Quantiles of Corruption

00317421 and 50861899*

29 April 2012

1 Introduction

> "The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slaves of some defunct economist."
>
> -J.M. Keynes, *The General Theory of Employment, Interest and Money*

Ideas created by the elite tend to *eventually* affect the practical man (or woman), who may forced to live by them. Especially careful interpretation should be given to ideas that may influence policy decisions. Corruption is a good example. Although inherently difficult to observe, measurements are plentiful. As one economist has suggested, it's "like pornography...you know it when you see it" Wei (1999). Whether it be pervasive bribery in India or sophisticated influence markets in the United States, Transparency International (2005); Abramoff (2011), corruption has probably been the subject of a recent domestic or international news story.

Like most observations in the social sciences, concept and measurement validity are inherent problems Adcock and Collier (2001). As a very 'normative' concept, corruption, has been the subject of much academic debate (see Nye (1967); Scott (1972); Thompson (1993); Johnston (1996); Philp (1997); Philp (2002)). Although the rent-seeking framework provides social scientists with a 'positivist' and presumably less value laden analytical tool (see Krueger, Anne (1974)) debates on measurement validity are rare, yet an apparently quantifiable definition is now typical: 'The abuse of public office for private gain' World Bank (1997).

This is the definition used by Transparency International's (TI) *Corruption Perceptions Index*, published annually since 1995. Now with an iPhone application [1], cross-country comparisons can be made at the dinner table or the cabinet meeting. Indeed, TI has a local chapter in most countries and when policy makers and the public think about corruption they tend to reference this index[2]. 'Rank-seeking', in which policy is designed to improve index scores rather than *real* performance is an emerging concern Høyland et al. (2012).

*Please note that results from the original paper Billger and Goel (2009) could not be replicated. The authors did not provide data but because they are publicly available we were able to collect for the relevant years. However, the data collected did not match the descriptive statistics reported in the original paper. For example, the Economic Freedom Index, which contains 10 sub-components and ranges from 0 to 100 takes on the values -5 to -1 in the original paper. One of the sub-components, corruption, is also the dependent variable. Another sub-component, Labor Freedom, is alluded to in the original paper but did not exist during the time period examined. When we refer to the 'original findings' in the paper we mean the ones we replicated (Table 1) using the data collected assuming listwise deletion was used to handle missing values. These results, as well as all subsequent findings are substantially different from the results reported in the authors' original paper.

[1] http://cpi.transparency.org/cpi2011/iphone/

[2] See, for example, a recent article on Mexico in the NY Times

Perhaps rank-seeking is not a cause for concern when the *real* closely matches what we observe. Unfortunately measurement error is ubiquitous in the social sciences and what we actually capture is rarely the 'true' quantity of interest. As Høyland et al. (2012) and Treier and Jackman (2008) have shown, the extensive literature on cross-country analysis using indices is especially vulnerable to measurement criticism. Corruption is no exception. Although agreement on causal mechanisms is rare and no unified model exists, many hypotheses have been proposed and examined using other indices, such as democracy and economic freedom, as covariates. The results have been equivocal. Researchers examining the same concepts have drawn completely different conclusions depending on how those concepts were measured and what covariates were included in their models La Porta (1999); Fisman (2002); Montinola (2002); Graeff (2003); Gerring (2005); Goel (2005).

Results have also been driven by the data consulted and the countries sampled. Early empirical findings that 'big government' results in high levels of corruption, for example, depended on data that excluded Denmark, Finland, Iceland, Norway and Sweden LaPalombara (1994). Although much concern has focused on omitted variable bias and 'robustness checks' using different indices, little attention has been paid to missing data, measurement error and the changes in statistical relationships across countries. In this article, we show that the substantive consequences of ignoring these issues can be severe. To demonstrate this, we examine the findings of a recent study, Billger and Goel (2009), and show the causal claims that more democratic countries have lower levels of corruption and that democracy has a stronger effect on corruption in the most corrupt countries are data dependent.

First, we use Amelia II, a flexible multiple imputation method Honaker, King and Blackwell (2009), to recover 84% of the observations removed with listwise deletion. We then expand the dataset to a 10 year period and include control variables for regional and temporal correlation. Third, we use alternative measurements of democracy and corruption to illustrate the consequences of measurement error. The results are revealing. Although multiple imputation replicates the finding that democracies tend to have lower levels of corruption, we find that this relationship diminishes in the most corrupt countries. Dividing the Freedom House Democracy Index into its individual components reveals that 'Political Rights' are associated with high perceptions of corruption whereas 'Civil Liberties' are associated with lower perceptions of corruption. Using the Polity Index, we also find that democratic countries have higher levels of corruption as measured by the Transparency International Index. None of these relationships are strongest for the 'most corrupt' countries. In fact, the relationship between corruption perceptions and democracy is *lower* in countries with the highest corruption scores. Our findings suggest that substantive conclusions and inferences previously drawn from observational studies on the 'causes' of corruption are data dependent.

2 Background

Social scientists have long questioned the relationship between economic prosperity and democracy. However, little evidence exists to suggest an increase in income is a sufficient condition for institutional improvement Robinson (2006); Acemoglu et al. (2008). A natural question seems to be, what about the relationship between *corruption* and democracy? Cross-country searches have thus far been unsuccessful in uncovering general theories to answer this question and most results have been sensitive to measurement choice and data selection (see Lambsdorff (2006);Lambsdorff (2007); Jain (2001) for an overview).

Bringing institutions and policy closer to the democratic ideal could plausibly increase or decrease the level of corruption. It might be the case, for example, that as a wider proportion of citizens determine leadership succession, they are increasingly able to punish corrupt leaders. Quantitative studies have found support for this argument (Lambsdorff (2007); Rose-Ackerman (1999); Bohara et al. (2004); Chowdhury (2004); Fisman (2002); Goel (2005); Shen and Williamson (2005); Tavares and Wacziarg (2001); Treisman (2000)). On the other hand, movements toward democracy might increase perceived levels of corruption in some cases. In autocratic systems, it may be that

only a small portion of the elite need to be 'bought off'. As a greater portion of the population is enfranchised, including, perhaps, previously marginal sub-state power brokers or ethnic-elites, these groups may also demand, and now receive, side-payments in exchange for political support. Rulers and their executive may be required to buy off a larger number of people as the political elite expands. As Morris (2009: 8) notes, 'democratization creates new rules for institutions and new means of acquiring and exercising power and wealth, conditions that may also open new opportunities for corruption'. Indeed, a series of studies has found that democratization in Latin America during the 1980s was accompanied in many cases by an increase in what was labelled 'new corruption' Brown (2004).

It is important to understand how corruption differs *across* countries conditional on common covariates and ask whether conditional mean estimation is an appropriate analytical tool. It is possible, for example, that where leaders and political elites are involved heavily in corruption (countries where the perception of corruption might plausibly be high) elections enable voters to eject the most egregious and public representatives of corruption. As the public side of corruption is dispensed with, however, democratic institutions and norms may have more difficulty rooting out lower level practices. In this case, democratization would have a visible impact on corruption in the most corrupt countries, but its marginal effect might decline as countries become less corrupt. As Billger and Goel (2009) point out, however, it may be the opposite. Corruption may be so entrenched in some countries that democratization has a small marginal impact in the most corrupt countries, but is much more effective in reforming countries with moderate or low levels of corruption.

In a novel departure from the traditional conditional expectation approach, Billger and Goel (2009) used quantile regression to estimate the relationship between corruption conditional on some commonly used covariates. Quantile regression is a useful approach as it allows the researcher to model variations in the effect of an independent variable on a dependent variable along segments of the dependent variable's range Koenker (2005). It is likely that institutional changes will have differential impacts on the level of corruption depending on whether a country experiences a high, moderate or low level of corruption. Indeed, this is what Billger and Goel report. While movements toward democracy are generally associated with reductions in corruption (sometimes in statistically significant ways, depending on the control variables included in the model), the more corrupt a country is may influence the degree to which democracy impacts corruption (again, sometimes in significant ways, depending upon control variables). We might conclude from this research that democratization is a useful tool to reduce corruption in the most corrupt nations.

Unfortunately, a sample of only 105 countries measured over three years (2001-2003) is insufficient for evaluating these hypotheses. Further, listwise deletion, which is only appropriate when data is missing completely at random (MCAR) can bias standard errors and in some cases result in incorrect conclusions King et al. (2001), Honaker and King (2010) Gemici et al. (2011).[3] Although methodologically sophisticated, quantile regression fails to add any value beyond the traditional OLS approach when the sample is small. Moreover, results indices such as Freedom House's Democracy and Polity Marshall and Jaggers (2010) are sensitive to changes in measurement and aggregation. In fact, this is also a problem for the Heritage Foundation's index of economic freedom, which also includes a measure of 'Freedom from Corruption' derived from Transparency International's Index. Failure to disaggregate the independent variable means regressing corruption on corruption.[4]

[3] Note that in the original paper the authors report 99 observations whereas we were able to recover 105 from the sources cited. We are also assuming that the authors used listwise deletion to deal with missing data. We are only speculating here. The authors did not respond to e-mail enquiries.

[4] We cannot be completely sure on this as the authors did not respond to e-mail queries on this issue. In our attempt at replication we assumed they removed this component of the Economic Freedom Index.

2.1 Quantiles of Corruption

The 'quantiles of corruption' approach divides the population of countries into five parts.[5] The 0.1 quantile contains countries with corruption perceptions scores lower than 90% of the population. Countries in the 0.5 quantile perform better than half the population, but worse than the other half. As a heuristic, Table 1 provides some examples of countries in different quantiles:

Table 1: Countries by Quantile – Corruption Perceptions Index

Quantile	Country	Score (2005)
0.10	Austria	1.29
	United Kingdom	1.40
	Spain	2.99
0.25	Bahrain	4.24
	South Africa	5.47
	Greece	5.74
0.5	Turkey	6.47
	Mexico	6.50
	Poland	6.59
0.75	Argentina	7.23
	Bolivia	7.46
	Sierra Leone	7.59
0.90	Cameroon	7.76
	Papua New Guinea	7.70
	Bangladesh	8.32

Notes: drawn from sample of 181 countries with multiple imputation. Values listed here match those reported by Transparency International

One might argue that Bangladesh and Papua New Guinea are quite different than Austria and the United Kingdom or Turkey and Mexico. In the linear world, the conditional expectation function allows statements like, "a one standard deviation increase in democracy results in a 10% decrease in corruption, *on average*." Unfortunately, the *on average* qualification may not be very useful for cross-country analysis. The relationship between corruption and democracy in Cameroon is likely to be influenced by different confounding factors than that same relationship in Spain. Quantile regression allows one to ask, conditional on being in quantile 0.90 what is the relationship between democracy and corruption? More generally, if 90% of the countries in the sample are less corrupt than X, is the relationship between democracy and corruption different for X than for countries where only 25% of the samples is less corrupt?

To understand this, it helps to move slightly beyond the 'black box' interpretation of quantile regression. Fortunately, the basic intuition is rather simple and does not require derivation from first principles or a lesson in linear programming.[6] When modeled as an optimization problem, linear regression involves minimizing the sum of squared residuals whereas quantile regression seeks to minimize the sum of asymmetrically weighted absolute residuals Koenker and Hallock (2001).[7] As Koenker explains, the equation that defines the unconditional quantiles is written:

[5] Since we are only focusing on five quantiles, we might call this 'quintiles of corruption' instead. One can divide the population into any number of quantiles and obtain estimates with associated measures of uncertainty. Selection of quantiles may be motivated by some substantive theory. In this application, we are not necessarily motivated by any theory but instead mimic the selection of quantiles from Billger and Goel (2009). In Figure X-Y we graphically illustrate the differences across all quantiles of the conditional distribution

[6] Koenker and Bassett (1978) provide the formal treatment. A textbook version Koenker (2005) is thorough and contains several applications and extensions to other estimation methods.

[7] Koenker's website is also a valuable resource and contains examples using R code: `http://www.econ.uiuc.edu/~roger/`. Angrist and Pischke (2009) also provide a useful description of quantile regression in Chapter 7

$$\min_{\xi} \in \Re \mathfrak{e} \sum \rho_{\tau}(y_i - \xi)$$

Where ρ_{τ} gives different weights to the residuals and the solution is the τth quantile. The optimization problem for linear regression,

$$\min_{\mu} \in \Re \mathfrak{e} \sum_{i=1}^{n} (y_i - \mu)^2$$

Produces an estimate of the unconditional population mean – the sample mean. Substituting a function $\mu(x, \beta)$for μ,

$$\min_{\beta} \in \Re \mathfrak{e} \wp \sum_{i=1}^{n} (y_i - \mu(x, \beta))^2$$

And solving this analytically more complicated optimization problem yields the conditional expectation function, $E[Y_i|X_i = x]$. Likewise, we can substitute $\xi(x, \beta)$for ξ,

$$\min_{\beta} \in \Re \mathfrak{e} \wp \sum \rho_{\tau}(y_i - \xi(x, \beta))$$

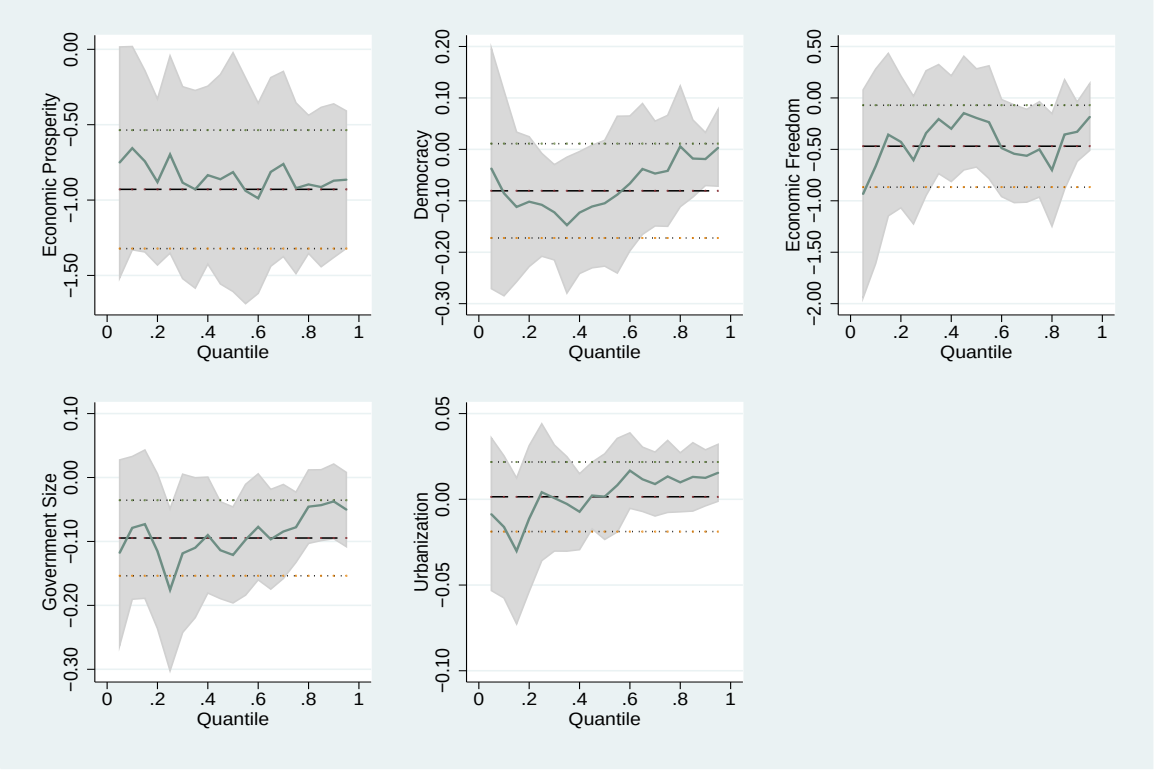
And, assuming $\xi(x, \beta)$ is a *linear* function of the parameters we want to estimate, linear programming can find a solution for the conditional quantile, $Q_y(\tau|x) = \beta_0 + x\beta_1 + F_u^{-1}(\tau)$. Where $\tau = .50$ estimates the conditional median.

In the following section we replicate the Billger and Goel findings using quantile regression to demonstrate the consequences of listwise deletion. Second, we implement multiple imputation as an alternative. Next, we extend the sample to a 10 year period and explore measurement issues associated with the covariates–economic freedom and democracy. Finally, we introduce an alternative index to examine measurement error in the outcome variable.

2.2 Listwise deletion

Model 1a (Table 2) shows our replication of the Billger and Goel study. These results suggest the relationship between corruption and democracy is *not* more important for the most corrupt countries. The results from Model 1a (Table 2) suggests that the OLS model provides more precise estimates than the quantile regression specification. This is not surprising given the small sample size. Figure 1 provides a graphical illustration of how the estimates for the covariates change across the quantiles of corruption. Indeed, as Figure 1 shows, the confidence intervals are too large to discern any benefit of quantile regression over the traditional OLS approach when analyzing the relationship between *any* of the covariates.

Figure 1: OLS and Quantile Regression Estimates from Model 1a



Notes: Attempted replication of Billger and Goel (2009). Dependent variable is Corruption Perceptions Index (0-10). Higher values indicate higher corruption perceptions. OLS parameter estimates represented by black dotted lines with corresponding 95% confidence intervals above and below. Grey shaded area represents 95% confidence interval for quantile regression parameters.

According to these results, an increase in government size is associated with a reduction in corruption perceptions. This is supported by previous findings Elliot (1997); Graeff (2003); Adsera (2000) and contradicts others LaPalombara (1994); Goel (1998). Although these findings are 'statistically significant' in the OLS regression and at the 0.25 and 0.50 quantiles, it's hard to draw any substantive conclusions. This implies, for example, that a 3 standard deviation increase ($\approx$ 21%) in government consumption expenditure corresponds to a 1.71 ($\approx$ 17% or .53 points short of 1 standard deviation) reduction in corruption perceptions. To put this in context, for the period 1960 to 1980, the modern 'boom' in public expenditure, average government consumption expenditure as a % of GDP in rich countries increased by $\approx$ 54% from 27.9 % to 43.1% Mueller (2003).

Still, the results suggest that all quantiles are not created equal. If the conditional distribution of corruption on the covariates is *symmetric* the OLS coefficient and the median (0.5 quantile) coefficient will be the same Angrist and Pischke (2009). Although the results provide evidence for asymmetry, we should be skeptical in light of the information we have – 105 of potentially 543 observations. A comparison to results using multiple imputation may provide us with a better picture of what we're missing with listwise deletion.

TABLE 2: Quantile Regression v. OLS on Corruption Perceptions Index (Model 1a)

	Obs.	Desc. Stats		OLS	Quantiles				
		μ	σ	Coeff.	0.1	0.25	0.5	0.75	0.9
Economic Prosperity	105	8.79	1.24	-0.929	-0.656	-0.698	-0.815	-0.921	-0.871
				(0.198)	(0.388)	(0.293)	(0.305)	(0.263)	(0.231)
Democracy	105	6.05	3.54	-0.081	-0.085	-0.108	-0.105	-0.042	-0.019
				(0.046)	(0.095)	(0.064)	(0.071)	(0.050)	(0.038)
Economic Freedom	105	0.22	0.90	-0.469	-0.664	-0.604	-0.195	-0.500	-0.329
				(0.201)	(0.479)	(0.384)	(0.282)	(0.232)	(0.234)
Government Size	105	15.70	5.02	-0.095	-0.079	-0.176	-0.121	-0.078	-0.037
				(0.030)	(0.059)	(0.063)	(0.040)	(0.037)	(0.028)
Urbanization	105	59.96	20.82	0.001	-0.016	0.004	0.002	0.013	0.013
				(0.010)	(0.024)	(0.020)	(0.014)	(0.010)	(0.010)

Notes: Results from quantile regression with 1000 bootstrapped replications using dataset with listwise deletion. Dependent variable is Corruption Perceptions Index rescaled for interpretation (0-10). Higher values indicate higher corruption perceptions. Democracy is the addition of Freedom House Civil Liberties and Political Rights scores rescaled for interpretation (0-14). Higher values indicate more democracy. Economic Freedom is the average of Business Freedom, Financial Freedom, Fiscal Freedom, Investment Freedom, Monetary Freedom and Trade Freedom components of the Heritage Foundation's Economic Freedom Index as described in Billger and Goel (2009). Economic Prosperity is the natural log of GDP per capita (PPP). Urbanization is the % of population living in an urban area. Government size is government consumption expenditure as % of GDP. Urbanization, Government Size and GDP data come from the World Bank.

2.3 Multiple Imputation

Honaker and King (2010) outline four sufficient conditions for using listwise deletion, *all* of these must hold in order for listwise deletion to be preferable to multiple imputation:

- The analysis model is conditional on X and the functional form is known to be correctly specified.
- There is non-ignorable missingness in X and no external variables can be used in the imputation stage.
- Missingness in X is not a function of Y, and unobserved omitted variables that affect Y do not exist.
- The number of observations left after listwise deletion should be so large that the efficiency loss from listwise deletion does not counterbalance the biases induced by the other conditions.

None of these conditions are met in this case. First, the functional form is not known to be correctly specified, which is explicitly recognized in Billger and Goel (2009). Although corruption is argued to be a 'function' of economic prosperity, democracy, economic freedom, government size and urbanization, the form of this function cannot be specified in the absence of some *a prior* theory, which does not currently exist. In the absence of a formal hypothesis, statistical relationships are purely observational. Second, although we can never know when data are non-ignorable Gemici et al. (2011), we do have external variables that can be used in the imputation stage. Economic freedom, for example, is also influenced by the extent to which regulation serves as a barrier to business start ups, which we can measure relatively 'objectively' with the World Bank's days required to start a business index.[8] Third, whether or not missing in X is a function of Y is not known for *all* covariates[9]; however, one can imagine countless omitted variables which may affect corruption. Fourth, the number of observations left after listwise deletion (105) is only 84% of the total number of possible observations.

[8] See appendix for imputation diagnostics.

[9] Apparently there is a test for this, Little (1988) which we may explore at a later date.

Three 'reasons for missingness' are relevant in this context: wave non-response, unit non-response and non-response by some survey participants Gemici et al. (2011). Wave non-response means that observations are available for some years and not others. Albania, for example, does not have a TI score for 2001 but records observations for years prior (1999) and after (from 2002). Multiple imputation allows for estimation for 2001 using information from the observables and the covariates. The second issue, unit non-response, can also be remedied using multiple imputation methods. In 2001, for example, Albania has observations on all the covariates even though it's missing an observation on the dependent variable for that year. Listwise deletion removes this country from the sample completely. The third issue, non-response by survey participants is a particular problem with the Transparency International Index. [10] Because the index is an average of several surveys from risk agencies and business people (see Lambsdorff (2007) for an extensive overview of the CPI methodology), more surveys may be used in some years than others. In 2002, for example, Albania's corruption perceptions score relied on three surveys whereas in 2003 the score relied on 5 surveys.[11]

Non-response by some survey participants is not an issue addressed in this application. The other issues, wave non-response and unit non-response, can be remedied. To do so, we use the open-source software Amelia II Honaker, King and Blackwell (2009) to recover 438 observations via multiple imputation King et al. (2001), Schafer (1997) and Little and Rubin (2002). Given the nature of missingness, ignoring the multiple imputation remedy means discarding potentially useful information which results in standard errors that make the results meaningless when compared to OLS (this is demonstrated graphically in Figure 1) Gemici et al. (2011).

Table 3: Quantile Regression v. OLS on Corruption Perceptions Index (Model1b)

	Obs.	**Desc. Stats**		**OLS**	**Quantiles**				
		μ	σ	Coeff.	0.1	0.25	0.5	0.75	0.9
Economic Prosperity	543	8.43	1.31	-0.861	-0.669	-0.836	-0.811	-0.780	-0.742
				(0.093)	(0.179)	(0.121)	(0.123)	(0.113)	(0.165)
Democracy	543	3.44	1.90	-0.120	-0.146	-0.116	-0.120	-0.076	-0.082
				(0.123)	(0.049)	(0.067)	(0.066)	(0.060)	(0.102)
Economic Freedom	543	-0.12	1.03	-0.624	-0.528	-0.528	-0.468	-0.503	-0.387
				(0.214)	(0.110)	(0.165)	(0.170)	(0.104)	(0.163)
Government Size	543	16.24	6.94	-0.081	-0.059	-0.078	-0.068	-0.048	-0.035
				(0.019)	(0.015)	(0.021)	(0.018)	(0.017)	(0.018)
Urbanization	543	52.69	23.08	-0.020	-0.002	-0.009	-0.000	0.004	0.004
				(0.010)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)

Notes: Results from quantile Regression with 1000 bootstrapped replications using dataset with multiple imputation for years 2001-2003. Dependent variable is Corruption Perceptions Index rescaled for interpretation (0-10). Higher values indicate higher corruption perceptions. Democracy refers to the average of Freedom House Civil Liberties and Freedom House Political Rights scores. Economic Freedom refers to the average of Business Freedom, Financial Freedom, Fiscal Freedom, Investment Freedom, Monetary Freedom and Trade Freedom components of the Heritage Foundation's Economic Freedom Index as described in Billger and Goel (2009). Economic Prosperity is the natural log of GDP per capita (PPP). Urbanization is the % of population living in an urban area. Government size is government consumption expenditure as % of GDP. Urbanization, Government Size and GDP data come from the World Bank.

Model 1b (Table 3) shows a replication using multiple imputation. One thing to note is the reduction in estimation uncertainty. For example, the standard error associated with the OLS estimate for economic prosperity has been reduced by approximately 53 %. Recall that Billger and Goel (2009) report a negative relationship between democracy and corruption that diminishes as countries become less corrupt. Here, as in the replication with listwise deletion, we find the opposite. Although the relationship between democracy and corruption remains negative overall, the coefficient on demo-

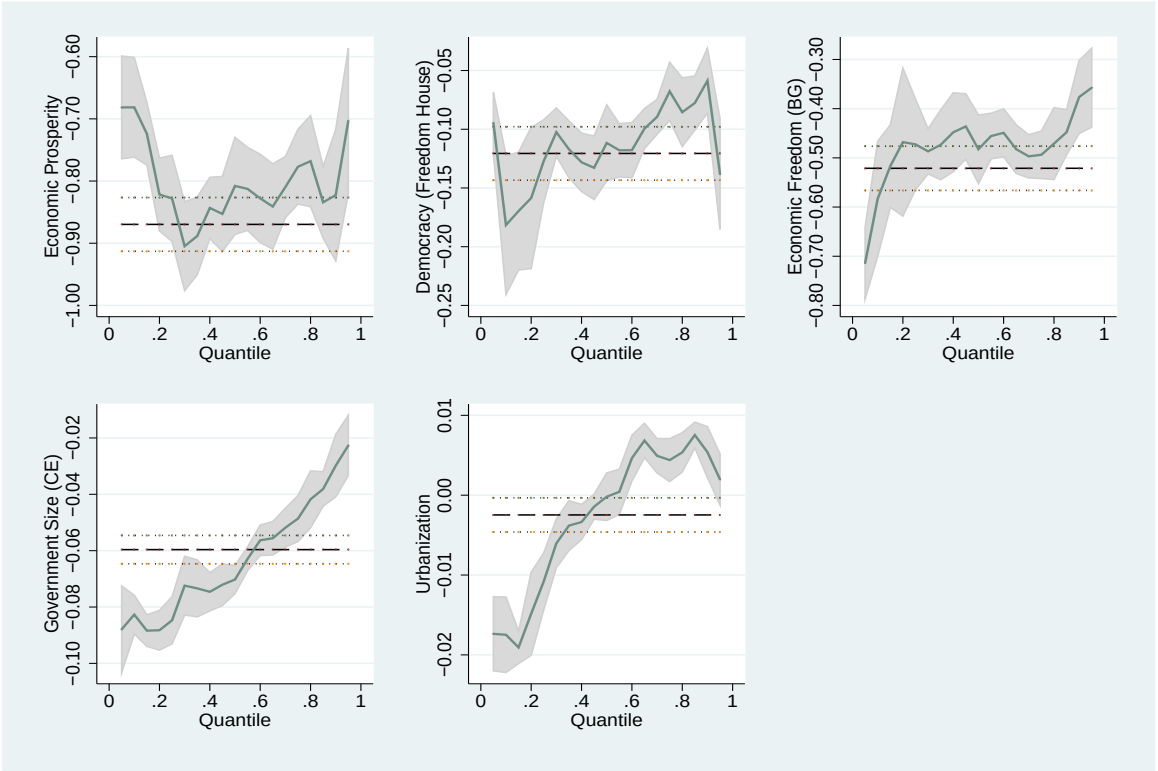
[10] This may also be an issue with the Freedom House Index but we don't know because, unlike TI, they don't release their survey statistics

[11] Given this fact, it might be useful to incorporate confidence intervals for each TI estimate. This is, possibly a task for another paper.

cracy is strongest in magnitude, albeit not statistically significant, for quantiles below the median. The results also suggest that OLS estimates overemphasises the relationships between corruption and economic prosperity for the 0.1 quantile – those countries with the lowest corruption levels.

In terms of estimation uncertainty, multiple imputation results in substantial improvements for all covariates, as illustrated graphically in Figure 2. Indeed, the relationship between corruption and the covariates for government size and urbanization fall well beyond the 95 percent confidence intervals for the standard OLS analysis. This is a tremendous improvement when compared to Figure 1, which, given high levels of uncertainty, showed no discernible difference between traditional OLS and quantile regression estimates.

Figure 2: OLS and Quantile Regression Estimates from Model 1b



Notes: Dependent variable is Corruption Perceptions Index (0-10). Higher values indicate higher corruption perceptions. OLS parameter estimates represented by black dotted lines with corresponding 95% confidence intervals above and below. Grey shaded area represents 95% confidence interval for quantile regression parameters.

The improvements in uncertainty and changes in coefficient magnitudes suggest the missing data were a substantive problem. In terms of bias, for example, the new results suggest the relationship between economic freedom and corruption was underestimated by approximately 35% for the OLS regression and approximately 140% for the conditional median (the 0.5 quantile). The substantive finding that "among the most corrupt countries, increases in government size do not reduce corruption. This suggests that government machinery fights corruption after a minimum threshold or that larger governments are unable to check corruption in the most severe instances" Billger and Goel (2009) is not supported by either specification above. Indeed, the estimates from both models suggest the opposite: increases in government size reduce corruption perceptions across all quantiles, albeit not in an easily interpretable way. For example, an increase in government size by three standard deviations ($\approx$ 20.82 % increase in government consumption expenditure) is needed for a

reduction in corruption perceptions of approximately 1.41 points, or just over half of one standard deviation.

3 Measurement, Controls and Disaggregation

Having shown that missing data are problematic for the original study, we move on to test the findings with alternative specifications of democracy and corruption. This is party motivated by the warnings from Jackman (2004); Treier and Jackman (2008); Høyland et al. (2012) and partly by the indecisive and contradictory findings in cross-country corruption research, often conditional on the data and measurement instruments used. First, we briefly discuss measurement error issues that render causal conjectures problematic. Second, we increase the sample period to 1999-2009 for a total of 1991 observations and apply quantile regression to the *individual components* of the economic freedom and democracy indices and on an alternative measure of democracy, the Polity Index Marshall and Jaggers (2010). Finally, we use an alternative measure for the dependent variable to test construct validity Adcock and Collier (2001)

3.1 Caution: Measurement Error Ahead

> "If you run a regression of lung cancer death rates on the purchasing power of the dollar, the data will follow the line very closely. Inflation, however, neither causes nor prevents lung cancer." Freedman (2005)

Even if our imputations worked perfectly and we are no longer worried about the missing data problem, we are not out of the woods yet. To illustrate why we review two concepts: measurement error in the dependent variable and measurement error in the covariates. Suppose we have a simple regression of corruption (Y) on democracy (X) and we use a linear specification to estimate the relationship:

$$Y = X\hat{\beta} + \epsilon$$

Assume we are only concerned with measurement error in the simplest sense. If our measure of corruption (Y) is subject to random error this can be absorbed by ϵ. Stated differently, the simplest measurement error problem can be decomposed as follows:

$$\epsilon = Y - X\hat{\beta}$$

In which case our residuals will be incorrect, either too large or too small. Estimates might have less precision and although the coefficient may not be 'statistically significant', the estimate $\hat{\beta}$ will on average be the true value. Given the richness of the data after imputation we might feel ok about this. On the other hand, if we also have measurement error in X this creates imprecision and bias in our coefficient estimate $\hat{\beta}$ which can go *either way*. For example, assume:

$$X = \tilde{X} + \gamma$$

Where $\tilde{X}$ is the true value and we observe X. Our real measurement error is some combination of γ and ϵ. The "Iron Law of Econometrics" suggests that the $\hat{\beta}$ will be less than the true $\tilde{\beta}$ that would have been calculated using $\tilde{X}$ rather than X Hausman (2001). In other words, we will underestimate the true relationship between X and Y. But what if γ is also correlated with $\tilde{X}$ and or ϵ? As the variance in γ becomes large compared to the variance in $\tilde{X}$ the bias increases and can become a very serious problem. Estimates of the relationship between education and earnings, for example, have been found to underestimate the true relationship by more than 25% Card (2001).

Imagine we have another covariate, $X2$, which is also correlated with X. This is certainly the case with economic freedom, GDP Per Capita and democracy. Additional covariates and the transition to hyper dimensional space make things increasingly complicated and uncertain. Although solutions have been proposed to address these issues Freedman (1987); Bollen (1989); Abrevaya and Hausman (1999); Freedman (2004), the discussion is beyond the scope of this analysis. Instead, we seek to test the issues by 1) controlling for some observable confounders, 2) disaggregating the indices as suggested by Høyland et al. (2012) and 3) testing similar ones for robustness as suggested by Treier and Jackman (2008).

3.2 Endogeneity

Controlling for unobservable confounding factors can mitigate one source of 'endogeneity'–omitted variable bias. Because we have used multiple imputation to recover values and have expanded the sample to years 1999-2009 we are less worried about aggravating the trade off between missing data and omitted variable bias King et al. (2001). To address this source of endogeneity, we add additional covariates to control for some commonly identified confounders.

One potential confounder is geography. As Przeworski (2004a) cautions, "if everything is endogenous, identification is impossible: everything is simply determined by the initial conditions, which may, in turn, be shaped only by geography." In a recent application of quantile regression to political science data Alexander et al. (2011) argue that certain regions or countries have institutional peculiarities that stem from colonial history or social structure that introduce bias. This is certainly the case with corruption. Since at least Treisman (2000), colonial history and religious tradition have been identified as such factors. A lack of state-building in colonial Africa, for example, may have made these countries more dependent upon side-payments to maintain elite cohesion than other countries that enjoyed greater legitimacy within the elite or within the population at large (See Herbst (2000); Reno (1999)).

Another potential confounder is time. Temporal correlation is not as much of an issue for the 2001-2003 dataset as year to year changes in the dependent variable may reflect changes in methodology rather than substance. Indeed, the dependent variable is essentially a ranking and is not designed to be compared over time.[12] Nevertheless, variation in the covariates across time may be important. It may be the case, for example, that the ability of international financial institutions to compel leaders to 'crack down' on corruption has changed. It is also possible that the very success over time of indices have led leaders to create politics to improve their rank, known as 'rank-seeking' Høyland et al. (2012). We address this in a relatively simple way, by adding a single variable to control for time elapsed since 1999 as a linear confounder.

A third potential confounder is "development". For example, Peyton and Belasen (2012) have shown that including an indicator for "emerging and developing economies" in cross-country analysis may reveal that increased political freedom or democratization only reduces corruption in developed countries. Indeed, democracy may *increase* corruption in developing countries. Further, there is the problem of selection on observables, such a GDP per capita. As Przeworski (2004b) explains, "since almost all poor countries are dictatorships and most wealthy countries are democracies, if we look only at the observed world we will conclude that economies grow faster under dictatorships". To address this, we have included a control for whether the country in question was an 'emerging and developing' economy as defined by the International Monetary Fund. [13]

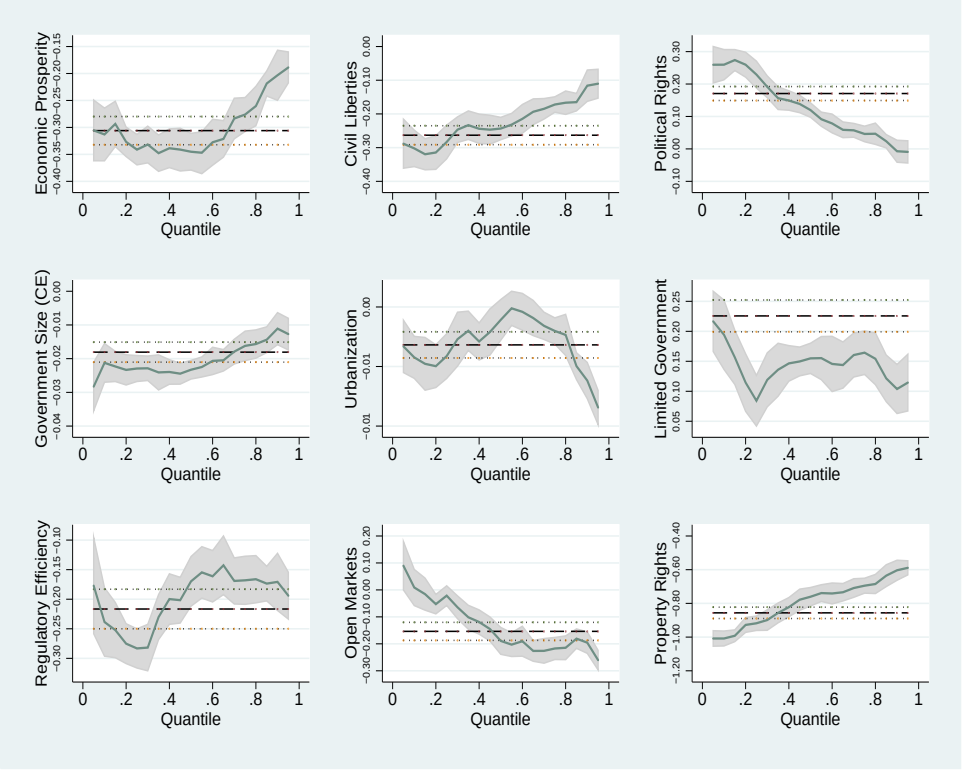
3.3 Aggregation Bias

Figure 1 shows the results from Model 2a. Here, both Freedom House and Economic Freedom have been divided into their sub-components. Recall that in Models 1a and 1b, increases in democracy were associated with *decreases* in the corruption perceptions index. Now the relationship appears

[12] See Transparency International's "Frequently Asked Questions" at `http://cpi.transparency.org/`

[13] Geographical controls also come from the IMF. See Appendix for additional explanation of data and sources.

Figure 3: OLS and Quantile Regression Estimates from Model 2a



Notes: Dependent variable is Corruption Perceptions Index (0-10). Higher values indicate higher corruption perceptions. OLS parameter estimates represented by black dotted lines with corresponding 95% confidence intervals above and below. Grey shaded area represents 95% confidence interval for quantile regression parameters.

contradictory. Although increases in Civil Liberties are associated with decreases in corruption perceptions across all quantiles, the opposite is true for Political Rights. An increase in political rights by 1 standard deviation implies a 0.51 point *increase* in the corruption perceptions score ($\approx$ 5%) for countries in the 0.10 quantile. Such an increase in Estonia circa 2005, for example, is enough to move it to the 0.25 quantile. A corresponding increase in Civil Liberties, on the other hand, would results in a change of 0.41 points ($\approx$ 5%) in the opposite direction resulting in a net change of 0.1 points ($\approx$ 1%) leaving the position in quantile unchanged. The implication is that increases in Civil Liberties *and* Political Rights doesn't matter much whereas changes in one component "all else equal" may result in substantive movements.

Further, when decomposed into its individual components: Limited Government, Regulatory Efficiency, Open Markets and Property Rights, the Economic Freedom Index also tells a different story.[14] In contrast to Models 1a and 1b, where increases in Economic Freedom were associated with decreases in corruption scores, Model 2a reveals that not all elements of 'Economic Freedom' are created equal. This may result in different substantive interpretations. For example, according to the disaggregated model, increases in the attributes of 'Limited Government', defined as Fiscal Freedom and freedom from Government Spending[15] are associated with increases in corruption perceptions. In Niger, a country in the 0.75 quantile of corruption, an increase in Limited Government

[14]Note that we have removed the 'Freedom From Corruption' measure from the 'Rule of Law' index leaving only Property Rights. For a complete outline of the indices and their components see the appendix.

[15]Note that freedom from government spending contains a transformation of government consumption expenditure that heavily penalizes countries when expenditures exceed more than 30 % of GDP which is the case for many of the least corrupt

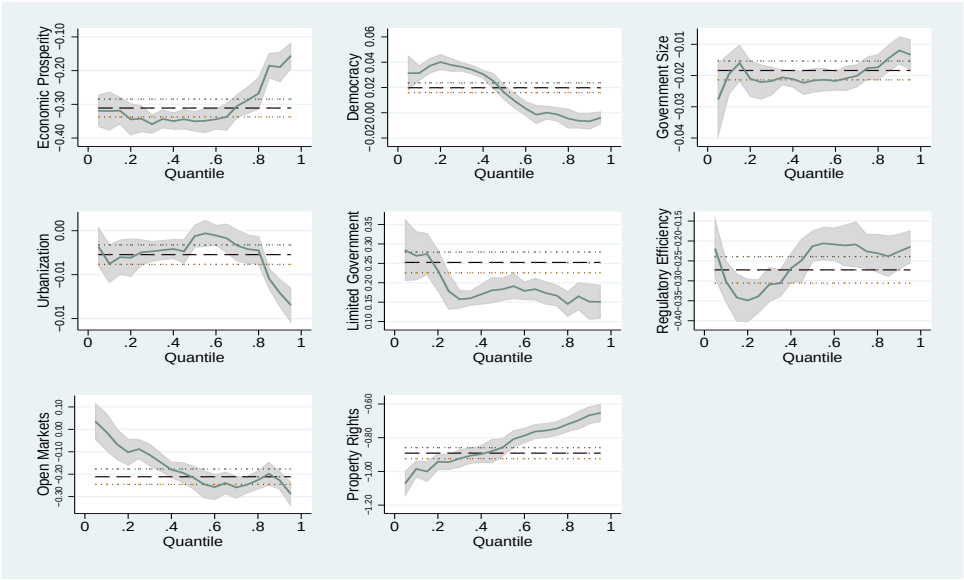
by 1 standard deviation is associated with a .14 point increase in the corruption perceptions index. Such an increase, however, could be offset by a corresponding increase in Regulatory Efficiency or *more than offset* by an increase in Open Markets or Property Rights.

Table 4: Quantile Regression v. OLS on Corruption Perceptions Index (Model2a)

	Obs.	Desc. Stats		OLS	Quantiles				
		μ	σ	Coeff.	0.1	0.25	0.5	0.75	0.9
Economic Prosperity	1991	8.56	1.32	-0.309	-0.272	-0.324	-0.344	-0.291	-0.275
				(0.036)	(0.058)	(0.051)	(0.044)	(0.059)	(0.066)
Civil Liberties	1991	3.64	1.76	-0.254	-0.261	-0.264	-0.234	-0.182	-0.161
				(0.042)	(0.073)	(0.052)	(0.050)	(0.059)	(0.075)
Political Rights	1991	3.43	2.13	0.162	0.238	0.210	0.110	0.054	0.018
				(0.027)	(0.055)	(0.040)	(0.036)	(0.046)	(0.060)
Government Size	1991	16.01	6.22	-0.018	-0.021	-0.022	-0.023	-0.019	-0.016
				(0.008)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)
Limited Government	1991	-0.05	1.00	0.217	0.227	0.125	0.153	0.140	0.100
				(0.034)	(0.086)	(0.069)	(0.042)	(0.046)	(0.054)
Regulatory Efficiency	1991	-0.08	0.99	-0.215	-0.250	-0.270	-0.175	-0.188	-0.230
				(0.0531)	(0.072)	(0.062)	(0.056)	(0.064)	(0.075)
Open Markets	1991	-0.01	0.99	-0.162	0.007	-0.044	-0.189	-0.244	-0.206
				(0.047)	(0.093)	(0.070)	(0.056)	(0.059)	(0.054)
Property Rights	1991	-0.05	1.00	-0.841	-1.030	-0.924	-0.752	-0.630	-0.490
				(0.043)	(0.059)	(0.055)	(0.062)	(0.060)	(0.064)

Notes: Results from quantile Regression with 1000 bootstrapped replications using dataset with multiple imputation for years 1999-2009. Dependent variable is Corruption Perception Index rescaled for interpretation (0-10). Higher values indicate higher corruption perceptions. Democracy is measured with Freedom House combined Civil Liberties and Political Rights scores scaled for interpretation (0-7). Higher values indicate more democratic regimes. Economic Prosperity is the natural log of GDP per capita (PPP). Urbanization (not presented here due to substantive insignificance) is the % of population living in an urban area. Government size is government consumption expenditure as % of GDP. Urbanization, Government Size and GDP data come from the World Bank. Additional covariates from Heritage Foundation Economic Freedom Index. Higher values indicate more economic freedom. Covariates not reported include time as well as regional controls for European Union, Central and Eastern Europe, Middle East and North Africa, Sub-saharan Africa, Western Hemisphere and Emerging and Developing Economies as defined by the International Monetary Fund (imf.org).

Figure 4: OLS and Quantile Regression Estimates from Model 2b



Notes: Dependent variable is Corruption Perceptions Index (0-10). Higher values indicate higher corruption perceptions. OLS parameter estimates represented by black dotted lines with corresponding 95% confidence intervals above and below. Grey shaded area represents 95% confidence interval for quantile regression parameters.

countries in the sample including Denmark, Finland, Sweden, Norway, Netherlands, Switzerland, Canada and Australia. See heritage.org for details

Indeed, Table 2a implies that the strongest reductions in corruption occur in the most corrupt countries (those in the 0.10 quantile) with increases in Open Markets, Property Rights and Economic Prosperity. Clearly, these are different concepts than Limited Government. The results from Model 2a are troubling. Given that the Freedom House measure is simply the average of scores on Civil Liberties and Political Rights, we would expect that both indicators move in the same direction. But how believable are these results?

As discussed in Lambsdorff (2007), divergent findings are not at all uncommon. As a simple test, we use an alternate measure of democracy, the Polity Index Marshall and Jaggers (2010). As Figure 4 shows, although the overall patterns are broadly the same, the new measure of democracy again changes course. Now increases in democracy are associated with increases in corruption perceptions in all but the most corrupt quantiles. Moreover, the coefficients for all covariates have changed. Increases in Property Rights, for example, one of the more robust findings, are now associated with a greater reduction in corruption perceptions than in Model 2a. Are we left to conclude that property rights are in fact more important? Does this finding provide evidence to support claims made by Billger and Goel (2009) that democracy reduces corruption in the most developed countries? An affirmative for either question is unlikely as small changes in the magnitude of coefficients are more likely the result of measurement error in the democracy variable.

Table 5: Quantile Regression v. OLS on Corruption Perceptions Index (Model2b)

	Obs.	**Desc. Stats**		**OLS**	**Quantiles**				
		μ	σ	Coeff.	0.1	0.25	0.5	0.75	0.9
Economic Prosperity	1991	8.56	1.32	-0.316	-0.323	-0.337	-0.354	-0.289	-0.188
				(0.037)	(0.062)	(0.055)	(0.046)	(0.056)	(0.079)
Democracy	1991	3.50	6.16	0.019	0.032	0.036	0.015	-0.002	-0.006
				(0.006)	(0.010)	(0.007)	(0.008)	(0.006)	(0.069)
Government Size	1991	15.97	6.28	-0.019	-0.020	-0.021	-0.021	-0.020	-0.015
				(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Limited Government	1991	-0.05	1.00	0.244	0.254	0.180	0.183	0.153	0.133
				(0.034)	(0.069)	(0.050)	(0.042)	(0.050)	(0.053)
Regulatory Efficiency	1991	-0.08	0.99	-0.272	-0.306	-0.350	-0.218	-0.231	-0.241
				(0.049)	(0.070)	(0.054)	(0.066)	(0.070)	(0.095)
Open Markets	1991	-0.01	0.99	-0.220	-0.030	-0.108	-0.228	-0.250	-0.206
				(0.046)	(0.097)	(0.056)	(0.067)	(0.057)	(0.066)
Property Rights	1991	-0.05	1.00	-0.875	-0.982	-0.931	-0.825	-0.721	-0.661
				(0.043)	(0.070)	(0.052)	(0.077)	(0.066)	(0.064)

Notes: Results from quantile Regression with 1000 bootstrapped replications using dataset with multiple imputation for years 1999-2009. Dependent variable is Corruption Perceptions Index rescaled for interpretation (0-10). Higher values indicate higher corruption perceptions. Economic Prosperity is the natural log of GDP per capita (PPP). Urbanization is the % of population living in an urban area. Government size is government consumption expenditure as % of GDP. Urbanization (not presented here due to substantive insignificance), Government Size and GDP data come from the World Bank. Additional covariates from Heritage Foundation Economic Freedom Index. Higher values indicate more economic freedom. Covariates not reported include time as well as regional controls for European Union, Central and Eastern Europe, Middle East and North Africa, Sub-saharan Africa, Western Hemisphere and Emerging and Developing Economies as defined by the International Monetary Fund (imf.org).

Nevertheless, a few substantive findings are worth noting. First, components of the Economic Freedom Index are correlated with corruption in different ways. For example, results from the previous models suggest that increases in Regulatory Efficiency, Open Markets, and Property Rights may all have a substantive impact on corruption perceptions across quantiles. The negative relationship between corruption and Open Markets, however, contradicts previous findings that Foreign Direct Investment is unrelated to corruption Lambsdorff (2007). One can only speculate on causality, however. It could very well be the case that countries with Open Markets are simply less corrupt. Or it could be the case that this relationship is driven by 'rank-seeking' Høyland et al. (2012). In 2011, for example, Georgia took out a 1 page advertisement in *The Economist* claiming to be "the World's number 1 in fighting corruption" according to corruption data from Transparency International, presumably to attract investors (February 5-11, 2011).

The divergent findings in Model 2a and 2b point us towards an interesting conjecture regarding democracy. According to Bollen (1990) the Freedom House indicator of Political Rights measures the extent to which "the national government is accountable to the general population and each individual is entitled to participate in the government directly or through representatives" whereas Civil

Liberties "exist to the extent that the people of a country have the freedom to express any political opinions in any media and the freedom to form or to participate in any political group". Put another way, the 'Political Rights' variable may be an approximation of the presence of democratic *institutions* that facilitate some level of popular participation in political decision-making while the 'Civil Liberties' variable is a measure of the extent to which *society* can freely express it's preferences. The Polity Index more closely approximates the Political Rights indicator rather than the Civil Liberties indicator Marshall and Jaggers (2010). Indeed, the Index aggregates three aspects of the institutional framework governing 'executive recruitment', the independence of executive authority and 'political competition and opposition' (see Treier and Jackman (2008) for a detailed treatment).

A generous interpretation might be that measures of Political Rights and the institutions of central government should yield similar results. Not only do both the Freedom House measure of Political Rights and the Polity Index indicate the general tendency for countries with liberal political institutions to have higher levels of perceived corruption, but this relationships diminishes in both magnitude and statistical significance at the 75th and 90th quantiles of corruption. Based on these results, it appears unlikely that democratizing the institutions of central government by opening the political process to popular participation, for example, will reduce the level of corruption in the most corrupt countries. In countries with moderate to low levels of corruption, however, these changes may increase the perceived level of corruption. Perhaps opening the political process to popular participation also opens the process to bribery. For countries that are perceived as grievously corrupt, this may be especially relevant.

Civil Liberties, on the other hand, operate differently. Countries that are more liberal with respect to the expression of public preferences through the media and public assembly have lower levels of corruption. Again, this relationships is strongest in the lowest and middle quantiles of corruption. The ability to embarrass leaders outside political institutions may be an important corruption-reducing mechanism and opposition parties that are capable of exposing the malfeasance of a ruling party or leader may have an important role to play in combating corruption.

3.4 Final Specification: Quality of Government

We are less concerned with the same measurement issues ascribed to the Freedom House and Polity measures in Høyland et al. (2012) and Treier and Jackman (2008) in the dependent variable. Indeed, Lambsdorff's (2007, Chapter 1) analysis of Transparency International's Index and methodology suggests it may be more robust for statistical analysis than the Polity and Freedom House indices. Nevertheless, it's worth testing whether or not measurement error in Y has an impact on the associations above. One option is to use a different measurement instrument for the dependent variable. This also allows us to test for construct validity in the dependent variable. That is, are indicators empirically associated in a way that "conforms to theoretical expectations about their interrelationship" Adcock and Collier (2001). For example, Billger and Goel (2009) argue that "...a larger government might be associated with stronger checks and balances (i.e., better oversight) and in this case corruption might actually decrease with government size". To evaluate this claim, a broader measure, Quality of Government (QoG) is used Rothstein (2011). Does the positive correlation between political rights and corruption, for example, hold in both models?

Table 1 provides some examples of countries in different quantiles of QoG.

Table 6: Countries by Quantile – Quality of Government Index

Quantile	Country	Score (2005)
0.10	Cameroon	0.31
	Papua New Guinea	0.35
	Bangladesh	0.361
	Sierra Leone	0.28
0.25	South Africa	0.43
	Bolivia	0.44
	Russia	0.42
0.5	Egypt	0.47
	Mexico	0.53
	China	0.53
0.75	Baharain	0.56
	Greece	0.65
	Poland	7.59
	Turkey	0.55
	Argentina	0.55
0.90	Austria	0.94
	United Kingdom	0.88
	Spain	0.75

Notes: drawn from sample of 181 countries with multiple imputation. All data in this paper will be made available for replication.

Recall the scaling of QoG is 0-1 where 1 indicates higher quality of government. Hence, we expect the most corrupt countries measured by TI to be those with the lowest QoG scores. For example, those falling in the 0.75 quantile in QoG should fall in the 0.25 quantile of the TI Corruption Score. That is, countries in which 25% of the sample have higher quality of government should also fall such that 75% of the countries in the sample are more corrupt. Table 6 provides some cursory evidence that the QoG and TI Corruption Index have some conceptual validity as 10/15 countries follow this rule.[16] Exceptions include South Africa, Turkey, Poland, Argentina and Sierra Leone. For the extreme quantiles – the most corrupt and least corrupt in Table 1 – the match is perfect, those that have the highest quality of government (0.90 quantile of QoG) fall in the corresponding 0.10 quantile of corruption. Bangladesh, for example, is among the 'most corrupt' and also has among the lowest QoG score. The mapping does not work so well for some quantiles. South Africa, for example, appears in the 0.25 quantile according to both the TI and QoG. That is, 75% of the countries in the sample are *more corrupt* than South Africa yet 75% of the countries in the sample also have higher quality of government than South Africa.

Tables 6 and 7 demonstrate that most findings reported in Models 2a and 2b are robust to this change in dependent variable. Model 3a implies that increases in Civil Liberties are associated with increases in QoG across the board. Increases in Political Rights, however, decrease the Quality of Government, again with the strongest effects reserved for those countries scoring worst on the index. Finally, when we use the Polity Index, the findings of Model 2b are broadly upheld. As countries become more democratic, they also experience a deterioration in Quality of Government. Take, Estonia in 2005, for example, a country in the .75 quantile of Quality of Government in 2005 with a democracy score of 9. According to the model, an increase in democracy by one standard deviation (6.11 points) relative to the sample would result in a 0.02 reduction in QoG from .60 to .58, or 11 % of one standard deviation in QoG. Such a change would not alter it's relative position in a meaningful way – it would still be in the 0.75 quantile of QoG and better off than 75 % of the sample.

[16] As a more robust test of measurement validity we might compare the mapping from QoG quantiles to TI quantiles. This is perhaps a new study altogether

TABLE 7: Quantile Regression v. OLS on Quality of Government Index: (Model3a)

	Obs.	Desc. Stats		OLS	Quantiles				
		μ	σ	Coeff.	0.1	0.25	0.5	0.75	0.9
Economic Prosperity	1991	8.56	1.32	0.030	0.032	0.028	0.022	0.025	0.021
				(0.004)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Civil Liberties	1991	3.64	1.76	0.015	0.012	0.015	0.015	0.019	0.021
				(0.005)	(0.007)	(0.006)	(0.005)	(0.006)	(0.008)
Political Rights	1991	3.57	2.13	-0.011	-0.000	-0.008	-0.013	-0.017	-0.017
				(0.003)	(0.006)	(0.005)	(0.004)	(0.004)	(0.006)
Government Size	1991	15.97	6.28	0.002	0.004	0.003	0.002	0.001	0.001
				(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Property Rights	1991	-0.05	1.00	0.063	0.057	0.065	0.070	0.063	0.064
				(0.005)	(0.010)	(0.009)	(0.012)	(0.008)	(0.009)
Limited Government	1991	-0.05	1.00	-0.013	0.002	-0.013	-0.021	-0.022	-0.029
				(0.004)	(0.007)	(0.009)	(0.005)	(0.007)	(0.010)
Regulatory Efficiency	1991	-0.08	0.99	0.024	0.026	0.012	0.024	0.020	0.007
				(0.005)	(0.010)	(0.008)	(0.007)	(0.005)	(0.013)
Open Markets	1991	-0.01	0.99	0.004	0.008	0.016	0.005	-0.003	-0.010
				(0.005)	(0.010)	(0.008)	(0.006)	(0.006)	(0.008)

Notes:
Results from quantile Regression with 1000 bootstrapped replications using dataset with multiple imputation for years 1999-2009. Dependent variable is Quality of Government Index (0-1). Higher values indicate higher Quality of Government. Civil Liberties and Political Rights measured by Freedom House rescaled for interpretation (0-7). higher values indicate more freedom. Economic Prosperity is the natural log of GDP per capita (PPP). Urbanization (not presented here due to substantive insignificance) is the % of population living in an urban area. Government size is government consumption expenditure as % of GDP. Urbanization, Government Size and GDP data come from the World Bank. Additional covariates from Heritage Foundation Economic Freedom Index. Higher values indicate more economic freedom. Covariates not reported include time as well as regional controls for European Union, Central and Eastern Europe, Middle East and North Africa, Sub-saharan Africa, Western Hemisphere and Emerging and Developing Economies as defined by the International Monetary Fund (imf.org)

TABLE 8: Quantile Regression v. OLS on Quality of Government Index: (Model3b)

	Obs.	Desc. Stats		OLS	Quantiles				
		μ	σ	Coeff.	0.1	0.25	0.5	0.75	0.9
Economic Prosperity	1991	8.56	1.32	0.027	0.030	0.026	0.020	0.022	0.025
				(0.004)	(0.006)	(0.005)	(0.005)	(0.004)	(0.005)
Democracy	1991	3.50	6.16	-0.002	-0.001	-0.003	-0.003	-0.003	-0.002
				(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Government Size	1991	15.97	6.28	0.002	0.004	0.003	0.001	0.000	0.001
				(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Property Rights	1991	-0.05	1.00	0.065	0.061	0.071	0.066	0.065	0.063
				(0.005)	(0.010)	(0.008)	(0.009)	(0.008)	(0.008)
Limited Government	1991	-0.05	1.00	-0.015	-0.018	-0.012	-0.023	-0.024	-0.030
				(0.005)	(0.008)	(0.010)	(0.005)	(0.007)	(0.010)
Regulatory Efficiency	1991	-0.08	0.99	0.027	0.034	0.026	0.028	0.023	0.010
				(0.005)	(0.011)	(0.007)	(0.007)	(0.006)	(0.009)
Open Markets	1991	-0.01	0.99	0.009	0.019	0.021	0.012	0.002	-0.007
				(0.005)	(0.012)	(0.009)	(0.006)	(0.006)	(0.008)

Notes: Results from quantile Regression with 1000 bootstrapped replications using dataset with multiple imputation for years 1999-2009. Dependent variable is Quality of Government Index (0-1). Higher values indicate higher Quality of Government. Economic Prosperity is the natural log of GDP per capita (PPP). Urbanization is the % of population living in an urban area. Government size is government consumption expenditure as % of GDP. Urbanization (not presented here due to substantive insignificance), Government Size and GDP data come from the World Bank. Additional covariates from Heritage Foundation Economic Freedom Index. Higher values indicate more economic freedom. Covariates not reported include time as well as regional controls for European Union, Central and Eastern Europe, Middle East and North Africa, Sub-saharan Africa, Western Hemisphere and Emerging and Developing Economies as defined by the International Monetary Fund (imf.org).

Quantile regression is clearly a useful analytical tool. It allows researchers to examine how relationships between corruption and some covariates change across quantiles of the distribution. It does not, however, in the absence of exogenous variation or formal hypotheses, establish causality when applied to cross-country analysis any more than OLS methods. Indeed, in a small sample size OLS may provide more precise estimates. Measurement is also an issue. As models with different dependent variables demonstrate, choice of instrument can substantially change the results. These

problems are aggravated by listwise deletion. Although valuable as descriptive tools, sophisticated methods applied to imperfectly measured observational data from surveys we know little about [17] are not a substitute for knowledge of the data generation process.

4 Conclusions

The question of whether democratization will lead to more or less corruption, especially in the world's most corrupt countries, is one of considerable importance. In this paper we subjected an innovative study of this topic Billger and Goel (2009) to three robustness tests. First, we used multiple imputation instead of listwise deletion to dramatically increase the number of valid observations and test the extent to which these findings were dependent upon the data used. We then expanded the time period to cover a decade, added regional and temporal control variables and disaggregated 'Economic Freedom' and 'Political Freedom' into their component parts. With these improvements we re-tested the democracy-corruption link by, first, splitting the Freedom House indicator into it's component parts (Political Rights and Civil Liberties) and, second, using an alternative measure of democracy (the Polity Scale). We then tested the relationship with an alternative dependent variable - the Quality of Government Index. Billiger and Goel's finding that democracy reduces perceptions of corruption, and does so differentially in the most corrupt countries, did not stand up to these tests.

In fact, we found the opposite result – the relationship between corruption and democracy is *weaker* in the most corrupt countries. We conclude that, like most studies of corruption, the results are data dependent. The findings from the disaggregated Freedom House index were striking. Improvements in Civil Liberties appear to decrease corruption perceptions while liberalization of Political Rights has the opposite relationship. Both these associations were strongest in the low to moderately corrupt countries and both were broadly robust to the change in dependent variable. We have interpreted this as a substantive finding worth further examination. Indeed, this may call into question previous findings that rely on the Freedom House measure of democracy.

As Johnston (2005) has cautioned, "rank-ordering countries from high to low corruption effectively defines the problem as the same everywhere." Quantile regression, as an analytic tool, suggests that little can be gained from conventional blanket solutions to corruption such a "cutting big government" Becker (1994). Indeed, our findings suggest these reforms may be harmful at worst or irrelevant at best. Social scientists doing cross-national analysis with indices should pay increased attention to missing data and measurement error. Decomposition may alter substantive conclusions and previous inferences drawn from observational studies may be data dependent. As such, they should be interpreted with caution. Future research using indices should place emphasis on these issues.

One obvious criticism is that we didn't control for enough confounders. We acknowledge this as valid, however, this is a task for another day. Another approach to dealing with a second source of endogeneity, 'simultaneity bias' between corruption and democracy, is to measure democracy as a 'stock' rather than a flow variable. As Przeworski and Limongi (1997) and Przeworski et al. (2000) have argued, democracy may be subject to path dependency. That is, once a country is democratic, it tends to stay democratic unless some minimum threshold of economic prosperity is breached. Treisman (2000); Gerring (2004) ; Gerring (2004) have provided evidence that although corruption may not appear to be related to the *current* level of democracy, *established democracies* are in fact less corrupt. This too requires a rich data source to model in a meaningful way–a task for the future. Other identification strategies include instrumental variables Acemoglu et al. (2008). Indeed, Alexander et al. (2011) provide extensions to the quantile regression framework which allow for the inclusion of instrumental variables and fixed effects. This is also a strategy for the future, and one that requires additional data sources and substantive knowledge.

[17] Freedom House, for example, does not release the survey questions they develop to the public

References

Abramoff, Jack (2011), *Capitol Punishment*. Washington, DC: WND Books.

Abrevaya, Jason and Jerry Hausman (1999), "Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells", *Annales D'Economie et de Statistique*, 55-56: 243-275.

Acemoglu, Daron, Simon Johnson, James A. Robinson and Pierre Yared (2008), "Income and Democracy", *American Economic Review*, 98(3): 808-842.

Adserà, Alícia, Carles Boix and Mark Payne (2001), "Measurement Validity: A shared Standard for Qualitative and Quantitative Research", *American Political Science Review*, 95(3): 529-546.

Adserà, Alícia, Carles Boix and Mark Payne (2003), "Are you being served? Political accountability and quality of government", *Journal of Law, Economics, & Organization*, 19(2): 445-90.

Alexander, Marcus, Matthew Harding and Carlos Lamarche (2011), "Quantile Regression for Time-Series-Cross-Section Data", *International Journal of Statistics and Management System*, 6(1-2): 47-72.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics*. Princeton, New Jersey: Princeton University Press.

Becker, Gary S. (1994), "To root our corruption, boot out big government", *Business Week*, 18, January 31.

Billger, Sherrilyn M. and Rajeev K. Goel (2009), "Do existing corruption level matter in controlling corruption? Cross-country quantile regression estimates", *Journal of Development Economics*, 90(2): 299-305.

Bohara, Alok K., Neil J. Mitchell and Carl F. Mittendorff (2004), "Compound Democracy and the Control of Corruption: A Cross-Country Investigation", *Policy Studies Journal* 32(4): 481-99.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, Kenneth A. (1990), "Political Democracy: Conceptual and Measurement Traps", *Studies in Comparative International Development* 25(1): 7-24.

Bollen, Kenneth A. (2004), "Neoliberal Reform, Governance and Corruption in Central America: Exploring the Nicaraguan Case", *Political Geography* 24(5): 601-30.

Card, David (2001), "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems", *Econometrica*, 69(5): 1127-1160.

Card, David (2001), "The Effect of Democracy and Press Freedom on Corruption: An Empirical Test", *Economics Letters*, 85(1): 93-101.

Elliot, Kimberly Ann (1997), "Corruption as an international policy problem: overview and recommendations", in K.A. Elliot (ed.),*Corruption and the Global Economy*, Washington, DC: Institute for International Economics, pp. 175-233.

Fisman, Raymond and Robert Gatti (2002), "Decentralization and corruption: evidence across countries", *Journal of Public Economics*, 83(3): 325-345.

Freedman, David A. (1987), "As Other See Us: A Case Study in Path Analysis", *Journal of Educational Statistics*, 12(2): 101-128.

Freedman, David A. (2004), "Graphical Models for Causation and the Identification Problem", *Evaluation Review*, 28(4): 267-293.

Freedman, David A. (2005), *Statistical Models: Theory and Practice*. New York: Cambridge University Press.

Gerring, John and Strom C. Thacker (2004), "Political Institutions and Governance: Pluralism versus Centralism", *British Journal of Political Science*, 34(2): 295-30.

Gerring, John and Strom C. Thacker (2005), "Do Neoliberal Policies Deter Political Corruption?", *International Organization*, 59(1): 233-254.

Gemeci, Sinan, Alice Bednarz and Patrick Lim (2011), *Getting Tough on Missing Data: A Boot Camp for Social Science Researchers*, National Centre for Vocational Education Research Technical Report, [online] Available from: .

Goel, Rajeev K. and Michael A. Nelson (1998), 'Corruption and government size: a disaggregated analysis', *Public Choice*, 97(1): 107-20.

Goel, Rajeev K. and Michael A. Nelson (2005), "Economic freedom versus political freedom: cross-country influences on corruption", *Australian Economic Papers*, 44(2): 121-33.

Graeff, Peter and Guido Mehlkop (2003), "The impact of economic freedom on corruption: different patterns for rich and poor countries", *European Journal of Political Economy*, 19(2): 605-20.

Hausman, Jerry (2001), "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left", *The Journal of Economic Perspectives*, 15(4): 57-67.

Herbst, Jeffrey (2000), "Economic Incentives, Natural Resources and Conflict in Africa", *Journal of African Economies*, 9(3): 270-294.

Honaker, James, Gary King and Matthew Blackwell (2009). AMELIA II: A Program for Missing Data. [online] Available from: http://j.mp/k4t8Ej.

Honaker, James and Gary King (2010), "What to do about Missing Values in Time Series Cross-Section Data", *American Political Science Review*, 54(2): 561-581.

Høyland, Bjørn, Karl Moene and Fredrik Willumsen (2012), "The tyranny of international index rankings", *Journal of Development Economics*, 97(1): 1-14.

Jackman, Simon (2004), "What Do We Learn from Graduate Admissions Committees? A Multiple Rater, Latent Variable Model, with Incomplete Discrete and Continuous Indicators", *Political Analysis*, 12(4): 400-424.

Jain, Arvind K. (2001), "Corruption: A Review", *Journal of Economic Surveys*, 15(1): 71-121.

Johnston, Michael (1996), "The Search for Definitions: The Vitality of Politics and the Issue of Corruption", *International Social Science Journal*, 48(3): 321-335.

Johnston, Michael (2005), *Syndromes of Corruption: Wealth, Power, and Democracy*. New York: Cambridge University Press.

King, Gary, James Honaker, Ann Joseph and Kenneth Scheve. (2001), "Analyzing Incomplete Political Science Data", *American Political Science Review*, 95(1): 49-69.

Koenker, Roger (2005), *Quantile Regression*. New York: Cambridge University Press.

Koenker, Roger and Gilbert Bassett, Jr. (1978), "Regression Quantiles", *Econometrica*, 46(1): 33-50.

Koenker, Roger and Kevin F. Hallock (2001), "Quantile Regression", *Journal of Economic Perspectives*, 15(4): 143-156.

Krueger (1974), "The Political Economy of the Rent-Seeking Society", *American Economic Review*, 64(3): 291-303.

Lambsdorff, Johann Graf (2006), "Causes and consequences of corruption: What do we know from a cross-section of countries?", in Susan Rose-Ackerman (ed.), *International Handbook on the Economic of Corruption*, Edward Elgar Publishing, Cheltenham, pp. 3-51.

Lambsdorff, Johann Graf (2007), *The Institutional Economics of Corruption and Reform: Theory, Evidence and Policy*. New York: Cambridge University Press.

La Porta, Rafael, Florencio Lopez-de-Silanis, Andrei Shleifer and Robert Vishny (1999), "The Quality of Government", *Journal of Law, Economics and Organization*, 15(1): 222-79.

LaPalombara, Joseph (1994), "Structural and institutional aspects of corruption", *Social Research*, 61, pp. 325-50.

Little, Roderick J. A. (1988), "A Test of Missing Completely at Random for Multivariate Data with Missing Values", *Journal of the American Statistical Association*, 83(404): 1198-1202.

Little, Roderick and Donald B. Rubin (2002), *Statistical Analysis with Missing Data*. Hoboken: John Wiley & Sons, Inc.

Marshall, Monty G. and Keith Jaggers (2010), *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010* (Data Users Manual), Center for Systemic Peace. [online] Available from: systemicpeace.org

Montinola, Gabriella R. and Robert W. Jackman (2002), "Sources of Corruption: A Cross-Country Study", *British Journal of Political Science*, 32(1): 147-170.

Mueller, Dennis C. (2003), *Public Choice III*. Cambridge: Cambridge University Press.

Morris, Stephen D. (2009), *Political Corruption in Mexico: The Impact of Democratization*. Boulder: Lynne Rienner.

Nye, Joseph S. (1967), "Corruption and Political Development: A Cost-Benefit Analysis", *American Political Science Review*, 61(2): 417-427.

Paldam, Martin (2002), "The Cross-Country Pattern of Corruption: Economics, Culture and the See-Saw Dynamics", *European Journal of Political Economy*, 18(2): 215-220.

Peyton, Kyle and Ariel R. Belasen (2012), "Corruption in Emerging and Developing Economies: Evidence from a Pooled Cross-Section", *Emerging Markets Finance and Trade*, 48(2): 29-43.

Przeworski, Adam (2004), "Institutions Matter?", *Government and Opposition*, 39(2): 527-540.

Przeworski, Adam (2004), "Geography vs. Institutions Revisited: Were Fortunes Reversed?", mimeo, New York University. [online] Available from: `http://politics.as.nyu.edu/docs/IO/2800/reversal.pdf`

Przeworski, Adam and Fernando Limongi (1997), "Modernization: Theories and Facts", *World Politics*, 49(2): 155-183.

Przeworski, Adam, Michael E. ALvarez, José Antonio Cheibub and Fernando Limongi (2000), *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge: Cambridge University Press.

Philp, Mark (1997), "Defining Political Corruption," in Paul Heywood, ed., *Political Corruption*. Oxford: Blackwell, pp. 20-46.

Philp, Mark (2002), "Conceptualizing Political Corruption," in Arnold J. Heidenheimer and Michael Johnston, eds., *Political Corruption: Concepts and Contexts*. New Brunswick, NJ: Transaction Publishers, pp. 41-57.

Reno, William (1999), *Warlord Politics and African States*. Boulder: Lynne Rienner.

Robinson, James (2006), "Economic Development and Democracy", *Annual Review of Political Science*, 9(1): 503-27.

Rose-Ackerman, Susan (1999), *Corruption and Government: Causes, Consequences, and Reform*. New York: Cambridge University Press.

Rothstein, Bo (2011), *The Quality of Government*. Chicago: University of Chicago Press.

Scott, James C. (1972), *Comparative Political Corruption*. Englewood Cliffs, NJ: Prentice-Hall.

Schaefer, J.L. (1997), *Analysis of Incomplete Multivariate Data*. Boca Raton: Chapman & Hall.

Shen, Ce and John B. Williamson (2005), "Corruption, Democracy, Economic Freedom, State Strength", *International Journal of Comparative Sociology*, 46(4): 327-345.

Tavares, José and Romain Wacziarg (2001), "How Democracy Affects Growth", *European Economic Review*, 45(8): 1347-78.

Teorell, Jan, Nicholas Charron, Marcus Samanni, Sören Holmberg and Bo Rothstein (2011), The Quality of Government Dataset, version 6Apr11. University of Gothenburg: The Quality of Government Institute, http://www.qog.pol.gu.se

Thompson, Dennis F. (1993), "Mediated Corruption: The Case of the Keating Five", *American Political Science Review*, 87(2): 369-381.

TI India (2005). *India Corruption Study 2005*. Transparency International India. [online] Available from: www.transparency.org/regional_pages/asia_pacific/newsroom/news_archive2/india_corruption_study

Treier, Shawn and Simon Jackman (2008), "Democracy as a Latent Variable", *American Journal of Political Science*, 52(1): 201-217.

Treisman, Daniel (2000), "The Causes of Corruption: A Cross-National Study", *Journal of Public Economics*, 76(3): 399-457.

Wei, Shang-Jin (1999), "Corruption in Economic Development: Beneficial Grease, Minor Annoyance, or Major Obstacle?" World Bank Policy Research Working Paper No. 2048. [online] Available from: papers.ssrn.com

World Bank (1997). *Helping Countries Combat Corruption: The Role of the World Bank*. Washington, DC: World Bank. [online] Available from: World Bank.org.

A Data and Imputations

Table 1: Data definitions

Variable	Description	Source
Corruption	the Corruption Perceptions Index	Transparency International
Democracy (Model 1a)	Sum of 'Political Rights' and 'Civil Liberties' Indices	Freedom House
Democracy (Models 1b)	Average of 'Political Rights' and 'Civil Liberties' Indices	Freedom House
Democracy (Models 2a & 3b)	Split'Political Rights' and 'Civil Liberties' Indices	Freedom House
Democracy (Models 2b & 3a)	Revised Combined Polity Score from QoG Standard Dataset	QoG Institute
Economic Freedom	Heritage Foundation's Economic Freedom Index	Heritage Foundation
Economic prosperity	GDP per capita (constant 2000 USD)	The World Bank
Government Size	Government Consumption Expenditure (% of GDP)	The World Bank
Quality of Government	QoG Standard Dataset	QoG Institute
Urbanization	Urban population (% of total)	The World Bank
IMF Groupings	World Economic Outlook	imf.org

Notes: According to Freedom House, the appropriate way to combine the sub-components is by averaging Civil Liberties and Political Rights rather than summation. IMF Groupings includes regional variables and "emerging and developing economies" classification. Variables not otherwise described come from QoG Standard Dataset

Freedom House's democracy score and Heritage Foundation's Economic Freedom Index can both be divided into multiple components components. The Polity Index can also be divided into multiple components, though in a more complicated way. See Treier and Jackman (2008) for a thorough explanation.

Table 2: Variable Sub-components

Variable	Components	Type
Political Freedom		
Political Rights	Electoral Process	Ordinal
	Political Pluralism and participation	Ordinal
	Functioning of government	Ordinal
Civil Liberties	Freedom of expression and belief	Ordinal
	Associational and organizational rights	Ordinal
	Rule of Law	Ordinal
	Personal autonomy and individual rights	Ordinal
Economic Freedom		
Rule of Law	Property Rights	Ordinal
	Freedom from Corruption	Continuous
Limited Government	Fiscal Freedom	Continuous
	Government Spending	Continuous
Regulatory Efficiency	Business Freedom	Ordinal
	Labor Freedom	Continuous
	Monetary Freedom	Continuous
Open Markets	Trade Freedom	Continuous
	Investment Freedom	Ordinal
	Financial Freedom	Ordinal
Quality of Government		
Corruption	6 Points	Unknown
Law and Order	6 Points	Unknown
Bureaucracy Quality	4 Points	Unknown

Notes: The overall freedom house score is the average of the individual sub-components, Political Rights and Civil Liberties. The overall score and the sub-components range from 0(most free) to 7(least free). The overall Economic Freedom score is the mean of individual sub-components. The Index and the sub-components are on a scale of 0-100 with 100 representing "most free". Heritage Foundation's Freedom from Corruption comes "primarily" from Transparency International's Corruption Perceptions Index although it is a continuous variable in the primary source. Labor freedom was not measured prior to 2004. Business freedom became an continuous variable from 2006. The QoG index is composed of mean values of the components "Corruption", "Law and Order" and "Bureaucracy Quality". 0-1 Scale. Higher values indicate higher quality of government.

Billger & Goel Dataset (2001-2003) results from multiple imputation using Amelia II with linear time effects for time-series cross section data. Note that the sample was too small to add intersections. ti_cpi was bounded between 0-1, 5% ridge priors used to shrink covariances between variables.

Peyton & Butcher Dataset (1999-2009) results from multiple imputation using Amelia II with linear time effects with intersections for time-series cross section data. ti_cpi was bounded between 0-1, 5% ridge priors used to shrink covariances between variables.

Table 3: Billger & Goel Listwise Deletion

	μ	σ	min	max
Corruption(TI)	4.38	2.27	1.30	9.70
Quality of Government	0.55	0.20	0.19	1.00
Control of Corruption	0.17	1.04	-1.27	2.41
ln(GDP per capita)	8.79	1.24	5.31	10.65
Democracy (Polity)	4.97	6.05	-10.00	10.00
Civil Liberties (FH)	3.10	1.66	1.00	7.00
Political Rights (FH)	2.95	1.94	1.00	7.00
Democracy (BG)	6.05	3.54	2.00	14.00
Government size	15.70	5.02	5.35	27.78
Urbanization	59.96	20.82	12.34	100.00
Business start	49.45	33.30	2.00	168.00
Economic Freedom (No CP)	0.22	0.91	-2.51	2.67
Economic Freedom (HF)	0.21	0.92	-2.19	2.72
Economic Freedom (BG)	0.22	0.90	-2.38	2.21
Business Freedom	0.13	0.81	-1.56	2.45
Freedom from Corruption	0.18	1.03	-1.55	2.52
Financial Freedom	0.22	0.99	-2.05	1.93
Fiscal Freedom	-0.05	0.98	-2.94	1.95
Government Spending	-0.06	1.01	-2.88	1.26
Investment Freedom	0.28	0.91	-2.22	2.00
Monetary Freedom	0.19	0.90	-4.48	1.11
Property Rights	0.15	0.96	-1.62	1.75
Trade Freedom	0.02	0.96	-3.43	1.21
Emerging & Developing	0.74	0.44	0.00	1.00
European Union	0.12	0.33	0.00	1.00
Central & Eastern Europe	0.10	0.29	0.00	1.00
Middle East & North Africa	0.10	0.29	0.00	1.00
Sub-saharan Africa	0.18	0.39	0.00	1.00
Western Hemisphere	0.18	0.39	0.00	1.00

Notes: Categorical classifications come from IMF.org. European Union includes Austraia, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Portugal, Slovakia, Slovenia and Spain

Table 4: Billger & Goel Multiple Imputation

	μ	σ	min	max
Corruption(TI)	3.90	2.24	0.05	9.90
Quality of Government	0.52	0.19	0.11	1.00
Control of Corruption (WB)	-0.02	1.01	-2.12	2.90
ln(GDP per capita)	8.43	1.31	5.31	11.01
Democracy (Polity)	3.58	6.46	-10.00	10.00
Civil Liberties (FH)	3.46	1.75	1.00	7.00
Political Rights (FH)	3.43	2.13	1.00	7.00
Democracy (BG)	6.89	3.81	2.00	14.00
Government Size	16.24	6.94	-6.29	83.16
Urbanization	52.69	23.08	8.54	100.00
Business Start	53.44	34.54	2.00	202.00
Economic Freedom (No CP)	-0.09	1.06	-4.73	2.67
Economic Freedom (HF)	-0.10	1.06	-4.20	2.72
Economic Freedom (BG)	-0.12	1.03	-4.23	2.21
Business Freedom	-0.11	0.93	-3.57	2.45
Freedom from Corruption	-0.06	1.03	-2.04	2.56
Financial Freedom	-0.08	1.02	-2.58	1.93
Fiscal Freedom	-0.14	1.03	-4.32	2.62
Government Spending	0.01	1.01	-5.67	1.75
Investment Freedom	0.05	0.99	-2.51	2.00
Monetary Freedom	0.01	1.07	-5.52	2.21
Property Rights	0.01	0.99	-2.43	1.75
Trade Freedom	-0.23	1.00	-4.43	1.44
Emerging & Developing	0.82	0.38	0.00	1.00
European Union	0.09	0.28	0.00	1.00
Central & Eastern Europe	0.08	0.27	0.00	1.00
Middle East & North Africa	0.08	0.27	0.00	1.00
Sub-saharan Africa	0.24	0.43	0.00	1.00
Western Hemisphere	0.18	0.38	0.00	1.00

Notes: Categorical classifications come from IMF.org. European Union includes Austraia, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Portugal, Slovakia, Slovenia and Spain

Table 5: Peyton & Butcher Listwise Deletion

	μ	σ	min	max
Corruption(TI)	4.32	2.16	1.60	9.63
Quality of Government	0.55	0.19	0.22	1.00
Control of Corruption (WB)	0.08	0.99	-1.35	2.58
ln(GDP per capita)	8.99	1.22	5.22	11.29
Democracy (Polity)	5.05	5.96	-10.00	10.00
Civil Liberties (FH)	2.91	1.58	1.00	7.00
Political Rights (FH)	3.06	2.01	1.00	7.00
Democracy (BG)	5.97	3.53	2.00	14.00
Government Spending	14.96	4.98	4.58	27.19
Urbanization	60.25	21.21	12.20	100.00
Business Start	53.44	34.54	2.00	202.00
Economic Freedom (no CP)	0.18	0.85	-2.62	2.70
Economic Freedom (HF)	0.26	0.88	-2.79	2.16
Economic Freedom (BG)	0.19	0.87	-2.33	2.76
Business Freedom	0.16	0.97	-2.58	2.45
Freedom from Corruption	0.12	0.94	-1.29	2.43
Financial Freedom	0.16	0.96	-2.05	1.93
Fiscal Freedom	0.12	0.92	-2.73	1.95
Government Spending	-0.06	0.98	-3.01	1.28
Investment Freedom	0.07	1.00	-2.22	2.00
Labor Freedom	0.02	0.95	-2.12	2.47
Monetary Freedom	0.17	0.64	-5.52	1.30
Property Rights	0.01	0.95	-1.62	1.75
Trade Freedom	0.36	0.84	-4.43	1.54
Emerging & Developing	0.76	0.43	0.00	1.00
European Union	0.12	0.32	0.00	1.00
Central & Eastern Europe	0.09	0.29	0.00	1.00
Middle East & North Africa	0.08	0.27	0.00	1.00
Sub-saharan Africa	0.20	0.40	0.00	1.00
Western Hemisphere	0.19	0.39	0.00	1.00

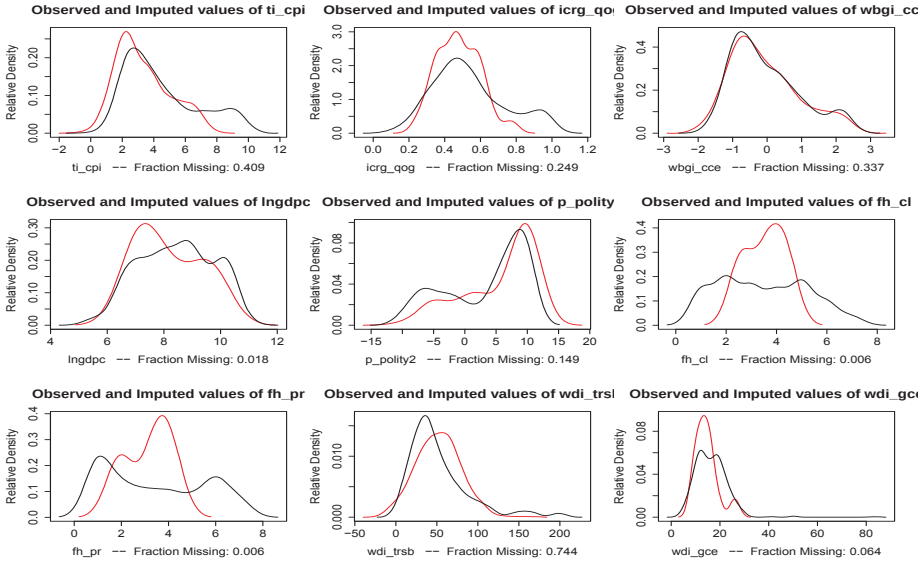
Notes: Categorical classifications come from IMF.org. European Union includes Austraia, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Portugal, Slovakia, Slovenia and Spain

Table 6: Peyton & Butcher Multiple Imputation

	μ	σ	min	max
Corruption (TI)	3.97	2.09	0.23	10.00
Quality of Government	0.54	0.18	0.09	1.01
Control of Corruption(WB)	-0.05	1.00	-2.22	2.62
ln(GDP per capita)	8.56	1.32	5.05	11.42
Democracy (Polity)	3.52	6.11	-10.00	10.00
Civil Liberties (FH)	3.35	1.76	1.00	7.00
Political Rights (FH)	3.43	2.13	1.00	7.00
Democracy (BG)	6.78	3.82	2.00	14.00
Government Size	16.01	6.22	1.67	83.16
Urbanization	53.27	22.84	5.15	104.43
Business Start (WDI)	47.38	60.18	0.00	694.00
Economic Freedom (No P)	-0.09	1.03	-4.73	2.86
Economic Freedom (BG)	-0.09	1.02	-4.23	2.28
Economic Freedom (HF)	-0.09	1.01	-4.20	2.79
Business Freedom	-0.05	1.00	-3.84	2.45
Freedom from Corruptin	-0.04	0.96	-1.72	2.56
Financial Freedom	-0.08	0.99	-2.45	1.93
Fiscal Freedom	-0.01	0.99	-4.53	2.07
Government Spending	-0.05	1.00	-3.01	1.95
Investment Freedom	-0.08	0.99	-2.98	2.00
Labor Freedom	0.01	0.93	-2.85	2.77
Monetary Freedom	-0.07	1.05	-5.65	1.66
Property Rights	-0.05	0.97	-2.16	1.96
Trade Freedom	-0.04	0.98	-4.43	1.60
Emerging & Developing	0.82	0.38	0.00	1.00
European Union	0.09	0.28	0.00	1.00
Central & Eastern Euope	0.08	0.27	0.00	1.00
Middle East & North Africa	0.08	0.27	0.00	1.00
Sub-saharan Africa	0.24	0.43	0.00	1.00
Western Hemisphere	0.18	0.38	0.00	1.00

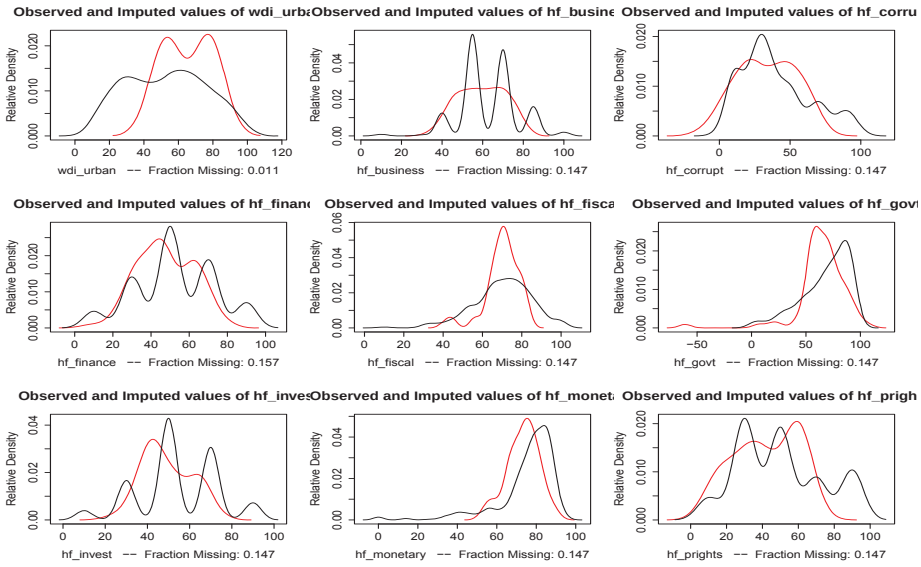
Notes: Categorical classifications come from IMF.org. European Union includes Austraia, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Portugal, Slovakia, Slovenia and Spain

FIGURE 1: BG Multiple Imputation Diagnostics Series 1



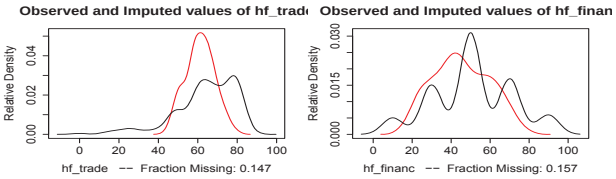
Notes: Notes: ti_cpi: transparency international's corruption perceptions index, *icrq_qog* = international country risk guide's quality of government index, *wbgi_cce* = world bank's control of corruption index, lngdpc = *log*(GDP per capita, PPP) from world bank, *p_polity*2 = revised combined polity score from the QoG database, fh_cl = freedom house civil liberties, fh_pr = freedom house political rights, *bg_dem* = Billger & Goel democracy score ($fh_cl + fh_pr$), *wdi_trsb* = time required to start a business (days),*wdi_gce* = government consumption expenditure as percentage of GDP (World Bank).

FIGURE 2: BG Multiple Imputation Diagnostics Series 2



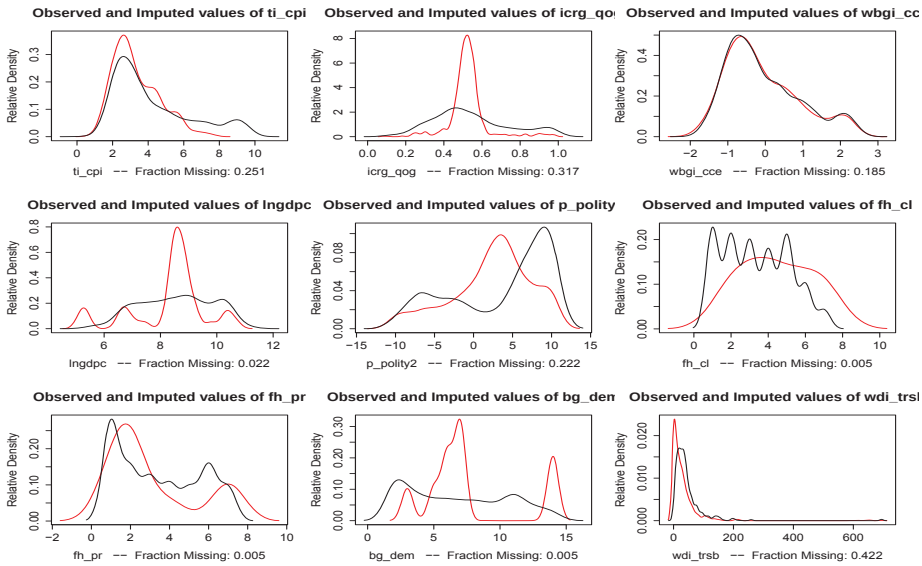
Notes: wdi_urban = % of population living in urban area (World Bank), *hf_business* = business freedom, *hf_corrupt* = corruption component of economic freedom index, *hf_financ* = financial freedom, *hf_fiscal* = fiscal freedom, *hf_govt* = government spending, *hf_invest* = investment freedom, *hf_monetary* = monetary freedom, *hf_prights* = property rights

FIGURE 3: BG Multiple Imputation Diagnostics Series 3



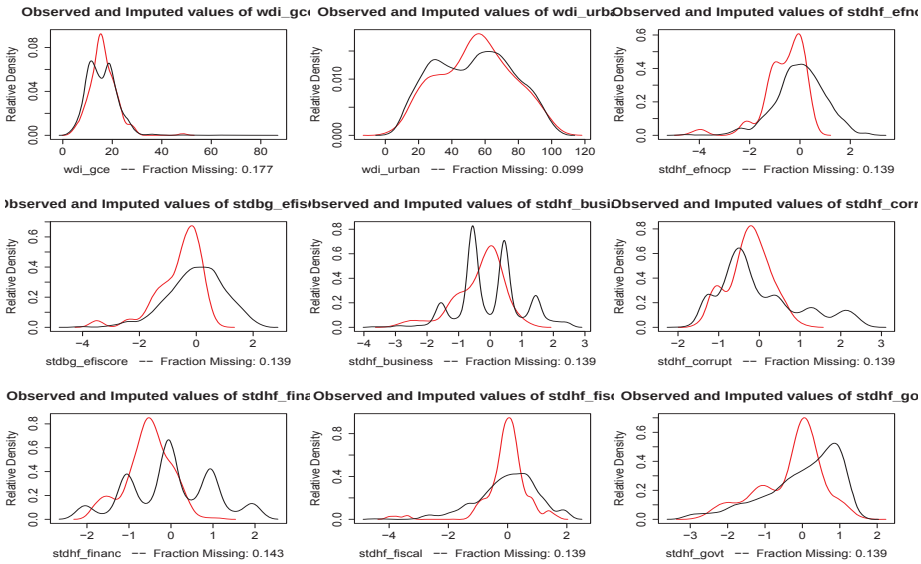
Notes: hf_trade = trade freedom, hf_financ=financial freedom

FIGURE 4: PB Multiple Imputation Diagnostics Series 1



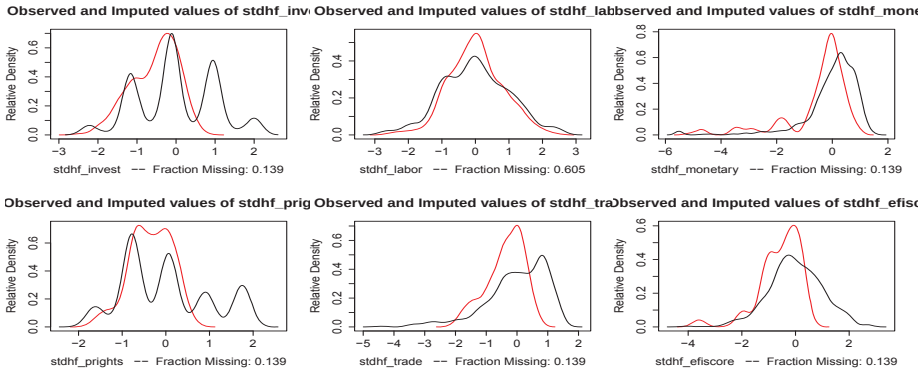
Notes: ti_cpi: transparency international's corruption perceptions index, $icrq_qog$ = international country risk guide's quality of government index, $wbgi_cce$ = world bank's control of corruption index, lngdpc = log(GDP per capita, PPP) from world bank, $p_polity2$ = revised combined polity score from the QoG database, fh_cl = freedom house civil liberties, fh_pr = freedom house political rights, bg_dem = Billger & Goel democracy score ($fh_cl + fh_pr$), wdi_trsb = time required to start a business (days)

FIGURE 5: PB Multiple Imputation Diagnostics Series 2



Notes: "std" prefix denotes standardized variables with $\mu \approx 0$ and $\sigma \approx 1$. wdi_gce = government consumption expenditure as percentage of GDP (World Bank), wdi_urban = % of population living in urban area (World Bank), $stdhf_efnocp$ = economic freedom index with corruption perceptions index removed, $stdbg_efiscore$ = Billger & Goel economic freedom index, $stdhf_business$ = business freedom, $stdhf_corrupt$ = corruption component of economic freedom index, $stdhf_financ$ = financial freedom, $stdhf_fiscal$ = fiscal freedom, $stdhf_govt$ = government spending

FIGURE 6: PB Multiple Imputation Diagnostics Series 3



Notes: $stdhf_invest$ = investment freedom, $stdhf_labor$ = labor freedom, $stdhf_monetary$ = monetary freedom, $stdhf_prights$ = property rights, $stdhf_trade$ = trade freedom, $stdhf_efiscore$ = aggregate economic freedom score