110–132.
Yohou, Hermann Djedje and Michaël Goujon. (2017). "Reassessing Tax Effort in Developing Countries: A Proposal of a Vulnerability-Adjusted Tax Effort Index (VATEI)." Working Papers P186, FERDI.
Wang, Hung-Jen (2002). "Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model". In: Journal of Productivity Analysis 18(3): 241–253.
Wang, Hung-Jen and Chia-Wen Ho. (2010). "Estimating Fixed-Effect Panel Stochastic Frontier Models by Model Transformation." *Journal of Econometrics* 157 (2): 286–296.

Appendix

Table A1: Country list

Angola, Benin, Botswana, Botswana, Burkina Faso, Burundi, Cabo Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Côte d'Ivoire, Equatorial Guinea, Ethiopia, Gabon, Gambia, The, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Madagascar, Malawi, Mali, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Seychelles, Sierra Leone, South Africa, Swaziland, São Tomé and Principe, Tanzania, Togo, Uganda, Zambia, Zimbabwe

Table A2: Data sources and definition

Variables	Definition	Sources
Total non-resource tax (% GDP)	Total tax revenues excluding resource rent	Tax revenue dataset, Mansour (2019)
Corporate income tax (% GDP)	Tax imposed on corporate income in countries that have a corporate tax	
Total direct taxes (% GDP)	Taxes on all income sources (i.e. business profits, wages, portfolio income, income from real property, capital gains, etc.)	
Personal income taxes (% GDP)	Taxes on individual income	
Total resource rent (% GDP)	Sum of oil rents, natural gas rents, coal rents (hard and soft), mineral rents, and forest rents over GDP.	World Development Indicators, World Bank
GDP PC (constant 2010 US $)	Sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products	
Total Trade (% GDP)	Volume of imports and exports over GDP	
Agriculture, value added (% GDP)	Net output of forestry, hunting, and fishing, as well as cultivation of crops and livestock production after adding up all outputs and subtracting intermediate inputs, divided by GDP	
Gini index	Index measuring the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution	
Net official development assistance and official aid received (% GDP)	Net official development assistance is disbursement flows (net of repayment of principal). Net official aid refers to aid flows (net of repayments).	
Financial development index	Aggregate of nine indices that summarize how developed financial institutions and financial markets are in terms of their depth, access, and efficiency.	Svirydzenka (2016)

Table A3: Descriptive statistics

Variables	Obs.	Mean	Std. Dev.	Min	Max
Total Taxes	1473	16.19	8.97	0.57	53.33
Total non-resource taxes (% GDP)	1473	13.22	7.09	0.55	50.81
Total income taxes	1473	3.89	2.74	0.18	18.69
Corporate Income Tax	1373	1.64	1.24	0.00	9.06
Personal Income Tax	1368	1.84	1.79	0.00	13.33
GDP per capita (constant 2010 US$)	1474	1892.50	2780.36	131.65	20333.94
Total trade (% of GDP)	1323	71.3	36.48	6.32	265.98
Agriculture, value added (% of GDP)	1345	27.42	15.70	0.89	72.03
Total natural resources rents (% of GDP)	1431	11.61	11.92	0.00	89.17
Financial development (% of GDP)	1435	0.11	0.08	0.00	0.64
Net official development assistance and official aid received (% of GDP)	1448	10.79	10.49	-0.25	94.44

Table A.4: Summary of tax effort estimation results - Baseline estimation [2]-[4]

Dependent	Tax effort components	Obs.	Mean	SD	Min	Max
Direct	Time-varying tax effort	1165	0.756	0.109	0.223	0.936
	Persistent tax effort	1165	0.620	0.157	0.064	0.866
	Total tax effort	1165	0.470	0.139	0.031	0.755
CIT	Time-varying tax effort	1086	0.701	0.114	0.092	0.919
	Persistent tax effort	1086	0.443	0.221	0.045	1.000
	Total tax effort	1086	0.311	0.165	0.019	0.834
PIT	Time-varying tax effort	1081	0.648	0.148	0.031	0.940
	Persistent tax effort	1081	0.539	0.182	0.035	0.834
	Total tax effort	1081	0.350	0.142	0.006	0.676

Table A.5: The three-stage estimation results controlling for financial development

	[1]	[2]	[3]	[4]
Variables	NRTAX	Direct	CIT	PIT
Log GDP per cap. (Cst. 2010 USD) $_{(-1)}$	0.228***	0.337***	0.288***	0.339***
	(0.0309)	(0.0459)	(0.0617)	(0.0716)
Total trade (%GDP) $_{(-1)}$	0.001***	0.004***	0.008***	0.004***
	(0.0004)	(0.0006)	(0.0010)	(0.0011)
Agriculture value added (%GDP) $_{(-1)}$	-0.003**	-0.007***	0.002	-0.021***
	(0.0014)	(0.0021)	(0.0029)	(0.0033)
Total natural resource rent (%GDP) $_{(-1)}$	-0.001	-0.000	-0.000	-0.007*
	(0.0014)	(0.0021)	(0.0029)	(0.0035)
Financial development $_{(-1)}$	0.776***	1.259***	0.666	1.070**
	(0.1821)	(0.2718)	(0.4225)	(0.4803)
Constant	0.739***	-1.500***	-2.411***	-1.912***
	(0.2303)	(0.3418)	(0.4534)	(0.5273)
Observations	1165	1165	1,086	1,081
R-squared	0.170	0.245	0.140	0.181
Number of countries	39	39	38	38

Panel A: Stage 2 - estimation of the time-varying tax inefficiency (stochastic frontier)

				Number of obs.		1165
				Wald chi2(1)		318.74
Log likelihood = 111.52				Prob > chi2		0.0000
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
frontier (one)	0.214	0.012	17.850	0.000	0.191	0.238
usigmas (_cons)	-2.605	0.106	-24.640	0.000	-2.813	-2.398
vsigma (_cons)	-3.773	0.100	-37.740	0.000	-3.969	-3.577

Panel B: Stage 3 - estimation of the persistent tax inefficiency (stochastic frontier)

				Number of obs.		1165
				Wald chi2(1)		1568.67
Log likelihood = -339.46				Prob > chi2		0.0000
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
frontier (one)	0.445	0.011	39.610	0.000	0.423	0.467
usigmas (_cons)	-1.278	0.058	-22.180	0.000	-1.391	-1.165
vsigma (_cons)	-4.016	0.125	-32.070	0.000	-4.261	-3.770

Panel C: Summary of tax effort estimation results

	Obs.	Mean	Std. Dev.	Min	Max
Time-varying tax effort	1165	0.818	0.091	0.322	0.968
Persistent tax effort	1165	0.698	0.171	0.108	0.946
Total tax effort	1165	0.572	0.156	0.061	0.848

Notes: Significance: *** p<0.01, ** p<0.05, * p<0.1; Standard errors in parentheses. NRTAX: Total non-resource tax revenue; Direct: Total income tax revenue; CIT: Corporate income tax revenue; PIT: Personal income tax revenue.

Table A.6: Summary of tax effort estimation results – Robustness controlling for financial development [2]-[4]

Dependent	Tax effort components	Obs.	Mean	SD	Min	Max
Direct	Time-varying tax effort	1165	0.757	0.108	0.209	0.941
	Persistent tax effort	1165	0.637	0.152	0.080	0.874
	Total tax effort	1165	0.483	0.136	0.038	0.766
CIT	Time-varying tax effort	1086	0.700	0.115	0.088	0.919
	Persistent tax effort	1086	0.456	0.217	0.054	1.000
	Total tax effort	1086	0.320	0.164	0.019	0.840
PIT	Time-varying tax effort	1081	0.646	0.149	0.030	0.941
	Persistent tax effort	1081	0.555	0.177	0.045	0.838
	Total tax effort	1081	0.360	0.141	0.006	0.676

Table A.7: The three-stage estimation results controlling for official development assistance and aid

Variables	[1] NRTAX	[2] Direct	[3] CIT	[4] PIT
Log GDP per cap. (Cst. 2010 USD) $_{(-1)}$	0.253***	0.367***	0.280***	0.324***
	(0.0325)	(0.0483)	(0.0643)	(0.0755)
Total trade (%GDP) $_{(-1)}$	0.002***	0.005***	0.009***	0.004***
	(0.0004)	(0.0006)	(0.0010)	(0.0012)
Agriculture value added (%GDP) $_{(-1)}$	-0.003**	-0.007***	0.001	-0.022***
	(0.0014)	(0.0021)	(0.0030)	(0.0034)
Total natural resource rent (%GDP) $_{(-1)}$	-0.002	-0.001	-0.000	-0.006*
	(0.0014)	(0.0020)	(0.0029)	(0.0035)
Net off. Dev. assistance and off. aid received (%GDP) $_{(-1)}$	-0.001	-0.006***	-0.008***	-0.009***
	(0.0011)	(0.0017)	(0.0025)	(0.0027)
Constant	0.669***	-1.539***	-2.209***	-1.603***
	(0.2491)	(0.3696)	(0.4939)	(0.5760)
Observations	1137	1137	1062	1057
R-squared	0.159	0.251	0.155	0.175
Number of countries	39	39	38	38

Panel A: Stage 2 - estimation of the time-varying tax inefficiency (stochastic frontier)

				Number of obs.		1137
				Wald chi2(1)		305.96
Log likelihood = 106.30				Prob > chi2		0.0000
	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
frontier (one)	0.213	0.012	17.490	0.000	0.189	0.237
usigmas (_cons)	-2.617	0.108	-24.270	0.000	-2.828	-2.406
vsigma (_cons)	-3.752	0.099	-37.750	0.000	-3.947	-3.557

Panel B: Stage 3 - estimation of the persistent tax inefficiency (stochastic frontier)

				Number of obs.		1137
				Wald chi2(1)		1387.37
Log likelihood = -376.46				Prob > chi2		0.0000
	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
frontier (one)	0.451	0.012	37.250	0.000	0.427	0.475
usigmas (_cons)	-1.233	0.060	-20.640	0.000	-1.350	-1.116
vsigma (_cons)	-3.819	0.120	-31.760	0.000	-4.055	-3.583

Panel C: Summary of tax effort estimation results

	Obs.	Mean	Std. Dev.	Min	Max
Time-varying tax effort	1137	0.819	0.090	0.319	0.967
Persistent tax effort	1137	0.694	0.170	0.099	0.941
Total tax effort	1137	0.570	0.155	0.056	0.842

Notes: Significance: *** p<0.01, ** p<0.05, * p<0.1; Standard errors in parentheses. NRTAX: Total non-resource tax revenue; Direct: Total income tax revenue; CIT: Corporate income tax revenue; PIT: Personal income tax revenue.

Table A.8: Summary of tax effort estimation results – Robustness controlling for official development assistance and aid [2]-[4]

Dependent	Tax effort components	Obs.	Mean	SD	Min	Max
Direct	Time-varying tax effort	1137	0.763	0.104	0.232	0.935
	Persistent tax effort	1137	0.619	0.155	0.066	0.863
	Total tax effort	1137	0.474	0.137	0.033	0.758
CIT	Time-varying tax effort	1062	0.699	0.115	0.089	0.917
	Persistent tax effort	1062	0.419	0.214	0.042	1.000
	Total tax effort	1062	0.294	0.160	0.018	0.823
PIT	Time-varying tax effort	1057	0.647	0.148	0.030	0.940
	Persistent tax effort	1057	0.554	0.172	0.045	0.829
	Total tax effort	1057	0.360	0.138	0.006	0.655

Table A.9: The three-stage estimation results: baseline, controlling for financial development and official development assistance and aid – Log-log specification

Variables	[1]		[2]		[3]	
	NRTAX					
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Log GDP per cap. (Cst. 2010 USD) (-1)	0.329***	(0.0350)	0.287***	(0.0349)	0.358***	(0.0370)
Total trade (%GDP) (-1)	0.279***	(0.0289)	0.282***	(0.0294)	0.296***	(0.0301)
Agriculture value added (%GDP) (1)	0.068**	(0.0339)	0.098***	(0.0335)	0.098***	(0.0349)
Total natural resource rent (%GDP) (-1)	-0.006	(0.0148)	-0.000	(0.0151)	0.004	(0.0149)
Financial development (-1)			0.204***	(0.0292)		
Net off. Dev. assistance and off. aid received (%GDP) (-1)					-0.001	(0.0112)
Constant	-1.222***	(0.3231)	-0.580*	(0.3404)	-1.600***	(0.3393)
Observations	1155		1126		1125	
R-squared	0.229		0.261		0.238	
Number of countries	39		39		39	

Panel A: Stage 2 - estimation of the time-varying tax inefficiency (stochastic frontier)

	[1]	[2]	[3]		[1]	[2]	[3]
				Number of obs.	1155	1126	1125
				Wald chi2(1)	226.97	254.58	224.92
Log likelihood	141.42	169.20	149.50	Prob > chi2	0.0000	0.0000	0.0000

	[1]		[2]		[3]	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
frontier (one)	0.192***	(0.0127)	0.194***	(0.0121)	0.1896***	(0.0126)
usigmas (_cons)	-2.828***	(0.1254)	-2.812***	(0.1184)	-2.852***	(0.1258)
vsigma (_cons)	-3.680***	(0.0943)	-3.794***	(0.0981)	-3.6974***	(0.0940)

Panel B: Stage 3 - estimation of the persistent tax inefficiency (stochastic frontier)

	[1]	[2]	[3]		[1]	[2]	[3]
				Number of obs.	1155	1126	1125
				Wald chi2(1)	1511.43	1382.52	1515.57
Log likelihood	-452.54	-385.76	-473.43	Prob > chi2	0.0000	0.0000	0.0000

	[1]		[2]		[3]	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
frontier (one)	0.478***	(0.0123)	0.447***	(0.0120)	0.488***	(0.0125)
usigmas (_cons)	-1.125***	(0.0581)	-1.240***	(0.0597)	-1.056***	(0.0580)
vsigma (_cons)	-3.661***	(0.1092)	-3.710***	(0.1089)	-3.638***	(0.1098)

Panel C: Summary of tax effort estimation results

		Obs.	Mean	Std. Dev.	Min	Max
[1]	Time-varying tax effort	1155	0.834	0.078	0.379	0.965
	Persistent tax effort	1155	0.683	0.175	0.090	0.940
	Total tax effort	1155	0.571	0.158	0.056	0.865
[2]	Time-varying tax effort	1126	0.833	0.082	0.365	0.965
	Persistent tax effort	1126	0.696	0.169	0.113	0.941
	Total tax effort	1126	0.581	0.154	0.070	0.866
[3]	Time-varying tax effort	1125	0.836	0.078	0.382	0.963
	Persistent tax effort	1125	0.676	0.179	0.085	0.939
	Total tax effort	1125	0.566	0.161	0.054	0.866

Appendix: Replications of Gupta (2007)

Table A.10: Replication results of Gupta (2007) – Common Correlation Coefficients (Verification)

	I	II	III	IV	VII	VIII	IX	X
Log of Per capita GDP	3.624***	3.874***	3.080***	3.360***				
	(0.398)	(0.363)	(0.401)	(1.103)				
Agriculture share					-0.208***	-0.181***	-0.125***	-0.243***
					(0.0362)	(0.0382)	(0.0352)	(0.0754)
Import share		0.0330	0.112***	0.0484*		0.0338**	0.0821***	0.0741**
		(0.0213)	(0.0207)	(0.0260)		(0.0158)	(0.0212)	(0.0310)
Aid share			-0.00778	0.0371			-0.00711	0.0419
			(0.0252)	(0.0645)			(0.0442)	(0.0771)
Debt share			0.00512	0.0308*			-0.0111**	0.00927
			(0.00332)	(0.0164)			(0.00485)	(0.0153)
Government stability				0.231				0.297*
				(0.173)				(0.170)
Corruption				-0.305				-0.433
				(0.553)				(0.539)
Law and order				0.348				0.476
				(0.352)				(0.321)
Tax on goods and services				0.337**				0.0674
				(0.169)				(0.181)
Tax on income, profit and capital gain				0.521***				0.441**
				(0.171)				(0.179)
Tax on trade				1.151***				0.852***
				(0.199)				(0.191)
Constant	-8.578***	-12.16***	-9.976***	-21.63*	25.51***	23.50***	20.82***	12.97***
	(3.270)	(2.907)	(3.014)	(11.09)	(0.678)	(0.843)	(1.039)	(2.104)
Observations	954	926	595	120	1,046	1,022	677	118
R-squared	0.419	0.443	0.534	0.662	0.312	0.311	0.364	0.614
Number of countries	93	89	62	20	85	83	59	19

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

Table A.11: Replication results of Gupta (2007) – Panel Specific Correlation Coefficients (Verification)

	I	II	III	IV	VII	VIII	IX	X
Log of Per capita GDP	3.668***	4.304***	3.259***	3.543***				
	(0.365)	(0.354)	(0.395)	(1.216)				
Agriculture share					-0.215***	-0.188***	-0.174***	-0.299***
					(0.0301)	(0.0343)	(0.0313)	(0.0686)
Import share		0.0306	0.128***	0.0268		0.0153	0.0624**	0.0436
		(0.0220)	(0.0209)	(0.0230)		(0.0163)	(0.0297)	(0.0267)
Aid share			-0.0169	0.0338			-0.00306	0.0792
			(0.0244)	(0.0626)			(0.0500)	(0.0669)
Debt share			0.00674*	0.0404***			-0.00510	0.0121
			(0.00352)	(0.0151)			(0.00578)	(0.0156)
Government stability				0.275				0.348**
				(0.181)				(0.159)
Corruption				-0.297				-0.353
				(0.463)				(0.384)
Law and order				0.173				0.403
				(0.328)				(0.350)
Tax on goods and services				0.577***				0.309*
				(0.143)				(0.168)
Tax on income, profit and capital gain				0.571***				0.309*
				(0.166)				(0.166)
Tax on trade				1.135***				0.690***
				(0.217)				(0.186)
Constant	-8.602***	-15.32***	-12.04***	-24.46**	26.21***	25.12***	23.07***	14.33***
	(2.993)	(2.636)	(2.829)	(11.11)	(0.660)	(0.860)	(1.340)	(2.159)
Observations	954	926	595	120	1,046	1,022	677	118
R-squared	0.678	0.685	0.773	0.801	0.606	0.594	0.602	0.860
Number of countries	93	89	62	20	85	83	59	19

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

Table A.12: Replication results of Gupta (2007) – Common Correlation Coefficients (Robustness)

	I	II	III	IV	VII	VIII	IX	X
Log of Per capita GDP	2.330***	2.409***	1.993***	0.0881*				
	(0.207)	(0.206)	(0.197)	(0.0485)				
Agriculture share					-0.169***	-0.162***	-0.103***	-0.00172
					(0.0117)	(0.0120)	(0.0122)	(0.00275)
Import share		0.0166***	0.0537***	0.00266**		0.0280***	0.0555***	0.00315**
		(0.00453)	(0.00691)	(0.00122)		(0.00430)	(0.00670)	(0.00132)
Aid share			-0.0206**	-0.00231			-0.0218**	-0.00264
			(0.00829)	(0.00159)			(0.00973)	(0.00179)
Debt share			-0.00464***	6.78e-05			-0.00450**	1.76e-05
			(0.00143)	(0.000302)			(0.00194)	(0.000411)
Government stability				0.00737				0.00693
				(0.0124)				(0.0116)
Corruption				-0.0110				-0.0248
				(0.0348)				(0.0344)
Law and order				-0.0305				-0.0313
				(0.0356)				(0.0331)
Tax on goods and services				1.045***				1.039***
				(0.0156)				(0.0159)
Tax on income, profit and capital gain				0.969***				0.983***
				(0.0175)				(0.0166)
Tax on trade				1.015***				1.015***
				(0.0174)				(0.0169)
Constant	-3.919**	-5.465***	-3.817**	-0.226	19.15***	17.67***	14.32***	0.515***
	(1.735)	(1.710)	(1.576)	(0.338)	(0.374)	(0.405)	(0.431)	(0.189)
Observations	3,874	3,729	2,317	1,256	4,195	4,073	2,552	1,321
R-squared	0.321	0.337	0.348	0.953	0.307	0.324	0.341	0.956
Number of countries	187	183	116	73	181	177	112	71

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

Table A.13: Replication results of Gupta (2007) – Panel Specific Correlation Coefficients (Robustness)

	I	II	III	IV	VII	VIII	IX	X
Log of Per capita GDP	2.027***	2.097***	1.984***	0.00538				
	(0.169)	(0.159)	(0.197)	(0.0509)				
Agriculture share					-0.154***	-0.145***	-0.106***	-0.00234
					(0.0117)	(0.0126)	(0.0130)	(0.00262)
Import share		0.0202***	0.0612***	0.00529***		0.0326***	0.0622***	0.00510***
		(0.00423)	(0.00712)	(0.00116)		(0.00486)	(0.00730)	(0.00128)
Aid share			-0.0198**	-0.00344**			-0.0237**	-0.00340*
			(0.00912)	(0.00168)			(0.0103)	(0.00190)
Debt share			-0.00383***	-0.000193			-0.00518***	-8.41e-07
			(0.00133)	(0.000346)			(0.00191)	(0.000430)
Government stability				0.00524				0.00243
				(0.0110)				(0.0102)
Corruption				-0.0272				-0.0332
				(0.0328)				(0.0322)
Law and order				-0.0471				-0.0378
				(0.0354)				(0.0332)
Tax on goods and services				1.056***				1.042***
				(0.0132)				(0.0146)
Tax on income, profit and capital gain				0.985***				0.999***
				(0.0186)				(0.0176)
Tax on trade				1.009***				1.020***
				(0.0197)				(0.0182)
Constant	-1.206	-2.878**	-3.451**	0.352	18.69***	16.87***	14.40***	0.431***
	(1.382)	(1.289)	(1.504)	(0.373)	(0.352)	(0.447)	(0.488)	(0.154)
Observations	3,874	3,729	2,317	1,256	4,195	4,073	2,552	1,321
R-squared	0.613	0.620	0.633	0.984	0.559	0.541	0.555	0.984
Number of countries	187	183	116	73	181	177	112	71

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

Appendix: Replications of Fenochietto and Pessino (2013)

Table A.14: Parameter of the SF tax function for Non-resource dependent countries

	Verification		Robustness			
	BC Half Normal	BC Truncated Normal	BC Half Normal	BC Truncated Normal	BC Half Normal	BC Truncated Normal
Log of per capita GDP	0.523**	0.534***	-7.745*	10.66**	5.883***	2.494***
	(0.203)	(0.206)	(4.680)	(5.161)	(0.450)	(0.490)
Agri. added value (% GDP)	-0.005***	-0.005***	-0.03	-0.03	0.017	-0.057
	(0.002)	(0.002)	(0.04)	(0.04)	(0.036)	(0.038)
Pub. expenditure in Edu. (%GDP)	0.031***	0.031***	1.268***	0.809***	1.184***	0.888***
	(0.006)	(0.006)	(0.136)	(0.142)	(0.134)	(0.134)
Trade	0.0006**	0.0006**	0.04***	0.005	0.037***	0.008
	(0.0002)	(0.0002)	(0.007)	(0.007)	(0.007)	(0.006)
GINI index	-0.008***	-0.008***	-0.140***	-0.125***	-0.159***	-0.128***
	(0.001)	(0.001)	(0.025)	(0.028)	(0.026)	(0.028)
Log of per capita GDP squared	-0.022**	-0.023**	0.772***	-0.463		
	(0.011)	(0.011)	(0.264)	(0.290)		
Constant	0.966	0.925	32.35	-16.60	-24.63***	23.90
	(0.961)	(0.968)	(20.23)	(23.33)	(5.035)	(0)
Sigma	-1.880***	-2.066***	4.273***	2.983***	4.575***	2.868***
	(0.185)	(0.455)	(0.220)	(0.180)	(0.205)	(0.151)
Gamma	3.558***	3.367***	3.196***	1.840***	3.514***	1.678***
	(0.204)	(0.475)	(0.243)	(0.226)	(0.224)	(0.192)
Mu	(omitted)	0.0892	(omitted)	20.25***	(omitted)	25.09***
		(0.184)		(5.366)		(5.152)
Eta	-0.004*	-0.004**	-0.024***	-0.002	-0.019***	-0.002
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.0008)
Observations	533	533	561	561	561	561
Number of countries	68	68	70	70	70	70
sigma2	0.153	0.127	71.73	19.75	97.01	17.60
gamma	0.972	0.967	0.961	0.863	0.971	0.843
sigma_u	0.385	0.350	8.301	4.129	9.706	3.851
sigma_v	0.065	0.065	1.679	1.645	1.675	1.664

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses

Table A.15: Parameter of the SF tax function for Non-resource dependent and resources dependent countries

	Verification		Robustness			
	BC Half Normal	BC Truncated Normal	BC Half Normal	BC Truncated Normal	BC Half Normal	BC Truncated Normal
Log of per capita GDP	0.599***	0.615***	-9.203*	7.646	5.454***	2.188***
	(0.193)	(0.197)	(4.740)	(4.732)	(0.441)	(0.493)
Agri. added value (% GDP)	-0.005***	-0.005***	-0.071**	-0.0758**	-0.029	-0.093***
	(0.002)	(0.002)	(0.034)	(0.036)	(0.032)	(0.033)
Pub. expenditure in Edu. (%GDP)	0.035***	0.034***	1.310***	0.838***	1.215***	0.857***
	(0.006)	(0.006)	(0.137)	(0.132)	(0.134)	(0.130)
Trade	0.0007***	0.0007***	0.028***	0.002	0.023***	0.002
	(0.0002)	(0.0003)	(0.007)	(0.006)	(0.007)	(0.006)
GINI index	-0.008***	-0.008***	-0.130***	-0.096***	-0.144***	-0.097***
	(0.001)	(0.001)	(0.024)	(0.026)	(0.026)	(0.026)
Log of per capita GDP squared	-0.026***	-0.027***	0.829***	-0.306		
	(0.010)	(0.011)	(0.267)	(0.266)		
Constant	0.592	0.536	41.10**	-3.334	-20.41***	20.64***
	(0.909)	(0.923)	(20.47)	(20.87)	(4.878)	(6.686)
Sigma	-1.888***	-2.171***	4.332***	3.161***	4.647***	3.127***
	(0.181)	(0.417)	(0.205)	(0.146)	(0.187)	(0.136)
Gamma	3.487***	3.196***	2.938***	1.723***	3.279***	1.677***
	(0.200)	(0.437)	(0.232)	(0.188)	(0.209)	(0.174)
Mu	(omitted)	0.132	(omitted)	20.66***	(omitted)	20.52***
		(0.158)		(2.212)		(3.569)
Eta	-0.003	-0.003	-0.018***	-0.0007	-0.014***	-0009
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)
Observations	566	566	681	681	681	681
Number of countries	73	73	95	95	95	95
sigma2	0.151	0.114	76.13	23.59	104.3	22.82
gamma	0.970	0.961	0.950	0.849	0.964	0.842
sigma_u	0.383	0.331	8.503	4.474	10.03	4.384
sigma_v	0.067	0.067	1.957	1.890	1.946	1.896

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses