A New Cross-National Time Series Indicator of Bureaucratic Quality

Andrew Williams∗
Abstract

In this paper, a new indicator of bureaucratic quality is developed that has extensive coverage across countries (175) and time (1960-2000), based on data collection from the World Development Indicators and the International Finance Statistics databases. Both anecdotal and econometric evidence suggest this new indicator could prove a useful institutional tool for empirical researchers, particularly with respect to panel data analyses.

Key words: bureaucracy, institutional quality, economic growth.

JEL Classification: O17, O47, P14.
Section 1: Introduction

Empirical research on the links between institutional quality and economic development has grown almost exponentially over the past twenty years, and economists have gained many useful insights as a result. This empirical research, however, has been hampered to a large extent by problems with the existing institutional data that is available (see Williams and Siddique, forthcoming, for a review of some of these issues). One of the major problems relates to the lack of adequate institutional data with a relatively long time series component. This has meant that the vast majority of studies have been, by necessity, cross-sectional analyses. Despite the impressive work that has been undertaken in cross-sectional studies, there are a number of potential methodological problems inherent within this type of estimation, particularly in relation to the endogeneity of the institutional variable used.

This paper is an attempt to at least partially fill this gap. Focussing on one specific area of institutional quality, the bureaucracy, a new measure is developed that has extensive coverage across both countries (175) and time (1960-2000). The paper proceeds as follows. Section 2 will take a brief look at the existing indicators of bureaucratic quality currently used in empirical research. Section 3 will then develop this new indicator of bureaucratic quality, which is based on the data collection abilities of governments around the world. Having developed this new measure, Section 4 incorporates a visual comparison between this new bureaucratic quality index with the most commonly-used current measure, the International Country Risk Guide (ICRG) from Political Risk Services (2003), and then in terms of a small number of countries that have experienced particularly acute instances of political (and consequently bureaucratic) turmoil. Section 5 employs a more formal econometric analysis of this indicator, using both cross-sectional and panel data analysis. Finally, Section 6 will summarise these results and point towards future areas of research.

Section 2: Overview of existing measures of bureaucratic quality

Recent research on the role of the bureaucracy in economic development has tended to focus on the deleterious effects of a poorly-functioning bureaucracy on development, such as corruption. With respect to what makes a ‘good’ bureaucracy, most research has focussed on the ‘Weberian’ meritocracy, whereby civil servants are rewarded and promoted based on their talents, rather than any tribal or ethnic affiliation. Where rewards are based on merit, the bureaucracy will be run on a more efficient basis and will facilitate,
rather than obfuscate, market mechanisms (see, for example, Rauch and Evans, 2000, and Evans and Rauch, 1999).

This paper, however, does not take a sociological perspective on this issue. Rather, it is the empirical measurement of this quality that is the issue. Overwhelmingly, empirical research in this field has by necessity used a very narrow range of sources. The most common one is the measure of bureaucratic quality developed by Political Risk Services, the *International Country Risk Guide* (ICRG). Here, countries are given a score (1-4), which is derived by country experts. This has been used extensively in the literature, commonly as part of a composite index encapsulating the more general ‘institutional quality’ of a country (most notably by Knack and Keefer, 1995). Mauro (1995) used a measure from another private firm, Business International (BI), that included ‘bureaucracy and red tape’ and ‘corruption’ to get an overall indicator of ‘bureaucratic efficiency’.1

As research on institutions gained momentum, there was a concomitant increase in the number of indicators developed that purported to measure various aspects of institutional quality. Arising from this proliferation of indicators were the composite indicators created by Kaufmann et al (1999), who attempted to combine these wide range of empirical measures of institutions into several composite indicators (the ‘KKZ’ indicators). Although there is no single specific ‘bureaucratic quality’ measure, their sub-category of Government Effectiveness is a reasonably close approximation. Kaufmann and Kraay (2002: 6) define this as the:

“... perceptions of the quality of public service provision, the quality of the bureaucracy, the competence of civil servants, the independence of the civil service from political pressures, and the credibility of the government’s commitment to policies into a single grouping”

Although the empirical literature on institutions has focussed more on their ‘Rule of Law’ composite indicator,2 Government Effectiveness has been used in a combined index of institutions by Easterly and Levine (2003), among others.

However, despite the existence of these measures of the bureaucracy, there are some inherent problems with each of them. The ICRG measure, for example, although it now has extensive coverage across countries (140), has data that only goes back to 1984. More importantly, however, the index suffers
somewhat from the fact that it is constructed by country experts. Irrespective of how expert these people are (and I am not doubting their knowledge here), the problem is that generally they can only perceive and report on a problem when it becomes public knowledge. In other words, the rating score can sometimes lag the major events it is purporting to predict. For example, Linder and Santiso (2002: 14), who in a broader context investigated the predictive powers of the ICRG’s Economic, Financial and Political Risk Ratings in Brazil, Argentina and Peru in the late 1990s suggested that although the Economic and Financial Ratings performed reasonably well:

“A closer look at ... the political risk rating, which typically is based on survey data and individuals’ perceptions, is particularly vulnerable to misinterpretation, as it appears to have reacted to actual events rather than predicted them. This finding thus leads us to question whether the political risk indicator of the ICRG model behaves more as a lagging indicator rather than a leading indicator of crises.”

In cross-sectional analyses where long-run averages are taken this probably does not matter too much, as they are at least picking up these political factors at some point. It does, however, matter if one wants to see whether institutional changes predate changes in economic activity. The criticism surrounding the generation of indicators based on subjective opinions also relates to some of the other indicators in the literature, such as those compiled by Business International (BI) and Business Environment Risk Intelligence (BERI).

The newer indicators, such as the KKZ indicators, have the advantage that they have been compiled from a wide variety of sources, and they cover an extremely wide number of countries, however they suffer from the fact that they are a relatively recent construct, with data dating back only to 1996. Again, this makes them unsuited to time series analysis.

Despite the existing paucity of institutional measures with adequate time series dimensions, the question remains over whether such a measure is needed in the first place. Does institutional quality, or any of its sub-components, change so slowly over time that a time series approach is largely redundant? Although it may be true that institutional quality for many countries has altered little (or very slowly) over the past forty years, there are also examples where institutional quality has changed, both for the better and for the worse. Writing with respect to the World Bank’s KKZ Indicators, Kaufmann et al (2005: 4), reviewing the scores for countries over the eight years of the indices’ existence (1996-2004), noted how around 10% of
countries showed an appreciable and statistically significant change in institutional scores, even over this short period of time. This led them to state that:

“...while in general institutional quality changes only gradually, there are also countries where one can point to sharp improvements or deteriorations over an eight-year period. This finding is of particular interest given the common perception that, while deterioration in a particular country can take place rather quickly, improvements are always very slow and incremental.”

That this can be observed over such a short period of time suggests that there is the potential for much to be learnt by extending an institutional quality indicator to cover the medium term.

**Section 3: Development of the bureaucratic quality indicator**

At the present point in time there has been no attempt to derive an indicator of bureaucratic quality that has extensive coverage over time, or at least an existing adequate *proxy* for bureaucratic quality, due in part to data limitations. In other institutional-related areas, such as property rights, some have tried to overcome the lack of direct historical information by using indirect, existing historical data to proxy for some institutional measure. For example, Clague *et al* (1996, 1999) have used a proxy for contract or property rights known as Contract Intensive Money (CIM). Their argument followed the Williamson (1995) hypothesis that the existence of long-term contracts was a sign of a developed economy, as it showed confidence in dealing with other parties. If this trust existed, then investment would be higher. They argued that if this were true, then this would be reflected in a high proportion of the money supply being held in financial institutions (indicating long term, high value transactions were taking place). Conversely, the greater the proportion of money held in currency, the less faith people had in making these transactions. Furthermore, they felt that during times of instability, more people would hold their wealth in currency due to the uncertainty over economic conditions.

The benefit of using CIM as a proxy for contract rights is that the data is available for many countries over a relatively long period of time (from 1960). Higher values of CIM indicate a greater reliance on or preference for long-term contracts. Other relatively recent measures include the Budgetary and Revenue Source Volatility indicators, which uses existing data from the IMF’s Government Finance Statistics to look at the volatility of public spending and revenue decisions (see Knack, Kugler and Manning, 2003).
The more volatile they are, the poorer is the institutional quality of the country. The measure developed below, while not using any of these sources, is in the same vein, in that it uses existing data to proxy for bureaucratic quality.

One of the problems that arises in developing proxies for institutional quality using existing data is that, for many countries, the data itself is quite limited. Data limitations in general are a common problem in many areas of empirical macroeconomic research. For example, OECD countries are often over-represented in samples, while many other countries are often omitted from analyses entirely because no data exists for them. As Knack, Kugler and Manning (2003: 4) note:

“...countries ranked near the bottom on indicators ... conceivably are not actually among the most poorly-governed countries in the world, but instead may just be the most poorly-governed among those with a reasonable capacity for statistical reporting. Countries without such minimum capacity could be rewarded, in effect, for their inability or unwillingness to report data.”

This lack of data availability, however, warrants further examination. It is obviously only ever possible to include countries in an empirical analysis if data exists for them. However, perhaps the degree of data coverage can in itself tell us something about the administrative capacity (and quality) of a government. For example, the reason OECD countries are often over-represented in research is precisely because they have the administrative capacity to collect, collate and distribute information across a wide range of issues, which therefore allows researchers to run empirical tests.

One may suspect that a particular country has poor bureaucratic capacity and quality, but if no data exists then it is almost impossible to tell. Therefore, this new indicator introduced here is a measure of administrative capacity, based on the collection of data. The less data that is available, the weaker the administrative capacity of the government. One of the benefits of this, of course, is that a score can technically be developed for every country over every period (from 175 countries, nearly 6,000 observations are recorded, giving an average of 34 annual observations for each country between 1960-2000).

One of the immediate possible objections to the rationale behind using data collection as an indicator of bureaucratic quality is that it may not necessarily reflect bureaucratic quality, but rather a lack of public
resources devoted towards their collection – poor countries have poor statistics because they are poor. Although this is undoubtedly an issue, Chart 1 below shows that income is certainly not the defining characteristic of statistical capacity. Drawing on the Statistical Capacity Indicator (SCI) developed by the World Bank (for further details see Section 3.3 below) and GNI (in $US), the correlation between the two for low income countries (less than $5,000 per capita) is quite low. As the World Bank (2004: 7) notes: “…higher income does not guarantee better statistical capacity, and a relatively high level of statistical capacity can be attained in a low-income environment.”

![Chart 1: Relationship of Statistical Capacity Indicator (SCI) with GNI Per Capita](image)


It should also be reiterated at this point that the suggestion here is not that a country has a competent bureaucracy just because it collects a lot of data. Rather, the data collection process is a signal, or proxy, of bureaucratic competence, because it shows that the civil administration has the ability to organise the appropriate mechanisms to enable this information to be collected in a consistent and timely basis. The collection of statistics is rarely, if ever, the sole responsibility of one public body. Specific data will be put together by different departments and agencies across the whole of the government, and so data collection provides a useful example of bureaucratic co-ordination. If the bureaucracy can competently organise and disseminate this data, then they should also be competent in other areas of bureaucratic responsibility as well. Writing specifically on the situation of statistical capacity in Cambodia, the IMF (2005: 24) notes:
“...the current formal coordination mechanisms between RGC [Royal Government of Cambodia] institutions and ministries for statistical activities and capacity building have proven to be severely constrained. Communication among ministries about information sharing generally requires a written request by one minister to the other minister — a process that can typically take about a month to get the desired result.”

In addition to this lack of timely co-ordination, in an earlier passage (IMF, 2005: 12), it is also noted that:

*The NIS [National Institute of Statistics] and the few functioning RGC statistics units do not have the legal and institutional requisites to effectively collect the range of quality statistics required to monitor and measure development results.*

For many countries, it is certainly not unreasonable to assume that this lack of efficiency, poor co-ordination and concomitant institutional difficulties can be multiplied across the many functions and responsibilities of a bureaucracy, not just statistical collection.

There may, additionally, be a political element to the process of data collection. The modern focus on the collection of economic statistics can be traced back to Keynes during the Second World War (Maddison, 2005). Keynes felt that in order to direct resources to the war effort, Britain needed to know exactly what resources it had at its disposal. This led to the first substantial effort to measure a country’s national income. The success of this had a flow-on effect after the end of the war, as economists (and politicians) saw the benefits to policy-making in improving the stock of information available. Since that time there has been a proliferation in both the quantity (and quality) of data available from OECD countries and, increasingly, to other countries around the world.

In other words, politicians who have a genuine desire to promote economic development in their country may place more emphasis on the collection of data in their economy in order to better achieve their goal. For example, Chen, Haggard and Kang (1998: 103), discussing the role of the bureaucracy in South Korea, note:

“...the close interest in economic developments was visible after 1964 in the construction of an ‘economic situation room’ that allowed Park [President Park Chung Hee] to monitor progress on specific projects on a daily, even hourly, basis...”
This desire to know what is happening in the economy requires not only the preference for obtaining this information, but also ensuring the bureaucracy has the competence and organisation to provide it to them.5

If one accepts the underlying premise of this argument, then the next obvious question, of course, is ‘What data?’ In order to capture as wide a variety of data as possible, I have taken the two main international databases that are currently used extensively in economic analyses: the World Development Indicators (WDI) produced by the World Bank, and the International Financial Statistics (IFS) database, constructed by the International Monetary Fund. Although each country has in some form its own statistical agency, and there are also many other regional data collection bodies (for example, the OECD, or the Asian Development Bank), these two databases ensure a commonality in methodology across all countries. At a general level, the IFS data can be thought of as examining the quality of the national bodies charged with the collection of financial data, which is largely the central bank and the Finance departments, whereas the WDI captures the data collection abilities across the whole of the government.

Although the easiest approach would be to just sum up the number of observations recorded for a country in a particular year over both databases, there are several reasons why this may be overly simplistic. Firstly, categories in both databases include many instances of ‘doubling up’, both across and within the respective databases. For example, GDP data is often expressed in terms of current Local Currency Units (LCUs), but is also expressed in constant LCUs, current US dollars, constant US dollars and so on. As long as a country has data recorded in current LCUs, the other transformations can be performed without any input from the domestic government, and so says nothing about their administrative capacity. This would have had the effect of magnifying the scores for countries which had data for these, as the same issue is recorded five or six times. Therefore, where an economic or social indicator appears in a multitude of transformations, I have only counted it once (generally this was the observation recorded in current LCUs). This also extends to data that appears in both the WDI and IFS databases, which was counted only once as well.

It should also be pointed out at this stage that a lot of the information from both databases are collected from a variety of primary sources. For example, data on balance of payments is taken from the IMF’s Balance of Payments Statistics database, while data on government spending and revenues are taken from the IMF’s Government Finance Statistics. The World Development Indicators also draw heavily
on data from a wide variety of sources, including various UN agencies, the International Labor Organisation, World Health Organisation and so on. As much care as possible has been taken to ensure that the data ultimately used here has required some form of domestic input from the country. For example, a number of categories in the WDI come from the International Telecommunications Union (ITU). The ITU collects their data by sending annual questionnaires to governments, which the government fills out and returns. Although this certainly does not guarantee the quality of the data, it does mean that it is the domestic government’s responsibility to collect the data and does not, for example, rely on the data being collected in each country by the relevant NGO or agency through some form of survey (in which case the ultimate scores would be more a reflection of the capabilities of the NGO that is collecting the data, rather than the government itself).

There were also issues relating to each individual database that required attention.

3.1 International Financial Statistics

The IFS is constructed using a 13-digit code. These 13 digits comprise the following sections:

- positions 1-3 is the country code;
- position 4 is the topic code;
- position 5 is the sub-section of that topic;
- positions 6-7 are the classification codes;
- position 8 is the qualification code;
- position 9 is the version code;
- position 10 is the source code;
- positions 11-13 is the partner or commodity market code.

The main data has been drawn using positions 4 and 5 (that is, the topic and sub-section codes). The topics covered by the IFS are:

- Topic . (dot): Exchange rate, Fund position or international liquidity;
- Topic 1: Monetary Authorities;
• Topic 2: Deposit Money Banks;
• Topic 3: Monetary Survey;
• Topic 4: Other Banking / Non-Bank Financial Institutions;
• Topic 5: Banking / Financial Survey;
• Topic 6: Interest, Prices, Production and Labour;
• Topic 7: International Transactions;
• Topic 8: Government Finance;
• Topic 9: National Accounts and Population.

However, not all topics were used, as there are a couple of survey topics that are largely summaries of data from other topics. For example, Topic 3 is the Monetary Survey, which incorporates information from Topics 1 and 2. This is also true for Topic 5, which is the Banking / Financial Survey. In order to simplify matters and avoid double-counting, I have excluded the data from the two surveys (Topics 3 and 5). Furthermore, I have tried to maintain the goal of only using data that requires at least some input from the domestic government. This means that Topic . (dot) has been excluded, as the information on exchange rates and Fund position could be collected without any assistance from the domestic government. The final adjustment was to exclude several categories from the qualification code (Position 8). These included qualifications:

• B: seasonally adjusted by IMF;
• E: transactions in convertible currencies;
• F: US dollars seasonally adjusted;
• S: SDRs
• U: SDRs, seasonally adjusted.
• T: for high inflation countries
• K, M, N, O: development institutions
• F, G, H, I, J: qualification for non-bank institutions;6
• X: exchange rate index, % change

This again follows from the rationale that where possible only data in its raw form would be used.

3.2 World Development Indicators Data

The collection of data from this source was considerably easier than for the IFS, but the same principles were applied, namely that no source from the other database was included here as well, nor was there any ‘double-dipping’ from data repeated in different transformations.

Having decided which categories from both databases to include, scores were then derived by taking the proportion of data coverage for each country for each individual year. Because there has been a general increase in data coverage over this time, the proportions were taken by dividing a country’s raw score in time $t$ (that is, the sum of the number of data observations from both databases) by the number of categories that had data for at least one country for that year.\footnote{7}

A final caveat was also imposed at this stage, whereby a score was only registered for a country if it was an independent nation in that year. That is, if a country was still a colony at some point between 1960-2000 (or, indeed, did not exist at all), then they were omitted from the analysis for the years they were not an independent nation.

3.3 Incorporating the Quality of Data

One of the major problems with using data coverage as a proxy for bureaucratic quality is that the data only represents the quantity of the data, and not the quality. Expressed slightly differently, a country with a well-functioning, competent bureaucracy would not just produce more information, but more accurate information as well. This is, unfortunately, an extremely difficult concept to measure, and I make no claim that my solution here is a complete one.

In order to get a sense of the quality of the data, a crude measure based on three indicators was derived. The first is derived in part from the PENN World Tables (6.1), which incorporates three indicators to arrive at a measure of data quality. These three indicators are:

• Variance of Price Level Estimates: This measure looks at the difference between short-cut estimates, extrapolated benchmark estimates, and current (1996) benchmark consumption
price level estimates. If there is only one estimate, the variance is zero, and hence the
country is ranked 0 for no information; otherwise, a country is ranked 1 for high variances
up to 5 for low variances between the estimates. Therefore higher numbers correspond to
more reliable estimates.

- Whether data was ‘benchmarked’ in successive versions of the data.\(^8\)

- A score based on income, with the rationale being that higher income countries have better
quality data.

Countries are then given a ranking A to D based on these factors. However, here I only use the price
variance estimate, as I don’t wish to use the information regarding per capita incomes, as that would
unnecessarily bias my results in finding a statistical relationship between this sub-index and income. In
addition, their measure of benchmarking was somewhat too simplistic, in that countries were given a
score of 0-2, depending on whether countries have respectively (i) never had, (ii) occasionally had, or
(iii) often had benchmarked data. The price variance index has been transformed from a 0-5 scale to a 0-
1 scale for comparability with the other measures below.

The second measure is based on an index on Statistical Capacity, developed by the World Bank, which
is divided into three sub-sections: ‘Statistical Practice’, ‘Data Collection’, and ‘Indicator Availability’
(World Bank, 2004). Within each category there are a number of criteria, and an overall score (out of
100) is derived from this.\(^9\) Here, I am only using the first two sub-sections, which go towards data
quality, rather than the entire index, as the ‘Indicator Availability’ is geared more towards the quantity
of data. This index has to date been developed for 2004 and 2005, and so I took the average of these two
periods (again re-scaled to give values between 0-1).

The final measure is included largely because the Statistical Capacity Indicator only covers 144
developing countries, with the assumption (presumably) that scores for the developed countries would
equal 100. Although this might well be true, I have chosen not to arbitrarily allocate a score of 100 to
these countries, instead preferring to include an indicator used by the World Health Organisation (2006)
that measures the degree of coverage of vital registrations. This is designed to measure the efficacy of
core population and demography statistics. The greater the degree of coverage, the more reliable
(therefore better quality) these statistics will be.\(^{10}\)
To derive the overall measure of ‘data quality’, I took a simple arithmetic average of Price Variance (PV), Statistical Practice (SP), Data Collection (DC), and Coverage of Vital Registrations (VR), with the proviso being that countries must have scores for at least two of these four categories (again, this was mainly to capture the high income countries not included in either the Statistical Practice and Data Collection measures).

Therefore, the score for data quality can be expressed as:

$$\text{Quality}_i = \frac{\text{PV}_i + \text{SP}_i + \text{DC}_i + \text{VR}_i}{n_i}$$

Where:

- $\text{PV}_i$ = price variance for country $i$;
- $\text{SP}_i$ = Statistical Practice score for country $i$;
- $\text{DC}_i$ = Data Collection score for country $i$;
- $\text{VR}_i$ = Coverage of Vital Registrations for country $i$;
- $n_i$ = the number of categories country $i$ has a score for.

$0 < \text{Quality}_i \leq 1$

Having calculated the data quality scores they were incorporated into the overall bureaucratic quality index by multiplying the proportional score for data quantity for each country in each year by that country’s score for data quality.

As mentioned above, the resultant scores for data quality are fairly crude. The main issue for the purposes here is that this methodology results in only one score per country. In other words, countries are assumed to have the same data quality over the entire period, which is often not going to be the case. Nevertheless, it is important to at least try and account for both quantity and quality, and this is probably the most feasible way to currently do this.11

3.4 Bureaucratic Quality (‘BQ’) Descriptive Statistics

Table 1 presents some brief descriptive statistics for Bureaucratic Quality between 1960-2000, while Table 2 lists the top and bottom twenty countries, averaged between 1960-2000. As can be seen, Canada
has the highest average score over the period, while the remainder of the top twenty countries are dominated by OECD countries. The bottom of the list includes many countries from Sub-Saharan Africa, however, it also includes ‘closed’ economies such as North Korea, Bhutan and Iraq.

Table 1: Descriptive Statistics for Bureaucratic Quality, 1960-2000

<table>
<thead>
<tr>
<th>Bureaucratic Quality</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.33</td>
</tr>
<tr>
<td>Median</td>
<td>0.29</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.88</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.01</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.21</td>
</tr>
<tr>
<td>Observations</td>
<td>5,953</td>
</tr>
</tbody>
</table>

Table 2: Bureaucratic Quality, Top and Bottom Twenty Countries, Average 1960-2000

<table>
<thead>
<tr>
<th>RANK</th>
<th>COUNTRY</th>
<th>RANK</th>
<th>COUNTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Canada</td>
<td>140</td>
<td>Guinea</td>
</tr>
<tr>
<td>2</td>
<td>United States</td>
<td>141</td>
<td>Guinea-Bissau</td>
</tr>
<tr>
<td>4</td>
<td>Netherlands</td>
<td>143</td>
<td>Lao PDR</td>
</tr>
<tr>
<td>5</td>
<td>Italy</td>
<td>144</td>
<td>Cambodia</td>
</tr>
<tr>
<td>6</td>
<td>Finland</td>
<td>145</td>
<td>Sudan</td>
</tr>
<tr>
<td>7</td>
<td>Australia</td>
<td>146</td>
<td>Samoa</td>
</tr>
<tr>
<td>8</td>
<td>Spain</td>
<td>147</td>
<td>Solomon Islands</td>
</tr>
<tr>
<td>9</td>
<td>Austria</td>
<td>148</td>
<td>Angola</td>
</tr>
<tr>
<td>10</td>
<td>Japan</td>
<td>149</td>
<td>Vanuatu</td>
</tr>
<tr>
<td>11</td>
<td>United Kingdom</td>
<td>150</td>
<td>Sao Tome and Principe</td>
</tr>
<tr>
<td>12</td>
<td>France</td>
<td>151</td>
<td>Djibouti</td>
</tr>
<tr>
<td>13</td>
<td>Norway</td>
<td>152</td>
<td>Comoros</td>
</tr>
<tr>
<td>14</td>
<td>Korea, Rep.</td>
<td>153</td>
<td>Bhutan</td>
</tr>
<tr>
<td>15</td>
<td>Israel</td>
<td>154</td>
<td>Liberia</td>
</tr>
<tr>
<td>16</td>
<td>Germany</td>
<td>155</td>
<td>Afghanistan</td>
</tr>
<tr>
<td>17</td>
<td>Portugal</td>
<td>156</td>
<td>Korea, Dem. Rep.</td>
</tr>
<tr>
<td>18</td>
<td>New Zealand</td>
<td>157</td>
<td>Equatorial Guinea</td>
</tr>
<tr>
<td>19</td>
<td>Denmark</td>
<td>158</td>
<td>Iraq</td>
</tr>
</tbody>
</table>

Note: only countries with more than ten years of data represented.

Table 3 also shows that this indicator is particularly well-suited to time series analysis, as there are 128 countries that have at least 31 years of data, while only 47 countries have fewer than 30 years of data (and this is largely due to the fact that they were not independent countries for some of this forty-one year period).
### Table 3: Country Coverage of Bureaucratic Quality

<table>
<thead>
<tr>
<th>Number of Observations (Years)</th>
<th>Number of Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 10</td>
<td>16</td>
</tr>
<tr>
<td>11 to 20</td>
<td>11</td>
</tr>
<tr>
<td>21 to 30</td>
<td>20</td>
</tr>
<tr>
<td>31 to 41</td>
<td>128</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>175</strong></td>
</tr>
</tbody>
</table>

Chart 2 has a more detailed summary showing both the country coverage for each year, as well as the mean scores across countries for each year. Unsurprisingly, there is a ‘jump’ in the country coverage after 1990, as the break-up of the former Soviet Union led to an increase in the number of independent countries. In terms of the mean scores, one can see a slight, gradual increase in the mean scores over this forty-one year period, rising from 0.20 in 1960 to 0.38 by 2000.

**Chart 2: Country Coverage and Mean Scores, 1960-2000**

---

**Section 4: Comparison to existing measures of bureaucratic quality**

As mentioned in Section 2, the most commonly-used existing measure is the bureaucratic quality indicator from ICRG (hereafter known as *ICRG*), with data available from 1984. The correlation between ICRG and BQ is shown in Chart 3 below, which shows a simple scatter plot of average scores between 1984-2000. As can be seen, there is a reasonably high (though not perfect) degree of correlation between the two.\(^{12}\)
Although the ICRG measure is the indicator most researchers have used in empirical analysis, it is not the only one. Table 4 shows the pair-wise correlation between the Bureaucratic Quality indicator developed here with the ICRG indicator of bureaucratic quality, the KKZ measure of Government Effectiveness (KKZ) and the measure of Bureaucratic Delays from Business Environment Risk Intelligence (BUR_DELAYS). The correlations with this new indicator are all roughly comparable, from a low of 0.64 for the Bureaucratic Delays indicator to 0.69 with the KKZ Government Effectiveness measure.

Table 4: Correlation between BQ, ICRG, KKZ and BERI

<table>
<thead>
<tr>
<th></th>
<th>BQ</th>
<th>ICRG</th>
<th>KKZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQ</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICRG</td>
<td>0.65</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>KKZ</td>
<td>0.69</td>
<td>0.77</td>
<td>1.00</td>
</tr>
<tr>
<td>BUR_DELAYS (BERI)</td>
<td>0.64</td>
<td>0.79</td>
<td>0.86</td>
</tr>
</tbody>
</table>

a Average 1984-2000
b Average 1996-2000
c Average 1972-2000
In addition to looking at the average levels of bureaucratic quality, it is also important to get an appreciation of its volatility within individual countries over time. In the first section it was noted that one of the rationales for an institutional measure with a relatively long time series dimension is that institutions may not be as invariant over time as is often thought. Therefore, to get a sense of these changes over time, I have constructed a pair-wise correlation matrix for BQ that compares the scores of countries that had data for the full sample period (104 countries) in 1960 with their scores in 2000 (Table 5). If bureaucratic quality really is invariant (making time series analysis largely redundant), then the result should be a correlation near or at unity. The further away from unity the correlation is, the more that countries have experienced a change in scores over this 41-year period. Of course, if the correlation approaches zero, this indicates that the bureaucratic quality of a country in 1960 bears almost no relation to the bureaucratic quality in 2000, which is not plausible and would indicate a problem with the index.

Table 5: Correlations Between Bureaucratic Quality Scores Over Time

<table>
<thead>
<tr>
<th></th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQ 1960 TO 2000 (104 Countries)</td>
<td>0.84</td>
</tr>
<tr>
<td>ICRG 1984 TO 2000 (117 Countries)</td>
<td>0.66</td>
</tr>
<tr>
<td>IQ 1984 TO 2000 (117 Countries)</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Although what may constitute an ‘appropriate’ figure for these correlations is unclear, the results from Table 5 at least indicate nothing untoward. As can be seen, the simple pair-wise correlation between BQ scores in 1960 and 2000 is less than one, indicating that bureaucratic quality does change over time. However, the score (0.84) is still quite high, suggesting that for many countries the change in bureaucratic quality over time is modest. I also calculated the correlations for the ICRG measure between 1984-2000 as a comparative tool, and the BQ measure also between 1984-2000 (using the common sample for the two indicators). Interestingly, the ICRG measure suggests a much greater variation in scores over this period (0.66) compared to the BQ measure over the same period (0.95).

Overall then, the variations in scores over time appear plausible, in the sense that while the correlations are extremely high (as one would expect), they are not perfect, which indicates that over this period countries’ scores have moved both up and down, and so bureaucratic quality is not completely invariant.
4.1 Comparison of Bureaucratic Quality with the ICRG measure of Bureaucratic Quality over time

Forty countries have been selected to give a visual interpretation of the relationship over time between the ICRG measure of Bureaucratic Quality and the one developed here (see Chart 4 below). This is certainly not an exhaustive list, however, it includes countries from every major region of the world, as well as income groups. Although there is clearly more ‘noise’ in the annual scores for the new indicator’s data, these charts show that in many cases changes in scores for these two measures are quite closely related. Indeed, a close inspection of the charts representing the scores for the Democratic Republic of Congo (Zaire), Ethiopia, Hungary, South Korea, Kuwait, Myanmar (Burma), Panama, Spain, Tanzania, Uruguay and Zimbabwe show that changes in this new Bureaucratic Quality index appear some one to three years before these changes are reflected in the ICRG scores. This suggests that, for at least some countries, the new BQ measure here is picking up meaningful changes in bureaucratic quality before it becomes apparent through the ICRG index. This may be a small indication that this measure of BQ constructed here is getting closer to the ‘coalface’ of problems (and improvements) within the bureaucracy, which manifests itself relatively quickly into changes in the effort and abilities towards gathering and releasing data. However, because these lags are not uniform across all countries it would be wrong to take this claim too far (at least with the anecdotal evidence presented so far).
Chart 4: Selected Comparisons Between ICRG and Bureaucratic Quality
Aside from these comparisons to the ICRG measure of bureaucratic quality, I have also selected a small number of visual case studies that focus on specific tumultuous events that have occurred in some countries. Although this does not necessarily ‘prove’ anything, it is important to see whether, when there are clear situations in which the bureaucracy is affected, that the scores actually reflect this.

To that end, I have chosen Rwanda in 1994, Iran in 1979 and Uganda in 1979 as initial examples. As Charts 5, 6 and 7 show, the scores for bureaucratic quality were substantially affected during these periods of crisis. In the case of Rwanda, the scores show a clear decline beginning in 1990, with a more substantial drop in 1994. This is not, of course, to suggest that the decline in the quality of the bureaucracy before 1994 directly led to the genocide of that year. Rather, it suggests that there was a build-up in problems before 1994, some of which were reflected in problems within the bureaucracy.13

With respect to Iran, in February 1979 Mehdi Bazargan became the first Prime Minister of the new revolutionary regime after the Shah went into exile. The new Prime Minister, however, did not immediately have control of the country or the bureaucracy, and central authority was extremely weak. This is reflected in the continued decline in bureaucratic quality after 1979. Although Bazargan and his successor, Bani Sadr tried to gradually gain control of the government and administrative systems, one of the major problems to emerge was a struggle for control of the government between those loyal to Bani Sadr, and the clerics of the Revolutionary Council, who favoured a more theocratic style of government. Associated with this issue was a purge of the bureaucracy of people still considered to be ‘royalists’. In 1980 around 4,000 civil servants lost their jobs and in 1982 special committees within government departments were set up to examine staff members’ beliefs and political inclinations (US Library of Congress 2006). It is in this context, rather than the Iran-Iraq War that began in 1980, through which the decline (and volatility) in bureaucratic quality scores should be seen.

Finally, the example of Uganda has been chosen for the opposite reason. It was in 1979 that Tanzania invaded Uganda and forced Idi Amin to flee into exile. The bureaucratic quality score for Uganda jumped significantly between 1979 and 1980 (from an admittedly low base). In this
situation, therefore, the deleterious effects that Amin had on the civil service during his rule was lifted, and their performance show a concomitant increase.\textsuperscript{14}

Another recent case study that might be interesting to examine is the Asian Financial Crisis of 1997/98. Although there were undoubtedly many causal influences on this turbulent period, one that has been cited in the past is the relatively poor institutional framework that existed in many of these countries, particularly with respect to a lack of regulatory control over the financial sector that failed to act on the increasing number of bad loans given to political ‘cronies’ (for example, see Chang, 2000 and Sau, 2003). Chart 8 shows that, for two out of the four countries examined,\textsuperscript{15} their BQ scores started to decline \textit{before} the crisis hit. For example, the BQ scores for Malaysia peaked in 1992, and then declined over the following six years, while for Korea scores began declining in
In Indonesia scores only began to fall in 1997. Only Thailand, of the four countries used here, did not experience a decline in scores prior to 1997.

Chart 8: BQ Scores for Malaysia, Thailand, Korea and Indonesia

The anecdotal evidence presented to this point suggests that this new indicator of bureaucratic quality may prove to be quite useful. Aside from having a relatively high correlation with existing measures (in terms of average levels), Chart 4 suggests that for many countries the changes in BQ scores follow the change in scores derived from the ICRG measure of bureaucratic quality quite closely. Nevertheless, it is also important that this measure stand up to more formal econometric testing, rather than just relying on anecdotal evidence. To that end, the following section will briefly introduce this measure into empirical economic growth regressions, both cross-sectionally and, more importantly, into panel data estimations.

Section 5: Empirical analysis

Given the limited space, this analysis will only be able to give a brief overview of the relationship between this measure of bureaucratic quality and economic growth, and so more detailed analyses will have to be left to another occasion. Although the rationale behind the development of this index is to add a temporal dimension to the empirical study of institutions, it is still important to see how this
index initially performs in comparison to existing measures of bureaucratic quality in a cross-sectional analysis.

5.1 Cross-sectional analysis

The approach taken here largely follows that of Rodrik, Subramanian and Trebbi, 2004 (hereafter known as RST), Dollar and Kraay (2003) and Easterly and Levine (2003), in which a long-run approach to growth is used. In these papers, countries were assumed to have had roughly the same pre-industrial level of per capita incomes, and so therefore if one used per capita income today then this incorporates growth rates going all the way back to this pre-industrial period. They then ran a ‘horse race’ between institutional quality, trade and geography. In the RST paper, institutions were consistently significant, while trade and geography were insignificant, and often had the wrong sign. However, Dollar and Kraay (2003) pointed out that these results owed more to econometric problems, rather than the primacy of institutions (specifically, multicollinearity between the institutional and trade variables and their respective instruments with geography).

It is not the intention here to definitively solve this debate, however, it does provide a simple vehicle through which to examine this new bureaucratic quality indicator, and (potentially) allows for the endogeneity of bureaucratic quality with respect to per capita incomes to be accounted for through the use of instrumental variables. Although I am not aiming for an exact replication of their results, I use many of the same principles and variables.

In order to see how this new indicator performs compared to other measures of bureaucratic quality, both the BQ and ICRG measure of bureaucratic quality will be employed. Furthermore, to test the robustness of these results, three distinct sets of instruments will be used. The instruments used here will be:

- The proportion of the population speaking either English [ENGFRAC] or a major European language [EURFRAC]. This has been used by Hall and Jones (1999), Dollar and Kraay (2003), Rodrik, Subramanian and Trebbi (2004) and many others. In essence, this is a form of colonial instrument, in that countries with majorities speaking these languages are either the Western European countries themselves, or their colonial offshoots.
The log of settler mortality, from Acemoglu, Johnson and Robinson (2001). Higher values indicate that colonial powers (because of the higher mortality rates) set up ‘extractive’ states, rather than settling them. Consequently, higher values represent lower institutional quality \( LNSETMOR \). While a relatively new construct, this instrument has been used extensively in the literature (for example, Easterly and Levine, 2003, Rodrik, Subramanian and Trebbi, 2004 and Dollar and Kraay, 2003).

Ethno-linguistic fractionalisation \( (ELF) \), which is defined as the probability that two randomly-chosen people will not belong to the same ethno-linguistic group and a dummy colony variable \( (COLOYN) \), used by Mauro (1995);\(^{18}\)

In addition to the institutional instrument, I will also follow the approach of RST and use the Frankel-Romer measure of predicted trade shares \( (LNFR) \)\(^{19}\) as the trade instrument, while trade itself is the measure of ‘real’ trade openness \( (RT) \), as defined by Alcala and Ciccone (2004) and used in Dollar and Kraay (2003). This is defined as the sum of imports and exports in exchange rate $US relative to GDP in purchasing power parity $US. The rationale behind this is that this eliminates distortions due to cross-country differences in the price of non-tradables. That is, a country that engages in specialisation should experience an increase in openness due to higher imports, however, this specialisation may also lead to an increase in price in the (non-tradable) service sector, which would tend to lower openness (see Alcala and Ciccone, 2004 for more information). The exogenous geography variable used here is the distance from the equator \( (DISTEQ) \), measured in absolute latitude, as in RST. The BQ variable is the average between 1960-2000, as is the trade openness measure, while the ICRG variable is averaged between 1984-2000. One slight difference here to their paper is that I will use per capita incomes in 2000 (whereas they used per capita incomes in 1995). All variables are in natural logs, and the explanatory variables have been standardised for a more direct comparison. The sample size is essentially as large as the (common) data allows.

The equation to be estimated therefore is:

\[
\ln(y_i) = \mu + \alpha BQ_i + \beta RT_i + \gamma DISTEQ_i + \epsilon_i ,
\]

where \( \ln(y_i) = \) the natural log of per capita real GDP in 2000.\(^{20}\)
In the first stage, the bureaucratic quality and trade variables are regressed on all of the exogenous variables:

\[
BQ_i = \alpha + \delta BQ_{\text{INST}_i} + \phi T_{\text{INST}_i} + \psi \text{DISTEQ}_i + \varepsilon_{BQ_i} \\
RT_i = \theta + \sigma T_{\text{INST}_i} + \tau BQ_{\text{INST}_i} + \omega \text{DISTEQ}_i + \varepsilon_{RT_i}
\]  

(2)

Where:

- \(BQ_i\) = the institutional variable (either BQ or ICRG);
- \(BQ_{\text{INST}_i}\) = the bureaucratic quality instrument(s), as listed above;
- \(\text{DISTEQ}_i\) = the geography variable (distance from the equator);
- \(RT_i\) = the trade variable ('real’ trade openness);
- \(T_{\text{INST}_i}\) = the Frankel-Romer trade instrument (LNFR).

The results can be found in Table 6 below.

Columns 1-4 summarise the results using a standard OLS estimation, using the 102-country sample. Not surprisingly, both bureaucracy variables are highly significant, with t-statistics over eight. Although both trade and geography have smaller coefficients (and t-statistics), they are also significant in both regressions (Columns 3-4). Differences do start to become apparent when the instruments are employed in the 2SLS estimations, however (Columns 5-10). With ENGFRAC and EURFRAC as instruments, both bureaucracy measures remain significant, although the t-statistic for the ICRG variable is appreciably smaller than that for the BQ variable (2.457 versus 6.320 for BQ). Moreover, while trade and geography suffer the same fate in Column 6 with ICRG as that found in the RST paper, where both become insignificant, this is not the case when the new BQ variable is used (Column 5). Here, trade is significant at the 1% level, while geography is also (marginally) significant.

Using ELF and the COLONY dummy as instruments (Columns 7-8), BQ remains highly significant (as does trade), however, the ICRG variable becomes completely insignificant (with a t-statistic of only 0.659). Finally, with the log of settler mortality as the instrument in Columns 9-10, BQ remains significant (with a t-statistic of 4.585), while ICRG is again significant (albeit only at the 5% level).

Overall then, this new indicator of bureaucratic performs extremely well. Not only is it consistently significant, but its coefficient remains quite stable over the three different sets of instruments and

30
varying sample sizes (1.233 – 1.575). The same cannot be said for the ICRG variable. The problem here is similar to that found in the Dollar and Kraay paper, where it appears as though the instruments for bureaucracy and trade have a high degree of multicollinearity both with the other explanatory endogenous variable, as well as the exogenous geography variable. The new BQ indicator does not appear to have such a strong correlation with the trade variables, and so this has ‘allowed’ the trade and (to a lesser extent) geography variables to remain significant.21

However, these results should not be construed as an attempt to claim that ‘my institutional variable rules’. Although it is pleasing that it remains robust to these different instruments, the use of instrumental variables here is problematic and so, as was the case in the Dollar and Kraay paper, I now wish to approach this issue using panel data estimations.22

Table 6: OLS and 2SLS

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>OLS</th>
<th>OLS</th>
<th>OLS</th>
<th>2SLS</th>
<th>2SLS</th>
<th>2SLS</th>
<th>2SLS</th>
<th>2SLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQ</td>
<td>1.252</td>
<td>0.875</td>
<td>1.233</td>
<td>1.515</td>
<td>1.575</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.553</td>
<td>6.295</td>
<td>6.320</td>
<td>5.811</td>
<td>4.585</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICRG</td>
<td>0.746</td>
<td>0.455</td>
<td>1.594</td>
<td>6.402</td>
<td>1.852</td>
<td>6.402</td>
<td>1.575</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>0.375</td>
<td>0.311</td>
<td>0.207</td>
<td>0.507</td>
<td>-0.833</td>
<td>0.964</td>
<td>1.103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.673</td>
<td>3.786</td>
<td>2.457</td>
<td>3.069</td>
<td>-0.357</td>
<td>1.693</td>
<td>0.658</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISTEQ</td>
<td>0.334</td>
<td>0.437</td>
<td>0.195</td>
<td>0.106</td>
<td>-1.670</td>
<td>0.160</td>
<td>0.383</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.340</td>
<td>4.934</td>
<td>0.920</td>
<td>0.814</td>
<td>-0.467</td>
<td>0.923</td>
<td>0.968</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.52</td>
<td>0.42</td>
<td>0.69</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs</td>
<td>102</td>
<td>102</td>
<td>60</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Instruments:

<table>
<thead>
<tr>
<th>ENGFRAC</th>
<th>EURFRAC</th>
<th>ELF</th>
<th>COLONY</th>
<th>LNSETMOR</th>
<th>LNFR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Notes: In Columns 1-4 estimation is by OLS, and Columns 5-10 using 2SLS, both corrected for heteroscedacity. T-statistics are in italics, while ***, **, and * represent significance at the 1, 5, and 10% levels respectively.

5.2 Panel data

Although these initial results are promising, this indicator was not developed specifically with the intention of using it for cross-sectional analysis. Therefore, this section will incorporate this new BQ indicator into a panel data analysis. The first question relates to the type of estimation procedure to use. Islam (1995), for example, uses a number of different methods: a pooled estimation, based on OLS; a Minimum Distance estimator with correlated effects; and a Least Squares Dummy Variables (LSDV)
estimator with fixed effects. All are relatively common panel data techniques, however, the situation is
complicated here by the fact that there is the lagged value of per capita GDP on the right hand side.
This makes LSDV an inconsistent estimator,\(^{23}\) while the additional problem with OLS is that it has only
the single intercept term, and so cannot take into consideration the unobserved country-specific effects,
which will result in the coefficient being both biased and inconsistent (that is, in equation (3) below, it
ignores \(\eta_i\)). A common approach in dealing with the country-specific effects is to first-difference the
variables, which removes these constant individual intercepts.

\[
y_{it} = \beta y_{i,t-1} + \eta_i + \nu_i ,
\]

(3)

The major issue, however, with these growth equations in panel data form is still the problem arising
from their inability to deal with the endogeneity of the explanatory variables. In the previous section it
was shown that the cross-sectional approach was to find appropriate instruments for these variables,
and the same is true for panel data studies. One of the more popular recent methods has been through
the first-differenced General Method of Moments (GMM) estimator developed by Arellano and Bond
(1991). This GMM-DIFF estimator has been used increasingly over the past few years in empirical
growth studies (see, for example, Caselli, Esquivel and Lefort, 1996, and Easterly, Loayza and Montiel,
1997, as well as the Dollar and Kraay, 2003 paper alluded to in the previous section), however, as
noted by Blundell and Bond (1998), this estimator has been found to have poor finite sample properties
when the lagged levels of the variable are only weakly correlated with subsequent first differences,
particularly when the number of periods is relatively small. In economic growth models, where
traditionally only five or six periods have been available, this is an acute problem.

Blundell and Bond therefore proposed what they called a GMM System Estimator (GMM-SYS). The
GMM estimator uses a stacked system of equations in first differences, and levels. Even though they
found that there were considerable gains in precision and reductions in finite sample biases in Monte
Carlo simulations, the validity of the instruments can be empirically tested using a Sargan test for over-
identifying restrictions.\(^{24}\) Finally, this estimator relies on there being no second order serial correlation
in the residuals (although there is likely to be first order serial correlation in the first-differenced
residuals), and so M1 and M2 tests can be conducted for whether there is any evidence of first and
second order serial correlation respectively (see Arellano and Bond, 1991, for a detailed description of
these tests). As is customary with this estimator, the results reported will be based on the one-step
procedure, while the Sargan, M1 and M2 tests are based on the two-step procedure. Furthermore, it has been suggested in the literature that orthogonal deviations (GMM-OD) rather than first differences could also be used (see Arellano and Bover, 1995 for further details). In this method, each observation is expressed not as the first difference, but as the deviation from the average of future observations for each individual, and then each deviation is weighted to standardise the variance.

Following Islam (1995), Caselli, Esquivel and Lefort (1996) and many others, the data here is divided into non-overlapping five-yearly intervals between 1960 and 1999. The use of five-year panels, while conventional, is not mandatory. They do, however, strike a middle ground between panel data using annual data, and cross-sectional data, which effectively uses one period. Using annual data is problematic because short-term disturbances are likely to be substantial with such a short time period. This is true not only for the economic variables used here, but for the bureaucratic variable as well. There is also likely to be a degree of ‘noise’ in year-to-year changes in the bureaucratic quality index, and so taking five-year averages would appear a more sensible path to take. The variables used are largely the same as in the previous section, however, a more traditional model of economic growth will be employed here, with the dependent variable being the log change in per capita income over each five-year period ($GY$), and initial income ($Y_{t-1}$) from the first year in each period (that is, 1960, 1965 and so on). The geography variable cannot be included in a panel data analysis, however, I will also include a number of additional variables here. These include:

i. Human capital, proxied by Gross Secondary School Enrolments in the first year of each period (taken from the Global Development Network Growth dataset) [$H$];

ii. Government expenditure as a proportion of GDP (using data from the PENN World Tables 6.1) [$G$];

iii. Population growth rate (using data from the World Development Indicators) [$POPG$];

iv. Investment as a share of GDP (using data from the PENN World Tables 6.1 [$KI$].

The observations for all variables (with the exception of the initial human capital and income variables) are five-year averages and all are again expressed in natural logs. Additionally, the data has been demeaned from its period averages, which also follows Caselli, Esquivel and Lefort (1996) and Blundell and Bond (1998). In order to test the sensitivity of the bureaucratic quality results, Table 7
employs a number of different samples, while Table 8 examines a number of different panel data estimation techniques.

Table 6 below, using the GMM-SYS estimation procedure, has four alternative samples, and in each sample growth is regressed initially against BQ and initial per capita income, and then with all additional variables entered collectively.\textsuperscript{27} Columns 1-2 uses a sample of 75 countries, and is the maximum number of countries for which data across all periods is available (that is, a balanced sample). Columns 3-4 extends the sample by 25 countries, with the proviso here being that each country must have at least six periods of data (that is, thirty years). Columns 5-6, which contains 129 countries, effectively includes all countries with at least two periods of data (given that using the GMM-SYS estimation means losing one period, this means that this is the absolute largest sample available).\textsuperscript{28} Finally, Columns 7-8 remove the high-income OECD countries from the 129-country sample (leaving 106 countries).

The first thing to note here is that the coefficient on Bureaucratic Quality is remarkably consistent across the entire range of samples. When entered on its own, its coefficient ranges from a low of 0.144 in the 100-country sample (Column 3) to a high of 0.176 with the high-income OECD countries removed (Column 7). The relatively high coefficient in the non-OECD sample also suggests strongly that it is not these high-income countries that are driving these results. More importantly, while the size of the coefficient falls when the other variables are included, BQ remains significant at the 1% level in each sample, with the range of coefficient values being between 0.105-0.126. Of the other variables, trade and investment are also significant across each sample, while government expenditure is significant for the two smaller samples (Columns 2 and 4) but not for the other two. Also, in keeping with results obtained in previous studies (see, for example, Islam, 1995), human capital fails to be significant in any of the samples.
Table 7: Panel data with different samples

<table>
<thead>
<tr>
<th>Sample size (countries):</th>
<th>75</th>
<th>75</th>
<th>100</th>
<th>100</th>
<th>129</th>
<th>129</th>
<th>NON OECD (106)</th>
<th>NON OECD (106)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Y_{t-1}</td>
<td>-0.007</td>
<td>-0.131</td>
<td>-0.002</td>
<td>-0.084</td>
<td>-0.011</td>
<td>-0.101</td>
<td>-0.013</td>
<td>-0.096</td>
</tr>
<tr>
<td></td>
<td>0.025</td>
<td>0.029 ***</td>
<td>0.027</td>
<td>0.026 **</td>
<td>0.030</td>
<td>0.029 ***</td>
<td>0.031</td>
<td>0.031 ***</td>
</tr>
<tr>
<td>BQ</td>
<td>0.175 **</td>
<td>0.126 **</td>
<td>0.144</td>
<td>0.105</td>
<td>0.166</td>
<td>0.107</td>
<td>0.176</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>0.059 ***</td>
<td>0.027 ***</td>
<td>0.058 **</td>
<td>0.040 ***</td>
<td>0.048 ***</td>
<td>0.029 ***</td>
<td>0.052 ***</td>
<td>0.033 ***</td>
</tr>
<tr>
<td>H</td>
<td>0.038</td>
<td>0.025</td>
<td>0.009</td>
<td>0.003</td>
<td>0.021</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>0.025</td>
<td>0.019</td>
<td>0.009</td>
<td>0.003</td>
<td>0.021</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>RT</td>
<td>0.046</td>
<td>0.019 **</td>
<td>0.070</td>
<td>0.079</td>
<td>0.027 ***</td>
<td>0.028 ***</td>
<td>0.077</td>
<td>0.028 ***</td>
</tr>
<tr>
<td></td>
<td>0.024 **</td>
<td>0.020 ***</td>
<td>0.027 **</td>
<td>0.033</td>
<td>0.007</td>
<td>0.019 ***</td>
<td>0.007</td>
<td>0.019 ***</td>
</tr>
<tr>
<td>G</td>
<td>-0.070</td>
<td>0.024 **</td>
<td>-0.049</td>
<td>-0.033</td>
<td>0.027</td>
<td>-0.039</td>
<td>-0.039</td>
<td>-0.039</td>
</tr>
<tr>
<td></td>
<td>0.019 ***</td>
<td>0.023 **</td>
<td>0.027 **</td>
<td>0.033 **</td>
<td>0.027 **</td>
<td>0.033 **</td>
<td>0.027 **</td>
<td>0.033 **</td>
</tr>
<tr>
<td>POPG</td>
<td>-3.357</td>
<td>1.688 **</td>
<td>-0.367</td>
<td>-3.075</td>
<td>-1.660</td>
<td>2.283</td>
<td>1.660</td>
<td>2.283</td>
</tr>
<tr>
<td></td>
<td>1.736</td>
<td>2.283</td>
<td>1.660</td>
<td>2.283</td>
<td>1.660</td>
<td>2.283</td>
<td>1.660</td>
<td>2.283</td>
</tr>
<tr>
<td>KI</td>
<td>0.098</td>
<td>0.019 ***</td>
<td>0.083</td>
<td>0.085</td>
<td>0.023 ***</td>
<td>0.023 ***</td>
<td>0.081</td>
<td>0.023 ***</td>
</tr>
<tr>
<td></td>
<td>0.019 ***</td>
<td>0.021 ***</td>
<td>0.023 ***</td>
<td>0.023 ***</td>
<td>0.023 ***</td>
<td>0.023 ***</td>
<td>0.023 ***</td>
<td>0.023 ***</td>
</tr>
</tbody>
</table>

All regressions use the system GMM estimator. Coefficients and robust standard errors based on one-step procedure, Sargan p-values, m1 and m2 tests based on the two-step procedure. The instruments for Bureaucratic Quality used for the first-differenced equations are BQ_{t-2} and all further lags (with the same process for all other explanatory variables), while the levels equations use the additional instruments of ΔBQ_{t-1} for the levels equation (with the same process for all other explanatory variables).

With respect to the different estimation procedures, previous analysis on growth models in a dynamic panel data context, such as Bond, Hoeffler and Temple (2001) and Hoeffler (2001), have noted that the coefficient on initial income in the OLS model is likely to be biased upwards, while the Within Groups estimator is likely to be biased downwards (this is due to the fact that OLS does not take into consideration country-specific effects, while as Nickell (1981) explains, Within Groups is also likely to be both biased and inconsistent). They therefore used these two estimators as upper and lower bounds. If the GMM-System estimator (or, for that matter, the GMM-Difference estimator) is a better estimator to use, then the coefficient on initial income should fall somewhere between these two extremes. In both of their papers they found this to be the case, and similar results are obtained here as well. Note that in the preferred GMM-SYS specification (Column 1), the coefficient on initial income (-0.101) falls within the OLS and Within Groups coefficients for initial income (-0.061 and -0.650 respectively). Using the GMM-Difference estimator, one can see the coefficient on initial income also falls within the coefficient found using the Within Groups estimator (-0.241).
Column 3 is used to examine a different issue, in that one possible check for whether the results obtained using the GMM-System estimator are plausible is by employing orthogonal deviations rather than first differences. If the results are widely different between the two methods this may indicate that the first-differenced GMM-System estimator is an unreliable methodology to employ. However, a comparison between Columns 1 and 3 show that the results for the two are virtually identical. This gives further confidence that using the GMM-System estimator is an appropriate one in this context.

With respect to the results obtained, one can see that irrespective of the estimation methodology employed, BQ remains significant at the 1% level across each one, with the exception of Column 4 (the pooled OLS estimation), where it is significant at the 5% level. Trade and investment once again are generally significant, although trade is insignificant with the GMM-DIFF estimator, while human capital again performs quite poorly.

<table>
<thead>
<tr>
<th>Dependent variable: GYit</th>
<th>GMM-SYS</th>
<th>GMM-DIFF</th>
<th>GMM-OD</th>
<th>OLS (Panel)</th>
<th>LSDV</th>
<th>WITHIN</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_{t-1}</td>
<td>-0.101</td>
<td>-0.241</td>
<td>-0.105</td>
<td>-0.061</td>
<td>-0.224</td>
<td>-0.650</td>
<td>-0.080</td>
</tr>
<tr>
<td>BQ</td>
<td>0.107</td>
<td>0.242</td>
<td>0.108</td>
<td>0.036</td>
<td>0.182</td>
<td>0.188</td>
<td>0.055</td>
</tr>
<tr>
<td>H</td>
<td>0.003</td>
<td>-0.061</td>
<td>0.004</td>
<td>0.017</td>
<td>-0.047</td>
<td>-0.012</td>
<td>0.011</td>
</tr>
<tr>
<td>RT</td>
<td>0.079</td>
<td>0.033</td>
<td>0.020</td>
<td>0.011</td>
<td>0.019**</td>
<td>0.023</td>
<td>0.009</td>
</tr>
<tr>
<td>G</td>
<td>-0.033</td>
<td>-0.025</td>
<td>-0.038</td>
<td>-0.027</td>
<td>-0.048</td>
<td>-0.050</td>
<td>-0.035</td>
</tr>
<tr>
<td>POPG</td>
<td>-3.075</td>
<td>-1.199</td>
<td>-3.386</td>
<td>-0.534</td>
<td>0.865</td>
<td>0.974</td>
<td>-0.530</td>
</tr>
<tr>
<td>KI</td>
<td>0.085</td>
<td>0.075</td>
<td>0.080</td>
<td>0.085</td>
<td>0.089</td>
<td>0.111</td>
<td>0.091</td>
</tr>
<tr>
<td>Sargan test</td>
<td>1.000</td>
<td>0.833</td>
<td>1.000</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>m1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>m2</td>
<td>0.156</td>
<td>0.298</td>
<td>0.157</td>
<td>....</td>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>Countries</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>Obs</td>
<td>865</td>
<td>736</td>
<td>865</td>
<td>865</td>
<td>865</td>
<td>736</td>
<td>865</td>
</tr>
</tbody>
</table>

Column 1 is the system GMM estimator, while Column 2 is the first-differenced GMM estimator and Column 3 uses orthogonal deviations, rather than first differences. Coefficients and standard errors in Columns 1-3 based on one-step procedure, Sargan p-values, M1 and M2 tests based on the two-step procedure. The instruments for Bureaucratic Quality used for the GMM-DIFF estimator are BQi_{t-2} and all further lags (with the same process for all other explanatory variables), while the levels equations use the additional instruments of ΔBQi_{t-1} for the levels equation in the GMM-SYS estimator (with the same process for all other explanatory variables). OLS (Column 4) is a pooled estimator with a single intercept (not reported). Column 5 uses the Least Squares Dummy Variables estimator with country-specific intercepts. The Within Groups estimator (Column 6) employs fixed effects panel estimation, while Columns 7 uses a Maximum Likelihood Estimator. *, **, and *** represent significance at the 10, 5 and 1% levels respectively. Standard errors in italics.
Section 6: Concluding comments and issues for future research

This paper has introduced a new method of measuring bureaucratic quality in a large number of countries over a relatively long period of time. Of course, ultimately the usefulness of this index rests largely on whether one accepts the rationale behind the formation of the index itself. Using the quantity and quality of data collected by countries as a basis for measuring the quality of the bureaucracy that is charged with its collection is certainly not without its problems, and no institutional measure is perfect. The evidence presented here, however, suggests that the ability of bureaucracies to collect, collate and disseminate this economic and social data may indeed be a reasonable proxy for their overall capabilities.

In an anecdotal sense this indicator appears to have two issues in its favour. Firstly, in terms of the levels of scores it is closely correlated with the most common current measure of the bureaucracy from the ICRG. Secondly, the charts from Section 4 suggest that for many countries it also closely approximates the changes over time that are observed from the ICRG measure (at least from 1984 onwards).

Although the econometric analysis of this new indicator with respect to economic growth was by necessity quite brief, the results appear to be both plausible and sensible. In the cross-sectional results, the new indicator of bureaucratic quality remained significant across three separate sets of institutional instruments, as well as having the benefit of not being highly correlated with the trade instrument, which thereby allowed the trade variable to also appear as a significant determinant of per capita incomes in the long run. When this new indicator was employed in a panel data analysis, which is its true raison d’être, the results were also quite promising. Across a number of different samples the coefficient on BQ was quite stable and significant, and this held when run with different estimation techniques as well.

If one accepts the efficacy of this new indicator, then the potential research applications are extremely broad. Given the increasing use of panel data analyses in economic growth research, it means that researchers now have an institutional measure to use, whereas in the past this option has been extremely limited. This applies not only to contemporaneous panel data analyses, but also to causal analyses, which may allow the researcher to examine some of the important transmission mechanisms through which institutions affect economic growth (or vice versa). Furthermore, it may also be possible
to incorporate this indicator into a more general institutional framework, similar to what Knack and Keefer (1995) have done in their cross-sectional index using the ICRG data. One possible example could be combining Clague et al’s (1999) Contract Intensive Money index of property rights with this measure of bureaucratic quality. In this sense, economists may be able to add substantially more depth to their analysis of the impact that institutions have on countries over time, and not just across countries.
References


Notes

1 Business International has now been taken over by the Economist Intelligence Unit, who also now report institutional-type measures for a wide range of countries.

2 See for example Dollar and Kraay, 2003, and Rodrik, Subramanian and Trebbi, 2004, which are also discussed further in a later section.

3 Specifically, CIM requires data for the money supply (defined as the sum of currency outside banks, demand deposits, time deposits, and time and savings deposits) and currency. Both of these are widely available from the IMF’s International Financial Statistics database.

4 This issue is in some part being addressed now through initiatives such as the PARIS-21 group (founded in 1999 by the OECD, World Bank, European Commission, International Monetary Fund and the United Nations), which aims to improve the statistical capacity of governments through specific programs for developing countries. For more information on their progress, visit the web site:

http://www.paris21.org/

5 Despite (or perhaps because of) corruption in the Korean bureaucracy over this period, entry into the civil service was extremely competitive, and attracted a very high calibre of applicants. See Chen, Haggard and Kang (1998: 105).

6 These appear in totals for NBFIs though, and so this just removes double counting.

7 The reason for this is that not all of the data has been collected annually since 1960, and so the number of eligible categories rose steadily over the forty-one year period. For example, in 1960 there were 189 eligible categories, rising to 297 in 2000, covering a variety of economic and social statistics (categories available from the author on request). This ensures that the scores for each year are proportional to the data that could have potentially been collected at that time, not compared to the year 2000.

8 As noted in the Appendix to Version 6.1: “The data set consisted of 32 heading parities and expenditure shares that were put together by the World Bank for 115 countries from various regional UN ICP comparisons. The underlying data set combined the benchmark comparisons of the EU, OECD, and other European and former Soviet Union countries for 1996, a total of 52 countries; the World Bank then updated the 1993 benchmark ICP comparisons for 14 ESCAP countries, 22 African countries, 12 Caribbean countries and 8 ECWA countries to 1996 and combined these with the results for 9 South American countries for 1996. (The total of 117 double counts Japan in the OECD and
ESCAP and Egypt in Africa and ECWA). The data then combine International Comparison Programme benchmark comparisons in different regions for either 1993 or 1996, with the former being brought forward to 1996. The linking of the various regions was done in different ways, usually with a link country like Japan for ESCAP, and the United States for Africa, and South America and the Caribbean.”

9 For example, the category ‘Statistical Practice’ gives a score of one if the country produces consolidated central government accounts (zero otherwise), while the ‘Data Collection’ category gives scores based on the regularity of various census data being undertaken by the government. For more information, see World Bank (2004).

10 Data for this can be found at:
http://www3.who.int/whosis/menu.cfm?path=whosis.search.mort&language=english

11 It should be noted at this point that it is not the data quality scores that are solely driving subsequent results in the econometric section below. When the (proportional) raw scores averaged between 1960-2000 are regressed (with initial per capita income) on average per capita growth between 1960-2000, the resultant t-statistic is 3.88, and the R² is 0.13. When the data quality scores are substituted for the raw scores into the same regression, the t-statistic is 3.24, and the R² is again 0.13. This suggests, that, while the quality of data is important, it isn’t dominating the subsequent results obtained.

12 The pair-wise correlation between the ICRG measure of Bureaucratic Quality and the one developed here over the entire period (covering 2,077 observations across 134 common countries between 1984-2000) is 0.63.

13 Although not shown, this decline from 1990 is also not simply a reflection of a decline in GDP per capita, which only started to fall after 1992.

14 In other words, this definition of bureaucratic quality also seems to reflect how well the civil service is allowed by their political masters to do their job, as well as their natural ability (obtained through education and training, for example).

15 It should be pointed out that for historical and political reasons neither the IMF nor World Bank publish data for Taiwan, while Hong Kong is considered in this study to be a part of Britain initially, and then China after 1997. Data for Hong Kong, however, are available on request.
Indeed, the reason for the decline in the scores for Korea and Malaysia can be almost completely explained by the relative decline in data from the IMF’s IFS database, which contains largely financial data. Who is responsible for the data not being reported to the IMF (and why), however, is unclear.

This does, however, highlight the fact that this financial crisis was a function of many factors, as Thailand was ultimately probably the worst-hit by this crisis.

Although it is used as an instrument here, ELF has also been used as an independent variable in its own right, as some researchers believe it to be an important determinant of growth overall (see, for example, Easterly and Levine, 1997).

The Frankel-Romer measure is the most common trade instrument used in the literature, and is derived by first regressing bilateral trade flows between two countries against various geographical attributes, such as land mass, distance between trade partners, whether they share a common border, and whether they are landlocked. They then construct the predicted trade share based on the coefficients obtained. For further details, see Frankel and Romer (1999).

Specifically, the RGDPCH measure, taken from the PENN World Table 6.1.

For example, from the regression in Column 8, when the fitted values of ICRG from the first stage of the regression are run against geography and the fitted trade values, an $R^2$ value of 0.97 is obtained, suggesting almost perfect collinearity between the right hand side variables in the second stage regression. This is virtually identical to the result achieved in the Dollar and Kraay paper, albeit with different instruments. In contrast, the resultant $R^2$ using the BQ variable (Column 7) results in an $R^2$ of 0.74.

The fact that this cross-sectional instrumental variables approach becomes problematic with only two endogenous variables on the RHS means that extending this model to include other common explanatory variables, such as investment, government expenditure, demography, economic policies or human capital simply multiplies this problem, as many of these could also be considered endogenous as well.

This is when asymptotics are considered in the direction of $N \rightarrow \infty$.

See Arellano and Bond, 1991 for a discussion of this test.

The reason for this is that it has been shown that the standard errors in the two-step procedure are biased downwards, which tends to lead to unreliable inferences (that is, because of the small standard errors, one finds a significant statistical relationship ‘too often’ when using the two-step procedure).
Specifically, because I am taking logs of these variables it is $\ln(1+\text{POPG})$ to account for population declines.

These regressions with the GMM estimator, and all subsequent regressions, have been run using the DPD 1.24 software using Ox 4.02.

These samples represent 79%, 84% and 91% of the world’s population (in 2000) respectively.