

Torrance, Samantha; Morrissey, Oliver

Working Paper

Taxation and indigenous institutions in sub-Saharan Africa

CREDIT Research Paper, No. 14/04

Provided in Cooperation with:

The University of Nottingham, Centre for Research in Economic Development and International Trade (CREDIT)

Suggested Citation: Torrance, Samantha; Morrissey, Oliver (2014) : Taxation and indigenous institutions in sub-Saharan Africa, CREDIT Research Paper, No. 14/04, The University of Nottingham, Centre for Research in Economic Development and International Trade (CREDIT), Nottingham

This Version is available at:

<https://hdl.handle.net/10419/126443>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Taxation and Indigenous Institutions in Sub-Saharan Africa

by

Samantha Torrance and Oliver Morrissey

Abstract

This paper contributes to the literature on tax performance in sub-Saharan African countries. A standard model of the determinants of tax revenue is augmented to include measures of indigenous pre-independence institutional structure constructed from anthropological data on the characteristics of ethnic group organisation. We posit that if the three largest ethnic groups characterised by a clan-based organisational structure are a sufficiently large share of the population they are more likely to be able to reach a political consensus that allows a higher revenue to GDP ratio. We find that indigenous institutions have an effect on tax performance in SSA that diminishes over time (as the economy grows and new institutions emerge).

JEL Classifications: H20, O23, O55

Keywords: Tax Revenue, Institutions, sub-Saharan Africa

**Centre for Research in Economic Development and International Trade,
University of Nottingham**



Taxation and Indigenous Institutions in Sub-Saharan Africa

by

Samantha Torrance and Oliver Morrissey

Outline

1. Introduction
2. Background and Literature Review
3. Indigenous Institutional Variables
4. Data
5. Empirical Model and Specification
6. Results and Discussion
7. Summary and Conclusion

The Authors

Samantha Torrance is currently with Adam Smith International stationed in Burundi and was a PhD student at Nottingham during the period the research was undertaken. Oliver Morrissey is Professor of Development Economics in the School of Economics, University of Nottingham. Both are CREDIT Fellows.

Contact e-mail: oliver.morrissey@nottingham.ac.uk

Acknowledgements

This paper was presented at the *Third International Workshop on Mobilising Domestic Revenue in Developing Countries*, DIE, Bonn, 24–25 September 2013. The authors are grateful to participants for comments.

1. Introduction

This paper contributes to the literature on tax performance in sub-Saharan African (SSA) countries. A standard model of the determinants of tax revenue that captures a country's 'tax handle' or 'tax capacity' (Gupta *et al.*, 2004; Clist and Morrissey, 2011; Benedek *et al.*, 2012) is augmented to include a measure of indigenous institutional quality, capturing their pre-independence or pre-colonial features. Mkandawire (2010) provides the initial impetus for this analysis. He classifies SSA countries in three categories reflecting the economic structure during the colonial period (cash crop economies, labour reserve economies and concession companies) and finds evidence that this colonial heritage in part explains systematic differences in tax to GDP ratios across a sample of SSA countries. However, a limitation of this analysis is that these countries are classified into one of three groups and this does not allow for any within-group heterogeneity. We exploit anthropological data on the characteristics of ethnic group organisation to construct country-specific measures of indigenous institutions and assess whether 'institutional organisation' helps to explain cross-country differences in tax performance in SSA.

The new variables that capture institutional quality of an 'indigenous' or pre-independence nature are created from African ethnographic and historical population data, which describe ethnic group characteristics. In addition, the degree of ethnic fragmentation within SSA countries is also accounted for, specifically the degree to which ethnic groups are fragmented across the political boundaries of present-day states. These variables aim to proxy for indigenous institutions or the initial institutional environment, which reflects deeper, more historic factors that are not usually accounted for by more contemporary governance and institutional indicators. This model is applied empirically using revenue data that covers the period 1970-2010. Given the time invariant nature of the indigenous institutional variables that are created, a series of cross-country regressions are estimated for a sample of 36 sub-Saharan African countries. Using these variables we find evidence that indigenous institutions have an effect on tax performance in SSA which diminishes over time. Further institutional organisation tends to dominate effects driven by ethnic group fragmentation. After considering the performance of existing institutional measures within this model, we conclude that these new variables are additional and complementary to the wide-range of institutional variables already being used in empirical analysis.

This paper is organised as follows, section 2 provides an outline of the existing literature and motivation for the paper. Section 3 describes the new institutional variables that are constructed, whilst section 4 provides a summary of other data sources. The model and specification used is covered in section 5. Key results are discussed in section 6 where we show that the effect of indigenous institutional quality has a statistically significant impact on the revenue to GDP ratio of a country and that this effect works predominantly through its interaction with GDP per capita. The share of agriculture in the economy, the ratio of import duties to GDP and in later periods GDP per capita are shown to be robust determinants of tax effort. Section 7 provides summary conclusions and proposed extensions.

2. Background and Literature Review

Institutions and Contemporary Measures

The importance of institutions to economic development was established as early as Adam Smith but the recent popularity in economic analysis has been attributed to North (1990, 1994). Since then the literature has pursued a number of avenues, in particular a focus on property rights as a proxy (McMillan, Rauser and Johnson, 1991; Barro, 1991; Acemoglu, Johnson and Robinson, 2001; Levine, 2005) or on political and democratic institutions (Rodrik, 1999; Persson and Tabellini, 2006; Bardhan, 2005; Besley, Persson and Sturm, 2005). Given the data constraints that are inherent, as well as the issue of dual causality between institutions and economic growth, much of the literature is committed to finding an appropriate instrument for institutional quality. The list of variables includes those that capture ethno-linguistic fractionalisation (Mauro, 1995); differences in law systems (Porta *et al.*, 1997, 1999); percentage of law students (Knack and Keefer, 1997), social infrastructure index (Hall and Jones, 1999); colonial origins (Rodrik, 1999); predicted trade shares (Rodrik *et al.*, 2004) and settler mortality – the seminal contribution of Acemoglu, Johnson and Robinson (2001). However, as Bardhan (2005) points out much attention has been made to finding the perfect instrument and less to actually establishing an adequate and satisfactory causal explanation.

More typically, controls for institutional quality take the form of constructed indices such as those from the World Bank Governance Indicators (Kaufmann, Kraay and Mastruzzi, 2009), the International Country Risk Guide (ICRG, published by Political and Risk Services) and the Polity IV dataset (Marshall, Gurr and Jaggers, 2013) amongst others. However, using composite indicators of institutions can be misleading – they are empirical indices that are limited in temporal coverage, subject to interpretation bias by the compilers and are subject to aggregation problems. These indices (as well as other constructed variables) fail to isolate the causal effect of any single institution, but at the same time cannot include the entire array of institutions that may affect growth for instance, which raises the issue of omitted variable bias (Pande & Udry, 2006). Perhaps a much stronger criticism lies with the fact that most indicators of institutions are invalid as they do not reflect institutions as inputs into production, rather they represent the outcomes that institutions are meant to effect (Glaeser *et al.*, 2004; Fedderke *et al.*, 2011). The variables created for the purposes of this analysis attempt in part to address these shortcomings, in particular the latter.

Of relevance to this analysis, the literature has also attempted to capture the underlying characteristics of indigenous groups, early state-structures and the effect of colonisation, all of which have been shown to be significant in the institutions-growth relationship. These include Morrison *et al.* (1989)'s characterisation of 'state-like' nature; Easterly and Levine's (1997) index of ethno-linguistic fractionalisation; Gennaioli and Rainier (2007)'s measure of political organisation; Michalopoulos and Papaionnou (2011a, 2012)'s spatial distribution of African ethnicities and Easterly and Levine (2012)'s density of colonial European settlement¹.

¹ A number of these measures will be used to assess the robustness of our results and are generally found to be in line with those that are presented in the original papers and support the hypotheses purported here. However, direct comparison of the magnitude of the coefficients cannot be made due to differences in sample coverage and the construction of the variables.

From the tax revenue perspective, in addition to more structural features of the economy, measures of the quality of governance and of political and legal institutions have been integrated into empirical models and been found to have positive and statistically significant effects. These include: political 'voice and accountability' (Bird, Martinez-Vasquez and Torgler, 2007), common law legal systems (Keen, 2012) and parliamentary systems of governance (Persson and Tabellini, 2003).

Pre-Colonial African Institutions

Fortes and Evans-Pritchard (1940) provides the basis for much of the work carried out on classifying political systems and administrative structure in pre-colonial Africa (prior to 1885). Their initial classification of political systems consisted of two main groups: primitive states and those considered stateless. Primitive states were those societies with centralised authority, some form of administrative framework and a judicial system. Social cleavages of wealth, privilege and status corresponded to the distribution of power and authority. In order to ensure stability the powers of the paramount ruler were counter-balanced through the regional devolution of powers and privileges. As Fried (1960) highlights, a state only emerges in stratified societies, but even with the presence of an aristocracy or system of castes in more politically centralised states, the bargaining powers of the masses were strong relative to the elites (Bates, 1983) and there was equality between chiefs and their subjects (Goody, 1971). On the other hand, stateless societies were territorial units not defined by an administrative system, but rather represented local communities which were linked depending upon lineage ties and bonds of direct cooperation. There was no dominant class in the political structure and no organised force. With a lack of a centralised and persistent political authority, these societies tended to be in a state of continual change and instability. However the common acceptance of social values by the members of the society, rather than more formal rules and sanctions worked to prevent widespread conflicts of interest (Middleton and Tait, 1958).

Davidson (1992) shows that some of the largest and most established African societies were characterised by constitutional checks and balances regulating the abuse of power including the devolution of power and pre-colonial states were implicitly distrustful of executive power. The societies that endured were those that continued to ensure their legitimacy through participatory styles of government (Schapera, 1956).

What determined the evolution to primitive statehood? A number of avenues have been pursued including, population density (Stevenson, 1968), regional and long-distance trade (Vansina, 1962; Hodder, 1964), and resource endowments and ecological diversity (Gray and Birmingham, 1970; Bates, 1983; Fenske, 2010). The development of markets and trade appears to be the dominant factor, although once colonisation began in Africa, the control of trade was assumed by the colonising European powers and as a result weakened the basis for political control (Hodder, 1964; Stevenson, 1968; Gray and Birmingham, 1970)².

However, the formation of states and the elaboration of institutions in Africa was very different to that within Europe and this was particularly true of property rights. Crucially, the distinction

² For example the Bemba state rose and prospered through the localised monopoly of the trade in slaves and ivory. When the British, in 1890, colonised that region of Africa, this monopoly ended and the basis of Bemba power was destroyed, both within itself and over other tribes. Simply, the Bemba developed trade, which led to an increased population density, as the British destroyed the basis of the state, this led to migration and thus the population declined (Stevenson, 1968).

between private land and communal land was not made predominantly due to the fact that land was so abundant in Africa and consequently, population density was equally minimal (Goody, 1971; Colson, 1975; Ahene, 2000). Although the relative quality of the land proved an important factor and areas that were more densely populated also benefited from higher quality soil and associated natural resource endowments (Herbst, 2000). In West Africa long-term claims to land seem to have been driven by the mutually incompatible uses envisaged for a particular tract of land (Colson, 1975). In addition, the lack of technological innovation in early African agriculture (Goody, 1971; Diamond, 1997) meant that agriculture remained unproductive and there was little incentive to demand exclusive rights to the use and ownership of land if the economic rents were low (Schapera, 1956; Sundstrom, 1965). African communities only began to alter their views regarding land during the colonial period due to, amongst other factors, the presence of a settling, foreign population.

There is little within this literature specifically related to taxation, but Sundstrom (1965) notes that it does appear as a form of early redistribution of the surplus by the chief whose power was derived from his ability to provide for his people. Equally, the people who chose to pay the taxes and tributes (commonly customs duties and road tolls) levied by the chief acknowledged his over-lordship or protection (Wright, 1999).

Colonial Heritage

A number of studies within the new institutional economics literature aim to account for the effect that colonisation has had on economic growth (La Porta *et al.*, 1999; Engerman and Sokoloff, 2005; Acemoglu, Johnson and Robinson, 2001, 2002). With a more specific focus on tax revenues as the outcome variable of interest, Mkandawire (2010) provides the initial motivation for our analysis. To test the economic relevance and statistical significance of indigenous institutions for economic performance in sub-Saharan Africa (SSA) 35 SSA countries were classified in three groups (*cash crop*, *labour resource* and *concession company economies*) based on their colonial experience.

The *cash crop economies* (17 countries located mainly in West Africa)³ relied on the production for export of cash crops such as palm oil, rubber and cocoa. The presence of colonial governments in these countries tended to be limited in the extent to which the production and logistics for export could be organised. This presence was even more limited in the case of the *concession company economies* (6 countries located around the Congo Basin)⁴. In fact, there was very little colonial government presence, if any, with much of the land in these countries having been licensed to large resource-motivated private enterprises (e.g. mining, forestry, cash-crops), which had complete control over its governance. The *labour resource economies* (12 countries mostly located in Southern Africa)⁵ were characterised by relatively high labour endowments used to facilitate economic activities such as farming and mining. These economies had much larger, more invested colonial governments that tended to replicate the institutions that were present in their home countries.

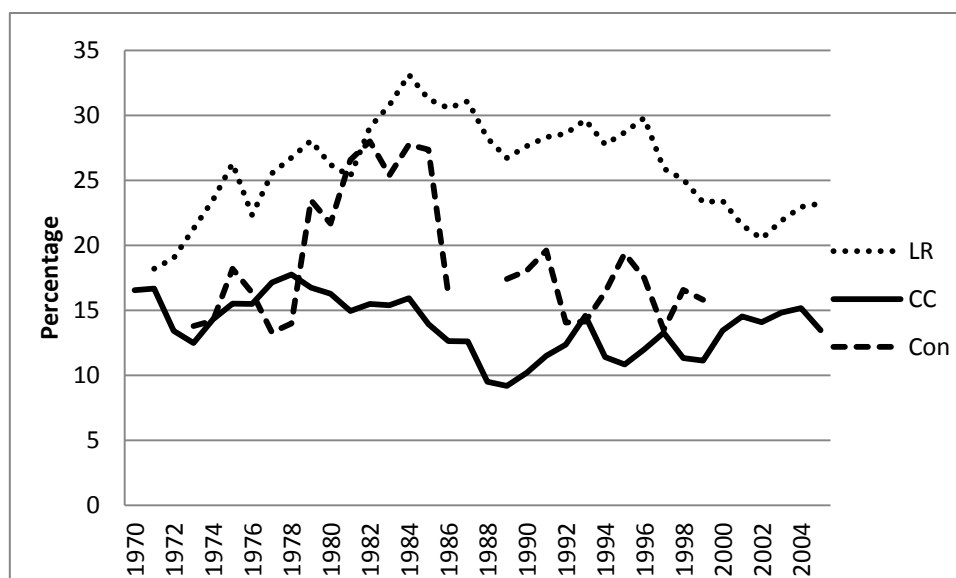
³ *Cash-crop economies*: Benin, Burkina Faso, Cameroon, Côte d'Ivoire, The Gambia, Ghana, Guinea, Guinea-Bissau, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, Tanzania, Togo, Uganda.

⁴ *Concession company economies*: Congo DRC, Congo Rep., Gabon, Central African Republic, Rwanda, Burundi

⁵ *Labour resource economies*: Angola, Botswana, Kenya, Lesotho, Madagascar, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia, Zimbabwe.

Mkandawire's (2010) main finding was that tax revenue to GDP ratios (tax/GDP) tended to be significantly higher in the *labour resource* economies compared to the *cash crop* economies and that this could be attributed to the institutional features inherited from the colonial period (it transpired that the *concession company* group was very disparate with no significant effect). Applying the classification discussed above to our revenue data (see Graph 1) highlights both the disparity in tax revenues observed, as well as an early indication of the effect that colonial heritage seems to have on this initial performance.

Graph 1: Colonial Heritage and Tax Performance

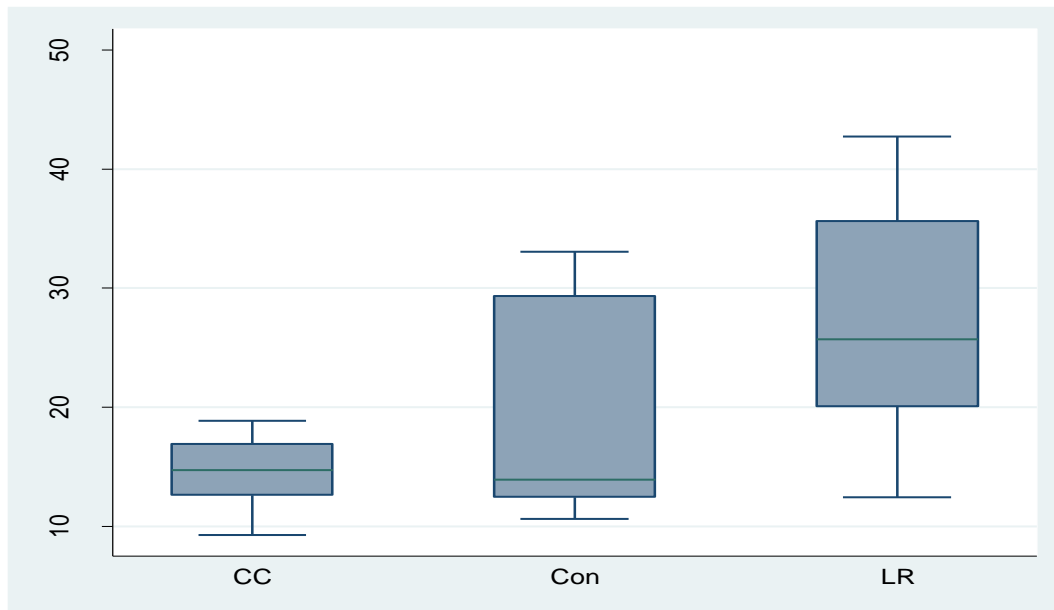


Note: 'LR' – labour resource economies; 'CC' – cash crop economies; 'Con' – concession company economies.

Source: Mkandawire (2010) classification applied to data from Clist & Morrissey (2009).

However, there is still a substantial amount of heterogeneity amongst countries even within the same colonial-economy group, highlighting a limitation of this classification. From Figure 1 one can observe the large variance around the average for concession and labour resource economies, whilst for the cash crop economies a much narrower distribution centred on the median is observed. This paper aims to use newly constructed measures of indigenous institutions to determine whether differences in institutional organisation can explain this heterogeneity in cross-country tax performance.

Figure 1: Box Plot - Revenue to GDP ratio by Colonial Classification, 1970-2010



Note: CC – cash crop economies; Con – concession company economies; LR – labour resource economies. The box plot: whiskers indicate the maximum and minimum, the line in the box is the median and the size of the box indicates the distribution between the 25th and 75th percentile.

3. Indigenous Institutional Variables

One of the key data sources made use of in this study is Murdoch's (1967) *Ethnographic Atlas* – a compilation of ethnographic evidence used to classify 862 societies across the world, including 239 in Africa. The Murdoch dataset provides a wealth of information equating to approximately 60 variables that capture societal, economic and political characteristics of ethnic groups, although with varying data. As would be expected, larger or more prominent groups are better represented in the data. For the purposes of this analysis, the sample is limited to just sub-Saharan African (SSA) societies and covers the period of their colonisation, prior to their independence. Many of the surveys used to compile this data were undertaken as a form of early census carried out by colonial powers, as well as those with more academic motivations – some as early as 1870. In addition, Murdoch's (1959) *Ethno-linguistic Map* provides a visual distribution of the ethnic groups in Africa.

Combining both the map-based representation and the geographic coordinates provided by Murdoch (1967), we use geographic information system (GIS) software to determine how these societies relate to the contemporary countries of Africa (see Figure 2). In addition to observing the groups and societies that constitute the nation-states of Africa, we are able to determine which groups are fully encapsulated by a country's national border or whether they overlap with other neighbouring countries. This distinction has proved important in the literature as societies that are fragmented or fragmented across borders.

Figure 2: Ethnolinguistic Map (Murdoch, 1959) and Current Political Boundaries of Africa



From the wealth of information that the Murdoch dataset provides interest is focused on those factors which are believed to provide the best indication of the indigenous institutional environment for state formation and taxation, namely: community organisation and the nature of settlement. This analysis focuses on just two newly variables: clan-based organisation and permanent settlement, both benefiting from better coverage in the original data sources and allowing for sufficient cross-country variation for empirical analysis⁶. Dummy variables are created based on the classification information provided by Murdoch (1969). Clan-based organisation is determined from the community organisation variable. The presence of a clan structure implies a more formal hierarchy with some form of centralised decision making. Settlement distinguishes between whether societies are considered migratory or settled. The interest in this variable is centred on the fact that if societies are less transitory then there is more opportunity for institutional structures to develop. Further settlement enables agricultural cultivation and the storage of food, allowing for a diversification of labour away from hunting and gathering, and thus facilitates the development of a bureaucracy. Intuitively, there is a degree of correlation between these two variables and this is established statistically⁷.

⁶ It is recognised that the data may be subject to sample selection bias as there is a tendency for the larger, as well as the more fragmented groups to be better represented. It is not clear whether the missing data is simply not available or actually not relevant for the group in question. As a result, focus is directed towards those groups for which there is adequate data and one can view it with some certainty.

⁷ Pairwise correlation coefficient of 0.45 significant at the 5 per cent level.

In addition to the Murdoch (1967) dataset, the Atlas Narodov Mira (1964) from the Miklukho-Maklai Institute of Anthropology and Ethnography⁸ provides population data on world ethnic groups organised by countries (correct at the time of publication)⁹. This Atlas allows for the geographic dispersion of ethnic groups to be validated and also permits the construction of population weighted variables for empirical analysis.

As mentioned earlier, the distinction between fragmented and non-fragmented groups is one that has been investigated within the literature with measures typically identifying the degree of ethnic fragmentation *within* a country, i.e. the number of different groups and the subsequent population distribution (Easterly and Levine, 1997; Posner, 2004). In this analysis, the focus is somewhat different, with fragmentation *across* national borders being of interest; reflecting the arbitrariness of colonially-imposed political borders and the ensuing implications for state legitimacy (Englebert, 2000).

These variables are selected on the basis that those groups or societies that had more experience with formalised organisational structures were better able to adopt Western-style governance following the colonial period and/or a greater ability for the consensus-building necessary for the formulation of government policy, particularly in relation to taxation. A lack of state legitimacy due to the fragmentation of groups across borders renders government policymaking somewhat redundant if barriers to implementation result due to the non-acceptance of the government and their policies.

Our overall analysis is at the country-level and thus the data collated on the ethnic groups requires aggregating and this is done by constructing concentration ratios¹⁰. Formally:

$$CR_n = \sum_{i=1}^n S_i$$

Where S_i is the population share of society i (population of society i divided by the total population of the country) and n defines the largest societies (by share). The concentration ratio for the 3 largest societies is calculated. The three variables/concentration ratios constructed represent the 3 largest societies that are¹¹:

- characterised by permanent settlement (*set*);
- characterised by a clan-based community structure (*clan*); and
- fragmented across national borders (*frag*).

Whilst the concentration ratio method provides us with the means to create country-level variables, the subsequent analysis of these ratios is more complicated, particularly if attempts

⁸ Given the rarity of this particular Atlas, gathering this data involved a trip to the British Library, St. Pancras, London as well as transcribing and translating from the original Russian to English. The English translation of the Atlas Narodov Mira by Telberg (1965) was consulted but did not include a translation of the section detailing the population data by society.

⁹ A number of steps were taken to validate the data between Murdoch (1967) and the Atlas Narodov Mira (1964), available on request from the authors.

¹⁰ A number of other methods of aggregation were also considered, including constructing a Herfindahl Index, using principal component analysis (PCA) and simple population shares of various combinations of the characteristics. These methods are not reported given that they performed poorly in subsequent empirical analysis and in some circumstances suffered from data availability concerns.

¹¹ See Appendix 1 for the values of these concentration ratios by country.

are made to provide inference regarding the level of political competition and thus potential conflict within the respective countries¹². On one hand, the hypothesis applied is that higher population proportions of similarly organised groups increases the likelihood of consensus-building within the country. However, on the basis that the three largest ethnic groups are used to create the ratio, this also suggests a higher likelihood of dominance by one or a few ethnic groups within the country as a whole. In a society characterised by fewer groups, there may be more opportunity for consensus and thus less conflict. But whether or not this is the case in practice can be disputed, as has been evidenced in a number of African countries, when the presence of a few dominant ethnic groups have led to internal conflict, e.g. Rwanda (Hutus and the Tutsis), Kenya (Odinga and the Kikuyu), South Sudan (Lou Nuer and the Murle) and Liberia (Gio /Mano and the Krahn). It may be the case that more groups indicates more competition, diluting the potential for conflict, or that a more fragmented society may create problems of state legitimacy and thus have a negative overall effect on the institutional environment, as suggested by Englebert (2000). Certainly, Easterly and Levine (1997), Posner (2004) and Campos, Saleh and Kuzeyev (2009) all show that the degree of ethnic fractionalisation has a negative effect on economic outcomes¹³.

4. Data

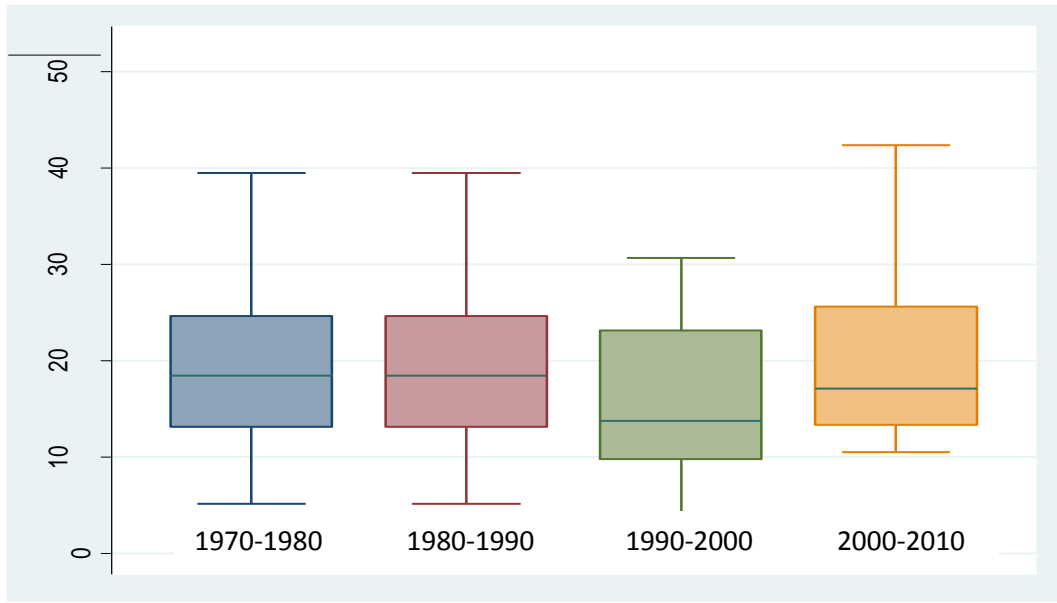
A sample of 36 sub-Saharan African countries is used with control variables capturing features important to determining the 'tax handle' or 'tax capacity' of developing countries. The share of agriculture (*agr*) and industry (*ind*) in the economy, GDP per capita (*GDPpc*) and the share of imports (*imports*) and exports (*exports*) as a percentage of GDP are sourced from the World Bank's World Development Indicators. Government revenue data is from the IMF's Government Financial Statistics database¹⁴. Full descriptions of the variables and descriptive statistics are provided in Appendix 2 and Appendix 3 respectively.

Figure 3 shows the evolution over time of government revenue as a share of GDP for the sample over four cross-sections. There is little change in the median and distribution between the 25th and 75th percentile, as well as overall variance between the two earlier periods (1970-1980 and 1980-1990). However, in the period 1990-2000, the variance of the distribution significantly reduces, as does the median value – potentially reflecting the actual implementation of tariff reductions through the World Trade Organisation. In the most recent period though, whilst the median tax share is lower than in the first two periods, the distribution of tax revenue shares between the 25th and 75th percentile has narrowed and is similar to what was observed earlier.

¹² In addition, the variables fail to capture other underlying factors which may have had a subsequent effect on the performance of contemporary institutions. These include the geographic location of these groups, their political engagement during the colonial period, the effect of nationalisation and state-building and the degree to which the blood or ethnic ties have been diminished over time, through migration, marriage or other socio-political influences.

¹³ Whilst Easterly and Levine (1997) treat ethnic fractionalisation as an exogenous, static variable, Campos, Saleh and Kuzeyev (2009) treat ethnicity as an endogenous variable that changes over time – they find a robust, negative correlation to economic growth over the period 1989-2007.

¹⁴ Much of this data was compiled by Clist and Morrissey (2009) and subsequently updated by the authors.

Figure 3: Box Plot - Revenue/GDP by Period

The box plot: whiskers indicate the maximum and minimum values, the line in the box is the median and the size of the box indicates the distribution between the 25th and 75th percentile.

Simple scatter plots suggest a positive correlation between the new indigenous institutional variables and the revenue to GDP ratio (see Appendix 4), the relationship being strongest for the variable *clan*, followed by *set* and *frag*. In addition, from observation it appears that the variable *frag* may be characterised by two sub-samples, where countries cluster both above and below some threshold value. This may also be the case for the variable *clan*, with clustering at lower levels. Potential threshold effects were investigated and reported in Section 6. Crucially, these scatter plots highlight the considerable heterogeneity that is present when comparing this sample of countries¹⁵.

5. Empirical Model and Specification

This analysis uses a standard model of the determinants of tax revenue (Gupta *et al.*, 2004; Clist and Morrissey, 2009; Benedek *et al.*, 2012) which seeks to capture a country's 'tax handle' or 'tax capacity'. Specifically, the ratio of current revenue to GDP is determined by the current structure of the economy – so a contemporaneous rather than a causal relationship.

$$\ln\left(\frac{rev}{GDP}\right) = \beta_0 + \beta_1 GDPpc + \beta_2 GDPpc^2 + \beta_3 agr + \beta_4 ind + \beta_5 imports + \beta_6 exports + \beta_7 I + (1)$$

Equation 1 shows this basic model, with the addition of a vector *I* for the indigenous institutional variables. The dependent variable is the natural log of the ratio of revenue to GDP;

¹⁵ For instance, Botswana has consistently higher revenue than would be predicted from the indigenous institutional variables (i.e. 'above the line') and this is likely attributed to the favourable agreement that the government agreed on with regards to diamond revenues. In addition, Angola would appear to benefit from the oil revenues and Lesotho, from the fact that the economy is almost solely driven by aid.

logs are taken in order to correct for the skew in the distribution of the data and is common practice in the literature (Gupta *et al.*, 2004; Clist and Morrissey, 2009). Income is measured by GDP per capita (*GDPpc*) in current US dollars and expected to have a positive coefficient ($\beta_1 > 0$), the larger the economy, the better off its citizens, the higher the expected revenues from taxation (Musgrave, 1969), as well as the higher the demand for public services (Tanzi, 1987). In addition it is also often taken as a proxy for administration and compliance capacity¹⁶. The quadratic effect of GDP per capita on revenue ($GDPpc^2$) takes into account the fact that as the economy increases, its effect on revenue will increase, but at a decreasing rate ($\beta_2 < 0$)¹⁷. *Agr* and *ind* are the shares of agriculture and industry in the economy as a percentage of GDP, respectively. Given the economic structure of sub-Saharan African countries it is expected that these two sectors capture the majority of taxable productive output. As agriculture tends to be organised on a more informal, subsistence basis, a negative relationship with tax revenues ($\beta_3 < 0$) is anticipated as collection and enforcement of tax policy is difficult. The opposite is true for the industrial sector which is more formal and urban-based, thus a positive relationship with tax revenue ($\beta_4 > 0$) is expected. Trade taxes have historically been a dominant contributor to government revenues in SSA, and in spite of liberalisation, remain important¹⁸. As a result we include imports (*imports*) and exports (*exports*) as shares of GDP and we expect the coefficients to be positive ($\beta_5 > 0$; $\beta_6 > 0$)¹⁹.

The institutional vector, *I*, relates to the constructed indigenous institutional variables, *cent*, *clan* and *frag*. These variables will be used individually and sequentially in the empirical analysis. It is expected that the revenue to GDP ratio will be positively correlated with *cent* - formal organisation allowing for political consensus when it comes to raising revenue. For *clan*, the expected direction of the sign of the coefficient is less clear. Being clan-based inherently implies some organisational structure and thus a positive effect would be expected. But equally, the presence of clans could imply an increased likelihood of conflict. Similarly, a negative sign could be expected in the case of *frag* - a higher degree of ethnic fragmentation implying that political consensus is harder to attain and thus a negative effect on tax revenues. In fact, it could be the case that the coefficients on all of the variables in *I* are negatively signed - on the basis that if indigenous institutional arrangements persist, they could provide a source of conflict with more modern, contemporaneous institutions. Furthermore, the nature of the relationship between the indigenous institutional variables and revenues may differ depending on the value of the observation in relation to some threshold value.

¹⁶ Wagner's Law implies that the share of government increases as income levels increase, however a number of studies find evidence of an income elasticity of less than 1 and sometimes even negative. In these cases it is possible that the distortionary costs of taxation result in lower revenue shares at higher incomes (Keen & Lockwood, 2010).

¹⁷ Should the signs on *GDPpc* and *GDPpc*² be consistent with these expectations, then that would add further support the argument that low income countries are already employing the greatest possible effort in terms of raising domestic revenue (Keen and Simone, 2004). Revenue effort will be considered in more detail in Section 5.

¹⁸ The average applied tariff in 2007 in low-income countries was 12 per cent in comparison to the global average of 8.8 per cent - UNCTAD TRAINS database http://r0.unctd.org/trains_new/database.shtml

¹⁹ Although the direction of the sign on trade variables is ambiguous in the literature, imports and exports are easy to tax as they take place in specific locations, and higher openness leads to larger revenues. But low protection levels (so lower tariff rates) may increase openness and thus, taxes and openness are inversely related. Clist and Morrissey (2009) demonstrate the validity of disaggregating the trade share to GDP in terms of imports and exports, and in their regressions find statistical evidence of an opposing effect on tax revenues i.e. imports being positive and exports negative.

The model is estimated using data over the years 1970-2010 organised into 10 year averages, which results in four cross-sections relating to the periods 1970-1980, 1980-1990, 1990-2000 and 2000-2010. An ordinary least squares (OLS) estimator with robust standard errors (to control for heteroskedasticity) is used. As the institutional variable is time-invariant panel methods are not employed (*I* would be equivalent to a fixed effect). Using cross-section averages however, does have its benefits: first, it smoothes data volatility in the annual observations, a characteristic of SSA data, and accounts for missing annual observations for some years. Second, estimating a succession of cross-sections allows an investigation into whether the significance and magnitude of the effects of indigenous institutions is dynamic. The hypothesis being that as contemporary institutions develop the effect of indigenous institutional structures on tax performance will decrease over time. This is due to gradual and organic institutional change, as well as more concerted reforms being implemented. In addition, as economies become richer and more integrated into the global economy, traditional or indigenous structures are likely to be increasingly diluted.

Given the lack of availability of consistent tax-specific data for the countries in this sample, analysis is focused on revenue, excluding grants, at the central government level; however tax-specific data (excluding revenues from resource taxes) is made use of in robustness tests carried out, noting that there is a shorter time-series beginning in 1980. Equation 1 is a reduced form version of models that have been applied in the literature and may be subject to omitted variable bias. However, as noted by Clist and Morrissey (2009), over 20 control variables were analysed as potential determinants to the ratio of tax to GDP by Gupta (2007) and few others in addition to those included here were found to be statistically significant²⁰. Whilst the full model as specified by equation 1 is estimated, a more parsimonious approach is undertaken following recognition of the relatively small sample sizes of the cross-sections. The results below show that there is little loss in the explanatory power of the model.

6. Results and Discussion

Estimation of the full model as shown in equation 1 (omitting the institutional variable at this stage) yields expected results: agriculture (*agr*) has a negative and statistically significant effect in three out of the four cross-sections, with the magnitude of the coefficient diminishing in the most recent period²¹. The industrial share of the economy (*ind*) is statistically insignificant across specifications. GDP per capita, the proxy for the size of the economy, is statistically significant only in the last period (2000-2010), the effect of which is positive and quadratic in nature, implying that the positive contribution to revenues from the size of the economy is subject to diminishing returns. The coefficient of the variable capturing import share (*imports*) is positive and statistically significant across specifications, highlighting the importance of trade taxes to government revenue. Exports are statistically insignificant across all cross-sections²².

In order to conserve degrees of freedom in our empirical estimations, the variables *ind* and *exports* are dropped and a more parsimonious model pursued. The same pattern of results is

²⁰ Other variables include population size and population growth, inflation, public debt, financial sector depth, extent of decentralisation, aid and natural resources (IMF, 2011).

²¹ Results available on request from the authors.

²² When re-estimated using cross-sections averaged over 1970-2010, 1970-1990 and 1990-2010, the same pattern of results is evident (available on request).

observed, with the coefficients on *agr* and *imports* significant at the 1 per cent level and these are robust to specification. There is no real increase in the explanatory power of the model (given by the R^2); however the F statistics of joint significance of the variables has increased, supporting the model.

Using this parsimonious model, the indigenous institutional variables are then included in the model individually and sequentially²³, and then interacted with GDP per capita and its quadratic term. Regressions with the variable *set* (capturing permanent settlement) show in the period 1970-1980, the coefficient of the variable *set*GDPpc* is positively signed and statistically significant – the effect of which is quadratic (see Table 1). This is also the case in the cross-section 1990-2000 where, the individual variable *set* is also statistically significant and positive. This implies that higher concentrations of groups characterised by permanent settlement are associated with higher revenue to GDP ratios, however as the respective country grows, this institutional or ‘settlement’ effect diminishes.

The concentration of groups characterised by clan-based organisational structures (*clan*) is statistically significant in the first two cross-sections (1970-1980 and 1980-1990) and this is despite relatively small sample sizes (see Table 2). The variable itself is positive, whilst the coefficient on the interaction term with GDP per capita (*clan*GDPpc*) is negative and quadratic. However, in the most recent period (2000-2010), *clan* and its interaction terms are statistically insignificant. This result seems to support the hypothesis that whilst indigenous institutions do have a role to play in determining revenues, their effect does diminish due to time itself, as well as a due to the growth of the economy.

Lastly, results using *frag* show the variable itself to be positive and statistically significant in the first cross-section as are the interaction variables (signs as expected), whilst the latter are also statistically significant in the period 1980-1990 and 2000-2010 (see Table 3). The positive sign on the *frag* variable, which at first appears contrary to expectations, can be explained. Given *frag* is the fragmentation across borders – and taking into account that this is true for the majority of groups in the sample – the variable itself can be considered a simple concentration ratio i.e. the larger its value the smaller the degree of the fragmentation of groups *within* the country. Thus, a positive sign would support the hypothesis that a lower degree of internal fragmentation, with the majority of the population being represented by fewer groups has beneficial effects on domestic revenue generation due to the increased ability to build political and policy consensus.

²³ The indigenous institutional variables are correlated and thus not included in the model together.

Table 1: Parsimonious Model with Institutional Variable – set

	1970-1980		1980-1990		1990-2000		2000-2010	
	1	2	3	4	5	6	7	8
agr	-0.017 (3.05)***	-0.015 (3.02)***	-0.012 (1.50)	-0.011 (1.54)	-0.023 (3.15)***	-0.018 (3.49)***	-0.009 (3.14)***	-0.009 (2.75)**
GDPpc	-0.008 (0.69)	0.153 (2.04)*	0.042 (1.34)	0.197 (1.89)*	0.016 (0.68)	0.227 (3.53)***	0.018 (3.48)***	0.049 (1.95)*
GDPpc ²	2.11e ⁻⁰⁸ (1.17)	5.515e ⁻⁰⁸ (1.05)	-6.34e ⁻⁰⁸ (1.16)	-5.83e ⁻⁰⁷ (1.05)	-2.39e ⁻⁰⁸ (0.54)	-5.84e ⁻⁰⁷ (2.99)***	-1.16e ⁻⁰⁸ (2.91)***	5.89e ⁻⁰⁸ (1.45)
imports	0.015 (6.58)***	0.016 (4.69)***	0.011 (5.22)***	0.012 (4.42)***	0.008 (2.79)***	0.014 (5.16)***	0.011 (4.73)***	0.011 (5.39)***
set	0.000 (0.06)	0.009 (1.51)	0.000 (0.09)	0.012 (1.57)	0.001 (0.27)	0.015 (2.69)**	0.001 (0.61)	0.003 (1.55)
set*GDPpc		-0.003 (2.11)**		-0.003 (1.59)		-0.004 (3.40)***		-3.96e ⁻⁰⁶ (1.36)
set*GDPpc ²		-1.39e ⁻⁰⁹ (0.93)		1.09e ⁻⁰⁸ (0.95)		1.05e ⁻⁰⁸ (2.91)***		5.56e ⁻¹⁰ (1.28)
_cons	2.850 (9.62)***	2.150 (4.54)***	2.600 (5.63)***	1.932 (3.41)***	2.963 (7.86)***	1.842 (4.86)***	2.507 (13.37)***	2.307 (11.09)***
<i>F</i>	24.24	65.89	20.54	37.04	13.93	111.90	72.40	62.21
<i>P</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>R</i> ²	0.69	0.75	0.67	0.73	0.69	0.79	0.82	0.82
<i>N</i>	28	28	30	30	32	32	35	35

Notes: Dependent Variable is ln(rev/gdp); OLS estimator with robust s.e.; 10 year average cross-sections; * denotes statistical significance: * p<0.1; ** p<0.05; *** p<0.01. Test of joint significance of variables (F), associated p-value (P); R² is the coefficient of determination. Coefficient of GDPpc rescaled: multiplied by 100.

Table 2: Parsimonious Model with Institutional Variable – *clan*

	1970-1980		1980-1990		1990-2000		2000-2010	
	1	2	3	4	5	6	7	8
agr	-0.015 (2.31)**	-0.011 (1.31)	-0.010 (1.17)	-0.008 (0.86)	-0.018 (2.70)**	-0.016 (2.33)**	-0.008 (2.35)**	-0.007 (1.81)*
GDPpc	-0.007 (0.55)	0.098 (1.58)	0.053 (1.65)	0.177 (2.72)**	0.028 (1.18)	0.047 (0.86)	0.019 (3.78)***	0.030 (1.98)*
GDPpc ²	2.60e ⁻⁰⁸ (1.29)	-9.43e ⁻⁰⁸ (0.94)	-7.93e ⁻⁰⁸ (1.45)	-5.36e ⁻⁰⁷ (2.50)**	-4.11e ⁻⁰⁸ (0.96)	-2.57e ⁻⁰⁸ (0.15)	-1.26e ⁻⁰⁸ (2.87)***	-2.96e ⁻⁰⁸ (1.12)
imports	0.015 (7.33)***	0.016 (7.89)***	0.011 (5.11)***	0.011 (5.51)***	0.009 (3.00)***	0.012 (4.89)***	0.010 (3.24)***	0.010 (3.14)***
clan	0.002 (0.64)	0.038 (1.94)*	0.001 (0.32)	0.022 (2.18)**	-0.002 (0.69)	0.014 (1.53)	0.002 (0.72)	0.004 (1.54)
clan*GDPpc		-0.011 (1.88)*		-0.005 (2.59)**		-0.002 (1.77)*		-2.42e ⁻⁰⁶ (0.99)
clan*GDPpc ²		1.43e ⁻⁰⁸ (1.73)		1.48e ⁻⁰⁸ (2.80)**		2.45e ⁻⁰⁹ (0.69)		2.69e ⁻¹⁰ (1.17)
_cons	2.678 (9.66)***	2.165 (4.92)***	2.411 (5.24)***	1.860 (3.32)***	2.786 (9.42)***	2.447 (7.49)***	2.486 (17.82)***	2.395 (15.67)***
F	16.09	28.59	17.90	20.10	13.53	12.50	65.91	33.52
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R ²	0.74	0.79	0.73	0.81	0.71	0.77	0.82	0.83
N	23	23	25	25	27	27	30	30

Notes: Dependent Variable is ln(rev/gdp); OLS estimator with robust s.e.; 10 year average cross-sections; * denotes statistical significance: * p<0.1; ** p<0.05; *** p<0.01. Test of joint significance of variables (F), associated p-value (P); R² is the coefficient of determination. Coefficient of GDPpc rescaled: multiplied by 100.

Table 3: Parsimonious Model with Institutional Variable – *frag*

	1970-1980		1980-1990		1990-2000		2000-2010	
	1	2	3	4	5	6	7	8
agr	-0.022 (3.47)***	-0.025 (3.41)***	-0.014 (1.95)*	-0.013 (1.82)*	-0.021 (3.01)***	-0.022 (3.27)***	-0.009 (3.18)***	-0.008 (2.67)**
GDPpc	-0.013 (0.95)	0.022 (1.15)	0.041 (1.38)	0.137 (2.57)**	0.026 (1.06)	0.098 (1.43)	0.018 (3.96)***	0.051 (2.80)***
GDPpc ²	1.75e ⁻⁰⁸ (0.94)	3.18e ⁻⁰⁸ (0.80)	-6.76e ⁻⁰⁸ (1.26)	-3.94e ⁻⁰⁷ (2.52)**	-4.33e ⁻⁰⁸ (0.95)	-2.16e ⁻⁰⁷ (1.09)	-1.12e ⁻⁰⁸ (3.29)***	-6.37e ⁻⁰⁸ (2.10)**
imports	0.012 (4.46)***	0.014 (4.66)***	0.009 (4.31)***	0.010 (4.52)***	0.007 (2.23)**	0.011 (3.84)***	0.011 (4.64)***	0.012 (5.35)***
frag	0.003 (1.33)	0.007 (2.05)*	0.002 (0.99)	0.006 (1.50)	0.002 (0.76)	0.006 (1.42)	-0.000 (0.22)	0.002 (1.27)
frag*GDPpc		-9.41e ⁻⁰⁹ (2.39)**		-1.39e ⁻⁰⁵ (1.92)*		-1.47e ⁻⁰⁵ (1.49)		-4.03e ⁻⁰⁶ (2.16)**
frag*GDPpc ²		1.13e ⁻⁰⁹ (1.51)		4.34e ⁻⁰⁹ (2.30)**		3.23e ⁻⁰⁹ (1.24)		6.02e ⁻¹⁰ (1.92)*
_cons	2.951 (12.68)***	2.867 (9.39)***	2.634 (6.75)***	2.288 (5.45)***	2.849 (9.12)***	2.530 (6.86)***	2.528 (17.08)***	2.283 (12.88)***
<i>F</i>	24.00	12.72	23.03	16.02	13.81	21.56.	69.10	37.19
<i>P</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
<i>R</i> ²	0.72	0.75	0.68	0.73	0.69	0.74	0.80	0.82
<i>N</i>	29	29	31	31	33	33	36	36

Notes: Dependent Variable is ln(rev/gdp); OLS estimator with robust s.e.; 10 year average cross-sections; * denotes statistical significance: * p<0.1; ** p<0.05; *** p<0.01. Test of joint significance of variables (F), associated p-value (P); R² is the coefficient of determination. Coefficient of GDPpc rescaled: multiplied by 100.

Table 4: Threshold Effects, Institutional Variable - *frag*

	1970-1980		1980-1990		1990-2000		2000-2010	
	1	2	3	4	5	6	7	8
agr	-0.022 (3.83)***	-0.020 (2.96)***	-0.014 (1.91)*	-0.013 (1.99)*	-0.020 (2.90)***	-0.020 (2.99)***	-0.009 (3.18)***	-0.008 (2.78)***
GDPpc	-0.015 (1.11)	0.031 (0.58)	0.043 (1.43)	0.149 (3.15)***	0.024 (0.98)	0.116 (3.19)***	0.018 (4.02)***	0.047 (3.20)***
GDPpc ²	2.17e ⁻⁰⁸ (1.21)	2.79e ⁻⁰⁹ (0.05)	-7.04e ⁻⁰⁸ (1.30)	-4.30e ⁻⁰⁷ (3.17)***	-3.86e ⁻⁰⁸ (0.82)	-2.93e ⁻⁰⁷ (2.94)***	-1.13e ⁻⁰⁸ (3.42)***	-6.05e ⁻⁰⁸ (2.41)**
imports	0.012 (4.62)***	0.013 (4.03)***	0.009 (4.44)***	0.009 (4.18)***	0.008 (2.48)**	0.010 (3.56)***	0.011 (4.45)***	0.011 (4.67)***
frag_thres	0.236 (1.54)	0.397 (1.56)	0.171 (1.22)	0.671 (2.36)**	0.064 (0.46)	0.510 (2.21)**	-0.014 (0.18)	0.180 (1.69)
frag_thres* GDPpc		-5.08e ⁻⁰⁴ (0.94)		-1.42e ⁻⁰³ (2.97)***		-1.38e ⁻⁰³ (2.94)***		3.29e ⁻⁰⁴ (2.40)**
frag_thres* GDPpc ²		2.47e ⁻⁰⁸ (0.42)		4.34e ⁻⁰⁷ (3.18)***		3.45e ⁻⁰⁷ (2.95)***		5.26e ⁻⁰⁸ (2.15)**
_cons	3.048 (13.63)***	2.760 (7.04)***	2.648 (6.86)***	2.268 (5.85)***	2.866 (9.01)***	2.504 (7.95)***	2.521 (18.27)***	2.327 (15.78)***
F	24.30	26.06	25.83	121.61	14.26	18.92	67.23	27.15
P	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
R ²	0.73	0.74	0.69	0.76	0.69	0.74	0.80	0.82
N	29	29	31	31	33	33	36	36

Notes: Dependent Variable is ln(rev/gdp); OLS estimator with robust s.e.; 10 year average cross-sections; * denotes statistical significance: * p<0.1; ** p<0.05; *** p<0.01. Test of joint significance of variables (F), associated p-value (P); R² is the coefficient of determination. Coefficient of GDPpc rescaled: multiplied by 100. Threshold value 45.07.

Robustness Tests

A number of robustness checks are undertaken to validate the main results presented above²⁴.

i. Fragmentation

We redefine the other indigenous institutional variables to take into account fragmentation in addition to the institutional characteristics they capture. Thus, the variable *set_f (clan_f)* is redefined as the concentration ratio of the three largest societies that are characterised by permanent settlement (clan-based structure) and fragmented across borders. The results are qualitatively similar to those reported above; *clan_f* is generally positive and statistically significant, whilst *set_f* is insignificant, although the interaction terms are statistically significant and negative in sign. These results support the fact that given that most ethnic groups included in the Murdoch (1967) dataset are fragmented across borders, the organisational characteristics of the groups are what is important to subsequent revenue collection.

ii. Threshold Effects

To investigate whether the institutional effects observed are determined by the actual value of the institutional variables themselves, an analysis of potential thresholds is undertaken. This is done by assuming a quadratic distribution to the indigenous institutional variables, from which the turning points, or threshold values are calculated²⁵. Using these thresholds dummy variables are created; equal to 1 where the respective country's observation is greater than the threshold and equal to zero otherwise. In addition they are also interacted with GDP per capita.

The only variable that is statistically significant is *frag_thres*, which is positive in the period 1980-1990 and 1990-2000 (see Table 4 above). The magnitude of the coefficient being greater than those previously observed, although still decreasing over time. This suggests that countries where the three largest ethnic groups command a higher proportion of the population, and where this proportion is greater than the threshold of 45.07 per cent, countries exhibit higher revenue to GDP ratios - an important result in explaining the observable heterogeneity in revenue performance in the sample. But again as the economy grows, these indigenous institutional factors have a reduced effect on revenues.

iii. Tax-specific data

As mentioned previously, tax-specific data of an adequate quality is somewhat lacking, and, observations are only available from 1980 and for a smaller sample of countries. Tax revenues are one component of the sources of domestic revenue, and as such the mean value for our sample is 13.5 per cent of GDP, in comparison to 19.66 for general revenue. The variance in observations is also less. Tax revenue to GDP ratios follow a similar pattern to total government revenue as discussed earlier, with a general decline in the period 1990-2000 and subsequent recovery.

Focusing on the three 10 year cross-sections from 1980, the parsimonious tax determinants model is re-estimated using the natural log of the tax to GDP ratio (*Intaxgdp*) as the dependent variable with and without the institutional variables. From the parsimonious model (without institutional variables) *imports* is the only statistically significant control variable – positive in effect and increasing in magnitude over time (similar to the estimations with *lnrevgdp* as the dependent variable).

²⁴ Results available on request from the authors.

²⁵ For frag this is 45.07; clan 45.92; set 56.94.

On including the indigenous institutional variables and the interaction terms, only *set* is robustly significant, the interaction term (*set*GDPpc*) is negatively signed and quadratic in effect across all three periods (although the magnitude of its contribution to explaining tax revenues is small). With regards to the other institutional variables, *frag* is statistically insignificant and *clan* displays some significance, signs as expected, in the first cross-section only.

iv. Revenue Effort

In a large proportion of the literature, tax effort and tax performance are used interchangeably. However, there are studies that approach tax effort more accurately. These papers calculate tax effort by creating a ratio between actual and predicted tax ratios, which reflects the variance in taxable capacity of a country (Lotz & Morss, 1967; Leuthold, 1991; Teera and Hudson, 2004; Wang *et al.*, 2009). In this way, one can make an assessment of the extent to which revenues are being fully exploited by the country given the tax capacity of economy²⁶. As such, a tax effort ratio of less than unity (<1) implies that the country is exploiting its tax capacity less than the average, whilst a ratio of greater than unity (>1) suggests that the country is exploiting its tax capacity greater than the average. Using this ratio in conjunction with the actual tax/GDP performance data, should a country have a high tax effort ratio (i.e. greater than unity) as well as a high actual tax/GDP ratio then this implies that the country may have limited opportunities to increase tax revenues further²⁷.

Within this literature a structural model of the determinants of tax revenue is estimated similar to equation 1, and often including national debt and government expenditure to GDP, population growth and measures of institutional quality (Bird *et al.*, 2005). Following the estimation of the model, the predicted value of the tax to GDP ratio is then calculated. This method is implemented here; again general revenue rather than tax revenue is our dependent variable.

In order to generate the predicted revenue to GDP ratio, equation 2 below is estimated:

$$\frac{rev}{GDP} = \beta_0 + \beta_1 agr + \beta_2 ind + \beta_3 GDP + \beta_4 GDP^2 + \beta_5 imports + \beta_5 exports + \varepsilon \quad (2)$$

Thus the revenue effort of a particular country (reported in Appendix 4) is given by:

$$Revenue\ Effort = \frac{revenue/GDP}{predicted\ revenue/GDP}$$

Appendix 6 shows a box and whisker plot of the distribution and dynamics over time of the revenue effort ratios for each country over the 10 year cross-sections. Whilst the median value of the index for the sample was around unity in the first two periods, in the years between 1990 and 2000 this had reduced with an associated narrowing in the distribution between the 25th

²⁶ IMF (2011) uses a stochastic production frontier approach to model revenue as a function of exogenous variables and policy choices, where effort is statistically an error term. The resulting indices are positively correlated with ours

²⁷ As Chelliah, Baas and Kelly (1975) state, tax effort indices are designed to be used as complementary evidence for the analysis of fiscal performance of a country and highlights whether there is scope to raise tax revenue further.

and 75th percentile. In the most recent period, 2000-2010, the distribution of the index had narrowed and the median had increased towards unity, suggesting improved revenue effort and a reduction in the variation amongst countries, however there were still a number of countries whose revenue effort ratios were below the average.

Simple bivariate regressions of the indigenous institutional variables on the revenue effort do not lead to statistically significant results. This suggests, that indigenous institutional characteristics, whilst helping to explain revenue performance, do not seem to play a role in determining revenue effort (although this analysis is severely constrained by the small number of observations).

7. Summary and Conclusion

Whilst the relationship between the quality and nature of institutions and economic outcomes has been previously established in the literature, the question on what has determined or influenced the performance of these contemporaneous institutions has received less attention. This analysis provides some insight to addressing the latter, and in particular, despite common colonial heritage and thus imposition of similar institutions, why countries in sub-Saharan Africa exhibit heterogeneous outcomes.

Despite data constraints, this analysis provides evidence of a statistically significant role for variables that capture the deeper determinants of institutions in explaining current economic performance. All three indigenous institutional variables show a positive relationship with revenue to GDP ratios and this is robust across time period aggregations. The variables capturing the permanency of settlement (*set*) and the presence of clan-based organised (*clan*) tend to display more statistically significant results.

The positive sign on the coefficients of *clan* and *frag* may appear contrary to expectations, however given the use of concentration ratios of the three largest ethnic groups to construct the variables, economic rationale can be provided. In terms of the former, whilst clans are often referred to in a negative context, the fact that they are organised and have governance structures in place appears to dominate the possibility for intra-clan conflict. For *frag*, it is important to emphasise that this variable captures cross-national fragmentation of ethnic groups and so is focused on the legitimacy of the state, and not on the number of ethnic groups or within-country fragmentation. As the majority of ethnic groups within Africa are fragmented across national borders, the variable is in effect a simple concentration ratio – thus, higher values imply dominance of fewer groups, suggesting an increased potential for reaching political consensus on economic policy.

The magnitude of the coefficients of these variables diminishes over time, supporting the hypothesis that these underlying factors become less important to determining revenue outcomes. This can be attributed to changes in the composition of the economy, as well as institutions adapting both organically and through reforms. This is further supported by the fact that the interaction terms of the indigenous variables with GDP per capita are typically negative and quadratic in effect. As economies become richer the effect of indigenous institutional structures becomes less important for domestic revenue and taxation. Thus indigenous institutional structures affect government revenues both directly and indirectly through their effect on GDP.

The effect of the indigenous institutional variables is robust to a number of changes in specification. First, the organisational characteristics of the ethnic groups dominate the potential negative effect of fragmentation across borders. Second, the effect of the variable *frag* on revenues appears to be subject to a threshold effect. Specifically, in countries where the three largest groups, characterised by cross-border fragmentation, have population shares greater than 45.07 per cent, a greater positive effect on revenues is observed. Although as noted previously, this effect does diminish over time. Third, the indigenous institutional variables maintain some explanatory power in the regressions using tax-specific revenue data; the variable depicting permanent settlement (*set*) being particularly robust. However, these estimations suffer from a lack of tax-specific revenue data. Fourth, simple bivariate regressions of the indigenous institutional variables on revenue effort do not lead to statistically significant results, but this analysis is generally exploratory in nature.

As previously noted, the size of the sample is somewhat restricted and this is due to a number of constraints. With regards to data, the indigenous institutional variables are generated from ethnographic and historic population data that suffers from missing observations. In addition, the sample is restricted to 36 sub-Saharan African countries for which this data is available. Data scarcity also applies to cross-country data on general government and tax revenue, which is particularly sparse in the 1960s and 1970s. In terms of model specification, given the data constraints encountered, a parsimonious model is specified in order to conserve degrees of freedom. Again omitted variables may be an issue, but tests of the joint significance of variables (F-test) as well as the coefficient of determination (R^2) do support the statistical relevance of the model in explaining revenues.

There is an evident decline in both revenue and tax to GDP ratios in the 1990s and an examination of the residuals generated through the estimation of the model (both with and without the indigenous institutional variables) shows that they are greater in this period than in the others. This supports the notion that there are factors that were important to determining tax performance particularly relevant in the 1990s that are excluded from the model. One hypothesis is that this relates to the actual implementation of the tariff reductions that were agreed as membership conditions to the World Trade Organisation, applicable to the vast majority of sub-Saharan African countries in our sample. However this remains informed conjecture.

It is possible that the importance of particular indigenous institutional structures differ according to the types of taxes that are being raised. For instance one may assume that permanent settlement (*set*) is more conducive to the collection of direct and indirect taxes, as seen above. Whilst the variable *frag*, which proxies for potential group conflict, may be more relevant in models where resource rents are the dependent variable. This is not investigated in great detail in this analysis due to constraints in accessing disaggregated tax data.

Perhaps one of the greatest constraints to the analysis is the inability to use panel data methods. In an attempt to capture the dynamics of the relationship successive cross-sections are estimated, but we are unable to take into account country fixed effects. Thus the estimates suffer from the unobserved country heterogeneity, common with this empirical approach. In addition, no distinction is made between the political relevance of particular ethnic groups within a country. Whilst certain groups may represent relatively large proportions of the population, it is not given that these groups are necessarily politically active and engaging in

policymaking. Equally, possible changes in the dynamics of relationships or political prominence or not captured either.

The newly created variables used in this analysis are proxies for indigenous institutional structures and are based on an informed review of a branch of the ethnographic and anthropological literature. Attempts have been made to provide valid economic rationale for the role that these institutions may play, but equally, the variables may be capturing other factors associated with ethnic group organisation and thus omitted variable bias is an issue. However, using existing, contemporary institutional indicators in the model does not yield better performing results and thus we consider these new variables additional and complementary to the wide-range of institutional variables already being used in empirical analysis.

REFERENCES

- ACEMOGLU, D., JOHNSON, S. & ROBINSON, J. A. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91, 1369-1401.
- ALESINA, A., EASTERLY, W. & MATUSZESKI, J. 2006. Artificial States. *NBER Working Paper* 12328.
- BARDHAN, P. 2005. Institutions Matter, but Which Ones? *Economics of Transition*, 13.
- BARRO, R. 1991. Economic Growth in a Cross-Section of Countries. *Quarterly Journal of Economics*, 106, 407-444.
- BATES, R. 1983. *Essays on the Political Economy of Rural Africa*, Berkeley, University of California Press.
- BENEDEK, D., CRIVELLI, E., GUPTA, S. & MUTHOORA, P. 2012. Foreign Aid and Revenue: Still a Crowding Out Effect? *IMF Working Paper* [Online].
- BESLEY, T., PERSSON, T. & STURM, D. 2005. Political Competition and Economic Performance: Theory and Evidence from the United States. *NBER Working Paper Series* [Online].
- BIRD, R., MARTINEZ-VAZQUEZ, J. & TORGLER, B. 2007. Tax Effort: The Impact of Corruption, Voice and Accountability. *International Tax Program Papers* [Online].
- BRUK, S. & APENCHENKO, V. 1964. *Atlas Narodov Mira*, Moscow.
- CAMPOS, N., SALEH, A. & KUZEYEV, V. 2009. Dynamic Ethnic Fractionalization and Economic Growth. *Journal of International Trade and Economic Development*, 20, 129-152.
- CLIST, P. & MORRISSEY, O. 2009. Aid and Tax Revenue: Signs of a Positive Effect Since the 1980s. *Journal of International Development*, 10.
- COLSON, E. 1975. *Tradition and Contract: The Problem of the Social Order*. London: Heinemann.
- DAVIDSON, B. 1992. *The Black Man's Burden*, London, Times Books.
- DIAMOND, J. 1997. *Guns, Germs and Steel: The Fates of Human Societies*, W.W. Norton & Company.
- EASTERLY, W. & LEVINE, R. 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics*, 112.
- EASTERLY, W. & LEVINE, R. 2012. The European Origins of Economic Development. *NBER Working Paper Series* [Online].
- ENGERMAN, S. L. & SOKOLOFF, K. L. 2005. Colonialism, Inequality and Long-Run Paths of Development. *NBER Working Paper* 11057.

- FEDDERKE, J. W., LOURENCO, I. & GWENHAMO, F. 2011. Alternative Indices of Political Freedoms, Property Rights and Political Instability for Zambia.
- FENSKE, J. 2010. Ecology, Trade and States in Pre-Colonial Africa. *MPRA Paper* [Online].
- FORTES, M. & EVANS-PRITCHARD, E. E. 1940. *African Political Systems*, London, Oxford University Press.
- FRIED, M. H. 1960. *On the Evolution of Social Stratification and The State*, Indianapolis, Bobbs-Merrill.
- GENNAIOLI, N. & RAINER, I. 2007. The Modern Impact of Precolonial Centralization. *Journal of Economic Growth*, 12, 185-234.
- GLAESER, E., PORTA, R. L., LOPEZ-DE-SILANES, F. & SCHLEIFER, A. 2004. Do Institutions Cause Growth? *NBER Working Paper Series* [Online].
- GOODY, J. 1971. *Technology, Tradition and the the State in Africa*, London, Oxford University Press.
- GRAY, R. & BIRMINGHAM, D. 1970. *Essays on Trade in Central and Eastern Africa before 1900*, Oxford University Press.
- GUPTA, S. 2007. Determinants of Tax Revenue Efforts in Developing Countries. *IMF Working Paper* [Online].
- GUPTA, S., CLEMENS, B., PIVOVOVARSKY, A. & TIONGSEN, E. 2004. Foreign Aid and Revenue Response: Does the Composition of Aid Matter? In: GUPTA, S., CLEMENS, B. & INCHAUSTE, G. (eds.) *Helping Countries Develop: The Role of Fiscal Policy*. Washington DC: IMF.
- HALL, R. & JONES, C. 1999. Why do Some Countries Produce so Much More Output per Worker Than Others? *Quarterly Journal of Economics*, 114, 83-116.
- HODDER, B. W. 1964 Some Comments on the Origins of Traditional Markets in Africa South of the Sahara. *Transactions of the Institute of British Geographers*, 36, 97-105.
- KAUFMANN, D., KRAAY, A., & MASTRUZZI, M. 2009. Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996-2008. *World Bank Policy Research Working Paper No.4978*
- KEEN, M. 2012. Taxation and Development - Again. *IMF Working Paper* [Online].
- KNACK, S. & KEEFER, P. 1997. Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics*, 112, 1251-1288.
- MARSHALL, M., GURR, T., & JAGGERS, K. 2013. Political Regime Characteristics and Transitions, 1800-2012, Polity IV Project.
- MAURO, P. 1995. Corruption and Growth. *Quarterly Journal of Economics*, 110, 681-712.
- MCMILLAN, J., RAUSSER, G. & JOHNSON, S. 1994. Economic Growth, Political and Civil Liberties. *Occasional Paper International Center for Economic Growth* [Online].
- MICHALOPOULOS, S. & PAPAIOANNOU, E. 2011a. The Long-Run Effects of the Scramble for Africa. *CEPR Discussion Paper* [Online].
- MICHALOPOULOS, S. & PAPAIOANNOU, E. 2012. Pre-Colonial Ethnic Insitutions and Contemporary Africa Development. *NBER Working Paper Series* [Online].
- MIDDLETON, J. & TAIT, D. 1958. *Tribes Without Rulers*, London, Routledge.
- MKANDAWIRE, T. 2010. On Tax Efforts and Colonial Heritage in Africa. *Journal of Development Studies*, 1-23.
- MORRISON, D., MITCHELL, R. & PADEN, J. 1989. *Black Africa: A Comparative Handbook*, New York, Free Press.
- MURDOCH, G. 1959. *Africa: Its Peoples and Their Culture History*, New York, McGraw-Hill.
- MURDOCH, G. 1967. *Ethnographic Atlas*, University of Pittsburgh Press.
- NORTH, D. C. 1990. *Institutions, Institutional Change and Economic Performance*, New York, Cambridge University Press.
- NORTH, D. C. 1994. Economic Performance Through Time. *The American Economic Review*, 84, 359-368.
- PANDE, R. & UDRY, C. 2005. Institutions and Development: A View from Below. *Yale University Economic Growth Center Discussion Paper* [Online].
- PERSSON, T. & TABELLINI, G. 2003. *The Economic Effects of Constitutions*, MIT Press.

- PERSSON, T. & TABELLINI, G. 2006. Democratic Capital: The Nexus of Political and Economic Change. *American Economic Journal: Macroeconomics*, 1, 88-126.
- PORTA, R. L., LOPEZ-DE-SILANES, F., SCHLEIFER, A. & VISHNY, R. 1997. Legal Determinants of External Finance. *Journal of Finance*, 52, 1131-1150.
- PORTA, R. L., LOPEZ-DE-SILANES, F., SCHLEIFER, A. & VISHNY, R. 1999. The Quality of Government. *Journal of Law, Economics and Organisation*, 15, 222-79.
- POSNER, D. N. 2004. Measuring Ethnic Fractionalization in Africa. *American Journal of Political Science*, 48, 849-863.
- RODRIK, D. 1999. Democracies Pay Higher Wages. *Quarterly Journal of Economics*, 114, 707-738.
- RODRIK, D., SUBRAMANIAN, A. & TREBBI, F. 2004. Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth*, 9, 113-165.
- SCHAPERA, I. 1956. *Government and Politics in Tribal Societies*, London, Watts.
- SMITH, A. 1776. *An Inquiry Into the Nature and Causes of the Wealth of Nations*.
- STEVENSON, R. F. 1968. *Population and Political Systems in Tropical Africa*, Columbia University Press.
- SUNDSTROM, L. 1965. *The Exchange Economy of Pre-Colonial Tropical Africa*, London, C.Hurst.
- VANSINA, J. 1964. Long-Distance Trade-Routes in Central Africa. *Journal of African History*, 3, 375-390.
- WRIGHT, D. 1999. "What Do You Mean There Were No Tribes in Africa?": Thoughts on Boundaries - And Related Matters - In Precolonial Africa. *History in Africa*, 26, 409-426.

APPENDICES

Appendix 1: Indigenous Institutional Variables, Concentration Ratios

	<i>set</i>	<i>clan</i>	<i>frag</i>		<i>set</i>	<i>clan</i>	<i>frag</i>
Angola	70.84	70.84	13.86	Madagascar	.	.	0.18
Benin	74.63	0.88	74.63	Malawi	78.03	.	82.18
Botswana	14.86	12.00	14.86	Mali	29.40	0.12	68.33
Burkina Faso	69.55	5.45	7.5	Mozambique	78.35	6.39	81.95
Burundi	98.65	18.51	98.79	Namibia	53.37	8.05	67.98
Cameroon	49.53	23.35	18.51	Niger	53.37	.	75.41
Chad	38.43	4.48	16.04	Nigeria	18.03	5.90	36.43
Congo, DRC	39.82	24.96	28.14	Rwanda	97.64	.	97.64
Congo, Rep.	83.33	76.67	83.33	Senegal	80.50	44.17	80.5
Côte d'Ivoire	55.91	6.52	22.88	Sierra Leone	69.59	5.51	42.65
Eq. Guinea	95.68	95.14	95.68	South Africa	48.01	20.48	31.51
Gabon	47.99	35.49	89.29	Sudan	19.20	5.13	8.44
Gambia, The	40.32	14.52	47.99	Swaziland	93.99	90.23	93.99
Ghana	81.52	4.10	81.52	Tanzania	30.52	15.03	17.65
Guinea	57.74	1.23	77.10	Togo	60.81	.	87.5
Kenya	54.76	54.76	16.95	Uganda	39.81	14.76	35.06
Lesotho	98.28	10.76	98.28	Zambia	60.08	4.84	27.42
Liberia	61.90	28.57	57.62	Zimbabwe	6.67	.	70.95

Appendix 2: Variable Descriptions and Data Sources

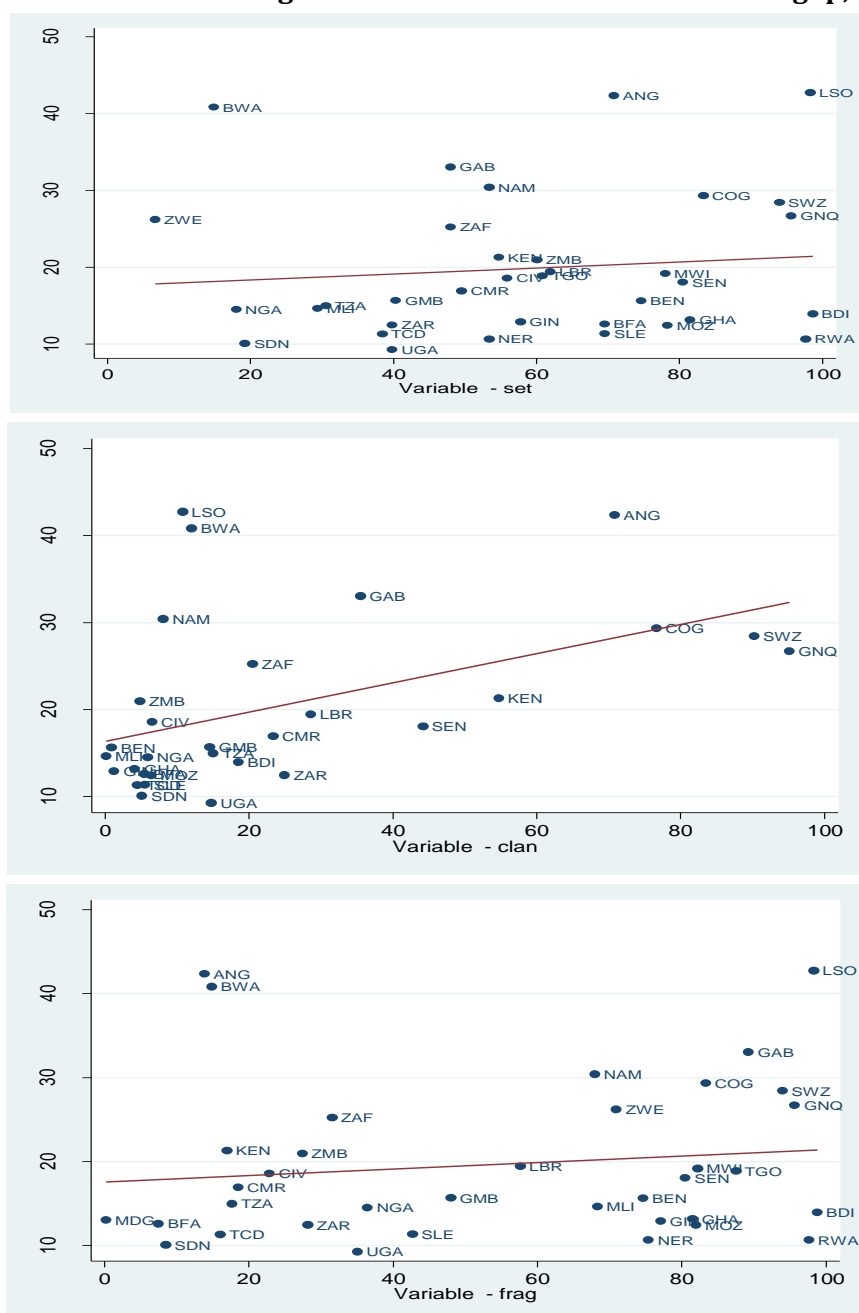
Variable Name	Variable Description	Data Source
revgdp	Government revenue excluding grants, % of GDP	ICTD database
agr	Share of agriculture (value added), % of GDP	World Bank, WDI
ind	Share of industry (value added), % of GDP	World Bank, WDI
GDPpc	GDP per capita, current US\$	World Bank, WDI
imports	Total imports, % of GDP	World Bank, WDI
exports	Total exports, % of GDP	World Bank, WDI
set	Concentration ratio of the 3 largest societies whose system of governance is characterised by permanent settlement.	Constructed by authors.
clan	Concentration ratio of the 3 largest societies that are characterised by a clan-based community structure.	Constructed by authors.
frag	Concentration ratio of the 3 largest societies that are fragmented across national borders.	Constructed by authors.

Appendix 3: Descriptive Statistics

Variable	Observations	Mean	Std. Mean	Min	Max
<i>revgdp</i>	36	19.66	9.31	9.26	42.74
<i>taxgdp</i>	35	13.5	6.37	5.04	38.61
<i>agr</i>	36	30.26	13.28	5.03	54.96
<i>ind</i>	36	27.43	12.85	13.00	57.85
<i>GDPpc*</i>	36	845.93	1098.23	153.37	4566.99
<i>imports</i>	36	39.35	19.13	18.81	107.73
<i>exports</i>	36	30.65	16.64	9.09	73.82
<i>set</i>	35	58.60	25.24	6.67	98.65
<i>clan</i>	30	23.63	27.25	0.12	95.14
<i>frag</i>	36	55.21	32.53	0.18	98.79

* The minimum value attributed to Burundi, the maximum to Gabon.

Appendix 4: Scatter Plots - Indigenous Institutional Variables and revgdp, 1970-2010

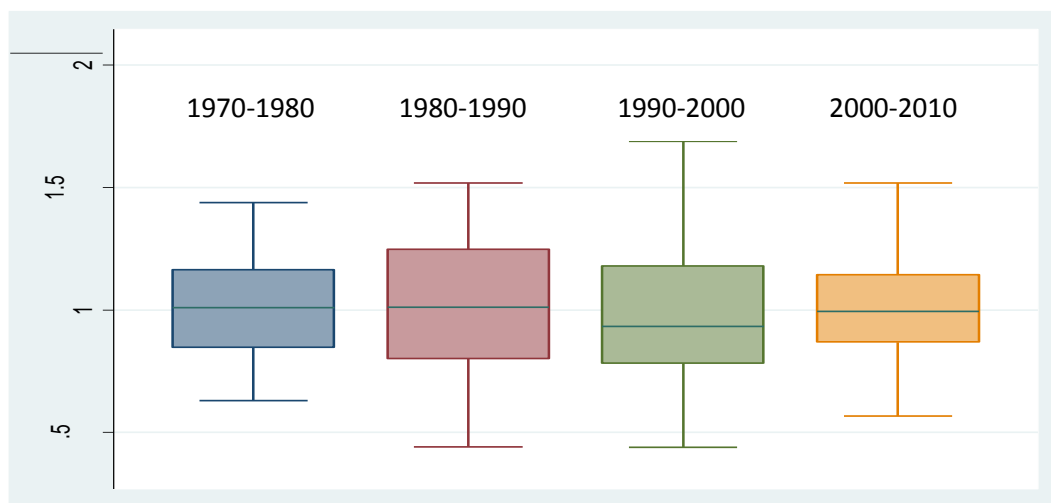


Appendix 5: Revenue Effort Index, 1970-2010

Below Average Performers		Above Average Performers	
Guinea*	0.63	Niger*	1.00
Mozambique*	0.73	Ghana	1.00
Chad*	0.76	Cameroon	1.04
Nigeria	0.77	Togo*	1.05
Sudan*	0.81	Cote d'Ivoire	1.05
Swaziland	0.82	Liberia*	1.06
Eq. Guinea*	0.83	Uganda*	1.08
Burkina Faso*	0.83	Lesotho*	1.10
Sierra Leone*	0.87	Tanzania*	1.11
Zambia*	0.87	Benin*	1.14
Congo DRC*	0.88	Gabon	1.15
Gambia, The	0.88	Botswana	1.16
South Africa*	0.91	Mali*	1.22
Rwanda*	0.93	Zimbabwe	1.22
Namibia	0.93	Angola*	1.26
Congo, Rep*	0.94	Malawi*	1.32
Senegal*	0.94	Kenya	1.44
Madagascar*	0.97	Burundi*	1.66

Note: countries in bold are members of the African Petroleum Producer's Association (APPA); * indicates LDC status.

Appendix 6: Revenue Effort Index by Period



The box plot: whiskers indicate the maximum and minimum values, the line in the box is the median and the size of the box indicates the distribution between the 25th and 75th percentile.