University of Leicester

DEPARTMENT OF ECONOMICS

Reconsidering the Fiscal Effects of Constitutions

James Rockey, University of Leicester, UK

Working Paper No. 10/16
May 2010

Reconsidering the Fiscal Effects of Constitutions

James Rockey*

University of Leicester

May 15, 2010

Abstract

This paper reconsiders Persson and Tabellini's (2003,2004) analysis of the causal effect of constitution type on government size, it addresses the concerns of Acemoglu(2005) and makes some further refinements to argue that there is a qualitatively large, and statistically significant relationship between constitution type and government size. The age of a democracy is of increased importance in the new identification strategy, but existing measures are shown to be flawed. Two new measures of the age of a democracy are introduced. The first details when a country first had a genuinely democratic election, the second when its current constitution was promulgated.

*james.rockey@le.ac.uk., I thank Jonathan Temple for many helpful discussions and Simon Johnson, Torsten Persson, and Stefan Voigt for making their datasets available. The usual disclaimer applies.

Introduction

Constitutions vary: the US constitution is a classic of political philosophy, whilst the UK has no formal constitution. Yet, constitutions tend to vary little over time. In many cases the decisions made when drawing up a constitution as to the separation of powers or the rights of individuals echo down the centuries. This paper revisits the question of whether these echoes include macroeconomic outcomes, specifically the size of government, and argues that they indeed do.

The study of constitutions both normative and positive has a long and distinguished history. Formal analyses are a more recent addition to the literature and two key contributions are those of Persson et al. (1997, 2000) and Persson and Tabellini (2000). Generally speaking, their models suggest that countries which have a presidential system, or a first past the post electoral system are likely to have smaller governments. This paper considers the previous empirical evidence for this hypothesis and addresses some important concerns about the methodology and some of the data used to provide new support for their findings.

Persson and Tabellini (2003) (henceforth PT) was a major step forward in the empirical analysis of the relationships between constitutional type and government size. They provide evidence that presidential systems have a large, negative, and statistically significant impact on the size of central government. They also suggest that first past the post electoral systems are also associated with a smaller government share of output, but the evidence is weaker for this hypothesis. PT treat a country's choice of constitution as possibly endogenous and use instruments premised on the idea that historical timing of democratisation and the degree of European influence are good predictors of constitutional type. These instruments will be seen to be problematic but using new instruments, and alternative, improved, estimation methods this paper argues that there is indeed a quantitatively large negative effect of presidential democracy on the average size of government.

This paper will focus on chapter 6 of PT, which overlaps to some degree with Persson and Tabellini (2004). Both provide evidence that countries with proportional electoral systems are expected to have more central government expenditure as a proportion of output, and *a fortiori* that Presidential democracies are associated with a smaller

government. The key criticisms of Acemoglu (2005) will be considered and addressed. Furthermore, new data will be described which attempt to date more accurately when a country first permanently became a democracy. This is motivated by some notable discrepancies in the existing data and also due to the demonstrated relevance of the age of democracy as an instrument for constitutional type. More generally, both of these objectives gain impetus from the importance of the contribution made by PT. Acemoglu writes that:

> "[...] I believe that overall PT have largely achieved their ambitious aim of revolutionizing comparative political economy, and this book is the most significant contribution to this field since Lipset's work almost 50 years ago".

Only time will tell if this is true, but on the available evidence it would be churlish to rule it out. Establishing causality is often difficult in social science, and in few cases more so than in any attempt to disentangle the complex web of institutions, social mores, and ideology that contribute to determining the size of the state. Yet, efforts to establish a genuinely meaningful, understanding of constitutions and their effects, require consideration of causal effects. Consequently, the focus of this paper are issues of causal identification.

Not surprisingly given the complexity of the task, PT's attempts to do so suffer from several problems. They pursue an instrumental variables approach (IV), but Acemoglu criticizes both the estimation method, and also the relevance of the excluded instruments. This paper will address Acemoglu's criticisms and show that by using an improved methodology and more suitable instruments that it is possible to provide additional evidence for a causal relationship.

This paper proceeds as follows. In the next section PT's set of instrumental variables will be outlined, as will Acemoglu's critique of their relevance. This will be followed by a discussion of the motivation for the new instruments used. The second part of section two will provide a brief overview of the evidence as to the relative merits of different estimation strategies, and in particular the advantages of using Limited Information Maximum Likelihood (LIML) based estimators rather than Two-Stage Least Squares (2SLS). The third section discusses some of the new data used, and in particular the relevance of the excluded instruments. Section four provides evidence that using LIML

based methods, and a different set of instrumental variables, that the criticisms of the estimation approach and instruments can, to a large extent, be overcome. Estimates are obtained using both the original dataset of PT and also the extended dataset of Blume, Möller, Voigt and Wolf (2009). Contrary to the general message of their paper, the IV estimates suggest that even when using their expanded dataset, there is still a large negative impact of presidentialism on government spending, and little evidence of an effect associated with the type of electoral system. Section four also reviews the results using alternative estimates of the age of democracies. Results obtained using the new data lend additional support to the claims of PT. Section five concludes.

1. PT's estimation strategy and its problems

1.1. The choice of instruments

This section shall only briefly outline the parts of PT that focus on the effect of presidentialism and majoritarian electoral rule on fiscal policy. Here PT aim to test the empirical validity of theoretical work outlined in their previous book, Persson and Tabellini (2000), which makes two central claims.

Firstly, it suggests that nations with majoritarian elections will have a smaller government than those with proportional representation. The theoretical logic for this, developed in Persson, Roland and Tabellini (2000), is founded upon models in which the larger electoral districts associated with proportional representation make electoral competition more diffuse instead of being focused in a few marginal districts. This means that support requires government transfers to a larger swathe of the population, thereby increasing government spending.

Secondly, as first discussed in Persson, Roland and Tabellini (1997) it is posited that presidential systems will have lower government spending than parliamentary systems, as the implied separation of powers prevents politicians from colluding in extracting rents and hence makes them more accountable. [1]

[1] There are several other important approaches, such as that of Milesi-Ferretti, Perotti and Rostagno (2002) present and empirically test a model in which, under proportional representation, voters have an incentive to elect politicians more prone to public good provision. Besley and Case (2003) consider variation in electoral rules across US state legislatures, emphasising the importance of more-subtle variations in state constitutions.

In the regressions of interest, in PT, the dependent variable is central government expenditure as a percentage of GDP, *cgexp*. Using data on around 80 democracies, PT consider whether *ceteris paribus* the type of constitution a country has is responsible for, in part, determining the size of government. However, it is potentially the case that a country's choice of constitution is partly endogenous, driven by national preferences that also partially determine the level of government expenditure. Evaluating the effects of constitutions on government spending under the assumption of random constitutional assignment will result in severely biased estimates and prohibit causal inference. PT take these concerns seriously and as Persson (2004) notes, they

> "exploit systematic co-variation between the relative frequency of alternative constitutional rules and their broad time period of introduction".

As discussed above this is done using the variables describing the age of democracies as instruments. Section 3 considers some potential weaknesses in the construction of these measures and proposes two alternatives. The first part of this section focuses on the choice of the other instruments used, and the second part considers some potential problems with the estimators used. In particular, it considers PT's 2SLS analysis. PT consider a range of other econometric techniques to circumvent the problems of endogeneity, specifically Heckman selection-correction models and matching estimators. However, here the focus is restricted to the 2SLS estimates since this approach has garnered most attention, and also allows both of the PT constitutional variables to be treated as endogenous simultaneously. PT instrument for whether a country's electoral system is majoritarian or not, and whether it has a presidential system. The extent to which they do so successfully is the focus of the following discussion. Of course, the approach taken here could be readily adapted for other related questions.

Binary variables denoted *maj* and *pres* are defined based upon these criteria: a country is considered to be majoritarian (*maj*=1) if elections to the national legislature (or the lower chamber in a multi-cameral system) are conducted using exclusively a plurality rule, that is, each constituency elects the single candidate who gains the highest proportion of the vote. Similarly, a country is considered to be a presidential democracy (*pres*=1) if the executive is independent of the legislature, that is, not subject

to a confidence vote.

Both are instrumented for using the same seven variables, four describing when a country became a democracy, and three variables describing colonial influence. The age of a democracy is measured using the variable *age* and three binary variables *con2150, con5180, con81*. The inclusion of variable describing the age of democracies is premised on the idea that when a constitution was written is a plausible predictor of its type, since as PT note there have been changes in 'constitutional fashion' over time. These variables are constructed based upon the variable *demage* which describes how long a country has been a democracy. *demage* is defined by the start of a continuous set of positive *polity* values excluding any interruptions due to foreign occupation. The *polity* variable records the difference between the score given by the Polity IV database for the extent of institutionalized autocracy and the degree of institutionalized democracy with a score between of –10 (very autocratic) to +10 (very democratic). *age* is defined as 2000-*demage*. As is argued in Section 2. this is likely to be an imperfect measure of the age of a democracy, despite its virtue of objectivity. The variables *con2150, con5180, con81* are indicator variables based on *demage* which describe whether the current constitution was promulgated between 1921 and 1950, 1951-1980 or post-1981 with 1920 or earlier the omitted category. To capture colonial influence PT use the variables *Engfrac, Eurfrac* and *Lat01* based upon the work of Hall and Jones (1999) and (more loosely) Acemoglu, Johnson and Robinson (2001).*Lat01* describes distance from the equator. *Engfrac* describes the proportion of the population speaking English as a first language, *Eurfrac* is the same but for the major European languages.

However, Acemoglu (2005) argues that this instrument set is unsatisfactory, in particular that some of the instruments are not excludable, and that the others are weak. He argues that this is the case for two main reasons. Firstly, that Hall and Jones (1999) use the variables *Engfrac, Eurfrac*, and *lat01* as a measure of Western colonial influence. They argue that countries further from the equator are more likely to have climates similar to those encountered in Europe and were therefore more appealing to European settlers. Similarly, those countries with a high percentage of speakers of a European language can be reasonably expected to have experienced more European influence. Hall and Jones (1999) argued that these measures of Western influence

were a good instrument for the presence of good legal and/or democratic institutions (”social infrastructure”).

But, Acemoglu, Johnson and Robinson (2001) argue that the nature of colonization was related to the environment encountered in the area being colonized more specifically than can be described by latitude. That is, in countries where colonists encountered adverse conditions (high rates of settler mortality) large scale settlement was unlikely, and this led to absolutist regimes focused on extractive activities, often employing forced labour and with no system of property rights. In colonies where mortality rates were lower, larger settlements were more likely and these settlers demanded, and in general got, institutions similar to those in their home country. For example, contrast the Belgian Congo with Australia or New Zealand. Hence, whether the European colonization was beneficial to a particular country is dependent upon the nature of the colonial process and as such, although the Hall and Jones instruments are correlated with settler mortality rates, the instruments cannot be seen to be related to beneficial Western influence in the form of high quality institutions.

Secondly, the Hall and Jones instruments, even if they described the extent of European influence, would not be able to explain the variation in the forms of democracy which PT wish to instrument. Since a large proportion of the sample are themselves European democracies, or were not colonized at all, we should not expect the Hall and Jones instruments to be informative for these. If they are not informative about countries that weren't colonized or were themselves colonial powers then there is no reason to expect that the instruments are relevant for these countries and hence in general are not useful instruments.

Perhaps more importantly, even if they are good instruments for the quality or presence of democratic institutions they cannot explain why certain countries have adopted particular democratic institutions. Acemoglu provides an extremely thorough treatment of the pitfalls of IV estimation in his paper, and *inter alia*, demonstrates that if instruments that are related to a “cluster of institutions” are used to predict a specific institution then the effect of all the other, ignored, institutions is included in the effect for the specific institution of interest. Hence, it is clear that using the Hall and Jones instruments are not appropriate for the specific features of democracies. Unfortunately, the other instruments PT use, the measures of democratic age, have

little explanatory power on their own. This means there is (potentially) a weak-instruments problem, if only they are used. The problem of weak instruments will be discussed in more detail in the next section.

The remainder of this section will now consider some alternative instruments. An obvious approach is to include age^2 as well as *age* since there is no reason to think that the variations in 'constitutional fashion' have occurred linearly.[2] The other instruments that will be considered relate to countries' colonial experiences as well as some describing the antiquity of states and societies.

The central premise underlying the choice of the rest of the instruments is that colonization did matter, but for the different reason that colonized countries were likely to inherit similar institutions to those of their colonial rulers either directly or indirectly. The widespread presence of UK-style Parliamentary democracies in ex-British colonies is anecdotal evidence of direct influence. More generally, there are several different but related indirect reasons why certain colonial powers might have induced particular constitutional rules. European colonization can be broadly conceived as having taken place in two phases. The first roughly coincides with the discovery of the Americas and the subsequent colonization. What is clear for this, early, process of colonization is that the different colonial nations had different objectives. The focus of Spanish and Portuguese colonialists was the extraction of mineral wealth and the conversion of indigenous peoples to Catholicism. See for example, Olsson (2004)

This necessarily engendered different institutions to those in the British, Dutch, and French colonies, which were focused on more permanent settlement and trade. This is not a claim that there was not both substantial rent extraction by the British, Dutch, or French colonists and attempts by them to spread Christianity amongst the indigenous peoples. Indeed, much of the economic logic of British colonies was founded on obtaining slaves in West Africa for exploitation in plantations in North America and the Caribbean. Rather, the different, more permanent, emphasis of this colonial activity necessarily led to different institutions.

The second phase of colonization coincided with the colonization of parts of Africa

[2]This has the advantage compared to *con2150*, etc. of not needing to assume when the key changes in constitutional fashion took place. A cubic term was considered but added little additional explanatory power. Using the first-stage F-statistics designed to identify potential weak identification suggested by Angrist and Pischke (2009) suggests that using *age2* rather than *con2150*, etc., and the remainder of PT's original excluded instrument set, improves the identification of both equations.

and Asia, culminating in the 'Race for Africa', following which almost all of the African continent, alongside large amounts of Asia, was ruled by a European power by the outbreak of the First World War. A notable feature of this wave of colonization in comparison to the first phase was that France and the UK were pre-eminent, in contrast to the earlier importance of Spain and Portugal. Also, there was the growth of what might be termed "settled colonies" such as Australia, Canada, and New Zealand. This is not to deny the presence of the pre-existing populations of these countries, rather to emphasize the large numbers of Europeans who migrated to them.

This variation between the incentives and methods pursued by the different countries at different times is interesting, since the process of decolonization also differed not only between different colonialists but also across time. The peaceful way in which Brazil obtained its independence from Portugal stands in stark contrast to the violent wars of liberation fought in the countries that are now Bolivia and Venezuela, and in general across the rest of South America. Similarly, the struggle for independence in Algeria stands in stark relief to that of the comparatively peaceful transitions in many other French colonies. It is not the purpose of this paper to provide an account of the different paths to independence of the myriad different colonies, nor to delineate the reasons for this variation.

The argument instead is that countries are likely to have inherited political systems similar to those in the occupying colonial power, and that this tendency is expected to be more pronounced in countries where there was not a large-scale war of liberation. The argument about colonial influence is far from new, see for example, La Porta, Lopez-de Silanes, Shleifer and Vishny (1998), and Hall and Jones (1999). Moreover, those countries which did have to fight for their independence might well be influenced by the form of the broader societal institutions the colonialists had put in place. For example, Acemoglu, Johnson and Robinson (2001), Acemoglu and Robinson (2006) argue that high levels of inequality are likely to engender more political instability and that this is due to the importance of small elites. Furthermore, countries with great mineral wealth are more likely to exhibit high degrees of inequality (see Leamer, Maul, Rodriguez and Schott (1999)) mainly due to large rents that are concentrated amongst a small elite as suggested by Jensen and Wantchekon (2004). This is even more likely in countries in which colonialists or pre-existing elites used forced labour. These results

taken together suggest that if countries have higher degrees of inequality, they are more likely to have majoritarian regimes.

To try and capture the variety of colonial experiences, variables were included describing whether a country was colonized by the UK (*coluka),* Spain (*colespa*), or another country (*colotha*), having been colonized by France was the omitted category. Also included were whether a country was colonized at some point which is described by *excolony*[3] and the percentage of the population who were Catholic in 1980 is also included *(catho80).*[4]

In sum, it is argued that these alternative instruments are not vulnerable to the same criticisms that Acemoglu made of the Hall and Jones (1999) instruments. As a set of instruments they explicitly consider the different influences of different colonial powers, the degree of susceptibility of a country to these influences, and via *age* and age^2, the impact of constitutional "fashion" on those countries that were not colonized. Moreover, unlike the analysis in Hall and Jones (1999) they are concerned with the form of government rather than with the existence of high or low quality institutions.

1.2. Persson and Tabellini's Methodology

In small samples the conventional challenges associated with instrumental variables approaches, identifying relevant and excludable instruments, are heightened by related concerns about the number of regressors compared to the number of observations and that particular observations may be driving the results. By instrumenting using a set of variables measuring the age of a democracy, as well as distance from the equator and the proportion of the population speaking English or another European language, PT hope that the residual variation in constitutional choice is random (i.e. that there is selection on observables). As mentioned above, following this method, they obtain estimates which suggest that a presidential system is associated with smaller government and a proportional electoral system is associated with a larger one.

The dependent variable in the second stage regression is *cgexp*, which is regressed

[3] From Acemoglu, Johnson and Robinson (2001)

[4] Whilst *catho80* cannot be seen as being predetermined and hence plausibly exogenous, it is claimed that non-European countries with large Catholic populations in general were converted prior to becoming democracies. Hence, it is believed that as well as being a good indicator of a religious focus to colonization, it is plausibly exogenous.

upon *maj*, *pres*, *age*, *lyp* (log income per capita - richer countries may have larger governments) *trade* (openness defined by the share of imports and exports as a share of GDP, as Rodrik (1998) argues more open economies tend to have bigger governments), *prop1564* and *prop65* (the proportions of the population between the ages of 15 and 64 and the proportion 65 or older, an higher dependency ratio is often associated with larger government). [5]

Due to concerns about sample size and weak instruments (discussed below) PT do not include all of the exogenous variables in the first stage. As well as his critique of the relevance and validity of the actual instruments used, Acemoglu (2005) also presents a theoretical critique of this approach. His theoretical criticisms deserve careful consideration not least since they motivate some of the suggestions below. Given its importance for the following discussion it is worth formally outlining the basic instrumental variable (IV) estimator for a single endogenous variable, which can be stated as follows:

A conventional OLS estimator might be applied to:

$$Y = X_1\beta + X_2\gamma + u \tag{1}$$

However if $Cov(X_1, u) \neq 0$ then OLS is inconsistent. But assume that a variable Z can be found such that $Cov(Z, u) = 0$ and $Cov(Z, X_1) \neq 0$. Then the model can be written as follows:

$$Y = X_1\alpha + X_2\beta + u \tag{2}$$

$$X_1 = Z\lambda + X_2\zeta + v \tag{3}$$

Where Y is a $N \times 1$ vector containing the dependent variable, X_1 is the $N \times 1$ vector describing the endogenous variable. X_2 is an $N \times K$ matrix of exogenous variables. Z

[5]The other right-hand side variables are *gastil* is the average of Freedom House's index of civil and political rights measured on a 1-7 scale (the form of democracy may matter more in better functioning democracies) . Finally indicator variables are included for whether a country has a federal government (*federal* since central government spending might be expected to be lower in federal systems), or is a member of the OECD (*oecd*, for similar reasons to *lyp*). These variables and the others used in this paper are summarized in Table 1. The logic for the choice of controls is explained more thoroughly by PT(pp39-43).

is an $N \times K$ matrix containing the set of instruments for X_1. Then the IV estimator is:

$$\widehat{\beta}_{IV} = (Z'X)^{-1}Z'Y \tag{4}$$

PT for the reasons mentioned above use something closer to the following estimator:[6]

$$Y = X_1\alpha + X_2\beta + u \tag{5}$$

$$X_1 = Z\lambda + v \tag{6}$$

This estimator is only consistent if $Cov(Z,u) = Cov(Z,u \mid X_2) = 0$ whereas the conventional IV estimator only requires that $Cov(Z,u) = 0$. Abstracting from theory it is unlikely to find a variable which meets the conditions that it is both sufficiently correlated with the endogenous variable and unconditionally independent of the error term. Thus, unless X_2 has no explanatory power for X_1 then $Cov(Z,u) \geq Cov(Z,u \mid X_2) \geq 0$. Hence, the estimates are likely to be inconsistent. For this reason this paper uses an alternative approach.

But, PT's concerns about weak instruments are likely to be well founded and therefore it is logical to consider estimators that perform better if instruments are weak, rather than simply to use conventional 2SLS or to disregard the results entirely. The remainder of this section shall now consider alternative estimators such as LIML, and their relative advantages and disadvantages.

An instrument is said to be weak when $Cov(X_1, Z)$ is low or alternatively $\lambda \approx 0$. Hahn and Hausman (2003) state that weak instruments mean that

> " (i) 2SLS is badly biased toward the OLS estimate and alternative "unbiased" estimators such as LIML may not solve the problem and (ii) the standard (first order) asymptotic distribution does not give an accurate framework for inference."

There has been a large recent literature on the problem of weak instruments, and in particular testing for whether instruments are weak and the performance of

[6] It should be noted that PT(p164) acknowledge the assumption made, and suggest (as reported below) that the estimates are stronger when the included instruments are included.

different estimators under these conditions. Here, the analysis will be restricted to the specific case of PT's estimates but for more general and detailed analyses see Angrist, Imbens and Krueger (1999), Blomquist and Dahlberg (1999), Hahn and Hausman (2003), Stock, Wright and Yogo (2002) Honoré and Hu (2004). The main conclusions of this literature seem to be two-fold. Firstly, that 2SLS performs particularly badly with weak instruments, and that estimators without finite moments such as LIML, Jackknife Instrumental Variables (JIVE) or the Nagar estimator are potentially problematic in finite samples. Consequently, Hahn and Hausman (2003) advocate the use of either the Fuller (1977) modified LIML estimators or jackknife 2SLS estimation.[7] Fuller's modified LIML estimator has finite sample moments but requires the researcher to specify a constant b. Conventionally, b is either given a value of 1 or 4. If b is equal to 1 then the estimator is mean-unbiased, a choice of 4 provides an estimator that minimizes the mean squared error. The focus here is on these Fuller estimators since they are more readily implementable, and Hahn and Hausman (2003) results from Monte Carlo simulations suggest that Fuller(4) is expected to be the best performing estimator given the sample size, and first-stage R^2s.

Given that this paper is, in part, concerned with robustness two further estimators were considered. The first of these is the Continuously Updating GMM estimator (CUE) first proposed by Hansen, Heaton and Yaron (1996). This estimator can be thought of as generalization of LIML to the case of non-spherical disturbances analogous to the relationship between 2SLS and GMM. It's also a special case of the Generalized Empirical Likelihood (GEL) estimator. [8] In this respect, as discussed by Smith (2005) it's asymptotic bias is expected to be less than that of any (feasible) GMM estimator,

[7] It is important to note that JIVE and jackknifed 2SLS are different. Both are designed to address the bias of 2SLS in small samples. This bias, due to the correlation of u and v, becomes larger as the number of instruments increases for a given sample size. JIVE as proposed by Angrist, Imbens and Krueger (1999) estimates $\hat{\beta}_{JIVE} = (\hat{X}'_{JIVE}X)^{-1}(\hat{X}'_{JIVE}Y)$ where $\hat{X}_{JIVE}$ are predicted values of the endogenous regressors obtained using the jackknife to construct instruments orthogonal to the error term in finite samples. Hence, making the correlation between u and v zero and eliminating the bias. However, JIVE lacks finite sample moments and Monte Carlo simulations suggest that it performs badly in small samples. Jackknife 2SLS as described by Hahn, Hausman and Kuersteiner (2004) uses the jackknife to estimate the bias of the 2SLS estimate and subtracts this from the 2SLS estimate. The advantage of this approach is that as the estimated bias of the 2SLS estimate is a combination of N 2SLS estimates the estimator has finite moments if the degree of over-identification (the difference in the rank of Z and X) is greater than 2.

[8] Alternatively, it can be seen as in Hansen, Heaton and Yaron (1996) as a GMM estimator where the objective function is minimized for the parameter vector in both the expected moments and the weight matrix. .

and importantly it is efficient in the presence of arbitrary heteroskedasticity unlike the Fuller estimators. However, the finite-sample properties are more ambiguous. Guggenberger (2008) notes that it is difficult to analytically establish whether GEL estimators possess finite-sample moments, but using a Monte-Carlo analysis suggests that the finite sample properties of GEL, including CUE, are similar to LIML. That is, as discussed above, they suffer a moment problem. Specifically, he shows that the probability of an extreme estimate, and the standard deviation of the estimates is much higher for both GEL estimators and LIML. On this basis, he argues that GEL estimators shouldn't be used for the linear IV model. However, the lack of an efficient alternative in the presence of heteroskedasticity suggests a compromise of reporting the CUE estimates in tandem the Fuller estimates. Given neither estimator is entirely suitable.[9]

Another consequence of weak instruments, other than biased estimates, is that the size of the tests on coefficients in finite samples may differ arbitrarily from their asymptotic size. Moreira (2003) proposes a score (LR) test that is of the correct size, even when instruments are arbitrarily weak. However, it is only derived for the case of a single endogenous variable, whereas PT's setup has two (*Pres* and *Maj*). Applying the Moreira (2003) approach to the endogenous variables in turn supports the hypothesis that *Pres* is significant and negative, and *Maj* is not at all significant.[10]

In summary, Acemoglu's two central criticisms of PT's approach have been addressed. The new instruments proposed are both more likely to be relevant and excludable, and by using alternative estimators concerns about weak instruments are reduced. The next section considers some potential flaws in PT's measure of the age of democracies.

2. The age of democracies

Given that the age of democracies is central to the instrument set proposed in the previous section, and to the original approach of PT, it is worth considering how

[9] Estimates for both the Fuller and CUE estimators were obtained using the IVREG2 package for STATA provided by Baum, Schaffer and Stillman (2002)

[10] The particular specification employed is to include the other endogenous variable, *pres* or *maj*, as an exogenous predictor. The estimates were obtained using the CONDIVREG package of Mikusheva and Poi (2006).

accurately it has been measured.[11] As described in section 2.1 PT's measure *demage* records when the start of a continuous string of positive *polity* scores occurred. A casual inspection of *demage* immediately reveals some potential weaknesses. The most obvious of these is South Africa, which as described in Appendix B, is dated as having been a democracy since 1910. A country in which the majority of the population are precluded from voting on ethnic grounds is clearly not a democracy in any meaningful, modern, sense.

Japan is another interesting example. The PT variable *demage* records that it became a democracy in 1868. However, universal suffrage was only introduced in 1925 and individual and political freedoms were still limited by the "Peace Preservation Law" that followed shortly afterwards. More importantly, this transition to democracy was reversed during the 1930s as the military and nationalists seized effective control of the country and assassinated leading civilian politicians. After the end of the Second World War a very different democratic constitution was introduced that was avowedly pacifistic and which reduced the role of the emperor to a figurehead. That throughout the 1930s and 1940s the *polityiv* measure still records positive values suggests that this method of dating democracies leaves much to be desired. There are other examples, including the USA, South Africa, and Costa Rica, where the age of a democracy, as measured by *age*, would seem to be measured with error, and these are described in the Web Appendix .

It has often been the case that the transition to democracy from autocracy has been a gradual evolution, particularly in countries that have not been colonized or involved in major conflicts. At which point a country should be seen as fully democratic is often more difficult to judge in these cases as *de facto* changes are sometimes as important as *de jure* changes. Consider the case of Sweden, which PT records based upon the *polityiv* data, as having been a democracy since 1917. A crucial point in the development of Swedish democracy was the passing of the "Instrument of Government" in 1809 which divided power between the monarch and "Riksdag of the Estates" which represented the four Swedish social groups. This was replaced by a bicameral parliament in 1866 and universal male suffrage was formally introduced in 1907 and was followed

[11] Although, Acemoglu (2005) notes that it is the other variables *lat01, engfrac,* and *eurfrac,* that have the majority of the explanatory power in the first stage.

closely by universal suffrage in 1921. However, whilst increasingly the King's role was merely a formality, it was not until the constitution of 1975 that his constitutional role was reduced to head of state. Hence, in some respects Sweden can be seen to have been a democracy since 1907; alternatively it could be considered as having become a democracy when the first elections with universal male suffrage took place in 1911, or the first with universal suffrage in 1921. It is clear that Sweden was a democracy in a meaningful sense before 1975 but in other countries the introduction of a genuinely democratic constitution has been a necessary, if not always sufficient, condition for democracy.

Hence, given the disadvantages of using *age* as a measure of the age of democracies, and its importance in PT's analysis, a strong case can be made for considering some other measures. However, there is an absence of readily available alternatives to *age*. Moreover, the often great variations between when a country became a democracy *de jure*, and when it was in fact a *de facto* democracy motivated a collation of two new variables describing the dates of a country's first democratic election, and the promulgation of its first democratic constitution. These two new variables, constructed by the author, will be denoted *dateelections* and *dateconstitution*.

The new measures were constructed using primarily the International Constitutional Law Project website Tschentscher (2009), national constitution websites, Encyclopaedia Britannica, the CIA World Fact Book, and www.rulers.org. As noted above, in some cases this is a slightly subjective process. For example in the case of the UK, there was no clear transition to democracy, but instead the franchise was slowly extended as power was transferred from the aristocracy to democratically elected representatives. Similar issues exist with countries that have no formal constitution, or in which the constitution has been changed in a fundamental way subsequent to having become a democracy. Facing such difficulties the most appropriate strategy is inevitably a combination of using some *a priori* criteria whilst retaining an appreciation of the idiosyncrasies of a given country's transition to democracy.

The method used here is premised on the basis that a small degree of subjectivity is more than compensated for by a more accurate picture than can be arrived at by purely mechanistic methods. Consequently, when considering the date of a country's first election the criteria sometimes vary. For a country such as Uganda the important

criteria relate to the extent to which the election was free, fair, and pluralistic. Alternatively, when considering the evolution of democracy in more established democracies such as the USA or the UK, the salient issues are concerned with universal suffrage.[12] Similarly, in countries which lack a written constitution, a judgement has to be made as to when the country became democratic in a modern sense. For example, ancient Athens would not qualify due to property, citizenship, and gender restrictions. In this paper's measure, the criteria for the date of first democratic elections (*dateelections*) were each considered to be a necessary condition and are as follows:

1. A country is considered to have held democratic elections if they were pluralist, and not characterised by widespread voter-intimidation or ballot stuffing. In most cases this is evident from the standard historical narratives for a given country, however in some (particularly recent) cases the records of outside election observers are required.

2. The elections were or are expected to be succeeded by further free elections at regular intervals.

3. A country is considered to have held a democratic election if the franchise is defined by universal male suffrage. Whilst it is obviously not in any modern sense "democratic" to have elections in which only around 50% of the adult population may vote, the decision not to require universal suffrage was based upon a need to preserve the richness inherent in the global history of democracy. Requiring universal suffrage would have artificially compressed the data, implying that almost every country that is now a democracy with a full franchise, held its first democratic election, at the earliest, at the start of the twentieth century.[13]

4. For the purpose of this paper, a country is only considered to have achieved the democratic transition when there is no subsequent return to autocracy except that imposed by short-term foreign invasion.

[12] This is because whilst the franchise was often extended slowly in many older democracies it is increasingly rare for a newly democratic country to hold elections with a restricted franchise.

[13] This is perhaps the most controversial of the criteria, in part due to its logic following from a concern for the sufficiency of the data rather than a standard democratic principle. In doing so it neglects an important historical movement. But, if it were to focus on universal suffrage then in many respects it would ignore instead the, perhaps equal, importance of the myriad movements that constitute the history of democratisation.

Similarly for the date of constitution (*dateconstitution*):-

1. A country was considered to have a democratic constitution when it first promulgated a constitution that provided for free and plural democratic elections.
2. The date of constitution refers to the first democratic constitution that provided for a structure of government broadly similar to that currently in force. Therefore, in a parliamentary system, the promulgation of a new constitution or an amendment to the current constitution is only seen as the first constitution in cases where the transition to the new constitution was characterised by martial law, popular unrest, or military intervention.
3. The replacement of a monarch with an elected, but largely symbolic, president is not considered to represent a real change in either the structure of government or the advent of real democracy.
4. If several constitutions are promulgated and rapidly replaced as part of the democratic transition then it is the final (and current) one, which is considered for the purposes of establishing the date of constitution. In reference to Criterion Two, this means that a country that undergoes a period of political instability as part of the democratic transition is not considered a democracy until it achieves a workable constitution.
5. A constitution is only considered to date from the beginning of its current period of continuous enforcement. Hence, ex-USSR countries which resurrected their pre-WW2 constitutions are considered to have had a democratic constitution since its promulgation date post-1989.
6. If a country lacks a specific written constitution, the date of the constitution will refer to when the specific body of laws governing the nature of the political system as it currently is were passed. However, if again there is no such specific body of law, or its introduction is spread out over a long period then the constitution is dated based upon when the elected representatives of the people first became legally politically pre-eminent.

The two variables that will be used in the analysis are *mthelct* and *mthconst* and were obtained by transforming the *dateelections* and *dateconstitution* variables into

a figure recording the number of months between the inception of democracy as measured and the start of the year 2004.

It is useful to compare *age, mthconst,* and *mthelct.* It is clear from Figure 1 that although there is a marked correlation between *mthconst*, *mthelct*, and *age* there are also some notable discrepancies. This conclusion is reinforced by the following correlations:

Variable	age	mthconst	mthelct
age	1	0.78	0.89
mthconst		1	0.81
mthelct			1

And also these Spearman's Rank coefficients:

Variable	age	mthconst	mthelct
age	1	0.74	0.89
mthconst		1	0.78
mthelct			1

Some examples of how these criteria were implemented and the most notable discrepancies, are contained in Web Appendix A. The actual data are provided in Web Appendix B.

3. Results

This section first considers new estimates obtained in the light of Acemoglu's criticisms, using Fuller and CUE estimators and the new instrument set. Secondly, it will consider the results obtained using the new measures of the the age of democracies. The results suggest improved support for PT's claims, with larger estimated coefficients, limited bias, and weak-instrument robust results. Of course, it is impossible to ever prove robustness, rather the purpose of this section and generally this paper, is to address the

key econometric and conceptual concerns discussed above. Results are reported using both the original dataset of PT and the expanded data provided by Blume, Möller, Voigt and Wolf (2009), who provide data for larger set of democracies. These democracies are in general more recent than those in the PT dataset, and include many that became democracies subsequent to the early-1990s date for which PT's data were created. On the basis of this extended dataset they argue that PT's consistent result of a negative and statistically significant OLS coefficient on *Pres* becomes consistently positive.

Table 2 considers whether PT's results are driven by their choice of a potentially biased estimator. Column 1 reports results identical to those in PT using their methodology for both their original sample, and the Blume et al. (2009) (henceforth, the expanded sample). Columns 2 and 3 suggest that for this specification, using potentially invalid instruments, that the main consequence of using 2SLS or the Fuller estimators is to increase the estimated coefficient on *Pres* by around 30% to one and one half standard deviations of the sample variance of *cgexp*. The Moreira (2003) weak-instrument robust confidence set (CCCS) suggests that the negative coefficient on *Pres* is indeed significant, and that the coefficient associated with *Maj* is not. Results using the expanded data set are reported in Columns 4-8.[14] The estimates of the effect of Presidentialism for the expanded sample are slightly smaller, but not significantly so. The estimates obtained using the CUE are larger for both the PT sample and the expanded sample.[15] Indeed, a consistent result across all of the specifications considered in this paper, is that the CUE consistently suggests a larger negative coefficient for *Pres* and higher t-statistics. This perhaps reflect the joint estimation of the coefficients and covariance matrix. Whilst, the results are consistently larger than the Fuller estimates that they qualitatively similar and consistent across specifications suggests the lack of finite-sample moments has not proved overly problematic. The CCCS, is similar to the results obtained using other PT's data although the p-value for *Pres* falls to 0.01.

Whilst, the results of Table 2 suggest that PT's choice of estimator if anything led to them underestimating their impact of presidential democracy, Acemoglu (2005)'s

[14] A lack of data for *engfrac and eurfrac* is why the sample is smaller than the full Blume et al. (2009) sample.

[15] CUE estimates corresponding to the specification in columns 1-3 are not reported but the coefficient of *Pres* is estimated as -13.11 with standard error 5.28

key criticism remains: the set of excluded instruments is unable to predict variation in the form of democracy. Table 3 reports results using the improved set of excluded instruments discussed in Section 1.1.. Column 1 limits the sample to that used by PT, and in Table 2, to facilitate comparison with the previous estimates. The coefficient of *Pres* is similar to that obtained in Table 2, but one potential cause for concern is that the F-statistic of the excluded instruments in the first-stage is low. As discussed above, formally, instruments can be considered weak when the matrix of instruments is of reduced-rank, previous tests had a variety of limitations and often the the fist stage F-statistic was considered the best heuristic. However, Kleibergen and Paap (2006) propose the 'rk' statistic which has a χ^2 limiting distribution and makes no assumptions about the structure of the error terms. As well as providing a Wald-test for whether the instruments are weak, a related LM test again following Kleibergen and Paap (2006) (KP) tests whether the model is identified. These tests suggest that whilst the new instruments are potentially weaker than those used by PT, as suggested by the KP tests, the p-value that the model is under-identified is only 0.03 Interpretation of the Wald statistics is based on the critical values tabulated in Stock and Yogo (2005), they do not provide critical values for CUE estimates, but do for LIML and the use of these seems reasonable both given that one interpretation of the CUE estimator is the generalisation of LIML to the case of possible heteroscedasticity Hansen, Heaton and Yaron (1996). Given the potential effects of weak instruments, inference requires joint consideration of estimated coefficients and their potential biases. Accordingly, the remainder of this section will pay both equal attention.

As a benchmark, it is worth noting that the KP Wald statistics reported in Table 2 suggest we are only able to reject the hypotheses that the relative bias of the 2SLS estimates is greater than 30% and 20% for the PT and the expanded samples respectively. The superior performance of the Fuller estimates is revealed by the fact that we are able to reject relative biases greater than 10% and 5% respectively, for the Fuller(1) estimators.

The performance of the new instruments, from the perspective of potential bias due to weak-identification is mixed. For the PT sample, only a maximum bias of 30% cannot be rejected, whilst for the sample of all observations in the original PT dataset the KP Wald-statistic associated with the Fuller estimators, fails to reach the necessary

critical to reject even this level of bias. However, when the expanded sample is used, that whilst the model is better identified for both the pre-1992 and full samples, there is little robust evidence for an effect of presidentialism in full sample. Specifically, for both samples the maximum bias is now estimated as 10%. But, the CCCS p-value is now 0.16 for the full-sample. That the estimated coefficients are qualitatively smaller as well suggest that the effect of presidentialism is smaller in more recent democracies. This finding is consistent with those of Blume et al. (2009). Smaller estimates for more recent democracies are perhaps to be expected. If, as Persson, Roland and Tabellini (2000) argue, smaller governments in presidential democracies, is an equilibrium effect of the separation of powers it is expected that the impact of constitutional type takes time to emerge. This may be specifically due to the relative infrequency of budget-setting and the large number of confounding factors.[16] The CCCS estimates, suggest that this is not just a consequence of weak identification. Whilst, weak identification could in theory bias the estimated Fuller or CUE coefficients downwards, the CCCS p-values are correct for arbitrarily weak-identification and suggest a lack of significant effect in the full sample.

To facilitate comparison, Columns 1 and 2 of Table 4 correspond to columns 2 and 3 of Table 3. The results suggest that the choice of age measure has little impact on the estimated coefficients, perhaps increasing them slightly. This suggests that, despite its limitations, the use of *age* is not driving the results. However, the estimated maximum bias of the Fuller coefficients is now smaller, suggesting that consistent with the overall identification strategy, that better measures of democratic age, predict *pres* and *maj* better. Crucially, the rk Wald statistic is now sufficient to reject bias of more than 20% for the estimates employing *mthelct* and *mthconst*. The size test associated with the CUE estimates suggests there is only a very minor improvement with a move from 15% to 10% of the maximal LIML size. This suggests that a 5% confidence intervals of *pres* would at worst correspond to a 10% confidence interval. Equivalently, this means that at worst the p-value of the CUE estimates would increase from 0.01 to 0.02, but in general the estimates remain comfortably significant at the 1% level.

[16] As Blume et al note, the average country in the expanded dataset has a higher Gastil (quality of democracy) index, is more likely to be in Africa, is closer to the equator, and more likely to speak English. They are also significantly smaller, and more likely to be islands. Whether, it is differences in these characteristics instead of their shorter history of democracy that accounts for the smaller impact of *pres* is unclear.

Columns 7 and 8 report results for an alternative sample that restricts the sample to countries that became democracies prior to 1992 according to the two measures. These are not a sub-sample of the APT countries, rather from the extended sample. The results suggest a smaller impact of presidentialism for this sample, and the CCCS only gives a p-value of around 0.05. The smaller estimated *pres* coefficient is consistent with the results obtained in columns 4-7 of Table 3 and that obtained in column 8 of Table 2.

Overall, based on results obtained using a variety of samples, estimators, excluded instrument sets, and measures of democratic age, the general finding must be that there is indeed consistent evidence for a quantitatively large negative effect of presidentialism on the size of government. A key exception is that there is no such result if countries that became democracies after the early 1990s contained in the expanded dataset of Blume et al. (2009) are included. Of central importance to the results obtained is that both the CCCS and the kp-Wald test suggest that the results are not an artefact of a weak-instruments problem, whilst the use of the new excluded-instrument set is motivated by the requirements of exogeneity as well as an ability to explain differences between institutions.

Less emphasis has been placed on the results concerning the impact of majoritarian elections (*maj*). This is in part because it was statistically insignificant in PT's original analyses and in no specification considered here does that change. However, this is not an argument for disregarding it, including *maj* as an exogenous variable gives rise to significant negative coefficients, underlining the importance of treating it as potentially endogenous, and giving it equal *a priori* importance, when choosing the excluded instruments.

4. Conclusions

This paper has addressed three problems in the approach of PT: the use of a potentially biased estimator, the irrelevance of the excluded instruments, and the weakness of the measure of the age of democracies used. The first two issues were raised by Acemoglu (2005) and the Fuller and CUE estimators and an alternative instrument set was proposed in response to his concerns. The results obtained with these new

estimators and using the new instruments provide support for the conclusions of PT that presidential democracies are associated with smaller governments as a percentage of GDP.

The paper also introduced two new measures of the inception of democracy. The results obtained using these improved measures lend further credibility to the PT results. However, it remains the case that no significant relationship is identified for an expanded dataset, when the sample includes countries that became democracies subsequent to the early 1990s. If the consequences of constitutions take time to emerge then this might be expected. Other than these exceptions, the consistency in results would suggest that the debate can only intensify as to whether Acemoglu (2005) is correct when he claims PT have made "the most significant contribution to this field since Lipset's work almost 50 years ago."

Table 1: Summary Statistics

	N	$Mean$	Std	Min	Max
cgexp	84	28.28	10.53	9.74	51.18
pres	84	0.39	0.49	0.00	1.00
maj	84	0.42	0.50	0.00	1.00
lyp	83	8.44	0.99	6.27	9.94
trade	84	79.61	49.16	17.56	343.39
age	84	1.20	1.11	0.29	5.17
prop1564	83	61.44	5.88	49.05	71.70
prop65	83	7.85	4.88	2.30	17.43
gastil	84	2.38	1.20	1.00	4.89
federal	84	0.16	0.36	0.00	1.00
oecd	84	0.27	0.45	0.00	1.00
con2150	84	0.11	0.31	0.00	1.00
con5180	84	0.25	0.44	0.00	1.00
con81	84	0.44	0.50	0.00	1.00
lat01	84	0.31	0.19	0.00	0.71
engfrac	84	0.15	0.33	0.00	1.00
eurfrac	84	0.40	0.44	0.00	1.00
coluka	84	0.32	0.41	0.00	0.93
colespa	84	0.07	0.15	0.00	0.79
excolony	84	0.62	0.49	0.00	1.00
mthconst	76	1.08	1.03	0.13	4.58
mthconst2	76	2.21	4.15	0.02	21.01
mthelct	76	1.09	1.02	0.13	4.26
mthelct2	76	2.21	4.15	0.02	21.01
mm	76	0.57	0.51	0.09	2.09
mm2	76	0.58	1.00	0.01	4.37

Table 2: Government Size and Constitutional Type: Alternative Instrumental Variable Estimators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	PT	2SLS	Fuller(1)	PT	2SLS	Fuller(1)	Fuller(4)	CUE
pres	-8.65**	-11.29**	-11.32**	-7.39**	-9.68**	-10.07**	-9.15**	-11.47***
	(3.63)	(5.26)	(5.28)	(3.52)	(4.29)	(4.61)	(3.89)	(3.68)
maj	-3.90	-4.23	-4.24	-2.03	-3.71	-3.84	-3.56	-5.37
	(3.46)	(3.14)	(3.15)	(3.88)	(3.96)	(4.29)	(3.54)	(3.58)
Excluded Instruments	PT	PT	PT	PT	PT	PT	PT	PT
1st-Stage F-Statistic of	9.82	19.27	19.27	11.33	10.09	10.09	10.09	10.09
excluded Instruments *Pres*	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
1st-Stage F-Statistic of	6.68	12.70	12.70	5.25	5.88	5.88	5.88	5.88
excluded Instruments *Maj*	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
KP LM Test		14.46	14.46		16.51	16.51	16.51	16.51
statistic		(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Hansen J		1.68	1.68		2.29	2.27	2.33	2.09
statistic		(0.79)	(0.79)		(0.68)	(0.67)	(0.68)	(0.72)
KP Wald		5.80	5.80		6.26	6.26	6.26	6.26
statistic								
Coverage Corrected		[-29.78,-0.74]					[-25.44,-2.70]	
Confidence Set *Pres*		(0.04)**					(0.01)***	
Coverage Corrected		[-10.79,5.53]					[-12.70,17.42]	
Confidence Set *Maj*		(0.46)					(0.97)	
R^2	0.59	0.63	0.63	0.57	0.57	0.56	0.58	0.54
N	75	75	75	82	82	82	82	82
Dataset	PT(2004)	PT(2004)	PT(2004)	Expanded	Expanded	Expanded	Expanded	Expanded

The excluded instrument set 'PT' contains: $con2150, con5180, con81, lat01, engfrac,$ and $eurfrac$. The excluded instrument set 'NEW' is: *age,age2, coluka, excolony, catho80,*and *colespa*. The included instruments are *lyp , prop1564, prop65, Gastil,oecd* and *federal*. Values in parenthesis are robust standard errors for the coefficients associated with *pres* and *maj* and p-values elsewhere.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Government Size and Constitutional Type: Alternative Excluded Instruments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimator	Fuller(4)	Fuller(4)	CUE	Fuller(4)	CUE	Fuller(4)	CUE
pres	-11.42***	-11.75***	-14.66***	-9.99***	-12.03***	-5.61*	-8.89***
	(3.54)	(3.89)	(3.70)	(2.83)	(2.78)	(3.37)	(2.92)
maj	-2.25	-1.97	-3.36	-1.75	-2.81	1.56	0.34
	(2.42)	(2.45)	(2.18)	(2.67)	(2.52)	(3.18)	(2.98)
Excluded Instruments	NEW	NEW	NEW	NEW	NEW	NEW	NEW
1st-Stage F-Statistic of	5.64	4.97					
excluded Instruments *Pres*	(0.00)	(0.00)					
1st-Stage F-Statistic of	11.73	11.74					
excluded Instruments *Maj*	(0.00)	(0.00)					
KP LM Test	10.70	10.45	10.45	10.88	10.88	16.69	16.69
statistic	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)
Hansen J	3.36	3.71	3.17	3.07	2.80	4.90	4.16
statistic	(0.34)	(0.29)	(0.37)	(0.38)	(0.42)	(0.18)	(0.25)
KP Wald	4.90	4.30	4.30	6.97	6.97	6.95	6.95
statistic							
Coverage Corrected	[-28.21,4.08]	[-31.10,-3.74]		[-24.11,-2.85]		[-16.00,2.75]	
Confidence Set *Pres*	(0.01)***	(0.01)***		(0.01)**		(0.16)	
Coverage Corrected	[-7.08,6.94]	[-6.76,7.19]		[-7.38,8.73]		[-3.94,12.11]	
Confidence Set *Maj*	(0.84)	(0.93)		(0.98)		(0.43)	
R^2	0.62	0.58	0.53	0.57	0.55	0.52	0.49
N	75	80	80	80	80	92	92
Dataset	PT(2004)	All PT(2004)	All PT(2004)	Pre-1992	Pre-1992	Expanded	Expanded

The excluded instrument set 'NEW' is: *age,age2, coluka, excolony, catho80,*and *colespa.* The included instruments are *lyp* , *prop1564, prop65, Gastil, oecd* and *federal.* Values in parenthesis are robust standard errors for the coefficients associated with *pres* and *maj* and p-values elsewhere.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Government Size and Constitutional Type: Alternative Measures of the Age of Democracy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	Fuller(4)	CUE	Fuller(4)	CUE	Fuller(4)	CUE	CUE	CUE
pres	-11.75***	-14.66***	-13.39**	-16.87***	-11.24*	-16.85***	-11.84***	-11.71***
	(3.89)	(3.70)	(6.03)	(4.98)	(6.12)	(5.02)	(4.34)	(3.98)
maj	-1.97	-3.36	-3.40	-5.77*	-1.70	-4.46	-2.39	-3.06
	(2.45)	(2.18)	(4.24)	(3.25)	(4.21)	(3.22)	(3.77)	(3.20)
Excluded Instruments	NEW	NEW	NEW	NEW	NEW	NEW	NEW	NEW
Age Variable	*age*	*age*	*mthelct*	*mthelct*	*mthconst*	*mthconst*	*mthelct*	*mthconst*
KP LM Test	10.45	10.45	7.23	7.23	7.57	7.57	8.19	9.76
statistic	(0.03)	(0.03)	(0.12)	(0.12)	(0.11)	(0.11)	(0.09)	(0.05)
Hansen J	3.71	3.71	3.37	2.85	2.92	4.05	1.10	2.25
statistic	(0.29)	(0.29)	(0.34)	(0.42)	(0.40)	(0.26)	(0.78)	(0.52)
KP Wald	4.30	4.30	5.00	5.00	5.33	5.33	5.76	5.32
statistic								
Coverage Corrected	[-31.10,-3.74]		[-57.91,-6.94]		[-59.94,-4.84]		[-42.68,-1.14]	[-30.22,-0.29]
Confidence Set *Pres*	(0.01)***		(0.00)***		(0.01)***		(0.04)**	(0.05)**
Coverage Corrected	[-6.76,7.19]		[-6.31,18.54]		[-5.30,16.61]		[-6.79,15.61]	[-7.66,14.78]
Confidence Set *Maj*	(0.93)		(0.55)		(0.48)		(0.64)	(0.76)
Dataset	APT	APT	APT	APT	APT	APT	Elections Pre-1992	Constitution Pre-1992
R^2	0.58	0.53	0.58	0.51	0.59	0.49	0.53	0.57
N	80	80	80	80	80	80	73	73

The excluded instrument set 'NEW' always contains: *coluka, excolony, catho80,*and *colespa.* It also one of the following pairs *age, age2, mthelct, mthelct2,* or *mthconst, mthconst2* as indicated in the fourth row of the table. The included instruments are *lyp , prop1564, prop65, Gastil, oecd* and *federal.*
Values in parenthesis are robust standard errors for the coefficients associated with *pres* and *maj* and p-values elsewhere.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

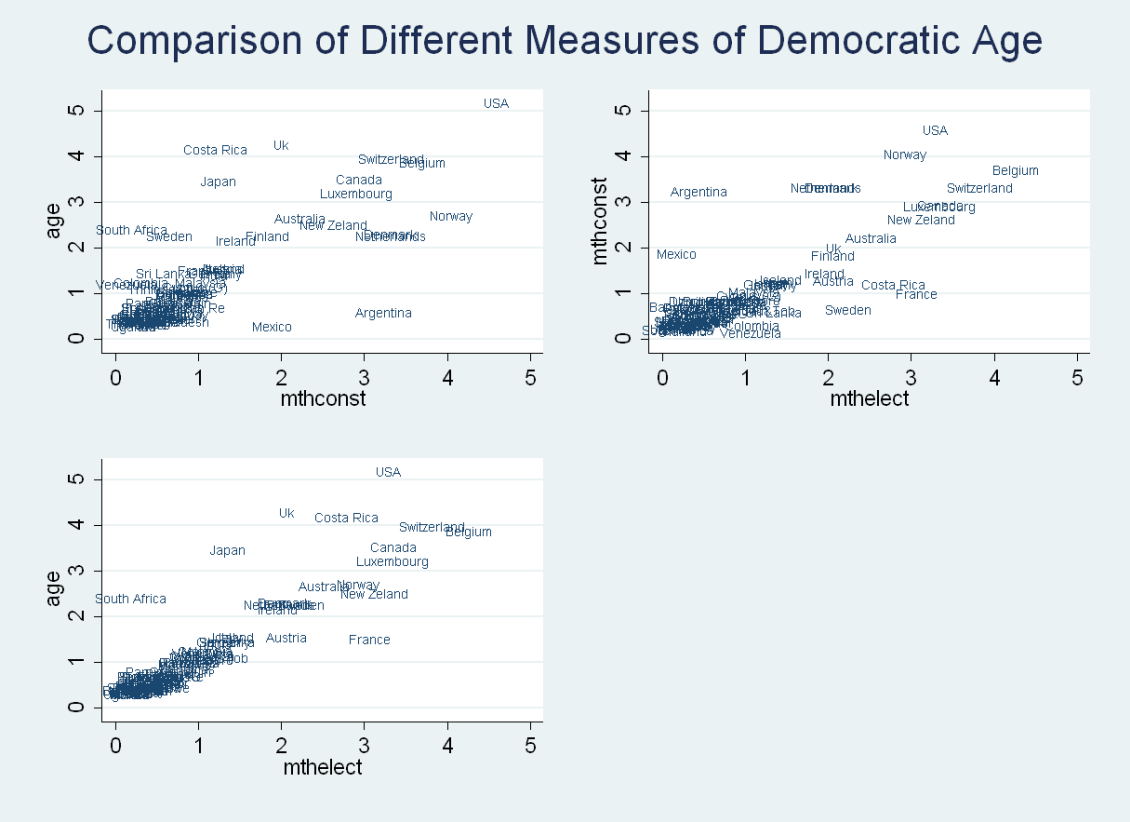

Figure 1: Scatter plots comparing age, mthconst, and mthelct

References

Acemoglu, Daron, "Constitutions, politics, and economics: A review essay on Persson and Tabellini's The Economic Effects of Constitutions," *Journal of Economic Literature*, 2005, *43* (4), 1025–1048. 48.

__ **and James A. Robinson**, *Economic Origins of Dictatorship and Democracy* number 9780521671422. In 'Cambridge Books.', Cambridge University Press, February 2006.

__ **, Simon Johnson, and James A. Robinson**, "The colonial origins of comparative development: An empirical investigation," *American Economic Review*, 2001, *91* (5), 1369–1401. 82.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st ed., Princeton University Press, 2009.

__ **, Guido W. Imbens, and Alan B. Krueger**, "Jackknife instrumental variables estimation," *Journal of Applied Econometrics*, 1999, *14* (1), 57–67.

Baum, Christopher F, Mark E Schaffer, and Steven Stillman, "IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation," Statistical Software Components, Boston College Department of Economics April 2002.

Besley, Timothy and Anne Case, "Political institutions and policy choices: Evidence from the United States," *Journal of Economic Literature*, 2003, *41* (1), 7–73.

Blomquist, Soren and Matz Dahlberg, "Small sample properties of LIML and jackknife IV estimators: Experiments with weak instruments," *Journal of Applied Econometrics*, 1999, *14* (1), 69–88.

Blume, Lorenz, Jens Möller, Stefan Voigt, and Carsten Wolf, "The economic effects of constitutions: replicating and extending Persson and Tabellini," *Public Choice*, April 2009, *139* (1), 197–225.

Fuller, Wayne A., "Some Properties of a Modification of the limited information estimator," *Econometrica*, 1977, *45* (4), 939–953.

Guggenberger, Patrik, "Finite Sample Evidence Suggesting a Heavy Tail Problem of the Generalized Empirical Likelihood Estimator," *Econometric Reviews*, 2008, *26* (4-6), 526–541.

Hahn, Jinyong and Jerry Hausman, "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *AEA Papers and Proceedings*, 2003, *93* (2), 118–125.

_ , _ , and Guido Kuersteiner, "Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal*, 2004, *7* (1), 272–306.

Hall, Robert. E. and Charles. I. Jones, "Why do some countries produce so much more output per worker than others?," *Quarterly Journal of Economics*, 1999, *114* (1), 83–116. 58.

Hansen, Lars Peter, John Heaton, and Amir Yaron, "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, July 1996, *14* (3), 262–80.

Honoré, Bo and Luojia Hu, "On the Performance of Some Robust Instrumental Variables Estimators," *Journal of Business & Economic Statistics*, 2004, *22* (1), 30–39.

Jensen, Nathan and Leonard Wantchekon, "Resource wealth and political regimes in Africa," *Comparative Political Studies*, 2004, *37* (7), 816–841. 39.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 2006, *133* (1), 97 – 126.

Leamer, Edward E., Hugo Maul, Sergio Rodriguez, and Peter K. Schott, "Does natural resource abundance increase Latin American income inequality?," *Journal of Development Economics*, 1999, *59* (1), 3–42.

Mikusheva, Anna and Brian P. Poi, "Tests and confidence sets with correct size when instruments are potentially weak," *The Stata Journal*, 2006, *6* (3), 335–347.

Milesi-Ferretti, Gian Maria, Roberto Perotti, and Massimo Rostagno, "Electoral Systems and the Composition of Public Spending," *Quarterly Journal of Economics*, 2002, *117*, 609–657.

Moreira, Marcelo J., "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 07 2003, *71* (4), 1027–1048.

Olsson, Ola, "Unbundling Ex-Colonies: A comment on Acemoglu, Johnsonm, and Robinson (2001)," 2004.

Persson, T., G. Roland, and G. Tabellini, "Comparative politics and public finance," *Journal of Political Economy*, 2000, *108* (6), 1121–1161. 50.

Persson, Torsten, "Presidential Address: Consequences of Constitutions," *Journal of the European Economic Association*, 04/05 2004, *2* (2-3), 139–161.

__ **and Guido Tabellini**, *Political Economics: Explaining Economic Policy* Zeuthen Lecture Books, 1st ed., Cambridge, MA: The MIT Press, 2000.

__ **and** __ , *The Economic Effects of Constitutions* Munich Lectures in Economics, 1st ed., Cambridge, MA: MIT Press, 2003.

__ **and** __ , "Constitutional rules and fiscal policy outcomes," *American Economic Review*, 2004, *94* (1), 25–45. 29.

__ **, Gérard Roland, and Guido Tabellini**, "Separation of powers and political accountability," *Quarterly Journal of Economics*, 1997, *112* (4), 1163–1202. 32.

Porta, Rafael La, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny, "Law and finance," *Journal of Political Economy*, 1998, *106* (6), 1113–1155.

Rodrik, Dani, "Why Do More Open Economies Have Bigger Governments?," *Journal of Political Economy*, October 1998, *106* (5), 997–1032.

Smith, Richard, "Weak instruments and empirical likelihood: a discussion of the papers by DWK Andrews and JH Stock and Y Kitamura," CeMMAP working papers CWP13/05, Centre for Microdata Methods and Practice, Institute for Fiscal Studies October 2005.

Stock, James. H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression.," in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge, UK: Cambridge University Press, 2005, p. 588.

__ **, Jonathan. H. Wright, and Motohiro. Yogo**, "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business & Economic Statistics*, 2002, *20* (4), 518–529.

Tschentscher, Axel, "International Constitutional Law Project," www.verfassungsvergleich.de December 2009.