

"MEASURING THE QUALITY OF GOVERNMENT AND SUBNATIONAL VARIATION"

Report for the
European Commission
Directorate-General Regional Policy
Directorate Policy Development

The following report has been prepared by the research team at
the Quality of Government Institute, Department of Political
Science, University of Gothenburg
Sweden



UNIVERSITY OF
GOTHENBURG

December 2010

Table of Contents	Page
Authors and acknowledgements	4
Table of figures and tables	5
Project Overview	8
Executive Summary (short version)	9
Executive Summary (long version)	17
Part Ia: Introduction and Review of the Literature	56
1. Defining Quality of Government	57
2. Some EU Countries Better Than Others	60
3. Consequences of QoG: How QoG Affects Social Well-being	63
Part Ib: Existing National Indicators of QoG	67
Leading measures of QoG	68
Single Source vs. Composite Source Indicators	70
Expert vs. Random/ Representative Surveys	72
Analysis and Evaluation of Current National QoG Indicators	74
QoG Concept 1: Rule of Law	74
QoG Concept 2: Corruption	79
QoG Concept 3: Quality of the Bureaucracy	83
QoG Concept 4: Democracy & Strength of Electoral Institutions	86
Conclusions	91
Part II: Evaluation of the World Bank Data	92
1. Overview of the Data and the Rankings for EU Member States	92
2. Internal Consistency of the Data	93
3. Impact of the 4 Pillars on the Overall Index of QoG	98
4. Cluster Analysis	99
5. Uncertainly and Sensitivity Analyses	101
Pillar 1: Rule of Law (RL)	104
Pillar 2: Government Effectiveness (GE)	106
Pillar 3: Control of Corruption	108
Pillar 4: Voice & Accountability (VA)	110
6. External Validity: Examining Correlates of the Index	112
Conclusions	117
Part III: Measuring QoG at the Regional Level	120
Description of the Survey Data: Individual Level	120
The Survey Questions	121
Respondent Demographics	127
Building the QoG Regional Index: Multivariate Analysis	131
1. Correlation Matrix	131
2. Factor Analysis	132

3. Normalization of the Data	134
4. Aggregation of the data & weights	134
The Aggregated Data: QoG at the Regional Level	136
Within Country Variation	137
Margins of Error	138
Internal Consistency Checks of the Data	140
Sensitivity and Uncertainty Tests	140
Combining the National and regional Level Data	144
Rule of Law	145
Corruption	146
Government effectiveness	147
Voice & Accountability	148
External Consistency Tests of the Data	149
Bivariate relationships with QoG correlates	150
Conclusions	159
Part IV: Case Studies	
Introduction	1
Methodology	1
Summary of Findings - 8 Lessons	2
Discussion	12
1. Bolzano (ITD1)	14
2. Campania (IFT3)	30
3. Västra Götaland (SE232)	46
4. Wallonie (BE3)	62
5. Estonia	80
6. Jihozapad (CZ03)	95
7. Pais Vasco (ES21)	110
8. Norte (PT11)	125
9. Nord-Vest (RO11)	143
10. Severozapaden (BG31)	161
References	i
Appendix (part Ia)	xiii
Appendix (part Ib)	xv
Appendix (part II)	xxiii
Appendix (part III)	xxxix
Appendix (part IV)	lvi

Authors and Acknowledgements

Main Contributing Researchers*:

Nicholas Charron, PhD
Victor Lapuente, PhD
Bo Rothstein, Professor
Aiysha Varraich, MA
Mauricio Hernandez, MA
Laleh Kazemi Veisari, MA
Mihai Dinescu, MA
Dimitar Popovski, MA
Jonas Håkansson, MA
Sofia Jonsson, MA
Tabatha Morgado, MA
Oana Borcan, MA

**All main contributing researchers are based out of the University of Gothenburg in Sweden*

Acknowledgments:

We would like to thank Petra Olsson, Michaela Saisana, Jan Teorell, Mikael Gilljam, Sören Holmberg, Steffan Kulmin, Carl Dhalström, Peter Esaiasson, Henrik Oscarsson, Ulf Bjreld, Veronica Norell and Marianne Ednell Persson for their advice, help and contributions on this report. Moreover, we would like to thank the European Commission Directorate-General on Regional Policy and Policy Development for the opportunity to undertake such an interesting and important study such as this, and in particular special thanks to Dr. Lewis Dykstra for his assistance and insights that have gone into this project.

Table of Figures

Figure 1	The Combined E.U. QoG Index	14
Figure 2	Building the Index – The Roadmap	30
Figure 3	The Regional QoG index	34
Figure 4	Sensitivity Tests of the Regional QoG Index	36
Figure 5	The Combined E.U. QoG Index	39
Figure 6	QoG in the E.U. and Subnational Variation	40
Figure 7	EU QoG Index, GDP and Long-term Unemployment Rates	41
Figure 8	QoG Index and Infant Mortality and Heart Disease Deaths	43
Figure 9	Evolution of QoG for three groups of European countries	61
Figure 10	EU Member States by Cluster	99
Figure 11	Rule of Law in EU Member States	105
Figure 12	Government Effectiveness in EU Member States	107
Figure 13	Voice & Accountability in EU Member States	109
Figure 14	Corruption in EU Member States	111
Figure 15	GDP per capita and QoG	113
Figure 16	Social Capital and QoG	114
Figure 17	Income Inequality and QoG	115
Figure 18	Child Mortality and QoG	116
Figure 19	Population and QoG	117
Figure 20	The Aggregated Data: QoG at the Regional Level	136
Figure 21	Within-Country Regional Variation of QoG	137
Figure 22	Regional estimations with Margins of error	140
Figure 23	Sensitivity Tests of the Regional QoG Index	143
Figure 24	4 Pillars of QoG Combined with Regional Data	145
Figure 25	Combined Rule of Law Pillar with Regional Data	146
Figure 26	Combined Corruption Pillar with Regional Data	147
Figure 27	Combined Government Effectiveness Pillar with Regional Data	148
Figure 28	Combined Voice and Accountability Pillar with Regional Data	149
Figure 29	Regional QoG Index and GDP	150
Figure 30	Regional QoG Index and Unemployment	151
Figure 31	Regional QoG Index and Long-term Unemployment	152
Figure 32	Regional QoG Index and Work-Week Hours	152
Figure 33	Regional QoG Index and Male/Female Employment rates	153
Figure 34	Regional QoG & Percentage of Workforce with a University Degree	154
Figure 35	Regional QoG Index and Population	155
Figure 36	Regional QoG Index and the use of Technology	156
Figure 37	Regional QoG Index and Tourism	157
Figure 38	Regional QoG Index and Income Inequality	158

Table of Tables

Table 1	Levels of within-Country QoG Variation	13
Table 2	QoG Rankings in the E.U. 27	22
Table 3	Survey Questions Incorporated in the Regional QoG Index	25
Table 4	Satisfaction with Current Government and Economy	26
Table 5	Respondent Demographics	27
Table 6	Factor Loadings of 16 Indicators	31
Table 7	Weighting Scheme by Indicator and Pillar	33
Table 8	Levels of within-Country QoG Variation	35
Table 9	Bivariate Regressions	40
Table 10	Examples of Leading Measures of QoG for EU Countries	69
Table 11	Empirical Coverage of Rule of Law Datasets	78
Table 12	Empirical Coverage of Corruption Datasets	82
Table 13	Empirical Coverage of Bureaucratic Quality datasets	85
Table 14	Empirical Coverage of Democracy Datasets	90
Table 15	Principle Component Analysis	94
Table 16	Pearson's Correlation Coefficient Test	96
Table 17	The Effects of Weighting Scheme and Exclusion of 1 Pillar on QoG Ranks	98
Table 18	Clusters of EU Countries Based on 4 Areas of QoG	100
Table 19	Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Rule of Law	106
Table 20	Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Government Effectiveness	108
Table 21	Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Corruption	110
Table 22	Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Voice & Accountability	112
Table 23	Respondent Experience with 3 Public Services in 18 E.U. Countries	122
Table 24	Citizen Responses to Quality of Public Services:	123
Table 25	Citizen Responses to Corruption of Public Services	124
Table 26	Citizen Responses to Impartiality of Public Services	125
Table 27	Citizen Responses to Bribery of Public Services	126
Table 28	Citizen Responses to Effectiveness of Media and Fairness of elections	127
Table 29	Satisfaction with Current Government and Economy	128
Table 30	Factor Loadings of 16 Indicators	132
Table 31	Factor Analysis of Individual Pillars	133
Table 32	Person Correlations between the QoG Regional Index & Three Pillars	134

Project Overview

There is much variation with respect to social and economic development within and across EU countries. Many possible explanations have been suggested for such variation, ranging from cumulative causation and macro-economic conditions to brain drain. One additional explanation for faltering economic and social development is that the ‘quality of government’ (QoG) is not sufficiently strong. Although a recent proliferation of QoG data have emerged since the mid 1990s, *no quantification of the quality of government has been created or used in this process so far at the regional level.* Thus the primary task of this project was to create such data. Based on the combination of national level international expert assessments from the World Bank and the largest QoG survey to date to focus on regional variation, we construct the most complete quantitative estimates of QoG variation for all 27 EU countries and 172 NUTS 1 and NUTS 2 regions within 18 countries which are highly robust to a wide array of sensitivity tests. We complement these quantitative findings with ten in-depth case studies in order to elucidate common characteristics in high and low QoG regions throughout the EU to provide the most systematic analysis of QoG in this region to date.

Short Executive Summary

1. Why Study Quality of Government in Europe?

“The most significant achievement in governance during the 1990s was the shattering of the taboo that barred discussion of corruption, particularly in diplomatic circles and intergovernmental institutions.”

United Nations *Global Programme against Corruption* (cited in Holmberg, Rothstein and Nasiritousi 2008)

Many theoretical reasons have been put forth over many years as to why democracies should exhibit better quality of government than autocracies. However, the empirical evidence in favor of the democratic hypothesis is, at best, mixed. There are numerous accounts showing how quality of government decreases – and corruption increases – after moves towards democracy (Lemarchand 1972, Scott 1972, Wade 1985, Sayari 1977, Varese 1997 Weyland 1998). Many authors have explored what Harris-White and White (1996: 3) and Sung (2004: 179) define as the “contradictory” relationship between democracy and corruption – that is, that democratization is far from going hand in hand with improving the quality of government in a country (see, for example, Keefer 2007, Bäck and Hadenius 2008 or Charron and Lapuente 2010 for recent reviews of this debate). Contrary to the predictions of many scholars, especially after the expansion of democratic freedoms in 1990, as Larry Diamond (2007, 119) states, “there is a specter haunting democracy in the world today. It is bad governance.” There are also numerous pieces of evidence indicating that the world’s most democratic region, the EU, may also be affected by this “specter” of bad governance. The goal of this report is to map the specter of bad governance in the EU: which EU member states, and regions, have bad/good governance and why?

1.1. What is “Good Governance” or “Quality of Government”?

The most frequently used definition of governance in the scholarship is the World Bank’s. Governance is *“the traditions and institutions by which authority in a country is exercised. This includes: (1) the process by which governments are selected, monitored, and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions among them.* (Kaufmann, Kraay, and Mastruzzi 2004, 3). Many economists have narrowed the concept of good governance by defining it as “good-for-economic-development” (La Porta et al. 1999, 223). Yet in these definitions a major problem persists: they are difficult to measure or border on tautology (Rothstein and Teorell 2008). This report opts for a definition that is based on a normative criterion and, at the same time, can be operable empirically. This is the case of the definition proposed by Rothstein and Teorell (2008) who, instead of “good governance”, use the term “quality of government.” They link quality of government to the concept of *impartial* government institutions – that is, when public officials who implement policies do not take anything about the citizen/case into consideration that is not *beforehand* stipulated in the policy or the law.

It is important to note that quality of government does not equal democracy (Rothstein and Teorell 2008). Democracy, which concerns the access to government

power, is a necessary but insufficient criterion of quality of government (QoG). The reason is that, if QoG were merely to equal democracy, the importance of *how power is exercised* would be left out, since the emphasis would only be on *how power is achieved*. As Paul Collier (2007, 2009) has extensively documented, without limits to elected officials, democracies can end up imploding from within or with levels of governance as bad as tyrannies. The longer a state has been a democracy, the more likely it is going to be that its citizens benefit from high levels of QoG, yet this transition, as the evolution of some post-Communist regimes nowadays indicate, is not as straightforward and automatic as many thought just a couple of decades ago. That is, democracy may be a necessary cause to enjoy QoG, but it is far from being a sufficient one. QoG requires both democracy in the access to power (i.e. elected officials in charge of policy-making) and impartiality in the exercise of that power (i.e. elected officials limited in policy implementation, so as they cannot partially favour their constituencies).

If one looks at the evolution of any available indicator of quality of government (e.g. lack of corruption, rule of law, government effectiveness), one can see important differences within the context of EU countries. On average, European countries have not shown a clear improvement in their levels of quality of government during the latest two decades. In addition, there are some notable differences across Europe. First, there seems to be a group of countries, such as Germany, Sweden or the UK, the Netherlands, Denmark or Finland, which exhibits steadily high levels of QoG irrespective of the particular index used to capture good governance. They present low corruption levels, high government effectiveness and bureaucratic quality and, at the same time, the rule of law is perceived to be very high. In the second place, there are other EU countries, such as France, Italy, Spain, Portugal or Greece, that tend to present significantly lower levels of quality of government. If any, during the latest ten to 15 years, these countries have been diverging even more from the best performers in terms of quality of government. It can thus be argued that the idea of a gap between Northern and Southern Europe in terms of quality of government seems more evident now than two decades ago. In the third place, the lack of convergence is more remarkable in the case of many Central and Eastern European countries, such as Bulgaria, Poland and Romania, where the levels of QoG have not significantly improved after joining the EU.

In sum, the existing data indicate that, as Dani Kaufmann noted, “if anybody thought that the governance and corruption challenge was a monopoly of the developing world... that notion has been disposed completely” (quoted in Rothstein 2009:2). Nevertheless, the existent research should be completed with more in-depth studies to really assess the level and relevance of the differences in QoG and corruption across the EU, such as that presented in this study.

1.2 Why is QoG Important?

To start with, numerous academic studies and statements by international organizations, such as the World Bank and the United Nations, have emphasized that only with a high quality of government can a country reap the benefits of economic growth and social development (Holmberg, Rothstein and Nasiritousi 2008). Good governance is seen as a necessary requirement for countries to foster economic development (Acemoglu, Johnson, and Robinson 2002; Acemoglu, Johnson, and Robinson 2004; Clague et al.

1999; Easterly 2001; Easterly and Levine 2003; Hall and Jones 1999; Knack and Keefer 1995; Mauro 1995; Rodrik, Subramanian, and Trebbi 2004, Rothstein and Teorell 2008). In addition, the existence of high levels of QoG – measured using different and highly correlated proxies, such as rule of law, government effectiveness or control of corruption – have also been found to reduce income inequality and poverty (Gupta, Davoodi, and Alonso-Terme 2002) and to foster a country’s aggregate levels of education and health (Mauro 1998; Gupta, Davoodi, and Alonso-Terme 1998), subjective happiness (Frey and Stutzer 2000; Helliwell 2003), citizen’s support for government (Anderson and Tverdova 2003), a lower incidence of civil war (Fearon and Laitin 2003) and, in general, the consolidation of democratic institutions (Rose and Shin 2001; Zakaria 2003). QoG has also been pointed out as a significant factor contributing to environmental sustainability (Morse 2006), reducing pollution at all income levels (Welsch 2004). In summary, all else being equal, higher (lower) levels of QoG increase (reduce) human development – as measured by life expectancy, educational attainment and standard of living – in a given society (Akçay 2006, Holmberg, Rothstein and Nasiritousi 2008).

2. Measuring QoG in the EU

According to the contemporary data, most indicators point out that QoG, or ‘good governance’ is generally higher on average within the EU-27 member states as compared with other world regions, yet there is significant variation among the countries in the EU. While this is generally accepted by many, the measurements of such concepts as bureaucratic quality or corruption are difficult to capture to say the least.

That there is debate regarding how we should best measure QoG remains in the fields of academic and practitioner research is unquestioned. Some doubt that contemporary indicators present valid measures of salient concepts within the umbrella of QoG, such as corruption, rule of law or bureaucratic effectiveness. In addition, as noted, there is strong disagreement over how QoG might be evaluated with respect to objective, or ‘hard’, indicators as opposed to subjective perceptions, as is the dominant practice in the field today. In sum, there are still many questions regarding both *what* we measure to aptly capture QoG and *how* we go about doing it. Therefore, any undertaking of a new measure will be impossible to capture perfectly and will undoubtedly be subject to a degree of opposition.

Nevertheless, many in the field of comparative politics and development economics have reached a relatively high degree of consensus on certain concepts we should focus on. QoG has become a broad concept in the social sciences and, when quantified, it has generally been disaggregated into categories/concepts such as:

- 1) ‘corruption’,
- 2) ‘rule of law’,
- 3) ‘bureaucratic effectiveness’
- 4) ‘government voice and accountability’/ or ‘strength of democratic and electoral institutions’

Further, with some exceptions, these have been aimed at capturing the performance in the *public sector*. We therefore seek a measure that captures these four essential

concepts of QoG. After careful comparison both empirically and conceptually, we find that the World Bank Governance Indicators (WGI – Kaufmann, Kraay and Mastruzzi 2009) provide the best tools with which to make reliable and meaningful comparisons within the EU at the national level of the 11-plus sources currently available that measure some aspect of QoG in European countries. We combine these four QoG pillars into one index for each EU country. Upon extensive sensitivity tests of the data, we find the WGI to be remarkably robust to multiple changes in the weighting and aggregation scheme along with the exclusion of any individual underlying data source for all four QoG pillars. Further, the data are strongly correlated with numerous socio-economic variables that have been shown in recent publications to correlate highly with QoG, such as GDP per capita, social capital, income inequality and child mortality rates, adding external validity to the WGI.

2.1 The Rank Order of E.U. Countries by QoG and Cluster Groupings

As the authors of the WGI note, finding any existing measure of QoG that can reliably distinguish between Sweden and Denmark or Romania and Bulgaria would be an admittedly impossible task. However, we can use the WGI to accurately assess relative standing to other groups of countries. What we find is that there are several patterns in the data. Though there are geopolitical and historical similarities among the groups, they are not entirely driven by such factors. For example, NMS's of the former eastern bloc, Estonia and Slovenia, are grouped with EU-15 MS's such as Spain, Belgium and France. Other EU-15 MS's, Italy and Greece, belong to the third cluster grouping. However, the group of top nine performers with respect to QoG are all EU-15 states from Northern and Central (Germany, Luxembourg and Austria) parts of Europe. Cluster analysis reveals the following three broad groups:

Group 1: Denmark, Sweden, Finland, Netherlands, Luxembourg, Germany, Austria, U.K. and Ireland

Group 2: France, Spain, Belgium, Malta, Portugal, Cyprus, Estonia and Slovenia

Group 3: Czech Republic, Lithuania, Hungary, Slovakia, Poland, Latvia, Greece, Italy, Bulgaria and Romania.

3. The Regional QoG Survey and new Regional-Level QoG Index

Although the data capturing QoG at the national level are freely available from multiple sources, such data range from scarce to non-existent at the regional level for most EU countries. Thus this study makes a significant contribution, and is the first, to provide regional level data for 18 EU countries in 172 NUTS 1 and 2 regions simultaneously. We then use these survey data to build the index from a survey of approximately 34,000 EU citizens, which constitutes the largest survey ever undertaken to measure QoG at the sub-national level to date anywhere in the world. The EU regional survey was undertaken between 15 December, 2009, and 1 February, 2010, by Efficiencie 3, a French market research, survey company.

The survey consisted of 34 QoG and demographic-based questions to the approximately 200 respondents per region. Regarding the QoG questions, the

respondents were asked about three general public services in their regions – education, health care and law enforcement – which are often administered or financed by regional authorities to maximize the amount of within-country variation. In focusing on these three services, we asked respondents to rate their public services with respect to three related concepts of QoG – the *quality*, *impartiality* and level of *corruption* of said services¹. In addition we included two further questions in the index – one about the fairness of regional elections and the other about the strength and effectiveness of the media in the region to expose corruption. The complete survey can be found in the appendix to the report.

The regional level index is based on 16 questions from the survey, following the rules and guidelines set forth by the “Handbook on Constructing Composite Indicators” (henceforth ‘HCCI’), published jointly by the OECD and European Commission. After aggregating and standardizing the data for each survey question, we obtain 16 QoG indicators for each region. With a number of decisions needed to be made along the way in constructing such an index, we also undertake a series of extensive sensitivity tests to see whether changes in our model alter the final data in meaningful ways. We find that the data constructed here are highly robust to multiple changes in weighting and aggregation schemes, the removal of individual questions or alterations in the demographic make-up of the respondents. Moreover, based on several regressions, the regional level data are highly correlated with such indicators of socio-economic development as infant mortality rates, GDP per capita and levels of technological capacity.

A quick examination of the data reveals fairly predictable patterns among the regions with respect to QoG. All regions in the top performing EU members with regard to the national QoG index (Denmark, Sweden and the Netherlands) are in the top 15% of all 172 regions. Among the NMS, all but one of the regions are in the bottom 50% (i.e. have a score lower than ‘0’), with the only exception being Nord Vest (0.21) in Romania. On the other hand, most of the EU 15 regions are in the top 50%, with Portugal and Greece being the only exceptions, having all of their respective regions under the mean average. Moreover, several regions in France and Italy are under the EU mean, with the latter having two in the bottom 10%.

3.1 Within-Country Variation

The data reveal that within-country variation varies quite significantly from country to country. After constructing margins of error (95%) around the regional estimates, we find that some countries display very tight regional groupings, while others display a remarkably wide range in QoG across their given regions, and the differences are statistically significant. Table 7 displays a simple breakdown of EU countries by level of within-country regional variation in QoG.

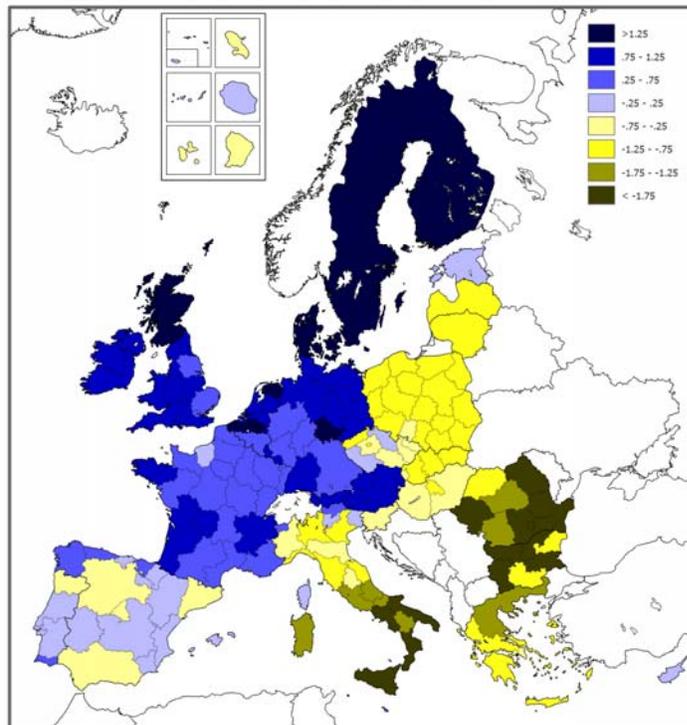
¹ These are related concepts that have come up frequently in the comparative QoG literature, and thus we try to include citizens’ opinion regarding all three.

Table 1: Levels of within-Country QoG Variation

<u>Within-Country Variation</u>		
<i>High</i>	<i>Moderate</i>	<i>Low</i>
BELGIUM	CZ. REPUBLIC	DENMARK
SPAIN	GERMANY	NETHERLANDS
ITALY	FRANCE	SWEDEN
PORTUGAL	GREECE	POLAND
ROMANIA	U.K.	SLOVAKIA
BULGARIA	HUNGARY	AUSTRIA

Several interesting findings emerge from these results that are worth noting. One is that the number of regions a country has is not an accurate predictor of within-country variation. For example, while Denmark and Slovakia have only five and four regions, respectively, in the data, Poland – which has 16 – has approximately the same amount of within-country variation, which is to say, very little. On the other hand, Belgium and Portugal, with only three and six regions, respectively, show relatively large gaps in QoG from region to region. France, with the largest number of regions at this level of analysis, has only a moderate level of within-country variation. Second, we find that whether a country is federal or highly centralized is not an accurate predictor of the level of regional variation. For example, federal countries such as Germany and Austria demonstrate only moderate to low levels of within-country variation in QoG, while federal/semi-federal countries like Italy, Belgium and Spain have high degrees of variation. Slovakia and the Netherlands, which are more centralized countries, show low within-country regional variation, while Romania and Bulgaria on the other hand demonstrate fairly wide within-country variation, even though most regions cluster in the bottom 15% of the data. This indicates that variation may have to do both with political and/or administrative choices (as shown in Italy, Belgium and Spain) or that within country QoG variation may also hinge on the distribution of quality employees – which also affects variation in centralized countries. Finally, whether a country was an original member state, EU 15 or NMS is not a good predictor of regional variation. The results show that the cases of high variation include both EU 15 (including two of the original six member states) and two countries from the NMS.

One aspect of QoG measurement that this type of within-country variation speaks to is the appropriateness of the national level to indicate levels of QoG in European countries. Our data show that, for countries such as Denmark, Netherlands, Poland or Slovakia, the national level estimates are more or less appropriate because we find no significant regional differences within those countries. However, in cases such as Belgium, Spain, Romania or particularly Italy, for example, we would argue that the national level estimates overestimate the levels of QoG in lower performing regions while underestimating the stronger regions, and thus our regional level data provide a much clearer picture of QoG levels in these countries.

Figure 1: The Combined E.U. QoG Index

In order to maximize the number of EU countries in the sample (e.g. including those countries with only one NUTS 2 region, such as Estonia or Slovenia) we elected to combine the WGI external assessment with our citizen-based, regional level data. Our aim was to come up with a method that most fairly includes the omitted EU countries from the survey while simultaneously maintaining the richness of the within-country variation in several of the countries surveyed in the regional level study. In this process, we combine the ‘external’ assessment from the WGI, based largely on expert or investment firm opinion, with the ‘internal’ assessment of the citizens from each of the 172 regions in the survey, or the ‘consumers’ of QoG in this case. For this, we take the QoG country average from the WGI, standardize it for the EU sample and set each country’s mean score to that level for each QoG pillar. We then use the new regional level data from the survey described in the previous section to capture the within-country QoG variation around the WGI mean score. A complete list of the rank order of regions and countries is found in the full report.

5. How to Improve Quality of Government: Lessons From Best Performing Regions

As a complement to the large, quantitative data we use to map QoG variation within and across countries in the EU, we undertake ten in-depth qualitative case studies to try to explain why some EU regions present higher QoG than others, by looking at which characteristics above-average regions in QoG have in common that, at the same time, make them different from below-average regions. The empirical material used in this

section is heterogeneous, from official statistics, scholarly and media reports, and the opinions of dozens of experts and professionals from ten regions:

Bolzano and Campania (Italy), Wallonie (Belgium), Severozapadan (Bulgaria), Västra Götaland (Sweden), Pais Vasco (Spain) Norte (Portugal), Nord Vest (Romania), Jihozapad (Czech Republic) and the country of Estonia

We find strong support that several broad factors in particular play a significant role in producing higher levels of QoG. One, regions that apply ‘bottom up’ pressure, in the form of a strong, independent media tend to outperform regions with weaker media institutions. Two, and especially in combination with a strong media, a region must provide sound ‘whistle-blowing’ protections for employees within the public administration. Three, regions with higher levels of QoG tend to ensure more merit-based hiring practices, as opposed to personal contacts or clientalism. Four, regions with less bureaucracy and ‘red tape’ tend to exhibit more flexibility in decision-making and higher QoG. Five, most experts agreed that giving more policy and administrative autonomy to the regions would lead to improved QoG. Finally, although the literature often cites ethno-linguistic diversity as a challenge for building QoG – creating more opportunities for collective action problems and intra-group mistrust – we find that several regions have used diversity to their advantage to create relatively higher QoG compared with other regions in their countries. This generally occurs in border regions, where cross-border sharing and learning can take place.

6. Conclusion

This report highlights the importance of QoG for economic and social development for both countries and regions within countries. Although known as one of the most advanced regions in the world, the EU contains much variation with respect to QoG. We present new and robust data that capture this variation in QoG for all 27 EU countries and 172 NUTS 1 and NUTS 2 regions. We complement the quantitative QoG data with ten in-depth, qualitative studies in order to elucidate common characteristics shared by both high and low performing regions from which we highlight several policy recommendations for improving QoG.

Although QoG is generally a topic discussed when referring to developing areas, such as Africa or Asia, we would like to emphasize the importance of focusing on it inside the EU. As this study clearly shows, there is much variation within and across countries of the EU-27. More importantly, too many EU residents still have first-hand experience of corruption and discrimination, and the share of residents confronted with these issues is far higher in some regions and countries. As a result, a consistent, joint and targeted effort to improve QoG in those countries and regions could substantially improve the economic prospects of these regions and the lives of their residents.

Long Executive Summary

1. Why Study Quality of Government in Europe?

“The most significant achievement in governance during the 1990s was the shattering of the taboo that barred discussion of corruption, particularly in diplomatic circles and intergovernmental institutions.”

United Nations *Global Programme against Corruption* (cited in Holmberg, Rothstein and Nasiritousi 2008)

Many theoretical reasons have been put forth over many years as to why democracies should exhibit better quality of government than autocracies. However, the empirical evidence in favour of the democratic hypothesis is, at best, mixed. There are numerous accounts showing how quality of government decreases – and corruption increases – after moves towards democracy (Lemarchand 1972, Scott 1972, Wade 1985, Sayari 1977, Varese 1997 Weyland 1998). Many authors have explored what Harris-White and White (1996: 3) and Sung (2004: 179) define as the “contradictory” relationship between democracy and corruption – that is, that democratization is far from going hand in hand with improving the quality of government in a country (see, for example, Keefer 2007, Bäck and Hadenius 2008 or Charron and Lapuente 2010 for recent reviews of this debate). Contrary to the predictions of many scholars, especially after the expansion of democratic freedoms in 1990, as Larry Diamond (2007, 119) states, “there is a specter haunting democracy in the world today. It is bad governance.” There are also numerous pieces of evidence indicating that the world’s most democratic region, the EU, may also be affected by this “specter” of bad governance. The goal of this report is to map the specter of bad governance in the EU: which EU member states, and regions, have bad/good governance and why?

1.1. What is “Good Governance” or “Quality of Government”?

The most frequently used definition of governance in the scholarship is the World Bank’s. Governance is *“the traditions and institutions by which authority in a country is exercised. This includes: (1) the process by which governments are selected, monitored, and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions among them.* (Kaufmann, Kraay, and Mastruzzi 2004, 3). Many economists have narrowed the concept of good governance by defining it as “good-for-economic-development” (La Porta et al. 1999, 223). Yet in these definitions major problem persists: they are difficult to measure or border on tautology (Rothstein and Teorell 2008). This report opts for a definition that is based on a normative criterion and, at the same time, can be operable empirically. This is the case of the definition proposed by Rothstein and Teorell (2008) who, instead of “good governance”, use the term “quality of government.” They link quality of government to the concept of *impartial* government institutions – that is, when public officials implementing policies do not take anything about the citizen/case into consideration that is not *beforehand* stipulated in the policy or the law.

It is important to note that quality of government does not equal democracy (Rothstein and Teorell 2008). Democracy, which concerns the access to government

power, is a necessary but insufficient criterion of quality of government (QoG). The reason is that, if QoG were merely to equal democracy, the importance of *how power is exercised* would be left out, since the emphasis would only be on *how power is achieved*. As Paul Collier (2007, 2009) has extensively documented, without limits to elected officials, democracies can end up imploding from within or with levels of governance as bad as tyrannies. The longer a state has been a democracy, the more likely it is going to be that its citizens benefit from high levels of QoG; yet this transition, as the evolution of some post-Communist regimes nowadays indicate, is not as straightforward and automatic as many thought just a couple of decades ago. That is, democracy may be a necessary cause to enjoy QoG, but it is far from being a sufficient one. QoG requires both democracy in the access to power (i.e. elected officials in charge of policy-making) and impartiality in the exercise of that power (i.e. elected officials limited in policy-implementation, so as they cannot partially favour their constituencies).

The available indicators of good governance or QoG – as well as the one presented here – are based on subjective perceptions of either experts or users of a country’s administration. The question is thus why should one pay attention to these perception-based indicators when we have “hard measures” with which to compare countries, such as literacy levels, school persistence rates, medical services, state of law and order or civil society empowerment? To start with, and as mentioned above, we would fall in a tautology if we define quality of government as a government that delivers the best policies. Secondly, as Kaufman et al. (2008:3) argue in a summary of this debate on quality of government indicators, “perceptions matter because agents base their actions on their perceptions, impression, and views.” If citizens or foreign firms perceive a given country’s administration to be plagued with corruption and public sector mismanagement, they are less likely to use its public services, with deleterious consequences for the country. An additional reason to use perception-based indicators of corruption or quality of government is that they tend to go closely “hand in hand” (Holmberg, Rothstein and Nasiritousi 2008). For example, three of the most widely used indexes of quality of government – the World Bank’s “Government Effectiveness Index”, Transparency International’s “Corruption Perceptions Index” and the “Rule of Law Index” also from the World Bank – correlate at the 0.90 level. These indicators seem thus to be capturing some very closely related phenomena. For these reasons, this report proposes that EU regional and country level differences in quality of government be measured mostly through perception-based indicators – and, in particular, through a EU-wide survey of households, that is users of public services.

If one looks at the evolution of any available indicator of quality of government (e.g. lack of corruption, rule of law, government effectiveness), one can see important differences within the context of EU countries. On average, European countries do not show a clear improvement in their levels of quality of government during the latest two decades. In addition, there are some notable differences across Europe. First, there seems to be a group of countries, such as Germany, Sweden or the UK, the Netherlands, Denmark or Finland, that exhibit steadily high levels of QoG irrespective of the particular index used to capture good governance. They present low corruption levels, high government effectiveness and bureaucratic quality and, at the same time, the rule of law is perceived to be very high. In the second place, there are other EU countries, such as France, Italy, Spain, Portugal or Greece, that tend to present significantly lower levels of

quality of government. If any, during the latest ten to 15 years, these countries have been diverging even more from the best performers in terms of quality of government. It can be argued thus that the idea of a gap between Northern and Southern Europe in terms of quality of government seems more evident now than two decades ago. In the third place, the lack of convergence is more remarkable in the case of many Central and Eastern European countries, such as Bulgaria, Poland and Romania, the levels of QoG of which have not significantly improved after joining the EU.

In summary, the existent data indicate that, as Dani Kaufmann noted, “if anybody thought that the governance and corruption challenge was a monopoly of the developing world... that notion has been disposed completely” (quoted in Rothstein 2009:2). Nevertheless, the existing research should be completed with more in-depth studies to really assess the level and relevance of the differences in QoG and corruption across the EU, such as the one presented in this report.

2.2. Quality of Government Affects Quality of Life

Do the existing cross-country perception-based indicators of quality of government reveal to us something that could matter for the people living in those countries? There is a general acceptance among scholars and policy-makers as to the crucial role of government institutions for the welfare of its citizens. On the one hand, “good governance”, “state capacity” or “quality of government” foster social and economic development (Charron and Lapuente 2009: 2) and, on the other, dysfunctional and corrupt government institutions play a central role in many of the world’s most pressing economic and social problems (Rothstein and Teorell 2008: 166).

To start with, numerous academic studies and statements by international organizations, such as the World Bank and the United Nations, have emphasized that only with a high quality of government can a country reap the benefits of economic growth and social development (Holmberg, Rothstein and Nasiritousi 2008). Good governance is seen as a necessary requirement for countries to foster economic development (Acemoglu, Johnson, and Robinson 2002; Acemoglu, Johnson, and Robinson 2004; Clague et al. 1999; Easterly 2001; Easterly and Levine 2003; Hall and Jones 1999; Knack and Keefer 1995; Mauro 1995; Rodrik, Subramanian, and Trebbi 2004, Rothstein and Teorell 2008).

In general, many economists have shifted the focus away from the traditional variables in economics for explaining economic growth, such as physical capital, natural resources and human capital, to quality-of-government-related matters (Holmberg, Rothstein and Nasiritousi 2008). Economists have also departed from the traditional explanatory variables of development in sociology or political science (e.g. Putnam 1993), such as the importance of long-term cultural traits or the concept of “social capital.”

In line with this recent trend in scholarship, an increasing number of studies point to the fact that the degree of social capital – or the level of generalized trust existing in a given society, as it is normally understood – seems to be determined by the quality of government institutions rather than the other way around, as it had traditionally been considered (Letki 2003; Rothstein 2003; Rothstein 2005; Rothstein and Stolle 2003). For instance, in one of those studies, Rothstein and Eek (2006) showed how contact with

corrupt officials – i.e. direct experience with bad governance – decreased the levels of trust amongst the participants in an experiment.

In addition, the existence of high levels of QoG – measured using different and highly correlated proxies, such as rule of law, government effectiveness or control of corruption – has also been found to reduce income inequality and poverty (Gupta, Davoodi, and Alonso-Terme 2002) and to foster a country's aggregate levels of education and health (Mauro 1998; Gupta, Davoodi, and Alonso-Terme 1998), subjective happiness (Frey and Stutzer 2000; Helliwell 2003), citizen's support for government (Anderson and Tverdova 2003), a lower incidence of civil war (Fearon and Laitin 2003) and, in general, the consolidation of democratic institutions (Rose and Shin 2001; Zakaria 2003). QoG has also been pointed out as a significant factor contributing to environmental sustainability (Morse 2006), reducing pollution at all income levels (Welsch 2004). In sum, all else being equal, higher (lower) levels of QoG increase (reduce) human development – as measured by life expectancy, educational attainment and standard of living – in a given society (Akçay 2006, Holmberg, Rothstein and Nasiritousi 2008).

Outside the scientific community, recent years have seen a surge in the public debate amongst media and policy-makers about QoG-corruption in general and in European countries in particular. For example, it has been noted that, in Romania, corruption allegations in recent years have included a prime minister, more than 1,100 doctors and teachers as well as 170 police officers and three generals according to Romanian anticorruption investigators (*The New York Times* March 9th 2009). In Bulgaria, advisors to the prime minister, Sergei Stanishev, suggested in February 2009 that the country should give the EU more power to monitor structural funds and even intervene in cases of domestic corruption where the government has failed to bring about change, following a series of cases of political corruption and organized crime that the Bulgarian judicial system had been slow to tackle (*The Economist* 3/21/2009).

The lack of QoG – especially, but far from exclusively – in Romania and Bulgaria is an increasing cause of concern for the EU as voiced by EC president Barroso: “Citizens in both countries and across the rest of Europe must feel that no one is above the law.” Similarly, one of the EU countries that has more steadily declined in the international indexes of QoG – e.g. control of corruption, bureaucratic quality – during the latest decade is Greece. This decline has been highlighted during the recent financial and economic crisis, with investor confidence plummeting and the country's budget deficit increasing to unsustainable levels. The central role of bad governance and, in particular, of corruption, for understanding Greece's difficulties is understood by the political elites. For example, at the EU summit on December 2009 Prime Minister Papandreou admitted that corruption was the main reason for Greece's serious economic problems (*The Economist* December 17th 2009).

In sum, both at the global level as well as within the EU, there are numerous indications that the “quality of government” is one of the main factors – on par or even more important than democracy – to explain the divergences in the quality of life in different countries. It is thus an issue of life and death, and thus deserves major policy attention.

Part 2. Measuring QoG in the E.U.

2.1 The Key Concepts of Quantification

According to the contemporary data, most indicators point out that QoG, or ‘good governance’ is generally higher on average within the EU-27 member states as compared with other world regions, although there is significant variation among the countries in the EU. While this is generally accepted by many, the measurements of such concepts as bureaucratic quality or corruption are difficult to capture to say the least.

That a debate regarding how we should best measure QoG remains in the fields of academic and practitioner research is unquestioned. Some doubt that contemporary indicators present valid measures of salient concepts within the umbrella of QoG, such as corruption, rule of law or bureaucratic effectiveness. In addition, as noted, there is strong disagreement over how QoG might be evaluated with respect to objective, or ‘hard’, indicators as opposed to subjective perceptions, as is the dominant practice in the field today. In sum, there are still many questions regarding both *what* we measure to aptly capture QoG and *how* we go about doing it. Therefore, any undertaking of a new measure will be impossible to capture perfectly and will undoubtedly confront a degree of opposition.

Nevertheless, in the field of comparative politics and development economics, a relatively high degree of consensus has been reached on certain concepts on which we should focus. QoG has become a broad concept in the social sciences and, when quantified, it has generally been disaggregated into categories/concepts such as:

- 1) ‘corruption’,
- 2) ‘rule of law’,
- 3) ‘bureaucratic/ government effectiveness’
- 4) ‘government voice and accountability’/ or ‘strength of democratic and electoral institutions’

Further, with some exceptions, these have been aimed at capturing the performance in the *public sector*. We therefore seek a measure that captures these four essential concepts of QoG.

In evaluating the many existing indicators of QoG at the national level for EU member states, we base our judgement of the most suitable data on the following criteria. First, we examine the quality and clarity of the descriptive information on which each of the data is constructed, or what we call the ‘internal criteria’. Second, we compare the indicators to one another, evaluating their coverage, their external validity, and how using different sources impacts their empirical relationship to such outcome variables as economic development and income inequality. Owing to time constraints, we seek to find the best and most robust existing indicator to build an index on which to rank EU countries’ QoG rather than ‘reinventing the wheel’ so to speak.

2.2 National Level Assessment for E.U. Countries

After careful comparison both empirically and conceptually, we find that the World Bank Governance Indicators (WGI – Kaufmann, Kraay and Mastruzzi 2009) provide the best tools with which to make reliable and meaningful comparisons within the EU at the national level. Of the 11 plus sources currently available that measure some aspect of QoG for European countries, we chose the WGI based on the following reasons.

1. *The ‘internal’ strength of the data* – The data were transparent and replicable, conceptually relevant and robust to sensitivity checks.
2. *The ‘external’ strength of the data* – The coverage (in terms of years and countries) along with the coverage of multiple concepts of QoG were superior to other sources.

Upon thorough review, we find that the WGI data would be the most appropriate source on which to build an indicator of QoG for EU countries. First, they cover all four pillars of QoG. Second, they cover all EU countries for at least ten years, going back to the mid 1990s, and are now published annually. Third, they are transparent in the way that they are built – publishing freely all underlying indicators on which they are built, along with a clear description of the conceptual meaning of each concept and the methodology used to create each variable. Fourth, the theoretical scope of each QoG concept is wide rather than narrow. Fifth, they form a composite indicator with data from a wide scope of sources, such as expert assessment, firm investment agencies, NGOs, IGOs and citizen surveys, although the data EU data are mostly expert and firm based. We believe that, unless specified, all aspects of corruption, rule of law etc. should be included rather than focusing on narrow aspects alone. This allows for more information to be included, which is good for reliability checks of the data for example.

2.3 QoG in the E.U. According to the WGI

Table 1 shows the rankings for each of the four ‘pillars’ of QoG along with the combined measure of QoG, which averages the four together. All data are from the most recent year available (2008).

An obvious caveat to bring forth is the fact that the World Bank authors maintain that the WGI data are a tool that scholars and practitioners can use to rate the *relative position* of countries (and not absolutely expect to rank countries with pinpoint accuracy). Finding any existing measure of QoG that can reliably distinguish between Sweden and Denmark or Romania and Bulgaria would be an admittedly impossible task. However, we can use existing measures to accurately point out relative standing to other groups of countries. What we find is that there are several patterns in the data. Although there are geopolitical and historical similarities among the groups, they are not entirely driven by such factors. For example, NMS’s of the former eastern bloc, Estonia and Slovenia, are grouped with EU-15 MS’s such as Spain, Belgium and France. Other EU-15 MS’s Italy and Greece belong to the third cluster grouping. However, the top nine performers with respect to QoG (group 1) are all EU-15 states from Northern and Central (Germany, Luxembourg and Austria) parts of Europe. Cluster analysis of the data reveals the following three broad groups:

Group 1: Denmark, Sweden, Finland, the Netherlands, Luxembourg, Germany, Austria, U.K. and Ireland

Group 2: France, Spain, Belgium, Malta, Portugal, Cyprus, Estonia and Slovenia

Group 3: Czech Republic, Lithuania, Hungary, Slovakia, Poland, Latvia, Greece, Italy, Bulgaria and Romania.

Table 2: QoG Rankings in the E.U. 27

1. Government Effectiveness	2. Control of Corruption	3. Rule of Law	4. Voice & Accountability	5. QoG Combined Rank
1 DENMARK	1 FINLAND	1 DENMARK	1 SWEDEN	1 DENMARK
2 SWEDEN	2 DENMARK	2 AUSTRIA	2 NETHERLANDS	2 SWEDEN
3 FINLAND	3 SWEDEN	3 SWEDEN	3 LUXEMBOURG	3 FINLAND
4 NETHERLANDS	4 NETHERLANDS	4 FINLAND	4 DENMARK	4 NETHERLANDS
5 AUSTRIA	5 LUXEMBOURG	5 LUXEMBOURG	5 FINLAND	5 LUXEMBOURG
6 UNITED KINGDOM	6 AUSTRIA	6 NETHERLANDS	6 IRELAND	6 AUSTRIA
7 GERMANY	7 GERMANY	7 IRELAND	7 BELGIUM	7 GERMANY
8 LUXEMBOURG	8 IRELAND	8 GERMANY	8 AUSTRIA	8 UNITED KINGDOM
9 FRANCE	9 UNITED KINGDOM	9 UNITED KINGDOM	9 GERMANY	9 IRELAND
10 IRELAND	10 FRANCE	10 MALTA	10 UNITED KINGDOM	10 FRANCE
11 BELGIUM	11 BELGIUM	11 FRANCE	11 FRANCE	11 BELGIUM
12 MALTA	12 SPAIN	12 BELGIUM	12 MALTA	12 MALTA
13 CYPRUS	13 PORTUGAL	13 SPAIN	13 PORTUGAL	13 SPAIN
14 ESTONIA	14 CYPRUS	14 CYPRUS	14 SPAIN	14 PORTUGAL
15 PORTUGAL	15 MALTA	15 PORTUGAL	15 ESTONIA	15 CYPRUS
16 CZECH REPUBLIC	16 SLOVENIA	16 ESTONIA	16 CZECH REPUBLIC	16 ESTONIA
17 SLOVENIA	17 ESTONIA	17 SLOVENIA	17 SLOVENIA	17 SLOVENIA
18 SPAIN	18 HUNGARY	18 CZECH REPUBLIC	18 HUNGARY	18 CZECH REPUBLIC
19 SLOVAKIA	19 SLOVAKIA	19 HUNGARY	19 CYPRUS	19 LITHUANIA
20 HUNGARY	20 POLAND	20 GREECE	20 ITALY	20 HUNGARY
21 LITHUANIA	21 CZECH REPUBLIC	21 LITHUANIA	21 SLOVAKIA	21 SLOVAKIA
22 LATVIA	22 LATVIA	22 LATVIA	22 GREECE	22 GREECE
23 GREECE	23 LITHUANIA	23 POLAND	23 LATVIA	23 LATVIA
24 POLAND	24 ITALY	24 SLOVAKIA	24 POLAND	24 POLAND
25 ITALY	25 GREECE	25 ITALY	25 LITHUANIA	25 ITALY
26 BULGARIA	26 ROMANIA	26 ROMANIA	26 BULGARIA	26 BULGARIA
27 ROMANIA	27 BULGARIA	27 BULGARIA	27 ROMANIA	27 ROMANIA

Note: data comes from the World Bank Governance Indicators (Kaufmann, Kraay and Mastruzzi 2009) for the year 2008.

2.4 Testing the Stability and Internal Consistency of the Rankings and Groupings

The WGI is a ‘composite index’, meaning that it takes data from other sources and combines them into one number. Naturally, many choices, such as how the data are averaged together, how they are standardized or how much weight is assigned to each individual indicator can have meaningful consequences for the outcome of each pillar,

and thus the combined QoG index itself. We tested the sensitivity of the data by first downloading all indicators of the four pillars used in this study from the World Bank homepage and then replicated their findings for 2008. We tested the validity and robustness of the rankings by employing several statistical tests of uncertainty and sensitivity. We altered the original weighting, standardization and aggregation schemes used by the authors of the WGI, along with removing each indicator individually from the underlying data for each pillar for a total of 264 simulations to test how sensitive the data were to any number of changes in the original model. Overall, we did not find that any member state was significantly biased by the combined index. The corruption data were in particular remarkably stable; for example, in none of the 60 simulations in this pillar did the mean rank change go above one place, and every country stayed within only four places of their original rank throughout the simulations. Finally, while there was some degree of movement in the rankings across the simulations *within* the three groups themselves, the groups established by the cluster analysis were remarkably stable in that there was very little movement *between* the groupings.

In addition to tests of the sensitivity of the data, we ran tests of internal consistency. We find that all 54 individual underlying data sources among the four pillars (save two), are positively correlated with the overall QoG index. Factor analysis of each pillar revealed strongly unified underlying data, and that the sources that make up each pillar statistically represent a single concept.

2.5 What is QoG in the EU Correlated With?

We perform several bivariate regression analysis with other national level data to test whether there is a statistically significant relationship with QoG, along with testing whether the QoG index from the WGI is externally valid. The overall index relates well externally to several variables that have been shown in recent publications to correlate highly with QoG. GDP per capita, social capital, income inequality and child mortality rates all significantly correlate with the World Bank QoG index in their directions. As regards to the relationship between population size and QoG, at least among the EU sample, there appears to be no advantage in neither large nor small states.

3.1 The Regional QoG Survey

Although the data that capture QoG at the national level are freely available from multiple sources, such data range from scarce to non-existent at the regional level for most EU countries. Thus this study makes a significant contribution in its attempt to be the first to provide regional level data for 18 EU countries in 172 NUTS 1 and 2 regions simultaneously. We use a survey of approximately 34,000 EU citizens to build a QoG index for each region, which constitutes the largest survey ever undertaken to date to measure QoG at the sub-national level anywhere in the world.

3.2 Description of the Method

The EU regional survey was undertaken between 15 December, 2009, and 1 February, 2010, by Efficienc 3, a French market research survey company². A full description of the survey method and other details about how the survey was conducted is given in the appendix along with the full survey itself. The respondents, who were 18 years of age or older, were contacted randomly by telephone in the local language via the ‘birthday method’ by a trained telephone survey expert. The next-birthday method is an alternative to the quotas method. The ‘next-birthday’ method, which simply requires the interviewer to ask the person who answers the phone who in their household will have the next birthday, still obtains a reasonably representative sample of the population according to survey experts. The interviewer must take the person who has the next coming birthday in the household (if this person is not available, the interviewer makes an appointment), thus not relying on whomever might be available to respond in the household. So, where the quota method is stronger in terms of a more even demographic spread in the sample, the next-birthday method is stronger at ensuring a better range of opinion. After consulting with E3, we decided upon the next-birthday method because we feel that what we might lose in demographic representation in the sample, we will make up for with a better distribution of opinion.

3.3 The Survey Questions

We posed 34 QoG and demographic based questions to the approximately 200 respondents per region. Regarding the QoG questions, the respondents were asked about three general public services in their regions – education, health care and law enforcement. While not perfect – as noted, regions in different countries having different administrative and fiscal responsibilities - we find that, in general, these are the public services that are most likely to be either administered or financed by regional governments (or both), thus maximizing regional variation where applicable. Furthermore, even in cases where these areas are highly centralized, there can of course be regional variation in the quality in terms of how local public servants execute and provide these services, based on the level of quality of employees that each region attracts. We thus attempt to capture all forms of regional variation with these questions. However, publicly administered areas such as immigration, customs or national security were intentionally avoided because these are dealt with at the national or even

² <http://www.efficence3.com/en/accueil/index.html>

supranational level. In focusing on these three services, we asked respondents to rate their public services with respect to three related concepts of QoG – the *quality*, the *impartiality* and the level of *corruption* of said services³. In addition we included two further questions in the index – one about the fairness of regional elections and the other about the strength and effectiveness of the media in the region in exposing corruption. The complete survey is given in the appendix. The 16 QoG questions from the survey used in building the QoG regional index are shown in Table 2.

Table 3: Survey Questions Incorporated in the Regional QoG Index

Rule of Law-Focused Questions

"How would you rate the quality of the police force in your area?" (low/high, 0-10)

"The police force gives special advantages to certain people in my area." (agree/disagree, 0-10)

"All citizens are treated equally by the police force in my area" (Agree, rather agree, rather disagree or disagree, 1-4)

"Corruption is prevalent in the police force in my area" (agree/disagree, 0-10)

Government Effectiveness-focused questions

"How would you rate the quality of public education in your area?" (low/high 0-10)

"How would you rate the quality of the public health care system in your area?" (low/high 0-10)

"Certain people are given special advantages in the public education system in my area (agree/disagree, 0-10)

"Certain people are given special advantages in the public health care system in my area." (agree/disagree, 0-10)

"All citizens are treated equally in the public education system in my area" .” (Agree, rather agree, rather disagree or disagree, 1-4)

"All citizens are treated equally in the public health care system in my area" .” (Agree, rather agree, rather disagree or disagree, 1-4)

Voice & Accountability-focused questions

"In your opinion, if corruption by a public employee or politician were to occur in your area,

how likely is it that such corruption would be exposed by the local mass media?" (unlikely/likely, 0-10)

"Please respond to the following: Elections in my area are honest and clean from corruption" (agree/disagree, 0-10)

Corruption-Focused Questions

"Corruption is prevalent in my area's local public school system (agree/disagree, 0-10)

"Corruption is prevalent in the public health care system in my area" (agree/disagree, 0-10)

"In the past 12 months have you or anyone living in your household paid a bribe in any form to:

Health or medical services?" (yes/no)

"In your opinion, how often do you think other citizens in your area use bribery to obtain public services?" (never/very often, 0-10)

In addition to demographic and QoG questions, we questioned each citizen's satisfaction with the current state of the economy and national government (0-10, higher numbers being more positive). The numbers here can be interpreted as any figure above a '5' equalling some level of positive satisfaction with the economy or government while any number under '5' is a negative sentiment overall. We find respondents in countries such as Denmark, the Netherlands, Sweden and Austria to be the most satisfied among all 18 countries surveyed here. Conversely, citizens from Hungary, Spain, Greece and Romania on average were the least satisfied with the current state of the economy and/or performance of their government.

³ These are related concepts that come up frequently in the comparative QoG literature. Thus we try to include citizens' opinion regarding all three.

Table 4: Respondent's Satisfaction with Current Government & Economy

<i>Unit</i>	<i>Obs.</i>	<i>Satisfaction of e Economy (0-10)</i>	<i>Satisfaction of Government (0-10)</i>
E.U. average	33540	3.24	3.69
France	5070	3.13	3.81
Germany	3120	3.75	3.80
Italy	4095	3.52	4.02
Spain	3315	2.01	2.43
UK	2340	3.33	3.79
Hungary	585	2.12	2.42
Cz. Rep.	1560	2.28	3.06
Slovakia	780	3.45	4.08
Portugal	1365	2.53	3.58
Romania	1560	2.16	2.76
Sweden	585	4.66	5.06
Denmark	975	5.17	5.23
Belgium	585	4.12	4.24
Austria	1755	4.92	4.58
NL	780	5.11	5.24
Poland	3120	3.46	3.12
Bulgaria	1170	2.61	4.53
Greece	780	1.91	3.92

note: Standard deviations for E.U. averaged economic and government satisfaction are 2.63 and 2.76 respectively

All country standard deviations range from 2.09 (NL) to 3.04 (Cz. Rep.).

*Both variables range from 0-10, with higher scores equating to higher satisfaction. **High & Low figures in Bold***

We report demographic information on the citizen respondents in Table 5. We find that a majority of citizens polled have had direct contact with public education (60%) and health care services (78%) in the past year, while one quarter have had direct contact with law enforcement in some capacity. Other distributions, such as income, education, age, population of the region, religion and gender, are listed, along with whether the respondent has their region's official language as the mother tongue. In terms of gender, 55% of the respondents were female and 45% were male and in only one country (Austria) did we receive a majority of male respondents.

Table 5: Respondent Demographics

Experience in last 12 months	
Education	0.60
Law enforcement	0.25
Health Care	0.78
All 3	0.14
None of the 3	0.09
Income	
Low	0.38
Medium	0.37
High	0.14
Education	
Secondary or less	0.49
University	0.35
Advanced degree	0.14
Population	
10,000 or less	0.39
10,000 - 100,000	0.36
100,000 - 1 million	0.18
1 million +	0.06
Age	
18-29	0.19
30-49	0.35
50-64	0.27
65+	0.19
Religion	
Christian	0.56
Secular	0.17
Muslim	0.02
Other	0.26
Gender	
Female	0.55
Male	0.45
Language: mother tongue same as official regional language?	
yes	0.95
no	0.05

Upon constructing the regional level index based on the 16 questions in Table 3 (construction to be discussed in the following section) we run several multivariate analyses to elucidate which demographic and satisfaction questions are systematically related with higher or lower perceptions of QoG. In summary, several factors stand out in the regressions. First, when multivariate regression is done with a very large number of observations (N), significant relationships between the independent and dependent variables are expected. It is almost more informative when a relationship is *non-significant* in this case, thus showing almost definitively that there is no relationship between the explanatory variable in question and the dependent variable. With this in mind, it is certainly noteworthy that, among a sample of 33,540 across 18 European

countries, the ‘minority’-type characteristics, such as the indicator for Muslim religion, whether the respondent has a different first language than that of the official regional language or whether the respondent is a woman, have very little systemic impact on perceptions of QoG. Regarding the latter, women tend to think that the quality of services is lower on average, yet that the public sector in their region is more impartial compared with male respondents. There is no significant difference in gender when looking at attitudes about corruption or the combined 16 question index.

The demographic characteristics that stand out as most systemically impacting perceptions of QoG based on Table 3 are *population* (in particular respondents in large cities), *income* and *age*. There are several factors that lead to very straightforward interpretations. By far, the most negative respondents throughout the survey were citizens from large cities; thus we will expect regions with large urban areas, or those regions that are in fact cities, such as Barcelona, Inner London, or Prague, to have lower QoG scores on average than other regions in their country. On age and income, the most positive respondents were those over 65 along with middle and especially high income earners (relative to respondents under 30 and low income earners, respectively). Education is slightly more difficult, with attitudes toward quality of services being a negative function of education, while those with higher levels of education tend to think that their regional public sectors are more impartial and less corrupt than respondents with a secondary education or less.

Moving to respondents who have experienced the three public services in question, we find several interesting results. One, there is an apparent gap between the perception and ‘reality’ of two services in particular – education and law enforcement service. The perception of the former (captured by those without direct experience in the past 12 months) is that aspects of QoG are generally higher than ‘reality (measured by those with direct experience in the past year). What we find is that respondents who have had recent direct contact with their education system generally rank it lower with respect to quality, impartiality and corruption questions as well as the combined index, and the relationship is significant at the 95% level of confidence or higher in each model. With respect to law enforcement, the perception works in the opposite direction. Those without direct contact in the past 12 months generally perceive this particular service as worse with regard to QoG than those who have had contact with law enforcement recently. The relationship is at the 99% level of confidence in each model. With respect to the health care services, we find that those with experience believe that this public service provides slightly less overall quality than those who have not used their health care services recently; however, those with direct contact perceive health care services as being slightly less partial than those with no experience. On corruption and overall QoG, there is no significant difference between respondents with and without direct contact in the past 12 months.

Finally, as anticipated, satisfaction with the current state of the economy and government are positively related to attitudes about QoG. In other words, citizens that are optimistic about or satisfied with the present state of the economy and politics in their region believe that their regional public sectors are more impartial, less corrupt and have better quality services than those respondents who are less satisfied.

4.1 Sub-National Data: The Regional-Level Index

This section elucidates the methods, techniques and results of building the regional level QoG index from the survey questions discussed in the previous section. In doing so, we follow the rules and guidelines set forth by the “Handbook on Constructing Composite Indicators” (henceforth ‘HCCI’), published jointly by the OECD and European Commission. It is thus the goal in this section to transform the 33,540 observations into 172 (one number per region). With a number of decisions needed to be made along the way in constructing such an index, we also undertake a series of sensitivity tests to see whether changes in our model alter the final data in meaningful ways. We find that the data constructed here are highly robust to multiple changes in weighting and aggregation schemes, the removal of individual questions or alterations in the demographic make-up of the respondents. We report final data in visual form (maps), although the specific rankings and numbers can be found by region in the appendix.

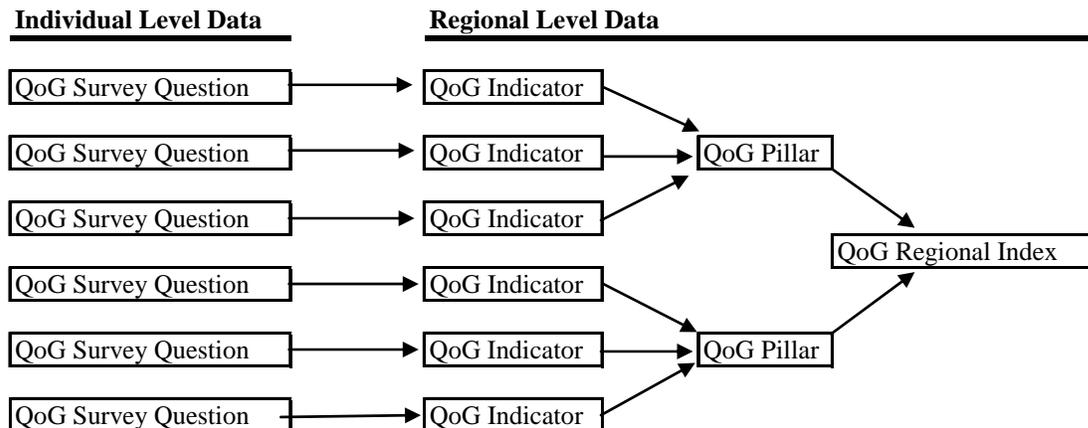
4.2 Building the QoG Regional Index: Multivariate Analysis

The HCCI recommends in combining multiple data into a single index that all underlying data be internally consistent with one another. In other words, if the questions in the survey are aimed at capturing certain aspects of the broad concept of ‘QoG’, then we would ideally like them to be correlated to one another. However, we do want some degree of ‘lack of correlation’ in the underlying data so that we can say with some degree of certainty that each individual indicator (e.g. survey question) is capturing some unique aspect of QoG. In this section we discuss briefly the results of several tests employed in this analysis:

- Correlation Matrix (Pairwise Person Correlations) & Chronbach’s Alpha
- Factor Analysis (Principle Component)
- Normalization of the Data
- Aggregation of the Data & Weights

Figure 2 shows the ‘roadmap’ we follow in building the index from the survey questions discussed in the previous section using just six questions as an example. Each arrow represents a step where we must ‘aggregate’ the data in some way to combine multiple sources. The process thus begins with each region having 16 questions with approximately 200 responses per question and ends with each region having one number representing its estimated level of QoG.

Figure 2: Building the Index – The Roadmap



4.2.2 Correlation Matrix and Chronbach’s Alpha

Upon the first aggregation step, we first test the underlying relationship between the 16 QoG indicators to one another. We have corrected all data so that higher numbers equal higher estimate levels of QoG. In essence, since they are meant to be components of the same overall concept, we would hope that they all correlate with one another in the same basic direction. With only two exceptions (the insignificant and negative relationship between *Quality of Education* and *Impartiality of Law enforcement* as well as between *General Bribery* and *Impartiality of Education*) we find that all 16 variables have a positive relationship to one another. Moreover, that the average pair-wise covariance is 0.48 among the 16 indicators and 89% of the correlations are significant at the 95% level of confidence.

One further test of the internal consistency of the underlying data is Chronbach’s Alpha coefficient (Chronbach 1951), which in basic terms measures the extent to which the variables tested are capturing a common concept. We find the Alpha coefficient of reliability is 0.94. The ‘rule of thumb’ cut-off reliability threshold of the coefficient varies from discipline to discipline (from a lenient 0.6 up to a more stringent 0.8), thus the .94 coefficient exceeds all accepted reliability thresholds. Therefore, on taking a closer look at the basic underlying structure of the data, we feel it is entirely appropriate to proceed further using these 16 QoG indicators.

4.2.3 Factor Analysis – Principle Component

Factor analysis (FA) is intended to “describe a set of Q variables x_1, x_2, \dots, x_Q in terms of a smaller number of m factors and to highlight the relationship between these variables” (OECD 2008: 69). In essence, we are assuming that there are underlying groups (factors) in the data in which we can combine individual indicators so as to make several larger factor groupings. We therefore seek to discover which of the 16 indicators have common enough variance according to FA to be considered appropriate to be placed in a single grouping, as shown in Figure 1 where multiple ‘QoG Indicators’ become ‘QoG Pillars’. We use the common method of Principle Component Analysis (PCA) to derive the

number of appropriate principle components (e.g. number of factors in this case). We follow the standard rules of thumb as a ‘cut-off’ measure in order to be as pragmatic as possible without losing too much information in that data. The Kaiser criterion suggests that a factor must have an Eigenvalue of at least ‘1’, while many experts suggest that such a factor also explains at least 10% of the total variance. We follow both rules of thumb here. Of the potential 15 principle components discovered in the underlying data, three independent factors are revealed by the FA, all meeting the Kaiser criterion and explaining a suitable amount of variation. Together, the three factors explain over three-fourths of the variation in the data. We employ Varimax rotation in order to see which variables load onto which factor. Table 6 displays the rotated factor loadings for the individual 16 QoG indicators.

Table 6: Factor Loadings of 16 Indicators (Principle Component, Varimax Rotation)

EdQual			0.69
HealthQual			0.51
LawQual			0.89
Media			0.67
Election			0.48
EdImpartial1	0.93		
HealthImpartial1	0.80		
LawImpartial1	0.91		
EdImpartial2	0.79		
HealthImpartial2	0.52		
LawImpartial2	0.91		
EdCorrupt		0.67	
HealthCorrupt		0.80	
LawCorrupt		0.38	
OthersCorrupt		0.76	
GeneralBribe		0.53	
Expl. Var	0.33	0.25	0.17
Expl. Var/ Total	0.4177215	0.3164557	0.2151899

The block of six impartiality variables load cleanly onto one principle component (Impartiality), which explains roughly 42% of the total variance in the model. The next most consequential block of variables is the set of corruption questions (Corruption), which together explain approximately 31.5% of the total variance in the model, while the third group (Quality) has a combined total variance of about 21.5%. We label these three factors the three ‘pillars’ of the regional QoG index – e.g. ‘corruption’, ‘impartiality’ and ‘quality’. We check the internal consistency of the three pillars individually to see whether running the five to six indicators by pillar results in one clear factor. The FA from all three individual pillars indicates that they are internally consistent (e.g. only one significant factor is found).

4.2.4 Normalization of the Data

As several of the questions have different scales in the original data, we must of course find a common scale for all indicators to be able to aggregate them into a single indicator. Two methods appear to be the most appropriate based on the data that we have – standardization and ‘min-max’. For the former, for each individual indicator, we calculate the average across all EU regions and the standard deviation. We believe that the method of standardization has several advantages. One, it renders direct comparison quite easy between regions, as we can say that region ‘X’ is a full standard deviation higher in QoG than region ‘Y’ by simply glancing at the data. Two, the range in variance is allowed to be a bit more extensive in standardization than in min-max. However, the choice is still somewhat arbitrary, and thus we test the sensitivity of our choice of normalization in the section on uncertainty and sensitivity testing.

4.2.5 Aggregation of the Data and Weights

There are essentially two stages of aggregation with this particular data. In the first step, we employ a very simple and replicable method to transform each survey question into the 16 ‘QoG indicators’ as shown in Figure 1. We take simple mean scores from each region’s question. We then standardize each indicator, (so that the EU regional mean equals ‘0’ with a standard deviation of ‘1’). An alternative method would be to compensate regions with disproportionately high or low demographic responses compared with the ‘actual’ demographic distribution by giving each region individual weights on certain demographic variables. We elect not to do this for two reasons. One, we do not have perfect information on all demographic data for each region, and thus would be forced to leave out certain variables, such as income or religion. Second, the time constraints prevent us from undertaking this alternative and, furthermore, any future attempts to measure QoG at the regional level will be more easily comparable with simple means aggregation than with demographic weights. However, the individual survey data will be made public and thus any researcher can alter or weight the data at this stage as he or she sees fit.

In the second step, at the regional level, we normalize the data by standardization and then aggregate them into the three pillars and then into a single composite index. The three pillars are each weighted $1/3^{\text{rd}}$ and combined into a single QoG index for each region. We find after running a correlation test that the relationships between each of the three pillars and the combined QoG regional index are significant at the 99% level of confidence.

An alternative weighting scheme would be based on the factor weights from the PCA. Table 7 shows the weighting scheme by pillar and indicator.

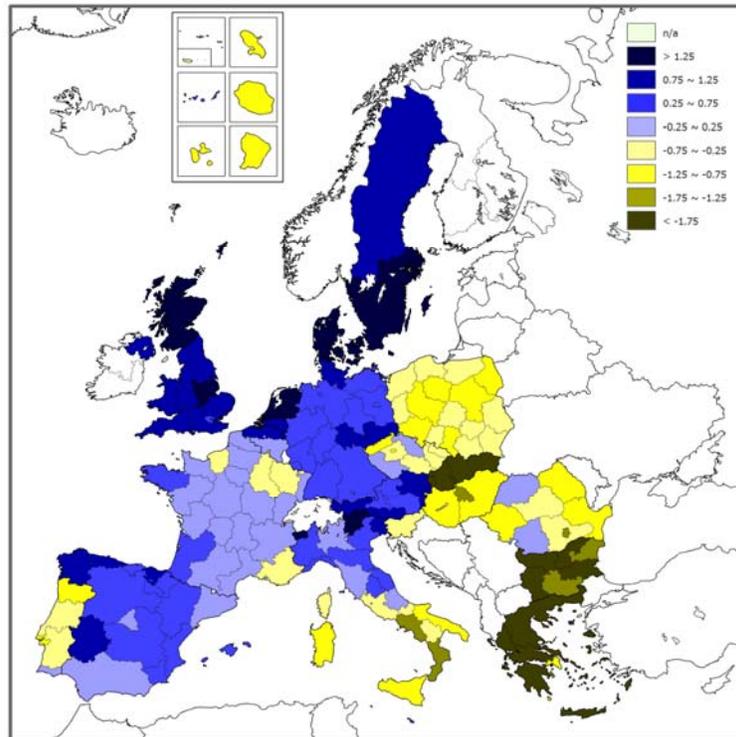
Table 7: Weighting Scheme by Indicator and Pillar

Indicator	Pillar	Pillar weight	Factor weight		
			Equal Weight W/in Pillar weight	w/ Pillar of Total	
EdQual	Quality	33.33%	20%	22.93%	4.89%
HealthQual			20%	12.52%	2.67%
LawQual			20%	38.15%	8.14%
Media			20%	21.62%	4.61%
Election			20%	11.09%	1.64%
EdImpartial1	Impartiality	33.33%	16.67%	21.33%	9.51%
HealthImpartial1			16.67%	15.78%	7.04%
LawImpartial1			16.67%	20.42%	9.10%
EdImpartial2			16.67%	15.38%	6.86%
HealthImpartial2			16.67%	6.67%	2.97%
LawImpartial2			16.67%	20.41%	9.10%
EdCorrupt	Corruption	33.33%	20%	21.46%	8.21%
HealthCorrupt			20%	30.59%	11.70%
LawCorrupt			20%	6.91%	2.64%
OthersCorrupt			20%	27.62%	10.56%
Health Sector Bribe			20%	13.43%	5.14%

In the aggregation method we thus keep with our desire to build the regional QoG indicator in a parsimonious way. We aggregate using the common ‘linear aggregation method’ when we combine the indicators into a single number. A more detailed explanation of this process is given in the appendix.

4.3 The Aggregated Data: QoG at the Regional Level

After aggregating and standardizing the data for each survey question, we obtain 16 QoG indicators (which then combine into the three pillars) for each region. Table 10 (in the full report) shows a complete list of the 172 regions in the study according to their final QoG index rankings. Country names represent the average scores of all regions for that country included in the data. The data range from Midtjylland in Denmark (DK04) which is ranked the top region in the EU with respect to overall QoG (score of 2.13) to Severozepaden in Bulgaria (BG31), which is ranked lowest, with a QoG score of -2.96, or just over three standard deviations below the mean EU score. Figure 3 shows a visual distribution of the regional-level QoG index, where darker scores represent more extreme values on either end and red (blue) indicates that a region is above (below) the EU mean.

Figure 3: The Regional QoG index

A quick examination of the data reveals fairly predictable patterns among the regions with respect to QoG. All regions within the top performing EU members with regard to the national QoG index (Denmark, Sweden and the Netherlands) are in the top 15% of all 172 regions. Among the NMS's, all but one of the regions are in the bottom 50% (i.e. have a score lower than '0'), with the only exception being Nord Vest (0.21) in Romania. Other relatively strong NMS regions are Sud Vest and Centru in Romania along with Severovýchod and Jihozápad in the Czech Republic (all just under the mean average of the EU). On the other hand, most of the EU 15 regions are in the top 50%, with Portugal and Greece being the only exceptions with all of their respective regions under the mean average. Moreover, several of the regions in France and Italy are under the EU mean, with the latter having two in the bottom 10%.

4.3.2 Within-Country Variation

The data reveal that within-country variation varies quite significantly from country to country. After constructing margins of error (95%) around the regional estimates, we find that some countries display very tight regional groupings while others display a remarkably wide range in QoG across their given regions which are statistically significant. Table 7 displays a simple breakdown of EU countries by level of within-country regional variation of QoG.

Table 8: Levels of within-Country QoG Variation

<u>Within-Country Variation</u>		
<i>High</i>	<i>Moderate</i>	<i>Low</i>
BELGIUM	CZ. REPUBLIC	DENMARK
SPAIN	GERMANY	NETHERLANDS
ITALY	FRANCE	SWEDEN
PORTUGAL	GREECE	POLAND
ROMANIA	U.K.	SLOVAKIA
BULGARIA	HUNGARY	AUSTRIA

What the data reveal is that, for countries such as Denmark, Slovakia or Poland, the national estimates essentially capture the level of QoG in the country, as there is no significant within-country variation from region to region. However, in countries such as Belgium, Spain, Portugal or Italy, the variation is quite significant, and thus the national level data tend to overestimate poorer performing regions while underestimating better performing regions. Others, like Germany, the U.K. or France, show moderate levels of within-country variance.

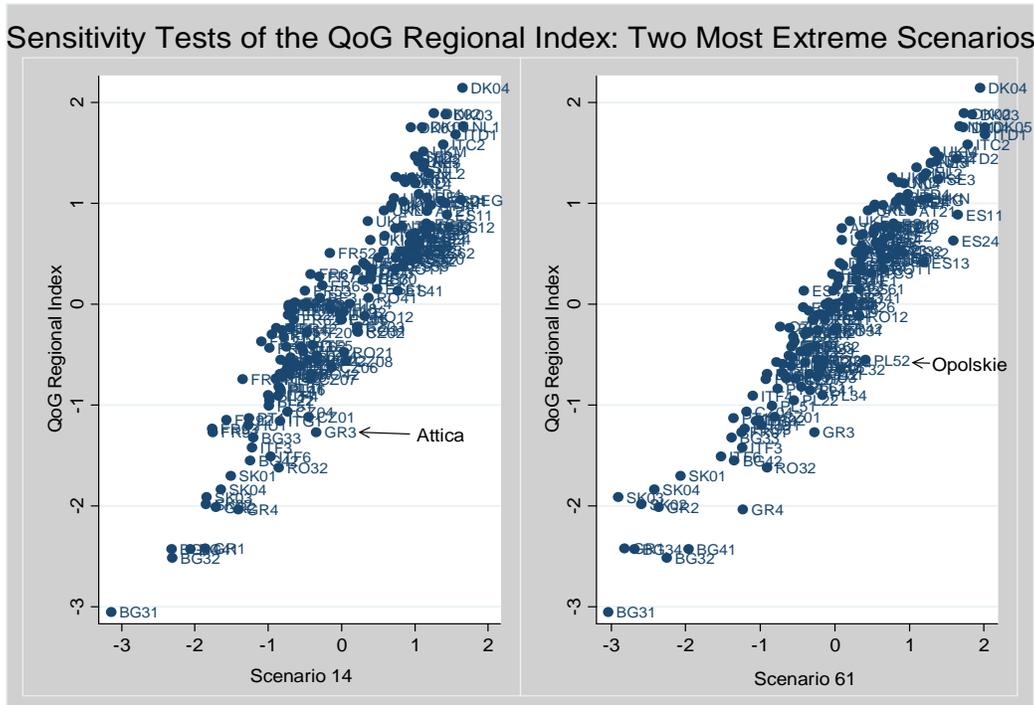
Several interesting findings emerge from these results that are worth noting. One is that the number of regions a country has is not an accurate predictor of within-country variation. For example, while Denmark and Slovakia have only five and four regions, respectively, in the data, Poland - which has 16 - has approximately the same amount of within-country variation, which is to say, very little. On the other hand, Belgium and Portugal, with only three and six regions, respectively, show relatively large gaps in QoG from region to region. France, with the largest number of regions at this level of analysis, has only a moderate level of within-country variation. Second, we find that whether a country is federal or highly centralized is not an accurate predictor of the level of regional variation. For example, federal countries such as Germany and Austria demonstrate only moderate to low levels of within-country variation of QoG, while federal/semi-federal countries such as Italy, Belgium and Spain have high degrees of variation. Slovakia and the Netherlands, which are more centralized countries, show low within-country regional variation, while Romania and Bulgaria on the other hand demonstrate fairly wide within-country variation in terms of absolute score, even though the regions in these two countries are all found in the bottom 15% of the regions/countries in the data with respect to QoG. However, on the basis of the margins of error, we find the variation in these two countries to be significant. This indicates that variation may have to do both with political and/or with administrative choices (as shown in Italy, Belgium and Spain) or that within-country QoG variation may also hinge on the distribution of quality employees - which also affects variation in centralized countries. Finally, whether a country was an original member state, EU 15 or NMS is not a good predictor of regional variation. The results show that the cases of high variation include both EU 15 (including two of the original six member states) and two countries among the NMS's.

One aspect of QoG measurement that this type of within-country variation speaks to is the appropriateness of the national level to indicate levels of QoG in European countries. Our data show that, for countries such as Denmark, the Netherlands, Poland or Slovakia, the national level estimates are more or less appropriate because we find no

significant regional differences within these countries. However, in cases such as Belgium, Spain, Romania or particularly Italy, we would argue that the national level estimates overestimate the levels of QoG in lower performing regions while underestimating the stronger regions; thus our regional level data provide a much clearer picture of QoG levels in these countries.

4.4 Sensitivity Tests and Internal Consistency Checks of the Data

Figure 3



We performed 62 simulations in which the sensitivity of the data was tested. For example, factor weights and geometric aggregation were employed in addition to equal weighting and additive aggregation, while we removed each individual question as well as whole question groups (e.g. all questions pertaining to ‘quality’, ‘impartiality’ or ‘corruption’). In addition, for several simulations, we re-aggregated the data from the individual level, whereby we excluded certain demographic groups, such as men, high income respondents, young respondents, higher educated respondents and those who did not have any interaction with any of the public services in question within the last 12 months. We found that, even in the most extreme scenarios, the Spearman Rank Coefficient never fell below 0.90 and that the median shift in the rankings was never above 11. Of the 62 simulations, we show the visual effects of the two most extreme cases in Figure 3. A perfect ‘1’ in the Spearman Rank Coefficient would of course look like a perfectly straight line from the bottom left corner to the top right corner of the graph. Scenario 14 (FA weighting scheme, min-max normalization and leaving out the Corruption pillar) along with scenario 61 (EW, including only respondents with relatively low education) revealed the lowest two Spearman Rank Coefficients (.919 and .921

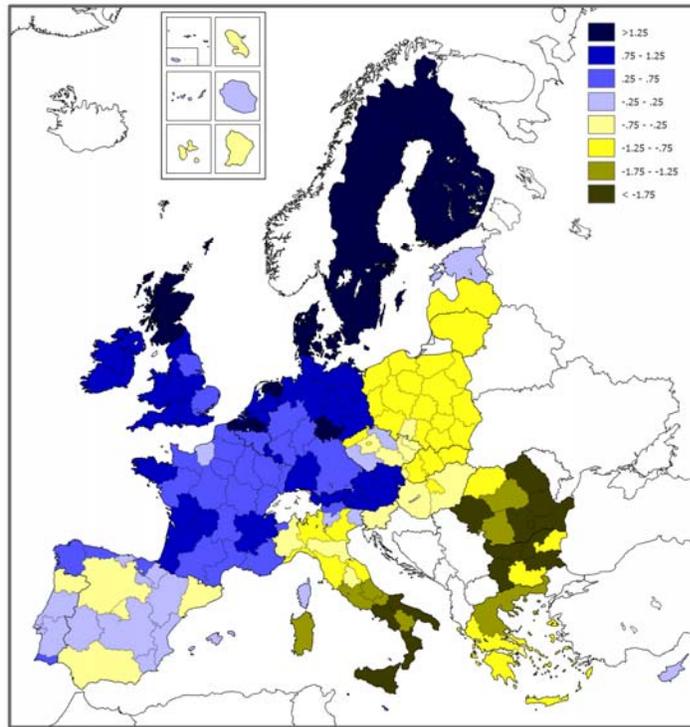
respectively). Scenario 14 had the highest median shift (14 places) while scenario 61 the greatest single region shift (68 places). As we can see, even the extreme outlier cases (Attica (Athens) in #14 and Opolskie in #61) do not look to be so out of line from their original position as to cause alarm, as both increase roughly one full standard deviation to end up just over the mean score of '0'.

Following the national-regional level comparisons that we discussed in section 2, we test further for external validity by performing a number of bivariate regressions and scatter plots with regional-level variables for which we would expect (or not expect) to find an empirical relationship with QoG. We find that the regional QoG index is strongly related to such factors as GDP per capita, various measure of unemployment, education levels of the workforce, use of technology, level of non-resident tourism and income inequality. Similar to the national level QoG index, the regional level index is not related empirically to population – either total population or population density. The full analysis can be found in the appendix to the report.

5. Combining the External and Internal Assessments: The E.U. QoG Index

In order to maximize the number of EU countries in the sample (e.g. including those countries with only one NUTS 2 region, such as Estonia or Slovenia) we elect to combine the WGI external assessment with our citizen-based, regional level data. Our aim was to devise a method that mostly fairly includes the EU countries omitted from the survey while simultaneously maintaining the richness of the within-country variation in several of the countries surveyed in the regional level study. In this process, we combine the ‘external’ assessment from the WGI, based largely on expert or investment firm opinion, with the ‘internal’ assessment of the citizens from each of the 172 regions in the survey or, in this case, the ‘consumers’ of QoG.

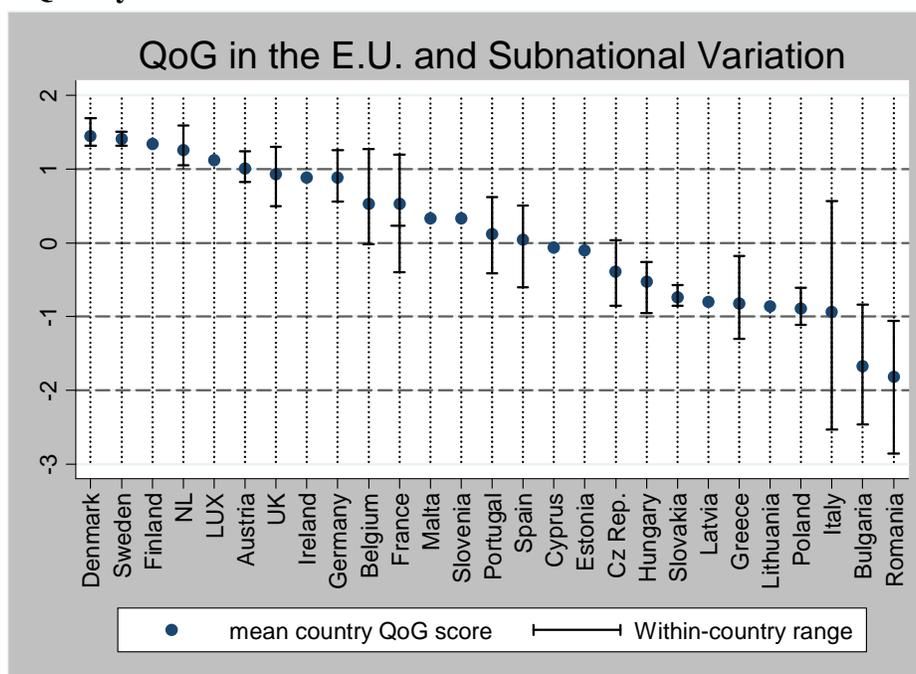
We essentially take the QoG country average from the World Bank data from Table 1, standardize them for the EU sample and set each country’s mean score to that level for each QoG pillar. For countries outside the survey, there is nothing to add to the WB country score, and thus these national level data are used as the QoG estimate alone, and regional variation is unobserved. With respect to countries that have regional data, we set the national average using the external assessment and explain the within-country variance using the regional level QoG data. Simply speaking, we take the unweighted, average regional QoG score for each of the 18 countries in the survey and subtract this from each region’s individual score, which is then added to the national level data for each pillar, thus giving each region an adjusted score. In doing this we combine the national level, World Bank data with the regional level data presented in Figure 3. We feel that this method makes the best use of both external and internal assessments, combining our citizen survey with the opinions of the main experts, NGOs and the IGO data of the World Bank. After calculating the scores for all of the regions, we then standardize them one more time, the results of which are shown in Figure 5, which constitute a total of 181 political units in the EU. The complete list of regions and countries in rank order with their respective QoG score, along with each individual pillar, is given in the appendix. Figure 6 shows a rank order of countries by QoG scores and sub-national variation within each country..

Figure 5: The Combined E.U. QoG Index

The data show that 11 of the 15 EU-15 states have all their regions and/or national scores above the EU average, while all regional and national level scores for the New Member States (NMS) are under the mean average; however, on specific combined pillars, such as RL or GE, several regions in the Czech Republic are above the EU mean score. Three countries – Italy, Spain and Portugal – have regions that are both above and below the mean score, while Greece is the only EU-15 country to have all of its regions below the mean level of QoG in the EU.

Interestingly, the data show that within-country QoG variation is at times equally or more important than cross-country variation. For example, in Figure 6 shows the rank-order of E.U. countries (again, using the WGI, national assessment as the country mean). Using the margins of error for the estimates provided by the World Bank, we see that differences between countries like the Netherlands and Denmark, Belgium and Portugal, or Czech Republic and Poland are negligible. However, in several countries, such as Italy, Spain, Belgium, Portugal, France, Romania, Czech Republic and Bulgaria have multiple significantly different regions according to the regional-level margins of error while Greece, U.K., Poland, Hungary and Germany have one significant difference (e.g. the only significant within-country different is between the high and low region). The other 14 countries (nine of which were not measured for regional variation) have no significant within-country QoG variation, meaning that all regions' margins of error overlap.

Figure 6: Quality of Government in the E.U. and Subnational Variation



Note: the two lower lines in France represent the range with (the lowest line) and without the overseas territories.

5.2 Correlates of QoG in the EU

We now check to see the strength of the empirical relationship between the QoG index and a number of economic and social indicators of development. For this we present only basic bivariate regression results, along with a few scatter plots. Of course we acknowledge the likelihood of edgoneity (e.g. two-way causality) between several of these variables, but the intent is primarily to show the external validity of the data in relation to several of the ‘usual suspects’ empirically related to QoG. Here we employ simple ordinary least squares (OLS) regression and report robust t-statistics in parentheses. The correlate variables we report here are:

- GDP per capita
- Long-term unemployment
- Infant mortality rates
- Heart disease deaths
- % of employees in research & development (R&D)
- % of households with internet access
- Population

Table 9: Bivariate Regressions

	<i>GDP p.c.</i>	<i>unemploy (L.T.)</i>	<i>% R&D</i>	<i>Infant Mort.</i>	<i>Heart Dis.</i>	<i>%internet</i>	<i>Population^a</i>
QoG Index	4850.7*** (11.09)	-1.14*** (4.95)	4.71*** (7.89)	-1.93*** (9.60)	-26.60*** (9.51)	13.07*** (12.92)	0.01 (0.24)
Obs.	181	181	86	181	123	115	181
Rsq.	0.36	0.09	0.48	0.36	0.43	0.65	0.002

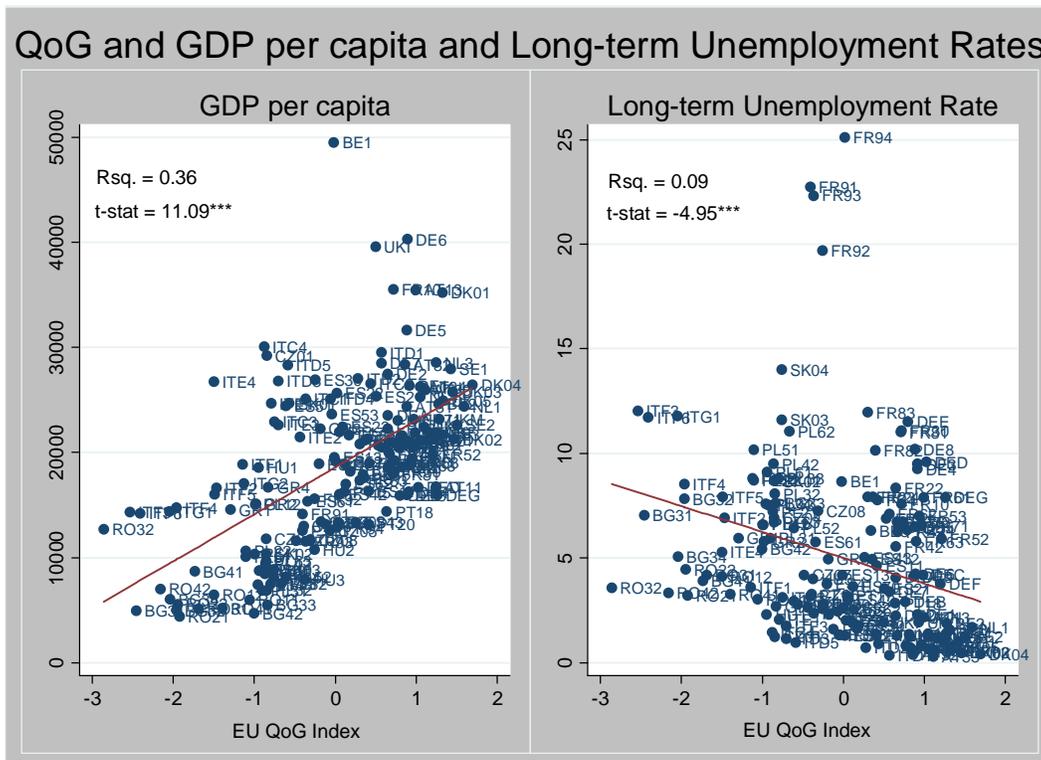
Note: robust t-statistics in parentheses. ‘GDP p.c.’ is the purchasing power standard (current market value, 10-year average). ‘Unemploy (L.T.)’ is the long term unemployment rate. ‘% R&D’ is the percentage of research and development workers in the region/country. ‘Infant mort’ rates are measured as the number of infant mortality deaths per 10,000. ‘Heart Dis.’ Are heart disease deaths per 100,000 inhabitants (from 2006-08). ‘%internet’ is the percentage of people with access to internet at home. Population is the log of the total population. All data taken from Eurostat are averaged over a 10-year period.

^aQoG index used as the dependent variable in this model.

*** $p < .01$, ** $p < .05$, * $p < .10$

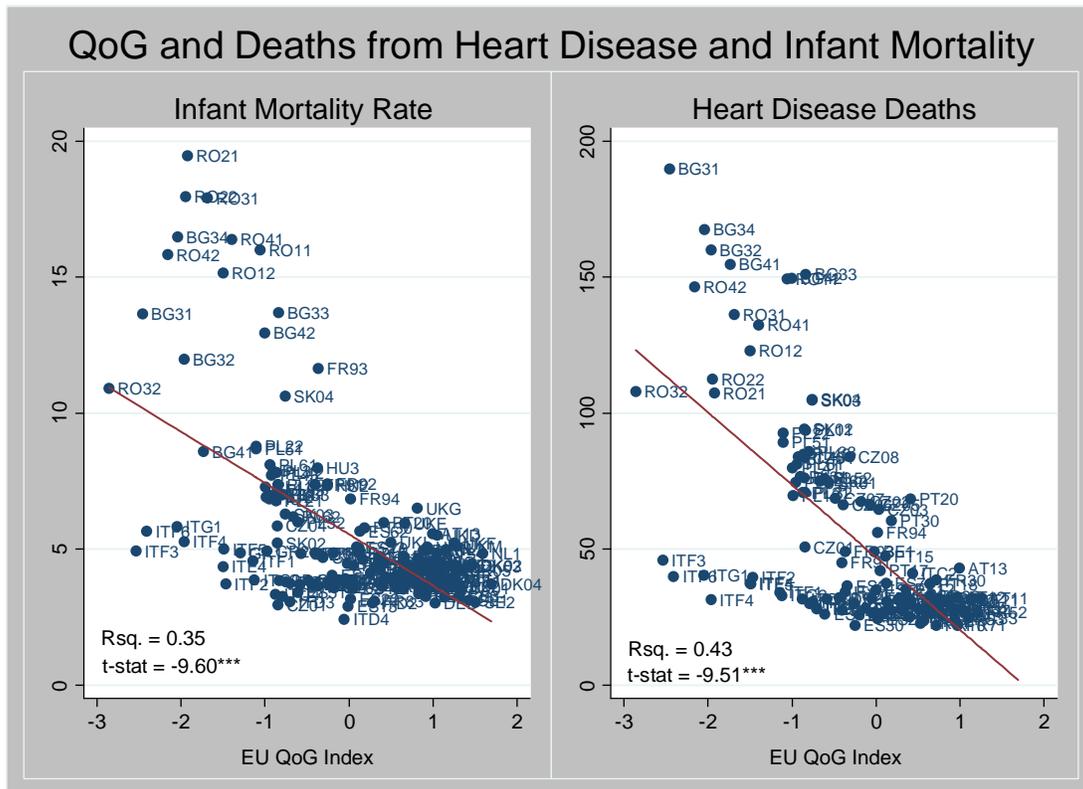
Table 9 shows the results for the bivariate regressions with the EU QoG index and the development indicators listed above. According to the data, the QoG index is associated with a greater GDP per capita of the region/country by 4851.7, while an increase from ‘0’ to ‘1’ in the index shows that the long-term unemployment rate is 1.14% lower. The regressions also reveal that that a one standard deviation increase in the QoG index is associated with almost 5% more employees in the research and development sector, while the same increase shows an association with 207 more employed nurses. We find that an increase of one standard deviation in the QoG index is associated with a lower infant mortality rate of roughly 2. A change in the QoG index from -1 to 1 is associated with a change in the per cent of households with access to internet of 26%. All of these bivariate relationships are significant at the 99% level of confidence. Population, however, is not a significant correlate of the QoG index, as the coefficient is far from statistical significance.

Figure 7



Note: GDP per capita is measured in purchasing power standard (PPS - current market value, 10-year average, 199-2008). Long term unemployment rate is from 2008 and both figures are taken from Eurostat.

Figure 8



Note: Infant mortality rates (from 2006-07) are deaths under 1 year of age / 1000 live births. Heart disease deaths (2006-08) are deaths per 100,000 inhabitants. Both figures taken from Eurostat.

Figure 7 shows the relationship between QoG and two salient economic indicators – GDP per capita and long term unemployment rates. Here we find that the relationships between both these variables and QoG are highly significant, and in particular with the case of GDP. Figure 8 shows the scatter plotted visual relationship between the EU QoG index and infant mortality rate and deaths by heart disease. It is clear that the pattern in the EU – although not perfectly – is that, the greater the level of QoG, the higher the level of economic development as measured in per capita wealth for a region in a country. As far as the distribution for infant mortality rates, we find that no region or country with a QoG level above ‘0’ (e.g. the mean score) has an infant mortality rate greater than 7.5, while no region or country with a QoG score above ‘1’ has a rate above 6. Similarly, no region with a QoG score above ‘1’ has a rate of 50 or higher in Heart Disease Deaths. However, in cases of a QoG score below ‘-1’, we find wide degrees of variation in infant mortality, mainly between Bulgarian and Riemannian regions, with the low-QoG Southern Italian regions, which perform relatively well on the two health indicators

6. How to Improve Quality of Government? Lessons from the Best Performing Regions

This section attempts to explain why some EU regions present higher QoG than others by looking at the characteristics that above average regions in QoG have in common that, at the same time, make them different from below average regions. The empirical material used in this section is heterogeneous, from official statistics, scholarly and media reports to the opinion of dozens of experts and professionals from ten regions:

Bolzano and Campania (Italy), Wallonie (Belgium), Severozapadan (Bulgaria), Västra Götaland (Sweden), Pais Vasco (Spain) Norte (Portugal), Nord Vest (Romania), Jihozapad (Czech Republic) and the country of Estonia.. The main “lessons” as to how to improve QoG are the following.

6.1 Independent and active media

First, it is important to remark that the countries with poorer levels of QoG belong to the EU countries with the lowest positions in the *Reporters Sans Frontiers's* 2009 World Press Freedom Index, such as Romania, which occupies position 50, or Bulgaria, ranked 68. This is common knowledge in the scholarly literature. According to the experts, one of the main triggers of high levels of QoG in a given country is the existence of a free and widely read press that keeps an eye on abuses of power. An example of the media playing a watchdog role - as “the people's eye on the state”- is in Sweden where, in April 2010, the investigative television programme *Uppdraggranskning*, of the state owned channel SVT1, uncovered a case of corruption in Gothenburg municipality departments and a municipality-owned housing association. While the case had earlier been reported internally, the case had been shut down because the internal investigation was inconclusive. However, the further investigative work of the television journalists, followed by their colleagues in major local and national newspapers, forced the public authorities to re-open it.

The existence of a free and independent media seems to be correlated more with the level of QoG in a region than its level of economic development. For instance, a founding member of the EU with an advanced G-8 economy, Italy, ranks 49 in the world in terms of press freedom – behind, among others, several African countries such as Cape Verde, Ghana, Mali and South Africa – while Estonia, a new member state with a relatively recent communist experience and a small economy, ranks sixth in the world. Similarly, while Freedom House rates Estonia as “free” with respect to press freedom – together with the vast majority of advanced democracies – it rates Italy as only “partly free” since 2004. These notable, and relatively paradoxical in a historical and socio-economic point of view, differences in media freedom may account for the also relatively paradoxical differences in QoG between Estonia, which ranks 16th among the EU-27 (and first among new member states), and Italy, which ranks 25th.

Experts and professionals consulted in Estonia regard its media as key for understanding its relatively high level of QoG. Especially in comparison with its Baltic neighbours, the Estonian media are considered to be more independent of political parties

and entrenched economic interests⁴. This independence would be the result of both an Estonian media ombudsman who closely monitors the media and fierce competition among the largest newspapers⁵. Similar to EU countries leading the world rankings of QoG, such as Sweden, Estonia has popular investigative journalists who on a regular basis report cases of potential public concern about government activities on a regular basis in both the written press and prime time TV shows⁶.

Quite the opposite, according to some accounts (*Reporters Sans Frontiers* 2010), Italian reporters may be threatened if they attempt to uncover certain stories about the political elite. Even some of the exceptions to the Italian average level of QoG can be partially explained with reference to the media. For instance, the bilingual region of Bolzano – the highest performer in Italy and which ranks among the top 10% of the 172 EU regions surveyed – has a unique media situation with both German and Italian newspapers and radio stations. Experts consider that this increases media’s ability to check power because both types of media complement each other: the Italian-speaking media is more critical of the institutions dominated by German-speaking individuals and vice versa (Charron 2010, 20). Similarly, in a region of Belgium, another EU founding member, such as Wallonie, regional experts regard the fact that “the media are probably not corrupt but they might have difficulties in reporting certain issues”⁷ may play an important role for understanding the persistence of a relatively low level of QoG.

In regions with relatively lower levels of QoG, the political dependence of the media has been found to affect both publicly and privately owned media. Experts in Wallonie⁸, in Severozapaden⁹ or in Jihozapad¹⁰ consider the political dependence to be a result of the financial support that political parties ultimately give to media groups. In the Norte region (Portugal), experts consider that the notable presence of Municipal Chambers as main sponsors of local newspapers and radio stations makes reports by the latter on QoG issues less reliable. In new member states’ regions, such as Severozapaden, in addition to the lack of independence, a traditional economic weakness must be added that impedes media from properly researching and informing on problems of QoG¹¹.

6.2 Policy of High Transparency

The regions studied indicate that public authorities can play a decisive role in stimulating media by virtue of adopting a policy of high transparency. Measures for transparency, allowing journalists immediate access to all types of information that have to do with

⁴Kazemi 2010, 251

⁵ibid, 252

⁶ibid, 253

⁷ Quote from an expert, 21 May 2010. Translated from French: « Je ne pense pas que les media soient corrompus mais je pense qu’ils manquent nettement de moyens et donc qu’ils ne peuvent pas toujours mener les enquêtes qu’ils voudraient. » (Håkansson 2010, 222).

⁸Håkansson 2010, 222

⁹Popovski 2010, 315

¹⁰Dinescu 2010, 281

¹¹Popovski 2010, 315

public activities, have traditionally formed the backbone of quality of government in Scandinavian countries. In the case of the particular region explored in Sweden, Västra Götaland (Sweden), the familiarity of local journalists with public information is so high that one of them acknowledges that civil servants often call him and tell him where and when to search for a certain document¹².

Similarly, the economic capital of Spain's best performer in QoG, País Vasco, was ranked in 2008 and 2009 as the most, and the second most, respectively, transparent municipality in the country, thanks to the extremely large number of administrative internal procedures and decisions that are available to citizens, journalists and civil organizations¹³. The regional government in País Vasco has also been a pioneer in offering information about its activities, not only in comparison with other Spanish administrations but also within the context of the EU. The País Vasco government has three different institutional mechanisms that offer an unusually high level of transparency: *Open Data Euskadi*, a web site in which all information not restricted by privacy, security or property reasons is available to citizens; *Ireki Open Government*, a direct channel of on-line communication and participation between decisions taken, or to be taken, by the regional government, and its citizens; and *Zuzenean*, the agency of citizen attention¹⁴. This unusually high level of transparency is regarded by the experts in País Vasco as an important factor for understanding the region's high level of QoG.¹⁵

The empirical evidence from these experiences at the local and regional levels of transparency and the overall positive effects in other countries forced the Spanish government to launch a Transparency and Citizens' Access to Information Act in 2010. The main innovation in transparency is that the default option across all Spanish administrations will be transparency and not secrecy, with the usual exceptions regarding privacy and national security¹⁶. Notable positive effects for quality of government in Spain are expected as a result of the implementation of this transparency act¹⁷.

6.3 Active Civil Society Organizations

Nevertheless, the existence of a free media with a transparent government is not enough to foster QoG. According to the regions explored, it can be argued that a free media is necessary but not sufficient. It needs to be complemented by either numerous and strong civil society organizations – such as NGOs or professional associations – or by an open economy that allows foreign investors or citizens to exert an “exit” option – that is, the possibility to move to another region with better QoG.

The Nord Vest region in Romania, which is the highest ranking region in either Romania or Bulgaria, is a good example of having a combination of different factors.

¹²Jonsson and Varraich 2010, 176

¹³Transparencia International 2010

¹⁴Gobierno Vasco 2009

¹⁵Hernandez 2010, 78

¹⁶*El Pais* 17-08-2010

¹⁷Manfredi 2010

One such relationship is that of the printed press and the active civil society organizations. According to the experts in Nord-Vest that were consulted, the trust in the regional press is reinforced by the active role of various NGOs. One of these measures is submission of newspapers to analysis by a non-profit organization called the Romanian Bureau for Circulation Auditing¹⁸, which ensures the monitoring and correct transmission of the publication's activity (e.g. the number of copies edited) to users of the press¹⁹. There is also a growing history of collaboration between press agencies and NGOs such as Transparency International, the Center for Independent Journalism in Bucharest and Active Watch, which help to preserve media independence. An apt example is the case in which 2009 Cluj's Association of Press Professionals revealed a venal agreement between the County Council and a local newspaper. The agreement was a mutually beneficial one that concerned the allocation of the Council's marketing budget. The affair attracted the attention of more NGOs, which resulted in national media coverage, bringing the whole deal to an end. The chances for stopping the deal would have been limited if the press association had exclusively counted with its own means. As the chairman of Cluj's Association of Press Professionals acknowledges, due to the involvement of several non-profit organizations in the investigation, the deal was aborted. This example highlights the importance of the collaboration between civil society organizations and media, a mutually beneficial check and balance system whose collaboration increases the possibility of higher QoG.

Another "pressure from below" that amplifies the effect of press freedom is a high degree of economic openness in the region. The qualitative evidence from two especially open regions in Central Europe, such as Nord Vest (Romania) and Jihozapad (Czech Republic), shows the relevance of trade – of goods, services and human capital – as an incentive to deliver QoG. Given the higher levels of foreign direct investment in Nord Vest – in comparison to other Romanian regions – public authorities have been forced to, first, adopt faster new (and higher) standards in public service delivery and, second, satisfy the increasing demands of foreign and local investors to improve public infrastructure for the growing export industry and engage in long-term projects having to do with human capital formation²⁰. In the Czech region of Jihozapad, experts detect a similar factor – by export-oriented business – in improving public services and tackling corruption²¹.

6.4 The Advantages of Ethnic Diversity

The conventional view of experts in political science and economics has been that ethnic fragmentation leads to overall lower levels of QoG. It is argued that cultural homogeneity strengthens the sense of belonging to the same community and generally makes all sorts of social interactions easier. Moreover, the more social groups a society

¹⁸ In Romania, BRAT is an acronym for "Biroul Roman de Audit al Tirajelor".

Information can be found at <http://www.brat.ro/index.php?page=home>

¹⁹Borcan 2010, 132

²⁰Borcan 2010, 153

²¹Dinescu 2010

has, the great the risk for problems associated with collective action. However, this view is challenged by more than one region in the EU, where ethnic diversity has offered a boost in QoG. The three EU regions included in this study that displayed more ethnic diversity, Nord Vest, Jihozapad and Bolzano, were those that also had a relatively high QoG. The experts consulted point out that ethnic diversity has benefits for QoG. The long term exposure to ethnic diversity helps to develop a spirit of tolerance and collaboration among the different ethnic groups²².

In the most ethnically fragmented region in Romania, Nord Vest – populated by Romanians (75%), Hungarians (19.3%), Roma (3.5%), Ukrainians (1.3%), Germans (0.4%) and other ethnic groups (0.5%) – local public authorities have traditionally supported the development of human capital in the university center of Cluj, which is regarded as a key for understanding the current levels of tolerance and collaboration among the different ethnic groups. In Jihozapad, the large public investments in infrastructure and the strength of economic ties with neighbouring Germany help transform past violent relationships between Czechs and Germans – such as the “ethnic cleansing” of Sudeten Germans in 1945-46 – into positive interactions and the absence of a perception of partiality or discrimination among a particular group²³.

The most advanced institutional integration of different ethnic groups was found in the case of Bolzano. A particular aspect that highlights this institutional integration pertains to the employment process in the public sector, where a strict *proportionality rule* is followed. What this entails is that potential public employees must declare their native language before applying for a position in the public sector. The selection is made according to the proportion of the total regional population and the language group – Italian, German or Ladin – that the candidate represents. Despite this, there is a widespread perception that, once inside, civil servants may encounter more or less opportunities depending on their language background – for example, top positions in bureaucracy and health care may predominantly go to German speakers, although the general view of public employment in the region is one of impartiality and meritocracy²⁴. According to the regional experts, the success of this policy of institutionalization of ethnic diversity is obvious on its own, especially in contrast to the problematic policies of cultural integration that were so frequent in the past, such as the forced Italianization of the region during Mussolini’s rule²⁵.

6.5 Meritocratic Public Organizations

One key factor that distinguishes above-average from below-average EU regions in QoG is the protection of merit-recruited civil servants from attempts at politicization. The extent to which merit – as opposed to political connections – is rewarded within the

²²Borcan 2010, 131

²³Dinescu 2010

²⁴Charron 2010, 18

²⁵Charron 2010, 26

public organizations of a country is more important than other traditionally more explored state-related factors, such as the electoral system or the political regime. The experts from the EU regions explored here agree that those regions with above-average QoG have more meritocratic civil service systems than the regions with lower levels of QoG.

The paradigmatic example would be Västra Götaland (Sweden), where experts and professionals acknowledge the merit-based nature of Swedish public organizations. Even experts admit that “being in the right circles” matters, and thus that contacts also play a role in Sweden; “that you have done the right thing”) is also key.²⁶ That is, unlike other contexts, personal contacts do not replace the merit qualifications of the candidates in Sweden but instead go hand in hand with merit; once merit is satisfied, only then will a contact matter. Similarly, although the levels of meritocracy are considered lower than in Sweden, the general perception among the experts in Nord Vest is that civil servants in the region are less selected and promoted according to their allegiance to a particular political or family network than in other Romanian regions²⁷.

In contrast, the hiring of public employees on the basis of a contact network instead of merit is prevalent in many EU regions, two of them that were included in this study: Severozapaden (Bulgaria) and Norte (Portugal). Officials in Severozapaden admit that a majority of public employees are hired and promoted on the basis of either political, personal or business contacts²⁸. Since these contacts, according to the experts, give priority to other considerations than competence requirements, merit is hardly rewarded in this region. There are very limited positive incentives for public employees, which has promoted many of the qualified potential public employment candidates to move to other EU countries to find work, resulting in a loss of qualified staff. Experts in the Norte region express similar concerns about the hiring and promotion of public employees: contacts trump merit²⁹. Jobs are largely awarded – experts believing this to be over 60% - according to family, friendship or political ties.

Experts agree that these forms of contacts affect the quality of the public services, because those who are more capable and well-prepared are not always the ones who in the end will get the job.

An example of the stark contrast between regions of the same country is found in Italy, i.e. between Bolzano and Campania. While hiring in the public sector in Italy is theoretically based on merit, experts in Bolzano and Campania largely disagree about how merit is rewarded in the two regions. Bolzano experts note the relatively high meritocracy of the different regional administrations whereas in Campania personal contacts would supersede the meritocratic rule more often than not. One piece of quantitative evidence that may support these differences is the percentage of women employed in public administrations, given that patronage networks tend to be male-dominated. While the employment gender gap in Bolzano is very small and, as one

²⁶ Jonsson and Varraich 2010, 183

²⁷ Borcan 2010, 149

²⁸ Popovski 2010, 310

²⁹ Morgado 2010, 109

source said, “women play a much larger role in both politics and in the public administration than in other regions,” Campania presents one of the largest employment gaps in the EU – with an astonishing 31% difference between eligible males and females. This example helps illustrate the importance of rewarding meritocracy for higher QoG.

6.6 Protection of whistle-blowers

Experts agree that, to prevent power abuses by public officials, whistle-blowing must be protected from potential attempts by organizational superiors to curb it. The most obvious examples of the importance of whistle-blowing protections come from the Scandinavian countries – that usually rank at the top of any world index on curbing corruption. The public administrations in these countries offer an encompassing protection of rank-and-file public employees that “whistle-blow” – that is, that denounce either a case of corruption or mismanagement of public funds in the organization they work for. The possibility that whistle-blowers can act with full confidence that their identity will not be revealed has long been regarded as one of the (or the) most important measure for curbing corruption in Sweden³⁰. The citizen (or public employee) who informs the press about some particular public activity has not only the right to remain anonymous (included in the Swedish constitution), but her/his public employer could also be prosecuted if there is an attempt to investigate who has reported X to the press³¹. In other words, the successful polities in curbing corruption pay special attention to the protection of those public employees that take the risk of whistle-blowing. An important distinction that needs to be made is that of “right” versus “duty” in public employment in Sweden. Public employees in the health care sector are duty-bound to report, i.e. whistle-blow, if there is any mismanagement or acts of corruption, as lives are at stake. Public employees outside the health care sector, on the other hand, have the option of exercising their right to report.

In addition to these protections, whistle-blowing is actively promoted in Sweden. An example is the annual initiative of Transparency International Sweden, which encourages whistle-blowing by awarding prizes to the most relevant whistle-blower of the year³². Similarly, another initiative of Transparency International, in this case of the Spanish chapter, of rewarding those municipalities that are more transparent may have a positive effect in fostering a culture of good governance across one administrative layer – local government – with more prevalence of corruption cases in Spain³³.

The qualitative evidence gathered from interviews with both Swedish public employees and journalists shows how whistle-blowing works in practice. One of the points of agreement among the different experts is that whistle-blowing is a last-resort

³⁰ Erlingsson, Bergh and Sjölin 2008

³¹ Utvecklingsrådet för den statliga sektorn (2002)

³² The following people have received the award; 2007 Åsa Sveds –journalist at Gotland, 2008 Christer Hansson –local glassblower exposing the European glass cartel, 2009 Ara Abrahamien –Elite sportsman exposing corruption and bringing more transparency to elite sport. <http://www.transparency-se.org/Varlden.html>

³³ Estefanía 2009, Transparencia Internacional 2010

measure. Generally, the first measure taken by an employee is to utilize the internal reporting system, expressing concern to a superior within the organization³⁴. If there is no positive feedback, it is then that the public employee resorts to the media. Nevertheless, the experts agree that the interest of the employee is being weighed against the interest of the organization. If the employee notices that the organization will be damaged, his or her coworkers will also be damaged and the outcome could be social exclusion. Social exclusion tends to be the strongest factor in keeping public employees from reporting. If the employee instead sees the result as a win-win situation, it will be easier for him/her to report the problem.

The case of Belgium also provides a good comparative example between two of its regions: Wallonie and Flanders. Wallonie ranks just above the EU regional average of QoG whereas Flanders ranks among the 15% best performers, which is also illustrative of the importance of whistle-blowers' protection. As the experts on QoG in Wallonie point out, an important deficit of their region vis-à-vis the Flemish regards the relative fragile legal situation of whistle-blowers. While Flanders has a longer tradition of whistle-blowing protection and enacted an advanced protection system in 2005, Wallonie has not implemented any protection measures. Similarly, experts on the Portuguese Norte region consider that would-be whistle-blowers are poorly protected and that, knowing that they will probably be harassed and terminated from their jobs, public employees are fearful to denounce a case of mismanagement or corruption in their organization³⁵.

According to the experts, more problematic are the contradictions experienced by civil servants in regions such as Campania, where public employees are obliged by law to report any malfeasance at work if they are employed within the public sector³⁶. Yet, at the same time, there are no legal protections for those who whistle-blow, and anonymity is far from guaranteed. In sum, would-be whistle-blowers live trapped between the possibility of penalty in case of not reporting malfeasance and the potential backlashes in their careers in case they report.

6.7 Adoption of Private Sector Practices

It is usually claimed that public organizations should remain isolated from private sector dynamics to assure legality and impartiality. Although that has not been questioned in the cases analyzed here, one of the relatively striking findings is that public organizations in high QoG regions have adopted many practices similar to their private sector counterparts. To start with, in the best performing regions, administrative procedures tend to be as flexible and quick as one expects in the private sector. This allows high QoG administrations to rectify unexpected situations when these occur – e.g. the revelation of a corruption scandal – than administrations that must closely follow rigid and lengthy public law procedures. A telling example is the recent corruption scandal uncovered in Gothenburg, Sweden. On April 28, 2010, the investigative television program *Uppdraggranskning*, uncovered bribery that had been occurring over the past few years

³⁴Jonsson and Varraich 2010

³⁵Morgado 2010, 111

³⁶Charron 2010, 41

within the Gothenburg municipality and one of its municipality-owned housing associations. The City Council acted quickly in collaboration with the private sector and, by May 10th, the firm Ernst&Young delivered its report as external auditor. Later in the month the civil servants involved in the mismanagement of public funds were terminated and by June 11th the state prosecutor at the National Anti-Corruption Unit announced the charges against the suspects. The way the Swedish public authorities dealt with the situation represents an example of quick response to a potentially devastating corruption case which could undermine citizens' trust in the reputed QoG of the city.

Similarly, the experts in regions with higher levels of QoG agree that mainstream public organizations in their regions have adapted tools traditionally associated with the private sector in its procedures and policies. That would be the case, for example, of the regional government in Pais Vasco which, during the last decades, has been importing management tools such as the Porter model of competitiveness or total quality certifications such as ISO and EFQM. From the Porter model the Pais Vasco Government initiated a public policy oriented to the formation of Industrial Clusters as a regional long-term strategy for economic development. This and the EFQM seem to play an important role in the high quality of government observed by the public administration, specifically the high quality perceived in public services, because the public administration procedures are subject to implementation and evaluation according to these tools³⁷.

The conventional response in how to prevent public organizations from being monopolized by patronage and nepotism has been the isolation of public employees from the flexible environment in which their private counterparts work. Regarding staff policy, that should be done through primarily three measures: special employment laws different from the general labour laws in a country; access to civil service based on highly regulated and formal examination systems; and guaranteeing secure tenure once the civil servant has achieved a position. On the contrary, in line with the latest research, the experts consulted in this study confirm that the same trends can be observed at the regional level: that is, it is not necessary to isolate public employees in strict regulations regarding hiring, promotion and incentives³⁸. According to experts, in Västra Götaland, the whole recruitment process can take, under certain circumstances, just a few

³⁷Hernandez 2010, 63, 67

³⁸For example, scholars indicate that, regarding corruption and good governance, a "closed" bureaucracy does not seem to have significant positive effects. Dahlstrom, Lapuente and Teorell (2009) in an analysis of 52 countries show that, first, a meritocratic bureaucracy does not need to have those mechanisms in order to be capable of curbing corruption. The Swedish bureaucracy is a clear example: one of the countries ranking at the top in all indexes of control of corruption has most of its public employees covered by general labour laws; that is, like their private-sector counterparts. Sweden, which has an "open" civil service system, recruits most of its employees like standard firms or NGOs, according to the rule of "best-suited candidate for each position" (OECD 2004: 4). Second, having a "closed" civil service system – that is, in which recruitment, in sharp contrast to private organizations, follows formal procedures to theoretically guarantee transparency, legality and impartiality – does not need to lead to lower levels of corruption. According to Dahlstrom, Lapuente and Teorell (2009), the reason could lie in the fact that a *de facto* meritocratic recruitment – i.e. merit counts – is poorly correlated with a *de jure* merit system – i.e. highly regulated formal exams. This could be the case in countries like Greece and Italy where, for the experts in central government consulted in Dahlstrom, Lapuente and Teorell (2009), political connections matter more than merit to get many positions in the public sector.

minutes³⁹. In contrast, in regions with theoretically strict merit procedures, such as Jihozapad (Czech Republic), experts consider that personal and political contacts frequently trump merit in the recruitment process in the public sector. Although the process follows the encompassing Czech law in terms of civil service entry, and, in principle anyone with the adequate competence and skills should have equal chances for the job, those who lack either personal, business or political contacts must clearly outperform candidates with contacts in the selection procedures. The reason is that the latter often get privileged information on the selection process. The region of Wallonie also serves as an example of the importance of contacts to get a position. One of the experts elaborated on his personal experience where there had been three candidates for a position and the one that was selected was the one that was affiliated with the “right” political party⁴⁰.

Nevertheless, on the basis of experts’ responses in different EU regions, two different patterns of *de facto* non-meritocratization occur in countries with *de facto* merit systems. On the one hand, in QoG average regions like Wallonie, political contacts mostly trump meritocratic procedures at the managerial levels of the administration and only marginally in street-level bureaucrats and professional groups delivering core services.

In addition, Bulgarian experts agree that the pervasiveness of personal networks increases in small municipalities where contacts are easier to establish and more solid. This fits the pattern found in studies of nepotism and corruption in other countries, such as Spain, where small municipalities are significantly less transparent in their activities than larger ones. Therefore, there is an increasing, and cross-cultural, body of evidence questioning a conventional view among policy-makers and institution designers: that small polities (e.g. municipalities) deliver better governance because they help preserve the local identity and, since there are fewer voters per elected incumbent, these polities are more democratic⁴¹.

6.8 The importance of self-government with clear chains of accountability

The professionals and experts agree in the importance of substantial regional self-government for fostering high levels of QoG. Italy serves as a good case for comparison. Of its 20 regions, there are five that, for different historical reasons, enjoy a status of special autonomy – Bolzano (studied here), Valle d’Aosta, Friuli-Venezia Giulia, Sardinia and Sicilia – which includes extensive law-making abilities and notable fiscal and administrative self-government. Looking merely at the quantitative rankings it is difficult to claim that regional government fosters good governance, since the autonomous regions show a wide degree of variation. Three regions are well above the EU average, such as Bolzano, Valle d’Aosta and Friuli-Venezia Giulia, but two are below the European mean: Sardinia (if only slightly) and Sicilia (more clearly). Yet the experts’ views allow us to make a more subtle qualification. For instance, in Bolzano, experts consider that the region’s level of autonomy is not, on its own, a sufficient factor for explaining its high

³⁹ Jonsson and Varrach 2010, 183

⁴⁰ Håkansson 2010, 220

⁴¹ (Skaburskis 2004, Sorensen 2006)

levels of QoG, although it does notably contribute to it. The reason is the ample margin of manoeuvre regional that policy-makers enjoy to re-design institutions to best meet citizens' demands and increase efficiency. Another factor that contributes to higher QoG in relation to regional autonomy is the regionally-tailored hiring practices exercised within the region⁴².

Similarly, all Pais Vasco experts rank its high level of self-government as a key factor for understanding the high QoG in the region. Furthermore, experts emphasize the importance of self-government when it comes to selecting civil servants. The recruitment process for Basque public administration is in the hands of an independent agency – the Basque Institute of Public Administration (IVAP) – with a record for innovation in staff policy. In addition, Pais Vasco experts point out one important difference between the levels of self-government of Pais Vasco and Catalonia – traditionally, the two economically more developed regions that have similar aspirations to self-government. Unlike Catalonia, Pais Vasco enjoys a unique *Concierto Económico*, or legal agreement between the region and Spain by virtue of which Pais Vasco collects most of its taxes – including income tax, added value tax and property tax.

One key factor as to why regional governance seems to contribute to good governance is that the chains of accountability are clear. This is not the situation in the case of Wallonie, however, which does have ample self-government capabilities. According to its experts, the fact that authorities are divided among different layers of government – i.e. Federal, Wallonie, French Community, Provinces and Communes – negatively affects the transparency and the effectiveness in the delivery of public policies. There is a notable lack of accountability since citizens do not know which layer does what under what circumstances. For instance, a 2004 survey by the Walloon Institute for Evaluation, Prospection and Statistics (I.W.E.P.S.) on users' satisfaction with public services found that one out of five citizens was not able to properly identify the level of government involved in the provision of a service. This lack of accountability as a result of a multi-level governance with both cross-time and cross-regional variations within the same country has also been pointed out in the studies of other regions, such as Spain. In the view of some of the Wallonie experts the reason why some policy-makers prefer not to merge almost overlapping administrative layers is that merging would imply a reduction in the number of positions to be filled by political appointees⁴³.

6.9 Discussion

The eight “lessons” presented here are far from an exclusive and exhaustive to-do list to improve QoG. Yet they should be conceived as starting points for discussion and public debate. If one had to summarize the backbone recommendation that emerges from the ten case studies is that, as important as it is to develop mechanisms to punish misbehaviour in the public sector (e.g. anti-corruption measures), it is equally important to protect and reward those civil servants that contribute to good governance. Three protections stand out in particular in this analysis of EU regions as decisive for achieving high levels of

⁴²Charron 2010, 17

⁴³ (Håkansson 2010, 218)

QoG: protecting whistle-blowing, protecting meritocratic civil servants and protecting an independent media.

Part I – Review of Quality of Government Indicators and Related Literature

“The most significant achievement in governance during the 1990s was the shattering of the taboo that barred discussion of corruption, particularly in diplomatic circles and intergovernmental institutions.”

United Nations *Global Programme against Corruption* (cited in Holmberg, Rothstein and Nasiritousi 2008)

“An overwhelming lesson that I think we have learned in the 1990s is... the transcendent importance of the quality of institutions and the closely-related questions of the efficacy of political administration.”

Lawrence Summers (cited in Besley 2007a, p. 571).

A brief review of the literature and available cross-country governance indicators reveals that this specter of bad governance is also haunting EU democracies. For many years many theoretical reasons have been suggested as to why democracies should exhibit a better quality of government than autocracies. However, the empirical evidence in favour of the democratic hypothesis is, at best, mixed. There are numerous accounts that show how quality of government decreases – and corruption increases – after moves towards democracy (Lemarchand 1972, Scott 1972, Wade 1985, Sayari 1977, Varese 1997 Weyland 1998). Many authors have explored what Harris-White and White (1996: 3) and Sung (2004: 179) define as the “contradictory” relationship between democracy and corruption – that is, that democratization is far from going hand in hand with improving a country’s quality of government (see, for example, Keefer 2007, Bäck and Hadenius 2008 or Charron and Lapuente 2009 for recent reviews of this debate). Contrary to the predictions of many scholars, especially after the expansion of democratic freedoms in 1990, as Larry Diamond (2007, 119) states, “there is a specter haunting democracy in the world today. It is bad governance.” There are numerous pieces of evidence indicating that the world’s most democratic region, the EU, may also be affected by this “specter” of bad governance. It is the goal of this report to make a diagnosis of what parts of the EU enjoy high levels of quality of government and what others suffer from bad governance.

This section provides a preliminary introduction to the problems created by bad governance in general and outlines how the literature review of this research project would appear for the particular case of EU countries. The aims of this section are to show, first, that there are notable differences in the level of quality of government across EU countries; second, that those differences, despite being inherently difficult to capture, can be measured with sufficient accuracy; third, that those differences in quality of government decisively matter for the economic and social development of countries and regions; and, fourth, that there is a long bibliography that deals with the topic of

comparative quality of government and can serve as the preliminary basis for writing a literature review.

1. Defining Quality of Government

A first word of caution before starting a study of this topic is that, as mentioned above, neither researchers nor practitioners have yet arrived at a standard definition of what “good governance” – or what this study prefers to call “quality of government” (QoG) – is (Rothstein, Holmberg and Nasiritousi 2009). A general problem in most research is that studies extend the concept of good governance or quality of government to almost all questions related to how states behave, which makes its scientific tractability very difficult given that almost all subjects in political science and political economy fall within the domain of governance (Keefer 2004, 5). In order to scientifically treat a concept, we need a working definition that we can, first, operationalize and, second, measure. What is absent in this debate is a consensual working definition of what exactly constitutes QoG. To address this issue it is imperative to delineate a definition of what QoG constitutes and what it does not (for example, the size or the “quantity” of government). The most frequently used definition of QoG in the literature is still that of the World Bank of governance:

“the traditions and institutions by which authority in a country is exercised. This includes: (1) the process by which governments are selected, monitored, and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions among them.” (Kaufmann, Kraay, and Mastruzzi 2004, 3)

Nevertheless, as Rothstein and Teorell (2008) argue, some parts of “politics” must matter more than others, because, “if governance is everything, then maybe it is nothing.” Many economists narrow the concept of good governance by defining it as “good-for-economic-development” (La Porta et al. 1999, 223), but then the problem is that this definition borders on tautology (Rothstein and Teorell 2008). A partial solution to these problems is to opt for a definition that is based on a normative criterion and, at the same time, can be empirically operable. This is the case in the definition proposed by Rothstein and Teorell (2008), where QoG is linked to the concept of *impartial* government institutions. They define an impartial government, or a high quality government, as one whose officials, when implementing policies, shall not take anything about the citizen/case into consideration that is not *beforehand* stipulated in the policy or the law. This working definition is measurable across countries, as the *Quality of Government Institute*’s pioneering expert-based survey⁴⁴ on the QoG of 56 countries shows, in which

⁴⁴ Most experts belong to these four international networks for public administration scholars: the Network of Institutes and Schools of Public Administration in Central and Eastern Europe (NISPACEE), the

countries are primarily ranked according to their level of impartiality. In some dimensions this survey represents one of the most encompassing datasets on bureaucratic structures at the cross-country level, with several measures regarding not only how the government delivers policies but also how it is structured. Still, it only covers central governments (unlike this EU report, which covers regions as well) and it surveys experts on public administration instead of users (as this EU report does).

A reflection by Rothstein and Teorell (2008) that deserves special attention is that quality of government does not equal democracy. Democracy, which concerns the access to government power, is a necessary criterion of QoG but insufficient in itself alone. The reason is that, if QoG were merely to equal democracy, the importance of *how power is exercised* would be left out, since the emphasis would only be on *how power is achieved*. As Paul Collier (2007, 2009) has extensively shown, although both dimensions are essential, the international community has focused its efforts – especially, although not exclusively, regarding developing countries – on fostering democratic means to achieve power, neglecting the key role that the limits to how power is exercised play in assuring social well-being. Looking at the empirical relationship between democracy and QoG, a curvilinear relationship can be seen representing that at times *some democracy* can be worse in terms of QoG in comparison with *no democracy* at all. The longer a state has been a democracy, the more likely it will be that its citizens benefit from high levels of QoG; yet this transition, as the evolution of some post-Communist regimes indicates, is not as straightforward and automatic as many thought only a couple of decades ago. One cannot overlook the fact that some of the traditionally greatest achievers in terms of “good governance” are not democracies – such as Hong Kong or Singapore. Excluding these particular countries – or outliers – from the sample of analysis, it can be stated that QoG generally requires both, democracy in the access to power and impartiality in the exercise of that power. That is, democracy may be a necessary requirement to enjoy QoG but it is far from being the sole sufficient one.

Apart from the above mentioned survey by the Quality of Government Institute, there are many cross-country indicators that are intended to capture the level of quality of government – or, alternatively, the level of corruption, which is frequently seen in the literature as the inverse of quality of government. No available indicator is free from measurement problems, as Task 2 will show at length. In this section we would simply like to mention that trying to measure corruption in the public sector across space and time is for the most part difficult due to its clandestine nature and difficulty to observe. Two main approaches have been taken in the recent literature. The first is to use a ‘hard measure’, employing indicators such as conviction rates or reports of corruption (Alt and Lassen 2003; Goel and Nelson 1998). The second and more prevalent approach,

European Group of Public Administration Scholars (EGPA), the European Institute of Public Administration (EIPA), and the Structure and Organization of Government (SOG) Research Committee at IPSA.

especially in cross-country studies, uses perception-based indicators to measure QoG or corruption. This has become the common norm for a number of reasons.

To begin with, a ‘hard measure’ might be a better test of a country’s legal system or in fact its ability to detect corruption – not its *actual* corruption – if used in a cross-section analysis. Thus ‘hard measures’ may lead to significant biased results (Charron and Lapuente 2009). Similarly, many authors, such as Robert Rotberg, argue that quality of government should be defined according to a set of objective outcomes, such as literacy levels, school persistence rates, medical services, state of law and order, and civil society empowerment. Nevertheless, as mentioned before, that approach amounts to a tautology, thus rendering it unworkable from a scientific point of view (Holmberg, Rothstein and Nasiritousi 2008).

Furthermore, as Kaufman et al. (2008:3) argue in a summary of this debate on quality of government indicators, “perceptions matter because agents base their actions on their perceptions, impression, and views.” If citizens or foreign firms perceive a given country’s administration to be plagued with corruption and public sector mismanagement, they are less likely to use its public services, with deleterious consequences for the country. An additional reason to use perception-based indicators of corruption or quality of government is that they tend to go closely “hand in hand” (Holmberg, Rothstein and Nasiritousi 2008). For example, three of the most widely used indexes of quality of government – the World Bank’s “Government Effectiveness Index”, Transparency International’s “Corruption Perceptions Index” and the “Rule of Law Index” also from the World Bank – correlate at the 0.90 level.

The general impression is thus that, despite their potential problems of subjectivity, they seem to be capturing very closely related phenomena. Treisman elaborates “...perceptions – even if not matched by reality – can have powerful effects. Mauro (1995) found that perceived corruption reduced growth via its effect on investment but found much weaker evidence of effects going via other paths, such as increased costs or reduced productivity. Other studies have confirmed that higher perceived corruption reduces foreign direct investment. Soares (2004) found that high perceived corruption discourages victims from reporting crimes to the police. So perceptions clearly matter.” For these reasons, this report proposes that EU regional and country level differences in quality of government be measured to the greatest extent through perception-based indicators – and, in particular, through an EU-wide survey of households in which regular “users of government” evaluate them according to this definition of QoG.

The aspect of QoG that surfaces in the media and has been of great concern to Europe of late is that of corruption. QoG is very often described as the *absence of corruption*. Despite this, QoG encompasses much more than lack of corruption (Rothstein 2008); as mentioned above, the idea of measuring corruption is an essential component and a definition should thus be provided. Generally, the literature defines

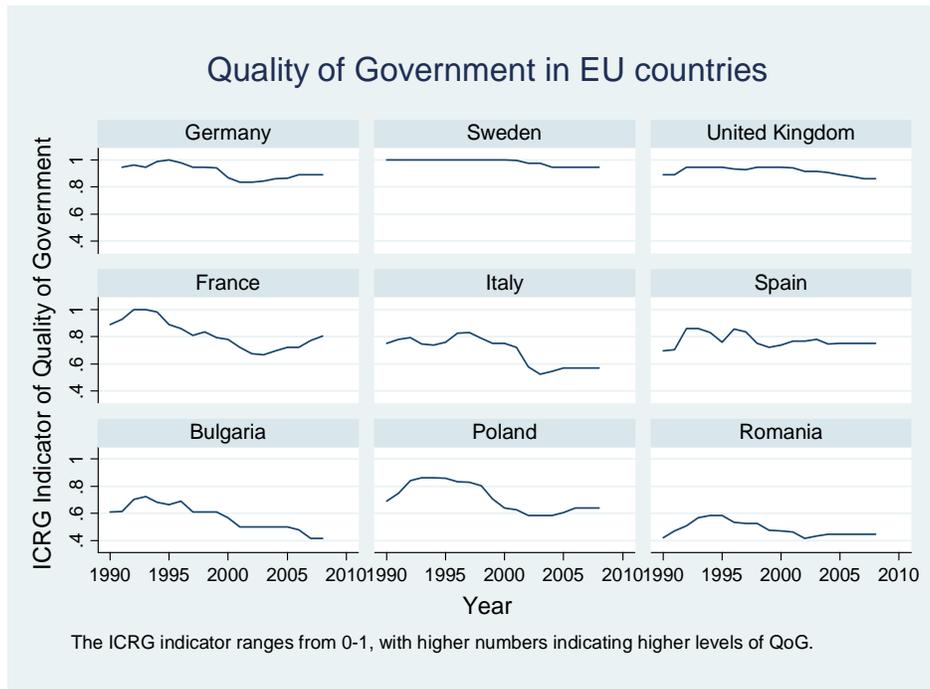
corruption as "abuse/misuse of public office for private gain." This definition suffers from "abuse/misuse" not being specified, giving leeway for the evaluator to interpret it in accordance with his or her own perception, which could vary in accordance with different cultural backdrops. Corruption therefore is considered here as a key, but not the unique, feature of QoG (as we will show in subsequent sections). In the context of the broader concept of QoG, corruption is defined here as a situation that "involves a holder of public office violating the impartiality principle in order to achieve private gain" (Kurer (2005, 230). In particular, there are two main forms of corruption that this report will deal with: grand corruption and petty corruption. The former pertains to corruption within the higher spheres of government – large sums of money in bribe taking/giving at high offices with the goal of, for example, getting a public contract – whereas the latter refers to corruption experienced by the citizen as part of their daily exchanges with the public administration – e.g. paying a small amount of money to the hospital orderly to jump ahead in the queue.

2. Some EU countries better than others

Looking at the evolution of any available indicator of Quality of Government, one can see important differences within the context of EU countries. Figure 1 plots the historical trend of one of the most all-encompassing and quoted measure of QoG – the *Political Risk Services'* ICRG indicator of Quality of Government – for a selected group of nine EU countries in the period 1990-2008. States receive a higher score when they are perceived to, first, be able to withstand a change in government without experiencing traumatic disruptions of services and day-to-day administrative functions and, second, when potential corruption in the form of excessive patronage, nepotism, job reservations, 'favor-for-favors', secret party funding, and suspiciously close ties between politics and business is perceived as being the exception and not the rule.

Although somewhat simplistic, this figure provides a basic preliminary overview of a phenomenon pointed out by many leading economists when they refer to the main political challenges of developed societies. As Dani Kaufmann noted: "if anybody thought that the governance and corruption challenge was a monopoly of the developing world... that notion has been disposed completely" (quoted in Rothstein 2009:2). The general trend that emerges from this extremely simple empirical exercise is far from optimistic. On average, European countries do not show a clear improvement in their levels of quality of government in the past two decades. As a matter of fact, the slope for most EU countries is, if anything, negative. Whether this general downward pattern is due to measurement issues (we discuss these questions below in this document) or really obeys to a true decline in the levels of quality of government needs further investigation. Nevertheless, what can be said is that, at first sight, there does not seem to be empirical support for the arguments that link democratic experience or membership in the EU to better governance.

Figure 9. Evolution of QoG for three groups of European countries



When we explore cluster-grouping differences among the EU countries in a later section in this analysis in much further detail, we can begin to see notable differences – and, what is especially relevant, persistent across time – among EU countries. First, there seems to be a group of countries, such as Germany, Sweden or the UK (although there are six others with a similar pattern, such as the Netherlands, Denmark or Finland) that exhibit steadily high levels of quality of government during the period under scrutiny here. These countries also rank at the top of any global indicator of good governance. They invariably score between 0.85 and 1 in the 0-1 scale of quality of government chosen in this graph. In the second place, there are other EU countries, such as France, Italy or Spain, with significantly lower levels of quality of government – normally ranging between 0.6 and 0.8 in the 0-1 scale. Looking at the ICRG data for other countries of Southern Europe, such as Portugal or Greece (not shown above), reveal similar levels of QoG. Note that these countries do not show convergence towards the levels of the countries at the top. If anything, during the most recent ten to 15 years, these countries have been diverging even more from the best performers in terms of quality of government. It can be argued – although, again, this should be subject to further inquiry – that the idea of a gap between Northern and Southern Europe in terms of quality of government seems more evident nowadays than two decades ago. In addition, the lack of convergence is even more remarkable in the case of many Central and Eastern European countries, such as the three selected here. Bulgaria, Poland and Romania present very low

levels of quality of government which, to say the least, do not seem to have significantly improved after joining the EU.

Looking at data over time from Transparency International's control of corruption index (CPI), we can see how the position of 27 EU countries in the world ranking has evolved during the latest decades⁴⁵. It is interesting to note that the top rankings over the years have steadily been held by the Nordic countries, Sweden, Finland and Denmark, whereas the Benelux belt states maintain a fluctuation around the same rankings, constituting the middle ranking countries. Mediterranean countries such as Portugal, Greece, Italy and Spain exhibit quite intense downward movements in the ranking during the latest decade. For example, Portugal and Spain evolve from being the 22nd and 23rd best countries in the world to becoming 35th and 32nd respectively. More dramatic is the impoverishment of Italy and Greece in this ranking: Greece goes down from position 36 to 71 in the world, while Italy moves from 39 to 63. Non-EU countries so diverse as Chile, Qatar and Botswana have become, at least in the eyes of the CPI, better performers in curbing corruption.

If we move from the general overview of quantitative cross-country aggregate data to the more specific studies of quality of government, we can also find important differences within the three different sub-groups of EU countries that seem to emerge at first sight. For example, within Central and Eastern European countries we not only find some dissimilar patterns in the outputs of the states – e.g. in their positions in the QoG international rankings – but also in the inputs they use to deliver policies. In that sense, there are remarkable divergences in what O'Dwyer (2004) defines as a common characteristic in post-socialist Europe: the “runaway state building” – that is, the paradox that post-communist state administrations have tended to grow after their anti-state revolutions. While in countries like Poland the number of personnel in the central state administration has outstandingly increased, in others, such as the Czech Republic and Hungary, state administrations have expanded little or even contracted. Task 2 extensively deals with these differences in the EU with the existing available indicators.

Nevertheless, it is important to note here the shortcomings of the existing indicators and that further research is needed to truly assess the level and relevance of the differences in quality of government and/or corruption in the EU. And that is the overall goal of this project. The lack of comparative research on this topic in the context of the EU is all the more surprising given the fact that several new parties in Europe, focusing on anti-corruption, have lately become extremely successful in their first contested election and that the corruption issue thus seems to play an increasingly important role in election campaigns. National Movement Simeon II in Bulgaria 2001, New Era in Latvia 2002, Res Publica in Estonia 2003 and Labour Party in Lithuania 2004 are the most prominent examples of such new parties that have won their first election. In Bulgaria,

⁴⁵ See the appendix for the full table.

Estonia and Latvia the leaders of the new parties also became prime ministers (Bagenholm 2009).

3. Consequences of Quality of Government: How QoG affects social well-being

Do the existing cross-country perception-based indicators of quality of government reveal something that could be relevant for the people living in those countries? There is less controversy over this issue in the literature since there seems to be a general acceptance among scholars and policy-makers as to the crucial role of government institutions for the welfare of its citizens. On the one hand, “good governance”, “state capacity” or “quality of government” foster social and economic development (Charron and Lapuente 2009: 2); on the other hand, dysfunctional and corrupt government institutions play a central role in many of the world’s most pressing economic and social problems (Rothstein and Teorell 2008: 166). In general, recent political science debates have been flooded with research on the importance of “good governance” – from Kaufman's (2004) findings that governance matters for basic human rights protection to Rothstein's study (2009) of how quality of government matters for understanding the recent financial crisis.

In the first place, especially during the past 15 years, numerous academic studies and statements by international organizations, such as the World Bank and the United Nations, have emphasized that only with a high quality of government can a country reap the benefits of economic growth and social development (Holmberg, Rothstein and Nasiritousi 2008). Good governance is seen as a necessary requirement for countries to foster economic development (Acemoglu, Johnson, and Robinson 2002; Acemoglu, Johnson, and Robinson 2004; Clague et al. 1999; Easterly 2001; Easterly and Levine 2003; Hall and Jones 1999; Knack and Keefer 1995; Mauro 1995; Rodrik, Subramanian, and Trebbi 2004, Rothstein and Teorell 2008).

This large number of economists has shifted the focus away from the traditional variables in economics for explaining economic growth, such as physical capital, natural resources, and human capital, to quality-of-government-related matters (Holmberg, Rothstein and Nasiritousi 2008). They also depart from traditional variables in sociology or political science (e.g. Putnam 1993), such as long-term cultural traits or the concept of “social capital.” On the contrary, as an increasing number of studies points out, the degree of social capital – or the level of generalized trust existing in a given society, as it is normally understood – seems to be determined by the quality of government institutions rather than the other way around (Letki 2003; Rothstein 2003; Rothstein 2005; Rothstein and Stolle 2003).

In one of those studies, Rothstein and Eek (2006) conducted two parallel experiments in two EU countries where the levels of corruption/quality of government and social trust are known to be different. Swedish subjects and one Romanian subject responded to a number of scenarios in which another person’s efforts to try to receive

immediate assistance from an authority were described. Subsequent to several scenarios in which participants were meeting either impartial or partial/corrupt officials, the levels of trust of the participants were measured. As expected, the Romanian sample had reliably lower initial levels of horizontal trust – that is, trust in others – than the Swedish sample, but in both cases the contact with corrupt officials decreased participants' levels of trust. These experiments thus support the idea that quality of government – or, more precisely, the level of corruption – affects the perceptions of the trustworthiness of others in general. Since the links between generalized trust and the economic development of a society have been found in the literature to be very robust, ensuring a high quality of government should become a priority for European policy-makers interested in economic growth.

Another noteworthy economic effect of quality of government is that it reduces income inequality and poverty (Gupta, Davoodi, and Alonso-Terme 2002). Further, it has also been argued to have substantial effects on diverse non-economic phenomena, such as education and health (Mauro 1998; Gupta, Davoodi, and Alonso-Terme 1998), subjective happiness (Frey and Stutzer 2000; Helliwell 2003), citizen support for government (Anderson and Tverdova 2003), the (lower) incidence of civil war (Fearon and Laitin 2003) and democratic consolidation (Rose and Shin 2001; Zakaria 2003).

For example, at the individual level, studies show that quality of government - as measured by the averages of six main World Bank indicators - explains a large part of the international differences in subjective well-being found through surveys (Helliwell 2003). At the aggregate level, for instance, strong positive relations have been found between several cross-country quality of government indicators and the most widely used measures of health outcomes, such as average levels of subjective health, life expectancy, infant mortality and child mortality (Holmberg, Rothstein and Nasiritousi 2008). In addition, inequity in health is higher in countries with more corruption (Eslava-Schmalback *et al.* (2008). Regarding the impact on other policy outcomes that at first sight could look less related to quality of government, such as environment protection, the results are similar. Corruption has a negative correlation with environmental sustainability (Morse 2006), enhancing pollution at all income levels (Welsch 2004)

Generally speaking, lower levels of quality of government indeed reduce human development – as measured by life expectancy, educational attainment, and standard of living (Akçay 2006, Holmberg, Rothstein and Nasiritousi 2008). As a result, scholars emphasize that the political legitimacy in a country, including those with higher levels of legitimacy, such as the Nordic European countries, depends at least as much on the quality of government as it does on the capacity of electoral systems to create effective representation (Rothstein 2008).

In summary, research consistently shows how QoG strongly determines many different aspects of social well-being in a community. The consequences that emanate from bad quality of government range from general economic, environmental or

educational outcomes to dire effects on health care. To gain a quantitative picture of the importance of QoG for a society's well-being, we can look into the impact of the three highly correlated and most widely used indexes in the comparative literature – the World Bank's "Government Effectiveness Index", Transparency International's "Corruption Perceptions Index" and the "Rule of Law", also by the World Bank – on a long series of indicators of social welfare. These three indexes are some of the closest proxies to the principle of impartiality, as defined by Rothstein and Teorell (2008), which, as mentioned above, constitutes the definition of QoG adopted by this report. Table 2 comes from Holmberg, Rothstein and Nasiritousi (2008)⁴⁶

The table shows the extent to which the variables are correlated with QoG. A positive association can be seen between QoG and both subjective health and life expectancy at birth – which constitute two of the most widely used indicators of health at the comparative level. Similarly, QoG reduces infant mortality and the prevalence of HIV. Other aspects of social well-being in a community are also highly correlated with the three selected indicators of QoG. Since correlations can not be evaluated as causal relations, it is not evident which one reinforces which but, generally speaking, QoG is highly positively associated with the major environmental, economic, social and even subjective measures of quality of life in a country. Aspects as diverse as air, water and forest quality, economic growth, social equality, employment levels, social protection, human development index, perceived happiness and satisfaction with life are clearly associated with the most prevailing proxies for quality of government.

The consequences that are constantly lifted to our attention by media are the ones that have to do with corruption. At this point, it is interesting to look at some examples of how pervasive QoG problems (and, in particular, corruption) are in some EU countries. A telling example of both grand corruption and petty corruption operating in parallel is that of Romania. According to some observers (again, we lack systematic empirical evidence, and that is precisely the goal of this study) the health care system could be so bedridden by corruption that medical care is hardly availed without a bribe being paid. Those who have faced corruption allegations in recent years have included a prime minister and more than 1,100 doctors and teachers, as well as 170 police officers and three generals, according to Romanian anticorruption investigators (*The New York Times* March 9th 2009). Only three months after accession to the EU the farm minister, Decebal Traian Remes, was caught on tape “taking delivery via a middleman of €15,000 (\$21,600), 20 kilos (44lb) of sausages and 100 litres (22 gallons) of plum brandy” as reported by *The Economist* (11/3/2007). If this was not important enough on its own, as the previous section has shown, corruption scandals affecting political elite and the interaction of citizens with petty corruption in their daily lives have their negative externalities in the form of further repercussions on quality of life and the levels of trust. Both horizontal

⁴⁶ A full table of correlation between QoG variables and quality of life indicators is found in the appendix.

trust – that is, generalized trust in other individuals – and vertical trust – or institutional trust – may decrease even more as a result of these practices.

With respect to Bulgaria, observers point out that bad governance may have taken a new turn when in February 2009 advisers to the prime minister, Sergei Stanishev, suggested that Bulgaria should give “the commission and other EU members” more power to intervene where “weaknesses may be qualified as structural and persistent and...cannot be resolved by the Bulgarian government alone.” The plan proposes that European officials and diplomats should be involved in monitoring the implementation of laws, managing EU funds and supervising courts, prosecutors and investigators. They would follow cases of political corruption and organized crime that the judicial system has been slow to tackle (*The Economist* 3/21/2009). The rule of law being undermined to the extent as in Bulgaria and Romania is a cause for concern, as voiced by EC president Barroso: “Citizens in both countries and across the rest of Europe must feel that no one is above the law.” Greece has steadily declined in its TI CPI ranking since 1998 and continues to fall. This decline has been highlighted during the recent financial and economic crisis, with investor confidence plummeting and the country’s budget deficit increasing to unsustainable levels. The central role of bad governance and, in particular, of corruption for understanding Greece’s difficulties is understood by the political elites. For example, at the EU summit in December 2009 Prime Minister Papandreou admitted that corruption was the main reason for Greece’s serious economic problems (*The Economist* December 17th 2009).

In this section we have seen some very preliminary indications that QoG, understood as impartiality, may differ quite a lot across EU countries. It is therefore interesting to assess, as this study will do in later stages, whether there are also differences within EU countries. It can be argued in general that, since quality of government appears to have positive effects on social well-being and there seems to be important (and, at least, not diminishing) differences among countries, further research in quality of government in the context of the EU is needed. In the light of an extensive literature – discussed above and referenced in the next section – the consequences of having (or not having) QoG are decisive for the quality of life in any given polity.

Part Ib – Existing National Indicators of QoG

According to the contemporary data, most indicators point out that QoG, or ‘good governance’, is generally higher on average in the EU-27 member states as compared with other world regions; yet there is significant variation among the countries in the EU. While this is generally accepted by many, it is difficult to measure such concepts as bureaucratic quality or corruption, to say the least.

That there is a debate about how we should best measure QoG remains in the fields of academic and practitioner research is unquestioned. Some doubt that contemporary indicators present valid measures of salient concepts under the umbrella of QoG, such as corruption, rule of law or bureaucratic effectiveness. In addition, there is strong disagreement over how QoG can be evaluated with respect to objective, or ‘hard’, indicators as opposed to subjective perceptions, as is the dominant practice in the field today. In summary, there are still many questions regarding both *what* we measure to aptly capture QoG and *how* we go about doing it. Therefore, any new measure will be impossible to capture perfectly and will undoubtedly confront a degree of opposition. In the following section (II) we discuss some of the advantages and pitfalls to the contemporary data in the field of QoG.

QoG has become a broad concept in the social sciences and, when quantified, has generally been disaggregated into categories such as: 1) ‘corruption’, 2) ‘rule of law’, 3) ‘bureaucratic effectiveness’ and 4) ‘government voice and accountability’ or ‘strength of democratic and electoral institutions’ (with respect to number 4, the ‘process by which government are selected and replaced’). Further, with some exceptions, these have been aimed at capturing the performance in the *public sector*.

In evaluating the many existing indicators of QoG at the national level for EU member states, we base our judgement of the most suitable data on the following criteria. First, we examine the quality and clarity of the descriptive information on which each of the data is constructed, or what we call the ‘internal criteria’. Second, we compare the indicators to one another, evaluating their coverage, their external validity and how using different sources impacts their empirical relationship with such outcome variables as economic development and income inequality. Due to time constraints, we seek to find the best and most robust existing indicator to build an index on which to rank EU countries’ QoG rather than ‘reinventing the wheel’ so to speak.

QoG is a broad concept that we break down into the following components:

1. Corruption
2. Bureaucratic effectiveness
3. Rule of law and impartiality
4. Strength of democracy and democratic institutions

1b.1 Leading Measures of QoG

In this section we elucidate several leading indicators based on survey data, their methods of construction and some of the strengths and weaknesses of their usage in comparing levels of QoG and/or corruption in a reliable and valid way. Indicators can be broken down and compared in many ways; however, two significant distinctions are of particular relevance to this project in evaluating strengths and weaknesses in the contemporary data.

- 1) Whether the data is a 'single-source' or 'composite' indicator
- 2) Whether the source is 'representative' (i.e. households or firms) or whether it is an 'expert' survey

Table 9 shows the distribution of QoG and corruption indicators that can be found in a majority of EU countries based on these two distinctions. As we can observe here, there is a good degree of variation in how the QoG measures are constructed, the sources used, the country and years of coverage and the number of different concepts under the umbrella of QoG that are captured by each source.

The recent quantitative literature on QoG and corruption studies in the fields of economics, political science and public administration demonstrates a willingness among scholars to employ one or multiple sources from the table above. However, it is quite rare to see a thorough account of the overall strengths and weaknesses of such indicators. Before breaking down such differences, we acknowledge that there are several weaknesses among *all* the leading indicators of QoG at this time.

Table 10: Examples of Leading Measures of QoG for EU Countries

<i>Data Source</i>	<i>Type of Evaluation</i>	<i>Start Year</i>	<i>Years included</i>	<i>EU Coverage^a</i>	<i>Type of QoG Concepts</i>
1) World Bank (WB)	Composite Index	1996	1996-2002 biannual 2003-2008 annual	All	Corruption, rule of law, political stability, bureaucratic effectiveness
2) Transparency International (TI - CPI)	Composite Index	1995	1995-2008 annual	All	Corruption
3) World Values Survey (WVS)	Household survey	1990 ²	1990, 1995, 2000, 2005	All	public trust, confidence in public services and political parties
4) Global Corruption Barometer (TI)	Household survey	2003	2003-2009 annual	All	Corruption
5) World Bank Investme Climate Assessments (BEEPS)	Firm survey	1999	1999, 2002 & 2005	CZ, Pol., Lith. Slov., Lat., Rom., Hun., Est., Bul. Slovak	Corruption
6) World Economic Forum's (WEF) Executive Opinion Survey	Firm survey	1979	1979-1999* 2000-2008 annual	All EU countries from 2001	Bureaucratic effectiveness, corruption
7) International Country Risk Guide (ICRG)	Expert Assessment	1984	1984-2008, annual	All	Corruption, rule of law, bureaucratic quality
8) Freedom House Nations in Transit	Expert Assessment	2003	2003-2008 annual	CZ, Pol., Lith. Slov., Lat., Rom. Hun., Est., Bul. Slovak	'Governance', rule of law, corruption, civil society
9) Freedom House	Expert Assessment	1972	1972-2008	All	Political rights, civil liberties rule of law ¹ , corruption ¹ , press freedom
10) Polity	Expert Assessment	1800	1800-2008	All	Strength of Democracy political stability
11) Fraser Institute	Expert Assessment	1970	1970-1999 once p years 2000-2004, annual	All	Rule of Law

^aFor latest year available.

*Indicates that a significant change in the data collection

²First survey launched by the European Values Survey, later replicated by WVS from 1995

¹these are sub-categories within the two indices 'political rights and civil liberties' that FH has published annually as of 2006

- The time frame of availability and country coverage. Apart from indicators of 'democracy' the vast majority of available data with few exceptions dates back only to the 1990s, making time series analyses and empirical tests of a 'causal mechanism' for change in such measures difficult. Further, no current indicator contains all countries in the world in a single year.
- As with any abstract concept in the social sciences, such as 'democracy' or 'minority rights', 'hard measures' of concepts such as corruption or bureaucratic efficiency are nearly impossible to decide upon. For example, if one chooses to employ the number of tried or convicted corruption cases annually as a measure

- of control of corruption, one might simply be measuring the strength of the rule of law or the effectiveness of the media.
- Thus the field has relied on perception-based measures as the primary mode of comparison across countries. This is problematic because of the notion that citizens' attitudes can be embedded in their general attitudes towards their government; firms or experts can be biased towards countries with low taxes or high growth.

1b.2 Single Source vs. Composite Source Indicators

Single Source Indicators

These indicators are derived by a single organization, which either assigns each country-year a ranking based on expert assessments, such as ICRG or Freedom House, or sends a survey to households, firms or country experts to gauge the yearly or semi-yearly levels of QoG, such as WVS (World Values Survey) or WEF (World Economic Forum).

On the plus side:

- We find that one of the primary strengths of the single source data, such as the ICRG (International Country Risk Guide), WVS, BEEPS (Business Environmental and Enterprise Performance Survey) or WEF⁴⁷ surveys is the precision with which the researchers can define and measure the desired concept. For example, while some may be interested in the notion of 'state capture'-type kleptocracy in measuring corruption, others might be more concerned with administrative or 'petty corruption' in a particular country.
- Second, as long as the system of rating does not change over time, such single sources can be more aptly used to make multi-year comparisons and not simply those across space due to internally consistent measurements.

On the negative side:

- Surveys of this kind may be irrelevant to certain users because their idea of QoG or corruption might be too narrow (or broad) for certain projects.
- There might be a good deal of measurement errors in some or all of the sources of this kind. We have established that no two sources measure corruption in QoG in the exact same way; thus any one of the sources might theoretically suffer from bias – either based on the respondents or the measurement tactics employed by the organization.
- Given that such surveys are costly and that obtaining a reliable number of respondents from a wide set of countries is difficult, such sources might be limited in their coverage across space and time.

⁴⁷ See Table 1 for an explanation of the abbreviated sources

Composite Indicators

At this time, there are two leading indicators that, instead of relying on one single source, pool together multiple sources in order to create an index of the concept in question. *Transparency International* was the first to do this, as attention to corruption was growing among many international actors in the 1990s. However, their index is limited to the concept of corruption. The second measure, launched by the World Bank originally in 1996 (Kaufmann, Kraay and Mastruzzi 2009), offers researchers data on issues such as corruption, ‘voice and accountability’, ‘rule of law’ and ‘political stability’ (they also aggregate these measures into one single ‘government effectiveness’ measure).

On the plus side:

- Reduces the measurement error of any single source. Any single source will have some measurement problems in attempts to capture difficult-to-measure ideas such as corruption or ‘rule of law’. However, given that such error is random, the composite indices will significantly minimize such problems relative to the single source method.
- Indicator will not be ‘too narrow’ in what it captures.
- Wider coverage of countries. In taking multiple sources for a given year, the composite index approach allows for more state-years to be ranked despite missing data from certain single sources.
- Sources can be weighted so that households or firms can play a larger role than an expert survey or vice versa. Further, the World Bank weights indicators that correlate closely while outliers are weighted less in order to avoid any single outlying indicator from having a substantive impact on the results.

On the negative side:

- There might be a greater degree of imprecision and inconsistency in the measurements. As discussed, different sources have different definitions of the concepts that make up QoG. In combining these concepts and taking for granted that the numbers used capture the same idea, they are implicitly accepting that their index captures multiple definitions.
- There may be less transparency in the construction of the data. Individual sources elucidate varying degrees of information with regard to the various weights and decisions that are included in their indicators. Pooling such data makes these indices far more nebulous.
- The sources may not be independent from one another. With the multiple sources available now, and with the popularity and attention of the TI and WB indicators, the single-source organizations – in particular those constructed of the decisions

- of ‘experts’ – could be basing their assessment on alternative sources or simply ‘free riding’ on the TI or WB data, which creates a feedback problem in the data.
- Different sources are used for countries in different years. In any given year, a country can have between one and, say, ten sources that are included its annual ranking for the World Bank (TI must have at least three) and the sources can change in any year. Thus there is a limited amount of internal consistency within these measures that makes year-to-year comparisons problematic. For example, none of the former socialist countries new to the EU had the same sources between 2004 and 2005 in TI’s data, making comparisons between these two years questionable.
 - Adding a new country can arbitrarily change the ranking of other countries. For example, the World Bank data are constructed yearly to have a mean of ‘0’ and TI’s system ranking countries in relation to one another. The inclusion of several new countries that have lower corruption than country ‘A’ from last year’s data would necessarily lower country ‘A’s rank in this year’s data even if corruption did not worsen. Inserting a high ranking country can arbitrarily reduce the scores of the previously included countries, while including low ranking countries can have the opposite effect.

1b.3 Expert vs. Random/ Representative Surveys

This distinction is made based on the source of the data obtained by the organization. Are experts determining the rankings of the countries in the sample behind closed doors or is the data obtained by random sample of household or firms? Such a distinction has potentially significant implications for the quality of the data employed. Table 1 notes the primary source of each indicator.

Advantages in using random samples:

- Data collected by WVS, Eurobarometer and the Global Corruption Barometer (TI) are based on household surveys. Since citizens are the primary consumers of government services, this can help to capture actual experiences of corruption or poor government performance in the bureaucracy.
- The average citizen or firm questioned will most likely not know the score of its country in the TI or WB data, therefore minimizing the ‘feedback problem’.
- They are determined by a larger opinion sample, compared with a small number of experts as in the case of the ICRG and Freedom House measures.
- Random surveys can capture a number of important questions and break down citizen’s opinions based on a number of salient distinctions in the public sector, as in the case of the TI India survey in 2005 that asked citizens about their perceptions and experiences regarding the quality of services in law enforcement,

schools, public utility companies and other public sector areas. As for business surveys, employees of firms can be asked directly about the extent to which they might have to pay a bribe to conduct certain business practices, the level of contract enforcement or the efficiency with which the bureaucracy operates.

Disadvantages in using random samples:

- The respondent might be uninformed and simply guess. If enough of the respondents employ this tactic, the results can be very unreliable.
- The citizen could be expressing his/her opinion of the current government and not the QoG concept in question, while the employee of a given firm could simply be basing his/her opinion on prospects for profit in the country in question.
- Over time comparisons are questionable because different people respond each year.

Advantages in using expert rankings:

- Scores are decided by full-time experts with a presumably deep knowledge of the countries and concepts in question.
- Since the organization decides the ratings internally, such as the Freedom House or ICRG measures, the over time comparisons are expected to be more reliable because of greater degrees of internal consistency.

Disadvantages in using expert rankings:

- Potentially less transparent. The organizations that employ this method do not publically release all their decision-making processes. This renders such data problematic because researchers are unable to replicate the data or choose their own weights for various components of the indicators.
- Less precision in the indicators. Because of the comparatively lower levels of transparency, what the data actually capture becomes more uncertain. For example, while the PRS Group lists a number of criteria that go into the expert construction of the data, the public is not privy to such information and what the researcher might be expecting to capture could be significantly biased by the uncertainty of what the organization is actually measuring.
- Data could be aimed towards specific audiences. For example, Freedom House is an advocacy NGO with an agenda for democratization and liberalization, while ICRG caters mainly to international investors seeking to profit in potentially new countries. If the goal of the researcher is not similar to the goal of the audience that the organization is aiming to please, then there may be implications for the empirical findings.

1b.3 Analysis and Evaluation of Current National QoG Indicators

We plan on analyzing the qualifications of the 11 indicators in Table 1 based on the following two broad criteria:

1. The ‘internal’ strength of the data

- *Conceptual relevance.* We review each of the indicators and elucidate the extent to which the variables capture the desired concept and the precision of the definition of the concept.
- *Clarity and replicability.* The ease with which the researcher can understand the construction of the variables and, based on the given definition and transparency of each organization, the extent to which the data are replicable by outside scholars.
- *Robustness of rankings:* We seek to test how sensitive the results are to alternative weighting schemes and aggregation methods. In addition, how sensitive are the results when certain underlying variables are excluded?

2. The ‘external’ strength of the data

- *Time availability and country coverage.* It is important that the indicator covers as many EU countries as possible for as far back as at least the year 2000 or earlier years.
- *Robustness of the data.* Are the indicators sensitive to a certain weighting in the coding of the variable? After conducting sensitivity and robustness checks, how do each of the indicators hold in their relation to economic and social indicators established in the political science/economics literature? Further, how do the data relate to one another? How closely do different sources correlate on the same QoG concept in the same country-year? We plan to make multiple comparisons to identify which source is strongest to multiple robustness checks.
- *Coverage of multiple concepts of Quality of Government by one source.* Some of the existing data measure only corruption or another single-issue concept under the umbrella of QoG. We will seek a wide scope of concepts that have been measured by the data sources.

We compare each of the indicators for these components of QoG separately.

QoG Concept 1: Rule of Law

With respect to the rule of law in the EU member states, we evaluate four contemporary indicators for which almost every EU member has been included in at least one year⁴⁸.

⁴⁸ All indicators except the ICRG data contain all 27 EU MS in at least one year. ICRG does not include Luxembourg and contains a maximum of 26 MS for certain years.

The four we examine are from the World Bank, ICRG, the Fraser Institute and Freedom House.

Internal Characteristics

A) World Bank

In their attempt to capture the rule of law (ROL), they define this concept broadly as:

“the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.” (Kaufmann, Kraay & Mastruzzi 2008: 7)

As noted in section B, the World Bank is a composite indicator that employs multiple sources to capture the QoG concept in question. For example, in the latest year for which the data are publically available (2008), a total of up to 25 different sources are used for the ROL⁴⁹ measure. The number of sources included in each MS varies from a low of eight different sources (Malta) up to 19 (Romania). The average number of separate data sources used in the 27 MS in 2008 was 14.4. The WB scholars argue that “each of the individual data sources we have provides an imperfect signal of some deep underlying notion of governance that is difficult to observe directly” (Kaufmann, Kraay and Mastruzzi 2008: 13). Thus each of the individual sources may have their own broad or narrow view of how to capture ROL and, so long as the World Bank experts view such measures to be under their broad definition, such measures are included. The underlying individual data of ROL include concepts such as: judiciary independence, property rights, the level of organized crime, respect for contracts, human trafficking, money laundering, and trust of police and the courts. For example, the World Bank uses the Cingranelli and Richards⁵⁰ measure as part of their ROL for ‘public sector data providers’, which has a narrow definition of ROL based solely on the independence of the judiciary, scored as 0, 1 or 2 (not independent to ‘generally independent’). Conversely, several of the business firm surveys employed by the WB such as the *Heritage Foundation* or the *Business Environment Risk Intelligence* focus their measurement of the ROL on property rights or contract enforcement by businesses and individuals. Further, household surveys such as the *Gallop World Poll* include questions that target individuals’ confidence in the police force or the judiciary or whether they have ever been the victim of a crime⁵¹.

⁴⁹ The World Bank employs a mix of business expert surveys, household surveys, NGO data providers and public sector data providers. The number of sources by category and the weights each category (in parentheses) for the ROL is 6 (.35), 7 (.19), 5 (.22) and 7 (.24).

⁵⁰ See http://ciri.binghamton.edu/documentation/ciri_variables_short_descriptions.pdf for more information on this particular dataset.

⁵¹ The World Bank’s annual report on their governance indicators carefully elucidates which sources capture which aspects of the variables in question. See: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1148386 for Kaufmann, Kraay and Mastruzzi’s (2008) full description of each concept measured.

If one is seeking to capture a narrow aspect of the ROL, such as the independence of the judiciary or the likelihood of property expropriation by the government, it is obviously best to utilize a single source. However, the WB employs all such measures in their composite index and is extremely transparent and clear with respect to which sources are employed for each country annually as well as the weights each measure is given.

B) ICRG

The PRS Group measures ROL as two separate concepts that are given equal weight - 'law' and 'order' - and gives each country a rating of 0-6 (higher scores equate with better ROL). The former concept is defined as "the strength and impartiality of the legal system", while the latter is "assessment of popular observance of the law". While both definitions are broad-based, the former seems to capture the treatment of individuals and firms in the legal judiciary system and whether there is discrimination of any sort in upholding the country's laws, while the latter measures the level of the actual crime rate in a country for a given year.

The PRS Group suffers from a lack of clarity and transparency in its ranking system. The researcher is very unclear about how the expert assessment arrived at a given score in a given year and this weakness consequently renders replicability impossible. Further, we do not know if the crime rate assessed is in regard to 'white collar' crime that might directly impact businesses or other crimes such as murder rates or robberies or a combination of both.

C) Fraser Institute

The Frasier Institute's ROL (called Legal Structure and Secure Property Rights) seeks to capture "the key ingredients of a legal system consistent with economic freedom (which) are rule of law, security of property rights, an independent judiciary, and an impartial court system." Thus their primary focus is how ROL best ensures market freedoms and the business environment in a given country, and not necessarily how the ROL relates to ordinary citizens per se. The construction of their ROL variable consists of seven equally weighted components, which they draw from three separate sources (in parentheses):

- Judicial Independence (*Global Competitiveness Report*)
- Impartial Courts (*Global Competitiveness Report*)
- Protection of Property Rights (*Global Competitiveness Report*)
- Military interference in rule of law and the political process (*ICRG*)
- Integrity of the legal system (*ICRG*)
- Legal enforcement of contracts (*WB's 'Doing Business'*)
- Regulatory restrictions on the sale of real property (*WB's 'Doing Business'*)

As in the World Bank data, the Frasier Institute is transparent about the sources, weighting and country-years in their data, which means that the researcher would be able

to obtain sub-components and shift the weights of the sub-components and/or replicate the data for any given year. Drawing multiple sub-components from three different sources can moreover be an advantage in that it increases external validity. However, one of the sources of the sub-components (ICRG) is not transparent in the assessment of the ROL. Further, the focus on how ROL impacts financial transactions and facilitates (or hinders) economic growth is somewhat narrow, in that there is little attention to ROL in the eyes of average citizens and how they view the impartiality of the legal system or the effectiveness of the police services and/or level of crime in the society. For example, the respondents of the first two sub-components (judicial independence and the impartiality of courts) are employees of international firms.

D) Freedom House

Similar to the PRS Group, the Freedom House measure is decided upon by a group of regional experts (all of whom are listed on their website). The rule of law measure captures four distinct concepts, which we can only assume are equally weighted in the final measure. They are:

- Independent judiciary
- The impartiality of the rule of law in court decisions and the level of civilian control of the police force
- Protections from political terror and war
- Level of impartiality of written laws with respect to ethnicity, religion and gender

Unlike the Fraser Institute's or ICRG measures of ROL, Freedom House focuses more on how the laws protect/affect individual citizens rather than directly impact firms. Their attention to protection of citizens from political terror and violence along with a focus on the impartiality of court outcomes and country laws that prevent discrimination make this measure distinct. However, the weaknesses that the ICRG measures suffer from – namely lack of transparency in the coding (though they are more transparent about the specific questions they ask their experts – see appendix) that makes replicability impossible – are shared by the Freedom House measures. Further, the four aspects of the rule of law contain several sub-questions and they are not explicit about the weighting of these data in the final measure. Thus, in terms of conceptual relevance, we find this measure strong, while, in transparency and replicability, it suffers several notable weaknesses⁵².

External Characteristics

Table 11 shows the number of observations and gives descriptive statistics for each of the aforementioned variables for EU member states from 1970-2008.

⁵² For a complete discussion of the FH rule of law measure, see: <http://www.freedomhouse.org>

Table 11: Empirical Coverage of Rule of Law Datasets

	World Bank	ICRG	Frasier	Freedom House
Number of MS	27	26	27	27
MS per year (mean)	27	22.5 ¹	22.1 ²	27
Time Period	1996-2008	1984-2008	1970-2004 ^a	2005-2008
Years per MS (mean)	10	20.8	9	4
MS-years (total)	270	563	243	108

^aFrasier Institute data from 1970-1999 is every 5 years, while 2000-2004 is annual data

¹Starting in 1999, the ICRG data contain 26 MS (no Luxembourg)

²Starting in 1995 the Frasier Institute data contain all 27 MS

With respect to the time period coverage, the Frasier Institute data has the largest time span (35 years); however, in terms of total annual coverage, the ICRG data have a clear advantage. Dating back to 1984, these data are published annually for between 17 and 26 EU (and future EU) member states, while the Frasier data stop at 2004. Freedom House coverage is clearly the most limited due to the lack of publication of the sub-components of its ‘political rights’ measure until 2005. Its data are therefore limited to only four years. The World Bank offers the advantage of publishing data on all 27 EU (and future EU) MS for each of the years that it has published the ‘governance indicators’, although the data are biannual from 1996-2002, which adds up to only ten years in total per MS.

We begin by looking at simple pairwise Pearson correlations for a mean, cross-sectional sample of all 27 EU member states during the years 2000-2008⁵³. A cross-sectional average is taken for the sake of visual simplicity and to allow a comparison between Freedom House and the Frasier Institute (no common years are available in these cases). Figure 2 shows how the four indicators of rule of law correspond to one another.

Not surprisingly, the WB and Frasier Institute ROL variables correlate extremely well, at .96 according to the Pearson correlation coefficient. This is largely due to the fact that two of the three sources used to build the Frasier Institute’s ROL measure (the ICRG ‘law and order’ sub-component and World Economic Forum’s ‘Global Competitiveness Report’) are also used in the World Bank measure. Thus, while the WB measure employs additional sources to the ROL data, these two are hardly independent

⁵³ See the Appendix for all visual pairwise Pearson correlations for each QoG pillar.

sources. Further, it is not surprising that the ICRG measure and the Frasier Institute correlate at .89 due to the use of the former in the latter's index. Moreover, the ICRG 'law and order' sub-component is also used as one of the sources in the World Bank ROL measure, although the correlation is slightly weaker because of the influence of anywhere between eight and 17 additional data sources used for each country as compared with only two other data sources by the Frasier Institute. Freedom House – while very limited in its time series availability – is the only source that appears independent from the other three. Even though this is the case, the correlations with the WB, ICRG and the Frasier Institute are still relatively high, at .86, .68 and .78, respectively.

QoG Concept #2: Corruption

Corruption clearly implies several difficulties for scholars with hopes of measuring the concept. First, the notion of illegality entails that actors that engage in it do so clandestinely. Second, there are many cultural differences across countries, such as traditions of 'gift giving', that might increase or decrease perceptions of survey respondents depending on how questions are phrased or how the concept of corruption is defined.

With respect to the concept of corruption in EU member states, we evaluate four contemporary indicators. The four we examine are from the World Bank, ICRG, Transparency International's CPI and Freedom House's 'Nations in Transit' data.

A) The World Bank

The World Bank data capture a very broad notion of the concept of corruption. They also provide the 'standard bearer' definition in the field of corruption studies most often cited by scholars, which is simply "the abuse of public power for personal gain". On the 'control of corruption' indicator they publish annually, they write for example that the indicator measures "*the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as 'capture' of the state by elites and private interests.*" The reason why the World Bank can claim such a wide score for corruption is of course the nature of the composite index. For corruption, they have up to 24 available underlying indicators at their disposal. The 27 EU MS range from five (Malta) to 17 (Romania) total sources for the latest year of the data, 2008. The average number of separate data sources used in the 27 MS in 2008 was 12.11, slightly less than the number of sources for the rule of law indicator.

Again, the underlying sources vary to a considerable degree in what they aim to capture. For instance, the *Business Environment and Enterprise Performance Survey* (BEEPS), which is undertaken in all former Eastern-bloc countries in the EU, asks firms "how common is it for firms to have to pay irregular additional payments to get things done?" and "how problematic is corruption for the growth of your business?" Another

firm survey included in the composite index, the World Economic Forum Global Competitiveness Survey (GCS), asks business leaders specific questions such as the frequency with which a firm might have to make extra payments in connection with trade permits, loan applications, taxes, and to obtain public contracts. On the other hand, the Bertelsmann Transformation Index collects data on the same sample of EU states as the BEEPS, but their data on corruption capture only the extent to which an anti-corruption agency has been established and how effective it is perceived to be in carrying out its mandate. The same goes for the Global Integrity Index (GII) survey incorporated in the corruption index. Differences in the underlying measures are further accentuated by household surveys, such as the Gallop World Opinion, which asks respondents broadly, “to what extent is corruption in government widespread?” Moreover, the IFAD Rural Sector Performance Assessments (IFD), an expert survey, focuses only on the transparency and accountability in rural areas in each country. Some surveys, such as the Institute for Management Development World Competitiveness Yearbook (WCY), are interested in the frequency of “bribing and corruption in the economy”, while others, such as World Bank Country Policy and Institutional Assessments (PIA), ask experts only about the perceived corruption in the public sector. Finally, the indicator receiving the largest weight according to Kaufmann et al. (2009) among all corruption data in the World Bank’s indicator is Freedom House, which focuses primarily on two areas of corruption in developing countries in the EU – the extent to which the government and market spheres overlap and the effectiveness that anti-corruption agencies, media and other whistleblowers in society exert over corrupt officials in government or the bureaucracy (see more details below).

Similar strengths and weaknesses clearly apply to the corruption indicator such as the World Bank’s measure of Rule of Law, in that one can replicate the composite index, check sensitivity by altering some of the weighting systems etc. Still, a good number of the underlying data sources are not transparent in how they produce the data and, as a result, full replication is not possible. Further, the concept which is captured by the WB is clearly very broad. Thus researchers seeking to focus on specific types of corruption in targeted areas of a country, such as customs or anti-corruption measures, should select a more specific measure.

B) ICRG

Again, the ICRG expert-guided measure shows that corruption within the political system, which may be damaging to the economic and financial environment, reduces the efficiency of government and business by enabling people to assume positions of power through patronage rather than ability and introduces an inherent instability in the political system. The types of corruption on which they focus clearly affect business the most, such as in “the form of demands for special payments and bribes connected with import and export licenses, exchange controls, tax assessments, police protection, or loans”.

However, experts are also asked to consider “actual or potential corruption in the form of excessive patronage, nepotism, job reservations, 'favor-for-favors', secret party funding, and suspiciously close ties between politics and business”.⁵⁴

As discussed in the preceding section, the PRS Group suffers from a lack of clarity and transparency regarding its ranking system. The researcher is very unclear as to how the expert assessment arrived at a given score in a given year, and this weakness renders replicability impossible. One clear strength of the ICRG measure for researchers with a specific interest in how corruption impacts business or investment is its relatively narrow focus - the measure clearly focuses on how corruption impacts international investors (the target sales audience) and therefore does not focus on anti-corruption measurements by the government or its impact on the lives of everyday citizens. Further, the concept of corruption is still rather vague – defined by using examples rather than a working definition.

C) Transparency International – Corruption Perception Index (CPI)

Like the World Bank indicators, the CPI is a composite index that utilizes up to 13 different business and expert surveys (a country must have at least three sources per year to be included) to construct the annual data. Unlike the World Bank indicator, the CPI focus on the concept of corruption is narrower. It is only concerned with ‘perceptions of corruption in the public sector’ and therefore does not include surveys or questions that concentrate on the level of corruption in the whole economy (private/public) or how such corruption impacts only business or the black market. The two primary themes of the questions included in the CPI are:

- 1. What is the perceived level of corruption in the public sector or, put another way, the ‘misuse of public office for private gain’?*
- 2. How serious is the government in combating corruption – i.e. are there anti-corruption measures in the country and how effective are they? Do convicted offenders receive appropriate punishment, etc.?*

Of the ten different sources from which the surveys are drawn in the past two years, three use current and past international business leaders as respondents, while the other seven are country expert surveys⁵⁵. There are therefore no household surveys or opinions from everyday citizens that contribute to the CPI, as is the case with the World Bank indicator.

Another difference is that the underlying sources of the CPI are not weighted. Whereas the World Bank weights the more ‘deviant’ cases less than those surveys that tend to correlate more strongly with one another, the CPI weights all surveys equally in their annual data. Thus it might be interesting to see whether certain analyses would hold

⁵⁴ http://www.prsgroup.com/ICRG_Methodology.aspx

⁵⁵ See http://www.transparency.org/policy_research/surveys_indices/cpi/2009/methodology for further details on the specific sources.

up to sensitivity tests given that CPI employed similar weights as the World Bank or vice versa. In addition, the CPI country scores themselves are not meant to be compared over time in assessing improvement with corruption in an individual country. Instead, the measure is a year-to-year ‘snapshot’ that ranks countries relative to one another and provides a confidence interval for valid cross-sectional comparisons. Thus, when the CPI adds new countries (or drops them due to a lack of three sources), the standing of any given country can move up or down irrespective of a change in its own individual score. Regarding the sources and availability for the MS, all 27 EU countries are included in the most recent CPI with an average of 6.3 sources per MS, with Cyprus having only three and a number of new MS in Eastern and Central Europe having eight sources.

D) Freedom House

The Freedom House’s *Nations in Transit* has focused on 29 ex-communist states in Eastern and Central Europe along with ex-Soviet states, with data going back to 1995. Their basic operationalization of corruption seeks to capture:

“Public perceptions of corruption, the business interests of top policy makers, laws on financial disclosure and conflict of interest, and the efficacy of anticorruption initiatives”

Similar to ICRG, this is an expert-driven survey, yet, unlike ICRG, Freedom House is an NGO and does not cater to international firms. It poses ten specific questions to country experts when assessing the level of corruption in each country⁵⁶. This indicator, based on these ten questions, specifically taps into the overlap of public officials in the private market along with the effectiveness of anti-corruption channels to detect malfeasance. While impressive in its conceptual aim, it only covers the so-called ‘transitioning states’ in the EU, consequently completely leaving out the EU-15 countries for example.

External Characteristics

Table 12: Empirical Coverage of Corruption Datasets

	World Bank	ICRG	CPI	Freedom House
Number of MS	27	27	27	27
MS per year (mean)	27	23.5	24	9
Time period	1996-2008	1984-2008	1995-2008	1999-2008
Years per MS (mean)	10	21.7	12.4	10
MS years (total)	270	587	336	90

⁵⁶ See appendix for the complete list of the 10 questions.

The four datasets vary in temporal coverage as well as number of countries. We can almost immediately drop the Freedom House indicator from the discussion since it only covers those countries that are ‘in transition’, and not the EU-15 countries for example. The World Bank measure is again impressive in that it covers every MS for each year that it is published, for a total of 270 observations from 1996-2008. The CPI has relatively similar coverage, yet it has more overall observations because it was published annually from the mid-1990s as opposed to bi-annually. The ICRG measure is clearly the most comprehensive with respect to time series, going back to 1984 for the majority of countries.

We move on to examine how closely these four measures correlate with one another, taking a cross-section average of EU countries from 2000-2008. While they vary with respect to the scope of the coverage of the concept that they seek to measure, they all intend to capture some aspect(s) of corruption; thus we expect relatively high correlations. Indeed we find this to be the case. The World Bank data for example show remarkably high Pearson correlation coefficients with the other three indicators – both ICRG and CPI correlate over .93 with the WB data, while the Freedom House data correlate at .85. The CPI not surprisingly has very similar correlations with the Freedom House and ICRG data, as the WB measure does. The weakest overall correlation is between the Freedom House data and the ICRG measure, 0.74. Overall, it seems that there is general consensus among these four leading indicators regarding the measurement of the concept of corruption.

Concept 3: Quality of the Bureaucracy

While its definition is somewhat elusive according to a number of public administration scholars (Cameron and Whetten 1983; Seldon and Sowa 2004), there is no doubt that the creation of a quality bureaucracy and civil service is of the utmost importance for a country’s economic growth, delivery of essential public services and creation of legitimacy and trust among citizens. We take from the literature that this concept intends to capture the effectiveness or impartiality of the civil sector. Do services favour certain groups in society or do they mainly benefit everyone? Are bureaucrats selected for civil service on their merits or are there political considerations in the hiring process? Are their decisions in carrying out public policy independent, or influenced by political interests? These are some of the ideas we take into consideration in this QoG pillar. We take a closer look at three sets of data that assess the quality of the bureaucracy in EU countries (among others) – the World Bank, the ICRG and the World Values survey.

A) World Bank

On the Government Effectiveness (GE) variable, Kaufmann et al. (2009) write “capturing perceptions of the quality of public services, the quality of the civil service

and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann, Kraay & Mastruzzi 2009: 6).

The WB employs 18 different underlying indicators with which they build their GE measure, of which 12 are used for various EU states. Bulgaria and Romania for example have 12 sources, while Malta has just five. The majority of countries have roughly nine sources each in the construction of the variable.

As with the previous two pillars of QoG, the WB employs a broad definition of the concept here. This is reflected in the fairly divergent survey questions from the various expert and citizen surveys that underlie the overall data. For example, the three questions that are aggregated into Gallup World Poll's contribution to the overall GE index focus exclusively on citizens' satisfaction with public education, public transit and with roads and highways. In the World Economic Forum's (GCS) survey, leaders of firms respond to questions about a country's infrastructure, along with how much time business leaders interact with government officials in the civil service. The Economist Intelligence Unit (EIU) asks experts to assess the amount of bureaucratic excess (red tape) and the 'institutional effectiveness' of a country's civil service. Another variable included in the GE measure, the Global E-Governance Index (EGV), focuses exclusively on a government's quality of online services – including such features as the amount of documents and publications, forms, languages, security features, ease of payment for fees and taxes etc. that a government publishes online for their citizens. Finally, the Business Environment Risk Intelligence (BRI) asks experts only to assess the amount of delay time for businesses when dealing with government bureaucracy.

Clearly there are many different conceptual foci among just these five variables discussed here, in addition to others included in the index. Further, some focus on outputs (effectiveness and quality of service) while others measure inputs (impartiality in recruiting for the civil service). However, only an analysis of the individual indicators themselves will tell us whether they indeed capture the same underlying concept.

B) ICRG

The ICRG data include a measure of 'bureaucratic quality' in the overall measure of political risk, which it provides to international firms seeking to invest in foreign countries. They tend to see this measure as capturing the independence of the civil service from the government. They write "The institutional strength and quality of the bureaucracy is another shock absorber that tends to minimize revisions of policy when governments change. Therefore, high points are given to countries where the bureaucracy has the strength and expertise to govern without drastic changes in policy or interruptions in government services."⁵⁷ Thus they focus on the stability of a civil service along with

⁵⁷ Quote obtained from the Political Risk Services website:
http://www.prsgroup.com/ICRG_Methodology.aspx

features that we would call impartiality – which decisions are made on a basis of stable, consistent policy, rather than by the whims of year-to-year shifts in the political control of a given country.

Again, experts determine the scores presumably on the basis of a good deal of personal experience with the particular country that they are assessing. Accordingly, the results are not transparent or replicable in that they do not publish a systematic set of criteria that all experts must follow in order to come to numeric conclusions for each country.

C) World Value Survey – Trust in the Civil Service

The World Value Surveys (WVS) are undertaken about every five years and try to gauge levels of, and changes in, attitudes and values in over 80 countries worldwide. They focus on the feelings and perceptions of everyday citizens rather than of experts or firms. In regard to this particular pillar in QoG, the WVS has asked respondents to rate their confidence in a number of areas, such as the armed service, the press, the courts and – for the interest of this section – the civil service. The question they pose to respondents is taken from the latest round of surveys (2005):

“Could you tell me how much confidence you have in (the civil service): is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?”

Thus respondents are asked to rate their perception of their civil service in terms of their confidence level, which is coded 1-4. The data are available for all EU states for at least two years between 1980 and 2005 (except the U.K.).

External Characteristics

While all three indicators report data for the entire group of EU 27 countries, only the World Bank and ICRG data could be used in a time series-type analysis. The WVS, while important to consider because of its emphasis on citizen opinions, simply does not cover enough time. Further, the years in which data points are reported vary from time period to time period; for example, the Netherlands has data for 1981, 2000 and 2007, while Austria has data for only 1990 and 1999. As with the corruption and rule of law indicators, the WB data has each EU member state for every year the data is published (ten years in total), while the ICRG data cover almost all EU countries going back to 1984.

Table 13: Empirical Coverage of Bureaucratic Quality datasets

	World Bank	ICRG	WVS
Number of MS	27	27	27
MS per year (mean)	27	23.4	2.5
Time period	1996-2008	1984-2008	1980-2008
Years per MS (mean)	10	21.67	2.7
MS years (total)	270	585	71

The Pearson correlation between the ICRG and WB data is strong, at .92 – but this is not surprising since the WB incorporates the ICRG data as one of its underlying indicators. However, while the WVS data report a positive correlation with both the ICRG and WB, the correlations are significantly weaker - .60 and .64 respectively. We take this to possibly mean that there might be a gap between public opinion and that of the so-called experts or firms, which are interested in evaluating bureaucratic effectiveness for international investors and that of the public with respect to this QoG concept.

Concept 4: Democracy and Strength of Electoral Institutions

Unlike the relatively recent interest on the part of scholars and policy-makers in conceptualizing and quantifying the previous three pillars of QoG, the concept of democracy is a longstanding one in most all modern social sciences. Among various conceptualizations of democracy, both ‘thick’ and ‘thin’ versions can be found. The latter tend to dominate the large-N quantitative studies because of easier generalizability and operationalization among many countries, while the latter are generally found in small-N cases studies, in which scholars can delve deeply into the specifics of one or two countries’ institutions. Most of the ‘thinner’, more generalizable versions of democracy tend to follow the Schumpeterian definition, which focuses on electoral institutions and republican representation: "The democratic method is that institutional arrangement for arriving at political decisions which realizes the common good by making the people itself decide issues through the election of individuals who are to assemble in order to carry out its will" (Schumpeter 1976: 252). In this section, we take a closer look at the criteria used by the five sets of data measuring democracy.

A) Freedom House

The Freedom House measure of ‘political rights’ (formerly the Gastil index) has provided researchers with empirical data on democracies for intra-state comparisons since 1972.

Each country is assigned a numeric rating from 1-7, with lower numbers indicating better democratic performance. Each region of the world is assessed by area studies experts, Freedom House staff and country analysts. The FH measure can certainly be described (at least relative to other measures) as a ‘deeper’ or ‘wider’ measure than what comes from other sources. For the political rights indicator, they pose ten different questions (ranging from 0-4 each) to the experts on a country’s:

- Electoral Process (3 questions) – including regular elections, competitive multipartism, universal adult suffrage and open campaign channels for all competitive parties.
- Political Pluralism and Participation (4) – including freedom to join political parties, run for office, freedom from the military or foreign powers, rights for ethnic and religious minorities and a realistic chance for the opposition to win a fair election.
- Functioning of Government (3) – including whether the freely elected head of government makes policy, whether the media is free, and whether there is impartiality and transparency between elections.

Those countries that receive higher ratings in this data tend to exhibit ‘free and fair elections’ and ‘competitive parties’, where the opposition has an important role in public policy and debate, and minority groups have reasonable access to government and are allowed channels for bargaining with the majority and other social groups.

B) Polity IV

The polity score of the type of regime (Marshall & Jaggers 2009) captures the extent to which a regime is either authoritarian or democratic using a spectrum on a 21-point scale ranging from -10 (hereditary monarchy) to +10 (consolidated democracy). Each country in the world (in some cases going back to 1800) is thus given a score that is based on six dimensions. Of these, there are three essential elements that these data try to capture⁵⁸:

- Executive recruitment – including competitiveness and openness.
- Constraints on executive authority - “the existence of institutionalized constraints on the exercise of power by the executive”.
- Political competition - “the presence of institutions and procedures through which citizens can express effective preferences about alternative policies and leaders”.

Based again on expert assessment that cannot be replicated perfectly by outside scholars, the polity team assigns what can be considered a fairly ‘thin’ version of democracy to countries. What they consider a ‘mature’ democracy is: “(an) internally coherent democracy, for example, that might be operationally defined as one in which (a) political

⁵⁸ For a more thorough reading of the Polity data, see:
<http://www.systemicpeace.org/inscr/p4manualv2007.pdf>

participation is unrestricted, open and fully competitive; (b) executive recruitment is elective; and (c) constraints on the chief executive are substantial” (Marshall and Jagger 2009: 14).

For example the mean score for the EU countries in 2007 was 9.48 for 25 countries (Luxembourg and Malta are not included), and the majority of countries were ranked as perfect democracies. Moreover, countries like Switzerland prior to universal suffrage in 1971 received a perfect 10 in the polity data, while South Africa during apartheid had a rather strong rating of 7. Thus its overly thin conceptualization of the concept and the lack of variation in this indicator in this specific sample might be problematic in its use.

C) The World Bank – ‘Voice and Accountability’

The WB data again bring a very broad conceptual framework of the concept of democracy, in what it calls ‘voice & accountability’. Kaufman et al. describe this pillar of QoG as “capturing perceptions of the extent to which a country’s citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media” (Kaufmann et al. 2009: 6).

As with the other QoG indices that the WB provides, the broad conception of this variable is based on the numerous and divergent underlying indicators they employ to build to data. For EU states, Romania and Bulgaria contain at least 14 sources each in recent years, while Malta has six. The majority of EU countries have about ten to 12 sources each in the latest year of the data (2008).

Included in the data are several different conceptual foci for what the measure captures. For example, some indicators employ a broad version of VA. The Gray Area Dynamics (GAD) expert assessment measures religious freedom, the military’s involvement in politics, political patronage and the role of the opposition – a fairly broad framework. The EIU uses five criteria to rate each country based on human rights, accountability of political figures, ‘vested interests’, freedom of association and ‘democracy index’. On the other hand, Gallup World (GWP) and the Reporters without Borders index (RSF) data used in the index have a more narrow focus – the former asking respondents about the fairness/freeness of elections and the absence of political violence, while the latter focusing exclusively on freedom of the press from the government. It should also be noted that the WB data incorporate the Freedom House data into their composite index.

A clear advantage of the WB data, which has been discussed in earlier sections, is transparency and replicability. Furthermore, if there are conceptual problems that a researcher has with any of the underlying indicators used to build the index, the data can be taken out after the sample is replicated for any given year.

D) Vanhanan

The data commonly known as the Vanhanen Index was a collaborative effort by Tatu Vanhanen and the International Peace Research Institute, Oslo (PRIO). Of the four so far, it is by and large the most easily replicable and transparent. However, it is considered by many to be a very ‘thin’ measure of democracy in its conceptual framework, focusing only on two primary components:

- Competition
- Participation

The first component is measured essentially by the electoral success of smaller and/or opposition parties in a political system during an election. Specifically, it is calculated by subtracting 100 from the winning party in a parliamentary or presidential election, with higher numbers presumably equating to greater levels of competition (this means the party that won received less overall support). The second measure is quantified by simply taking the total percentage of the eligible population that participated in the last election, with higher numbers meaning better democracy.

Clearly there are problems with these two sub-indicators, despite the advantages of transparency and replicability. For one, the right ‘not to vote’ is an essential part of most people’s freedom and can be used as a tool to show dissent from the government’s policies, not necessarily that a country is less democratic in principle. Another is that, despite the attempt to capture intra-party competition, there is no way of telling from these data whether the party in power has ever been voted out of office, as there is no indicator of government shifts.

E) Freedom House/Polity

The Freedom House/ Polity indicator was created by Hadenius and Teorell (2005), whereby the two measures are combined into a single index, ranging from 0-10, with higher scores equating to higher levels of democracy. In an empirical test by Hadenius and Teorell (2005) they show that the combined data are more reliable than their constituent parts and perform better on tests of external validity. We include this measure in order to compare its external performance with the other four indicators.

External Evaluation

We look at the coverage of five measures commonly used in the academic and policy literature that seek to capture the strength of democratic institutions. We count only data going back to 1980, although several of these datasets go back significantly further in time. The WB measure is certainly that which has the least coverage due to its relatively late entrance into the market, yet again, covers all 27 states for each year of publication. The FH measure is comprehensive but misses some EU states in the earlier part of the time series going back to 1980. Polity is the same but misses a few more states at times. The FH/Polity measure shares the same coverage as the FH measure.

Table 14: Empirical Coverage of Democracy Datasets

	World Bank	Polity	Freedom House	Vanhanen	FH/Polity
Number of MS	27	27	27	27	27
MS per year (mean)	27	21.3	23.6	23.8	23.6
Time period	1996-2008	1980-2008	1980-2008	1980-2004	1980-2008
Years per MS (mean)	9	22.3	25.3	22.1	25.3
MS years (total)	270	618	683	595	683

Polity time series is for some states 1800-2008, and most from 1946-2008

Freedom House times series goes back to 1972

Vanhanen data time series begins at 1946 for certain states

The correlations among the indicators seem to vary the most for the concept of democracy. Here we see correlations as low as .40 (between FH and Polity) as well as .43 and .46 – Vanhanen with FH and Polity, respectively. Not surprisingly, the FH/Polity measure has the strongest correlations with the other four indicators, while the WB data seems to match up relatively strongly as well according to the Pearson correlation coefficients. Notably, the Vanhanen measure does not correlate with any of the other indicators above .65, which suggests that they are in effect measuring quite different underlying concepts, considering the Vanhanen measure’s ‘thinner’ definition of democracy. The individual measures of FH and Polity display very little variation among the EU countries, with most states receiving the highest score possible in both measures throughout the last decade. The WB data, despite their relatively small amount of time, display an impressive amount of variation among the countries according to the scatter plots, which could be very useful in comparing and ranking the EU countries along this dimension.

1b.4 Conclusions

The data and sources that measure the multiple concepts of QoG began mainly in the mid-1990s and have expanded at an impressive rate since then with respect to coverage and the number of sources. As we noted earlier, we seek to take advantage of the existing measures of QoG – not necessarily to ‘reinvent the wheel’ so to speak – rather than presume that we could create something better in the time period allotted for this study. Upon thorough review, we find that the World Bank data would be the most appropriate source on which to build an indicator of QoG for EU countries. First, it covers all four pillars of QoG. Second, it covers all EU countries for at least ten years, going back to the mid 1990s, and is now published annually. Third, it is transparent in the way that it is constructed – publishing freely all underlying indicators on which it is built and giving a clear description of the conceptual meaning of each concept and the methodology used to create each variable. Fourth, the theoretical scope of each QoG concept is wide rather than narrow. We believe that, unless specified, all aspects of corruption, rule of law etc. should be included rather than focusing on narrow aspects alone. This allows for more information to be included, which is good for example for reliability checks of the data.

In the next section, we analyze the 2008 data for the four pillars of QoG aimed to capture the concepts we have discussed in this section – bureaucratic quality, corruption, rule of law and strength of democratic institutions. We check the data for reliability, robustness and internal consistency using variations of their methods to construct each variable.

Part II: Evaluation of the World Bank Data

2.1 Overview of the Data and the Rankings for EU Member States

Having described the four pillars of QoG for the World Bank data in some detail in the previous section, we only briefly discuss the data here. The World Bank Governance Indicators (Kaufmann, Kraay and Mastruzzi 2009) have been published freely online for scholars and practitioners to use since 1996⁵⁹. They were available bi-annually up to 2002 and the data have been published annually from 2002-2008. The data are a composite index that contains six pillars of governance:

1. *Control of Corruption*
2. *Rule of Law*
3. *Government Effectiveness*
4. *Voice & Accountability*
5. *Political Stability & the Absence of Violence*
6. *Regulatory Quality*

Each aggregate QoG pillar combines survey data from firms, experts and citizens from countries worldwide. The underlying data are collected from four main types of sources: non-governmental organizations (NGOs), firm and citizen survey data, public sector agencies and commercial interest/risk assessment providers for businesses.

Following the literature we employ the first four pillars to build a single ranking for EU member states. We have taken all underlying indicators for each country for the 2008 (most recent) data. The following sections seek to address the following questions:

1. Are the data (along with the rankings derived from it) internally consistent from a statistical point of view?
2. How sensitive are the data - do alterations in the assumptions used to build these pillars of QoG impact the initial results?

We employ a number of statistical techniques to address these important questions in the following sections. In particular, we follow the advice of JRC and OECD *Handbook on Constructing Composite Indicators*⁶⁰. We begin with the following rankings in the four pillars according to the 2008 World Bank data along with a total combined ranking⁶¹.

⁵⁹ Data can be found at: <http://info.worldbank.org/governance/wgi/index.asp>

⁶⁰ We are also grateful to Michaela Saisana and Stefano Tarantola for their help and advice on conducting sensitivity evaluations of the data.

⁶¹ When aggregating the pillars, we first conducted a principle component, factor analysis (FA) and weighted each one by their respective factor weights. The full rankings can be found in the appendix or on page 22 in the executive summary.

The rest of this analysis proceeds as follows: we first analyze the internal consistency of the data to test whether it is internally consistent from a statistical point of view. We test whether the underlying data in the four pillars of QoG are in fact consistent with one another, as well as all underlying data sources with the entire QoG index. In section 3, we test whether the pillars are consistent with the total QoG index along with the impact of an exclusion of each pillar from the index and two different weighting schemes in combining the pillars into the single QoG index. In section 4, we make a cluster analysis to place EU countries in appropriate groups, from which more realistic policy goals can be set. In section 5, we make a multi-modelled uncertainty and sensitivity analysis to see if the rankings derived from the index are significantly altered or if they remain robust to changes. More specifically, we challenge some of the initial assumptions made by the creators of the data with respect to aggregation and weighting schemes. We also test the impact of each individual indicator by removing it from each pillar one by one. In section 6, we show some evidence that the data are externally valid by showing bivariate relationships with five additional variables commonly correlated with QoG indicators from previous empirical studies. We make several concluding remarks in section 7.

2.2 Internal Consistency of the Data

The index built in this analysis relies heavily on the existing framework that researchers at the World Bank have published over the past 15 years. In addition, we built an index out of the four existing pillars of governance, combining them into a single composite indicator, which we can then use to devise a ranking of EU MSs and their relative QoG.

In combining the four pillars into a single index, we are undoubtedly simplifying a number of underlying issues that need to be discussed – namely, whether the underlying data support the notion of four separate and distinct pillars of QoG. Furthermore, is each of the individual pillars itself internally consistent?

To test this question, we use a method commonly employed in the literature – namely Principle Component Analysis (PCA). The World Bank classifies each of the pillars of governance independently from one another and does not specify any ‘sub-pillars’ that exist in building such indices as ‘rule of law’ or ‘control of corruption’ for example. The internal consistency checks in the data should demonstrate that this is the case.

We begin by testing the internal consistency of each of the four pillars individually. As noted, the World Bank data does not alert users of the data to any ‘sub-pillars’ – e.g. significant clusters in the data that might need to be accounted for in the weighting scheme – and we thus expect that each of the four measures of QoG will contain only one significant factor. Because of the need for common data sources, we were forced to drop the two to three sources that are region-specific to the former Eastern group and rely on

the sources that included all MSs⁶². Although the number of sources varies from country to country in the dataset covering all countries in the world (for example, some small island states have only one source for a given pillar while some states have more than 15) we are fortunate in that for this particular sub-set of countries – the EU – there are at least nine common sources for each pillar of QoG. Table 15 shows a PCA for each of the four pillars of QoG individually.

Table 15: Principle Component Analysis - Check of the Underlying Data Consistency

Pillar	Factor	Eigenvalue	Proportion
<i>Government Effectiveness</i>	1	5.41	0.61
<i>Control of Corruption</i>	1	6.96	0.78
<i>Rule of Law</i>	1	7.52	0.63
	2	1.41	0.12
<i>Voice & Accountability</i>	1	5.98	0.61
	2	1.17	0.12
QoG Index (w/ 4 pillars)	1	3.59	91.5

Note: see the appendix for the full analyses of each pillar

Government Effectiveness

- Using the standard rules of the Kaiser criteria, factors with Eigenvalues greater than ‘1’ that explain 10% or more of the variance, we find only one significant factor in the nine underlying indicators that make up this pillar. The single factor has an Eigenvalue of 5.41 and explains over 60% of the variance of the data.
- The next largest factor was close to the threshold established by the Kaiser criteria – explaining over 8% of the variance with an Eigenvalue of 0.9.
- This demonstrates strong internal consistency in the data, which corroborates the World Bank’s decision to condense this data into a single index. When aggregating the underlying data to form this pillar – as well as using GE to generate the total QoG composite index – we can be fairly certain that significant clusters in the underlying data do not exist.

Control of Corruption

⁶² We took only sources that included 24 or more MSs; thus at times we had to impute missing data, in particular for Malta and Luxembourg. We did so with the simple method of ‘mean substitution’.

- Only one significant factor was found in the data (nine underlying sources). The factor has an Eigenvalue of .69 and itself explains over three-fourths of the total variance.
- The next largest factor had an Eigenvalue under .7 and explained less than 6% of the variation.
- The results of the PCA point to a very internally consistent and related set of underlying indicators.

Rule of Law

- The Rule of Law pillar exhibited two significant factors in the underlying data. The first, which explains over 60% of the variance itself and has an Eigenvalue of 7.52, is clearly the most important factor in the data. Factor 2 has an Eigenvalue of 1.41 and explains 12% of the variance. Together, the two factors explain about three-fourths of the total variance in the data.
- After rotating the factor (Verimax method), we find that nine of the 12 indicators load onto factor 1, while Global Risk Services, Gray Area Dynamics and the 'U.S. State Department Trafficking in People Report' (DRI, GAD and TPR respectively) load onto factor 2.
- The latter three factors share a common thread in that they focus heavily on human trafficking and organized crime, rather than judicial independence or property rights, as many of the other indicators do.
- This suggests the need of testing the sensitivity of the RL indicator using two separately weighted factors instead of the single index. However, this may not be a serious problem in that factor 1 in and of itself has a relatively high Eigenvalue and explains over .6 of the proportion of variance – with all of the underlying variables positively loading on it with at least a weight of 0.3, suggesting a certain degree of internal consistency among all 12 indicators.

Voice & Accountability

- Two significant factors were also found in the VA pillar. Again, the first explains about 60% of the total variance and has an Eigenvalue of about 6, while the second factor is significant according to the Kaiser rule in that it exceeds the Eigenvalue threshold of 1 and explains over 10% of the total variance.
- After rotating the factor (Verimax method), we find that eight of the ten indicators load onto factor 1, while the other two (RSF & HUM) load onto factor 2.
- Although there is no direct, common theoretical thread in the underlying make-up of the two indicators in factor 2 (RSF asks exclusively about press freedom, while HUM deals with broader human rights issues) they are the underlying variables with the least amount of variance for the MSs – most of the EU 27 rank very high on both of these measures.

- This suggests the need of testing the sensitivity of the VA indicator using two separately weighted factors instead of the single index. Although the two factors were found, again, as with the RoL, all ten indicators are positively loaded onto factor 1 with weights of at least .35. Thus internal consistency of the data is present to a certain degree.

QoG Index

- The final line in Table 1 shows the PCA for the four pillars underlying the total QoG index. Here we find that the four components of QoG factor vary together in a single group, with the single factor explaining almost 92% of the total variance of the final index.
- The PCA reveals slightly different impacts among the four pillars, suggesting that slightly adjusted weights are more appropriate than a simple equal weighting scheme when GE, CC, RL and VA are combined into a single index. Based on the squared factor loadings of the principle components, the weights would be 0.27 for both GE and CC, 0.26 for VA and 0.20 for RL.

We move next to a test of the internal consistency of the framework of the total QoG index for the EU MSs. For this, we undertake a simple correlation test between the QoG index and each of its pillars and underlying indicators. If each of the indicators and pillars and indicators correlate positively with the QoG index, then we can be relatively certain that the underlying data are internally consistent in that they point to the same underlying concept.

**Table 16: Pearson's Correlation Coefficient Test:
QoG Index with 4 Pillars**

	GE	CC	RL	VA
QoG (total)	0.98	0.97	0.88	0.96
GE				
CC	0.96			
RL	0.84	0.79		
VA	0.95	0.95	0.8	

Note: all coefficients significant at 99% level ($p < 0.01$, $n = 27$)

Since the World Bank has re-coded all underlying indicators so that lower values equate to lower governance and higher values to higher levels of governance, we would certainly expect each of these pillars to be positively correlated. All four pillars of QoG point in the same direction as the total measure, which is a desirable aspect in any composite indicator. All of the Pearson correlation coefficients are statistically significant at the 99% level of confidence and range between .79 and .98.

We then explore the relationship between the underlying data and their respective pillar along with each indicator's bivariate relationship to the QoG index⁶³. To demonstrate 100% internal consistency of the underlying variables, each indicator would have a positive and significant relationship with both its pillar and the overall index. In fact, this is nearly the case. Of the 59 separate underlying indicators in the total index, only four are in the opposite direction with both their respective pillars and the total QoG index. However, their relationship appears to be random (indistinguishable from zero) as their p-values do not reach even the 90% level of confidence. These variables come from two sources, which are:

- BEEPS GE and CC scores (BPS), the Global Integrity Index for CC and RL (GII)

We believe that this is less of a problem than it might appear in that they only pertain to the sub-set of former Eastern-bloc states and are thus weighted less in the original data than are the common sources. Furthermore, weighting by factor analysis (FA) at the pillar level will also drop these sources, adding to a stronger internal consistency of the remaining underlying data.

Of the remaining data, five other indicators appear to be randomly associated with the QoG index and six sources are insignificantly related to their respective pillars. These data come from the following sources:

- The Bartelsman Index for GE (BTI), the BEEPS for RL (BPS), the Institutional Profile Database for RL (IPD), the Global Integrity Index for VA (GII) and the Media Sustainability Index for VA (MSI). The MSI is understandable, as it only has two observations. In addition, the Gray Area Dynamics measure for RL (GAD) is insignificantly related to the RL pillar.

This means that, of 59 total indicators underlying the QoG index, 50 are significant and correlated in the expected direction or, in other words, roughly 85%⁶⁴. Some practitioners would argue that the designers of the data should remove these variables with weak-positive or negative correlations to their pillar or overall index. This is due to the situation of when the underlying data are weighted by how strongly the data correlate together, an improvement in these non-significant variables would not result in a significant improvement in the overall ranking index, all things being equal. Moreover, cutting down the overall number of underlying indicators (in the EU sub-set, in any case) would not imply a significant change in the assessment of QoG. However, this does not mean that these nine indicators do not contain important information about a country's level of governance, but only that they are not consistent internally with the other measures of QoG.

⁶³ Full table with all correlations is found in the appendix.

⁶⁴ For a full list of the correlations by pillar, see the appendix.

2.3 Impact of the Four Pillars on the Overall Index of QoG

We also check the internal consistency of the measure of QoG for the EU MSs by checking the impact of the individual pillars on the total QoG scores for each country and the overall rankings themselves. For instance – how sensitive is the index to the exclusion of one of the four pillars? Are the rankings sensitive to changes in the weighting scheme?

The weight of each of the four pillars corresponds to the variance of the total index score that can be attributed to that particular pillar⁶⁵. Table 17 demonstrates nine simulations that we ran on the data which show the impact of the elimination of one of the pillars on the total QoG index along with the effect of two different weighting schemes for the pillars – equal weighting and weighting assigned by factor analysis (FA). We remove each pillar while weighting the other three by their factor weights and then weighting them equally. We show the median shift in the group, meaning that if at least 14 MSs remain in the same place as the original rank order, the median score is ‘0’. The max shift indicates the state that shifted the most drastically due to the underlying changes in the index (we label the state(s) that did so along with their direction) and, finally, we include the Spearman correlation coefficient, which compares how closely the original ranks correlate with the adjusted ranks – the closer to ‘1’, the more similar the adjusted rank order is to the original.

Table 17: The Effects of Weighting Scheme and Exclusion of 1 Pillar on QoG Ranks

Excluded Pillar	Weights	Median	Max Shift	Max MS	Spearman
none	1 EW	0	1	Cz Rep.(-), Lith.(+)	0.994
GE	2 FA	0	2	Pol(+), Irl(+)	0.993
GE	3 EW	0	2	Irl(+), Cz. Rep.(-)	0.995
CC	4 FA	0	2	Lux(-), Lith(+)	0.992
CC	5 EW	1	6	Lith(+)	0.981
RL	6 FA	0	4	Lux(+), Lith(-)	0.982
RL	7 EW	0	4	Lux(+), Lith(-)	0.983
VA	8 FA	0	2	Cyprus(+)	0.996
VA	9 EW	0	1	9 MS shift 1 place	0.998

The index remains remarkably stable irrespective of whether the pillars are ranked by FA or equally, demonstrating that no one pillar is dominant in deciding the overall index scores. The Spearman rank correlation coefficient never scores below .98 in any of the nine simulations. In addition, the absence of any one individual pillar does not significantly alter the rank order at all. Furthermore, it is worth noting that the top four

⁶⁵ A table of the factor weights by pillar and a complete formula can be found in the appendix.

states (Denmark, Sweden, Finland and the Netherlands) along with the bottom three (Italy, Bulgaria and Romania) do not shift positions at all in any of the nine simulations. In fact, in only one case, number 5 (dropping corruption and weighting equally), does the median shift equal 1 – implying that at least slightly more than half of the MSs move at least one spot. In this case we also observe the largest one-state shift – Lithuania moves up six places from 19 to 13. Judging from the list of states that exhibit more extreme volatility than other MSs due to the exclusion of pillars or weighting adjustment, there are two that stand out – namely Lithuania and Luxembourg. In the case of Lithuania, it is clearly most hurt in the rankings by its corruption score, while Luxembourg sees a jump of four spots when rule of law is excluded.

2.4 Cluster Analysis

Figure 10 – EU Member States by Cluster (lighter shades equal higher QoG)



Note: Malta (not seen) is grey (same as countries such as France and Spain)

What can we learn from cluster groupings? Although several EU states demonstrate similar QoG scores, there might be patterns in the underlying data that distinguish countries – suggesting different challenges despite sharing a similar overall score. As noted by Kaufman and others at the World Bank, the indicators of governance are a tool that scholars and practitioners can use to rate the relative position of countries (and not absolutely expect to rank countries with pinpoint accuracy). For example, while finding

any existing measure of QoG that can reliably distinguish between Sweden and Denmark or Romania and Bulgaria would be an admittedly impossible task, we can use existing measures to accurately point out the relative standing to other groups of countries. With the fact that almost all governance data is subjective in nature and thus an estimate of the true value of governance rather than representing the ‘real value’, we apply the cluster analysis to elucidate peer countries that are similarly ranked in the four pillars of QoG, as well as the total index itself. The cluster groupings can serve as a helpful tool to identify EU member states that share common challenges to building QoG at the national level.

We employed hierarchical clustering using Ward’s method and squared Euclidian distancing for the four pillars of QoG to identify the number of appropriate cluster groupings. Three distinct groups were detected in the analysis. We then used k-means clustering with squared Euclidian distancing to assign each country to a cluster⁶⁶. Although there are geopolitical and historical similarities among the groups, they are not entirely driven by such factors. For example, new MSs of the former Eastern bloc, Estonia and Slovenia, are grouped with EU-15 MSs such as Spain, Belgium and France. Other EU-15 MSs, Italy and Greece, belong to the third cluster grouping. However, the top nine performers with respect to QoG (Group 1) are all EU-15 states from the northern and central (Germany, Luxembourg and Austria) parts of Europe. Table 18 shows the distribution of states into the three groups, while Figure two lists them in groups by the rank of the total QoG scores.

Table 18 : Clusters of EU Countries (ranking order) Based on 4 Areas of QoG

Cluster 1 n=9	Cluster 2 n=8	Cluster 3 n=10
DENMARK	FRANCE	CZECH REPUBLIC
SWEDEN	BELGIUM	LITHUANIA
FINLAND	MALTA	HUNGARY
NETHERLANDS	SPAIN	SLOVAKIA
AUSTRIA	PORTUGAL	POLAND
GERMANY	CYPRUS	LATVIA
UNITED KINGDOM	ESTONIA	GREECE
IRELAND	SLOVENIA	ITALY
LUXEMBOURG		BULGARIA
		ROMANIA

⁶⁶ In addition, we clustered countries with a ‘maximum between cluster difference’ method and found that only France and the Czech Republic would belong to a different group (groups 1 and 2 respectively). After carefully looking at the data and their relative proximity to their neighbours within both groups, we made the judgment that the placement in the squared Euclidian clustering was the more appropriate choice.

Again, as we are not claiming with certainty that these groups are ‘etched in stone’, the pattern still tells us that cluster 1 countries exhibit relatively high levels of QoG, while cluster 2 and cluster 3 show good and moderate performance, respectively. Several states in the top group – in particular Denmark, Sweden, Finland and the Netherlands – rank consistently in the top four to five countries in Europe on most all pillars and underlying indicators and in the top 5% worldwide on almost all QoG data. It is also worth noting that, according to the World Bank data, all MSs are found in the top half of the world ranking in ‘voice and accountability’⁶⁷. On the other three pillars of QoG, there are times at which a few countries’ scores are indistinguishable from the 3rd quartile of countries in the world (the ones that rank between 25% and 50%) and the 2nd quartile (50% to 75% percent), such as Bulgaria, Romania, Greece, Latvia and Italy – particularly on corruption scores. However, none of the MSs are ever found in the 4th quartile (the lowest 25%) and are all safely distinguishable from this group on all four pillars.

Furthermore, we find that the pillar averages for the three cluster groupings are distinct from one another⁶⁸. Although the three groups clearly distinguish themselves in all four components of QoG, along with the total scores, the difference varies according to certain pillars. For example, the three cluster groups are most clearly distinguished with respect to ‘control of corruption’ scores, while the MSs’ scores converge in the measure for the ‘rule of law’. On both VA and GE, the mean cluster scores more or less resemble the overall QoG scores. On none of the individual pillars do any of the groups overlap – suggesting a strong internal consistency of cluster groupings.

2.5 Uncertainly and Sensitivity Analyses

In any composite indicator there are a number of assumptions and creative decisions that the modellers make in order to put together the various underlying indicators into a single number. As Cherchye et al. (2008) argue, “there is no recipe for building composite indicators that is at the same time universally applicable and sufficiently detailed.” Thus, in any composite index, modellers face potential criticism about what they ‘could have done’ when building the underlying variables and how basic changes in the assumptions of weighting variables or aggregating them together might alter the results.

In ranking EU MSs on the basis of such a composite index, it is important to utilize as many as possible of the existing alternative methods to build a composite measure in order to reduce the level of skewness or biases towards certain countries from the original assumptions of the index. We therefore address several questions about how stable the index is (e.g. how much the rankings shift). We then point out those countries that are

⁶⁷ Applying the 90% confidence intervals provided by Kaufmann et al. (2009), one can calculate the relative standing of a state to any one of the four ‘quartiles’ with relative ease, as scores are normalized with a mean of ‘0’ and an SD of ‘1’.

⁶⁸ A visual can be found in the appendix.

particularly prone to volatility based on shifts in the underlying assumptions of the QoG index.

In this section, we run several uncertainty analyses (UA) that entail evaluating the effect of alternative models on the ranking of countries in each of the four pillars of QoG. In particular, we focus on the choice of weights by the original modellers as well as the aggregation method. Of course, there is no way that we can establish the ‘true ranking’ for each individual country, as each indicator of the four pillars of QoG is only an estimate of the *perceived level* of corruption, rule of law etc. However, by testing a number of alternative assumptions about the weighting and aggregation schemes, we can evaluate the robustness of the measure used to calculate the rankings. Uncertainty tests, while still relatively new in the field, have been used by previous studies that evaluated such composite indices as the environmental Performance Index (Saisana & Saltelli 2008) and the Ibrahim Index of African Governance (Saisana, Annoni & Nardo 2009). As in these studies, we follow the guidelines and advice on UA set forth by the OECD (2008).

A brief description of the original data’s assumptions

The WB’s governance indicators are built with the use of an Unobserved Components Model (UCM) owing to the uneven amount of underlying indicators each country may contain in any given pillar of QoG. Each underlying indicator is normalized using a min-max method so that it is bound from 0-1, with higher scores equating to better QoG. The scores are then standardized by subtracting each country’s score in indicator ‘i’ from the mean value and then dividing it over the standard deviation for indicator ‘i’. The weights are essentially ‘country specific’ in that each indicator is weighted uniquely (outlier variables are weighted less for instance than those that are closer to other indicators). However, because each country potentially has a different number of underlying indicators, the adjusted weights can have a country-specific effect⁶⁹. Such a method allows for the use of regional-specific data as well as survey data with less than global coverage.

The indicators are then aggregated arithmetically, taking the sum product of the weights multiplied by the standardized indicator and then dividing each country’s score over the number of indicators it has multiplied by the weights in order to compensate for differences in the raw number of indicators. Once the indicators are aggregated, they are again standardized so that the global average for each year is ‘0’ with a standard deviation of ‘1’.

While this method allows for different countries to have different numbers of underlying indicators, it rewards a conformity of the data, so to speak, in that it minimizes the impact of outliers (OECD 2008: 101). Moreover, with highly correlated underlying indicators, identification problems could occur (OECD 2008: 101).

⁶⁹ See a more detailed description of UCM in the appendix of the final report.

Evaluation of the Data: A Multi-Modelling Approach

We gather all underlying data for the WB governance indicators for 2008. After replicating the Bank's estimates, we explore three key aspects of the model's assumptions in the UA. They deal with 1) the aggregation method, 2) the weighting method and 3) the number of indicators included in the pillar.

1. The aggregation method

The World Bank data are aggregated using a linear, additive method. Although this method is very common in contemporary indexes, it poses potential drawbacks. First, a strong assumption is made of the independence of the underlying indicators. In essence, the underlying variables are taken independently and thus a poor score in one indicator can be offset by a relatively high score in another indicator. Nardo et al. (2005: 79-80) give the following example: “if a hypothetical composite were formed by inequality, environmental degradation, GDP per capita and unemployment, two countries, one with values 21, 1, 1, 1; and the other with 6,6,6,6 would have equal composite if the aggregation is additive. Obviously the two countries would represent very different social conditions that would not be reflected in the composite.”

We take this warning seriously and therefore use the recommended geometric aggregation method as an alternative - which in the above example would give country A a score of 2.14 and country B one of 6. Thus one high score (which could potentially be a misleading outlier) does not skew the end result as much as occurs in the linear additive method. For geometric aggregation, we linearly transform the indicators so that all figures are positive (>'0') and aggregate the data using a multiplicative formula rather than additive.

2. The Weighting Scheme

As noted, the weighting scheme is based on the UCM, which can potentially have individual country-specific weights. We employ two alternative weighting schemes to test the robustness of the original data. First we use a simple equal weighting (EW) so that the correlation of the underlying indicators is not taken into account as in the original weighting scheme. Second we employ a new set of weights based on principle component analysis (PCA), including only common underlying data for each pillar. When necessary, we use factor rotation and square the factor loadings to derive the new weights. All weights for each indicator in each pillar are listed in the appendix.

3. Inclusion/Exclusion of Individual Indicators

For each pillar, we further test the robustness of the index by excluding each indicator relevant to the EU 27 by giving it essentially a '0' weight in the simulation. We do this to find out whether any one indicator's absence would significantly alter the results and

in what weighting scheme and aggregation context. We report all simulations for each pillar.

A Brief Note about Missing Data

In the simulations for the FA and geometric weighting schemes, we employ only common data sources, meaning that we remove the two to three region-specific indicators for former Eastern and Central Europe for each pillar and any data with less than 24 EU countries. The vast majority of EU MSs have between nine and 12 sources in common, depending on the pillar. The exception is mainly Malta (which has an average of six indicators per pillar), although Luxembourg and Cyprus also have one to two missing data points for each pillar. In order to maximize the number of underlying indicators for each pillar, we employ the method of mean substitution for each missing data point during these simulations so that they take the value of the EU 27 mean (as opposed to the world mean).

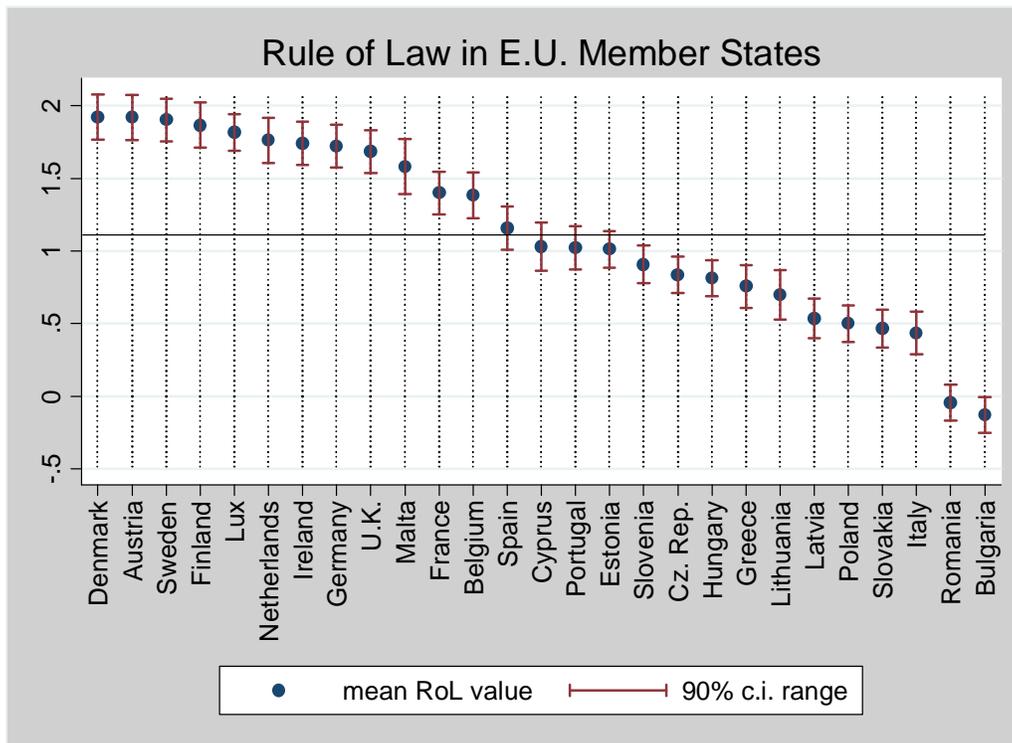
Uncertainty Tests⁷⁰

Pillar 1: Rule of Law (ROL)

Kaufmann et al. (2009) provide a 90% confidence interval for each country's QoG pillar in each year of the data. Figure 11 shows the 2008 rankings for the EU 27 along with the original confidence intervals. According to Kaufmann et al. (2009), we should only interpret significant differences between two countries in the case that the confidence intervals do not overlap. Thus the first significant difference between Denmark and the next highest ranking state would be France. On the other hand, although Denmark ranks number 1 in the original data, its score is all but indistinguishable statistically from other high ranking states such as Austria, Sweden and Finland. Bulgaria and Romania can be distinguished from all other states in the sample. The line in the middle represents the EU average, which all states starting with Belgium (and ranked higher) are significantly above, while all states from Slovenia and below are significantly under.

⁷⁰ See the appendix for detailed tables of the results of each pillar's uncertainty test.

Figure 11



We test the robustness of these rankings in Figure 11 with 78 simulations based on the different aggregation and weighting schemes along with removing each indicator one at a time. The results given in Table 19 show the frequencies of each country’s rank for the 78 simulations. For example, Denmark ranks in the top two positions 90% of the time and 10% of the time is either #3 or #4, while Bulgaria and Romania share the bottom two positions 100% of the time, irrespective of a change in any of the assumptions of the index or when any of the indicators are removed. The bold numbers indicate that a state was found to be in the same place in a majority of the simulations, and there are 12 such states, with seven of them belonging to group 1. Only six countries - Malta, Cyprus, Estonia, Hungary, Poland and Italy – were found to rank at least three places above or below their original ranking in a majority of cases. The other nine cases fell within two places of their original rank, examples being the U.K. and Spain. The two countries with the greatest volatility in the ROL pillar were the Czech Republic and Cyprus, which were found to span across seven to eight different rank positions each (both 13-20) in at least 5% of the simulations. The vast majority remained within two to three positions of their original rank, however, and all states (except possibly for Hungary) remained within their respective cluster blocs.

Sensitivity Results

We complement the uncertainty analysis with a sensitivity analysis for each pillar, beginning with the rule of law. Here we show the results of the ten cases that differ most

from the original rankings according to the Spearman rank coefficient (each of the 78 simulations can be found in the appendix). We report the median and maximum changes and a Spearman Rank Correlation Coefficient, comparing the altered data to the original RoL index. Table 19 shows the results.

We use the full data and then remove each indicator one at a time in the six different weighting/aggregation combinations. Overall, the results are remarkably robust and the original rankings do not appear to be significantly biased at all, with the lowest Spearman coefficient (simulation 30) being .931. Only six of the 78 simulations have a median shift of 2, and the vast majority has a median shift of 1. Eight of the ten most extreme cases use geometric aggregation, while 60% of the top ten cases are an equal weighting scheme. In cases in which the max shift of at least one country was five or more places from its original rank, we report that particular country. The largest shift of one state is Slovenia in simulation 30, with FA weighting and arithmetic aggregation when the Eiu variable is removed. Otherwise, the largest shift is five places, with Hungary and Italy demonstrating this on multiple occasions. It is clear that these two countries, for example, are at a relative disadvantage by the original weighting and aggregation in the RoL data, as they generally rank at least three to five places higher than their original ranking (25 for Italy, 19 for Hungary) in the majority of cases in the simulations.

Table 19: Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Rule of Law (10 Most Extreme Scenarios)

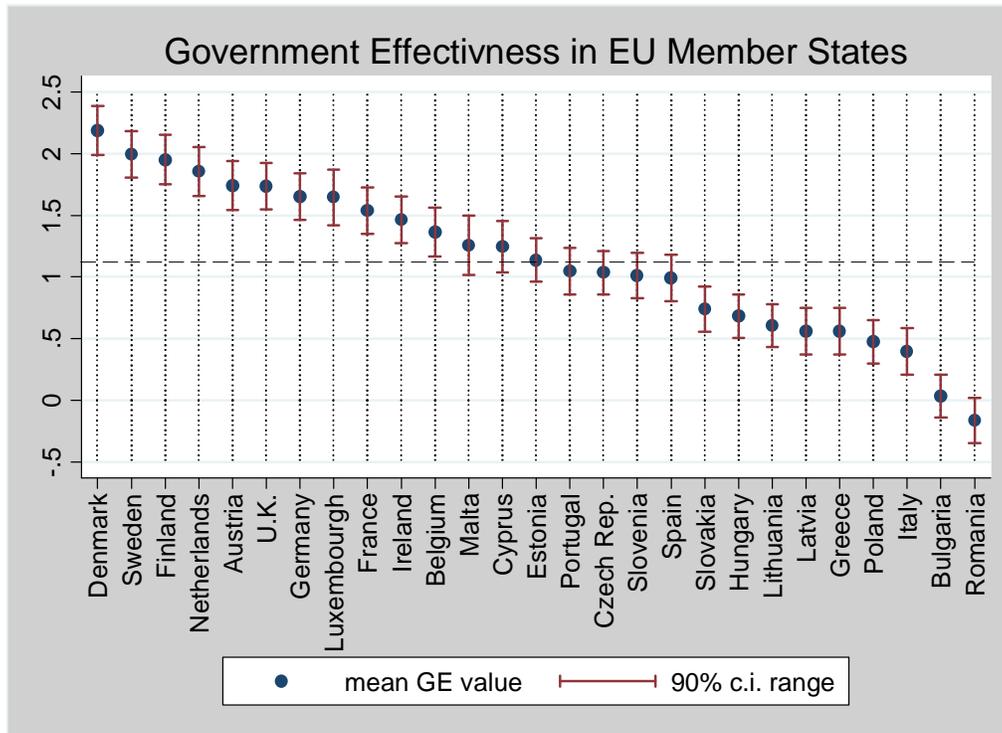
Scenario	Aggregation	Weighting	Excluded Indicator	Median	Max	Spearman Rank Coefficient
30	Arithmetic	FA	Eiu	1	6 (Slovenia+)	0.931
71	Geometric	FA	Prs	1	5 (Italy+)	0.943
45	Geometric	Original	Prs	1	5 (Italy+)	0.949
58	Geometric	Equal	Prs	1	5 (Italy+)	0.949
19	Arithmetic	Equal	Prs	2	5 (Italy+)	0.95
54	Geometric	Equal	Bri	1	4	0.951
22	Arithmetic	Equal	Gwp	1	5 (Hungary+)	0.954
46	Geometric	Original	Wmo	2	5 (Italy, Hun. +)	0.955
57	Geometric	Equal	Gad	1	5 (Italy+)	0.955
59	Geometric	Equal	Wmo	2	5 (Italy+)	0.955

Pillar 2: Government Effectiveness (GE)

We start with a distribution of the original scores, using the WB's confidence intervals to show significant differences between EU states. Again the top four countries – Denmark, Sweden, Finland and the Netherlands - are statistically indistinguishable from one another on the basis of the estimates and the confidence intervals surrounding them. However, Denmark is statistically higher than all EU states from Austria and downward,

while Sweden shows the same with states from Ireland and downward. All countries from #11 (Belgium) and up distinguish themselves from the mean EU score and from #19 (Slovakia) and down are clearly below the EU average, which re-enforces the cluster groupings reported in the previous section. States in the middle ranking, between 12 and 18 (Malta, Cyprus, Estonia, Portugal, Czech Rep., Slovenia and Spain), are virtually indistinguishable from one another.

Figure 12



Based on the results from the uncertainty tests for each country for the GE pillar, we observe results similar to the original rankings, as most countries ended up in their respective rank row in the majority or the near majority of cases after the 60 simulations. Denmark appears to have been slightly under-ranked in the original GE data, as it falls into 1st or 2nd place about three-fourths of the time.

We find that, overall, ten of the EU 27 countries end up in their original rank box at least 50% of the time after the 60 simulations, which points to somewhat more volatility as compared with the RL data. Portugal and Spain have a majority of the scenarios in which they are ranked at least three places higher than their original ranking, suggesting that they are hurt by some of the original assumptions of the index. States appearing to benefit from the original data are Slovakia, Hungary and Estonia, which have the majority of their respective scenario simulations at least three places under their original ranking. The state with the least certainty around its rank position is indeed Belgium in

the GE pillar, which saw at least 5% of its 60 scenarios range between 5th and 13th place, although 45% of the simulations had it in the 11-12 range, its original spot.

Table 20 reports the sensitivity analyses results for the ten cases that differ most from the original results according to the Spearman rank coefficient from the 60 simulations for the GE pillar. The overall results again show impressive robustness, as the median shift is never higher than two places and the Spearman coefficient does not fall below .9 in any of the 66 simulations (low is #47 and # 37 at .907). Based on alternative aggregation and weighting scenarios – in particular geometric aggregation, which nine of the ten most different cases use – several states show a pattern of shifting up or down in the rankings by about five or six places. Italy, Spain, Poland and Belgium all end up five or six places higher than their original rankings on several occasions, while the initial GE data seem to have benefitted Slovakia the most, with Hungary and Latvia also ending up lower than their original rank in multiple scenarios. The largest jumps occur in simulation 38 (Slovenia drops eight spots) twice, with Ireland moving from 10th place up to 3rd in numbers 17 and 47 and Italy jumps seven spots four times in the geometric aggregated simulations in both the original and equal weighting in scenarios 34, 39, 44 and 49. Despite these shifts in a handful of countries, the shift from one cluster to another is very rare - for example, Italy's frequent shifts up to rank #20 or #19 up from #25 is still within the bounds of group 3 – demonstrating that most of the variation of state rankings is *within* – and not *between* – clusters of EU states.

Table 20: Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Government Effectiveness (10 Most Extreme Scenarios)

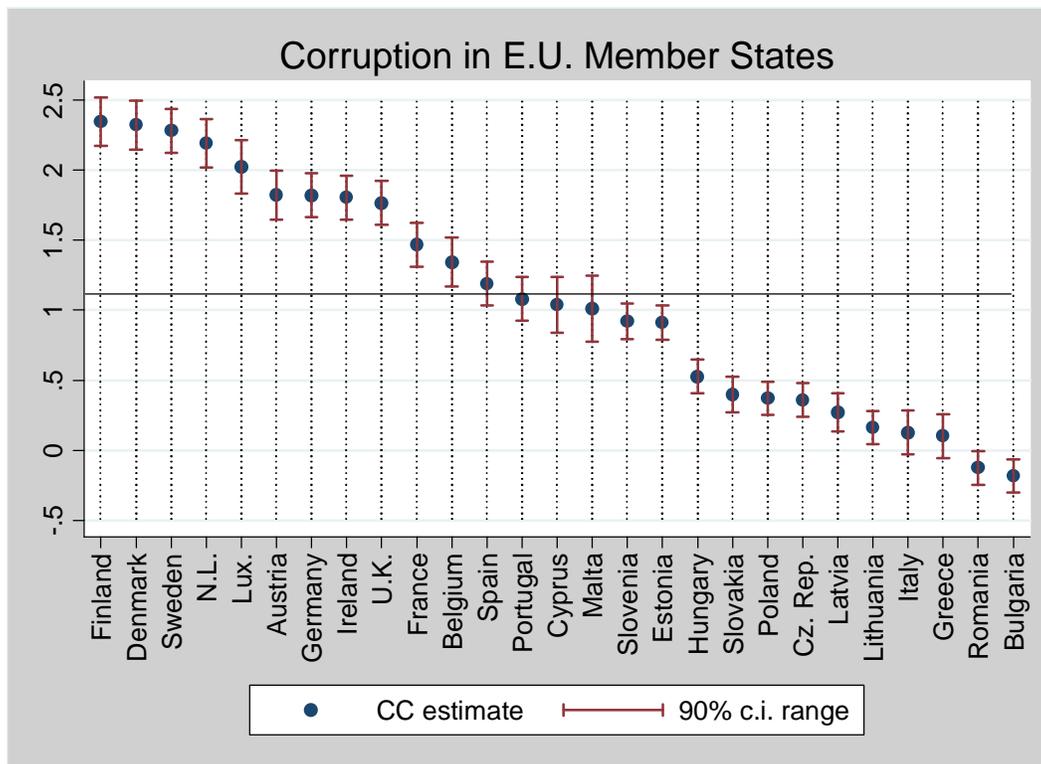
Scenario	Aggregation	Weighting	Excluded Indicator	Median	Max	Spearman Rank Coefficient
37	Geometric	Original	Gcs	2	7 (Ireland+)	0.907
47	Geometric	Equal	Gcs	2	7 (Ireland +)	0.907
17	Arithmetic	Equal	Gcs	2	7 (Ireland+)	0.914
34	Geometric	Original	Eiu	2	7 (Italy+)	0.925
44	Geometric	Equal	Eiu	2	7 (Italy+)	0.925
39	Geometric	Original	Wcy	2	7 (Italy+)	0.928
49	Geometric	Equal	Wcy	2	7 (Italy+)	0.928
59	Geometric	FA	Wcy	1	5 (Spain+)	0.928
36	Geometric	Original	Wmo	2	6 (Slovak-, Pol.+)	0.932
46	Geometric	Equal	Wmo	2	6 (Poland+)	0.932

Pillar 3: Control of Corruption

The results for the control of corruption pillar show that there is a more significant variation between countries than in the previous two pillars. Group 1 (from the cluster analysis) almost completely distinguishes itself statistically from both groups 2 and 3.

Moreover, the gap between Estonia and Slovakia (border states in clusters 2 and 3) is significant at the 90% level of confidence: thus group 2 completely distinguishes itself from group 3 as well. The top four countries in particular show remarkably high scores significantly over 2 in the original WB data, meaning that they are at least two standard deviations above the world mean score. Bulgaria and Romania perform relatively poorly relative to world rankings, however, as they are significantly below world averages according to the 90% confidence interval.

Figure 13



The results show that the three cluster groupings show remarkable robustness, as every country in the sample stays within its cluster grouping range. Further, 19 of the 27 states – including the top and bottom six ranked - are found in their original place at least 50% of the time (bold numbers), while those that show up below or above their original ranking in a majority of simulations (for example, the U.K., originally ranked 9, was found between 5th and 8th place 83% of the time) were only a few places above or below. The uncertainty tests show no signs of the original data negatively biasing any individual country. The largest range is Slovenia, which was found anywhere from 12th to 18th in the rankings, which are again within the second cluster grouping.

The sensitivity analysis results for each individual simulation demonstrate that the CC scores are, in comparison to the other three pillars, the most robust of the QoG data. In fact, in 23 simulations of the 60, the median state does not shift even one place in the

rankings and, in the other 37 cases; the median shift is only equal to 1. Moreover, the Spearman rank coefficient does not drop below .97 in any of the cases, and in no simulation did we find that the maximum shift was larger than four places from the original rankings. We are therefore quite confident of the robustness of the CC rankings from the World Bank data – there is apparently no bias from the original aggregation or weighting methods that hurt or help any states in any significant way from the initial rank order.

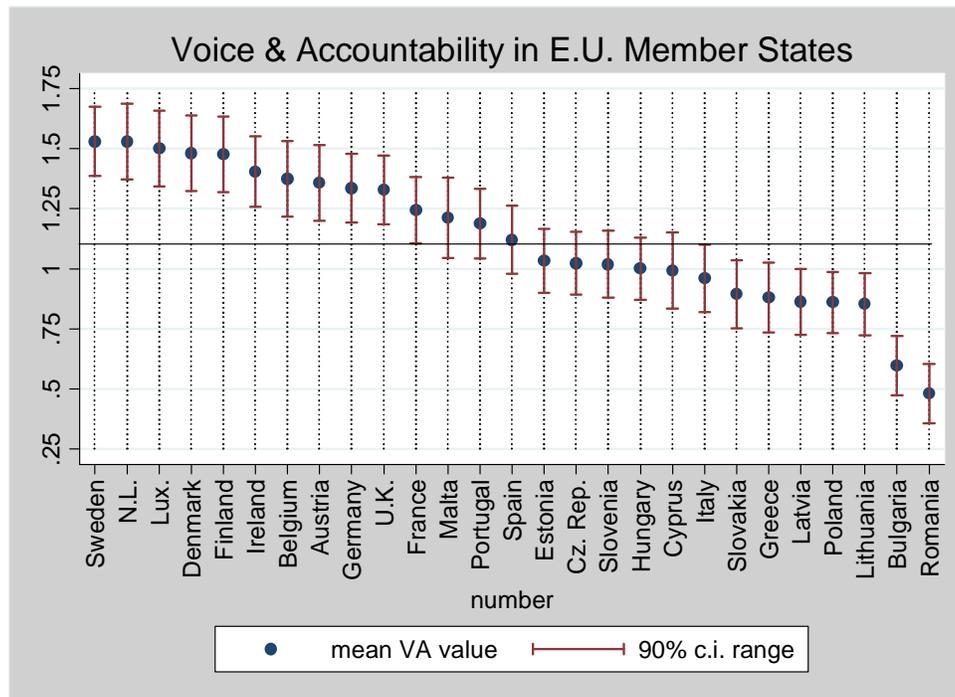
Table 21: Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Corruption (10 Most Extreme Scenarios)

Scenario	Aggregation	Weighting	Excluded Indicator	Median	Max	Spearman Rank Coefficient
36	Geometric	Original	Prs	1	4	0.971
58	Geometric	FA	Gcs	1	3	0.971
46	Geometric	Equal	Prs	1	4	0.972
26	Arithmetic	FA	Prs	1	4	0.974
16	Arithmetic	Equal	Prs	1	4	0.975
12	Arithmetic	Equal	Bri	1	4	0.981
54	Geometric	FA	Eiu	1	3	0.981
28	Arithmetic	FA	Gcs	1	3	0.982
34	Geometric	Original	Eiu	1	3	0.982
38	Geometric	Original	Gcs	1	3	0.982

Pillar 4: Voice & Accountability (VA)

The final pillar displays a rank order similar to the first three yet has a much smaller range of scores than these three pillars. All EU states are significantly over the world average (WB data has a mean of '0') yet the higher ranking states are not two standard deviations over the mean as they have been, but approximately 1.5 over the mean, thus making for a tighter grouping in the VA pillar. Thus the two highest ranking EU states, Sweden and the Netherlands, are statistically indistinguishable all the way down to the 11th ranked country (France). Spain's (ranked 14th) confidence interval overlaps with both Belgium's and Poland's, meaning for example that there is a 90% chance that Spain's 'true' VA estimate could be between 7th and 23rd in the EU.

Figure 14



Due to the tighter score groupings in the VA pillar, we would expect somewhat more volatility in the rankings than in the previous three pillars. We discovered this indeed to be the case. Only eight countries are found to be in their original rank at least 50% of the time after all simulations. We find that, after the 66 simulations, seven countries (Germany, Malta, Estonia, Cyprus, Hungary, Greece and Poland) ended up in four different boxes at least 5% of the time, meaning that they could plausibly be in one of eight different places in the rankings from where they were in the original VA data. Cyprus is an extreme case. Originally ranked 19th, it is only found in the 19th or 20th ranking in 11% of the simulations, while it reaches the 9th or 10th ranking 17% of the time and even the 8th position in the rankings in 6% of the simulations. Poland also appears to have been affected negatively by the original assumptions made according to the data, and ends up ranked either 19th or 20th in 60% of the simulations, about four to five places above its original ranking. Even Romania, which until the VA pillar has been in the bottom two spots in almost every simulation, ends up a couple of places higher in a majority of cases. Notable countries that seem to significantly benefit from the original weights and additive aggregation schemes (at least four to five spots ahead of the majority of the simulation outcomes) include Hungary and possibly Germany.

Table 22: Sensitivity Analysis: Impact of Assumptions of Weighting, Aggregation and Exclusion of Single Indicators on Voice & Accountability (10 Most Extreme Scenarios)

Scenario	Aggregation	Weighting	Excluded Indicator	Median	Max	Spearman Rank Coefficient
20	Arithmetic	Equal	GCS	2	9 (Cyprus+)	0.916
31	Arithmetic	FA	GCS	2	8 (Cyprus+)	0.916
29	Arithmetic	FA	HUM	2	10 (Cyprus+)	0.918
18	Arithmetic	Equal	HUM	2	9 (Cyprus+)	0.919
13	Arithmetic	Equal	WMO	2	11 (Cyprus+)	0.92
19	Arithmetic	Equal	WCY	2	10 (Cyprus+)	0.921
24	Arithmetic	FA	WMO	2	11 (Cyprus+)	0.921
30	Arithmetic	FA	WCY	1	11 (Cyprus+)	0.921
16	Arithmetic	Equal	EIU	2	10 (Cyprus+)	0.922
27	Arithmetic	FA	EIU	2	10 (Cyprus+)	0.922

It is obvious from the results of the sensitivity analysis that one state in particular, Cyprus, is clearly disadvantaged from the original weights and aggregation methods. For most all cases from simulation 12 and higher, we see Cyprus jump between six and 11 spots in all but a handful of cases, all with either FA or EW weighting schemes in arithmetic aggregation. Poland seems to benefit from geometric aggregation of the indicators, while Latvia jumps up six places in the rankings when the Gallop World Poll data are taken out in additive aggregate schemes. Surprisingly, Germany ends up ranked between 13th and 15th in about 29% of the simulations. Aside from those few cases, however, the remainder of the countries is again quite stable relative to the original ranking – in not one case does the median rank rise above two. Furthermore, the lowest Spearman Rank coefficient is still above .91 (simulation 31 with additive aggregation and the factor weight is .916).

In the case of all four pillars, we find that the World Bank data provide reliable estimates of QoG in EU states for the year 2008. We now move on to the next section in which we show how QoG relates to three outside variables: GDP per capita, income inequality and the infant death rate.

2.6 External Validity: Examining Correlates of the Index

In this section, we analyse the relationship between five potential correlates of QoG in the European Union. We take all data for four of these variables from Eurostat and the fifth, child mortality rates, come from the United Nations. For each variable, we show scatter plots of the bivariate relationship and Pearson's correlation coefficient. We also ran bivariate regression to test the significance between QoG and five indicators. Specifically, we look at the relationship of the QoG index and:

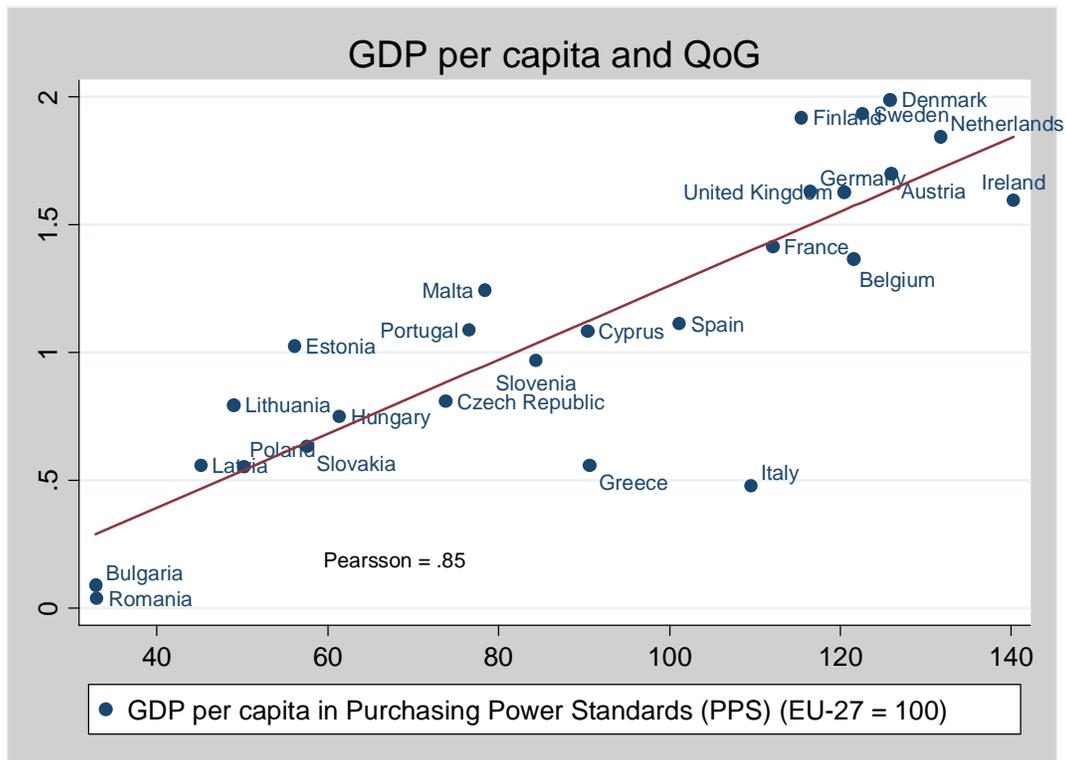
1. GDP per capita
2. Social capital
3. Income inequality (Gini index)
4. Child mortality rates
5. Population

We do not intend to make causal claims for any of these variables to QoG as this cannot be proven with simple bivariate, spatial correlations. We seek only to demonstrate that the QoG index demonstrates clear relationships with salient variables typically found in the literature, which we argue points to a degree of external validity for the governance data.

I. GDP per capita

We find that GDP per capita (measured in PPS, averaged over 2000-2007, taken from Eurostat) is highly correlated with the QoG index, as we would expect. The Pearson correlation coefficient reveals that the two variables are correlated at .85. The bivariate regression analysis shows that the t-statistic is 7.87; thus the relationship is strongly significant, at the 99% level of confidence. The EU 27 shows fairly wide variation in both GDP per capita as well as QoG, revealing that states with the best levels of QoG also enjoy the highest living standards across the union.

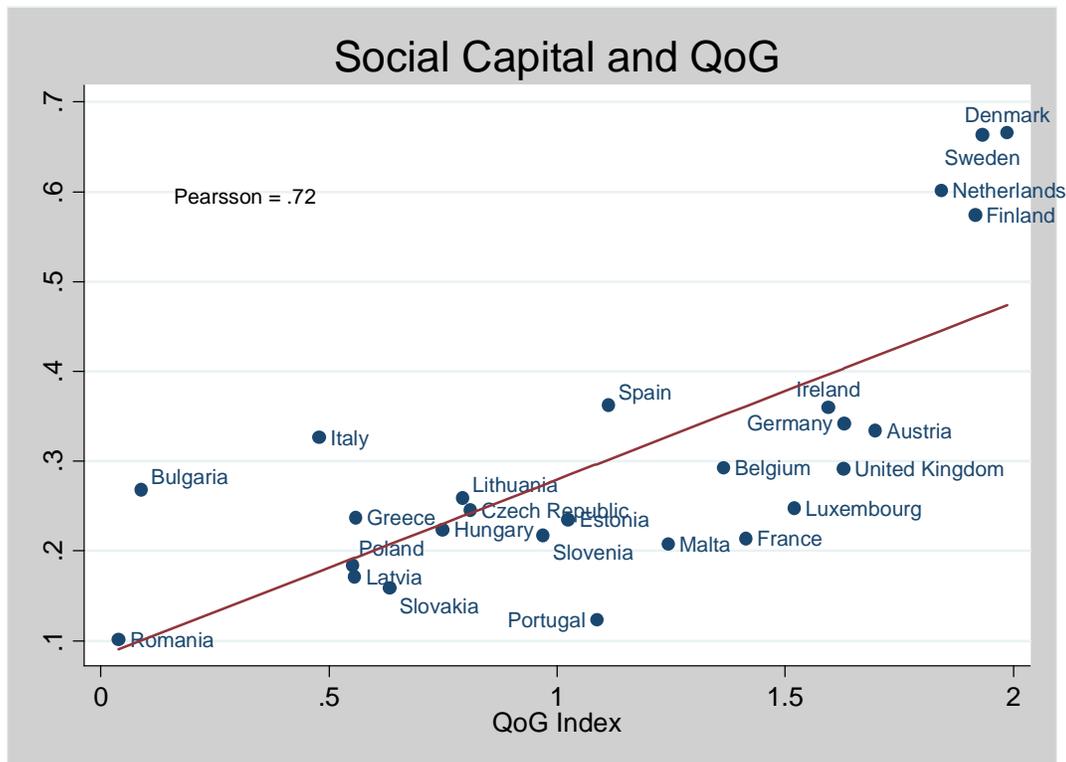
Figure 15



II. Social Capital

There is a relatively new burgeoning literature on the effects of QoG on social capital in a society, with the argument being that strong, impartial government institutions reduce sentiments that certain groups are being favoured or discriminated against and thus foster trust among citizens toward one another and their government (Rothstein and Stolle 2008; Rothstein and Eek 2009). We measure social capital by taking each country’s response to the World Value Survey question on interpersonal trust – namely, “would you say that most people can be trusted or that you need to be very careful in dealing with people?” Higher scores equate to higher levels of trust. The positive relationship between recent social capital and QoG is significant according to the t-statistic (5.10) in the bivariate regression. The Pearson coefficient is .72. The result of this bivariate relationship among EU countries supports the relationship previously found in several empirical studies. The top four countries in QoG, Denmark, Sweden, Finland and the Netherlands, display high levels of trust in fellow citizens, while lower ranking states in QoG such as Romania, Latvia, Poland and Slovakia display much lower levels of interpersonal trust. However, the absence of the four top QoG countries weakens the positive relationship somewhat between the two variables (the t-score drops to 2.69 and Pearson’s coefficient correlation drops to .55).

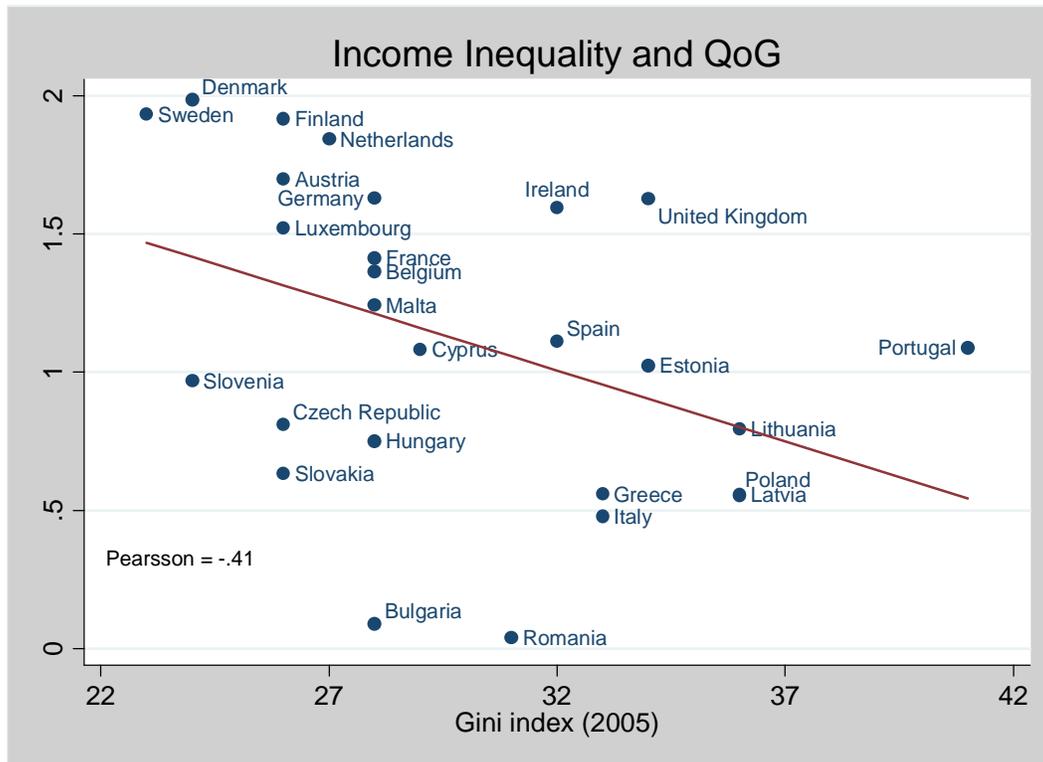
Figure 16



III. Income Inequality

Income inequality, as measured by the 2005 Gini Index⁷¹, has a negative relationship with QoG as we would expect. The relationship is significant at the 95% level of confidence and the Pearson coefficient shows a $-.55$ correlation. Although it is clear that the majority of the most equal states in terms of income are also those that have the highest levels of QoG, there are several EU countries, such as Slovenia, Slovakia and Bulgaria, that are far below the estimated linear prediction. However, one might argue that the income equality in these states is most likely a residual legacy from the previous socialist regimes. Overall, despite the fact that the variance of the relationship is fairly wide among the EU 27, we can see that the general relationship is such that: states with better QoG on average are more likely to have lower income inequality.

Figure 17



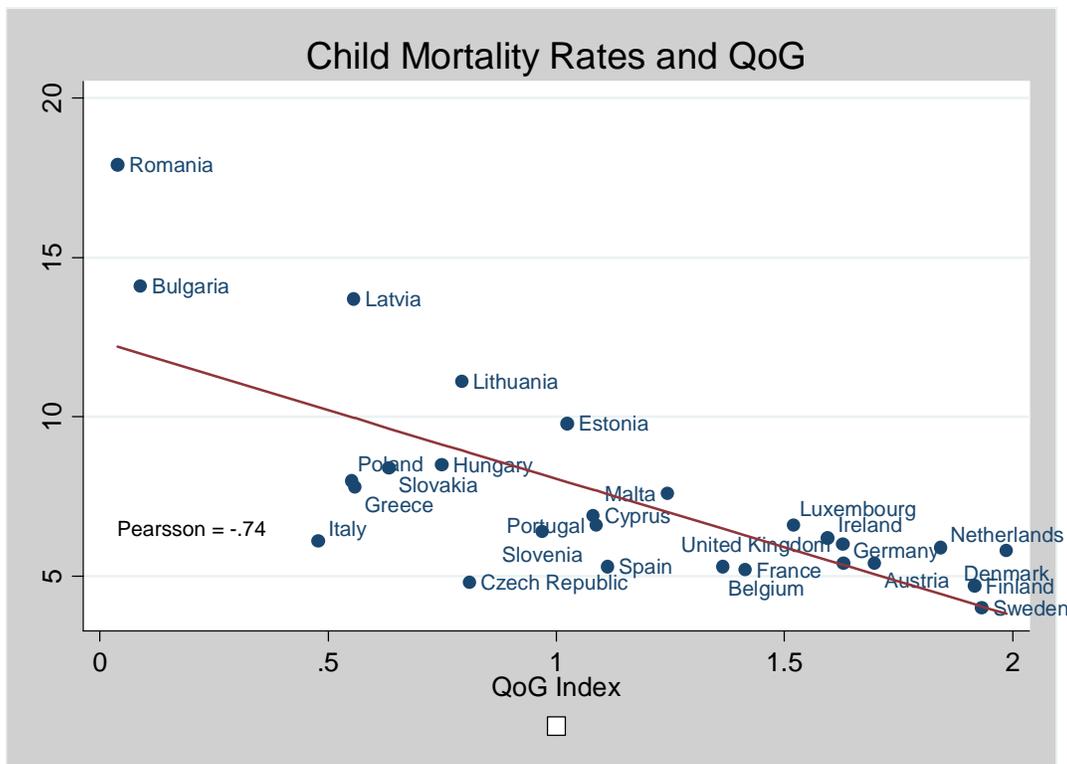
IV. Child Mortality Rates

We use the indicator of child mortality rates at the age of five and below (per 1,000 live births) from the *United Nations Population Division*. This indicator is argued by many to

⁷¹ 2000 Gini index figures were used to estimate Bulgaria's income inequality level.

be a robust indicator of a country’s health system and we would argue that QoG has consequences for this (thus we place QoG on the ‘Y’ axis here). There is a strong, negative correlation (t-statistic= -5.50 and significant at the 99% level of confidence) that suggests that QoG improves health conditions which in turn reduces the number of child mortalities. Sweden, Denmark and Finland are the top three in both categories, while Bulgaria, Romania and Latvia, with considerably lower QoG than the EU average, also display relatively high child mortality rates. The Czech Republic appears to be close to the top three in this important health outcome, despite being considerably lower in QoG. However, the relationship seems to fit quite strongly overall, with Pearson’s correlation coefficient showing -.74.

Figure 18

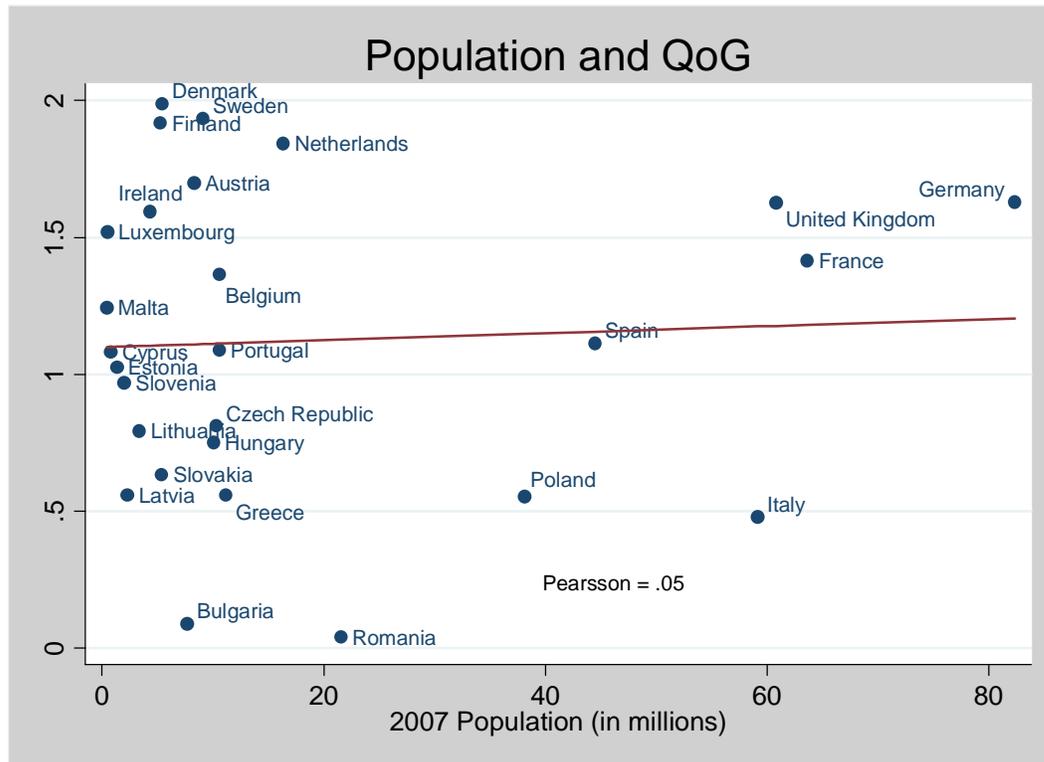


V. Population

While we have no strong theoretical reason to believe that population and QoG would be related, we show the relationship between the two variables simply to test whether smaller states perform better on average – as the top three or four countries on the QoG index would suggest. We find no such relationship in the EU. The t-test from the regression reveals that the relationship is nowhere close to statistical significance (0.27) while the Pearson coefficient shows a very small positive relationship (0.05). While indeed the best performing states with respect to QoG have smaller populations on

average, so do some of the relatively poorer performing countries, such as Slovakia, Latvia, Bulgaria and Greece. Smaller countries are as likely to rank high as low on QoG in the EU.

Figure 19



2.7 Conclusion

In this section we have rigorously analyzed the internal and external consistency, robustness and sensitivity of the data used to build the Quality of Government Index for the EU 27 member states. The QoG index that we have built from the World Bank Governance Indicators (Kaufmann, Kraay and Matruzzi 2009) consists of four pillars of QoG:

1. Government effectiveness (GE)
2. Control of corruption (CC)
3. Rule of law (RL)
4. Voice and accountability (VA)

Individually and collectively, in the original data, these four separate indices are made up of over 70 individual underlying indicators from over 30 separate sources ranging from expert assessments to firm and citizen surveys to NGO and governmental evaluations of various aspects of QoG. We acknowledge that this indicator captures QoG in a broad

sense and, unlike data that target very specific aspects of, say, judicial independence or petty corruption in the bureaucracy, this wide scope could be problematic for pointing out specific areas that states could address to improve their ranking or overall QoG performance. However, we argue that, as a starting point for a conversation about QoG, it is best not to leave out what could be vital elements of this concept, which is an advantage of using a ‘wide’ composite indicator such as the World Bank data, which are unique in their conceptual, temporal and country coverage.

Summary of Important Findings:

- The underlying data are almost without exception internally consistent. We find that all indicators among the four pillars save two (BPS and GII) are positively correlated with the overall QoG index. Further, the factor analyses show that both the common underlying indicators in CC and GE represent a single concept. The RL and VA reported two significant factors each, speaking for a possible conceptual diversity among the 12 and ten indicators the EU 27 have in common in these two pillars, respectively. Moreover, the pillars correlated strongly among themselves (Pearson correlations were between .84 and .98), showing internal consistency in their aggregation into a single index.
- After a hierarchical cluster analysis, the data show that there are three significant cluster groupings of states with respect to QoG. We acknowledge that it is very difficult to clearly distinguish two countries that consistently fall in a similar rank position from one another using *any* existing QoG data available, such as Sweden and Denmark, Spain and Portugal or Romania and Bulgaria. However, with the cluster analysis and the confidence intervals provided by the World Bank data, we can distinguish *groups of states* from one another rather clearly. We tested the strength of the cluster groupings over the four separate pillars of QoG and found that in no cases did the mean score of any group overlap with another group.
- We tested the validity and robustness of the rankings by employing several statistical tests of uncertainty and sensitivity. We altered the weighting and aggregation scheme, along with removing each indicator individually, of the underlying data for each pillar individually, for a total of 264 simulations. Among the four QoG pillars, we find CC to be the least volatile and most robust to any changes in the original assumptions of the building of the index. For example, in none of the 60 simulations did the mean rank change go above one place and every country stayed within only four places of their original rank throughout the simulations. We found VA, which showed a tighter original set of scores among the EU 27, to be the most volatile to change in the rankings, based on shifts in the weighting, aggregation or removal of variables, with Latvia,

Hungary and Poland moving five or six places on multiple occasions and Cyprus jumping up 11 places at times, the largest shift in any of the 264 simulations. Yet none of the Pearson rank correlation coefficients dipped below .90, nor did the median shift in rank go above 2 in any of the 66 VA simulations. GE and RL fell in between these other two pillars with respect to the sensitivity of the rankings. Overall, we did not find that any member state was significantly biased by the combined index. While admittedly, at times, certain states, such as Italy in GE and RL, Cyprus in VA, or Poland in VA, seemed under-ranked after assessing their rank for all simulations compared with the original, their overall standing within their respective cluster among the four pillars remained robust.

- The overall index relates well externally to several variables that that have been shown in recent publications to correlate highly with QoG. GDP per capita, social capital, income inequality and child mortality rates all significantly correlate with the World Bank QoG index in the directions expected according to the bivariate correlations and regressions. As regards the relationship between population size and QoG, at least among the EU sample, there appears to be no advantage, neither in large nor in small states. As concerns the causal direction of several of these outside variables – such as GDP per capital or income inequality – we leave this to further scholarship to establish. We sought only to show a significant correlation for the sake of external validity checks

The QoG index constructed here to assess relative positions among EU member states has passed the basic statistical hurdles used to evaluate the quality of a composite index. We acknowledge seriously the ‘conceptual-stretching’-type pitfalls of building an index out of four pillars which are themselves composite indices; this is exactly why we carefully tested each individual pillar throughout section 2. Taking into consideration the possible risks in using wide and generally subjective indicators of QoG that underlie the overall index, we find that the evidence presented in this section is in favour of an internally and externally consistent and robust set of rankings for EU countries.

Part III: Measuring QoG at the Regional Level

In this section, we construct a composite index score for 172 regions within 18 EU countries based on 16 underlying indicators of QoG. We build the index from a survey of approximately 34,000 EU citizens, which constitutes the largest survey undertaken to date to measure QoG at the sub-national level. In this section of the report, we describe the survey itself and show the results of the survey at the individual level, reporting regressions using the demographic data of the respondents. We then report step-by-step details about how the data are constructed. We discuss in detail many of the decisions that were made regarding the 16 indicators, how we normalized, aggregated and weighted them, and we report 62 alternative simulations in which we test the sensitivity of our choices to construct the index. We report the final scores for each region based on their relative rank, along with aggregated country means and in-country regional variations. We also report checks of internal and external validity of the data, showing how the index relates to national measures of QoG, and we show bivariate scatter plot relationships between the regional QoG index and economic development, levels of education and population, for example. As measuring QoG is a difficult task, and as doing so at the sub-national level is ‘uncharted territory’ for most EU states, we argue that, based on the results of the validity and sensitivity tests, the QoG regional index is a sound ‘first cut’ at measuring QoG at the regional level in the European Union.

3.1 Description of the Survey Data: Individual Level

Although there are obvious trade-offs between citizen-based surveys and so called ‘expert’ surveys, we decided the former was more appropriate for the following reasons. Firstly, citizens are the consumers of QoG and we trust that their opinions on their regional services are generally valid. Secondly, although many experts in public administration can share their opinions on certain countries in which they have expertise, we found the task of locating a sufficient number of sub-national level experts to be a daunting task. In certain regions where there exist no research institutes, universities or knowledgeable NGOs, we did not feel that the reliability of an opinion outside such a region would qualify as ‘expert’. Thirdly, the costs to a citizen-based survey were relatively cheap compared with the estimates for an expert-based survey. For the latter, according to several survey firms with which we had contact during this process, we were told we would be guaranteed only about five responses per region (this would be face-to-face interviews) and that such a process would require more time. For the citizen-based surveys, we were informed that phone interviews were reliable and efficient and that we could have a sample size of over 33,000 respondents rather than a few hundred. Finally, as discussed in part II of the report, the vast majority of QoG data is ‘expert-based’, or

based on assessments from investment risk services – thus our citizen-based data would serve as a nice complement to a wealth of existing expert-driven data in the field already.

The E.U. regional survey was undertaken between 15 December, 2009, and 1 February, 2010 by Effience 3, a French market-research, survey company⁷². A full description of the survey method and other details about how the survey was conducted is given in the appendix. The respondents, from 18 years of age or older, were contacted randomly via telephone in the local language by the ‘birthday method’. The next birthday method is an alternative to the quotas method. As with anything in social science research, there are trade-offs in every method. When using the quota method for instance, one obtains a (near) perfectly representative sample. However, as one searches for certain demographics within the population, one might end up with only ‘available’ respondents, or those that are more ‘eager’ to respond to surveys, which can lead to less variation in the responses. The ‘next-birthday’ method, which simply requires the interviewer to ask the person who answers the phone who in their household will have the next birthday, still obtains a reasonably representative sample of the population (Longstreth, and Shields 2009). The interviewer must take the person who has the next coming birthday in the household (if this person is not available, the interviewer makes an appointment), thus not relying on whomever might be available to respond in the household. So, where the quota method is stronger in terms of a more even demographic spread in the sample, the next-birthday method is stronger at ensuring a better range of opinion. After consulting with E3, we decided on the next-birthday method because we felt that what we might have lost in demographic representation in the sample would be made up for by a better distribution of opinion.

3.1.1 The Survey Questions: An Overview

In all, we posed 34 QoG and demographic-based questions to the respondents. The first three questions were intended to find out how much recent experience the respondents had of the three areas of public service about which we were questioning.

1. *‘Have you or any of your children been enrolled or employed in the public school system in your area in the past 12 months?’ (yes, no)*
2. *‘In the past 12 months have you used public health care services in your area?’ (yes/no)*
3. *‘Have you had any recent contact (positive or negative) with the security or police forces in your area in the past 12 months?’ (yes/no)*

⁷² <http://www.effience3.com/en/accueil/index.html>

Table 23: Respondent Experience with 3 Public Services in 18 E.U. Countries

	Obs.	Education	Law	Health
<i>Total</i>	33540	0.60	0.25	0.78
1 Austria	1755	0.57	0.39	0.70
2 Belgium	585	0.73	0.37	0.77
3 Bulgaria	1170	0.73	0.22	0.85
4 Czech rep.	1560	0.53	0.17	0.78
5 Denmark	975	0.66	0.25	0.76
6 France	5070	0.64	0.18	0.86
7 Germany	3120	0.57	0.43	0.73
8 Greece	780	0.69	0.56	0.87
9 Hungary	585	0.24	0.31	0.76
10 Italy	4095	0.63	0.18	0.76
11 NL	780	0.80	0.19	0.79
12 Poland	3120	0.58	0.25	0.74
13 Portugal	1365	0.69	0.21	0.85
14 Romania	1560	0.64	0.33	0.89
15 Slovakia	780	0.58	0.13	0.78
16 Spain	3315	0.36	0.17	0.68
17 Sweden	585	0.62	0.22	0.76
18 UK	2340	0.77	0.25	0.83

Note: numbers represent the proportion of respondents that have answered 'yes' to having direct contact with the service in question.

Blue numbers represent 'low' country in the sample.

Red numbers equal the 'high' response.

With respect to respondents who have had direct contact in the past 12 months with their local police, there is very little variation across countries – at least 70% of the respondents in each country answered that they had indeed had contact with police for one reason or another. Regarding education, we saw much more cross-country variation, with 80% of the respondents from the Netherlands having responded 'yes', while only 24% in Hungary had had contact with their local public education services in the past 12 months. The cross-country variation in respondents answering 'yes' on health services was somewhere in between – with 13% of citizens in Slovakia answering 'yes' on this question, while 56% in Greece had contact with their public health services in the past year. Most countries however ranged between 17% and 33% on this question.

In this section reports the mean, raw adjusted scores from the individual level responses – both the averages across the entire EU sample and the low and high country for each question. The next 19 questions on the survey deal directly with different aspects of QoG. In the first three, citizens were asked to rate the quality of these three regional services from 0 (extremely poor quality) to 10 (extremely high quality).

4. 'How would you rate the quality of public education in your area?'
5. 'How would you rate the quality of the public health care system in your area?'
6. 'How would you rate the quality of the police force in your area?'

Table 24: Citizen Responses to Quality of Public Services: Mean, High & Low Values by Country

	Education	Health	Law
<i>Total</i>	6.21 (2.09)	6.32 (2.31)	6.21 (2.29)
High Values	7.0 (1.7) U.K.	7.6 (1.5) Belgium	7.0 (2.5) Romania
Low Values	5.2 (2.2) Bulgaria	4.7 (2.4) Greece	4.9 (2.1) Bulgaria

Note: raw scores from 0-10 reported with standard deviation in parentheses.

We observe that, for the most part, EU citizens are on average fairly content with the quality of their regional public services. The average responses for education, health and law enforcement were 6.2, 6.3 and 6.2 respectively – all above the ‘average’ score of 5, indicating that citizens on the whole report positive quality. In table 24 we also report high and low country averages for each question. For example, citizens in the U.K. rate their public education highest (roughly 7), while citizens in Bulgaria rate the quality of their public education lowest in the EU (5.2). With respect to health care, the citizens of Belgium find their quality to be quite high (7.6), followed by Denmark (7.5) and Netherlands (7.47), while Greek respondents find the quality of their health services to be slightly on the poorer side of the scale (e.g. under ‘5’), at 4.7. We find that, with respect to citizen response to the quality of their police service, the two newest members are on extreme ends of the spectrum - Romanian citizens rate this service highest (7.0), while Bulgarian citizens believe their law services to be poorest (4.9).

Questions 7-12 are aimed at measuring the ‘impartiality’ of the services in question. Since ‘impartiality is an even more abstract concept for many people relative to ‘quality’ or even ‘corruption’, we took two complementary approaches to capture this in the data. In the first three questions, we asked citizens to rate whether they agreed that ‘certain people’ get special advantages when dealing with the public service in question from 0 (strongly disagree) to 10 (strongly agree). The second set of questions asks respondents whether all people in their region are ‘treated equally’ by the service in question on a four point scale (*Agree, rather agree, rather disagree or Disagree*). We use all six questions in the final index to allow for as much variation as possible while not letting either the ‘positively’ or ‘negatively’ framed question determine the impartiality data alone.

7. “Certain people are given special advantages in the public education system in my area.”

8. “Certain people are given special advantages in the public health care system in my area.”

9. “The police force gives special advantages to certain people in my area.”

10. “All citizens are treated equally in the public education system in my area”

11. “All citizens are treated equally in the public health care system in my area”

12. “All citizens are treated equally by the police force in my area”

Table 25: Citizen Responses to Impartiality of Public Services: Mean, High & Low Values by Country

	Education	Health	Law
<i>Total (questions 7-9)</i>	5.9 (3.2)	5.4 (3.2)	6.1 (3.1)
High Values	7.3 (2.8) Spain	7.2 (2.8) Spain	7.1 (2.7) Germany
Low Values	4.6 (2.3) Slovakia	3.8 (2.4) Slovakia	4.5 (2.7) Slovakia
<i>Total (questions 10-12)</i>	2.9 (1.1)	2.8 (1.1)	2.8 (1.1)
High Values	3.2 (1.0) Netherlands	3.2 (1.1) N.L. & U.K.	3.3 (.86) Ger. & Den.
Low Values	2.4 (.95) Bulgaria	2.4 (1.0) Bulgaria	2.3 (.98) Bulgaria

Note: scores have been adjusted so that higher values equal higher levels of QoG. (S.D. in parentheses)

Here we find that the two sets of questions are generally consistent with one another, with the average E.U. response indicating that people on average generally feel that their public services are more impartial than partial toward certain people. In the first three questions, we find that respondents generally believe their police institutions to be the least partial, while the health care services to be the most inclined toward favoritism of certain people or groups. Citizens in Spain, Germany, Netherlands, Denmark and Sweden generally responded that their three public services were the most impartial, while Slovakia, Bulgaria Romania, Hungary and Poland thought their institutions were the most inclined toward favoritism. In the second set of questions, phrasing our question in a ‘positive’ way, we find that the Netherlands, Denmark, U.K., Sweden and Germany (for police and education) are the top five countries while Bulgaria, Poland, Romania, Hungary and France have the lowest average scores with respect to the second set of impartiality questions.

The next set of questions deals with respondents’ opinion of the level of corruption in the three public service areas in their region. We begin with three questions (one for

each public service) simply asking them to rate how prevalent they believe corruption is in their region. We ask them how strongly they believe the following three statements from 0 (strongly disagree) to 10 (strongly agree).

13. “Corruption is prevalent in my area’s local public school system”

14. “Corruption is prevalent in the public health care system in my area”

15. “Corruption is prevalent in the police force in my area”

Table 26: Citizen Responses to Corruption in Public Services: Mean, High & Low Values by Country

	Education	Health	Law
<i>Total</i>	6.92 (2.96)	6.25 (3.14)	6.46 (3.09)
High Values	8.31 (2.33) Denmark	8.41 (2.31) Denmark	8.45 (2.35) Denmark
Low Values	5.18 (3.27) Greece	3.70 (2.37) Slovakia	4.47 (3.06) Greece

Scores are inverted so that higher scores mean less corruption (SD in parentheses)

Here we find that the EU average was positive in the sense that the average respondent across the 18 countries generally disagrees with the idea that corruption is prevalent in the three public sectors. However, we see significant variation in the sample when we look at the average responses in the countries themselves. For example, Denmark ranks ‘cleanest’ in all three public services – all of its averages are above 8.30. However, the low value in education services belongs to Greece, in which an approximately equal number of people think or do not think that corruption is prevalent in the education system. Slovakian respondents are quite clear that they believe corruption is rather prevalent in their health care services – as indicated by the low score of 3.70. Greece also ranks lowest with respect to the prevalence of corruption in the police force, with a score below the mean of 4.47.

The next two questions gauging the level of corruption in a region deal with bribery specifically. In the first of the two questions, we ask respondents if they themselves or anyone living in their household has paid a bribe to obtain one of the three public services in their region in the past 12 months (yes/no). In the second question, we ask respondents how often they believe others in their region utilize bribery to obtain a public service from 0 (never) to 10 (very frequently).

16. 'In the past 12 months have you or anyone living in your household paid a bribe in any form to: Education services?(yes/no) Health or medical services? (yes/no) Police? (yes/no)'

17. 'In your opinion, how often do you think other citizens in your area use bribery to obtain public services?'

Table 27: Citizen Responses to Bribery in Public Services: Mean, High & Low Values by Country

	Bribe	Others Bribe
Total	.947 (.223)	5.96 (2.85)
High Values	4 countries > .99	8.13 (2.19) Denmark
Low Values	.794 (.20) Romania	3.37 (2.43) Greece

Scores are adjusted so that higher scores mean less corruption (SD in parentheses)

When we break the bribe question down among the three different services, we find that the occurrence of bribe-paying in both the education and police services is extremely rare, with an EU average of about a 1.1% 'yes' response rate, with a 'low' country percent at less than 4% for either service (Romania at 3.7% for education and Greece at 3.9% for police services). However, the bribe question directed at the health care services was much more interesting with respect to variation among countries. Here we find that the 'yes' response rate is five times greater EU-wide than either of the other two services, and the standard deviation is 2.5 time greater. Thus health care services are clearly the public sector area that are most plagued by bribery in the EU, and we employ this as our measure of bribery.

Here we observe that, overall, the occurrence of an actual bribe paid by a respondent is quite infrequent, with only about 5% of the entire sample responding 'yes' to this question. In fact, less than half of one percent of the respondents in Sweden, the Netherlands, Portugal or Denmark said they or someone in their household had paid a bribe. In Romania, however, almost 21% of the respondents claimed to have paid some form of bribe in general to obtain a public health care service in their region in the past 12 months. The next two countries with the highest level of self-reported bribe-paying were Hungary and Slovakia, with 18.5% and 10%, respectively. The second question asks essentially what respondents thought of the practices of their fellow citizens when they sought to obtain a public service. Greek citizens were the most certain that 'others' employ bribery as a direct means of obtaining public services while, in Denmark, most people strongly disagreed with the idea that 'others' might be using bribery in this way.

The final two questions we include in the QoG index pertain to the strength of the local media in serving as a ‘watchdog’ along with the perception of the integrity of regional elections. Regarding the first question, we asked people how likely it would be that it would be reported in the local media that public official had committed an act of malfeasance between ‘0’ (extremely unlikely) to ‘10’ (extremely likely). Regarding elections, we sought to find out the level to which respondents believed their local elections are honest and clean from corruption, between ‘0’ (strongly disagree) and ‘10’ (strongly agree).

18. *‘In your opinion, if corruption by a public employee or politician were to occur in your area, how likely is it that such corruption would be exposed by the local mass media?’*

19. *‘Please respond to the following: Elections in my area are honest and clean from corruption’*

Table 28: Citizen Responses to Effectiveness of Media and Fairness of elections: Mean, High & Low Values by Country

	Media	Elections
<i>Total</i>	5.35 (3.19)	6.84 (3.11)
High Values	7.88 (2.64) Sweden	8.33 (2.22) Netherlands
Low Values	3.48 (2.92) Czech Rep.	5.71 (2.92) Bulgaria

Scores are adjusted so that higher scores mean better QoG (SD in parentheses)

The relative range in the confidence in the media to serve as a credible watchdog over public servants ranges from quite strong in Sweden (7.88) to quite low in the Czech Republic (3.48), with the majority of the respondents in the latter actually saying that they believe their regional media would *not* report corruption that occurred by public officials in their area. Regarding elections, the average response in all 18 countries was on the ‘positive’ side in terms of the respondents’ belief that their area’s elections are conducted honestly without corruption. However, this sentiment is strongest in the Netherlands (8.33) and weakest in Bulgaria (5.71).

3.1.2 Demographics of Respondents

Looking now at the individual level data, in all, there are 33,540 total respondents from 172 NUTS 1 and NUTS 2 regions in 18 EU countries, with approximately 200

respondents per region⁷³. This renders our study the largest survey conducted to date at the regional level on questions regarding QoG. Aside from questions pertaining to citizen opinions about regional governance, the respondents were also asked several demographic questions. A demographic breakdown of the sample is given in the appendix.

3.1.3 General Levels of Citizen Satisfaction

In addition to QoG questions, we asked respondents about their opinions of the state of the present government, the economy, their region's performance in delivering quality and incorrupt public services relative to others in the country and whether the present state of their regional public sector delivers better quality and does so with less corruption as compared with five years ago. We report the mean responses and number of citizen responses per country (obs.) from the survey regarding the last set of questions in Table 2 below.

Table 29: Satisfaction with Current Government & Economy

<i>Unit</i>	<i>Obs.</i>	<i>Satisfaction of e Economy (0-10)</i>	<i>Satisfaction of Government (0-10)</i>
E.U. average	33540	3.24	3.69
France	5070	3.13	3.81
Germany	3120	3.75	3.80
Italy	4095	3.52	4.02
Spain	3315	2.01	2.43
UK	2340	3.33	3.79
Hungary	585	2.12	2.42
Cz. Rep.	1560	2.28	3.06
Slovakia	780	3.45	4.08
Portugal	1365	2.53	3.58
Romania	1560	2.16	2.76
Sweden	585	4.66	5.06
Denmark	975	5.17	5.23
Belgium	585	4.12	4.24
Austria	1755	4.92	4.58
NL	780	5.11	5.24
Poland	3120	3.46	3.12
Bulgaria	1170	2.61	4.53
Greece	780	1.91	3.92

Note: Standard deviations for EU averaged economic and government satisfaction are 2.63 and 2.76 respectively.

All country standard deviations range from 2.09 (NL) to 3.04 (Cz. Rep.).

*Both variables range from 0-10, with higher scores equating to higher satisfaction. **High & Low figures in Bold***

⁷³ The NUTS 1 in the study are Germany, U.K., Sweden, Greece, Belgium, Hungary and the Netherlands. NUTS 2 countries are Italy, Spain, France, Poland, Slovakia, Austria, Czech Rep., Bulgaria, Romania, Denmark and Portugal.

We find that, on average, the 33,540 citizens that we surveyed were more dissatisfied than satisfied with both the current state of the economy and their respective governments. On a scale of 0-10, with scores under 5 representing relative dissatisfaction and scores above 5 representing relative satisfaction, the aggregated scores for economic and government satisfaction are 3.24 and 3.69, respectively. We find only two cases (Poland and Austria) where the respondents are slightly less satisfied with their current government than the state of the economy. In all other cases, citizens are relatively more satisfied with their governments than with the state of their national economy. There is also a fair amount of variation in the 18 countries polled, with three countries (Denmark, the Netherlands and Sweden) expressing (albeit slight) levels of positive satisfaction with their current governments, while countries such as Spain, Hungary and Romania demonstrate relatively low levels of government satisfaction (all between 2.43 and 2.76). As regards the economy, citizens are relatively most satisfied in Denmark, the Netherlands and Austria (all above 4.9), while the most dissatisfied citizens are in Greece, Spain, Hungary and Romania.

We now move to some basic multivariate regressions at the individual level in order to see whether there are systematic differences of opinion with respect to QoG based on age, gender, income etc. For purposes of direct comparison (some questions have different number scales), we standardize each answer so that the mean is equal to '0' and has a standard deviation of '1'. In a second step, we aggregate the data to take the average response score for each question for each of the 172 regions. On the basis of a PCA we perform with the aggregated regional data, we find three significant factors based on 16 underlying indicators that all correlate with one another. We call these pillars 'quality', 'impartiality' and 'corruption' (to be discussed in further detail in the following section). We combine the standardized scores of the individual questions (i.e. indicators) into their respective pillar, so that each respondent has a 'corruption' score, etc. Although the aggregated pillar data at the individual level does not exactly equate identically to the pillar scores aggregated at the regional level, they are indeed very highly correlated (although in the later part of this analysis only the aggregated data are reported for the regional level). Thus we are able to run multiple regression analysis using them as dependent variables in order to test whether certain demographic groups display systematically high or low responses in each of the pillars. Table 3 explores the relationship between the demographic and confidence variables in the pillars and the entire QoG variable itself.

We begin by standardizing each question at the individual level and then aggregating these to the regional level. We also create a combined 'corruption', 'quality' and 'impartiality' score for each individual and a QoG index score. The reason for this is simply to be able to run several regressions (as reported in the previous section) on the

combined indicators using the control factors, such as education, income and population, for example, which are reported in the previous section⁷⁴.

Several factors stand out in the regressions. First, when multivariate regression is carried out with a very large number of observations (N), significant relationships between the independent and dependent variables are expected. It is almost more informative when a relationship is *non-significant* in this case, thus showing almost definitively that there is no relationship between the explanatory variable in question and the dependent variable. With this in mind, it is certainly noteworthy that, among a sample of 33,540 across 18 European countries, the ‘minority’-type characteristics, such as the indicator for Muslim religion, whether the respondent has a different first language than that of the official regional language or whether the respondent is a woman, have very little systemic impact on perceptions of QoG. Regarding the latter, women tend to think that the quality of services is lower on average, but that the public sector in their region is more impartial, as compared with male respondents. There is no significant difference in gender in attitudes about corruption or a combined measure of the three pillars, called the ‘QoG Index’. Similarly, no significant difference in the respondents - Muslim or those with a minority mother tongue – regarding the level of quality of services or impartiality. However, we find that that Muslims and those with a minority mother tongue perceive corruption in their services to be higher and the overall QoG in their region to be slightly lower than respondents typically associated with the ‘majority group’.

The demographic characteristics that stand out as having the most systemic impact on perceptions of QoG based on Table 3 are *population* (in particular respondents in large cities), *income* and *age*. There are several factors that lead to very straightforward interpretations. The most negative respondents by far throughout the survey were citizens from large cities; thus we will expect regions with large urban areas, or regions that are in fact cities, such as Inner London or Prague, to have lower QoG scores on average than other regions in their country. On the other hand, the most positive respondents were persons over 65 years and middle, and especially high, income earners (relative to respondents under 30 and low income earners, respectively). The respondents perceptions toward education is slightly more difficult to interpret, with attitudes toward quality of services being a negative function of education, while those with higher levels of education tend to think that their regional public sectors are more impartial and less corrupt than respondents with a secondary education or less.

Moving to respondents who experienced the three public services in question, we find several interesting results. One, there is an apparent gap between the perception and ‘reality’ of two services in particular – education and law enforcement service. The perception of the former (that is, of those without direct experience in the past 12 months) is that aspects of QoG are generally higher than ‘reality’ (measured according to those with direct experience in the past year). What we find is that respondents who have had

⁷⁴ A full table of the regressions is given in the appendix.

direct recent contact with their education system generally rank it lower with respect to quality, impartiality, corruption and overall QoG, and the relationship is significant at the 95% level of confidence or higher in each model. With respect to law enforcement, the perception works in the opposite direction. Those without direct contact in the past 12 months generally perceive this particular service as worse as regard to QoG than those who have had contact with law enforcement recently. The relationship is at the 99% level of confidence in each model. With respect to the health care services, we find that those with experience believe that this public service provides slightly less overall quality than those who have not used their health care services recently; however, those with direct contact perceive health care services as being slightly less partial than those with no experience. In corruption and overall QoG, there is no significant difference between respondents with and with direct contact in the past 12 months.

Finally, as anticipated, satisfaction with the current state of the economy and government is positively related to attitudes about QoG. In other words, citizens that are optimistic or satisfied with the present state of the economy and politics in their region believe that their regional public sectors are more impartial, less corrupt and have better quality services than respondents who are less satisfied.

3.2 Building the QoG Regional Index: Multivariate Analysis

3.2.1. Correlation Matrix

We begin by running a simple correlation matrix using the 16 underlying questions (indicators) of QoG. In essence, since they are meant to be components of the same overall concept, we would hope that they all correlate with one another in the same basic direction. Table 4 shows the correlation matrix (bold numbers represent a significant level of confidence of at least 95%). Since all variables have been adjusted so that higher levels of each indicator mean a higher QoG, we expect positive correlations. In fact, this is exactly what we find. With only two exceptions (the insignificant and negative relationship between quality of education and impartiality of law enforcement and between general bribery and impartiality of education) we find that all 16 variables have a positive relationship to one another. Moreover, we find that the average pair-wise covariance is 0.48 among the indicators and that 89% of these pair-wise correlations are significant at the 95% level of confidence. The most obvious exception is the question pertaining to whether the respondent has ever had to pay a bribe to obtain any type of public service in his or her region, which is based on a skewed variance. Of the 172 regions, 44% had at least one respondent who had paid a bribe for a service, while 56% of the regions in the study reported having no bribery at all.

One further test of the internal consistency of the underlying data is Chronbach's Alpha coefficient (Chronbach 1951)⁷⁵. This test produces a statistic (Alpha coefficient)

⁷⁵ See the appendix for the formula for Chronbach's Alpha.

which one can use to measure how closely related a set of individual variables are as a group. In our case, the Alpha coefficient of reliability is 0.94. While the cut-off reliability threshold of the coefficient varies from discipline to discipline (from a lenient 0.6 up to a more stringent 0.8), the .936 coefficient exceeds all accepted reliability thresholds. Therefore, upon taking a closer look at the basic underlying structure of the data, we felt it to be entirely appropriate to proceed further using these 16 QoG indicators.

3.2.2. Factor Analysis

Factor analysis (FA) is intended to “describe a set of Q variables x_1, x_2, \dots, x_Q in terms of a smaller number of m factors and to highlight the relationship between these variables” (OECD 2008: 69). In essence, we are assuming that there are underlying groups (factors) in the data into which we can combine individual indicators in order to form several larger factor groupings⁷⁶. We thus seek to discover which of the 16 indicators have a common enough variance according to FA to be considered appropriate to be placed in a single grouping. We use the common method of Principle Component Analysis (PCA) to derive the number of appropriate principle components (e.g. number of factors in this case). We follow the standard rules of thumb as a ‘cut-off’ measure in order to be as pragmatic as possible without losing too much information in that data. The Kaiser criterion suggests that a factor must have an Eigenvalue of at least ‘1’, while many experts suggest that such a factor also explains at least 10% of the total variance. We follow both rules of thumb here.

Table 30 : Factor Loadings of 16 Indicators (Principle Component, Varimax Rotation)

EdQual		0.69
HealthQual		0.51
LawQual		0.89
Media		0.67
Election		0.48
EdImpartial1	0.93	
HealthImpartial1	0.80	
LawImpartial1	0.91	
EdImpartial2	0.79	
HealthImpartial2	0.52	
LawImpartial2	0.91	
EdCorrupt	0.67	
HealthCorrupt	0.80	
LawCorrupt	0.38	
OthersCorrupt	0.76	

⁷⁶ See the appendix for a formalized version of the FA analysis

GeneralBribe	0.53		
Expl. Var	0.33	0.25	0.17
Expl. Var/ Total	0.4177215	0.3164557	0.2151899

Of the potential 15 principle components discovered in the underlying data, three independent factors are revealed by the FA, all meeting the Kaiser criterion and explaining a suitable amount of variation⁷⁷. Together, the three factors explain over three-fourths of the variation in the data. We employ Varimax rotation in order to see what variables load onto which factor. Table 29 shows the rotated factor loadings for the individual 16 QoG indicators. While several indicators load positively onto multiple factors, we present a ‘cleaner’ table, with each variable loading onto only one factor.

The block of six impartiality variables load cleanly onto one principle component (Impartiality), which explains roughly 42% of the total variance in the model. The next most consequential block of variables is the set of corruption questions (Corruption), which together explain approximately 31.5% of the total variance in the model, while the third group (Quality) has a combined total variance of about 21.5%. We label these three factors the three ‘pillars’ of the regional QoG index. One variable in particular, the ‘elections’ indicator, could also have loaded onto the corruption pillar, as its factor loading was only slightly higher on the quality factor. We therefore test the sensitivity of the aggregated data adjusting for the elections variable moving from the quality pillar to the corruption pillar.

We perform one final stage of the FA, which is in a sense a reverse FA from the one presented above. We check the internal consistency of the three pillars individually to see whether running the five to six indicators by pillar results in one clear factor.

Table 31: Factor Analysis of Individual Pillars

	Quality		Corruption		Impartiality	
	Eigenvalue	Prop.	Eigenvalue	Prop.	Eigenvalue	Prop.
Factor 1	2.95	0.59	3.6	0.72	4.67	0.79
Factor 2	0.78	0.15	0.71	0.14	0.59	0.1
Factor 3	0.71	0.14	0.35	0.07	0.47	0.07
Factor 4	0.31	0.06	0.23	0.05	0.16	0.03
Factor 5	0.08	-	0.08	0.02	0.05	0.009
Factor 6					0.02	-

⁷⁷ The results of the original PCA are in the appendix

Table 7 shows that each pillar has a high degree of internal consistency, with one dominant factor with an Eigenvalue over '1' that explains about 60% of the total variation within the pillar alone.

3.3 Normalization of the Data

Because several of the questions have different scales in the original data, we must of course find a common scale for all indicators so that we can aggregate them into a single indicator. Two methods appear to be the most appropriate based on the data that we have – standardization and 'min-max'. For the former, we calculate the mean of each individual indicator across all EU regions as well as the standard deviation⁷⁸. The minimum and maximum values of each indicator, I_{qc}^l , are thus different depending on the level of variance and the distribution of responses that that particular indicator has across the regions. However, each indicator ends up with the same mean of '0' and standard deviation of '1'. In the second method, the 'min-max' normalization, the range is transformed for each indicator such that the lowest value is assigned '0' (\min_c) and the highest value '1'. We believe that the method of standardization has several advantages. For one reason, it renders direct comparison quite easy between regions, as we can say that region 'X' is a full standard deviation higher in QoG than region 'Y' by simply glancing at the data. Two, the range in variance is allowed to be a little more extensive in standardization than in min-max. However, the choice is still somewhat arbitrary, and thus we test the sensitivity of our choice of normalization in the section dealing with uncertainty and sensitivity testing.

3.4 Aggregation of the Data & Weights

There are essentially two stages of aggregation in these particular data, one that takes the individual level data and averages them to the regional level and a second that aggregates the 16 indicators into a single QoG index score for each region. The multivariate regression from the first part of this section was designed to give the user insight into the ways in which certain demographics systematically respond (if at all) to questions about QoG in different ways. This analysis can also be made at the individual level of each country, of course. Since we do not have perfect information about each region's distribution of education levels, income, gender, age etc., we aggregate the individual level data to the regional level using simple, mean average responses. Practitioners and researchers alike are certainly welcome to weight certain regions with various demographic characteristics in the future but, for the purposes of this study, we take a very parsimonious approach as we do not want to 'guess' too much about each region's makeup, which could possibly lead to 172 different demographic weighting schemes. At the regional level, we then normalize the data by standardization and aggregate the data

⁷⁸ The formula is given in the appendix.

into the three pillars and finally into a single composite index. The three pillars are each weighted $1/3^{\text{rd}}$ and combined into a single QoG index for each region. Although we test the sensitivity of the index using factor weights, we decided to keep the method as parsimonious as possible and apply an equal weighting scheme (EW) to the pillar level ($1/3^{\text{rd}}$ weight to each of the three pillars). Since the number of variables in each pillar is uneven, there will be a degree of weighting, regardless (i.e. if each variable were weighted $1/16^{\text{th}}$, impartiality would be weighted more heavily). We find that the difference is negligible, but we use the EW at the pillar level for the final QoG index data.

The relationship between each of the three pillars and the combined QoG regional index is shown in Table 31.

Table 32: Person Correlations between the QoG Regional Index & Three Pillars

Variable	QoG Reg.Index	Quality	Impartiality	Corruption
QoG Regional Index	1			
Quality	0.81	1		
Impartiality	0.84	0.44	1	
Corruption	0.90	0.65	0.71	1

All correlations in Table 31 are significant at the 99% level of confidence. Looking at the Pearson correlations between the three pillars and the QoG regional index itself, we find that the correlations range between .81 and .90. The relation between all other pillars is positive and significant, as expected, at .65 or better between the corruption pillar and the other two. However, the correlation is not quite as strong between the quality and impartiality pillars. A closer look at the data shows us that the majority of the Spanish and Romanian regions drive this weaker correlation. Most of the Spanish regions rank among the top 20% with regard to the impartiality of their public services but, when it comes to quality, citizens rank their regions much lower, so that most of the Spanish regions are around the 50% percentile in this pillar. The opposite is true for Romania, as respondents in several of the regions here placed the quality of their services very high but do not consider them to be entirely impartial. A visual of the scatter plot relationships from Table 8 is shown in the appendix.

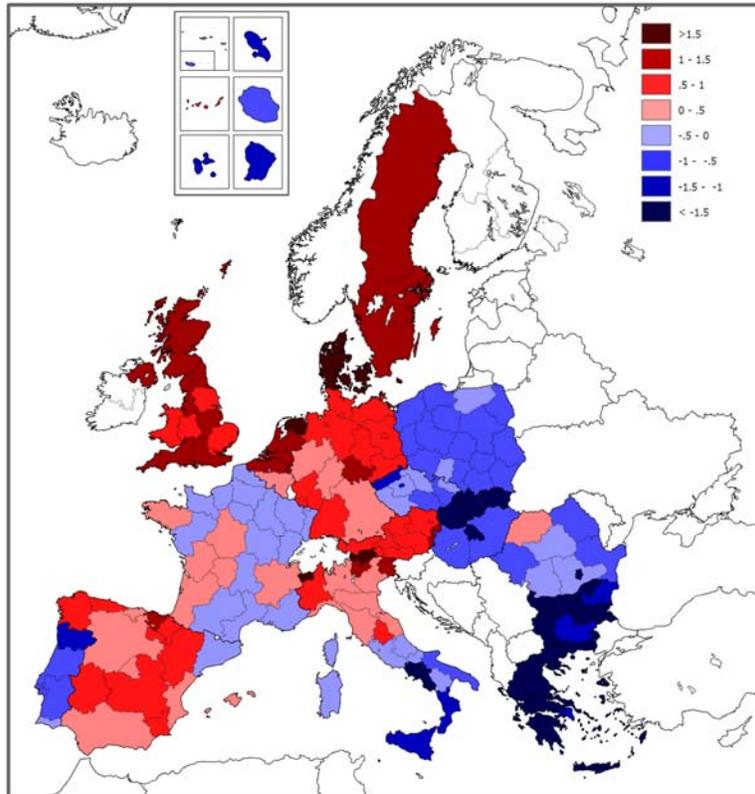
In applying factor weights at the pillar level and to check the internal consistency of each pillar, we employ the FA (principle component) from the multivariate section. We square the rotated factors and weight each indicator based on these results. We also show the resulting weighting scheme from the FA along with the EW scheme for each indicator⁷⁹.

In keeping with our desire to construct the regional QoG indicator in a parsimonious way, we decide on the common ‘linear aggregation method’ in combining the indicators

⁷⁹ The detailed weighting scheme is given in the appendix.

into a single number. We therefore take the summation of the weighted (EW and FA in several sensitivity tests) and normalized indicators (standardization, although we check for min-max in sensitivity tests)⁸⁰.

Figure 20: The Aggregated Data: QoG at the Regional Level



After aggregating and standardizing the data for each survey question, we obtain 16 QoG indicators (which then combine to be in the three pillars) for each region. Table 10 shows a complete list of the 172 regions in the study according to their final QoG index rankings. Country names represent the average scores of all regions for the particular country included in the data. The data have been standardized such that the EU regional mean is '0' and has a standard deviation of '1'. For example, Midtjylland in Denmark (DK04) is ranked as the top region in the EU with respect to overall QoG and is 2.13 standard deviations over the average of all 172 regions, while Severozepaden in Bulgaria (BG31) is ranked lowest, with a QoG score of -2.96, or just over three standard deviations below the mean EU score. Thus the data span over five standard deviations, from rank number 1 to rank number 172⁸¹.

⁸⁰ The formula is given in the appendix.

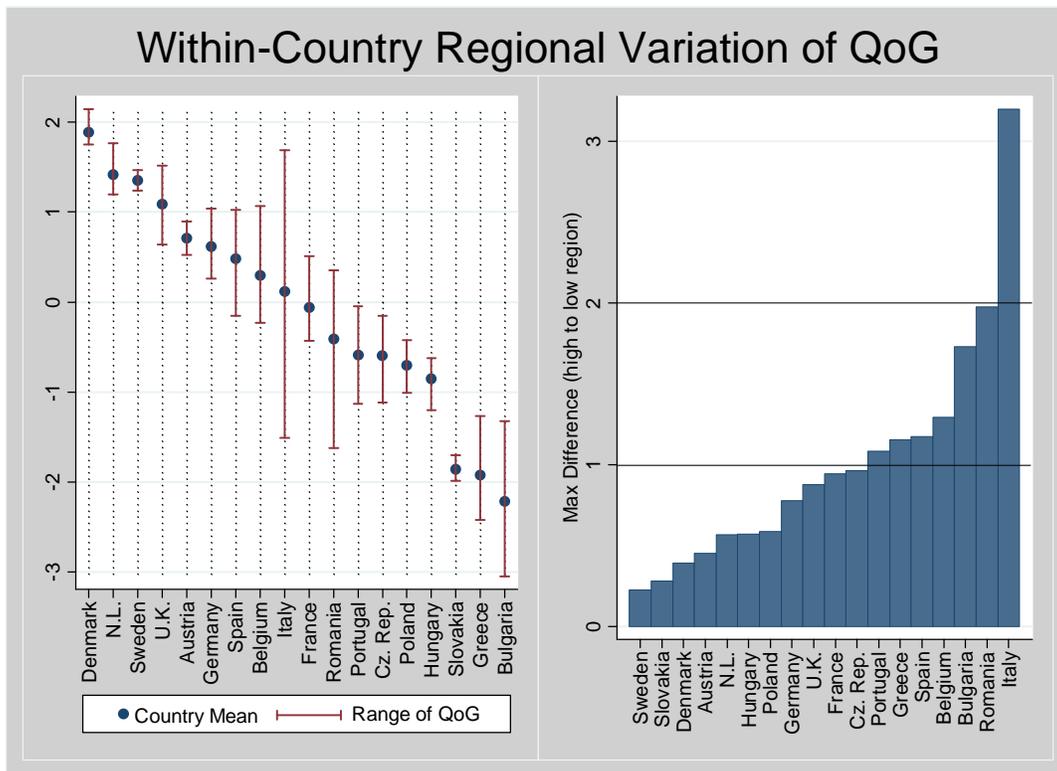
⁸¹ A complete list of all regions in rank order is given in the appendix.

A quick examination of the data reveals fairly predictable patterns among the regions with respect to QoG. All regions in the top performing EU members with regard to the national QoG index (Denmark, Sweden and Netherlands) are in the top 15% of all 172 regions. Among the NMS, all but two of regions are in the bottom 50% (i.e. have a score lower than '0'), with the only exceptions being Nord Vest (0.35) and Sud Vest (.06), both in Romania. The next three highest regions among the NMS are Centru in Romania and Severovýchod and Jihozápad in the Czech Republic (all just under the mean EU average). On the other hand, most of the EU 15 regions are in the top 50%, with Portugal and Greece being the only exceptions that have all of their respective regions under the mean average. Moreover, several regions in France and Italy are under the EU mean, with the latter having two in the bottom 10%.

3.5 Within-Country Variation

As regards within-country variation, it is clear that this varies considerably from country to country. Some countries display very tight regional groupings, while others display a remarkably wide range in QoG across their given regions. Figure 20 shows the average country score; here, the regional variation is on the left (i.e. between the two most extreme regions) in the 18 countries and the absolute distance (in standard deviations) from the low to high regions in each country is shown on the right.

Figure 21



Clearly, Denmark, Sweden, Slovakia and Austria have very little variation, as all of their regions are roughly 0.4 standard deviations (or less) of one another in the data. Hungary, Poland and the Netherlands have slightly larger extreme values among their regions, at about 0.6 standard deviations, yet still have relatively tight regional groupings. States with a moderate variation are Germany, the U.K, the Czech Republic and France, with values in the data among their regions that span between a standard deviation of 0.8 and 1 throughout the QoG index. Portugal, Spain, Belgium, Romania, Slovakia and Bulgaria all have a within-country variation between 1 and 2 standard deviations in the data. The extreme value range in Greece and Bulgaria is perhaps not as interesting as the other four, however, because all of the regions in these two countries are clustered in the bottom 10% of the 172 regions. In contrast, Belgium, with only three regions at the NUTS 1 level, proves to be an interesting case because it demonstrates that wide within-country variation does not depend exclusively on a large number of regions. Belgium has an average score of roughly .30 but has regions that range from under the EU mean of -0.24 (Brussels) up to 1.16 (Vlaams Gewest), the latter of which is in the top 15% of all the regions in the sample with respect to QoG. We find that Romania has the most within-country variation among the new member states (NMS) and has one region (Nord Vest) with a score of 0.22, which puts it in the top 45% of all regions, while its lowest ranking region, Bucharesti, has a score of -1.67, which places the latter region in the bottom 10%. Italy displays far and away the most within-country variation among the 18 EU countries in the sample according to the data, with regions spanning over three full standard deviations. Although Italy's mean score is just above the mean EU regional average at 0.11, it has three regions in the top 10% of all 172 regions (Bolzano, Valle d'Aosta and Trento) and two in the bottom 10% (Campania and Calabria). Figure 3 (in the appendix) demonstrates the within-country variance that we find in Italy, with its regions spanning over three standard deviations in the data. The line between regions under the mean average of the 172 EU regions and those that have positive scores is very distinct – all regions from Umbria and farther north have better QoG than the average EU region, while those from Lazio and south rank below average.

What we can understand in the within-country regional variation is as follows. For countries with low regional disparity in QoG, the national level indicators certainly tell a fairly accurate story, as they do not overlook a great deal of within-country variation. However, for the countries that contain high levels of within-country QoG variation, the national level indicators clearly do not elucidate the whole picture, as they tend to underestimate well-performing regions and overestimate poorly performing ones, as is the case with Portugal, Spain, Belgium, Romania and, particularly, Italy.

3.6 Margins of Error for the Regional QoG Estimates

The regional QoG index is based on data from a randomly selected group of respondents in each of the 172 regions. We thus do not claim to report the 'absolute'

value of QoG in any given region but rather an estimate of the total population. On the basis of the sample size of each region, we can use a fairly simple technique to calculate what is known as the ‘margin of error’. Although, in theory, any number can be chosen, we select a margin of error at the 95% confidence level. This level equates to the probability that a margin of error around the reported QoG estimate for each region would include the "true" value of QoG or, in other words, that the margin of error tells us that we can say with about 95% confidence that a region’s estimate of QoG can be found within a plus/minus 1 margin of error, which is inversely related to the sample size. After obtaining the margin of error based on our sample size, we then can calculate the distance around the estimates of QoG for each region.

To be precise, there are two ways to go about calculating the margin of error for survey data – an ‘exact’ confidence interval and an ‘approximate’ confidence interval. The former takes into account both sampling and non-sampling errors, while the latter only random sampling errors. An ‘approximate’ confidence interval, while much easier to calculate, does not take into account potential problems such as respondents lying in their answers, poorly phrased questions, exclusion of certain people who could not be contacted etc. Thus it essentially tries to elucidate the likelihood that the estimate obtained from the population might be off because of ‘pure luck’ (or lack thereof). While the ‘exact’ interval may be more precise, we find the advantages of the ‘approximate’ confidence interval to far outweigh the drawbacks, in particular with respect to the efficiency and time saved in the calculation. Moreover, we have no reason to suspect that there is any bias in certain groups being excluded or not being forthright in their responses, so compensating for such error is simply beyond our reach. Thus we report an ‘approximate’ confidence interval for each region’s QoG estimate⁸².

Three of the regions that showed the largest margins of error are Bucharesti (0.69), Sud Vest (0.55) and Galicia (0.52), which means that they all have a confidence interval that spans over a full standard deviation in the data due to relatively high/low scores compared with their final QoG regional index estimate. The lowest margin of error belongs to the Polish region of Kujawsko-Pomorskie (0.166), meaning that we can be about 95% sure that the true estimate lies within a confidence interval of about one-third of a standard deviation around our current QoG estimate for this region. The next figure looks at the regional variance in Spain, which has regions with high margins of error due to a general trend among respondents to rate their region very high with respect to impartiality questions, although relatively low regarding quality of services.

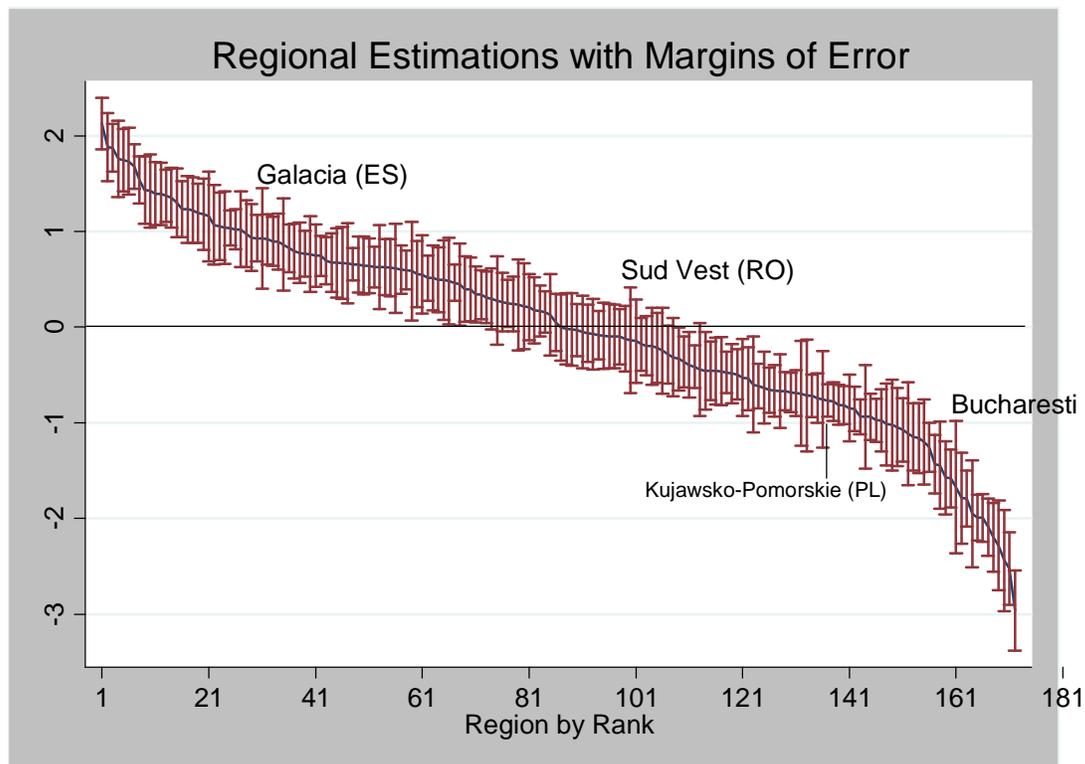
We also look at the significance of regional differences within each country based on margins of error. For example, in Italy, we observe significant differences at many levels, in that the set of very well performing regions, namely Bolzano, Valle d’Aosta and Trento, have estimates that are significantly higher than the fifth highest ranked region (Piemonte) down to the 21st ranked region. Further, we can say with 95%

⁸² A detailed explanation of the construct of the margins of error is given in the appendix.

certainty that the estimate for Toscana, which has a score about 0.25 over the EU mean and about 0.15 standard deviations above the Italian mean, is significantly higher than Molise and all regions ranked below it, although it cannot be statistically distinguished from other regions such as Lombardia or Abruzzo.

However, for a country like Denmark, there is no significant within-country variation. All five regions are clustered around the national average (1.87) and the highest region, Midtjylland (DK 04), with an estimate of 2.13, is indistinguishable from the lowest region's estimate, Hovedstaden (DK 01) at 1.73. Thus, in some countries, we find some or many significantly different regions based on our margin of error estimates while, in others, like Denmark, none of the regions can be significantly distinguished from one another⁸³.

Figure 22



3.7 Internal Consistency Checks of the Data

3.7.1 Sensitivity and Uncertainty Tests

In this section we test to see whether any changes of the preceding decisions that have been made in constructing the QoG regional index impact the data in a significant way. Table 12 displays 62 simulations in which alternative ways are tested in building

⁸³ A visual of the rank order and margins of error for Italian and Danish region is given in the appendix.

the QoG regional index⁸⁴. We report the median and max changes for each simulation, the region that is affected most by the alteration and the Spearman rank coefficient. Although many more combinations could indeed be run, we proceed in the first section with some basic tests, beginning with removing whole pillars, switching the weighting scheme from EW to FA, changing the normalization method from standardization to min-max and aggregating the data with a geometric formula rather than an arithmetic one. In scenarios 2-20, we find that the Spearman rank coefficient never drops below .919 (scenario 14), which is remarkably stable. The highest median and max changes in the overall rankings are 14 and 57, respectively, which occur when we use arithmetic aggregation, FA weighting and min-max standardization and exclude the entire corruption pillar. Looking at the regions that are most impacted by changes in the assumptions of the index, we do not find in scenarios 2-20 that any one region is singled-out, so to speak. It would appear that the region of Athens (which shifts upward between 51 and 57 spots in the rankings) benefits when either corruption or impartiality is removed, depending on the weighting scheme. Several regions from both Spain and Romania are often times quite vulnerable to different scenarios in numbers 2-20, including Madrid, Cataluña, Valencia and Sud-Vest.

In scenarios 21-52, we check to see whether any one of the 16 indicators drives any significant change when removed from the overall index. For the sake of time and space, we test the exclusion of each indicator with alterations only in weighting schemes – EW and FA. In these simulations, we find that no one particular indicator impacts the data in any meaningful way, as the highest median shift of 172 places in the rankings is only 6, while the greatest max change is 36 (in scenarios 23 and 30, respectively; both times the Czech region of Moravskoslezsko shifts upward). The Spearman rank coefficient that indicates the furthest distance from the original ranking order is .984 in scenario 51, when FA weighting is used and the indicator measuring how corrupt ‘other citizens in the region’ are perceived to be is excluded. However, the vast majority of scenarios in this section are over .990. Depending on the weighting scheme used in scenarios 21-52, only certain countries’ regions are impacted by the exclusion of individual variables. In the EW, we find that several Czech regions benefit when we leave out certain indicators, one of them being Moravskoslezsko, and that several regions in Spain are also consistently impacted (positively) by these changes, including Asturias, Madrid, Cataluña, Castilla y León and Castilla-La Mancha, which usually shift upward between 13-22 places. Portugal is another country with a number of regions that are impacted by changes in scenarios 21-52, yet almost always in a negative way, meaning that, while several regions in Spain seem to be at a disadvantage in the relative rankings of the original QoG index, the Portuguese regions mainly benefit. Moving to the FA weighting scheme, in particular in scenarios 37-52, we find that five of the seven regions in Portugal are particularly volatile, and only two (Norte and Centro) are relatively stable. In FA scenarios 37-52, no

⁸⁴ Table 12 presents the full results of the test and is given in the appendix.

other country is seemingly as systematically impacted as Portugal. In fact, the majority of countries' regions, such as Slovakia, Poland, Greece, Denmark and Austria, are remarkably stable, having no region move than even nine places in the rankings.

The third section of uncertainty tests deals with the adjustments in the demographics of the respondents. Here we try to elucidate how much the exclusion of certain age groups, income brackets etc. impact the overall rankings (e.g. if some regions that had a disproportionate amount of a certain demographic were advantaged or disadvantaged in the final rankings). We run each simulation with both EW and FA weighting schemes. In 53 and 54, we include only respondents that had used two or three of the three public services inquired about in the survey to see if those without recent experience have negatively or positively impacted certain regions. We find that many regions in Spain benefit when only 'experienced' respondents are included, pointing to a gap between the 'perception' of QoG in many regions in Spain and actual experience. It would seem that the reputation in many regions is worse than the actual services in these regions. However, the overall Spearman rank coefficients are quite strong (not going below .94) and the median changes in these simulations are only 7 and 8, respectively. In the next two simulations, we remove male respondents, as we learned from the multivariate regression that, on several occasions, women answered systematically differently than men, and thus regions that might have been oversampled by females (or males) might be impacted either positively or negatively. We find very little difference in scenarios 55 and 56, with the mean changes being 6 in EW and 8 in FA. Both Spearman rank coefficients are well above .90 as well. Only in a number of regions in Italy, Spain and Portugal do we see fairly significant downward changes when men are taken out, as there are no systematic country-level effects in the other 15 EU states.

In the remaining six simulations, we remove high-income earners, respondents under 50 years and anyone attending university (or higher). There are very few changes in scenarios 57 and 58, although Östra Sverige in Sweden and a few other regions in the U.K. dropped when high income earners were removed. Removing citizens 49 years and younger benefitted Sud Vest and Nord Vest in Romania somewhat significantly, as they both increase at least 40 positions and, in the latter case, 60 or more positions in scenarios 59 and 60. In the final two simulations, when only respondents with a secondary education level of education and lower were included, we find that several regions in Poland jump in the rankings; thus we can make the interpretation that, in Poland, the gap in opinion regarding QoG of regional services between those with lower and higher levels of education is wider than in other countries. Unfortunately, we were unable to remove certain population demographics because some regions are entirely rural or metro-urban and would thus have been left with few to no respondents.

Overall, what we learn from scenarios 53-62 is that there are particular countries in which the gap in opinion regarding QoG is fairly significant between different demographic groups, be it age, gender, income etc. However, in the majority of countries

(and regions), such significant demographic differences do not exist, as evidenced by relatively small median changes and Spearman coefficients. This shows that, despite the fact that some regions might have over/under sampling issues on certain demographics, such issues are not so severe as to significantly alter the final QoG regional index.

Figure 23

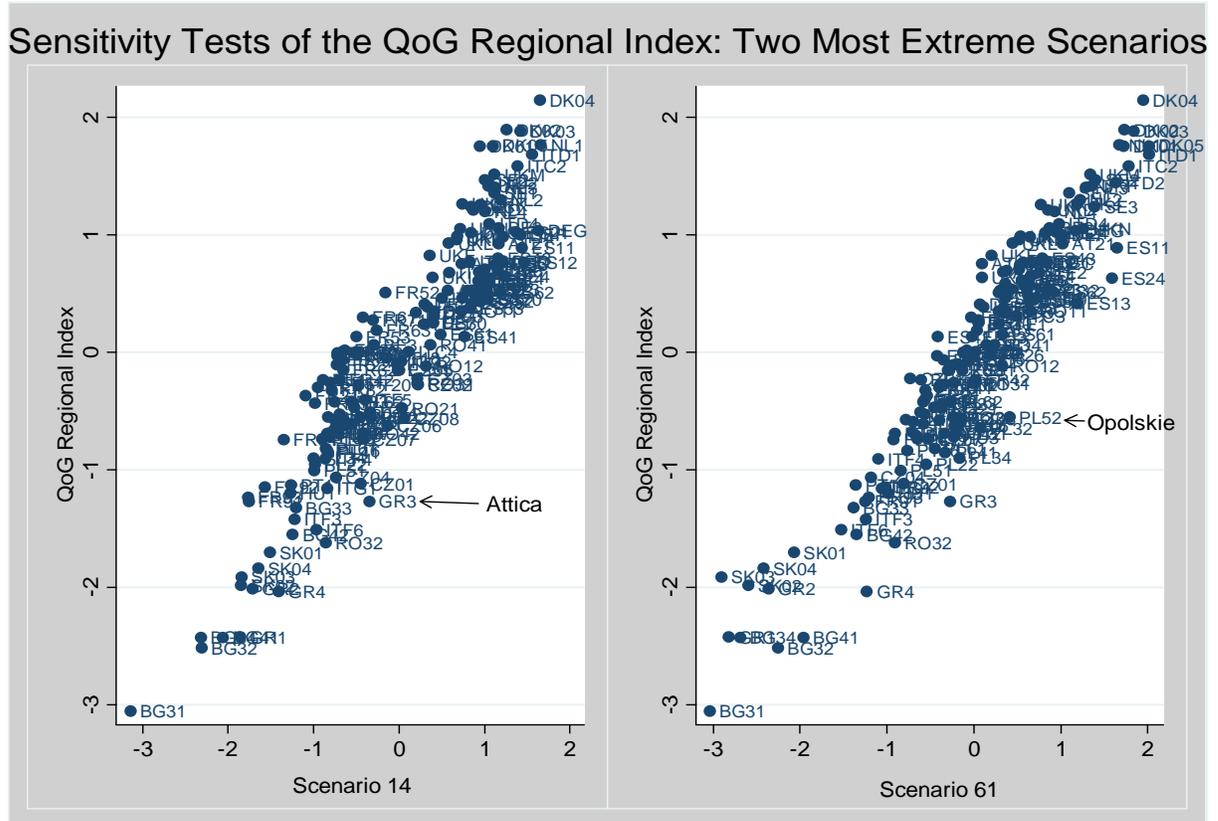


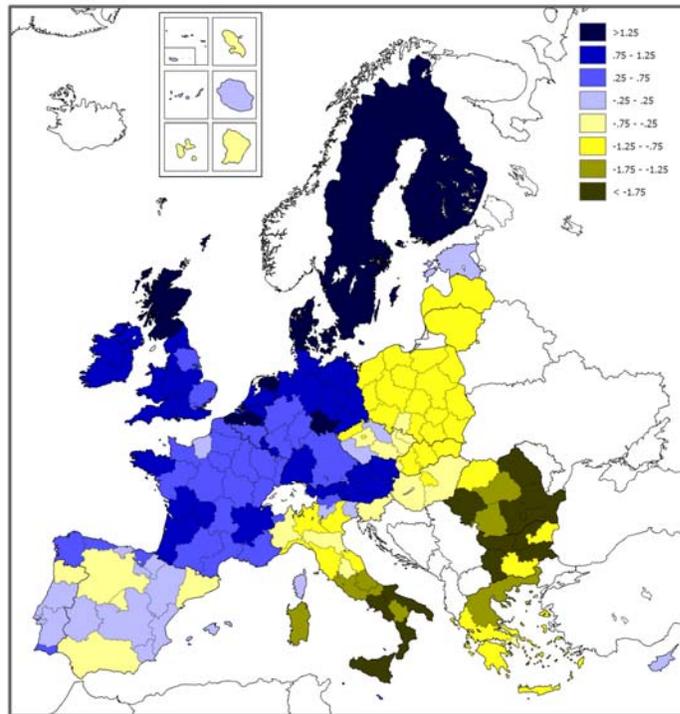
Figure 23 provides a visual of the two most extreme cases from the original QoG regional index among the 62 simulations. A perfect ‘1’ in the Spearman rank coefficient would of course look like a perfectly straight line from the bottom left corner to the top right corner of the graph. Scenario 14 (FA weighting scheme, min-max normalization and leaving out the corruption pillar) and scenario 61 (EW, including only respondents with relatively low education) revealed the two lowest Spearman rank coefficients (.919 and .921, respectively). Scenario 14 had the highest median shift (14 places) while scenario 61 the greatest single region shift (68 places). As we can see, even the extreme outlier cases (Attica in #14 and Opolskie in #61) do not look to be so out of line from their original position to cause alarm, as both increase roughly one full standard deviation to end up just over the mean score of ‘0’.

3.8 Combining the National and Regional QoG data

Here we take both sets of data already discussed – the first at the national level from the World Bank and the second from regional surveys. It is our aim to propose a method that mostly fairly includes the EU countries omitted from the survey while simultaneously maintaining the richness of the within-country variation in several of the countries surveyed. To make the comparisons most meaningful, we employ the following formula to calculate each unit's adjusted scores (whether regional or country):

$$WBCountryScore_{CountryY} + (QoGScore_{regionXinCountryY} - CountryScore_{CountryY})$$

Here we essentially take the country average from the four pillars (standardized) from the World Bank data, standardize the combined total within the EU and set each country's average score to that level, so that Austria's aggregate score can be directly compared with Latvia, even though the latter was not in the regional survey. For countries outside the survey, there is nothing to add to the WB country score; this is therefore used as the estimate, and regional variation is unobserved. Thus only the within-country variance is explained by the regional level QoG data; for countries that have a significantly higher or lower average on the regional level data as compared with the WB national level, this method thereby adjusts the placement of each country with the inclusion of the national, mostly 'expert survey'-driven data. We feel that this method makes the best use of both data while combining our citizen survey with the opinions of the main experts, NGO and IGO data of the World Bank. The regional 'country score' is simply an un-weighted mean of the aggregated regional scores for each adjusted pillar. After calculating the scores for all of the regions, we standardize them one more time. Figure 24 shows the combined data for all four pillars of QoG:

Figure 24: 4 Pillars of QoG Combined with Regional Data

The advantage of combining the full national QoG index with the total regional QoG index is that it allows us the ability to compensate for the fairly limited amount of respondents per region with the assessments of numerous international investment firms and NGO's, which focus on national-level assessments. Moreover, using this method allows us to keep the margins of error for each of the regional estimates. A complete list of regions and countries in rank order along with their margins of errors, and a visual figure of these estimates is located in the appendix, while a figure showing the complete national and sub-national variation is found on page 41 of this report.

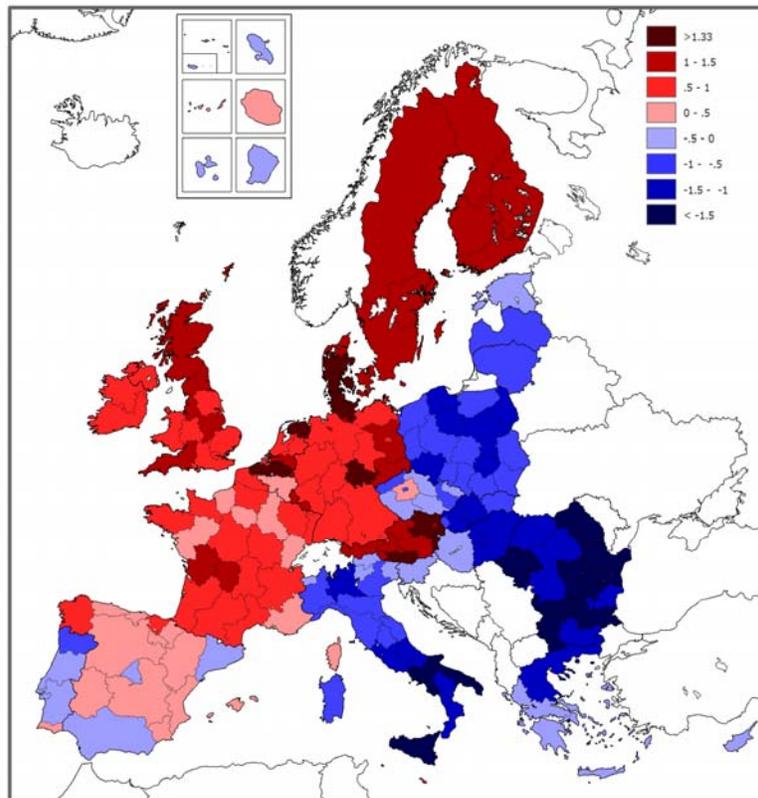
We now turn to how this method could be used to combine the four individual national level pillars with regional data. Although obviously not a perfect match, we attempt to match questions at the regional level with concepts that are also measured in the national level pillars. We begin with the rule of law pillar.

3.8.1 Rule of Law Pillar Combined

Although the majority of the data on the national level focuses on typically national level institutions, such as supreme courts, property rights, immigration law etc., we do find concepts such as objectivity and confidence in the police or law enforcement. Thus we take the four questions dealing with the respondents' opinions regarding the police force into account here. They are:

1. *‘How would you rate the quality of the police force in your area?’*
2. *“The police force gives special advantages to certain people in my area.”*
3. *“All citizens are treated equally by the police force in my area”*
4. *“Corruption is prevalent in the police force in my area”*

Figure 25: Combined Rule of Law Pillar with Regional Data

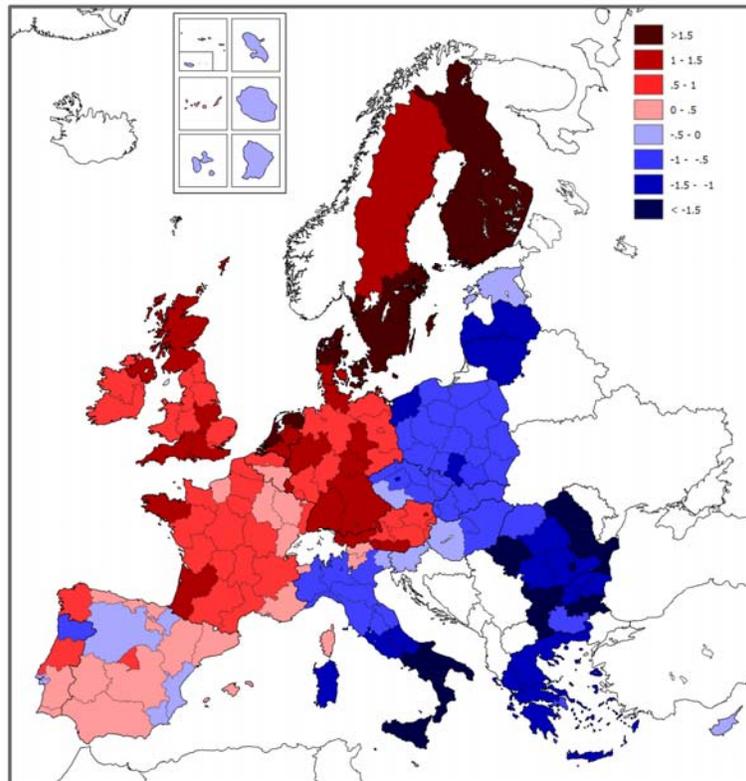


3.8.2 Corruption Pillar Combined

Next we move to the corruption pillar. We have a ‘corruption pillar’ in the regional QoG index that contains five questions. However, we have included the ‘corruption in the police force’ question in the rule of law pillar. To avoid double-counting this question, we take only the following four questions:

1. *“Corruption is prevalent in my area’s local public school system”*
2. *“Corruption is prevalent in the public health care system in my area”*
3. *‘In the past 12 months have you or anyone living in your household paid a bribe in any form to: health or medical services?’*
4. *‘In your opinion, how often do you think other citizens in your area use bribery to obtain public services?’*

Figure 26: Combined Corruption Pillar with Regional Data

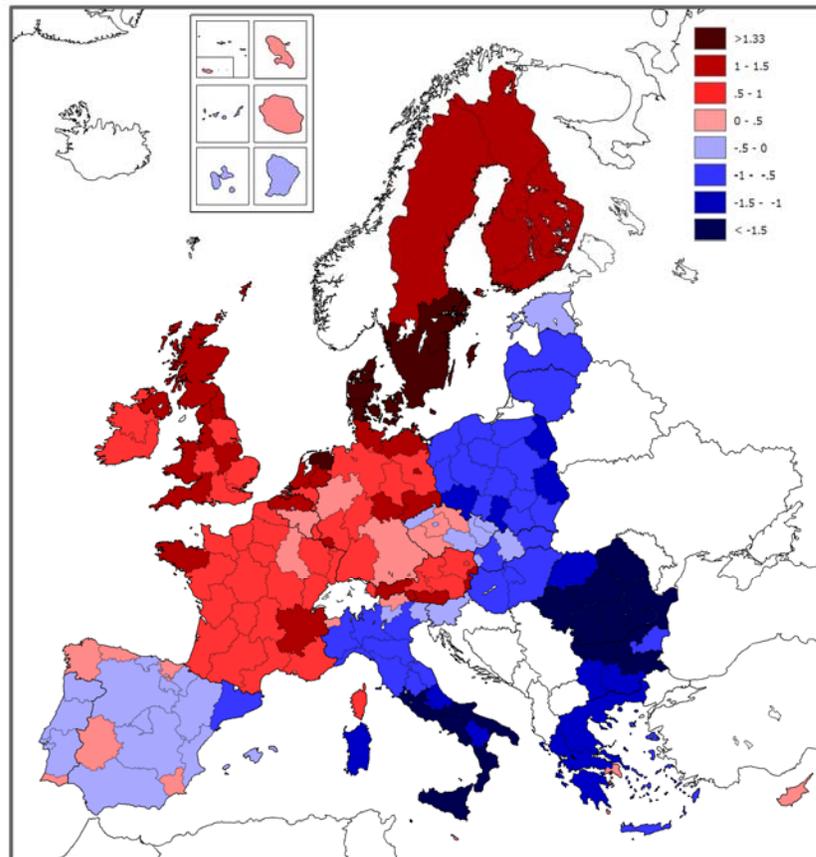


3.8.3 Government Effectiveness Combined

Government effectiveness is the next pillar that is combined, which captures a fairly wide score of concepts at the national level. ‘Quality of public services’ and impartiality of the bureaucracy are themes that come up in certain underlying data in the GE pillar. Thus we include the quality and impartiality questions about health care and education services. Again, we do not include the law enforcement questions because these have been counted in the RL pillar-combined. The following six questions are combined with the national level pillar:

1. ‘How would you rate the quality of public education in your area?’
2. ‘How would you rate the quality of the public health care system in your area?’
3. ‘Certain people are given special advantages in the public education system in my area.’
4. ‘Certain people are given special advantages in the public health care system in my area.’
5. ‘All citizens are treated equally in the public education system in my area’
6. ‘All citizens are treated equally in the public health care system in my area’

Figure 27: Combined Government Effectiveness Pillar with Regional Data



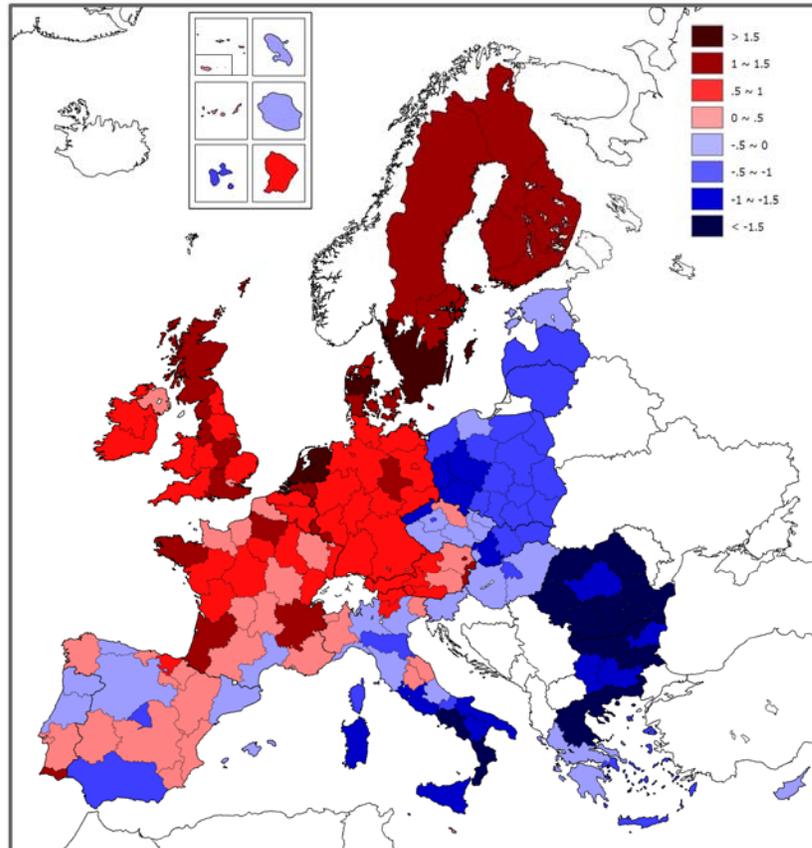
3.8.4 Voice & Accountability Pillar Combined

In calculating the combined VA index, we take the WB scores from Part II and standardize them so that the average in the EU is '0' with a standard deviation of '1'. Regarding the regional scores for the applicable countries and regions, we find that two indicators from our QoG regional data overlap nicely with the national VA data. The first is the question of the honesty of elections and the second is our indicator of the effectiveness of the regional media as a watchdog for corruption in the public sector. The World Bank VA data focus heavily on electoral integrity and media freedom; thus we are content that these two levels can be combined in a meaningful way. For a full list of the countries and regional scores for the combined VA pillar, see the appendix. Applying the above formula, we show a mapped visual in Figure () of the variance both across and within countries.

We thus use the two following questions here:

1. 'In your opinion, if corruption by a public employee or politician were to occur in your area, how likely is it that such corruption would be exposed by the local mass media?'
2. 'Please respond to the following: elections in my area are honest and clean from corruption'

Figure 28: Combined Voice and Accountability Pillar with Regional Data



3.9 External Validity Check : Bivariate relationships with QoG correlates

In this section, we report a number of bivariate regressions and produce the scatter plots for such regressions to check some basic empirical relationships between the regional QoG index and a number of other factors. With the exception of income inequality, we take all regional data from Eurostat and average them (where available) for the decade preceding the regional QoG index (e.g 1999-2008) so as not to rely on possibly misleading single years. Regional income inequality data come from the University of Texas Inequality Project (Galbraith 2009)⁸⁵. For each bivariate relationship, we report with coefficient, the level of significance and the amount of variance explained

⁸⁵ A more thorough description of the data can be found at: <http://utip.gov.utexas.edu/>.

by the correlate variable (e.g. R^2). In addition, we produce the corresponding number assigned to each region to be able to check specifically which regions are outliers etc.

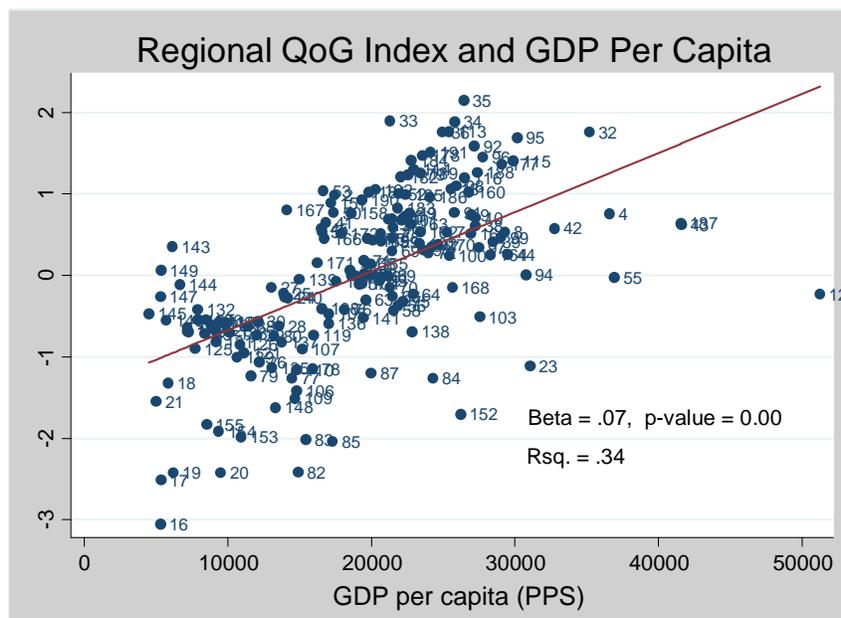
The seven correlates that we explore are:

1. GDP per capita
2. Unemployment/ employment patterns and work hours per week
3. Education level of the workforce
4. Population and population density
5. Use of technology
6. Non-resident tourism
7. Income inequality

Again, as in the bivariate analysis at the national level, we do not intend to make causal claims for any of these variables to regional QoG as this cannot be proven with simple bivariate, spatial correlations. We seek only to demonstrate that the regional QoG index accounts for clear relationships with salient variables typically found in the literature, which we argue points to a degree of external validity for the governance data.

3.9.1. QoG and GDP per capita

Figure 29



In general, the data show a wide variation in per capita income across the regions. We find that GDP per capita, as measured in Purchasing Power Standard (PPS) per capita (averaged between 1999-2008), is highly related to the regional QoG index. In addition to a significant, bivariate p-value and an R^2 of .34, the Pearson correlation coefficient is 0.58. The primary outlier in the model (region #12) is the capital region of Brussels in

Belgium, with the highest PPS, although it has a QoG rank just below the EU average. Other than this, most regions lie within the region of the 95% level of confidence.

3.9.2. QoG and Unemployment/ Employment and Work Trends

In these models, we examine the relationship between several types of unemployment and the regional QoG index. As in the case of GDP per capita, unemployment rates vary significantly from region to region. Figure 6 divides two age groups (15-24 and 25-64). We see a strong, negative relationship between both youth and general unemployment and QoG, both significant at the 99% level of confidence. In general, better governance is associated with higher levels of employment. The Pearson correlation coefficients are -0.57 and -0.44 respectively. Thus, although we see a general negative relationship between QoG and unemployment, poor governance is especially harmful for youth unemployment.

Figure 30

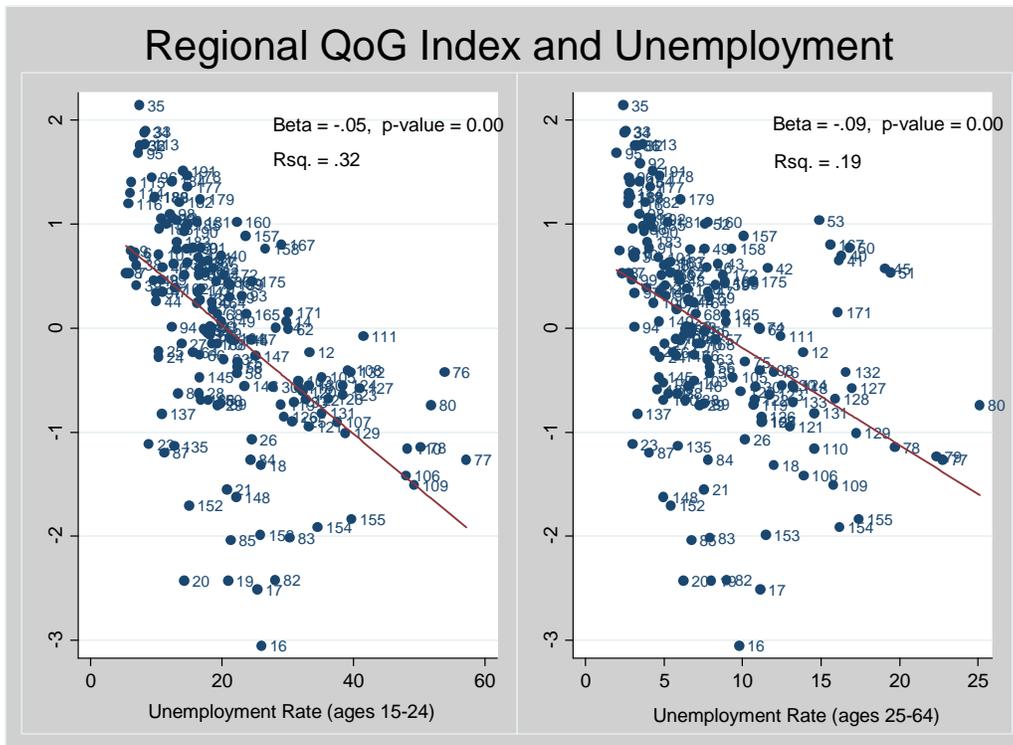
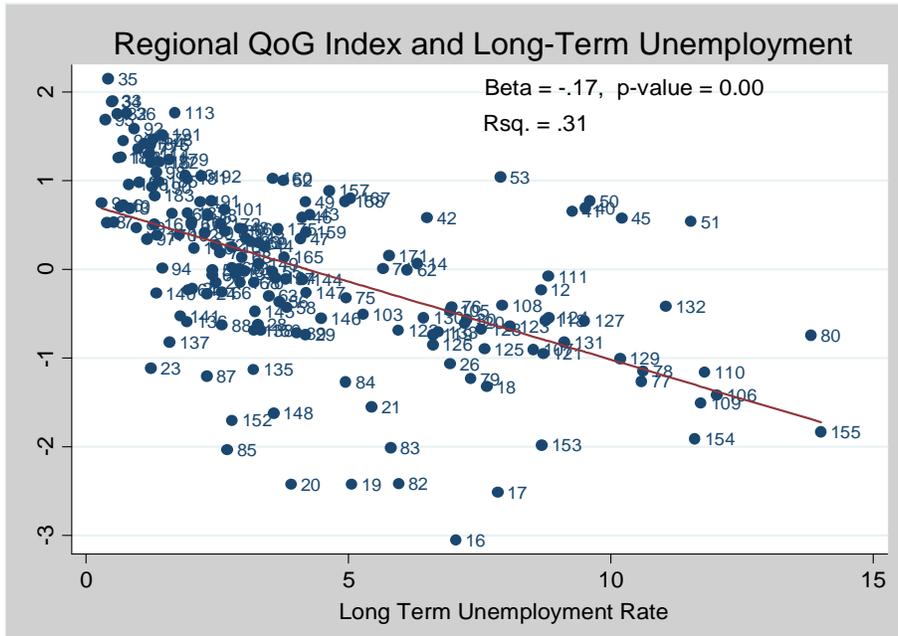


Figure 31



In Figure 31 we examine the relationship between long-term unemployment (LTU) and QoG. Again we find a significant and negative relationship, with a Pearson coefficient of -0.54 , and the LTU explains 31% of the variance of the regional QoG variable. Thus, on the basis of these initial findings, we would conclude that higher levels of quality of government are strongly associated with lower rates of unemployment, *ceteris paribus*.

Figure 32

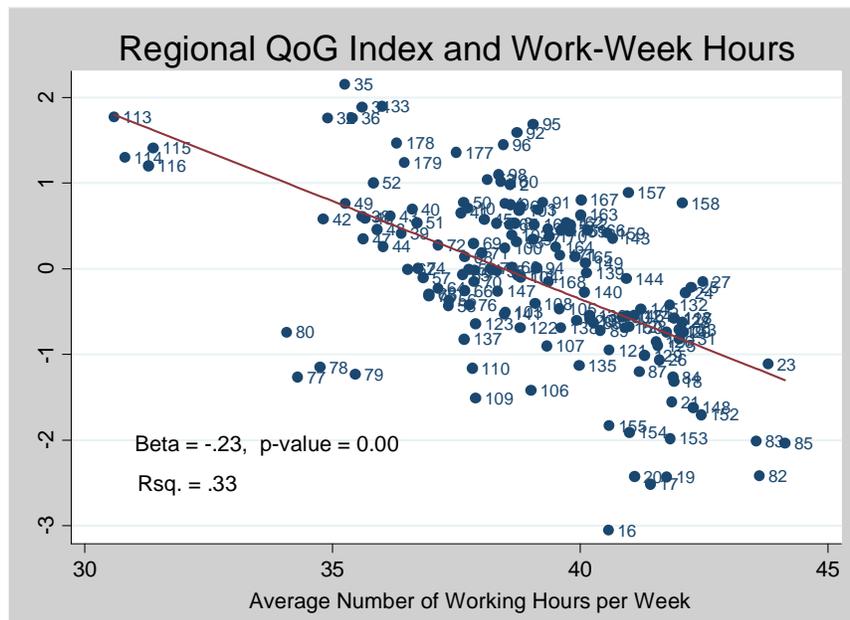


Figure 32 reports the relationship between the average number of hours worked per week (between 1999 and 2008) in the regions and QoG. The work week variable is strongly negatively related to QoG at the regional level (Pearson correlation coefficient = -0.58), as regions that report higher numbers of hours per week at work generally have lower QoG. In fact, all regions with an average of 42 work hours per week or higher are below the EU average in regional QoG. We are not sure that the lesson here is ‘work less and get better QoG’, yet there would appear to be a degree of efficiency that plays a role in this relationship. Regions with better QoG can perform tasks at a more efficient rate than ones with poorer quality of public services and higher levels of corruption. For example, four of the best regions with respect to QoG are the four NUTS 1 regions in the Netherlands (113-116), which also report the lowest average work week hours over the past decade.

Figure 33

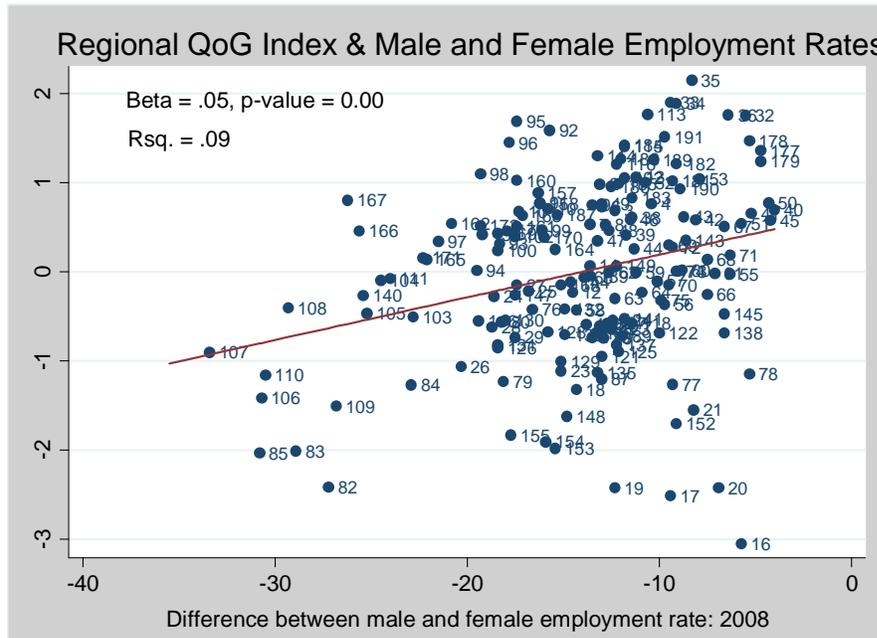
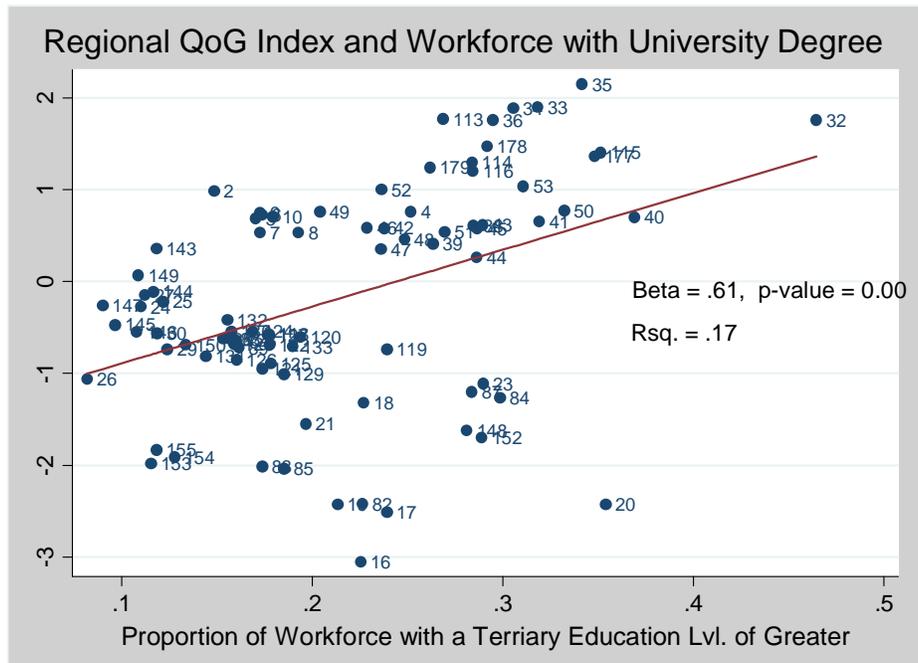


Figure 33 examines the relationship between recent employment patterns by gender and the regional QoG index. We would most likely expect that a more impartial, less corrupt region would have less disparity in employment between men and women. We do indeed find evidence for such a claim, with the bivariate relationship being significant at the 99% level of confidence. However, the R-squared suggests that only a little less than 10% of the variance is explained, and thus the pattern is not quite as strong as in the case of the general unemployment patterns in Figures 6 and 7.

3.9.3. QoG and Education Level in the Workforce

Figure 34

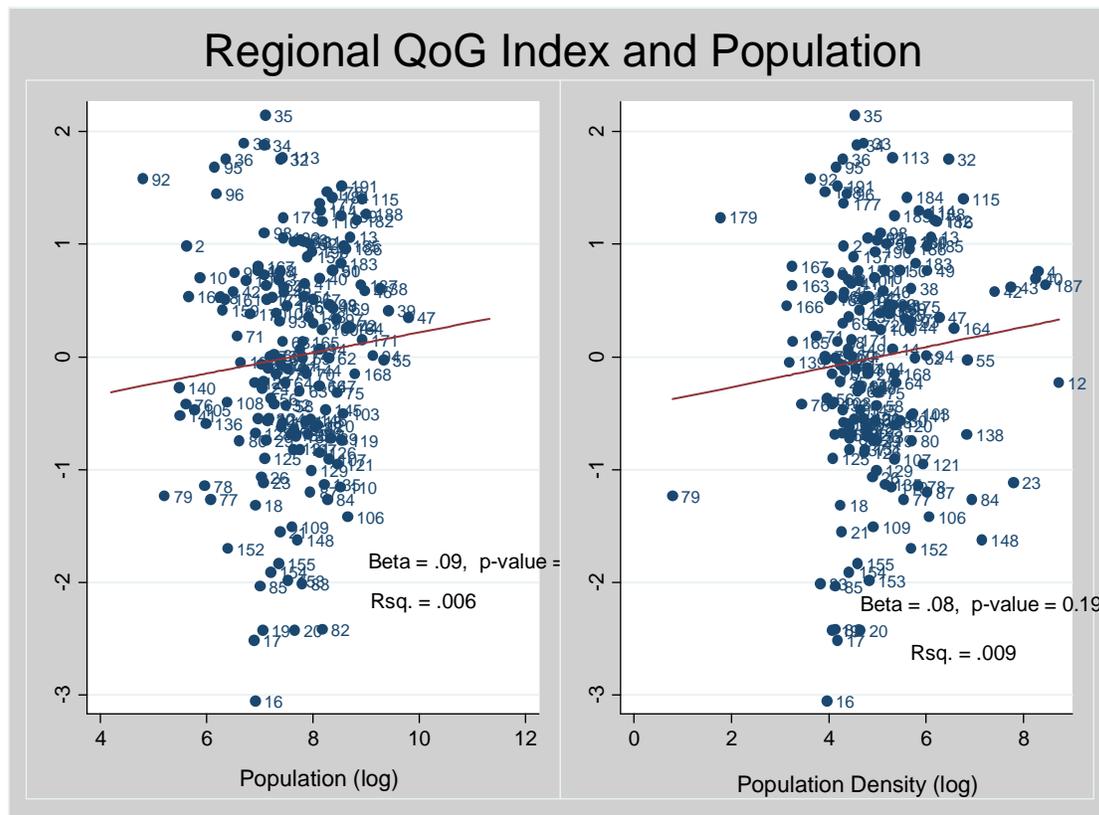


In this model we explore the relationship between how educated a region’s workforce is with its level of QoG. For this, we take the proportion of active workers with a university degree or higher from the total workforce population. We find a relatively strong and positive relationship between QoG and the proportion of university graduates in the workforce, with a Pearson coefficient of 0.41, and the beta coefficient is significant at the 99% level of confidence. The most notable outlier (#20) is the capital region of Bulgaria, Yugozapaden, which despite having about 35% of its workforce with a university education is plagued with relatively low QoG in the EU.

3.9.4. QoG and Population

The individual level regressions pointed to the fact that there might be a downward bias against regions with large populations. We test this relationship using two variables – total population (logged) and population density (logged). We find no such evidence at the regional level, with neither of the two variables being significantly related with QoG in the bivariate regressions. In fact, both coefficients are in the opposite direction, as one could anticipate (they are both positive), yet p-values of 0.30 and 0.19, respectively, reveal that this relationship is far from reaching an acceptable level of statistical significance. Just as we found the population variable to be insignificant at the national level of analysis, so is this variable at the regional level as well.

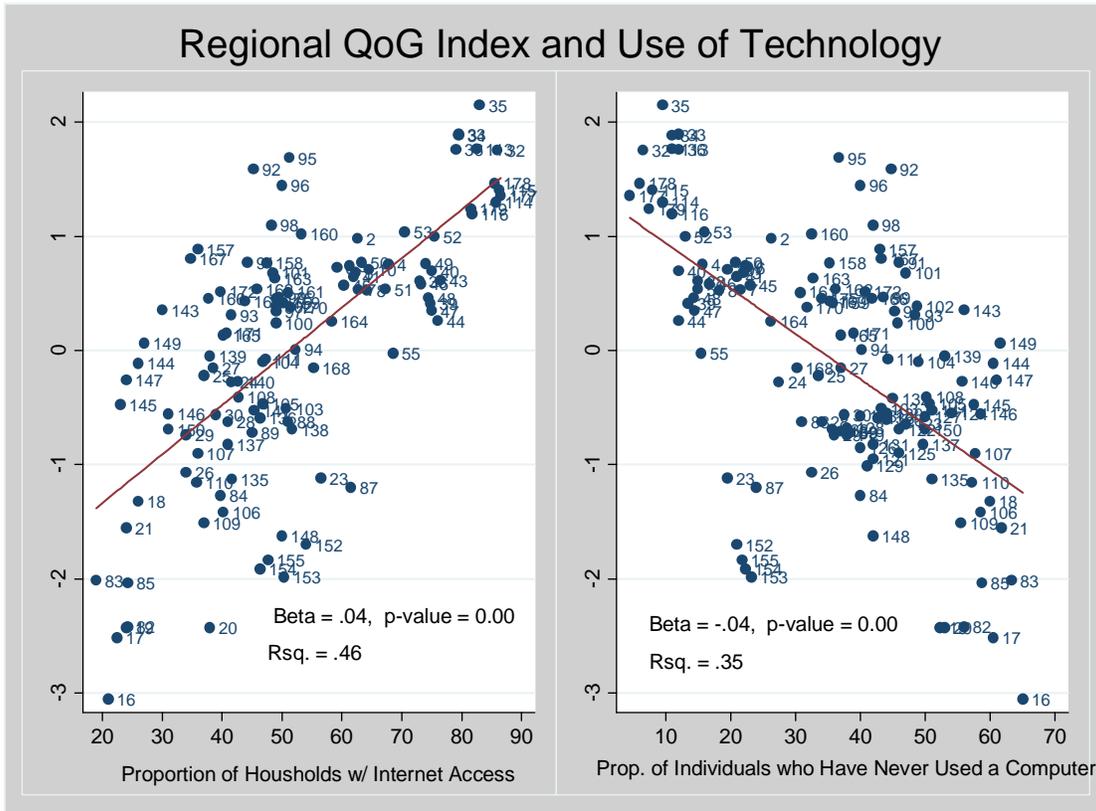
Figure 35



3.9.5. QoG and Technology

Here we look more deeply into the use (or lack thereof) of technology – specifically the proportion of household access to the internet and the proportion of individuals who do not use a computer. Interestingly, we find quite strong and significant relationships between these two variables and QoG at the regional level. First, regions with higher proportions of household internet access tend to have significantly higher QoG on average (Pearson correlation coefficient = 0.68). Second, regions that report higher proportions of individuals who have never used a computer have significantly lower QoG on average. Regions in Slovakia (#’s 152-155) can be considered somewhat of a group of outliers here; they have relatively few people who say that have never used a computer, yet still exhibit quite low QoG. Overall, we would conclude from such findings that technology is certainly connected with access to information, and thus the greater ease and access people in regions have to information, the closer they can monitor the performance of public officials and services.

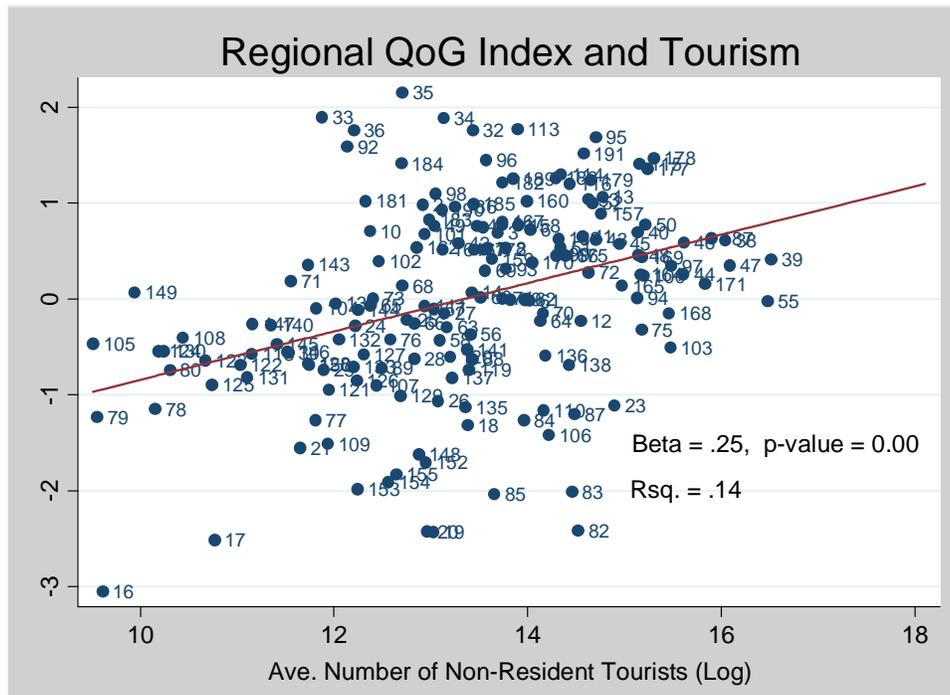
Figure 36



3.9.6. QoG and Tourism

The next analysis we run in this part looks at the relationship between the average number of non-resident tourists that visited the region and QoG. We use tourism essentially as a proxy for ‘reputation’ of the region. We assume that, all things being equal, tourists want to spend their money and travel to safer, more efficient regions where they can trust that their transactions will be valid. Regions with a reputation for high levels of corruption, discrimination or poor public services will most likely attract fewer tourists on average. Although not as strongly related as technology, GDP or unemployment, we find a statistically significant and positive relationship, as expected, between tourism and QoG (Pearson correlation coefficient = 0.37). Whether the driving force behind this relationship is that 1) higher QoG attracts more tourists or 2) the more ‘open dynamic’ tourism brings to the regions cannot be settled by spatial, bivariate analyses. However, the relationship exists and is of course worthy of reporting.

Figure 37



3.9.7. QoG and Income Inequality

In our final model, we test the empirical relationship between income inequality, as measured by the University of Texas Inequality Project, and our regional QoG measure. The inequality measure is a Theil Index and, for the purpose of this analysis, contains all or most regions from Austria, Bulgaria, the Czech Republic, France, Germany, Italy, Netherlands, Poland, Portugal, Spain and Slovakia. The data come from the years 1997-2000, and we take the average of that time period and regress it against our regional QoG index. Although it is measured almost approximately a decade before our QoG data, we feel it is still appropriate given that income inequality is generally a slow moving variable, and we would believe in general that most of the regions would be more or less in the same place with respect to this variable. Yet we acknowledge that this may not be the case for all regions. Given this caveat, just as we did with the national level QoG index, we find a strongly significant and negative relationship between income inequality and QoG, with 31% of the variation of the QoG variable being explained by income inequality.

3.10 Conclusions

In this part of the report, we presented a newly created regional QoG index built from a citizen-based survey of more than 33,500 respondents from 18 different EU countries. Again, this is a pioneer effort in generating a regional indicator for 172 EU regions with respect to quality of government. Moreover, this constitutes the largest survey ever undertaken that asks respondents specifically about regional level QoG. While we do not claim this is a perfect measure (for one, there is of course a limited number of respondents for each region), we are satisfied with the overall construction and outcome of the composite index, along with its internal and external validity. We summarize this section below.

- In general, we asked respondents about three public services in their regions – education, health care and law enforcement – three of the public services that are most often either administered or financed by regional governments (or both). Obviously not all NUTS 1 or 2 regions studied in the survey are equally meaningful in this regard; however, future studies can of course choose to omit certain regions for which QoG is a much less relevant concept. In focusing on these three services, we asked respondents to rate them with respect to three related concepts of QoG – quality of services, impartiality and level of corruption. In addition we included two further questions in the index – one about the fairness of regional elections and the other about the strength and effectiveness of the media in the region in exposing corruption.
- After many internal consistency checks and tests at both the individual and aggregate regional levels, we ran a factor analysis to determine whether the 16 underlying variables used in the composite index could be grouped together. We found that the questions that were intended to capture the three primary QoG concepts – quality, impartiality and corruption (i.e. the three ‘pillars’ of the regional QoG index) – grouped together according to the results of the factor analyses. We then built the final index by weighting the three pillars equally and deriving a final number, which we then standardize for each region so that the EU average is ‘0’ with a standard deviation of ‘1’ and report 95% margins of error for each region. We find many expected results – regions from the high-ranking group with respect to national level QoG, Denmark, Sweden, the Netherlands, U.K., Austria and Germany, all reported regions above the EU average and, in the case of the first three countries, all were at least one standard deviation over the EU average. On the other hand, regions from low ranking countries in the national indicator, such as Bulgaria, Slovakia, Greece and Poland, were also ranked in the

bottom third of the 172 that were accounted for in this study. These countries were also marked by very little within-country variation in QoG in their respective regions. In addition to expected findings, we discovered that several countries, such as Italy, Spain, Romania, Portugal and Belgium, have quite high levels of within-country variation that the national level indicators fail to detect. Moreover, one can argue that within-country QoG differences in states such as these are more important than cross-country differences in many pair wise comparisons across the E.U. For example, while considered to be among the bottom five countries or so with respect to national level QoG according to the World Bank, Italy has three regions that score around many regions in France, Germany and Austria, which have much higher national-level QoG scores than Italy on whole. Romania, a country found to be in the bottom two or three of all indicators we explored at the national level in the previous section of the report, has significant within-country variation despite having regions that are administratively meaningless, pointing to the uneven distribution of public sector employees across the country. Finally, the only one of the regions (Jihozapad - CZ03) among the new member states (NMS) is ranked above the EU mean.

- We tested the robustness of the findings by employing a sensitivity analysis where we altered the construction of the composite index in 62 different simulations. We use a different aggregation method, weighting scheme, normalization of the data, and inclusion/exclusion of whole pillars along with individual indicators. In addition, we run several simulations in which we exclude certain demographic groups from the aggregation from the individual to the regional level, such as gender, income and age. We find that regions in Spain, Romania, the Czech Republic and Portugal are most often impacted by changes in the model, while most other regions are relatively stable. While there are several regions that are impacted somewhat significantly (at times up to 68 places in the rank order), none of the simulations reported a Spearman rank coefficient under .90. Even the most extreme scenarios (see Figure 4) point to the regional QoG index being quite stable and robust to changes in the weights, aggregation, normalization, exclusion of certain indicators or pillars and demographic changes.
- As regards external validity, we find that the regional QoG index, when aggregated to the national level, corresponds quite closely to the national QoG index that we constructed in the previous section using the World Bank data. Although not entirely comparable owing to our focus on the regional level, this points indeed to some level of external validity. Aside from those

countries that exhibit high degrees of within-country regional variation, such as Italy and Romania, we find that the other countries fall within a place or two of the national level indicator for both the overall index and the corruption pillar, lending support to external validity. We also take advantage of this comparison between the national and regional level data to include the nine countries excluded from the regional data in an EU-wide QoG index by combining regional and national level data for all four pillars (see Table 14), with the 172 regions. Following the national-regional level comparisons, we further test for external validity by performing a number of bivariate regressions and scatter plots with regional level variables for which we would expect (or not expect) to find an empirical relationship with QoG. We find that the regional QoG index is strongly related to such factors as GDP per capita, various measures of unemployment, education levels of the workforce, use of technology, level of non-resident tourism and income inequality. Similar to the national level QoG index, the regional level index is not related empirically to population – neither total population nor population density.

The regional QoG index built here is intended to assess the relative positions among EU regions in 18 member states. It has been subjected to the standard internal and external tests to evaluate the quality of a composite index. As a result, we recommend its use so long as the researcher is knowledgeable of its construction, strengths and weaknesses and is cautious about the empirical conclusions drawn from this index.