81274

Does the Internet Reduce Corruption? Evidence from U.S. States and across Countries

Thomas Barnebeck Andersen, Jeanet Bentzen, Carl-Johan Dalgaard, and Pablo Selaya

We test the hypothesis that the Internet is a useful technology for controlling corruption. In order to do so, we develop a novel identification strategy for Internet diffusion. Power disruptions damage digital equipment, which increases the user cost of IT capital, and thus lowers the speed of Internet diffusion. A natural phenomenon causing power disruptions is lightning activity, which makes lightning a viable instrument for Internet diffusion. Using ground-based lightning detection censors as well as global satellite data, we construct lightning density data for the contiguous U.S. states and a large cross section of countries. Empirically, lightning density is a strong instrument for Internet diffusion and our IV estimates suggest that the emergence of the Internet has served to reduce the extent of corruption across U.S. states and across the world. JEL Classification codes: K4, O1, H0

Corruption is commonly perceived to be a major stumbling block on the road to prosperity. Aside from retarding growth (Mauro 1995), corruption entails fiscal leakage, which reduces the ability of poor countries to supply essential

Thomas Barnebeck Andersen (corresponding author) is professor at the Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M. Jeanet Bentzen is Ph.D. student, Carl-Johan Dalgaard is professor, and Pablo Selaya is assistant research professor at the Department of Economics, University of Copenhagen, Studiestraede 6, DK-1455 Copenhagen K, Denmark. Contact: Thomas Barnebeck Andersen (barnebeck@sam.sdu.dk), Jeanet Bentzen (jeanet.bentzen@econ.ku.dk), Carl-Johan Dalgaard (carl.johan.dalgaard@econ.ku.dk), and Pablo Selaya (pablo.selaya@econ.ku.dk). We owe a special thanks to William A. Chisholm for generously sharing his expertise on the subject of lightning protection. We also thank three anonymous referees, Phil Abbott, Shekhar Aiyar, Mark Bills, Christian Bjørnskov, Matteo Cervellati, Areendam Chanda, Oded Galor, Chad Jones, Sam Jones, Phil Keefer, Pete Klenow, David Dreyer Lassen, Ross Levine, Norman Loayza, Ben Olken, Ola Olsson, Elena Paltseva, Nancy Qian, Martin Ravallion, Yona Rubinstein, Paolo Vanin, Dietz Vollrath, David Weil, Nadja Wirz, and seminar participants at the University of Bergen, Brown University, University of Copenhagen, the Jerusalem Summer School in Economic Growth at Hebrew University, ETH Zürich, the ZEW Conference on Economics of Information and Communication Technologies, the 2008 Nordic Conference in Development Economics, the Growth Workshop at the 2008 NBER Summer Institute, the Southern Economic Association Annual Meeting 2008, the Latin American Econometric Society Meeting 2009, and the World Bank. Errors are ours.

THE WORLD BANK ECONOMIC REVIEW, VOL. 25, NO. 3, pp. 387–417 doi:10.1093/wber/lhr025
Advance Access Publication May 27, 2011

public services such as schooling and health care (Reinikka and Svensson 2004; World Development Report 2004). Corruption is unquestionably a governance failure one would like to dispose of, yet combating it has not proven easy.

In the present paper we hypothesize that the Internet is a useful technology for combating corruption around the world. We test this hypothesis using data for the 48 contiguous U.S. states as well as for a cross section of countries. Our estimates support the proposition that the Internet has worked to reduce corruption since its inception.

There are several reasons why the Internet could serve as a corruption-reducing technology. First, the Internet is among a select group of innovations that are considered General Purpose Technologies (e.g., Jovanovic and Rousseau 2005). General Purpose Technologies are fundamental and pervasive innovations that (with time) hold a first order impact on economic growth. Insofar as economic growth works to lower corruption, the Internet is likely to act as a corruption suppressor due to its positive impact on growth.[1] Second, rapid technological change usually encourages investment in human capital, which may instigate lower levels of corruption.[2] Accordingly, human capital accumulation is another potential transmission channel linking Internet diffusion and reductions in corruption levels. Third, the World Wide Web is a major source of information. Spreading information about official wrong-doing inevitably increases the risk of detection for politicians and public servants, thus making corrupt behavior less attractive.[3] Fourth, the Internet is the chief vehicle for the provision of E-government worldwide (West 2005). By allowing citizens access to government services online, E-government obviates bureaucrats' role as intermediaries between the government and the public, thus limiting the interaction between potentially corrupt officials and the public. Moreover, online systems require standardized rules and procedures. This reduces bureaucratic discretion and increases transparency as compared to the

1. See Andersen *et al.* (2010) for evidence that the Internet has stimulated growth across U.S. states, and e.g. Gundlach and Paldam (2009) for evidence of a causal impact of income on corruption.

2. See Schultz (1975) and Foster and Rosensweig (1996) for evidence of a positive impact of technological change on the return to human capital investments, and Glaeser and Saks (2006) for evidence on the impact of human capital on corruption.

3. A nice illustration of this mechanism at work is found in a 2001 scandal from India, which nearly toppled the government. Reporters from the online news site <www.Tehelka.com> posed as arms dealers and documented negotiations with top politicians and bureaucrats over the size of required side payments to get a contract; in some instances the reporters even got the delivery of the bribe on camera. Consequently, numerous politicians and top officials had to resign, chief among them the defence minister George Fernandes. See The Sting That Has India Writhing by Celia W. Dugger, *The New York Times* (March 16, 2001).

arbitrariness available to civil servants when dealing with the public on a case-by-case basis.[4]

The present paper examines the *reduced form* effect of the Internet on corruption levels across the U.S. and across the World. Thus, our analysis addresses the key issue of whether the Internet has had a causal impact on corruption. At the same time, the analysis admittedly does not clarify the exact mechanism(s) through which the Internet affects corruption; it may be either one of the mechanisms mentioned above, or some combination of them.

The Internet is a new technology. Indeed, if we identify the Internet with the World Wide Web, it only emerged in 1991.[5] To examine the impact of the Internet we therefore estimate the impact of *changes* in Internet users on *changes* in corruption levels from the early 1990s to 2006. Using data for the 48 contiguous U.S. states as well as cross-country data we establish a strong partial correlation between the rate of changes in Internet users and the evolution of corruption, consistent with the proposed hypothesis. However, since the speed of Internet diffusion is likely endogenous, OLS estimates might be misleading.

In an effort to establish causality we develop an identification strategy designed to isolate exogenous variation in the speed of Internet diffusion. The theory underlying our instrument of choice is the following. Computer equipment is highly sensitive to power disruptions: power surges and sags lead to equipment failure and damage. Consequently, a higher frequency of power disruptions implies higher costs of IT equipment, either through elevated IT capital depreciation or due to incurrence of additional costs in order to protect equipment from power disruptions. Frequent power disturbances are also likely to reduce the productivity of IT capital (or its marginal benefit), as power disturbances produce downtime, generate data glitches, etc. A natural phenomenon which produces power disturbances is lightning activity. In fact, one third of all power disruptions in the U.S. are related to lightning activity. We therefore hypothesize that higher lightning intensity is a viable candidate instrument for Internet diffusion. Using state level measures of lightning density (ground strikes per square km per year) and global satellite data on lightning activity,

4. The celebrated Bhoomi program (located in the state of Karnataka in India) constitutes a good example of the effectiveness of E-government in limiting the interface between civil servants and the public. Starting in 1998 the program aimed to computerize land records, and at the time of writing more than 20 million landholdings belonging to the state's 6.7 million landowners have been registered. Before online registration was available citizens had to seek out village accountants to register, a process which involved considerable delays and the need for bribes to be paid; a typical bribe could range from Rs.100 to Rs. 2,000 (US $2 to $40) (Bhatnager 2003). With the online system there is no longer a need for the official middlemen, implying that the online system probably has saved locals hundreds of millions of Rs. in bribes. See also Andersen (2009).

5. TheWorld Wide Web was launched in 1991 by CERN (the European Organisation for Nuclear Research). See Hobbes' Internet Timeline v8.2 <http://www.zakon.org/robert/internet/timeline/>. In this paper, we define the Internet/WWW as the network of networks using the TCP/IP/HTTP protocols, which was spawned by the launch of WWW.

we establish its influence on Internet diffusion: areas (states and countries) with a higher flash density have experienced a slower speed of Internet adoption.

A concern of first-order importance is whether, conditional on Internet diffusion, lightning density acts as a stand-in for factors that are correlated with corruption. To fix ideas, suppose areas with high lightning density just happen to be characterized by an abundance of natural resources as well, which influences corruption in its own right (i.e., conditional on Internet penetration). If so, then lightning is not a valid instrument for the Internet.

Notice, however, that the proposed instrument is most likely to be (spuriously) correlated with factors that are exerting a time persistent effect on corruption. To stay with the example from before: lightning prone areas that were rich in natural resources in (say) 1970 most likely remain so, and if the presence of natural resources fueled corruption in the 1970s, it would probably also spur corruption at the beginning of the 21st century. Accordingly, if lightning density is picking up this sort of influence we would expect to see a fairly time invariant correlation between lightning and (changes in) corruption over the last three or four decades. However, if lightning density is picking up the influence from processes instigated by the emergence of the Internet (be that a growth surge, an acceleration in human capital accumulation, improved diffusion of information, or E-government), one would expect to see a time varying correlation between lightning and (changes in) corruption. In fact, instrument validity would require that there is *no* correlation between lightning density and changes in corruption before 1991 (the founding year of the World Wide Web). These considerations form the basis of a falsification test: If lightning is correlated with (changes in) corruption prior to 1991, it is unlikely to satisfy the exclusion restriction.

As documented below, lightning density exhibits a *time-varying correlation* with corruption; the reduced form relationship between lightning density and corruption does not exist prior to the inception of the World Wide Web. Using U.S. state-level data, as well as cross country data, we are able to establish that the lightning/corruption correlation only exists after 1991. This falsification test makes probable that the lightning instrument is capturing processes instigated by the emergence of the Internet, and it therefore supports the use of lightning as an instrument for Internet diffusion.

Against this background we employ lightning density as an instrument for Internet diffusion. Our 2SLS estimates corroborate the OLS results: Rising Internet use over the 1990s reduced corruption in the U.S. and across countries. Our results are admittedly stronger, statistically speaking, for the U.S. Still, our cross-section analysis does suggest that the U.S. state-level results generalize to an international setting.

The present research is related to the literature which studies the determinants of the level of corruption. Notable contributions include Ades and di Tella (1999), Treisman (2000), Brunetti and Weder (2003), Persson, Tabellini, and Trebbi (2003), Glaeser and Saks (2006), and Licht, Goldscmidt, and Schwartz (2007).

Since the Internet is a central source of information, the paper is also related to the political economy literature that studies the impact of information on governance more generally. This literature suggests that a better informed public serves to discipline the political establishment, thus affecting governance (e.g. Besley and Burgess 2002; Strömberg, 2004; Reinikka and Svensson 2004; Eisensee and Strömberg 2007; Ferraz and Finan 2008).

Finally, the paper is related to the literature which studies the determinants of the spread of the personal computer (e.g. Caselli and Coleman 2001) and the Internet (e.g. Chinn and Fairlie 2007) across countries. We add to this literature by documenting a link between lightning density and the speed of Internet diffusion.

The paper is structured as follows. In Section I we present our empirical specifications of choice. Section II outlines the identification strategy in detail; in particular, we explain how lightning activity impacts digital equipment. Section III provides an analysis of how the Internet has affected corruption across the U.S. states, whereas Section IV provides cross-country evidence. Section V concludes.

I. Specification

As argued above, we wish to understand the impact of the Internet (or World Wide Web) on the evolution of corruption, from the emergence of the former and onwards. Since the Internet is a General Purpose Technology it seems infeasible to try and isolate any particular mechanism linking the Internet to corruption (e.g., growth spurts, human capital, enhanced dissemination of information, etc.), for which reason we focus on the reduced form.

Accordingly, in the analysis to follow we mainly rely on the following parsimonious specification:

$$\Delta C_i = \alpha_0 + \alpha_1 \Delta INTERNET_i + \alpha_2 C_{initial,i} + \varepsilon_i, \qquad (1)$$

where ΔC_i is the change in corruption levels between an initial and a final year, $C_{final,i} - C_{initial,i}$, and $\Delta INTERNET_i$ is the change in Internet penetration, $INTERNET_{final,i} - INTERNET_{initial,i}$.[6] When $\Delta C_i > 0$ means *increasing* corruption, the key hypothesis under investigation is whether $\alpha_1 < 0$ or not.

There are several reasons why we adopt this lagged dependent variables specification as our preferred specification. First, corruption is a naturally bounded variable. In the U.S. setting we employ corruption convictions as our measure of C, which is bounded from below at zero. Similarly in the

6. In the cross-state sample we also condition on state population size. State population enters as a control because we use total corruption convictions as our measure of state corruption; by including state size we thereby ensure that all scale effects are pruned from the data in a simple way.

cross-country analysis we employ a corruption index (the *ICRG* index), which is confined to a particular interval (zero to six) by construction. Obviously, for all country observations near either of the two endpoints, and in states near the zero boundary, one should expect the evolution of corruption to be subject to *mean reversion*. Unless we thus control for $C_{initial,i}$, our empirical model is likely to be misspecified. To spell it out formally, assume that equation (1) is the true population model, and suppose that $Cov(\varepsilon, \Delta INTERNET) = 0$. If we mistakenly ignore the lagged dependent variable and instead focus on the first difference equation, in an attempt to kill off an imagined fixed effect, we would perform OLS on $\Delta C_i = \alpha_0 + \alpha_1^{FD} \Delta INTERNET_i + \upsilon_i$, where $\upsilon_i \equiv \alpha_2 C_{initial,i} + \varepsilon_i$. The probability limit of α_1 obtained via the first differenced estimator is

$$\text{plim } \alpha_1^{FD} = \alpha_1 + \alpha_2 \frac{Cov(C_{initial}, \Delta INTERNET)}{Var(\Delta INTERNET)}. \tag{2}$$

Unless $Cov(C_{initial}, \Delta INTERNET) = 0$ we have that $\alpha_1^{FD} \neq \alpha_1$. Indeed, *a priori* $\alpha_1^{FD} \approx 0$ would be a likely outcome. To see this, note that the mean reversion process implies that $\alpha_2 < 0$. Moreover, if places with high initial corruption levels tend to see a slower diffusion of the internet then $Cov(C_{initial}, \Delta INTERNET) < 0$. Taken together this implies that $\alpha_2 \cdot Cov(C_{initial}, \Delta INTERNET) > 0$, which implies that the expected *negative* point estimate for α_1 is biased towards zero (see equation (2)).

Second, another virtue of including the initial level of corruption is that it automatically controls for (a potentially large set of) variables which may influence the evolution of corruption. To see the latter point more clearly observe that equation (1) is equivalent to a levels regression with a lagged dependent variable:

$$C_{final,i} = \alpha_0 + \alpha_1 \Delta INTERNET_i + (\alpha_2 + 1)\, C_{initial,i} + \varepsilon_i. \tag{3}$$

Accordingly, all time invariant structural characteristics affecting the *level* of corruption will be picked up by $C_{initial,i}$. This reduces the scope for omitted variable bias in contaminating the estimate of α_1 (Wooldridge 2003, p. 300).[7] Third, the core part of our analysis involves 2SLS estimation, where lightning density is invoked as an instrument for $\Delta INTERNET_i$. Since lightning density and the *level* of corruption are highly correlated, one may harbor legitimate concerns about the exclusion restriction if the latter is omitted from the empirical model; if the evolution of corruption is subject to mean reversion (i.e., if $\alpha_2 < 0$), trouble arises (see again equation (2)).

Naturally, one might also worry that the exclusion restriction is jeopardized by the omission of fixed effects in equation (1). But in the analysis below we

7. See also Angrist and Pischke (2009, Ch. 5.3), for a discussion of the virtues of using the specification in equation (1) *vis-à-vis* the first difference model.

are able to gauge the likely severity of this problem by performing the falsification test mentioned above: If the true empirical model involves fixed effects (i.e., α_0 in equation (1) should be country specific), and if our instrument is correlated with these unobservables, we would expect to see that lightning density is correlated with ΔC_i before as well as after 1991 (the founding year of the World Wide Web). As shown below, however, we are unable to reject the absence of a correlation between our instrument and changes in corruption *prior* to 1991. This holds true across U.S. states as well as across countries. As a result, we view the lagged dependent variables specification as the relatively safe choice when trying to elicit information about the causal impact of the Internet on corruption. Nevertheless, we will in the interest of completeness report the results from the first difference specification as well; that is, we also report the results from assuming $\alpha_2 = 0$ in equation (1).

II. Identification Strategy

There is good reason to believe that the adoption of new technologies, such as the Internet, is endogenous to governance. New technologies may create political as well as economic losers, for which reason incumbent entrepreneurs and politicians may try to block adoption (Mokyr 1990; Parente and Prescott 1999; Acemoglu and Robinson 2001). It seems plausible that places with widespread corruption, for example, may have adopted the Internet later, due to the influence of politicians, civil servants, or both. This mechanism rationalizes a positive impact of governance on the number of Internet users. Consequently, OLS estimation is unlikely to identify the impact of the Internet on corruption. To address this concern we employ an IV approach, the logic of which we now describe.

Computers are highly sensitive to even ultra brief power disruptions. Such disruptions are likely to cause down-time, though sudden power surges may also damage the equipment and randomly destroy or alter data. As observed in *The Economist*:[8] "For the average computer or network, the only thing worse than the electricity going out completely is power going out for a second. Every year, millions of dollars are lost to seemingly insignificant power faults that cause assembly lines to freeze, computers to crash and networks to collapse."

The reason why IT equipment is so sensitive to power disturbances is that computers are constructed to work under a clean electrical current, featuring a particular frequency and amplitude of voltage. The alternating power emanating from the commercial power plant is converted into direct current, after which transistors turn this small voltage on and off at several gigahertz during digital processing (Kressel 2007). However, if the input, in the form of the alternating current, is disturbed or distorted the conversion process is

8. The power industry's quest for the high nines, *The Economist*, March 22, 2001.

corrupted, which may in turn result in equipment failure and damage. Indeed, voltage disturbances measuring less than one cycle are sufficient to crash servers, computers, and other microprocessor-based devices; that is, at a 60 Hz frequency (the standard in the U.S.) this means that a power disturbance of a duration less than 1/60 th of a second is enough to crash a computer (Yeager and Stalhkopf 2000; Electricity Power Research Institute 2003). Importantly, this issue is unlikely to diminish over time as the sensitivity to small power distortions increases with the miniaturization of transistors, which is the key to increasing speed in microprocessors (Kressel 2007).

Accordingly, in areas with more power disturbances, the user cost of IT capital will be higher due to a higher rate of IT capital depreciation (Hall and Jorgenson 1967). By implication, the desired IT capital stock will be lower, reducing IT investments and the speed of Internet diffusion. Of course, steps may be taken to protect the equipment from power disturbances. A high-quality surge protector provides protection against voltage spikes, for example. High-tech companies install generators to supplement their power needs, thereby insuring themselves against power failure. They also add uninterruptible power sources relying on batteries to power computers until generators kick in. However, these initiatives will in any case increase the costs of acquiring digital equipment, and thereby the user cost of IT capital. The crux of the matter is that if one lives in an environment with low power quality, this adds to the costs of a computer.[9]

To this one may add that in areas with frequent power disruptions and outages, the marginal benefit of owning a computer is probably lowered as well. Obviously, in countries where firms and consumers face regular power outages it will be difficult to employ IT efficiently. But even if power disruptions are infrequent and of very short duration, power disruptions lead to glitches and downtime which serves to lower the productivity of IT equipment. Both mechanisms, higher marginal costs and lower marginal return/benefit, imply that poorer power quality should lead to a slower speed of Internet diffusion.

Naturally, power quality is not exogenous; it may well be determined by governance.[10] As a result, we employ a variable which generates exogenous variation in power quality, and thus IT costs and benefits, as an instrument for Internet diffusion. A natural phenomenon which interferes with digital equipment, by producing power failures, is lightning activity (e.g., Shim *et al.* 2000, Ch. 2; Chisholm 2000). By all accounts, the influence of lightning on power quality is substantial. According to some estimates, lightning is the direct cause

9. Besides, the above mentioned protective devises are not necessarily enough to ensure against damage. According to the National Oceanic and Atmospheric Administration (NOAA), a typical surge protector will not protect equipment from a nearby lightning strike. Generators, in turn, do not react fast enough and can deliver dirty power; batteries are expensive to maintain and may also not react fast enough. See e.g., The power industry's quest for the high nines, *The Economist* (March 22, 2001).

10. See Fredriksson *et al.* (2004) for evidence that corruption affects energy supply.

of one third of all power quality disturbances in the United States (Chisholm and Cummins 2006). Moreover, the probability of lightning-caused power interruptions or equipment damage scales linearly with lightning density (Chisholm 2000; Chisholm and Cummins 2006).[11] As a result, in areas with greater lightning density (strikes per square km per year) the (expected) rate of IT capital depreciation will tend to be larger. This implies higher IT investment costs, and possibly lower IT productivity as well.

The problems associated with lightning activity in the context of IT equipment have not escaped the attention of the popular press. A recent article in *The Wall Street Journal* highlights the practical relevance of lightning activity, and stresses the difficulty in shielding IT equipment:[12] "Even if electricity lines are shielded, lightning can cause power surges through unprotected phone, cable and Internet lines - or even through a building's walls. Such surges often show up as glitches. "Little things start not working; we see a lot of that down here," says Andrew Cohen, president of Vertical IT Solutions, a Tampa information-technology consulting firm. During the summer, Vertical gets as many as 10 calls a week from clients with what look to Mr. Cohen like lightning-related problems. Computer memory cards get corrupted, servers shut down or firewalls cut out."

Against this background we propose lightning density as an instrument for the speed at which Internet use per capita changed over the period in question. Schematically, we can express the theory underlying our identification strategy in the following way:

$$\text{LIGHTNING DENSITY} \rightarrow \text{POWER DISTURBANCES} \rightarrow \text{INTERNET USE}, \tag{4}$$

where the second arrow implicitly subsumes the impact of power disturbances on the costs and benefits of IT capital.

Lightning is certainly external in the sense of Deaton (2010). But this, of course, does not imply that it fulfills the exclusion restriction required for instrument validity. In particular, it could conceivably correlate with fixed factors (natural resource endowment, say) which themselves exert a persistent effect on the evolution of corruption. In order to examine whether this is likely to be the case or not, we perform a falsification test below: Under the hypothesis that lightning influences the evolution of corruption, via the Internet, changes in corruption should only be correlated with lightning after the emergence of the World Wide Web. We can in fact reject that lightning and changes in corruption are correlated prior to the invention of the World Wide Web

11. This linear scaling can be expressed precisely. Let N_S denote the number of strikes to a conducter per 100*km* of power line length, h the average height (in meters) of the conducter above ground level, and *GFD* the ground flash density, then $N_S = 3.8 \cdot GFD \cdot h^{0.45}$ (see Chisholm 2000).

12. There Go the Servers: Lightning's New Perils. *The Wall Street Journal*, August 25, 2009.

(i.e., prior to 1991) using data for both the U.S. States and our world sample. This suggests quite strongly that lightning is not – spuriously – correlated with country specific factors that persistently affect the path of corruption, thus supporting the exclusion restriction.

III. Cross-State Evidence

The cross-state analysis proceeds as follows: We first provide details on the data used for the analysis. Next we provide evidence on the partial correlation between changes in Internet usage and changes in corruption levels. Finally, before proceeding to our 2SLS estimates, we provide an independent check of the validity of our identification strategy.

Data

In measuring corruption in the U.S. we follow Glaeser and Saks (2006) by employing corruption convictions. The data derives from the Justice Department's Report to Congress on the Activities and Operations of the Public Integrity Section, a publication which provides statistics on the nationwide *federal* effort against public corruption, including the number of federal, state, and local public officials convicted of a corruption-related crime by state. As argued by Glaeser and Saks (2006), federal conviction levels capture the extent to which federal prosecutors have charged and convicted public officials for misconduct. There are potential problems with using conviction rates to measure corruption: in corrupt places, the judicial system is itself likely to be corrupt, meaning that fewer people will be charged with corrupt practices. This problem, however, is diminished when using federal convictions, the reason being that the federal judicial system is somewhat isolated from local corruption. Consequently, it should treat people similarly across states (Glaeser and Saks, 2006).

In concrete terms, we measure the *change* in corruption between 1991 and 2006 by the log difference in (one plus) the total number of corruption convictions in the two years; positive values in the rate of change reflect *increasing* corruption, and the choice of initial year follows from the fact that the Internet (in the sense of the World Wide Web) was introduced in 1990 (first Web page went online in 1991).[13]

The second key variable is Internet use, which we measure as the percentage of households with Internet access. It is based on data collected in a supplement to the October 2003 Current Population Survey (CPS), which includes

13. Note that an increased use of the Internet will both increase the *risk of detection* for a corrupt official (the detection technology is improved) as well as lower the *incentive* to commit corrupt acts. Hence, in theory, increased Internet use could increase the number of convictions if the former effect dominates. It might thus seem as if the Internet increases corruption. However, empirically the net effect is negative, implying that the incentive effect dominates, as documented below.

questions about computer and Internet use.[14] The CPS is a multi-stage probability sample with coverage in all states. The sample was selected from the 1990 Decennial Census files and is continually updated to account for new residential construction. To obtain the sample the United States is divided into 2,007 geographic areas, and about 60,000 households are eligible for interviews. Since U.S. corruption data goes back to 1991, the launch date of the WWW, we define the change in Internet use by state population as $\Delta INTERNET_i = INTERNET_{i,2003} - INTERNET_{i,1991} = INTERNET_{i,2003}$, since $INTERNET_{i,1991} = 0$ for all i (i.e., for all states).

In some specifications we also control for initial state population, or the growth of state population, which derives from the U.S. Bureau of Economic Analysis (BEA). The natural log of initial state population is denoted $\log(POP_{1991})$ and population growth is $\Delta \log(POP) = \log(POP_{2006} / POP_{1991})$. The purpose of including these controls is to make sure that all scale effects are pruned from the corruption data. Obviously, the *total* number of corruption convictions is likely greater in more populous states.

Finally, we measure lightning as the number of ground strikes per square kilometer per year. Accurate cloud-to-ground data exist for the 48 contiguous U.S. states, are measured by the U.S. National Lightning Detection Network (NLDN), and are provided by Vaisala for the period 1996-2005. NLDN consists of numerous remote, ground-based lightning sensors, which instantly detect (at a very high level of accuracy) the electromagnetic signals given off when lightning strikes Earth's surface.[15] Summary statistics are reported in Appendix Table A.1.

Partial Correlations

Table 1 documents the partial correlation between changes in Internet use and changes in corruption for the 48 contiguous U.S. states.

As a first step we report, in Columns 1 and 2, the results from the first difference specification where we thus omit initial corruption as a control; Column 1 is the basic specification, whereas population growth is included in Column 2 alongside Internet users. States that have seen larger increases in Internet use have also experienced larger reductions in corruption convictions, consistent with a corruption dampening effect of the Internet.

As explained in Section 2 we are concerned about the exclusion of the initial corruption level in the first difference specification, as changes in corruption convictions likely are subject to mean reversion. Specifically, in states with zero initial convictions one would almost inevitably expect an increase in convictions subsequently, regardless of Internet use.

In order to check whether this is an actual problem in the U.S. sample, we experiment, in Columns 3 and 4, with omitting states with zero corruption

14. <http://www.census.gov/population/socdemo/computer/2003/tab01B.xls>.

15. The data can be obtained, free of charge, here: <http://www.vaisala.com/thunderstorm>

TABLE 1. OLS Regressions, U.S. Sample

	Dependent variable: Change in corruption convictions 1991-2006					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔINTERNET	−0.058**	−0.057**	−0.060**	−0.057*	−0.058**	−0.051***
	(0.027)	(0.028)	(0.029)	(0.030)	(0.024)	(0.019)
Δlog(POP)		−0.243		−0.914		
		(1.068)		(0.972)		
log(POP_{1991})						0.849***
						(0.145)
log(1 + CC_{1991})					−0.386***	−0.873***
					(0.096)	(0.135)
Constant	3.667**	3.655**	3.577**	3.543**	4.392***	−7.790***
	(1.448)	(1.469)	(1.584)	(1.578)	(1.312)	(2.201)
Observations	48	48	40	40	48	48
R-squared	0.09	0.09	0.11	0.13	0.29	0.55

Notes: Robust standard errors in paranthesis. Asterisks ***, ** and * signify, respectively, p-value < 0.01, p-value < 0.05 and p-value < 0.1. All variables are defined in the main text.

Source: Authors' analysis based on data described in the text.

convictions in 1991. This should in theory reduce the extent of mean reversion in the sample. Further, in Columns 5 and 6, we employ the lagged dependent variable specification on the full sample: our preferred specification.

As seen from Columns 5 and 6, initial corruption is a significant correlate with subsequent changes in corruption; it adds considerably to the explanatory power of the regression model, as evidenced by the increase in R^2 when comparing Columns 1 and 2 with Columns 5 and 6.[16] In Columns 3 and 4 we truncate the sample by excluding states with zero initial corruption. If one includes initial corruption in this setting, it is insignificant (not shown). This shows that the significance of initial corruption in Columns 5 and 6 is due to mean reversion in the U.S. sample. Nevertheless, despite the presence of mean reversion in corruption we obtain nearly identical results for $\Delta INTERNET$ in all columns. The reason is simply that, statistically speaking, $Cov(C_{initial}, \Delta INTERNET) = 0$ in the U.S. sample. As a result, by equation (2), the first difference estimator is consistent; i.e., plim $\alpha_1^{FD} = \alpha_1$.[17] In any event, the bottom line is that the partial correlation is robust to the exact choice of specification.

The partial correlation is also robust to a long list of additional controls beyond what is reported in Table 1. Maxwell and Winters (2005) propose four sets of fundamental traits of U.S. states, which should have predictable effects

16. We have also experimented with including the growth rate of population, rather than its level, in the lagged dependent variables specification (Column 5 and 6), but much like in Column 2 and 4 it comes out insignificant.

17. By including the lagged corruption level, we can, however, reduce the error variance.

on corruption: (i) number of corruptible government bodies, (ii) the size of the state, (iii) socio-ethnic homogeneity, and (iv) civic-minded and well-informed political cultures. The authors consider seven additional control variables including income growth, general tax revenue and campaign expenditure restrictions. We have experimented with the inclusion of all the proposed corruption determinants (alongside initial corruption), and we find that changes in Internet use (1991–2006) remains significant. The results are available upon request.

Despite these encouraging results concerns about causality may legitimately be raised. We therefore further scrutinize the Internet/corruption nexus by way of instrumental variables estimation.

Instrument Falsification Test

2SLS estimates are obviously no better than the invoked instrument. In the present context we propose the use of lightning density as an instrument for Internet use, based on the theoretical argument we presented in Section II.

According to the identification strategy lightning is only allowed to influence changes in corruption via Internet use. If, by contrast, it is correlated with unobserved determinants of the evolution of corruption, the exclusion restriction fails, and the 2SLS estimates are no better than the OLS estimates discussed above.

Under the null of instrument validity, however, lightning should exert a time varying impact on the path of corruption. The reason is simply that the Internet is a new technology which cannot possibly have affected corruption prior to its inception and widespread use. By extension, determinants of Internet use should not affect corruption prior to the emergence of the World Wide Web.

Hence, in an effort to try to falsify our instrument we run the following regression:

$$\Delta C_i = \alpha_0 + \alpha_1 \log(LIGHTNING_i) + \alpha_2 C_{initial,i} + \varepsilon_i, \tag{5}$$

on two different time periods. First, we examine the link between lightning and changes in corruption during the Internet era: 1991 to 2006. In this period we expect a significantly positive point estimate for lightning, suggesting that high lightning states have experienced slower Internet diffusion and therefore smaller reductions in corruption levels than states with less lightning. Second, we examine the link between lightning and changes in corruption *prior* to the emergence of the World Wide Web. For instrument validity, the partial correlation between lightning and corruption should not be significantly different from zero. In the event of a significant correlation between lightning and changes in corruption we are forced to conclude that lightning is correlated with factors beyond the Internet (and the initial level of corruption), which influences changes in corruption. If so, the instrument is invalid.

The falsification test is potentially helpful in another respect: If changes in corruption are not correlated with lightning strikes in the pre-Internet age, the scope for lightning operating on corruption via other electronic technologies is limited. Indeed, it would appear unlikely that the impact date of such technologies coincides with the inception of the WWW. The introduction of the microprocessor, for instance, goes back to the early 1970s. By that time computers were widely used by the U.S. government. Therefore, if we think that increased and more efficient information storage capacity, say, reduces the scope for corruption, we should observe a correlation between lightning and corruption in the pre-Internet era. To the extent that this relationship is absent, the microprocessor *per se* is unlikely to influence the evolution of corruption.

Table 2 reports the results from this check. In Columns 1-4 we examine the pre-Internet periods 1976–1990, where we have missing observations for corruption in three states, and 1978–1990, where we have a full sample. The take-away is that lightning is not correlated with changes in corruption during these sub-periods, in keeping with the requirement that lightning is uncorrelated with other factors which exert a persistent impact on the evolution of corruption. Figure 1 provides a visual illustration of these findings.

In Columns 5 and 6 we shift attention to the period following the emergence of the World Wide Web. During this period lightning emerges as a strong correlate with changes in corruption levels: In states with more lightning we observe a slower rate of reduction in corruption compared to states with less lightning. This is consistent with the hypothesis that the Internet has worked to lower corruption, and that lightning has worked to slow down Internet diffusion. Figure 2 provides a visual illustration of the partial correlation between lightning and changes in corruption 1991–2006.

It is important to observe that the insignificance of lightning during the pre-Internet era is not simply a matter of imprecise estimates. As is evident from inspection of the results across columns, the size of the estimate itself rises many fold when moving from Columns 1–4 to Columns 5–6.

It may also be noted that the falsification test is consistent with lightning *not* operating through other electronic technologies, which would arguably have required that the correlation between lightning and corruption was in existence before 1991.

Overall, we find that these results constitute compelling evidence in favor of the exclusion restriction in the U.S. context. Hence, we now move to instrumental variables estimation where lightning is used as an instrument for Internet diffusion.

2SLS Estimates

Table 3 reports our 2SLS results; Panel A reports first stage results, whereas Panel B reports the second stage. In keeping with the approach taken in Table 1, we experiment with both the first difference specification and the lagged dependent variables specification. In the OLS setting (Table 1) we found

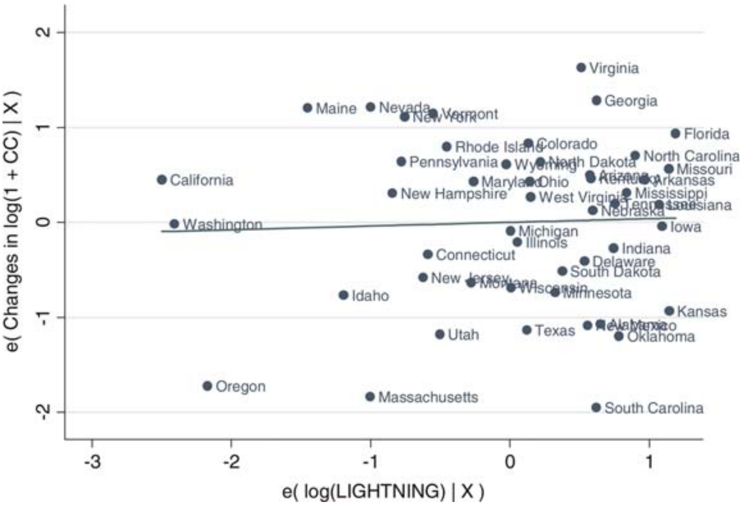
TABLE 2. Falsification test, U.S. Sample

	Change in corruption convictions 1976–1990		Change in corruption convictions 1978–1990		Change in corruption convictions 1991–2006	
Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
log(LIGHTNING)	−0.068	−0.027	0.019	0.016	0.344**	0.278**
	(0.183)	(0.110)	(0.198)	(0.113)	(0.145)	(0.109)
$\log(1 + CC_{1976})$	−0.388***	−0.954***				
	(0.143)	(0.162)				
$\log(POP_{1976})$		0.908***				
		(0.185)				
$\log(1 + CC_{1978})$			−0.352**	−0.930***		
			(0.170)	(0.185)		
$\log(POP_{1978})$				0.908***		
				(0.153)		
$\log(1 + CC_{1991})$					−0.452***	−0.916***
					(0.109)	(0.142)
$\log(POP_{1991})$						0.833***
						(0.147)
Constant	1.654***	−11.220***	1.469***	−11.304***	0.787**	−10.723***
	(0.313)	(2.733)	(0.358)	(2.235)	(0.347)	(2.051)
Observations	45	45	48	48	48	48
R-squared	0.15	0.49	0.10	0.44	0.29	0.54

Notes: Robust standard errors in paranthesis. Asterisks ***, ** and * signify, respectively, p-value < 0.01, p-value < 0.05 and p-value < 0.1. All variables are defined in the main text.

Source: Authors' analysis based on data described in the text.

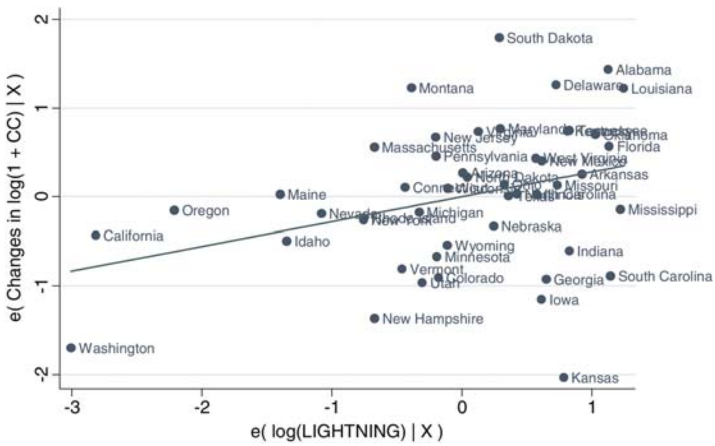
FIGURE 1. Reduced form in the pre-Internet era, U.S. sample



Note: The figure shows the association between lightning and changes in corruption convictions over the 1978–1990 period, with the influence of initial population partialled out.

Source: Figure is based on data described in the text.

FIGURE 2. Reduced form in the Internet era, U.S. sample



Note: The figure shows the association between lightning and changes in corruption convictions over the 1991–2006 period, with the influence of initial population partialled out.

Source: Figure is based on data described in the text.

virtually no difference in estimation results using either the first differenced or the lagged dependent variables specification. This is not true in the 2SLS setting.

In the first stage results are very similar across specifications. Lightning proves to be a strong instrument in explaining changes in Internet use; the first

TABLE 3. 2SLS regressions, U.S. sample

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: First stage — Dependent variable: Change in Internet users 1991–2006						
log(LIGHTNING)	−3.583***	−3.505***	−3.821***	−3.764***	−3.801***	−3.805***
	(0.604)	(0.606)	(0.756)	(0.758)	(0.651)	(0.678)
Δlog(POP)		3.443		3.947		
		(5.099)		(5.935)		
log(POP_{1991})						0.050
						(0.794)
log(1 + CC_{1991})					0.721	0.693
					(0.526)	(0.703)
Constant	60.54***	59.83***	61.41***	60.70***	59.55***	58.86***
	(1.080)	(1.336)	(1.403)	(1.547)	(1.378)	(10.870)
F-test (H0: log(lightning) = 0)	35.23	33.49	25.54	24.66	34.11	31.48
Panel B: Second stage — Dependent variable: Change in corruption convictions 1991–2006						
ΔINTERNET	−0.058	−0.056	−0.104***	−0.102***	−0.090**	−0.073**
	(0.043)	(0.046)	(0.035)	(0.037)	(0.036)	(0.030)
Δlog(POP)		−0.248		−0.606		
		(1.117)		(1.026)		
log(POP_{1991})						0.836***
						(0.146)
log(1 + CC_{1991})					−0.386***	−0.866***
					(0.097)	(0.137)
Constant	3.670	3.626	5.959***	5.928***	6.171***	−6.416**
	(2.343)	(2.413)	(1.924)	(1.924)	(1.949)	(2.708)
Observations	48	48	40	40	48	48
R-squared	0.09	0.09	0.05	0.07	0.26	0.54

Notes: Robust standard errors in paranthesis. Asterisks ***, ** and * signify, respectively, p-value < 0.01, p-value < 0.05 and p-value < 0.1. All variables are defined in the main text.

Source: Authors' analysis based on data described in the text.

stage F-statistic exceeds 10 in all columns, documenting that the 2SLS analysis in the U.S. sample is not plagued by weak identification (cf., Staiger and Stock 1997).

In the second stage results differ markedly across specifications (Panel B). In the first differenced specification Internet use is estimated imprecisely, with a point estimate virtually identical in size and sign to what we found in the OLS setting. In contrast, when we exclude either states with zero initial corruption (Columns 3 and 4), or when we control for initial corruption levels directly (Columns 5 and 6), changes in Internet use is estimated with high precision.

What should we make of the differences in size and significance of the Internet across specifications? The most straightforward explanation runs as follows: Recall from Section 1 that, under the maintained hypothesis that equation equation (1) is the true specification, the first difference equation is given by $\Delta C_i = \alpha_0 + \alpha_1^{FD} \Delta INTERNET_i + v_i$, where $v_i \equiv \alpha_2 C_{initial,i} + \varepsilon_i$. When we employ 2SLS with lightning as an instrument, the exclusion restriction is that $Cov(LIGHTNING, v) = 0$; since $Cov(LIGHTNING, \varepsilon) = 0$ by assumption, $Cov(LIGHTNING, v) = 0$ requires that $Cov(LIGHTNING, C_{initial}) = 0$. This latter condition is violated in the U.S. sample. The failure of the exclusion restriction implies that the 2SLS estimates are biased towards OLS, which explains the similarity of the results reported in Columns 1-2 in Tables 1 and 3.

In contrast, if we either eliminate the influence from initial corruption on changes in corruption by way of sample truncation (i.e., omission of states with zero initial corruption), or by including the initial level of corruption directly, the exclusion restriction is much more likely to be satisfied. The results reported in Columns 3-6 are therefore much more likely to identify the causal influence of Internet diffusion on changes in corruption. In all four columns where the exclusion restriction is plausible we find evidence that the Internet has worked to lower corruption in states where its diffusion was rapid, compared to states where Internet diffusion occurred at a slower rate.

What is the economic significance of the Internet? To answer this question we begin with the levels specification associated with equation (3):

$$\log(1 + C_{2006}) = \alpha_0 + \alpha_1 \Delta INTERNET + (\alpha_2 + 1) \log(1 + C_{1991}). \quad (6)$$

If we linearize, treating C_{1991} as a constant, the following simple approximation emerges:

$$\Delta C_{2006} \approx \alpha_1 (1 + C_{2006})\, INTERNET_{2003}, \quad (7)$$

where we have used that $\Delta INTERNET = INTERNET_{2003}$. Next, to gauge economic significance, consider moving from the median to the third quartile in the distribution of Internet users in 2003; this is equivalent to an increase of 3.1 Internet users per 100 people. Using the 2SLS results reported in Column 6 of Table 3 (the most conservative estimate) in equation (7) we find that $\Delta C_{2006} \approx (-0.073) \cdot (1 + 16) \cdot (3.1) = -3.85$ yearly convictions. This would correspond to moving from the median to the 33rd percentile in the U.S. state corruption convictions ranking in 2006. Accordingly, our results suggest that the introduction of the Internet has reduced U.S. corruption levels below what would otherwise have been observed absent this technology.

IV. Cross-Country Evidence

In this Section we examine whether the results obtained above generalize to a cross-country sample. After providing details on the data used in the cross-country setting, we discuss the partial correlation between changes in corruption and changes in Internet users, present an instrument falsification test, and discuss our 2SLS estimates.

Data

As our main measure of corruption we employ the well-known *ICRG* index. This index has the useful property that it is available from 1984, which enables us to perform the type of instrument falsification test employed in the context of the U.S. state-level analysis. The coverage for the pre-Internet era is shorter (six years, 1984-1990) than in the U.S. state-level analysis but hopefully long enough for us to clarify whether lightning is correlated with changes in corruption prior to the emergence of the World Wide Web. Changes in the *ICRG* index are calculated as the absolute change in the index between relevant years (1991-2006 and 1984-1990, respectively). The indicator is bounded between zero and six, with larger values of the index meaning *less* corruption.

As a matter of robustness we have also checked our main results (OLS and 2SLS) using the control of corruption index compiled by Kaufmann *et al.* (2007). This indicator is often used in cross-national studies (Treisman 2007). But it is unfortunately only available for the period 1996 onwards. Accordingly, the falsification test of our instrument cannot be performed with this indicator, for which reason we have chosen to stick with the *ICRG* index. But it should be noted that our OLS and 2SLS results for the Internet era are robust to the use of the control of corruption indicator; these results are available upon request.

Our key explanatory variable is the number of Internet users per 100 people. Increasingly, the number of Internet users is based on regular surveys. In situations where surveys are not available, an estimate can be derived based on the number of subscribers. Data is compiled by the International Telecommunication Union (ITU) and made available in the World Development Indicators (WDI) 2007. Changes in Internet users are calculated as in the context of our U.S. state-level analysis. When we examine the period 1991-2006, using the *ICRG* indicator, we use the approximation that $INTERNET_{i,1991} = 0$, as in the analysis of the U.S. states.

In the U.S. context data on lightning density derived from ground detectors. Naturally, such data is not available across the world. Instead we employ satellite data on lightning intensity. The raw data (strikes per km^2 per year) is provided by the National Aeronautics and Space Administration (NASA). Specifically, we rely on the data from the so-called Optical Transient Detector (OTD), a space based sensor launched on April 3, 1995. For a period of roughly 5 years the satellite orbited Earth once every 100 minutes at an altitude of 740 km. At any given instant it viewed a 1300 km × 1300 km region of Earth. Lightning is determined by comparing the luminance of adjoining frames of OTD optical data. When the difference was larger than a specified threshold value, an event was recorded.[18] These satellite-based data are archived and cataloged by the The Global Hydrology and Climate Center, where they are also made publicly available.[19] We apply the data from a high-resolution (0.5 degree latitude × 0.5 degree longitude) grid of total lightning bulk production across the planet, expressed as a flash density, from the completed 5 year OTD mission.[20] Figure 3 provides a world map of the average flash density over the 5 years period.

We construct average flash densities for each country by first mapping the corresponding geographic areas into the lightning data grid and then taking the average of flash densities within each of these areas. The coordinates describing the areas are taken from the GEOnet Names Server (GNS) at the U.S. National Geospatial-Intelligence Agency's (NGA),[21] and the U.S. Board on Geographic Names' (U.S. BGN) database of foreign geographic names and features.[22] We used the GNS database released on October 7, 2008.[23] The GNS data covers the entire planet with the exception of the U.S. and the Antarctica. The area for the U.S. was estimated using geographic features for the 48 contiguous U.S. states, contained in the database released on August 15, 2008 by the Geographic Names Information System at the U.S. BGN.[24]

A potential problem with the OTD data is that it only provides observations on *total* lightning events; i.e., intra-cloud, cloud-to-cloud, cloud-to-sky, and cloud-to-ground lightning. In other words, OTD data does not separate out *cloud-to-ground* lightning incidences. The pertinent characteristic of lightning in the evaluation of risk to electronic equipment and electric power systems is the cloud-to-ground flash density.

In an effort to examine whether the satellite data is likely to be a good proxy for ground strikes we compared our data on ground strikes for

18. Basically, these optical sensors use high-speed cameras designed to look for changes in the tops of clouds. By analyzing a narrow wavelength band (near-infrared region of the spectrum) they can spot brief lightning flashes even under daytime conditions.

19. <http://thunder.msfc.nasa.gov/data/#OTD_DATA>

20. <ftp://microwave.nsstc.nasa.gov/pub/data/lightning-satellite/lis-otd-climatology/HRFC/LISOTD_HRFC_V2.2.hdf>

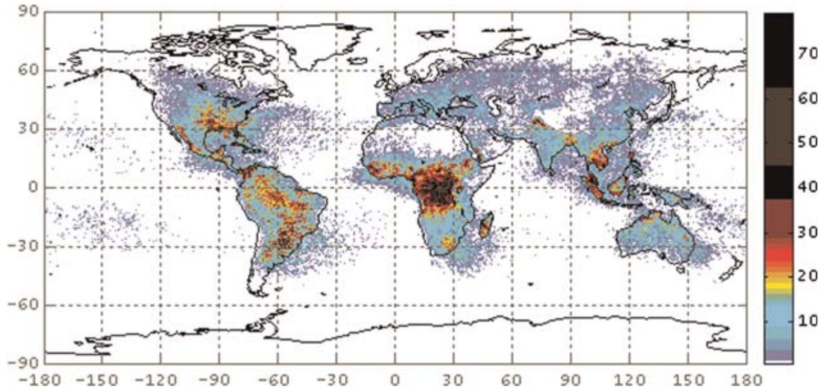
21. <http://earth-info.nga.mil/gns/html/namefiles.htm>

22. <http://geonames.usgs.gov/domestic/download_data.htm>

23. <ftp://ftp.nga.mil/pub2/gns_data/geonames_dd_dms_date_20081007.zip>

24. <http://geonames.usgs.gov/docs/stategaz/NationalFile_20080815.zip>

FIGURE 3. Global lightning map



Note: Average flash density (flashes per year per km^2). The figure is constructed using the OTD Global Lightning Distributions for the period April 12, 1995 to December 31, 1999.

Source: Figure is based on data described in the text.

individual U.S. states with corresponding satellite-based lightning data. Figure 4 provides a scatter plot of total lightning against cloud-to-ground lightning for the U.S. sample. Note that by definition total lightning should be at least as large as cloud-to-ground lightning, which is confirmed by the figure; all observations are below the 45 degree line.[25]

Figure 4 shows that there is agreement between the two measures of lightning. With a correlation above 0.95, total lightning appears to be a reasonable proxy for cloud-to-ground lightning in the contiguous U.S. states sample. This is also expected to be the case in the cross-country data (Chisholm and Cummins, 2006).

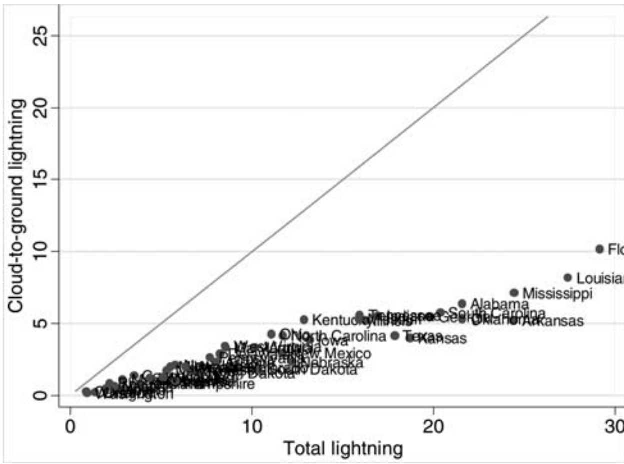
Partial Correlations

Table 4 reports the results from OLS estimation using cross-country data. As in Table 1 we experiment with various specifications: In Column 1 we use the first difference specification; in Column 2 we truncate the sample to limit the extent of mean reversion by excluding observations within one standard deviation from either endpoint of the 0-6 "*ICRG* interval"; in Column 3 we estimate the lagged dependent variables model; and finally in Column 4 we report the results from estimating our preferred model by way of an outlier robust estimator (Least Absolute Deviations, LAD).

Compared to the state-level analysis we find more variation in results when we move from the first difference specification to the lagged dependent variables specification. In Column 1 changes in Internet penetration is

25. NASA's flash densities of total lightning are calculated for the 1995-1999 period, while Vaisala's cloud-to-ground measures are calculated for the 1996-2005 period. In addition, Vaisala uses *mile*^{2} as the area unit; these where converted into *km*^{2} by dividing by the *mile*^{2} numbers by 1.609^2.

FIGURE 4. Total lightning versus cloud-to-ground lightning, U.S. sample



Note: The figure shows a scatter plot of total lightning flash rate density (flashes per km^2 per year) based on the Optical Transient Detector high resolution data (horizontal axis) versus cloud-to-ground flash rate density (flashes per km^2 per year) based on the U.S. National Lightning Detection Network (vertical axis). The full line is the 45-degree line. Number of observations is 48.

Source: Figure is based on data described in the text.

insignificantly correlated with changes in the *ICRG* index during the Internet era, whereas a positive correlation emerges when we employ the lagged dependent variables approach, thus suggesting a corruption suppressing influence of the Internet. It is worth observing that the initial *ICRG* level is highly significant and contributes considerably to the overall fit, as is evident from the increase in R^2 going from Column 1 to 3. The results from Column 2, where we truncate the sample, are somewhere in between. But even when we truncate the sample markedly to suppress mean reversion, the initial corruption level, *ICRG* in 1991, remains significant if included (not shown but results are available upon request); the partial correlation is weaker, as it should be in theory, but it is not rendered insignificant.

These results demonstrate that the evolution of corruption is subject to mean reversion in the world sample; perhaps even to a greater extent than in the U.S. sample, since even sample truncation is not sufficient to fully eliminate the correlation between initial corruption and subsequent changes in corruption. Now using both mean reversion ($\alpha_2 < 0$) and the fact that a high *ICRG* score in 1991 (low initial corruption) is associated with faster Internet diffusion (i.e., $Cov(C_{initial}, \Delta INTERNET) > 0$) in equation (2) immediately tells us that the first difference model is misspecified in the world sample; that is, since plim $\alpha_1^{FD} < \alpha_1$ the first difference model underestimates α_1. This is of course exactly what we observe in Table 4: the parameter estimate for $\Delta INTERNET$ rises monotonically when we go from Column 1 to 3.

TABLE 4. OLS and LAD regressions, world sample

Dependent variable:	Change in ICRG 1991–2005			
	(1)	(2)	(3)	(4)
ΔINTERNET	0.004	0.0138*	0.0361***	0.0352***
	(−0.003)	(−0.007)	(−0.006)	(−0.006)
$ICRG_{1991}$			−0.747***	−0.772***
			(−0.089)	(−0.097)
Constant	−0.923***	−1.044***	0.981***	1.072***
	(−0.140)	(−0.145)	(−0.235)	(−0.281)
Observations	102	67	102	102
R-squared	0.01	0.04	0.57	
Estimator	OLS	OLS	OLS	LAD

Notes: Robust standard errors in paranthesis. Asterisks ***, ** and * signify, respectively, p-value < 0.01, p-value < 0.05 and p-value < 0.1. In column 2 the sample is truncated; only countries with ICRG scores one standard deviation away from the ICRG index minimum (0) and maximum (6) value are included. Larger values for the ICRG index means *less* corruption. In column 4 the standard errors are bootstrapped with 1000 repetitions.

Source: Authors' analysis based on data described in the text.

There is a straightforward explanation of why the strong correlation between initial corruption and Internet diffusion, which we observe in the world sample, has no parallel in the U.S. sample. In the world sample, factors influencing the level of corruption, or the level of corruption *per se*, probably also influence the process of Internet diffusion; perhaps the over-all institutional quality of a nation influences not only corruption and its evolution, but also the speed of technology adoption. This seems plausible, and it can account for the positive correlation between the *ICRG* index and Internet diffusion in the world sample. It appears reasonable to expect that similar forces are *not* at play in the U.S. sample of relatively more homogenous states, for which reason a similar correlation between state-level corruption and state-level Internet use is absent. Therefore, in the world sample, but not in the U.S. sample, the end result of (erroneously) omitting the level of initial corruption is that the OLS estimate for Δ*INTERNET* is biased towards zero. The omitted variables bias is dampened when we truncate the sample (Column 2), as the correlation between the level of corruption and changes in corruption is reduced; the parameter estimate for Δ*INTERNET* therefore rises in absolute value. In sum, the lagged dependent variables specification should be considered the appropriate empirical model.

In an effort to check the robustness of the correlation we also report, in Table 4, the results from estimating our preferred model by way of an outlier robust estimator. As can be seen, comparing Column 3 and 4, the results are virtually the same whether we use OLS or LAD, implying that outliers are not carrying the correlation. This conclusion is further reinforced by invoking the Hadi (1992) outlier detection procedure; no observations are found to represent (multivariate) outliers.

As another check we have examined the robustness of the $\Delta INTERNET/\Delta ICRG$ correlation to the inclusion of the corruption determinants discussed in Treisman's (2007) survey. Treisman includes variables that constitute historical and controls; political controls; and finally a set of rents and competition controls. The correlation between $\Delta INTERNET$ and $\Delta ICRG$ remains when these variables are included in the lagged dependent variables model (one-by-one). Finally, the above results carry over if we employ the control of corruption indicator (Kaufmann *et al.*, 2007), instead of the *ICRG* index. These results are available upon request.

Instrument Falsification Test

Following the approach taken in the U.S. state-level analysis our 2SLS analysis on the World sample is preceded by an instrument falsification test.

Accordingly, Table 5 reports the results from estimating the reduced form on different time periods: pre-Internet (1984-1990) and post-Internet (1991-2005). As explained above, the pre-Internet period is limited to six years, due to lack of availability of the *ICRG* index beyond this. We report three sets of results on the two sub-periods: full sample OLS; OLS on a sample pruned for outliers detected by the Hadi (1992) method; full sample and outlier robust estimator (LAD).

Turning to the results, we find a statistically significant correlation between changes in *ICRG* and lightning during the Internet era (i.e., 1991 onwards). The partial correlation between lightning and changes in corruption reported in Column 1 can be inspected visually in Figure 5.

It is immediately obvious from the figure that Iceland (ISL) have considerable influence on the significance of lightning; i.e., Iceland may represent an outlier. Using the Hadi (1992) procedure, we confirm that Iceland is indeed an outlier; and the only one.

When we re-estimate the model, omitting Iceland (Column 2), we find essentially the same partial correlation as in the full sample. This remains true if we, rather than omitting Iceland from the sample, estimate the model by way of an outlier robust estimator (Column 3). The LAD point estimate is close to the OLS estimates. Hence, as in the U.S. sample, we find evidence that lightning prone regions have experienced lower reductions in corruption during the Internet era, consistent with the theory underlying the instrument.

In Columns 4-6 we turn to the pre-Internet era. At first sight the test seems to falsify the instrument as we detect a significant correlation between lightning and changes in *ICRG* during the pre-Internet era in Column 4. Taken at face value this means lightning is picking up forces that exert a time persistent influence on the evolution of corruption, which is inconsistent with the theory that lightning captures (solely) Internet diffusion. The OLS based partial correlation

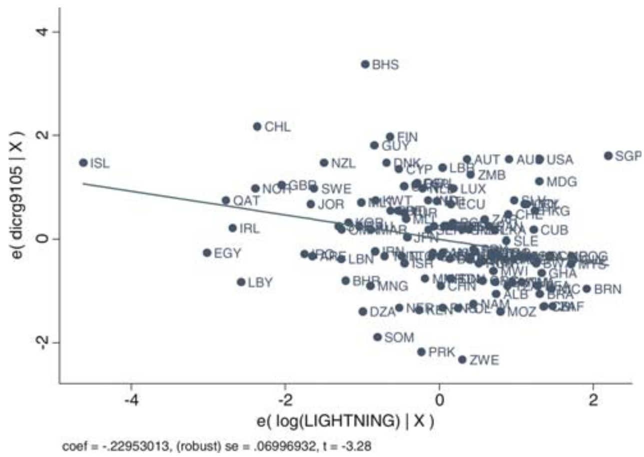
TABLE 5. Falsification test of instrument, world sample

	Change in ICRG 1991–2005			Change in ICRG 1984–90		
Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
log(LIGHTNING)	−0.230***	−0.216***	−0.295***	−0.0931*	−0.0938	−0.0164
	(−0.070)	(−0.0785)	(−0.0777)	(−0.0527)	(−0.0619)	(−0.0588)
ICRG, 1991	−0.531***	−0.530***	−0.565***			
	(−0.0825)	(−0.0825)	(−0.0952)			
ICRG,1984				−0.232***	−0.232***	−0.0267
				(−0.0492)	(−0.0494)	(−0.101)
Constant	1.362***	1.331***	1.517***	1.040***	1.042***	0.132
	(−0.377)	(−0.385)	(−0.387)	(−0.244)	(−0.256)	(−0.469)
Observations	124	123	124	99	98	99
R-squared	0.36	0.36		0.16	0.16	
Estimator	OLS	OLS no outlier	LAD	OLS	OLS no outlier	LAD

Notes: Robust standard errors in paranthesis. Asterisks ***, ** and * signify, respectively, p-value < 0.01, p-value < 0.05 and p-value < 0.1. Column 2 and 4 omit an outlier (Iceland) detected using the Hadi (1992) procedure. Larger values for the ICRG index means *less* corruption. In columns 3 and 6 the standard errors are bootstrapped with 1000 repetitions.

Source: Authors' analysis based on data described in the text.

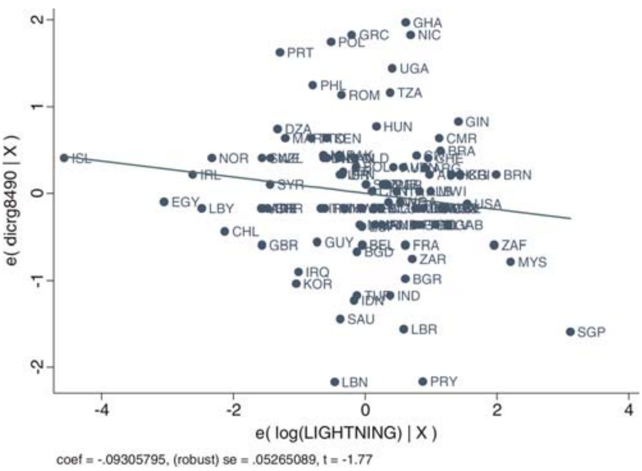
FIGURE 5. Reduced form in the Internet era, world sample



Note: The figure shows the association between lightning and changes in the ICRG index over the 1991–2005 period, with ICRG in 1991 partialled out.

Source: Figure is based on data described in the text.

FIGURE 6. Reduced form in the pre-Internet era, world sample



Note: The figure shows the association between lightning and changes in the ICRG index over the 1984–1990 period, with ICRG in 1984 partialled out.

Source: Figure is based on data described in the text.

between lightning and changes in the *ICRG* index from Column 4 is illustrated in Figure 6.

Visual inspection of the figure reveals that Iceland once again appears to represent an outlier, possibly even explaining the significance of lightning during the pre-Internet period. Employing the Hadi (1992) procedure one can confirm that Iceland indeed is a multivariate outlier; it is the sole observation (as in the post-Internet setting) that the Hadi (1992) test singles out.

When we subsequently re-estimate the model, omitting Iceland, the significance of lightning evaporates (Column 5). The fragility of the lightning /$\Delta ICRG$ correlation for the 1984-90 period is further illustrated by the LAD regression on the full sample (Column 6): In this setting, the size of the estimate is reduced considerably compared to the OLS result, and is far from being statistically significant at conventional levels.

Overall, these falsification tests are reassuring. When the outlier is omitted, or an estimation method that is robust to extreme observations is invoked, there is no statistically significant correlation between lightning and changes in corruption prior to the emergence of the World Wide Web. Moreover, moving beyond statistical significance, it is clear that lightning exhibits a much stronger correlation with changes in corruption during the post-Internet era, compared to the pre-Internet era. As can be seen from Table 5, the OLS estimate in Column 1 exceed that of Column 4 by no less than a factor of 2.5.

In sum, we find the evidence reported in Table 5 sufficiently supportive of the identification strategy (albeit perhaps less strong than in the U.S. case) so as to allow us to proceed to 2SLS estimation using lightning intensity as an instrument for Internet diffusion.

2SLS Estimates

Table 6 reports the 2SLS estimates of the impact from the Internet on the evolution of corruption around the World. Panel A reports the first stage, whereas Panel B displays second stage results. In keeping with the approach taken so far, we report results from three separate exercises: full sample, first difference specification (Column 1); truncated sample, first difference specification (Column 2); and our preferred lagged dependent variables specification on the full sample (Column 3).

Broadly speaking, the results mirror those attained using U.S. state-level data. First, the lightning instrument is statistically strong in the full sample of countries (Columns 1 and 3); the first stage features F-statistics well above ten. Second, the first difference specification produces insignificant estimates for $\Delta INTERNET$ in the second stage, whereas the lagged dependent variables specification leads to a statistically significant point estimate for $\Delta INTERNET$.

Our interpretation of this variation in results across specifications is the same as in the case of the U.S. state-level analysis. Changes in corruption display mean reversion, for which reason the initial level of corruption needs to be controlled for in the regression. When it is erroneously omitted, it jeopardizes identification since lightning and the level of corruption are correlated. A glance on the lightning map (Figure 3) above is enough to realize that high lightning intensity is more prevalent in the poorer regions of the World, near the equator, where countries also tend to be characterized by relatively poor

TABLE 6. 2SLS regressions, world sample

	Panel A: First stage		
	Dependent variable: ΔINTERNET		
	(1)	(2)	(3)
log(LIGHTNING)	−10.619***	−4.4499**	−6.9537***
	(1.2479)	(1.5087)	(1.2863)
$ICRG_{1991}$			6.5275***
			(1.1915)
Constant	38.985***	20.899***	10.252*
	(2.7145)	(3.4299)	(5.7635)
F-test (H0: log(LIGHTNING) = 0)	71.62	7.76	25.08
	Panel B: Second stage		
	Dependent variable: Change in ICRG 1991–2005		
ΔINTERNET	0.00011	0.0216	0.0435***
	(−0.006)	(−0.0285)	(−0.010)
$ICRG_{1991}$			−0.820***
			(−0.114)
Constant	−0.841***	−1.137***	1.079***
	(−0.176)	(−0.329)	(−0.244)
Observations	102	67	102

Notes: Robust standard errors in paranthesis. Asterisks ***, ** and * signify, respectively, p-value < 0.01, p-value < 0.05 and p-value < 0.1. In column 2 the sample is truncated; only countries with ICRG scores one standard deviation away from the ICRG index minimum (0) and maximum (6) value are included. Larger values for the ICRG index means *less* corruption.

Source: Authors' analysis based on data described in the text.

governance. Accordingly, identification is only plausible in the case of the lagged dependent variables specification.

The results from Column 3 of Table 6 corroborate our OLS findings (Table 4): From the early 1990s onward the Internet has worked to suppress corruption. Hence, it would appear that our main results, which pertain to the U.S., carry over to a cross-country setting. In closing, it should be noted that results are similar when we use the Kaufmann *et al.* (2007) corruption indicator instead of the *ICRG* data; these results are available upon request.

V. CONCLUDING REMARKS

In the present paper we have examined the hypothesis that technological change may cause improvements in governance. Specifically, we have studied the influence from the Internet on the evolution of corruption across U.S. states and across the World. The results are, statistically speaking, perhaps somewhat stronger in the U.S. setting but in both samples we find evidence that the Internet has worked to suppress corruption since its emergence.

Our results should be interpreted with some care: they are reduced form estimates of the total impact from the Internet on corruption. The analysis therefore does not speak to the mechanisms linking the Internet to corruption. As observed in the Introduction, the Internet is a General Purpose Technology, for which reason it probably influences the economy in a great many ways; growth, human capital accumulation, the dissemination of information, and E-government are the most obvious intervening factors that are both associated with the Internet and impinge on the level of corruption in a state or country. The task of disentangling these separate pathways of influence from the Internet technology to corruption is left open for future research.

The identification strategy developed in the present paper may be useful in other contexts. With a plausibly exogenous instrument for the Internet, researchers may potentially make new progress in the study of the impact from the Internet on other outcomes, such as the return to skills or productivity growth more broadly.

Appendix

Table A.1. Summary statistics

	Panel A: Cross-country sample					
	Obs.	Mean	Median	Std. Dev.	Min	Max
ICRG 2005	102	2.60	2.14	1.30	0	6
ICRG 1991	102	3.44	3	1.46	0	6
INTERNET 2005	113	22.56	15.24	21.66	0.21	86.94
Average lightning density	113	9.00	6.59	8.62	0.02	44.38
Panel B: U.S. sample						
CC 2006	48	20.15	9.50	21.48	0	83
CC 1991	48	13.44	6.00	21.14	0	108
INTERNET 2003	48	54.39	55.00	5.88	39.50	65.50
POP 1991	48	52.22	36.46	56.16	4.59	305.00
Average lightning density	48	10.54	8.58	7.31	0.89	27.34

Source: Authors' calculations based on data described in the text.

References

Acemoglu, Daron, and James Robinson, 2001. Political Losers as a Barrier to Economic Development. *American Economic Review*, 90, 126–130.

Ades, Alberto, and Rafael di Tella, 1999. Rents, Competition and Corruption. *American Economic Review*, 89, 982–993.

Andersen, Thomas B., 2009. E-government as an anti-corruption strategy. *Information Economics and Policy*, 21, 201–210.

Andersen, Thomas B., Jeanet Bentzen, Carl-Johan Dalgaard, and Pablo Selaya, 2010. Lightning, IT Diffusion, and Economic Growth across U.S. States. *Review of Economics and Statistics*, forthcoming.

Angrist, Joshua, and Stefan Pischke, 2009. Mostly Harmless Econometrics, Princeton University Press.

Besley, Tim, and Robin Burgess, 2002. The Political Economy of Government Responsiveness: Theory and Evidence from India. *Quarterly Journal of Economics*, CXVII, 1415–1451.

Bhatnagar, Subhash, 2003. E-government and access to information. *Global Corruption Report 2003*, Transparency International.

BSI, 2004. Bundesamt für Sicherheit in der Informationstechnik, IT-Grundschutz Manual. Available online at <http://www.bsi.de/english/gshb/index.htm>.

Brunetti, Aymo, and Beatrice Weder, 2003. A free press is bad news for corruption. *Journal of Public Economics*, 87, 1801–1824.

Caselli, Francesco, and Wilbur J. Coleman III, 2001. Cross-Country Technology Diffusion: The Case of Computers. *American Economic Review*, 91, 328–335.

Chinn, Menzie D., and Robert W. Fairlie, 2007. The determinants of the global digital divide: a cross-country analysis of computer and Internet penetration. *Oxford Economic Papers*, 59, 16–44.

Chisholm, William A., 2000. Lightning Protection. Chapter 4.10 in: The Electrical Power Engineering Handbook. *Edited by* L Grigsby., IEEE Press.

Chisholm, William A., and Kenneth Cummins, 2006. On the Use of LIS/OTD Flash Density in Electric Utility Reliability Analysis. Paper presented at *The Lightning Imaging Sensor International Workshop*. Available online at <http://lis11.nsstc.uah.edu/Papers/Proceedings.html>.

Deaton, Angus, 2010. Instruments, Randomization, and Learning about Development, *Journal of Economic Literature*, 48, 424–455.

Eisensee, Thomas, and David Stömberg, 2007. New Droughts, New Floods and U.S. Disaster Relief. *Quarterly Journal of Economics*, 122, 693–728.

Electricity Power Research Institute, 2003. Electricity Technology Roadmap: 2003 Summary and Syntheses. Available online at <http://www.iea.org/Textbase/work/2004/distribution/presentations/Gellings_bkgd_%20paper.pdf>.

Ferraz, Claudio, and Frederico Finan, 2008. Exposing Corrupt Politicians: The Effect of Brazil's Publicly Released Audits on Electoral Outcomes. *Quarterly Journal of Economics*, 123, 703–745.

Glaeser, Edward L., and Raven E. Saks, 2006. Corruption in America. *Journal of Public Economics*, 90, 1053–1072.

Foster, Andrew D., and Mark R. Rosenzweig, 1996. Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution. *American Economic Review*, 86, p. 931–53.

Fredriksson, Per G., Herman R. J. Vollebergh, and Elbert Dijkgraaf, 2004. Corruption and energy efficiency in OECD countries: theory and evidence. *Journal of Environmental Economics and Management*, 47, 207–231.

Gundlach, Erich, and Martin Paldam, 2009. The transition of corruption: From poverty to honesty. *Economics Letters*, 103, p.146–148.

Hadi, Ali S, 1992. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society* 54, 761–771.

Hall, Robert E., and Dale Jorgenson, 1967. Tax Policy and Investment Behavior. *American Economic Review*, 57, 391–414.

IEEE, 2005. *How to Protect Your House and Its Contents From Lightning*. IEEE Guide for Surge Protection of Equipment Connected to AC Power and Communication Circuits. Standards Information Network, IEEE Press.

Jovanovic, Boyan, and Peter L. Rousseau, 2005. General Purpose Technologies. In P. Aghion, and S. Durlauf (eds.) *Handbook of Economics Growth*, Elsevier, North-Holland.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, 2007. Governance Matters IV: Governance Indicators for 1996–2006. Available online at <http://www.worldbank.org/wbi/governance/wp-governance.html>.

Kressel, Henry, 2007. *Competing for the Future: How Digital Innovations are Changing the World*, Cambridge University Press.

Licht, Amir N., Chanan Goldschmidt, and Shalom H. Schwartz, 2007. Culture Rules: The foundations of the rule of law and other norms of governance. *Journal of Comparative Economics*, 35, 659–88.

Mauro, Paolo. 1995. Corruption and Growth. *Quarterly Journal of Economics* 110, 681–712.

Maxwell, Amanda E., and Richard F. Winters, 2005. Political Corruption in America. Working Paper (Dartmouth College).

Mokyr, Joel, 1990. *The Lever of Riches: Technological Creativity and Economic Progress*, Oxford University Press.

NASA, 2007. Lightning Detection From Space: A Lightning Primer. Available online at <http://thunder.msfc.nasa.gov/primer>.

National Energy Technology Laboratory, 2003. The Value of Electricity When It's Not Available. Available online at <http://www.netl.doe.gov/moderngrid/docs/The_Value_of_Electricity_When_It's_Not_Available.pdf>.

Parente, Stephen L., and Edward C. Prescott, 1999. Monopoly Rights: A Barrier to Riches. *American Economic Review*, 1216–1233.

Persson, Torsten, Guido Tabellini, and Francesco Trebbi, 2003. Electoral rules and corruption. *Journal of the European Economic Association*, 1, 958–989.

Reinikka, Ritva, and Jakob Svensson, 2004. The power of information: Evidence from a newspaper campaign to reduce capture. Policy Research Working Paper No. 3239, World Bank.

Schultz, Theodore W., 1975. The Value of the Ability to Deal with Disequilibria. *Journal of Economic Literature*, 13, p. 827–46.

Shim, Jae K., Anique Qureshi, and Joel G. Siegel, 2000. *The International Handbook of Computer Security*. Glenlake Publishing Company Ltd.

Staiger, Douglas, and James H. Stock, 1997. Instrumental variables with Weak Instruments. *Econometrica*, 65, 557–586.

Strömberg, David, 2001. Mass Media and Public Policy, *European Economic Review*, 45, 652–663.

———, 2004. Radio's Impact on Public Spending, *Quarterly Journal of Economics*, 119, 189–221.

Svensson, Jakob., 2005. Eight Questions about Corruption. *Journal of Economic Perspectives*, 19, 19–42.

Treisman, Daniel, 2000. The causes of corruption: a cross-national study. *Journal of Public Economics*, 76, 399–457.

———, 2007. What Have We Learned About the Causes of Corruption From Ten Years of Cross-National Empirical Research? *Annual Review of Political Science*, 10, 211–244.

Yeager, Kurt, and Karl Stahlkopf, 2000. *Power for a Digital Society*. Prepared for the Department of Energy CF-170/1-1-DOE. Available online at <http://www.rand.org/scitech/stpi/Evision/Supplement/yeager.pdf>.

WDI, 2007. *World Development Indicators 2007*, World Bank.

West, Darrell M., 2005. *Digital Government*, Princeton University Press.

———, 2008. *Improving Technology Utilization in Electronic Government around the World*. Brookings Governance Studies.

Wooldridge, Jeffrey., 2003. *Introductory Econometrics: A Modern Approach*, 3E, South–Western.

World Development Report 2004. *Making Services Work for the Poor*, The World Bank.