Using Interaction Terms as Instrumental Variables for Causal Identification: Does Corruption Harm Economic Development?*

Justin Esarey†

September 3, 2015

Abstract

Instrumental variables are an important tool for empirically identifying causal relationships among endogenous variables. Unfortunately, good instruments are often hard to find. In order to expand the universe of possible instruments, I propose and validate a new strategy for finding instrumental variables: if two exogenous, non-instrumental variables $v$ and $w$ have a conditional relationship with $x$ but an unconditional relationship with $y$, then the interaction term $vw$ serves as an effective instrument to identify the causal effect of $x$ on $y$. I use this strategy to address a debate in the corruption literature, where the causal interpretation of evidence linking corruption to economic development hinges critically on the validity of chosen instruments. After introducing three new interaction-based instruments for corruption, I use these instruments to determine that corruption lowers per capita GDP in contemporary democracies—but not as strongly as previous results would indicate.

Introduction

Policy makers and scholars often want to measure how much an intervention changes an outcome of interest using observational (non-experimental) data. But correlation is not causation: we cannot simply examine the empirical relationship between a putative independent and dependent variable and hope to conclude that the former caused the latter. Instrumental variable models help us to draw causal inferences by counteracting endogeneity, spurious

*I thank Tobias Heinrich, Cliff Morgan, Adrian Raftery, Tom Pepinsky, Leslie Schwindt-Bayer, and the participants in the 2015 University of Washington CSSS Seminar Series and the 2014 Annual Meeting of the Society for Political Methodology for providing helpful comments and feedback on previous drafts of this paper. I thank Katy Dessi for advice and research assistance provided during this project.

†Assistant Professor of Political Science, Rice University. E-mail: justin@justinesarey.com.
correlation, measurement error, and other threats to inference often present in observational
data (Angrist and Krueger, 2001). The primary difficulty with their use is that good instru-
ments are hard to find: an instrument \( v \) must be correlated with an endogenous independent
variable \( x \), but uncorrelated with an endogenous dependent variable \( y \) except through its
effect on \( x \) (Angrist and Pischke, 2009, p. 153-155). Opening up new and creative ways
to identify variables that meet this criterion is critically important for quantitative research
across all substantive areas.

As one applied example, there are many theoretical reasons to believe that corruption
(defined as “the misuse of public office for private gain” (Treisman, 2000, 399)) places a drag
on economic development. Corruption imposes additional financial and regulatory burdens
that especially harm innovation by new entrants (Murphy et al., 1993, p. 413) and might
thereby have an impact on economic development (Acemoglu et al., 2001), possibly by dis-
torting capital allocation (Easterly, 1993). It can result in inequitable application of the law
and inefficient allocation of tax burdens and government resources, possibly even underm-
ining the legitimacy and sovereignty of state institutions (Rose-Ackerman, 1999, Chapter 2).
It can also discourage investment, resulting in macroeconomic contraction (Mauro, 1995).
Unfortunately, there are also many theoretical reasons to believe that stronger economic
development lowers corruption. For example, economic development may encourage the
development of stronger and more democratic institutions that act to detect and punish
corruption (Boix and Stokes, 2003, Glaeser et al., 2004 Treisman 2000, p. 404). The re-
sulting literature has struggled to distinguish the causal effect of corruption on the economy
from the causal effect of the economy on corruption, ultimately culminating in a debate over
which instrumental variables are appropriate for empirically identifying how much corrup-
tion influences economic development. Treisman (2007, p. 225) ultimately concludes that
“to establish a direction of causation, one needs good instruments, which are unfortunately
in short supply.”

This paper is about promoting and validating a strategy to expand the universe of poten-
tial instrumental variables, and applying this strategy to measuring the effect of corruption on per capita gross domestic product (GDP). I demonstrate that if two exogenous, non-instrumental variables $v$ and $w$ have a conditional relationship with $x$ but an unconditional relationship with $y$, then the interaction term $vw$ serves as an effective instrument to identify $dy/dx$. That is, we need not look for a variable whose only effect on $y$ comes through $x$; we merely need to find a variable $v$ whose relationship with $x$ depends on $w$, but whose relationship with $y$ does not depend on $w$. Because $v$ and $w$ are not instruments for $x$, it is somewhat contrary to intuition that they can be used in combination to form a proper instrument. Nevertheless, I validate the approach using the causal path analysis methods of Pearl (2000) and with a Monte Carlo analysis.

I then propose and validate four new instruments for corruption and use them to measure the degree to which corruption lowers per capita GDP. The instruments are based on prior work establishing that the relationship between women’s representation in government and corruption is stronger in the presence of institutions that strengthen electoral accountability for corruption (Esarey and Chirillo, 2013; Esarey and Schwindt-Bayer, 2015). One of the instruments (a binary indicator of democratic-leaning institutions based on the Polity2 scale) indicates the existence of meaningful electoral pressures; it can be applied to the cross-sectional dataset collected by Acemoglu et al. (2001). The other three (freedom of the press, presidentialism, and personalistic vs. party-centric electoral rules) enhance the power of electoral accountability inside of states with generally democratic institutions; I use these instruments in a panel data set of democratic-leaning countries between 1990 and 2010. While women’s frequently-demonstrated\textsuperscript{1} greater aversion to risk (in this case, the risk of being held accountable for corruption) can explain why the women-corruption relationship is conditional on accountability, I have no reason to suspect that any empirical relationship between women’s representation and growth is conditional on the instruments. My analysis reveals a strong and substantively meaningful effect of corruption on economic performance;

\textsuperscript{1}See Croson and Gneezy (2009); Bernasek and Shwiff (2001); Sunden and Surette (1998); Watson and McNaughton (2007); Eckel and Grossman (2008); Byrnes et al. (1999).
however, this effect an order of magnitude smaller than the relationship previously measured by Acemoglu et al. (2001). This finding bolsters the claim that corruption actively harms economic development and that lowered corruption would improve prosperity. It also illustrates how interaction-based instruments can be a very useful tool for substantive researchers.

A $d$-separation argument for the use of interaction terms as instrumental variables

Consider the causal path diagram in Figure 1. As laid out by Pearl (2000, pp. 140-149), this diagram corresponds to the following set of structural equations:

\[
\begin{align*}
    y &= \alpha_0 + \alpha_1 x + \varepsilon_y \\
    x &= \beta_0 + \beta_1 vw + \varepsilon_x \\
    v &= \delta_0 + \varepsilon_v \\
    w &= \theta_0 + \varepsilon_w \\
    vw &= v \times w
\end{align*}
\]

where $\varepsilon_i$ indicates a stochastic error term specific to variable $i$ and  \{\alpha_j, \beta_j, \delta_j, \theta_j\} are scalar coefficients indexed by $j$. Each node of the graph (depicted as a circle) is a variable of interest (with the variable’s name in the center of the circle). Solid lines with arrows indicate that one variable deterministically influences the value of another in the structural equations (Pearl, 2000, 140-141). Bi-directional dashed lines indicate that stochastic error terms are correlated; this correlation can be the result an omitted variable that causes both variables connected by the dashed arc, because of simultaneity between the two variables, or for other reasons. In Figure 1, a variance-covariance matrix $\Sigma_\varepsilon$ of the vector of error terms $\varepsilon = [\varepsilon_y, \varepsilon_x, \varepsilon_v, \varepsilon_w]$.
Figure 1: Path diagram of generic network of causal relationships using an interaction instrument

\[\Sigma_\varepsilon = \begin{bmatrix}
\sigma_y^2 & \sigma_{yx} & \sigma_{yv} & \sigma_{yw} \\
\sigma_{xy} & \sigma_x^2 & \sigma_{xv} & \sigma_{xw} \\
\sigma_{vy} & \sigma_{vx} & \sigma_v^2 & \sigma_{vw} \\
\sigma_{wy} & \sigma_{wx} & \sigma_{wv} & \sigma_w^2 
\end{bmatrix}\]

The diagram of Figure 1 implies that variables \(v\) and \(w\) are not instruments: they are correlated with \(\varepsilon_y\) through \(\varepsilon_v\) and \(\varepsilon_w\). The interaction term \(vw\) indicates that the effect of \(v\) on \(x\) is conditional on the value of \(w\), but that the effect of \(v\) on \(y\) does not depend on \(w\). The same holds for the effect of \(w\) on \(x\) and \(y\).

Pearl (2000) demonstrates that \(\alpha_1\) can be identified in this diagram by using the instrument \(vw\). Begin by formally stating Pearl’s concept of \(d\)-separation:

**Definition 1.2.3 (Pearl, 2000, pp. 16-17):** A path \(p\) is said to be \(d\)-separated
(or blocked) by a set of nodes $Z$ if and only if (i) $p$ contains a chain $i \rightarrow m \rightarrow j$ or $i \leftarrow m \rightarrow j$ such that the middle node $m$ is in $Z$, or (ii) $p$ contains an inverted fork (or collider) $i \rightarrow m \leftarrow j$ such that the middle node $m$ is not in $Z$ and such that no descendant of $m$ is in $z$.

A “path” in this definition is a route that connects nodes to each other through vertices, such as the pathway that leads through $vw$ to $v$ to $x$ through the lines that connect them; note that a path need not follow the arrows in the plot.\(^2\) Pearl states a theorem that applies the concept of $d$-separation to a linear causal model like the one in Figure 1:

**Theorem 5.2.3 (Pearl, 2000, 142):** For any linear model structured according to a diagram $D$, which may include cycles and bidirected arcs, the partial correlation $\rho_{ab|Z}$ vanishes if the nodes corresponding to the set of variables $Z$ $d$-separate node $a$ from node $b$ in $D$. (Each bidirected arc connected by a dashed line is interpreted as a latent common parent $u$ with $i \leftarrow u \rightarrow j$.)\(^3\)

$\rho_{ab|Z}$ is the partial correlation between variables $a$ and $b$ holding a $1 \times k$ vector of variables $Z$ constant, equivalent to coefficient $\gamma_1$ in a regression $a = \gamma_0 + \gamma_1 b + Z\Gamma + \psi$ (where $\Gamma$ is the vector of coefficients associated with $Z$). It is important to note that, in Figure 1, variables $v$ and $w$ $d$-separate $vw$ from $x$ according to this definition. This theorem implies that, if $vw$ can be $d$-separated from $x$ in a graph with the edge between $vw$ and $x$ deleted, then all

---

\(^2\)Because $vw$ is functionally determined by $v$ and $w$, the closely-related concept of $D$-separation (Geiger et al., 1990, p. 516) can apply. As stated by Geiger et al. (with some minor adjustment to notation):

A node $i$ is (functionally) determined [a set of nodes] $Z$ iff $i \in Z$ or $i$ is a deterministic node and all its parents [nodes with a directed path whose arrow points toward $i$] are functionally determined by $Z$. If $i$ is a deterministic node with no parents, then it is functionally determined by $Z$. A set of nodes is determined by $Z$ if each of its members are determined by $Z$.

$D$-separation implies all independence relationships that are implied by $d$-separation, as well as new ones implied by an additional criterion: under $D$-separation, a path $p$ is blocked by $Z$ if a tail-to-tail node on $p$ is functionally determined by $Z$ (per their definition on $D$-separation on p. 516). I use the simpler concept of $d$-separation because it is functionally identical to $D$-separation in a graph like Figure 1, where the additional condition of $D$-separation does not apply.

\(^3\)I have changed some of Pearl’s variable names in this and other quoted theorems in order to avoid conflicting with my own notation.
sources of correlation between $vw$ and $x$ that are not attributable to the direct relationship $vw \rightarrow x$ can be blocked. Indeed, Pearl formally states this in a theorem:

**Theorem 5.3.1 (Pearl, 2000, pp. 150-151):** Let $G$ be any path diagram in which $\alpha$ is the path coefficient associated with link $b \rightarrow a$, and let $G_\alpha$ denote the diagram that results when $b \rightarrow a$ is deleted from $G$. The coefficient $\alpha$ is identifiable if there exists a set of variables $Z$ such that (i) $Z$ contains no descendant of $a$, and (ii) $Z$ $d$-separates $b$ from $a$ in $G_\alpha$. If $Z$ satisfies these two conditions, then $\alpha$ is equal to the regression coefficient $r_{ab|Z}$. Conversely, if $Z$ does not satisfy these conditions, then $r_{ab|Z}$ is not a consistent estimand of $\alpha$ (except in rare instances of measure zero).

$r_{ab|Z}$ is the coefficient on $b$ in a regression of $a$ that includes control variables $Z$, equal to $\gamma_1$ in the regression $a = \gamma_0 + \gamma_1 b + Z \Gamma + \psi$. Theorem 5.3.1 allows us to identify $\beta_1$, the coefficient for the effect of $vw$ on $x$, through the regression $x = \gamma_0 + \gamma_1 v + \gamma_2 w + \gamma_3 vw + \psi$. In this regression, $Z = \{v, w\}$ $d$-separates $vw$ from $x$ in the graph where the edge between $vw$ and $x$ is deleted; note that all back-door pathways from $vw$ to $x$ are intercepted by $Z$ in Figure 1. Consequently, $\gamma_3$ is a valid estimator for $\beta_1$.

Moreover, the total effect of $vw$ on $y$, $dy/dvw = (dy/dx)(dx/dvw)$, can be identified via the regression $y = \pi_0 + \pi_1 v + \pi_2 w + \pi_3 vw + \zeta$. The key insight is given in Pearl’s theorem 5.3.2, the “back-door criterion” for effect identification, which I re-state here:

**Theorem 5.3.2 (Pearl, 2000, p. 152):** For any two variables $a$ and $b$ in a causal diagram $G$, the total effect of $b$ on $a$ is identifiable if there exists a set of measurements $Z$ such that: (i) no member of $Z$ is a descendant of $b$, and (ii) $Z$ $d$-separates $b$ from $a$ in the subgraph $G'$ formed by deleting from $G$ all arrows emanating from $b$.

In the proposed regression, $Z = \{v, w\}$ $d$-separates $vw$ from $y$ in a graph of Figure 1 where all arrows emanating from $vw$ are deleted, and neither $v$ nor $w$ is a descendant of $vw$. Ergo,
we can conclude that $\pi_3$ is a valid estimator of $\beta_1\alpha_1$.

Consequently, the effect of $x$ on $y$ is identifiable as the ratio of the coefficients for the effect of $vw$ on $y$ and $vw$ on $x$:

$$\frac{dy/dvw}{dx/dvw} = \frac{\pi_3}{\gamma_3} = \frac{\beta_1\alpha_1}{\beta_1} = \alpha_1$$

This is definitionally an instrumental variables estimator, with $vw$ serving as the instrument (Heckman, 2000). The identifiability of $\alpha_1$ through use of $vw$ as an instrumental variable suggests the use of a classical two-stage least-squares (2SLS) strategy (Cameron and Trivedi, 2005, pp. 100-101) for more directly estimating $\alpha_1$:

$$\hat{\alpha} = (Z'X)^{-1} Z'y$$

where $Z = \begin{bmatrix} 1 & v & w & vw \end{bmatrix}$, $X = \begin{bmatrix} 1 & x & v & w & vw \end{bmatrix}$, and $\hat{\alpha}_1$ is the element of $\alpha$ corresponding to $x$.

In the event that there are additional pathways between $x$ and $vw$ that pass through intermediary variables, Theorem 5.3.2 applies. For example, suppose that $vw \rightarrow s \rightarrow x$, so that the structural equations are:

\[
\begin{align*}
y &= \alpha_0 + \alpha_1 x + \varepsilon_y \\
x &= \beta_0 + \beta_1 vw + \beta_2 s + \varepsilon_x \\
s &= \phi_0 + \phi_1 vw + \varepsilon_s \\
v &= \delta_0 + \varepsilon_v \\
w &= \theta_0 + \varepsilon_w \\
vw &= v \times w
\end{align*}
\]

Ergo, the total effect of $vw$ on $x$ is $\beta_1 + \beta_2\phi_1$. According to Theorem 5.3.2, we can measure this effect with a regression $x = \gamma_0 + \gamma_1 v + \gamma_2 w + \gamma_3 vw + \psi$ using coefficient $\gamma_3$. In the regression
\[ y = \pi_0 + \pi_1 v + \pi_2 w + \pi_3 vw + \zeta, \] \( \pi_3 \) measures the total effect of \( vw \) on \( y \): \((\beta_1 + \beta_2 \phi_1) \alpha_1\).

Consequently, the same ratio as before yields \( dy/dx \) even when \( s \) is omitted from both regressions:

\[
\frac{dy/dv}{dx/dv} = \frac{\pi_3}{\gamma_3} \frac{(\beta_1 + \beta_2 \phi_1) \alpha_1}{(\beta_1 + \beta_2 \phi_1)} = \alpha_1
\]

On the other hand, it is very important that there be no unblocked paths between \( y \) and \( vw \) that do not pass through \( x \). For example, if there is a pathway \( vw \rightarrow s \rightarrow y \) through a variable \( s \) that is not included as a control variable in the appropriate regression, then that regression’s estimate of \( dy/dvw \) will include correlation attributable to that pathway and the resulting estimate of \( dy/dx \) will be biased. This is simply another statement of the “exclusion restriction” that is required for instrumental variables such as \( vw \) (Angrist and Pischke, 2009, p. 153-155): \( vw \) must be associated with \( y \) only through its effect on \( x \).

**Simulation evidence**

A Monte Carlo simulation demonstrates that an interaction-based instrument allows for consistent identification of \( dy/dx \) using the model of equation 1 despite endogeneity and omitted variable bias. 1000 data sets of size \( n \in \{25, 50, 75, 100, 500, 1000, 2000\} \) are drawn from the DGP in:

\[
y = \beta_0 + \beta_1 x + \beta_2 v + \beta_3 w + \beta_4 u + \varepsilon \quad (2)
x = \alpha_0 + \alpha_1 y + \alpha_2 v + \alpha_3 w + \alpha_4 u + \alpha_5 vw + \eta \quad (3)
\]

\( \beta_1 = 2, \alpha_1 = 1, \{\beta_2, \beta_3, \alpha_2, \alpha_3\} \) are drawn from the uniform distribution between \([0, 1]\), \( \{\beta_4, \alpha_4\} \) are drawn from the uniform distribution between \([0.5, 1]\), and \( \alpha_5 = 0.5 \). The error terms \( \varepsilon \) and \( \eta \) are drawn from the normal density \( \Phi(\mu = 0, \sigma = 2) \). \( v, w, \) and \( u \) are drawn from the multivariate normal distribution with mean 2; the correlation \( \rho \) between \( v \) and \( u \) is varied inside \( \rho \in \{0, 0.25, 0.5, 0.75, 0.9\} \) but \( \text{cor}(u,w) = \text{cor}(v,w) = 0 \) and
An instrumental variable model that omits $u$ but is otherwise correctly specified is estimated for each data set and the 2SLS estimate for $dy/dx$ (or $\hat{\beta}_1$) is saved.

**Consistent 2SLS estimation of $dy/dx$**

Figure 2 shows the median and 95\% confidence boundaries of $\left(\hat{\beta}_1 - \beta_1\right)$ at varying values of $n$ and $\rho$ for the 1000 data sets (labeled “interaction” in the graph); the figure presents only $\rho = 0.2$ and $\rho = 0.9$ to save space, but other simulations had qualitatively similar results (see the appendix for details).

As the figure shows, the 2SLS procedure can accurately recover the relationship $dy/dx$ despite endogeneity and omitted variable bias. However, 2SLS is a consistent but not unbiased procedure. In small data sets, 2SLS estimates are biased downward and highly variable. Only in samples of $n \geq 500$ are the results reasonably on-target.

Figure 2 also includes results from a “standard” instrumental variable model generated out of a DGP with a non-interacted instrument:

\begin{align*}
  y &= \beta_0 + \beta_1 x + \beta_2 v + \beta_3 w + \beta_4 u + \varepsilon \\
  x &= \alpha_0 + \alpha_1 y + \alpha_2 v + \alpha_3 w + \alpha_4 u + \alpha_5 t + \eta
\end{align*}

where $vw$ from equation 3 is replaced with $t$, which I drew from the uniform distribution between $-1$ and $1$; all other parameters are the same as in the simulation with interaction instruments. As the figure shows, there is no apparent disadvantage in terms of bias of using an interaction instrument like $vw$ instead of a standard instrument like $t$.

**Efficiency properties**

The simulation indicates little evidence for efficiency disadvantages in using interaction instruments when compared to equivalent standard instruments. Figure 3 shows the proportion
Figure 2: Simulation evidence for consistent 2SLS estimation of $dy/dx = \beta_1$

The figure shows the median and 95% confidence boundaries of $(\hat{\beta}_1 - \beta_1)$ at varying values of $n$ and $\rho$ for 1000 models on 1000 simulated data sets. The circles show results for simulations using an interaction-based instrument from the data generating process in equations 2-3; the triangles show results for simulations using a standard instrument from the data generating process in equations 4-5.
The figure shows the proportion of the time that 95% confidence intervals for the $\hat{\beta}$ estimate cover the true value of $\beta$ at varying values of $n$ and $\rho$ for 1000 models on 1000 simulated data sets. The circles show results for simulations using an interaction-based instrument from the data generating process in equations 2-3; the triangles show results for simulations using a standard instrument from the data generating process in equations 4-5.

The figure shows the proportion of the time that 95% confidence intervals for the $dy/dx$ estimate cover the true value; this should be $\approx 0.95$ if confidence intervals are appropriately sized. My simulations indicate that confidence intervals are substantially too narrow for sample sizes $N < 500$, implying that 2SLS results will be overly confident for comparatively small sample sizes. On the other hand, the use of an interaction instrument is not disadvantageous compared to an equivalent non-interaction instrument; the simulations with the standard instrument $t$ are substantively indistinguishable from those for the interaction instrument $vw$.

Interestingly, the interaction instruments are slightly more efficient in the simulations
The figure shows the median standard error for estimated $\hat{\beta}$ at varying values of $n$ and $\rho$ for 1000 models on 1000 simulated data sets. The circles show results for simulations using an interaction-based instrument from the data generating process in equations 2-3; the triangles show results for simulations using a standard instrument from the data generating process in equations 4-5.

compared to standard instruments; this is shown in Figure 4. The figure shows the median estimated standard error of $\hat{\beta}_1$ for each of my simulations. The median SE of $\hat{\beta}_1$ when using the interaction instrument $vw$ is consistently lower than the equivalent median SE when using the standard instrument $t$. The implication of this analysis is that interaction instruments may provide estimates of $dy/dx$ that are at least as efficient, and possibly more efficient, than the equivalent analysis using typical (non-interaction) instruments.
Weak instruments

The quality of results in a 2SLS procedure is typically degraded when the instrument for the endogenous variable is weak. This is also true when using interaction instruments like $vw$. To show this, I repeated my simulations setting $\alpha_5 = 0.1$ instead of $\alpha_5 = 0.5$. The results of the simulations are shown in Figure 5, which shows the median and 95% confidence boundaries of $(\hat{\beta}_1 - \beta_1)$ at varying values of $n$ and $\rho$ for strong and weak instruments. As the figure indicates, convergence of $\hat{\beta}_1$ to its true value is much slower for weak instruments compared to strong instruments. Indeed, even when $N = 2000$, models with the weaker instruments exhibit considerable sample-to-sample variability in estimates of $dy/dx$. Consequently, as is standard for 2SLS, accurate results seem to depend on (i) large samples and (ii) instruments that are sufficiently strong to predict a substantial portion of the variation in the endogenous variable.

Application: does corruption lower per capita GDP?

Validating the use of interaction-based instrumental variables is important precisely because it opens up new possibilities for expanding and validating our knowledge of causal relationships among variables that are likely to be endogenous. In some cases, it may make it possible to address substantive controversies where the validity of an important causal finding hinges on the choice of instrument. Such findings are bolstered if they can be replicated using different instruments; here, the scarcity of reliable instrumental variables is acute. To illustrate the potential impact of interaction-based instruments on substantive knowledge, I deploy them to tackle an important controversy in the scholarship about corruption.

What impact does corruption have on economic development? The question holds self-evident importance for reform advocates and policy makers (who wish to improve outcomes for themselves and their fellow citizens) as well as scholars (who wish to understand the role that corruption plays in economic development). Unfortunately, even if perfect measures of
Figure 5: Simulation evidence for 2SLS estimation of \(dy/dx = \beta_1\), weak vs. strong instruments

The figure shows the median and 95% confidence boundaries of \((\hat{\beta}_1 - \beta_1)\) at varying values of \(n\) and \(\rho\) for 1000 models on 1000 simulated data sets. The simulations use an interaction-based instrument from the data generating process in equations 2-3. The circles indicate simulations with a “weak” instrument where \(\alpha_5 = 0.1\), while the triangles indicate simulations with a “strong” instrument where \(\alpha_5 = 0.5\).
corruption and prosperity can be devised, simply examining the relationship between the two is unrevealing. Treisman (2007, p. 225) summarizes the problem well:

If the correlation [between economic development and corruption] is strong as robust, the question remains what—if anything—it means. Does economic development reduce perceived corruption? Does corruption slow economic development? Are both caused by some third factor? A similar debate has pitted those who believe good institutions explain why some countries developed faster than others (Acemoglu et al., 2001; Rodrik et al., 2004) against those who think economic development explains why some countries acquired good institutions (e.g., Boix and Stokes, 2003) and others who believe the accumulation of human capital caused both economic development and superior institutions (Glaeser et al., 2004).

In short, economic development and corruption are likely to be endogenously related, such that the observed relationship between the two cannot be easily interpreted as causal in either direction.

Several instrumental variables to identify the causal effect of corruption on economic development (which I measure using log per capita GDP) have been proposed in the past. Unfortunately, all of these instruments have been subject to criticism. More and different instruments would help to resolve how much corruption lowers per capita GDP by validating pre-existing results from a new perspective. The possibility of using instrumental variables whose influence on corruption is conditional (but whose influence on GDP is not) opens up a world of new possibilities; after briefly summarizing prior instrument choices, I offer three new instruments based on interaction terms and estimate the causal impact of corruption on log GDP per capita using these instruments.

Prior instruments for corruption, and criticisms

Acemoglu et al. (2001) use colonial settler mortality as an instrument for the quality of
institutions (specifically, lower risk of expropriation by authorities) to measure their effect on log GDP per capita. The authors argue that European settlers migrated toward colonies with lower mortality rates and implemented European-style institutions there; colonies with higher mortality rates were simply exploited for the benefit of the mother country and left without strong institutions (pp. 1373-1375). An analysis with this instrument indicates “large effects of institutions on income per capita” (p. 1395); indeed, “it implies that improving Nigeria’s institutions to the level of Chile could... lead to as much as a 7-fold increase in Nigeria’s income” (p. 1371). However, Glaeser et al. (2004) argue that the Europeans also brought their human capital with them; indeed, they find that colonial settler mortality is more strongly correlated with human capital (specifically, years of schooling in the year 2000) than with expropriation risk (p. 290). This fact (among others) leads them to argue that “institutions have only a second-order effect on economic performance. The first order effect comes from human and social capital, which shape both institutional and productive capacities of a society” (p. 298). The reader is therefore left wondering whether policies designed to lower corruption would truly increase GDP per capita as Acemoglu et al. (2001) indicate; Glaeser et al. (2004) believe that the same evidence supports prioritizing the accumulation of human capital over reforming political institutions.

Other instruments have also been proposed. Mauro (1995) uses ethnolinguistic fractionalization (ELF) as an instrument for corruption’s effect on the growth of GDP per capita on the theory that it is “highly correlated with corruption... yet it can be assumed to be exogenous both to economic variables and to institutional efficiency” (p. 683). Alternatively, Hall and Jones (1999) uses distance from the equator as an instrument for “social infrastructure,” defined to mean “the institutions and government policies that determine the economic environment within which individuals accumulate skills, and firms accumulate capital and produce output” (p. 84); they are interested in social infrastructure’s effect on worker productivity. Hall et al. justify their choice of instrument in a manner that anticipates the later argument of Acemoglu et al. (2001), arguing that regions more densely settled by Western
Europeans (viz., far from the equator) were most powerfully influenced by Western European ideas about property rights and limited government (pp. 100-101). However, both of these choices of instrument have been criticized in the literature. Acemoglu et al. (2001, p. 1373) argues that both instruments are flawed because they might cause growth directly (i.e., not through an indirect effect on institutions): “William Easterly and Ross Levine (1997) argue that ethnolinguistic fragmentation can affect performance by creating political instability while... David E. Bloom and Jeffrey D. Sachs (1998) and John Gallup et al. (1998) argue for a direct effect of climate on performance.” Indeed, Treisman (2000) uses distance from the equator as an instrument for economic development’s effect on corruption—the opposite causal relationship—because “it is indeed highly correlated with development” and he “could not think of pathways by which distance from the Equator could affect corruption other than via economic development” (p. 430).

**Interaction-based instruments**

A recent line of research examining the relationship between women’s representation in government and corruption opens up the possibility of using interaction terms as instruments for corruption. Initial time-series cross-sectional studies in this area demonstrated that greater representation of women in government is consistently associated with lower perceived corruption in that government (Dollar et al., 2001; Swamy et al., 2001). More recent research demonstrates that this relationship is stronger in certain contexts, in particular where institutional arrangements allow the public to hold corrupt officials accountable for their actions (Esarey and Chirillo, 2013; Esarey and Schwindt-Bayer, 2015). Because the relationship between female representation and corruption is moderated by the strength of accountability institutions, this suggests using interactions between the proportion of women in government and various measures of accountability as instruments for corruption in studying corruption’s effect on economic performance.

In this paper, I focus on four measures of accountability used in this literature. The
first, simplest measure is whether a state is holistically considered a democracy or not. The other three measures are press freedom, presidentialism, and personalistic electoral rules. These factors have a negative relationship with perceived corruption (Treisman, 2000; Persson et al., 2003; Kunicova and Rose-Ackerman, 2005), with the theoretical explanation being that these institutions allow voters to monitor the behavior of officials and punish them for corruption. Esarey and Schwindt-Bayer (2015) extend this reasoning to argue that these institutions activate women’s resistance to corruption precisely because they expose officials to a greater risk of being caught and punished for corruption (e.g., by voters); women consistently demonstrate greater risk aversion than men inside and outside of the laboratory (Croson and Gneezy, 2009; Bernasek and Shwiff, 2001; Sunden and Surette, 1998; Watson and McNaughton, 2007; Eckel and Grossman, 2008; Byrnes et al., 1999) and are therefore more strongly incentivized to avoid corruption when it becomes riskier. An empirical analysis indicates that greater representation of women in government is associated with less corruption in democratic-leaning countries (Esarey and Chirillo, 2013); furthermore, inside the set of democratic-leaning countries, women’s representation in parliament is more strongly related to government when the press is freer, in a parliamentary (instead of presidential) democracy, and in voting system with personalistic (instead of party-centered) electoral rules (Esarey and Schwindt-Bayer, 2015). This also explains why the latter three accountability measures (press freedom, presidentialism, and personalistic electoral rules) apply only inside the set of democratic-leaning states: these factors enhance the strength of the electoral accountability intrinsic to democracies and have been empirically validated in this context.

Consequently, there is a theoretical and empirical rationale to propose four multiplicative interaction instruments for corruption in government:

1. in the set of all states:

   (a) % women’s representation in the lower house of parliament * democracy

---

4 As compared to party-centered electoral rules; see Carey and Shugart 1995.
2. in the set of democratic-leaning states:

   (a) % women’s representation in the lower house of parliament * press freedom
   (b) % women’s representation in the lower house of parliament * presidentialism
   (c) % women’s representation in the lower house of parliament * personalism

All of these instruments are predicated on the idea that certain institutions make corruption a riskier proposition, thereby creating a gender gap in behavior that only exists in the presence of these institutions. The theoretical logic is explained more fully in Esarey and Chirillo (2013, pp. 367-369) and Esarey and Schwindt-Bayer (2015, pp. 9-14), but I summarize the argument here. Democratic institutions in general make corruption a riskier proposition by, among other mechanisms, exposing officials to punishment by voters and allowing criticism and debate (of official corruption) as part of the election process. Inside of democracies, certain institutions enhance the mechanism of electoral accountability. Press freedom works by allowing journalists to expose corruption to the voting public. Presidential systems (unlike parliamentary systems) typically do not allow no-confidence votes that can hold governments immediately responsible for corruption scandals, and also diffuse responsibility for these scandals across multiple branches of government. Personalistic electoral rules (such as the degree of political party control over vote lists and the type of voting system in the country) give voters greater ability to individually target and punish corrupt officials; by contrast, party-centered rules allow responsibility for corruption to be diffused through the entire party. In all cases, the presence of the institution strengthens the relationship between women’s representation and government because women are more sensitive to the risks of corrupt activity that are created by the institution. Although any of these variables might have some individual effect on economic development, I have little reason to believe that any of these effects is conditionally dependent on the proportion of women in government in its effect on development. Therefore, each interaction term is a reasonable candidate instrument for corruption.
Data and Variables

I consider two data sets in this paper. The first data set is a cross-section of 64 countries collected by Acemoglu et al. (2001); I examine this data in order to provide the closest possible comparison between their results and my own. However, the small number of observations available in this data set imposes some limitations on the analysis. The second data set is a panel of 95 democratic-leaning countries collected by Schwindt-Bayer and Tavits (2015); this data is much more extensive, allows more statistically powerful hypothesis tests, and permits blocking of unit-level unobserved confounding with fixed effects.

Cross-Sectional Data\(^5\)

The number of complete observations in the cross-sectional data set varies between 50 and 64 depending on the variables entered into the model. The main independent variable is the average value of “protection from expropriation” between 1985-1995 as collected by Political Risk Services; this variable values between 0 and 10, with greater variation indicating stronger protection of property rights against government action. The main dependent variable is per capita GDP in 1995. As these variables are expected to be endogenous, the authors construct a measure of the logarithm of settler mortality to use as an instrument for protection from expropriation and cite two sources of historical information as the basis for this measure (Curtin et al., 1995; McEvedy and Jones, 1975). Dummy variables for observations from the continents of Asia, Africa, and all other non-American continents are also included. All of these variables come from the replication data set made available by Acemoglu et al. (2001).

To these variables, I add an alternative instrument: the percentage of women in the lower house of parliament in 1990 * democratic-leaning. A country is considered “democratic-leaning” when it has a Polity2 score > 0 based on the coding of Marshall et al., 2010.\(^6\) The percentage of women in the lower house of parliament in 1990 is measured by the Inter-

\(^5\)The cross-sectional data set is described in Acemoglu et al. (2001, pp. 1377-1378); the information in this section comes from and closely follows that source.

\(^6\)This variable comes out of the Quality of Government data set (Teorell et al., 2015b).
parliamentary Union (2012). I focus on this instrument because I can apply it to the entire data set; my other instruments apply only to democratic-leaning cases and therefore leave substantially fewer observations available.

I also add the International Country Risk Guide (ICRG) corruption measure to the data set. The ICRG is a 0-6 measure of “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” (PRS Group, 2012, pp. 4-5) that is determined by experts “on the basis of subjective analysis of the available information” (p. 2). The ICRG is fundamentally a measure of perceived corruption; the perception of corruption is important to economic development both in its role as a measure of corrupt behavior and also because perception directly informs the behavior of economic agents, like investors (Treisman, 2007). For the cross-sectional data set, I recode the ICRG so that higher values indicate greater corruption, then average each country’s scores between 1985 and 1995. I add the ICRG to this data set because (a) it is a widely used measure of (perceived) corruption, the key concept of interest, available over a long time period for many countries, and (b) the interaction-based instruments were studied and justified in the context of their specific relationship with this variable in Esarey and Chirillo (2013) and Esarey and Schwindt-Bayer (2015).

Panel Data

The panel data set covers 95 democratic-leaning countries (with a Polity2 score > 0) between 1990 and 2010 for a total of 1653 observations (with an average of 17.4 observations per country). This data set and all variables described below were assembled by Schwindt-Bayer.

---

7This variable comes from the World Bank data website (World Bank, 2015).
8I report the results of an analysis of the democratic-leaning countries in this data set with the alternative instruments in Appendix Tables 6 and 7. The results are, generally speaking, broadly comparable to the relationships I find in the panel data analysis (albeit with substantially greater uncertainty).
9This section closely follows the variable descriptions originally used in Esarey and Schwindt-Bayer (2015).
10To be included in the data set, a country had to (a) have a population greater than 500,000 in 2006, (b) have a polity2 score greater than zero, (c) have a score of 5 or lower on Freedom House’s average Civil Liberties and Political Rights scales, and have met the three criteria above for four years or more.
and Tavits (2015). I focus on democratic-leaning countries for two reasons. First, much of the previous work in this area (including Acemoglu et al., 2001) focuses on broad institutional differences between democracies and autocracies; by concentrating on subtler institutional variations within democracies, we get a new take on this question and avoid some of the problems with instruments encountered in previous work (e.g., how politically unfree states with strong property rights compare to politically free states with weaker property rights). Second, and more practically, Esarey and Chirillo (2013) find that democracies and autocracies have different relationships between women’s representation and corruption; focusing on democracies allows comparison of cases that are more alike on this critical dimension (and follows the analysis of Esarey and Schwindt-Bayer (2015)).

The dependent variable is log GDP per capita (World Bank, 2013). The key independent variable is the International Country Risk Guide (ICRG) corruption measure, again recoded so that higher values indicate greater corruption. The instrumental variables are interactions between the percentage of women in parliament (measured by the Inter-parliamentary Union) and each of three electoral accountability measures: freedom of the press, presidentialism, and personalism. Freedom of the press is measured by Freedom House on a scale of -80 to 0, with higher values corresponding to more freedom.\footnote{See \url{www.freedomhouse.org} for details. Freedom House’s measure runs from 0 to 100 with higher values indicating less freedom; I recode their measure as the negative of the original measure and note that there are no observations in the data set with press freedom greater than 80. Because Freedom House used a three-point scale before 1993, a country’s values for 1990-1992 are coded as being the same as those for 1993; this assumes a degree of short-term stability in the measure over this short period.} Presidentialism is coded as a binary variable (= 1 if presidential, = 0 otherwise).\footnote{Semi-presidential systems are classified as one or the other depending on which regime they are closer to; see footnote 18 in Esarey and Schwindt-Bayer (2015).} The degree of personalism is an ordinal variable from 1 to 13, with higher levels indicating that electoral rules promote personal (rather than party-centered) vote-seeking; the variable is measured by Johnson and Wallack (1997).

Two variables are included as controls. First, I include the polity2 score of democracy, an expert-coded measure which normally varies between -10 to 10 but in this restricted data set
varies between 1 and 10; higher values generally indicate freer elections and more restrained executive powers (Marshall et al., 2010). As indicated in the literature review, democracies tend to be consistently associated with lower corruption (though which causes which is a matter of debate); this might interfere with inference if the interaction terms are themselves associated with the strength of democracy, and I therefore block this potential problem by including the control. Second, I include trade imbalance (imports minus exports as a percentage of GDP); greater openness to trade is associated with lower corruption (Treisman, 2007, p. 239) and might be associated with interactions between accountability and women in government by exposing government officials to greater scrutiny from international actors (and therefore a higher risk of negative consequences from corruption).

Results

There are two sets of results to report: those for the cross-sectional data and those for the panel data. I begin with the cross-sectional analysis.

Cross-Sectional Analysis

I begin the cross-sectional analysis by reproduce the findings of Acemoglu et al. (2001); this is shown in Columns 1 and 3 of Table 1. When using log settler mortality as an instrument, I find a relationship between expropriation risk and log GDP per capita that is substantively identical to the one reported in the original publication. More interestingly, I also find a substantively similar relationship when I substitute the average ICRG corruption score measure of institutional quality for Acemoglu et al.’s original average appropriation risk measure. Acemoglu et al. illustrate the substantive meaning of their finding by comparing two cases in their data set, Nigeria and Chile: using the model of Column 1, Nigeria is predicted to be (counterfactually) about 755% richer if its risk of expropriation (≈ 5.5) were equivalent to Chile’s (≈ 7.8) (with a 90% confidence interval of 381% and 1421%). By

---

13 The standard errors are slightly different than those reported in the original publication, possibly because I use the ivreg2 routine to estimate the instrumental variable models (Baum et al., 2010).
Table 1: Cross-Sectional IV Models, log Settler Mortality instrument

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. Protection from</td>
<td>0.944***</td>
<td>0.982***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expropriation Risk, 1985-95</td>
<td>(0.154)</td>
<td>(0.288)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg. ICRG Corruption Score, 1985-1995</td>
<td>-1.626**</td>
<td>-0.918**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.516)</td>
<td>(0.355)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa dummy</td>
<td>-0.464</td>
<td>-1.253***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.344)</td>
<td>(0.264)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia dummy</td>
<td>-0.924*</td>
<td>-0.446</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.384)</td>
<td>(0.385)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other continent dummy</td>
<td>-0.941</td>
<td>-0.693</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.814)</td>
<td>(0.894)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>1.910</td>
<td>13.03***</td>
<td>2.032</td>
<td>11.46***</td>
</tr>
<tr>
<td></td>
<td>(1.011)</td>
<td>(1.599)</td>
<td>(1.931)</td>
<td>(1.122)</td>
</tr>
<tr>
<td>N</td>
<td>64</td>
<td>61</td>
<td>64</td>
<td>61</td>
</tr>
<tr>
<td>Anderson LM (weak ID)</td>
<td>17.29</td>
<td>8.129</td>
<td>6.115</td>
<td>6.157</td>
</tr>
<tr>
<td>Anderson LM p-value</td>
<td>0.0000321</td>
<td>0.00436</td>
<td>0.0134</td>
<td>0.0131</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
Dependent Variable: log GDP per capita, 1995
* p < 0.05, ** p < 0.01, *** p < 0.001

Models are two-stage least squares instrumental variable models estimated using the ivreg2 command of Baum et al. (2010) in Stata 14.0. Instrument: log Settler Mortality. Columns 1 and 3 replicate Columns 1 and 7 from Table 4 of Acemoglu et al. (2001).

Comparison, when using the ICRG corruption measure in the model of Column 2, Nigeria is predicted to be about 408% richer (90% CI between 118% and 1087%) if its corruption score (4) were the same as Chile’s (3). Although this relationship is slightly smaller, it is broadly comparable in magnitude (and objectively still very large).

However, I get different answers when I substitute my alternative instrument, the percentage of women in the lower house of parliament in 1990 * whether a state is democratic-leaning, for the log settler mortality instrument; these results are shown in Columns 1-4 of Table 2. None of these models show a statistically significant relationship between either expropriation risk or ICRG corruption score and log per capita GDP in 1995.
Table 2: Cross-Sectional IV Models, new instruments

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. Protection from Expropriation Risk, 1985-95</td>
<td>-0.221</td>
<td>-0.123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.207)</td>
<td>(1.183)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avg. ICRG Corruption Score, 1985-1995</td>
<td>0.0153</td>
<td></td>
<td>-0.102</td>
<td>-0.504**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.273)</td>
<td></td>
<td>(0.218)</td>
<td>(0.174)</td>
<td></td>
</tr>
<tr>
<td>Democracy</td>
<td>0.969</td>
<td>0.816**</td>
<td>-0.0996</td>
<td>-0.441</td>
<td>-0.660</td>
</tr>
<tr>
<td></td>
<td>(1.555)</td>
<td>(0.266)</td>
<td>(0.478)</td>
<td>(0.342)</td>
<td>(0.346)</td>
</tr>
<tr>
<td>% Women in Parliament, 1990</td>
<td>-0.0102</td>
<td>-0.00495</td>
<td>-0.0364</td>
<td>-0.0346*</td>
<td>-0.0321</td>
</tr>
<tr>
<td></td>
<td>(0.0533)</td>
<td>(0.0207)</td>
<td>(0.0425)</td>
<td>(0.0173)</td>
<td>(0.0178)</td>
</tr>
<tr>
<td>Africa dummy</td>
<td></td>
<td></td>
<td>-1.191</td>
<td>-1.481***</td>
<td>-1.649***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.625)</td>
<td>(0.365)</td>
<td>(0.373)</td>
</tr>
<tr>
<td>Asia dummy</td>
<td></td>
<td></td>
<td>-0.702</td>
<td>-0.852**</td>
<td>-0.824**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.605)</td>
<td>(0.293)</td>
<td>(0.303)</td>
</tr>
<tr>
<td>Other continent dummy</td>
<td></td>
<td></td>
<td>1.250</td>
<td>0.847</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.316)</td>
<td>(0.552)</td>
<td>(0.509)</td>
</tr>
<tr>
<td></td>
<td>(14.44)</td>
<td>(8.88)</td>
<td>(8.103)</td>
<td>(8.064)</td>
<td>(7.744)</td>
</tr>
<tr>
<td>N</td>
<td>51</td>
<td>50</td>
<td>51</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Instrument</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>Sargan stat. (over ID)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.934</td>
</tr>
<tr>
<td>Sargan p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00485</td>
</tr>
<tr>
<td>Anderson LM (weak ID)</td>
<td>0.130</td>
<td>6.843</td>
<td>0.321</td>
<td>7.357</td>
<td>12.29</td>
</tr>
<tr>
<td>Anderson LM p-value</td>
<td>0.719</td>
<td>0.00890</td>
<td>0.571</td>
<td>0.00668</td>
<td>0.00214</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
Dependent Variable: log GDP per capita, 1995
* p < 0.05, ** p < 0.01, *** p < 0.001

Models are two-stage least squares instrumental variable models estimated using the ivreg2 command of Baum et al. (2010) in Stata 14.0. Instruments: 1 = (% Women in Parliament in 1990) * (Polity2 Score > 0); 2 = log Settler Mortality.
In the case of models using average expropriation risk as the key independent variable, the lack of a strong result may be explainable because the interaction term is not an especially good instrument for the expropriation risk measure; the Anderson LM (Lagrange Multiplier) statistic cannot reject the null hypothesis that the model is underidentified. However, this explanation does not apply to models using the ICRG corruption score: this score is associated with the interaction term in the expected way (see the first stage models in Appendix Table 5) and the model rejects the null of underidentification. Furthermore, when both the log Settler Mortality and interaction instruments are included in the model for ICRG corruption (Column 5), the statistically significant relationship re-emerges (although with a substantially smaller coefficient). However, the Sargan test rejects the validity of the instruments \( p < 0.01 \), indicating that at least one of the instruments may be invalid.

My overall conclusion from a reanalysis of Acemoglu et al.’s (2001) cross-sectional data set is that the measured causal relationship between corruption and economic performance is sensitive to choice of instrument. Moreover, the Sargan test rejects the validity of the instruments in the model indicating a statistically significant relationship. The analysis might benefit from a larger dataset that enable more powerful hypothesis tests (and the efficient recovery of smaller relationships between corruption and growth). I therefore proceed to the panel analysis of democratic-leaning countries.

**Panel Analysis**

I estimated four models on the data: two ordinary least squares (OLS) linear regression models and two instrumental variable/two-stage least squares linear regression (IV/2SLS) models.  

\[ \text{See the help file for the Stata package referenced in Baum et al. (2010) under “tests of under- and weak identification.”} \]

\[ \text{This is also consistent with my findings when I analyze the democratic-leaning countries in the cross-sectional data set using my other three interaction-based instruments; see Appendix Tables 6 and 7.} \]
models. All models included fixed effects for year and either region\textsuperscript{16} or country\textsuperscript{17}. The results are depicted in Table 3.

The 2SLS results indicate a consistently stronger (negative) relationship between corruption and log GDP per capita than is indicated by the comparable OLS regression. Moreover, the Durbin statistic indicates that corruption is endogeneously related to log GDP per capita in both 2SLS models (the null of exogeneity is rejected at $\alpha = 0.01$ in both cases). These results confirm our intuition that simple examination of the GDP-corruption relationship is insufficient to determine how much corruption influences economic development. However, a test for overidentification suggests that the instruments are only valid conditional on the inclusion of country-level fixed effects; the null (that the instruments are valid) is rejected in the model with only regional fixed effects. I therefore focus on the 2SLS model with country fixed effects for interpretation of substantive relationships.

The 2SLS model with country fixed effects indicates that a one point increase in ICRG corruption score is associated with a 0.258 decrease in log per capita GDP with a 90% confidence interval of $[-0.454, -0.0625]$; this relationship is statistically significant at conventional levels ($p = 0.005$, one-tailed). In order to evaluate the substantive importance of this relationship, I emulate Acemoglu et al. (2001) by comparing Chile and Nigeria: if the perceived corruption levels of Nigeria (average ICRG score = 4.76) were the same as those of Chile (average ICRG score = 2.28), Nigeria’s GDP per capita would be 89.9% higher (with a 90% confidence interval of 26.6% and 184.8%). This is a much smaller effect than the one calculated by Acemoglu et al. (2001), who find that Nigeria would be about 755% richer if its risk of expropriation were equivalent to Chile’s. It is also smaller that the effect of the same ICRG corruption score on log GDP per capita when calculated using the Acemoglu et al. data set with their log settler mortality instrument; I found that Nigeria would be about

\textsuperscript{16}The regions are Sub-Saharan Africa, South Asia, East Asia, South East Asia, Pacific Islands/Oceania, Middle East/North Africa, Latin America, Caribbean and non-Iberic America, Eastern Europe/Soviet Union, and Western Europe.

\textsuperscript{17}Year and regional fixed effects were handled by simply adding dummy variables into a typical OLS or 2SLS model. Country fixed effects were handled via the xtivreg2 command developed by Schaffer (2010).
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
<td>IV</td>
</tr>
<tr>
<td>ICRG Corruption Score</td>
<td>-0.250***</td>
<td>-0.583***</td>
<td>0.00287</td>
<td>-0.258**</td>
</tr>
<tr>
<td></td>
<td>(0.0233)</td>
<td>(0.0936)</td>
<td>(0.0194)</td>
<td>(0.0994)</td>
</tr>
<tr>
<td>% Women in Parliament</td>
<td>0.00984***</td>
<td>-0.00323</td>
<td>-0.00254</td>
<td>-0.00231</td>
</tr>
<tr>
<td></td>
<td>(0.00235)</td>
<td>(0.00457)</td>
<td>(0.00296)</td>
<td>(0.00187)</td>
</tr>
<tr>
<td>Press Freedom</td>
<td>0.0231***</td>
<td>0.0182***</td>
<td>0.00741***</td>
<td>0.00735***</td>
</tr>
<tr>
<td></td>
<td>(0.00217)</td>
<td>(0.00254)</td>
<td>(0.00183)</td>
<td>(0.00132)</td>
</tr>
<tr>
<td>Presidentialism</td>
<td>-0.155*</td>
<td>-0.198**</td>
<td>-0.248</td>
<td>-0.107</td>
</tr>
<tr>
<td></td>
<td>(0.0615)</td>
<td>(0.0616)</td>
<td>(0.203)</td>
<td>(0.101)</td>
</tr>
<tr>
<td>Personalism</td>
<td>0.0270***</td>
<td>0.0228***</td>
<td>0.0135</td>
<td>0.0157*</td>
</tr>
<tr>
<td></td>
<td>(0.00589)</td>
<td>(0.00626)</td>
<td>(0.0124)</td>
<td>(0.00678)</td>
</tr>
<tr>
<td>Polity 2 Score</td>
<td>0.0639***</td>
<td>0.0359*</td>
<td>0.0186</td>
<td>0.00469</td>
</tr>
<tr>
<td></td>
<td>(0.0152)</td>
<td>(0.0176)</td>
<td>(0.0186)</td>
<td>(0.00996)</td>
</tr>
<tr>
<td>Trade Imbalance (% of GDP)</td>
<td>-0.00267***</td>
<td>-0.00182*</td>
<td>-0.00400***</td>
<td>-0.00275**</td>
</tr>
<tr>
<td></td>
<td>(0.000630)</td>
<td>(0.000710)</td>
<td>(0.00111)</td>
<td>(0.000889)</td>
</tr>
<tr>
<td>Year FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Region FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Country FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1653</td>
<td>1653</td>
<td>1653</td>
<td>1653</td>
</tr>
<tr>
<td>Sargan stat. (over ID)</td>
<td>22.68</td>
<td>1.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan p-value</td>
<td>0.0000119</td>
<td>0.605</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderson LM (weak ID)</td>
<td>66.04</td>
<td>16.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderson LM p-value</td>
<td>3.01e-14</td>
<td>0.000926</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Panel IV Models in Democratic-Leaning Countries

Models 1 and 3 are ordinary least squares regression models; model 1 is estimated using the *regress* command with region and year dummy variables, while model 3 is estimated using *xtreg* with the *fe* option (on country) and with year dummy variables. Models 2 and 4 are two-stage least squares instrumental variable models; model 2 is estimated using the *ivreg2* command of Baum et al. (2010) with region and year dummy variables, while model 4 is estimated using the *xtivreg2* command of Schaffer (2010) with the *fe* option (on country) and with year dummy variables. Dummy variables and intercepts are omitted from the table. All models are estimated in Stata 14.0. 2SLS model instruments: (% women)*(press freedom), (% women)*(presidential system), and (% women)*(personalism). Data set covers 95 democratic-leaning countries (polity score > 0) between 1990 and 2010 for a total of 1653 observations (with an average of 17.4 observations per country). Robust standard errors are reported.

Standard errors in parentheses
Dependent Variable: log GDP per capita
* p < 0.05, ** p < 0.01, *** p < 0.001
408% richer if its corruption score was the same as Chile’s. Nevertheless, this relationship calculated using the panel model still represents a substantial improvement in economic development and a much larger relationship than would be indicated by a fixed effects model without instrumental variables.

Conclusion

There are two conclusions I believe can be drawn from this paper, one methodological and one substantive. Methodologically, a mathematical argument (based on Pearl (2000)) establishes that, when the effect of a variable \( v \) on \( x \) depends on a conditional factor \( w \) but the effect of \( v \) on \( y \) does not, then a multiplicative interaction term \( vw \) can be used to recover the causal effect of \( x \) on \( y \) \( \frac{dy}{dx} \). A Monte Carlo simulation analysis verifies that the strategy is effective. This is important because good instrumental variables are both important and difficult to find. Expanding the universe of potential instruments is important for enabling the empirical identification of as many causal relationships as possible.

Substantively, I use this identification strategy to address a controversy in the corruption literature. Prior research in this literature has attempted to determine how much lowering corruption in a country (e.g., by institutional reform) would improve economic outcomes for that country’s citizens. But economic development is likely to be a powerful influence on the structure of a country’s institutions; the two outcomes are endogenously related. The controversy in the literature revolves around how to separate these two relationships; various instrumental variables have been proposed and then criticized (Acemoglu et al., 2001; Mauro, 1995; Treisman, 2000, 2007; Glaeser et al., 2004). The methodological findings of this paper suggest looking for a variable whose effect on corruption depends on context but whose effects (if any) on economic development are not contextually dependent. Some recent research suggests that women’s representation in parliament is such a variable (Esarey and Chirillo, 2013; Esarey and Schwindt-Bayer, 2015): the negative relationship between corruption and
women in government (Dollar et al., 2001; Swamy et al., 2001) tends to be strongest when institutions ensure politicians’ accountability to voters. But I have no reason to believe that women in government and accountability to voters interact in influencing economic development. Using three interaction-based instruments chosen according to this argument, I find that corruption substantially lowers per capita GDP: every one point improvement on the ICRG’s six point corruption scale raises per capita GDP by about 29%. However, this relationship is much smaller than the one previously measured by Acemoglu et al. (2001).

It is unlikely that this paper’s estimate of how strongly corruption lowers per capita GDP is in isolation any more definitive than earlier estimates (e.g., those of Acemoglu et al. 2001). However, earlier estimates were criticized because there was a plausible relationship between their instrument and economic development that did not pass through the independent variable of interest. In Acemoglu et al.’s case, the instrument of colonial settler mortality might be related to log GDP per capita through risk of expropriation, but it might also be related through initial levels of human capital Glaeser et al. (2004). The fact that my analysis also finds that greater corruption causes lower per capita GDP, despite a very different identification strategy, bolsters the qualitative conclusions of Acemoglu et al. (2001).

It is reasonable to believe that the smaller relationship between corruption and GDP measured in this paper is explained in part because the interaction-based instruments that I employ have fewer pathways to the dependent variable that do not pass through the independent variable. Ultimately, only future analysis of different data sets using different models will definitively clarify how powerfully corruption retards economic development. Perhaps the methodological insight that instruments can be built from one-sided conditional relationships will help to make some of these analyses possible.

References


Angrist, J. D. and A. B. Krueger (2001). Instrumental variables and the search for identifi-


Appendix A: Figures with Additional Values of Correlation Between \( v \) and \( u \)

Figures 6, 7, 8, and 9 repeat the analysis of Figures 2, 3, 4, and 5 from the main text but with additional values of \( \rho \). As described in the main text, the results for additional values of \( \rho \) are qualitatively similar to the results for \( \rho = -0.9 \) and 0.2.
The figure shows the median and 95% confidence boundaries of \( (\hat{\beta}_1 - \beta_1) \) at varying values of \( n \) and \( \rho \) for 1000 models on 1000 simulated data sets. The circles show results for simulations using an interaction-based instrument from the data generating process in equations 2-3; the triangles show results for simulations using a standard instrument from the data generating process in equations 4-5.
Figure 7: Simulation evidence, 95% CI coverage from 2SLS estimation of $dy/dx = \beta_1$, all values of $\rho$

The figure shows the proportion of the time that 95% confidence intervals for the $\hat{\beta}$ estimate cover the true value of $\beta$ at varying values of $n$ and $\rho$ for 1000 models on 1000 simulated data sets. The circles show results for simulations using an interaction-based instrument from the data generating process in equations 2-3; the triangles show results for simulations using a standard instrument from the data generating process in equations 4-5.
Figure 8: Simulation evidence, median standard errors for 2SLS estimation of $dy/dx = \beta_1$, all values of $\rho$

The figure shows the median standard error for estimated $\hat{\beta}$ at varying values of $n$ and $\rho$ for 1000 models on 1000 simulated data sets. The circles show results for simulations using an interaction-based instrument from the data generating process in equations 2-3; the triangles show results for simulations using a standard instrument from the data generating process in equations 4-5.
Figure 9: Simulation evidence for 2SLS estimation of $dy/dx = \beta_1$, weak vs. strong instruments, all values of $\rho$

The figure shows the median and 95% confidence boundaries of $(\hat{\beta}_1 - \beta_1)$ at varying values of $n$ and $\rho$ for 1000 models on 1000 simulated data sets. The simulations use an interaction-based instrument from the data generating process in equations 2-3. The circles indicate simulations with a “weak” instrument where $\alpha_5 = 0.1$, while the triangles indicate simulations with a “strong” instrument where $\alpha_5 = 0.5$. 
Appendix B: 2SLS First Stages from Tables 1 and 2

Tables 4 and 5 depict the first stages from instrumental variable models in Tables 1 and 2, respectively.
Table 4: First Stage from Table 1

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log Settler Mortality</td>
<td>-0.607***</td>
<td>0.341**</td>
<td>-0.432*</td>
<td>0.395*</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td>(0.113)</td>
<td>(0.173)</td>
<td>(0.158)</td>
</tr>
<tr>
<td>Africa dummy</td>
<td>-0.269</td>
<td>-0.485</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.413)</td>
<td>(0.366)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia dummy</td>
<td>0.333</td>
<td>0.331</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.498)</td>
<td>(0.427)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other continent dummy</td>
<td>1.241</td>
<td>-1.188</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.842)</td>
<td>(0.706)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>9.341***</td>
<td>1.470**</td>
<td>8.538***</td>
<td>1.427*</td>
</tr>
<tr>
<td></td>
<td>(0.611)</td>
<td>(0.550)</td>
<td>(0.783)</td>
<td>(0.706)</td>
</tr>
<tr>
<td>N</td>
<td>64</td>
<td>61</td>
<td>64</td>
<td>61</td>
</tr>
<tr>
<td>instrument</td>
<td>PRS</td>
<td>ICRG</td>
<td>PRS</td>
<td>ICRG</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001

These models are the first stage regressions from the model in Table 1; column numberings match those of the second stage table. The dependent (instrumental) variable is either the 1985-1995 average Political Risk Services risk of expropriation (PRS) or the 1985-1995 average ICRG corruption score (ICRG).
Table 5: First Stage from Table 2

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Women in Parliament, 1990</td>
<td>-0.0275</td>
<td>0.0546</td>
<td>-0.0403</td>
<td>0.0502</td>
<td>0.0368</td>
</tr>
<tr>
<td></td>
<td>(0.0375)</td>
<td>(0.0311)</td>
<td>(0.0394)</td>
<td>(0.0332)</td>
<td>(0.0321)</td>
</tr>
<tr>
<td>Democracy</td>
<td>0.469</td>
<td>1.065</td>
<td>-0.0913</td>
<td>1.107</td>
<td>0.757</td>
</tr>
<tr>
<td></td>
<td>(0.752)</td>
<td>(0.642)</td>
<td>(0.881)</td>
<td>(0.813)</td>
<td>(0.788)</td>
</tr>
<tr>
<td>% Women X Democracy</td>
<td>0.0271</td>
<td>-0.180**</td>
<td>0.0397</td>
<td>-0.176**</td>
<td>-0.161*</td>
</tr>
<tr>
<td></td>
<td>(0.0782)</td>
<td>(0.0665)</td>
<td>(0.0752)</td>
<td>(0.0646)</td>
<td>(0.0618)</td>
</tr>
<tr>
<td>log Settler Mortality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.479*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.204)</td>
</tr>
<tr>
<td>Africa dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.412</td>
<td>-0.160</td>
<td>-0.968</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.665)</td>
<td>(0.604)</td>
<td>(0.670)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.432</td>
<td>0.0233</td>
<td>0.0653</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.591)</td>
<td>(0.494)</td>
<td>(0.470)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other continent dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.913*</td>
<td>-1.817**</td>
<td>-0.908</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.788)</td>
<td>(0.651)</td>
<td>(0.731)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>6.601***</td>
<td>2.709***</td>
<td>6.941***</td>
<td>2.863***</td>
<td>1.163</td>
</tr>
<tr>
<td></td>
<td>(0.476)</td>
<td>(0.395)</td>
<td>(0.837)</td>
<td>(0.738)</td>
<td>(1.010)</td>
</tr>
<tr>
<td>N</td>
<td>51</td>
<td>50</td>
<td>51</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>instrument</td>
<td>PRS</td>
<td>ICRG</td>
<td>PRS</td>
<td>ICRG</td>
<td>ICRG</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001

These models are the first stage regressions from the models in Table 2; column numberings match those of the second stage table. The dependent (instrumental) variable is either the 1985-1995 average Political Risk Services risk of expropriation (PRS) or the 1985-1995 average ICRG corruption score (ICRG).
Appendix C: Analysis of Democratic-Leaning States in the Cross-Sectional Dataset of Acemoglu et al. (2001)

My main analysis of the cross-sectional data of Acemoglu et al. (2001) is limited to the interaction between women’s representation and democratic-leaning status because this instrument can be applied to most cases in that already small data set. However, the other three instruments can be applied to the democratic-leaning states in this data set. I have done so in Tables 6 and 7. Table 6 shows the results for models using the protection against expropriation risk as the key (endogenous) independent variable; Table 7 shows models using the ICRG corruption score in its place.

Substantively, the results of these models bear many similarities to the results from the panel analysis in Table 3. Consider, first, the results for average protection from expropriation risk (Table 6). The first column uses the log Settler Mortality instrument of Acemoglu et al. (2001). The second, third, and fourth columns input the press freedom, presidentialism, and personalism interaction instruments; each of these variables is measured in the year 1990. Each of the new instruments causes more cases to be dropped from the analysis due to missing data.

Two of the three models with new instruments (Columns 2 and 3) indicate a smaller relationship between protection from expropriation risk and log per capita GDP. The model in Column 4, which includes all three instruments, indicates a larger relationship; however, this model also fails to reject the null of an underidentified model in the Anderson LM test.19

There is a similar pattern in the findings of Table 7, showing models with the ICRG

18All three of these variables are drawn from the Quality of Government dataset (Teorell et al., 2015b); while they generally match the descriptions and sources referenced with regard to the Schwindt-Bayer and Tavits (2015) data set, there are some differences. First, the freedom of the press variable is measured with three categories (free, partly free, and not free); this reflects an earlier coding scheme from Freedom House (Teorell et al., 2015a, pp. 236-237). I recoded this variable so that freer presses corresponded to higher numerical values of the variable. Second, the presidentialism variable is coded as = 1 if Cheibub et al. (2010) classify the regime as a presidential democracy or mixed (semi-presidential) democracy, and = 0 if they classify the regime as a parliamentary democracy (Teorell et al., 2015a, p. 106).

19See the help file for the Stata package referenced in Baum et al. (2010) under “tests of under- and weak identification.”
Table 6: Cross-Sectional Models for Expropriation Risk in Democratic-Leaning Countries (Polity2 > 0)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. Protection from Expropriation Risk, 1985-95</td>
<td>0.741***</td>
<td>0.558*</td>
<td>0.454**</td>
<td>0.984*</td>
</tr>
<tr>
<td></td>
<td>(0.164)</td>
<td>(0.257)</td>
<td>(0.216)</td>
<td>(0.432)</td>
</tr>
<tr>
<td>% Women in Parliament, 1990</td>
<td>-0.00921</td>
<td>0.00614</td>
<td>-0.00126</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0337)</td>
<td>(0.0268)</td>
<td>(0.0352)</td>
<td></td>
</tr>
<tr>
<td>Press Freedom, 1990</td>
<td>-0.0970</td>
<td>0.394</td>
<td>-0.0381</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.449)</td>
<td>(0.360)</td>
<td>(0.565)</td>
<td></td>
</tr>
<tr>
<td>Presidentialism, 1990</td>
<td></td>
<td>0.446</td>
<td>-0.399</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.326)</td>
<td>(0.510)</td>
<td></td>
</tr>
<tr>
<td>Personalism, 1990</td>
<td></td>
<td></td>
<td></td>
<td>-0.222</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.133)</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.449**</td>
<td>4.990***</td>
<td>4.047***</td>
<td>3.044</td>
</tr>
<tr>
<td></td>
<td>(1.145)</td>
<td>(1.107)</td>
<td>(1.179)</td>
<td>(1.607)</td>
</tr>
<tr>
<td>N</td>
<td>34</td>
<td>30</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>Instruments</td>
<td>1</td>
<td>2</td>
<td>2, 3</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>Sargan stat. (over ID)</td>
<td></td>
<td>0.116</td>
<td>0.881</td>
<td></td>
</tr>
<tr>
<td>Sargan p-value</td>
<td></td>
<td>0.733</td>
<td>0.644</td>
<td></td>
</tr>
<tr>
<td>Anderson LM (weak ID)</td>
<td>9.586</td>
<td>4.777</td>
<td>4.542</td>
<td>2.983</td>
</tr>
<tr>
<td>Anderson LM p-value</td>
<td>0.00196</td>
<td>0.0288</td>
<td>0.103</td>
<td>0.394</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
Dependent Variable: log GDP per capita, 1995
* p < 0.05, ** p < 0.01, *** p < 0.001

Dependent variable: log GDP per capita in 1995. Instruments: 1 = log Settler Mortality; 2 = Press Freedom (three category); 3 = Presidentialism (binary); 4 = Personalism (0-13).
Table 7: Cross-Sectional Models for ICRG Corruption in Democratic-Leaning Countries (Polity2 > 0)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. ICRG Corruption</td>
<td>-0.791***</td>
<td>-2.010</td>
<td>-0.431</td>
<td>-0.453</td>
</tr>
<tr>
<td>Score, 1985-95</td>
<td>(0.236)</td>
<td>(3.314)</td>
<td>(1.105)</td>
<td>(0.359)</td>
</tr>
<tr>
<td>% Women in Parliament,</td>
<td>-0.260</td>
<td>-0.0604</td>
<td>-0.0620</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>(0.434)</td>
<td>(0.168)</td>
<td>(0.0708)</td>
<td></td>
</tr>
<tr>
<td>Press Freedom, 1990</td>
<td>-0.699</td>
<td>0.528</td>
<td>0.494</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.570)</td>
<td>(1.213)</td>
<td>(0.520)</td>
<td></td>
</tr>
<tr>
<td>Presidentialism, 1990</td>
<td></td>
<td>-0.0807</td>
<td>-0.107</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.257)</td>
<td>(0.509)</td>
<td></td>
</tr>
<tr>
<td>Personalism, 1990</td>
<td></td>
<td></td>
<td></td>
<td>0.0137</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0653)</td>
</tr>
<tr>
<td>Intercept</td>
<td>10.84***</td>
<td>18.11</td>
<td>8.975</td>
<td>9.064**</td>
</tr>
<tr>
<td></td>
<td>(0.696)</td>
<td>(19.17)</td>
<td>(7.696)</td>
<td>(2.987)</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>29</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>Instruments</td>
<td>1</td>
<td>2</td>
<td>2, 3</td>
<td>2, 3, 4</td>
</tr>
<tr>
<td>Sargan stat. (over ID)</td>
<td>3.060</td>
<td>4.130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan p-value</td>
<td>0.0803</td>
<td>0.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anderson LM (weak ID)</td>
<td>8.080</td>
<td>0.283</td>
<td>0.201</td>
<td>2.099</td>
</tr>
<tr>
<td>Anderson LM p-value</td>
<td>0.00447</td>
<td>0.595</td>
<td>0.904</td>
<td>0.552</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

Dependent Variable: log GDP per capita, 1995

* p < 0.05, ** p < 0.01, *** p < 0.001

Dependent variable: log GDP per capita in 1995. Instruments: 1 = log Settler Mortality; 2 = Press Freedom (three category); 3 = Presidentialism (binary); 4 = Personalism (0-13).
corruption independent variable. The first column uses the log Settler Mortality instrument, while the second through fourth columns add in new instruments. As before, two of the three models indicate a smaller relationship between corruption and log per capita GDP; the third model (with only the press freedom instrument) indicates a substantively larger relationship. However, none of these relationships is statistically significant. Moreover, none of the models in columns 2-4 can reject the null of underidentification in the Anderson LM test.

With such a small number of observations in a consistent (but not unbiased) model, it is hard to know what (if anything) we should conclude from these models; I focus on the democracy * women’s representation instrument in the main body of the paper because it allows more of the data set to be used in the analysis. However, insomuch that a conclusion can be supported, I believe that the models indicate uncertain, tentative evidence of a weaker relationship between corruption and economic performance than would be indicated using the log Settler Mortality instrument. This matches the conclusion that is produced from the panel data analysis.