On the Impact of Digital Technologies on Corruption

Andersen, Thomas Barnebeck; Bentzen, Jeanet Sinding; Dalgaard, Carl-Johan Lars; Selaya, Pablo

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Andersen, T. B., Bentzen, J., Dalgaard, C-J. L., & Selaya, P. (2008). On the Impact of Digital Technologies on Corruption: Evidence from U.S. States and Across Countries. Department of Economics, University of Copenhagen.

Discussion Papers
Department of Economics
University of Copenhagen

No. 08-11

On the Impact of Digital Technologies on Corruption: Evidence from U.S. States and Across Countries

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

Studiestræde 6, DK-1455 Copenhagen K., Denmark
Tel.: +45 35 32 30 82 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (online)

On the Impact of Digital Technologies on Corruption: Evidence from U.S. States and Across Countries*

Thomas Barnebeck Andersen Jeanet Bentzen Carl-Johan Dalgaard
Pablo Selaya†

April 23, 2008

Abstract

We hypothesize that the spread of the Internet has reduced corruption, chiefly through two mechanisms. First, the Internet facilitates the dissemination of information about corrupt behavior, which raises the detection risks to shady bureaucrats and politicians. Second, the Internet has reduced the interface between bureaucrats and the public. Using cross-country data and data for the U.S. states, we test this hypothesis. Data spans the period during which the Internet has been in operation. In order to address the potential endogeneity problem, we develop a novel identification strategy for Internet diffusion. Digital equipment is highly sensitive to power disruption: it leads to equipment failure and damage. Even very short disruptions (less than 1/60th of a second) can have such consequences. Accordingly, more frequent power failures will increase the user cost of IT capital; either directly, through depreciation, or indirectly, through the costs of protective devises. *Ceteris paribus*, we expect that higher IT user costs will lower the speed of Internet diffusion. A natural phenomenon which causes a major part of annual power disruptions globally is lightning activity. Lightning therefore provides exogenous variation in the user cost of IT capital. Based on global satellite data from the U.S. National Aeronautics and Space Administration (NASA), we construct lightning density data for a large cross section of countries and for the U.S. states. We demonstrate that the lightning density variable is a strong instrument for *changes* in Internet penetration; and we proceed to show that the spread of the Internet has reduced the extent of corruption across the globe and across the U.S. The size of the impact is economically and statistically significant.

Keywords: Public corruption; Internet; Information
JEL Classification codes: K4; O1; H0

*We owe a special thanks to William A. Chisholm for generously sharing his expertise on the subject of lightning protection. We also thank Phil Abbot, Shekhar Aiyar, Christian Bjørnskov, Areendam Chanda, Phil Keefer, Sam Jones, Norman Loayza, Ola Olsson, Elena Paltseva, Nadja Wirz and seminar participants at the University of Copenhagen and the World Bank. Errors are ours.

†All the authors are affiliated with the Department of Economics, University of Copenhagen, Studiestraede 6, DK-1455 Copenhagen K, Denmark. Contact: Thomas B. Andersen (thomas.barnebeck.andersen@econ.ku.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).

1 Introduction

Corruption is commonly perceived to be a major stumbling block for many poor countries in their struggle to prosper. Aside from retarding growth (Mauro, 1995), corruption entails "fiscal leakage", which reduces the ability of poor countries to supply public services like schooling and health care (e.g., Reinikka and Svensson, 2004). As a consequence, it is a government failure one would like to dispose of.

However, combating corruption is not an easy task. In theory, it is possible to offer government officials high powered contracts, which would make it too expensive to be caught taking bribes for it to be worth the risk. In practice, such course of action is rarely feasible in many poor countries due to its high cost. The marginal costs of being corrupt, as percieved by the civil servant, can also be increased by elevating the risk of detection. Unfortunately, this too can be a costly route to take as it requires considerable expenses to run an effective judicial system. More importantly, it may be difficult to find incorruptible officials to track down the corrupt ones. If the controllers are corrupt the consequences for bureaucrats of taking bribes are reduced, since the latter may avoid prosecution by bribing the former. In theory, this strategic complementarity in corruption leads to multiple equilibria, and it motivates why the corruption scourge can be hard to combat.[1]

The present paper provides evidence showing that the spread of the Internet has reduced the extent of corruption during the 1990s and early years of the 21st century.[2] A major reason why the Internet has had such an effect is very likely to be found in its role as a provider of information. By providing information about corrupt officials the risk of detection rises. Likewise, a more freely flowing stream of information may also increase the risks to politicians of taking bribes; if the public is informed about corrupt behavior the re-election chances of shady politicians tend to diminish (Ferraz and Finan, 2008). There is by now considerable informal evidence to suggest that the Internet serves such purposes.

The New York Times (May 24, 2005) reports on the self-appointed Chinese investigative journalist, Li Xinde, who maintains an anti-corruption Web site targeting official corruption in China.[3] This journalist travels around China with a laptop and a digital camera investigating official wrongdoing. He then writes about it on his Web site, making sure to leave town before local authorities can arrest him. The Web site has been instrumental in exposing a corruption case involving the deputy mayor of Jining, a large Chinese city. The site featured an investigative report and a series of photos showing the deputy mayor kneeling and crying, begging not to be reported to the police. The deputy mayor was subsequently arrested.[4]

[1] See Bardhan (1997) and Aidt (2003) for very useful surveys.

[2] Technically there is a distinction to be made between the Internet and the World Wide Web (WWW). The latter was launched in 1991 by CERN (the European Organisation for Nuclear Research), whereas the history of "the Internet" arguably is much older. 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.

[3] "Death by a Thousand Blogs" by Nicholas D. Kristof. The Washington Post (May 2, 2007) also ran a (different) story featuring Li Xinde exposing corruption in China.

[4] Another fascinating story from China concerns the village of Shengyou in Hebei Province, which saw a clash between village peasants and numerous armed thugs sent by property developers to grab their land. A video smuggled out by one of the villagers shows his fellow residents being beaten with staves and shovels. Six villagers were killed and around 50 were wounded. With copies of the video circulating the Internet, the authorities reacted promptly: A mayor and a Communist Party chief of the municipality to which the village belongs were sacked. The official media reported that 22 people had been arrested, including the bosses of a firm contracted by a local state-owned power plant to build a waste-processing plant on the village fields. See "Turning ploughshares into staves", The Economist (June 23, 2005).

A few years ago the Indian news Web site <http://www.Tehelka.com> nearly toppled the Indian government after documenting high-level corruption. The reporters posed as arms dealers and documented negotiations with top politicians and bureaucrats over the size of required side payments to get the order: 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.[5]

In 2004 judge John Gomery was heading a commission investigating a kickback program in which the Canadian Liberal government had given C$ 85 million to Montreal-based advertising firms. The funds were intended to be used for publicizing certain government programs, but were apparently instead funneled to political allies. While the hearings were public, the judge banned the official media from publishing reports on what was being uncovered until a final ruling was made. The official media respected the ban but a Canadian blogger was able to bypass it by collaborating with a U.S. based Web site <http://www.captainsquartersblog.com>.[6] The ensuing public anger from what was disseminated through the Web site probably cost the Liberals their parliamentary majority in the June 2004 election.[7]

The Internet has likely also worked to reduce corruption through a slightly different channel. That is, the WWW is the main vehicle for the provision e-government worldwide; in the U.S., for instance, more than 80% of all e-government is Internet based (West, 2005).[8] E-government can reduce corruption in two principal ways. First, by allowing citizens to directly access government services online, e-government obviates bureaucrats' role as intermediaries between the government and the public, thus reducing the number of interactions between potentially corrupt officials and the public. Second, by "digitalizing" the provision of government services, e-government reduces bureaucratic discretion (thereby curbing some opportunities for arbitrary action) and it increases transparency, which ultimately leads to more accountability. In particular, e-government may elevate chances of exposure by keeping detailed data on transactions, making it possible to track and link the corrupt with their wrongful acts.

The potential importance of e-government is also backed by much informal evidence (see Wescott, 2003; Global Corruption Report, 2003). In Pakistan, for instance, the entire tax department has introduced ICT systems with the stated purpose of reducing contact between tax collectors and tax payers. In the Philippines, the Department of Budget and Management has established an on-line e-procurement system that allows public bidding for suppliers. This system has increased transparency in transactions. In South Korea, the Online Procedures Enhancement for Civil Applications allows ordinary citizens to monitor applications for permits or approvals where corruption is most likely to take place; it also allows questions to be raised in case irregularities are detected. In the Indian state of Andhra Pradesh, where 40% of its 76 million people cannot read, 214 deed registration offices have been fully computerized. This has made the process of deed registration easy and transparent. The process started in April 1998 and by February 2000 about 700,000 documents had been registered. Before the introduction of online registration, opaqueness of procedures forced citizens to employ middlemen who used corrupt practices to obtain services. In several Asian countries,

[5] See "The Sting That Has India Writhing" by Celia W. Dugger, The New York Times (March 16, 2001).

[6] See "Shivering Mr Martin's timbers", The Economist (April 7, 2004).

[7] See also "Did blogosphere influence vote? Corruption inquiry covered only on Web might have tipped Canadian election" by David Kopel, The Rocky Mountain News (January 28, 2006).

[8] In Britain, *Directgov*, an official Webpage launced in 2004, aims to contain the entire British government in one place, <http://www.direct.gov.uk/en/index.htm>.

governments are introducing smart cards that help citizens access health-care services without having to provide corruption-prone cash payments for these services.[9]

The contribution of the present paper is to move beyond the anecdotal evidence and estimate the impact of the spread of the Internet (i.e. the *change* in Internet users) on *changes* in corruption levels from the 1990s to 2006. Using cross-country data as well as data for U.S. states we find that increasing Internet penetration lowers corruption. The impact is both economically and statistically significant.

In the regressions we control for the initial level of corruption when examining subsequent changes in the extent of corruption. By implication, the OLS estimates of the impact of changes in Internet use on changes in corruption will not be biased by time invariant determinants of the level of corruption, such as resource endowments, the extent of ethnic and linguistic fractionalization, etc. This leaves out time-varying factors, which could affect both the speed of Internet diffusion and corruption.

The fact that we are able to confirm our principal finding from the cross-country setting in the sample of U.S. states provides some assurance that results are not tainted by omitted variable bias; after all, the 50 American states represent a very homogenous sample. In addition, by moving to the U.S. sample we are able to test the impact of the Internet on corruption using better data. For example, when we focus on U.S. states, we do not have to rely on *indirect* data based on corruption perceptions. Instead, we employ *direct* data on the number of federal convictions as a measure of the intensity of corruption.

In addition, we develop an instrumental variables approach to the issue at hand. Our identification strategy is to find exogenous variation in the costs of using IT equipment in general, and the Internet in particular. As explained in detail below, computer equipment is highly sensitive to power disruptions: power surges can lead to equipment failure and damage. A higher frequency of power disruptions thereby leads to increasing costs of IT equipment, either by elevating capital depreciation or as a result of additional costs needed to be borne in order to protect equipment from power disruptions. A natural phenomenon which produces power surges is lightning activity. In the U.S. about one third of all power disruptions are related to lightning activity: in 1997 more than 100,000 computers were destroyed in the U.S. alone as a consequence of lightning activity. We therefore hypothesize that higher lightning intensity increases the costs of using IT equipment and thus lowers the speed of Internet diffusion, *ceteris paribus*.

Using global satellite data assembled by the U.S. National Aeronautics and Space Administration (NASA), we construct country and state level measures of lightning density. We then document, using cross-country data, that greater lightning density is associated with more frequent power distruptions per year, which in turn is associated with a slower speed of Internet diffusion. These findings support a causal link between lightning density and the speed of Internet penetration, making the former a viable candidate instrument for the latter. Indeed, lightning density turns out to be a strong instrument for the speed of Internet diffusion

[9] Other examples of e-government include Cristal in Argentina, a Web site aiming at disseminating online information concerning the use of public funds; the Central Vigilance Commission Web site in India, where the public among other things can report information about wrongdoings of public servants; an on-line Customs Bureau system in the Philippines, which has lessened the cost of trade for businesses, reduced opportunities for fraud and boosted revenue collection of the Customs Bureau; and several computerized interstate check posts in Gujarat, India, which has significantly reduced corruption at check posts. See <http://www1.worldbank.org/publicsector/egov/anticorruption&t.htm> for more information on these and other initiatives.

from 1996 to 2006 in the world at large, and in the U.S. over the period 1991-2006.[10] Our IV estimates confirm the OLS results: rising Internet use over the 1990s reduced corruption.

The present paper 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 et al. (2003), Glaeser and Saks (2006), and Licht et al. (2007). Conceptually, Brunetti and Weder (2003) is perhaps the closest precursor to the present paper. The authors find a corruption-reducing impact of a free press, and argue that this shows that an independent press works to increase transparency. In the present case, we expect the Internet to affect corruption for essentially the same reason. Evidence to the importance of mass media, in the context of combating corruption, is provided in the interesting study by McMillan and Zoido (2004). In the 1990s, Peru was *de facto* run by Vladimiro Montesinos Torres, chief of the secret police. In the course of exercising power, Montesinos bribed judges, politicians, and the news media. He kept records of these transactions in the shape of signed contracts, receipts, and videotapes. Invoking a revealed-preference type argument, McMillan and Zoido take the size of bribes as indicating just how much importance Montesinos attached to those who were in a position to check his power. Interestingly, the records of the various bribes paid during the years 1998 to 2000 reveal that television was paid US$ 3 million per month in bribes, approximately 10 times as much as payments to politicians and judges, respectively.[11] Arguably, this case study also hints at a potentially important difference between the traditional news media and the Internet: the latter is much harder to control by rulers. In fact, a famous metaphor compares trying to regulate the Internet to trying to "nail Jell-O to the wall".[12] This is not to deny that governments cannot make life difficult for Internet users. For instance, regulations in China require the "owner" of a blog to register with the government.[13] Nevertheless, regulation of this kind is far from perfect since users can take countermeasures.[14]

The political economy literature, which studies the impact of information on governance more generally, is also related. This literature broadly argues that a better informed public serves to discipline the political establishment, thus affecting governance. Besley and Burgess (2002), for example, provide a theoretical model which predicts that the effort level of an incumbent politician is increasing in media access and voter turnout. Using Indian data covering the 1958-2002 period, Besley and Burgess show that higher levels of newspaper circulation is associated with a higher degree of "responsiveness" by the state government, measured by how the individual states responded to food shortages via the public distribution of food.[15] Similarly,

[10] Our preferred corruption measure for the cross-country sample only goes back to 1996, explaining the shorter cross-country time span. We do, however, demonstrate that results reported in the paper also hold using an alternative corruption measure, which dates back to 1991.

[11] McMillan and Zoido stress that owners of television simply have more bargaining power than judges and politicians: Only a majority of judges or politicians need to be bribed to control the Supreme Court or to form a winning coalition, while all television channels have to be bribed in order to contain a story.

[12] "The party, the people and the power of cyber-talk", The Economist (April 27, 2006).

[13] In addition, and at the request of Chinese officials, Microsoft removed a popular Web site in December 2005 from its Chinese version of MSN spaces, a site used *inter alia* for blogs; and in 2005, Google launched a self-censored version of its search engine in China. See "The party, the people and the power of cyber-talk", The Economist (April 27, 2006).

[14] For instance, the non-governmental organisation "Reporters Without Borders" has written a Handbook for Bloggers and Cyber-Dissidents (found at <http://www.rsf.org/IMG/pdf/ Bloggers_Handbook2.pdf>). The book offers advice on how to blog anonymously and technical details on how to get around censorship. Also, an increasing number of re-routing Web sites such as <http://www.anonymouse.org> offer full anonymity when surfing the Internet.

[15] See also Eisensee and Strömberg (2007) on the impact of media on the extent of disaster relief in the U.S., and Strömberg

Reinikka and Svensson (2004) show that increasing information about spending policies, announced via local newspapers, reduced fiscal leakage in Uganda. Finally, Ferraz and Finan (2008) study the effects of releasing information about corruption practices on re-election chances. The analysis relies on an anti-corruption program in Brazil initiated in 2003, when the federal government began to audit municipal governments for their use of federal funds. Municipalities were randomly audited before and after the 2004 municipal elections, and the outcomes of the audits were posted on the Internet and released to the media. Ferraz and Finan estimate the effects of the audits on re-election probabilities. They find that the probability of getting re-elected was reduced substantially in municipalities where higher levels of corruption were revealed.

Finally, the paper is related to the literature which studies the determinants of the spread of the personal computer and the Internet across countries. Notable contributions include Caselli and Coleman (2001) on the spread of computers, and Hargittai (1999), Beilock and Dimitrova (2003), Oyelaran-Oyeyinka and Lal (2005), and Chinn and Fairlie (2007) on Internet penetration. This literature has documented a positive impact of GDP per capita and the electricity infrastructure on Internet penetration, and of human capital levels on the adoption of computers. Consequently, the level of Internet penetration should be viewed as endogenous.[16]

In the present paper, however, we examine the impact of changes in the level of Internet penetration on corruption. This choice of focus is inevitable since the Internet cannot possibly have affected corruption prior to its inception and subsequent spread. The speed at which the Internet has developed world-wide is illustrated in Figure 1, which shows the evolution of the number of Web sites from 1990 onwards. In 1996, which is the starting year for our main cross-country regressions, more than 600,000 Web sites were in existence; by 2006 the number had exploded to more than 100 million. Our analysis for the U.S., by contrast, starts immediately following the launch of the World Wide Web, i.e. in 1991. In this manner we are examining the impact of within country/state Internet diffusion on changes in corruption.

- Figure 1 about here -

While our 2SLS estimates of the impact of the Internet on corruption are considerably larger than the OLS estimates, we cannot reject that they coincide. That is, we cannot reject that the rate of change in Internet penetration is exogenous in our specification. This suggests that the diffusion process has been autonomous to *changes* in corruption levels over the period, *conditional* on the initial level of corruption, and therefore conditional on factors which determine the level of corruption.

The paper is structured as follows. In Section 2, we present our specifications of choice. Section 3 outlines the identification strategy in detail; in particular, we explain how lightning activity impacts on digital equipment. Section 4 contains a description of our data, whereas Section 5 presents our results, and

(2004) on the impact of the diffusion of the radio across U.S. states for spending patterns.

[16] There is strong theoretical basis for believing that the initial adoption of new technologies 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. A positive impact from governance indicators, or income per capita, on the level of Internet users can be rationalized by this mechanism.

discusses their economic significance and robustness. Section 6 offers concluding remarks.

2 Specification

Since governance indicators tend to be persistent over time, empirical work on the determinants of corruption usually seeks to explain differences in corruption levels across countries. That is, they focus on between-country variation in corruption. In the present case, however, the inevitable choice of focus is on changes in corruption. The Internet is a recent phenomenon, and, as such, cannot possibly have affected corruption levels prior to its inception, adoption and widespread use. As a result, we study whether changes in Internet penetration can explain changes in corruption over the time-period during which the Internet has been in operation.

For the most part we will rely on the following parsimonious specification:

$$DCCI_i = \alpha_0 + \alpha_1 DINTERNET_i + \alpha_2 CCI_{initial,i} + \varepsilon_i, \tag{1}$$

where $DCCI_i$ is the change in corruption levels between an initial and a final year, $CCI_{final,i} - CCI_{initial,i}$, whereas $DINTERNET_i$ is the change in Internet penetration during this period, $INTERNET_{final,i} - INTERNET_{initial,i}$.

Specification (1) has the virtue of automatically controlling for (a potentially large set of) variables which may affect the level of corruption. To see this more clearly, observe that (1) is equivalent to the following levels regression with a lagged dependent variable:

$$CCI_{final,i} = \alpha_0 + \alpha_1 DINTERNET_i + (\alpha_2 + 1)\, CCI_{initial,i} + \varepsilon_i. \tag{2}$$

Accordingly, any *time invariant* (or slow moving) structural characteristics affecting the *level* of corruption will be picked up by $CCI_{initial,i}$ (e.g., natural resource endowments or ethnic fractionalization). This reduces the scope for omitted variable bias.

Of course, while specification (1) reduces the likelihood that $\text{Cov}(DINTERNET_i, \varepsilon_i) \neq 0$, it does not rule it out entirely. *Time-varying* characteristics may be omitted, causing $\text{Cov}(DINTERNET_i, \varepsilon_i) = 0$ to fail nonetheless. Since previous research has documented an impact of GDP per capita on the level of corruption and Internet usage, it is an obvious candidate. Consequently, we also run regressions of the form

$$DCCI_i = \alpha_0 + \alpha_1 DINTERNET_i + \alpha_2 CCI_{initial,i} + \alpha_3 GYCAP_i + \varepsilon_i, \tag{3}$$

where $GYCAP_i$ represents real GDP per capita *growth*. Observe that the *level* of GDP per capita in the initial year is implicitly controlled for by the inclusion of $CCI_{initial,i}$. Hence, only the change in GDP per capita, from this initial year to the final year, should be included in an effort to explain $CCI_{final,i}$, and therefore $DCCI_i$.

We will be running various robustness checks by varying the underlying sample. For example, we estimate the above statistical models in a world-wide sample, on restricted samples excluding individual regions one

after another, and by running the regressions using data for the 50 U.S. states. The latter exercise is particularly revealing since we expect the U.S. sample to be fairly homogenous in many respects, thus limiting further the risk of omitted variable bias.

Nevertheless, in spite of these checks one may worry that the magnitude of the obtained estimate for α_1 could be misleading on account of $\text{Cov}(DINTERNET_i, \varepsilon_i) \neq 0$. To address this concern we employ an IV approach. The next section describes our identification strategy in detail.

3 Identification

In order to explain changes in Internet use, from the time of the Web's inception to present-day, one needs to explain IT investments. In this regard the user cost of IT capital plays a central role. A higher level of user cost implies a lower desired (long-run) IT capital stock, which *ceteris paribus* will imply lower IT investments during adjustment. By implication, the speed of Internet pentration will occur at a slower pace.

The identification strategy employed in the present paper is therefore to find exogenous determinants of the user costs associated with digital equipment in general, and therefore the Internet in particular. As is well known, the user cost of capital depends, among other things, on the rate of capital depreciation (Hall and Jorgenson, 1967). Capital depreciation is one important avenue through which our instrument affects the diffusion process.

Digital equipment, such as the computer, is highly sensitive to *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. In other words, power disruptions tend to increase the user cost of IT capital. A natural phenomenon which damages digital equipment, by producing power failures, is *lightning activity* (e.g., Shim et al., 2000 Ch. 2; Chisholm, 2000).

Perhaps surprisingly, the influence from lightning activity is far from trivial. The National Lightning Safety Institute, a nonprofit organization, reports that lightning in the United States alone accounted for more than 100,000 laptop and desktop computer losses in 1997.[17] NASA (2007) points out that the main impetus to lightning research in the late 1960s was the danger of lightning to aerospace vehicles and solid-state electronics used in computers and other electronic devices.[18] In order to understand why lightning is so problematic some remarks on the physics of power supply and distribution are required.

Power plants create commercial electric power with alternating voltage. Alternating voltage, $V(t)$, as a function of time, t, takes the form of a sinusoidal wave, i.e. $V(t) = V_0 \cdot \sin(2\pi \cdot f \cdot t)$, where f is the number of oscillations per second (f is usually expressed in hertz (Hz), where f Hz means f cycles per second). The time required for the pattern to be repeated is the period $T = 1/f$. Whenever this sinusoidal wave changes size, shape, frequency, develops notches, etc., technically, there is a power disturbance.

[17] <http://www.lightningsafety.com/nlsi_lls/nlsi_annual_usa_losses.htm>

[18] Solid-state electronics refers to an electronic device in which electricity flows through solid semiconductor crystals rather than through vacuum tubes. Transistors, made of one or more semiconductors, are at the heart of modern solid-state devices. In the case of integrated circuits, millions of transistors can be involved. Microprocessors are the most complicated integrated circuits. They are composed of millions of transistors that have been configured as thousands of individual digital circuits, each of which performs some specific logic function (see Kressel 2007 for an enjoyable discussion with a historical perspective).

It is important to appreciate that computers to this day remain extremely sensitive to such disturbances. For more than a century the reliability of the electricity grid has rested at 99.9%, or so-called "three nines" reliability. Three-nines reliability permits roughly 9 hours of outages per year. This was sufficient when the economy was built around light bulbs and electric motors. But microprocessor-based controls and computer networks demand at least 99.9999% reliability (or "six nines" reliability).[19] This amounts to only seconds of allowable outages per year.

The reason why such high reliability is required is that equipment based on solid-state electronics, such as computer chips, is constructed to work under conditions of a "clean" sinusoidal wave form. The power supply of a digital device converts the alternating current to direct current with a much reduced voltage. Digital processing of information works by having transistors "turn" this small voltage on and off at several gigahertz (Kressel, 2007). If the wave becomes disturbed or distorted, the conversion process can become corrupted. This may result in equipment failure (Conroy, 1999). In fact, sensitivity to small distortions increases with the miniaturization of transistors, which is the key to increasing speed in microprocessors (Kressel, 2007). Voltage disturbances measuring less than one cycle are sufficient to crash servers, computers, and other microprocessor-based devices (Yeager and Stalhkopf, 2000; Electricity Power Research Institute, 2003). To put it differently, at a 60 Hz frequency (the standard in the U.S.) this means that a power disturbance of a duration less than 1/60th of a second is enough to crash a computer!

By some estimates, lightning is the direct cause 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).[20] As a result, lightning-caused computer failure is a quite pervasive phenomenon.

Some of the most noted instances of lightning triggered computer failures relate to airports. For example, according to Germany's Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik, BSI), a major German airport experienced a lightning strike very close to the air traffic control tower. Despite the external lightning protection system that had been installed (a lightning conductor), the computerized fire extinguishing system in the IT area was triggered by the incidence, and, as a result, all airport operations were paralyzed for two hours (BSI, 2004). The New York Times (July 3, 1999) reports how air traffic across the northeastern United States was disrupted after lightning hit New England's main air traffic control center in Nashua, N.H.[21] After the lightning strike, the mainframe computer began reporting error messages to technicians, apparently because devices attached to it were not running properly.

In general, lightning discharges can enter electronic equipment inside a residence in four principal ways (IEEE, 2005). First, lightning can strike the network of power, phone, and cable television wiring. This network, particularly when elevated, acts as an effective collector of lightning surges. The wiring conducts the surges directly into the residence, and then to the connected equipment. In fact, the initial lightning impulse is so strong that equipment connected to cables up to 2 km away from the site of the strike can be

[19] See for example National Energy Technology Laboratory (2003). Yeager and Stalhkopf (2000) actually argue that "nine nines" reliability is needed. See also "The power industry's quest for the high nines", The Economist (March 22, 2001).

[20] 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).

[21] "Lightning Hits Control Center And Grounds Many Planes" by Matthew L. Wald.

damaged.[22] Second, when lightning strikes directly to or nearby air conditioners, satellite dishes, exterior lights, etc., the wiring of these devices can carry surges into the residence. Third, lightning may strike nearby objects such as trees, flagpoles, road signs, etc., which are not directly connected to the residence. When this happens, the lightning strike radiates a strong electromagnetic field, which can be picked up by the wiring in the building, producing large voltages that can damage equipment. Finally, lightning can strike directly into the structure of the building. This latter type of strike is extremely rare, even in areas with a high lightning density.

Accordingly, if computer equipment is left unprotected, high lightning intensity will work so as to increase the user cost of digital capital by elevating the rate of capital depreciation. Of course, steps may be taken to protect the equipment. 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 "uninterruptedly power sources" relying on batteries to power computers until generators kick in. These initiatives, however, 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 a high annual lightning density, this adds to the costs of using modern electronics, including a computer.[23]

In the discussion above we have focused on the impact of power quality on the marginal costs of IT capital. But this is not the only way in which lightning may matter for Internet diffusion. In areas with frequent and prolonged power disruptions and outages, the marginal *benefit* of owning a computer is surely diminished. This mechanism will also work to lower the speed of Internet diffusion. Consequently, lightning matters to Internet diffusion for two reasons. By inducing power disturbances, it increases the marginal costs as well as lowers the marginal benefits of IT capital. This is not to say that these two complementary channels nessesarily are equally relevant everywhere. In the U.S. sample we would conjecture that the benefit channel is of secondary relevance; power outages and disruptions are likely of too short a duration to seriously impact on the perceived benefits from IT. In the cross-country sample, however, both the cost channel and the benefit channel may be relevant; especially in poor countries. For this reason we would expect a stronger impact of lightning on Internet penetration in the cross-country context, compared to the U.S. sample.[24]

Against this background we propose lightning density as an instrument for the speed at which Internet

[22] According to BSI (2004, p. 258), with discharges of several hundred thousand volts, lightning strikes can achieve currents of up to 200,000 amperes. This enormous electrical energy is released and dies away within a period of 50-100 microseconds. A lightning strike of this order of magnitude originating from a distance of about 2 km will still cause voltage peaks that are capable of destroying sensitive electronic devices in the power lines of a building.

[23] 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 (see <http://www.lightningsafety.noaa.gov/ indoors.htm>). 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).

[24] Some may instinctively believe that the marginal cost of protecting the computer (completely) from lightning also is negligible in the U.S. context. In this regard William A. Chisholm has suggested to us in personal correspondence that a laptop computer with battery and wireless Internet access may indeed be completely protected from lightning at a cost of about $200 extra. This is hardly a trivial expense. However, and more importantly, these measures will not protect the nearby cell-phone tower providing wireless infrastructure. The point being that the cost of wireless service will reflect the local providers' costs in replacing equipment when the tall tower is struck repeatedly by lightning. Hence, either way, lightning density will elevate the costs of being online.

use per capita changed over the period in question. In Section 5.3 we employ cross-country data so as to corroborate the logic of the identification strategy. That is, we document that higher lightning density indeed is associated with more power outages per year. Moreover, a higher frequency of power disruptions has a marked impact on the speed of Internet diffusion, confirming previous findings of a link between the electricity infrastructure and the spread of the Internet. In reduced form, this motivates empirically the use of lightning density as an instrument for changes in Internet use over the 1990s.

Lightning is certainly exogenous in a deep sense. However, this does not ensure validity of the exclusion restriction. Validity rests on a redundancy condition in (1). Climate-related circumstances are likely to map into levels of corruption. For instance, to the extent that countries with a high lightning density have a higher endowment of natural resources, this is likely to affect levels of corruption (resource curse). As mentioned in Section 2, this circumstance is controlled for with the inclusion of $CCI_{initial.}$ Validity of the exclusion restriction therefore requires that lightning has no *direct* impact on changes in corruption over the period under study once we control for the level of corruption initially, $CCI_{initial}$. Formally, let z denote the lightning density; validity requires $\mathrm{Cov}(\varepsilon, z) = 0$, where ε is the error term in (1).

4 Data description

4.1 Corruption

4.1.1 Cross country

Global corruption is measured using the well-known Control of Corruption Index (CCI) compiled by Kaufmann et al. (2007). The CCI measure, which ranges from -2.5 (worst) to 2.5 (best), is available biannually from 1996 to 2002 and annually from 2002 onwards.[25] The CCI indicator attempts to measure "the extent to which public power is exercised for private gain, including both petty and grand forms of corruption as well as capture by elites and private interests" (Kaufmann et al., 2007, p. 4). The indicator is based on a large number of individual data sources, which are then aggregated into one measure by an unobserved components model. This means that the aggregate measure is a weighted average of the underlying individual data sources, with weights reflecting the precision of each of these underlying data sources. By virtue of being a solution to a statistical signal extraction problem, the aggregate CCI indicator is presumably more informative than any individual data source.[26] This makes the CCI the most comprehensive measure of corruption around.[27]

[25] These boundaries correspond to the 0.005 and 0.995 percentiles of the standard normal distribution. For a few countries, country ratings can exceed these boundaries when scores from individual data sources are particularly high or low (Kaufmann et al., 2007).

[26] Svensson (2005), however, notes that the aggregation procedures used by Kaufmann et al. presumes that subindicator measurement errors are independent across sources. In reality, errors may be correlated since producers of different indices read the same reports and most likely each other's evaluations. If the assumption of independence is relaxed, the gain from aggregating a number of different reports is less clear.

[27] The widely reported Corruption Perception Index (CPI) compiled by Transparency International forms part of the CCI measure (see Kaufmann et al., 2007, Table A13). Reasuringly, the simple correlation between CCI and CPI is 0.97. A high correlation is not unexpected since corruption reflects the underlying institutional framework (Svensson, 2005).

$DCCI$ is calculated as the difference between CCI in 2006 and 1996, i.e.

$$DCCI_i = CCI_{2006,i} - CCI_{1996,i}. \tag{4}$$

Ideally, we would like to go back to 1991, the launch date of the WWW (cf. Figure 1). Unfortunately, with the preferred CCI measure this is not possible. However, in Section 5.6 we will go back to 1991 in the cross-country context using a different measure of corruption.

As noted by Kaufmann et al. (2007), despite being somewhat persistent, governance indicators do change over relatively short periods such as a decade. This is illustrated in Figure 2, which plots the 1996 CCI score on the horizontal axis, the 2006 CCI score on the vertical axis, and a 45-degree line. Countries above the 45-degree line correspond to improvements in corruption, while countries below the line saw deteriorations in corruption.

- Figure 2 about here -

4.1.2 U.S. states

Turning to U.S. corruption data, we do not have to rely on a perception-based indicator of corruption.[28] Instead we use a more direct measure of corruption, viz. the number of government officials convicted for corrupt practices through the Federal Justice Department. We name this variable CC.

The corruption data is reported in the Justice Department's "Report to Congress on the Activities and Operations of the Public Integrity Section". This publication provides the number of federal, state, and local public officials convicted of a corruption-related crime by state. As argued by Glaeser and Saks (2006), conviction levels capture the extent to which federal prosecutors have charged and convicted public officials for misconduct in each of the 50 U.S. states. 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 focusing on 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).

DCC is calculated as the difference between CC in 2006 and 1991, i.e.

$$DCC_i = \log(1 + CC_{2006,i}) - \log(1 + CC_{1991,i}). \tag{5}$$

Note that for the U.S. sample we can go back to 1991. This is probably more important for the U.S. sample as compared to the cross-country sample since the U.S. was the frontrunner in terms of early Internet diffusion.

[28] Some are sceptical concerning the use of perception-based corruption data (Svensson, 2005). Interestingly, however, Olken (2006) has provided novel evidence on the relation between corruption perceptions and a direct measure of corruption in the context of villages across Indonesia. In a household survey, villagers were asked to state their beliefs about the likelihood of corruption in a road-building project in their village. Olken estimated the direct costs of corruption, called 'missing expenditures', by the difference between the actual and reported costs involved with the project. The actual costs were estimated by engineers after the roads were built by digging samples of the road to determine quantities, surveying local suppliers to determine prices, and interviewing villagers to determine wages. The empirical results show that villagers' perceptions of corruption do appear to be positively (albeit weakly) correlated with the direct missing expenditure measure. Olken finds that corruption in the project was predominantly hidden by inflating quantities, which are hard for villagers to detect, as opposed to marking up prices, which are easier for villagers to detect.

Note also that we did not employ a (natural) log-transformation in the global data since CCI is an index. In contrast, the CC measure is a levels variable, so using logs means taking out scale effects.[29] Importantly, note also that unlike CCI, *higher* values of CC corresponds to *more* corruption.

Figure 3 plots the (log-transformed) 1991 CC score on the horizontal axis, the (log-transformed) 2006 CC score on the vertical axis, and a 45-degree line. States *below* the 45-degree line now correspond to improvements in corruptions, while states *above* the line saw deteriorations in corruption.

- Figure 3 about here -

4.2 Internet

4.2.1 Cross country

Our key explanatory variable is the number of Internet users per 100 people. For a growing number of countries, 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.

Since global corruption data goes back to 1996, we calculate the change in Internet users over this period:

$$DINTERNET = INTERNET_{i,2005} - INTERNET_{i,1996},$$

where 2005 is the most recent year for Internet users in the WDI (2007).

4.2.2 U.S. states

Internet users per 100 people in the U.S. is based on data collected in a supplement to the October 2003 Current Population Survey (CPS), which included questions about computer and Internet use.[30] The CPS is a multi-stage probability sample with coverage in all 50 states and the District of Columbia. 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 occupied 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

$$DINTERNET = INTERNET_{i,2003} - INTERNET_{i,1991} = INTERNET_{i,2003},$$

since $INTERNET_{i,1991} = 0$ for all i (i.e. for all states).

[29] We use the $\log(1 + CC)$ since CC is zero in 14 out of 100 state-year observations.

[30] Released October 27, 2005, by the U.S. Census Bureau.

4.3 Lightning

4.3.1 Cross country

The raw data for flash densities (flashes per km^2 per year) comes from NASA. The Global Hydrology and Climate Center (GHCC) has designed, constructed, and deployed numerous types of groundbased, airborne, and spacebased sensors to detect lightning activity and to characterize the electrical behavior of thunderstorms as part of their research on atmospheric science. The GHCC's spacebased sensors detect all forms of lightning activity over land and sea 24 hours a day. Thus, such sensors allowed the development of the first global database of lightning activity, which has been used so far for severe storm detection and analysis, and for lightning-atmosphere interaction studies.

In this paper, we rely on the data from the 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 $km \times 1300$ km region of Earth. "Flashes" were 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.[31] These satellite-based data are archived and cataloged by the GHCC, where they are also made available, free of charge.[32]

We apply the data from a high-resolution (0.5 degree latitude $\times$ 0.5 degree longitude)[33] grid of total lightning bulk production across the planet, expressed as a flash density, from the completed 5 year OTD mission. Figure 4 provides a world map of the average flash density over the 5 years period.

- Figure 4 about here -

Average flash densities are constructed for the Minimum Bounding Rectangle (MBR) of each country. The MBR is the rectangular area covering the northernmost, southernmost, westernmost, and easternmost limits of an object. The MBR is a frequently used approximation for computing spatial indices on different geographical features (Caldwell, 2005).

The coordinates for each country are taken from the GEOnet Names Server (GNS) at the U.S. National Geospatial-Intelligence Agency's (NGA) and the U.S. Board on Geographic Names' (U.S. BGN) database of foreign geographic feature names.[34] We have used the version released on October 2, 2007.[35] The GNS data covers the entire planet with the exception of the U.S. and Antarctica. The MBR for the U.S. was constructed using data for the 50 individual U.S. states, where data is taken from the Geographic Names Information System at the U.S. BGN.[36]

Using these coordinates, we construct minimum and maximum latitude and longitude values for each country. Caldwell (2005) emphasizes two key issues that affect the accuracy of the MBR as an approx-

[31] 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.

[32] <http://thunder.msfc.nasa.gov/data/#OTD_DATA>

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

[34] <http://earth-info.nga.mil/gns/html/namefiles.htm>

[35] <ftp://ftp.nga.mil/pub2/gns_data/geonames_dd_dms_date_20071002.zip>

[36] <http://geonames.usgs.gov/domestic/download_data.htm>

imation of the area coverage of a feature. These are the presence of multiple component parts and the shape/orientation of the particular country. As an example of the former, the Bahamas consists of 2000 cays and 700 islands, which makes the Bahamas impossible to encapsulate with a MBR without including ocean. Concerning the latter, a country not well approximated by an MBR due to its diagonal shape is Japan. A third issue concerns small countries. Malta is only 316 km^2, which is much smaller than the area of one high resolution grid at this latitude, around 2.500 km^2. Furthermore, Malta is located in the crossing between two gridlines, which means that it is encapsulated by four grids. Therefore, the MBR around Malta is much larger than the actual size of Malta.

To assess the validity of the MBR as an approximation of the country area, we calculated the ratio of the area of the MBR to the actual area of a country. In our sample of 113 countries, this ratio has a mean of 3.8 and a median of 2.0. The Bahamas lies in the top of this range with a ratio of 59.2 due to the features mentioned above. Thereafter comes Malta with 31.2. In the bottom of the range lies Egypt with a ratio of 1.1. The larger ratios in this sample mostly represent instances where areas of the ocean are included when attempting to encapsulate the country. As lightning is less frequent over sea, this will bias the average flash rates downwards for these countries.

Consequently, we employ two additional measures of lightning activity: Average flash densities in country capitals and in the midpoint of each country's MBR. As it turns out, our results hold regardless of which of the three measures we employ.

4.3.2 U.S. states

For U.S. states we compile the state MBR as described above. As also mentioned above, coordinate data for U.S. states is from the Geographic Names Information System at the U.S. BGN.[37]

It should be noted that the OTD data provides observations on *total* lightning events, i.e. both 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. Since the mid-1980's, it has been possible to measure ground flash density more directly using networks of electromagnetic sensors. Such Lightning Location Systems (LLS) are able to resolve individual ground strikes comprising a flash with high spatial and temporal accuracy. However, many parts of the world, particularly the developing world, are not covered by the LLS data.[38]

Fortunately, accurate cloud-to-ground data does exist for the 48 continental U.S. states; it does not exist for Alaska and Hawaii. These cloud-to-ground lightning flashes, which are measured by the U.S. National Lightning Detection Network (NLDN), are provided by Vaisala for the period 1996-2005.[39] NLDN consists of more than 100 remote, ground-based lightning sensors, which instantly detect the electromagnetic signals given off when lightning strikes Earth's surface. Comparing the NLDN ground-based measures with the NASA satelite-based measures for the U.S. provides an indication of the extent to which the total amount

[37] <http://geonames.usgs.gov/domestic/download_data.htm>
[38] In addition, data is not freely available for the parts of the world actually covered.
[39] See <http://www.vaisala.com/thunderstorm>

of lightning is a reliable indicator of cloud-to-ground lightning.

Figure 5 provides a scatter plot of total lightning against cloud-to-ground lightning. 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.[40]

- Figure 5 about here -

In spite of the difference in time periods, Figure 5 clearly shows that there is agreement between the two measures of lightning. With a correlation above 0.94, total lightning is indeed a very good proxy for cloud-to-ground lightning in the U.S. sample. This can also be expected to be the case in the cross-country data as well, cf. Chisholm and Cummins (2006).

Nevertheless, in the regressions below we will use all available information to form three instruments: the cloud-to-ground lightning density provided by Vaisala; the MBR-based lightning density of a state, which we ourselves construct using OTD data; and the lightning density at the midpoint of the state MBR, which also involves OTD data.

4.4 GDP per capita

4.4.1 Cross country

Real GDP per capita, $YCAP$, for the global sample is taken from WDI (2007). Specifically, we calculate the growth rate of real GDP per capita as

$$GYCAP_i = (YCAP_{i,2005}/YCAP_{i,1996}) - 1,$$

where 2005 is the most recent year with observations on real GDP per capita in the WDI (2007).

4.4.2 U.S. states

For the U.S. our income measure is personal income per capita taken from the State Personal Income Database from U.S. Bureau of Economic Analysis (BEA). We deflate nominal income per capita using the implicit price deflator for personal consumption expenditures, index year 2000, also from BEA. We calculate the growth rate in real personal income per capita over the period 1991 to 2006 as

$$GYCAP_i = (YCAP_{i,2006}/YCAP_{i,1991}) - 1.$$

4.5 Electrical Outages

In order to examine whether indeed lightning is associated with more frequent power disruptions, which in turn affects the speed of Internet diffusion within countries, we need data on the frequency of power

[40] 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.

failures.[41] In spite of our best efforts, we have only been able to obtain data on power failures for a cross section of countries. The cross-country measure of the frequency of power disruptions is collected by the World Bank, and it is available in WDI (2007) for the years 2002-2006. Specifically, the variable measures the average number of days per year that establishments experience power outages or surges from the public grid. We will use this variable as a proxy for the intensity of power disturbances.

This variable, which we will refer to as OUTAGES, form part of the World Bank Group's Enterprise Surveys.[42] These surveys cover almost 58,000 firms in 97 countries for the period 2002–2006. For most countries electrical outages are measured only at one point in time during the period 2002-2006. However, for 26 countries there are two observations, one in 2002 and one in 2005. In these cases we use the 2005 values.

5 Results

5.1 Summary statistics

Table 1 reports summary statistics for the main levels variables.[43] Several features are worth noting. The cross-country sample illustrates the rapid rise of the Internet. The (unweighted) cross-country average fraction of the population having Internet access was 1.5% in 1996; by 2005, this number had risen to nearly 23%. During the process of Internet diffusion, the cross-country variation also rises enormously. The standard deviation in fact rises by a factor of 7, from 1996 to 2005, reflecting what some have called "the increasing digital divide" across the globe (e.g., Chinn and Fairlie, 2007). Accordingly, the spread of the Internet has clearly progressed at very different speeds around the world.[44]

- Table 1 about here -

In the U.S. sample this continues to hold true. The (unweighted) cross-state average fraction of the population having Internet access rises from zero in 1991 to roughly 60% in late 2003. However, this increase is far from uniform. In late 2003, the number of Internet users per 100 people varies from a low of 43 to a high of 72. In the cross-country data this corresponds to moving from number 25 (just behind France) to the 5th place in the 2005 world distribution (just behind Norway). Hence, even in the U.S. sample there is

[41] This independent check of the economic relevance of our identification strategy is conducted in Section 5.3.

[42] <http://www.enterprisesurveys.org/>

[43] The 113 countries in the cross-country sample are the following: Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, The Bahamas, Bahrain, Belarus, Belgium, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Canada, Chile, China, Colombia, Congo, Dem. Rep., Congo, Rep., Costa Rica, Cote d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Ethiopia, Finland, France, Gabon, Germany, Ghana, Greece, Guatemala, Guinea, Guyana, Haiti, Honduras, Hong Kong (China), Hungary, Iceland, India, Indonesia, Iran, Ireland, Italy, Japan, Kenya, Korea, Rep., Kuwait, Latvia, Lebanon, Lithuania, Luxembourg, Macedonia, Madagascar, Malaysia, Mali, Malta, Mexico, Mongolia, Morocco, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Poland, Portugal, Qatar, Russian Federation, Senegal, Slovak Republic, Slovenia, Somalia, South Africa, Spain, Sudan, Suriname, Sweden, Switzerland, Syrian Arab Republic, Thailand, Togo, Tunisia, Turkey, Uganda, Ukraine, United Arab Emirates, United Kingdom, Uruguay, Venezuela, Vietnam, Zimbabwe. The U.S. states include all fifty states.

[44] In 2005, Iceland is the country in the cross-country sample with the highest Internet penetation at nearly 87%; Niger ranks last at 0.2%.

evidence suggestive of substantial variation in the speed of Internet diffusion.

In the context of the cross-country data, it is also worth observing that the intensity of power failure varies tremendously across the globe. In the median country power outages or surges occur on average 9.6 days per year. Notice, however, that the standard deviation is huge: 46.6 days. Indeed the range varies from a low of 0.04 days (Korea) to a high of 248.96 days (Bangladesh). Our sample only contains 8 OECD countries (Belgium, Switzerland, Germany, Greece, Korea, Ireland, Spain and Portugal). Within this group the intensity of power failures is rather small, as expected. In this group the largest observation is 3.7 days (Switzerland).

5.2 Ordinary Least Squares

5.2.1 Cross country

Table 2 explores the partial association between changes in Internet use and changes in corruption for the cross-country data. The dependent variable is $DCCI$, defined by equation (4). Column 2 estimates the parsimonious specification (1), described in Section 2. The column shows that $DINTERNET$ is estimated with high precision: With a t-value above 4, significance is well below the 1% level. Including the growth rate of real GDP per capita, i.e. estimating (3), has no impact on the slope estimate associated with $DINTERNET$, cf. column 3. Moreover, the growth rate of real GDP per capita holds no explanatory power once we control for initial corruption and changes in Internet use.

- Table 2 about here -

The standardized coefficient associated with column 2 is slightly above 0.67.[45] This means that a one-standard deviation increase in $DINTERNET$ is associated with 0.67 standard deviations increase in $DCCI$.[46] In fact, the simple specification in column 2 can explain 22% of the variation in changes in levels of corruption of the period 1996 to 2006.[47]

The point estimate associated with changes in Internet use is remarkably robust. In a statistical sense, it remains constant when we exclude regions of the world one-by-one (using the WDI classification), as done in Table 3. The last row in Table 3 performs an F-test, where the null is that the slope estimate of $DINTERNET$ associated with each subsample is equal to the slope estimate associated with the full sample, i.e. column 2 in Table 2. The null cannot be rejected at any conventional level of significance. In addition, including $GYCAP$ does not change this conclusion, for which reason it has been excluded from the regressions reported in Table 3. Consequently, the partial association between changes in Internet use and changes in corruption over the period under consideration is very robust.

- Table 3 about here -

[45] The standardized (or beta) coefficient associated with α_1 in (1) is given by $(\sigma_{DINTERNET}/\sigma_{DCCI}) \times \alpha_1$, where σ_j is the standard deviation of variable j.

[46] We will address the economic significance in more detail in Section 5.5.

[47] Note that in the equivalent levels equation given by (2), more than 86% of the variation in corruption levels can be accounted for.

5.2.2 U.S. states

Turning to the 50 U.S. states, Table 4 reports results from OLS. The dependent variable is DCC, defined by equation (5). Column 2 estimates the specification (1). $DINTERNET$ is also estimated with high precision in the U.S. sample.[48] With a t-value of 2.43 in column 2, significance holds at the 2% level. The specification (1) accounts for more than 30% of the variation in DCC in the U.S. sample. Moreover, including the growth rate of real GDP per capita once again has no impact on the slope estimate associated with $DINTERNET$, and, as was also the case in the the cross-country sample, the growth rate of real GDP per capita holds no explanatory power in the U.S. sample once we control for initial corruption and the change in Internet users, cf. column 3. The standardized coefficient associated with column 2 is roughly -0.33.

- Table 4 about here -

5.3 Towards Two-Stage Least Squares: "The Zeroth Stage"

As explained in Section 3, we aim to obtain identification by using lightning density as an instrument for changes in Internet users over the period where the WWW has been in operation. The rationale for this strategy, outlined in Section 3, can be expressed in the following schematic fashion:

$$\text{LIGHTNING DENSITY} \longrightarrow \text{ELECTRICAL OUTAGES} \longrightarrow \text{INTERNET USE.}$$

where the second arrow implicitly subsumes the impact of power surges on the costs and benefits of IT capital. While we argue these links are highly plausible on physical grounds, it does not follow that they are quantitatively relevant in an economic context. In fact, absent evidence of these links, one may worry that an observed association between changes in Internet penetration and lightning could be spurious. By extension, the plausibility of the indentifying assumption that the lightning variable does not affect corruption, conditional on Internet penetration and initial corruption levels (cf. Section 2), would be hard to gauge. Clearly, the plausibility of the exclusion restriction is related to the empirical relevance of the full theory, which underlies the identification strategy. To address this issue, we explore the above chain of causation in the present section before turning to 2SLS regressions in the next section.

To proceed, we express the underlying logic of the identification strategy as a system of equations:

$$DINTERNET_i = \alpha_0 + \alpha_1 \cdot \log(OUTAGES_i) + \mathbf{X}_I'\gamma_I + \varepsilon_i, \quad (6)$$

$$\log(OUTAGES_i) = \beta_0 + \beta_1 \cdot \log(LIGHTNING_i) + \mathbf{X}_O'\gamma_O + v_i. \quad (7)$$

where $\mathbf{X}_O$ and $\mathbf{X}_I$ contain additional controls. The basic question is whether $\beta_1 > 0$ and $\alpha_1 < 0$.

Comparable data on electrical outages is only available for the cross-country sample; the data used

[48] Note that increased digital information 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. Better detection technology will work in the direction of more convictions, whereas lower incentives will work in the direction of fewer convictions. A priori, the net effect of more information on convictions is thus ambiguous. Our results show that empirically the net effect is negative, implying that the incentive effect dominates.

when estimating equations (6) and (7) are therefore those described in Sections 4.1.1, 4.2.1, 4.3.1, and 4.5. Specifically, in the regressions to follow we employ the MBR-based average lightning density for a country (i.e., what we refer to as "INSTRUMENT 1" in the next section). Using either of the alternatives described in Section 4.3.1 produces very similar results. In the interest of brevity, however, we only report the results from employing this variable in equation (7). Moreover, we will run regressions (6) and (7) both without controls and with controls, in which case $\mathbf{X}_O$ and $\mathbf{X}_I$ equal CCI in 1996.

In principle, a variety of variables could be included in $\mathbf{X}_O$ and $\mathbf{X}_I$. The choice of only including initial corruption as additional control is made for two reasons. First, governance indicators have previously been shown to affect energy efficiency (Frederikson et al., 2004) as well as Internet use (e.g., Chinn and Fairlie, 2006). Second, and more important in the present context, by controlling for corruption these specifications will have direct bearing on the validity of the 2SLS regressions in the next section: the identification strategy (implicitly) requires lightning to affect outages, and outages to affect Internet use *conditional* on CCI_{1996}. Consequently, this is the setting where confidence in the validity of the identification strategy, involving lightning as an instrument for changes in Internet use, is required.

Table 5 reports the results from estimating equations (7) (columns 1 and 2) and (6) (columns 3-5). In column 1 we report the bivariate association between lightning density and electrical outages and surges. The correlation is both highly significant and positive as required. Moreover, lightning can account for 13% of the global variation in electrical outages. In column 2 we add CCI_{1996}. The size of the point estimate for lightning declines but remains significant at the 5% level. Higher levels of corruption are associated with less power reliability. This is consistent with the findings of Frederikson et al. (2004). Figure 6 provides a visual impression of the partial correlation between $\log(LIGHTNING)$ and $\log(OUTAGES)$ (i.e., conditional on CCI_{1996}). Taking the point estimate at face value, the results imply that an increase in lightning density of one percent increases the number of days where power outages take place by about 0.4%.

- Table 5 about here -

- Figure 6 about here -

Turning to column 3 of Table 5, we find that a higher intensity of electric power failures hampers the spread of the Internet within countries. Indeed, electrical outages and surges can account for nearly 60% of the total variation in the Internet variable. Adding CCI_{1996} (column 4) does not change the impression of a first-order impact from outages to the spread of the Internet. In fact, outages appear to have a quantitatively stronger effect on the spread of the Internet than corruption. Specifically, increasing $\log(OUTAGES)$ by one standard-deviation is associated with a reduction in $DINTERNET$ of -0.61 standard-deviation units; in constrast, a similar exercise using CCI_{1996} would increase $DINTERNET$ by 0.29 standard-deviation units.[49] Figure 7 provides a visual impression of the strength of the relation between power failures and

[49] We also experimented with adding log of real (ppp) income per capita in 1996 in columns (2) and (4) instead of CCI_{1996}. The results are similar to those reported above. This is perhaps not too surpricing in light of the fact that the correlation between log of real (ppp) income per capita and CCI_{1996} is almost 0.8. We also found that the standardized coefficient associated with outages in column (4) remains numerically larger than the standardized coefficient associated with log of real (ppp) income per capita, when the latter is used instead of CCI_{1996}.

changes in Internet use.

- Figure 7 about here -

Finally, in column 5 we add $\log(LIGHTNING)$ as an explanatory variable. If lightning density affects the spread of the Internet in ways unrelated to electrical outages, we would expect it to be significant conditional on $\log(OUTAGES)$. If so, one may worry that lightning density is proxying for other variables affecting Internet diffusion, which would raise doubts concerning the validity of the exclusion restriction maintained in the next section. Fortunately, this is not the case: lightning is not a significant determinant of changes in Internet use over the period in question, *conditional* on power failures. This finding strongly suggests that the statistical correlation between $\log(LIGHTNING)$ and $DINTERNET$, which we document in the next section, arises due to lightning's influence on the frequency of power failures, which in turn affects the spread of the Internet within countries.

Overall, these findings show that the individual theoretical links underlying our identification strategy are present in the data; they are statistically as well as economically relevant. Against this backdrop, we now turn to 2SLS regressions, where lightning will be used as instrument for changes in Internet use.

5.4 Two-Stage Least Squares

5.4.1 Cross country

Turning now to instrumental variables methods, Table 6 reports 2SLS-based results for the cross-country sample. Panel A of Table 6 provides results from the first-stage regressions in the 2SLS procedure. INSTRUMENT 1 is the MBR-based average lightning density for a country, INSTRUMENT 2 is the lightning density in the capital of a country, and INSTRUMENT 3 is the lightning density at the midpoint of the country MBR. In accordance with the logic of the identification strategy, the lightning density is strongly and negatively related to the diffusion of the Internet. We also note that the three instruments are about equally powerful, and yield similar results in the second stage.

- Table 6 about here -

The association between average lightning density and the change in Internet use, once the influence of initial corruption is partialled out, is close to the "rule-of-thumb" value suggested by Staiger and Stock (1997). Specifically, if the F-value, associated with the null of zero explanatory power of the instrument, is above ten we need not concern ourselves with issues of weak identification: Inference based on 2SLS estimates will not be plagued with size distortions. This rule-of-thumb, however, requires that the error variance is homoskedastic (see assumption M, part (c) in Staiger and Stock, 1997). While this assumption is fulfilled in the second-stage regression (cf. Pagan-Hall test), it fails in the first-stage regression (cf. Breusch-Pagan/Cook-Weisberg test). Consequently, while inspection of Panel A reveals the F-value is "close" to ten in all columns, with non-constant error variance it is not entirely clear what this means for the actual size of significance tests. As a consequence, using confidence intervals that are robust to both heteroskedasticity and weak identification, the final paragraphs of this section will show that we need not worry about weak

identification.

Turning to Panel B, we see that 2SLS yield results that are broadly consistent with OLS, cf. Table 2. $DINTERNET$ remains significant at 5%; the slope estimates are roughly twice the size of the corresponding OLS estimates but standard errors also more than double. A formal test of the equality of OLS and 2SLS is the Hausman specification test. This test, which is reported in Table 5, Panel B, does not allow rejection of the null that $DINTERNET$ is exogenous. Put differently, we do not detect any systematic difference between OLS and 2SLS coefficient estimates. However, if our 2SLS esimates are weakly identified the overall properties of the test are somewhat unclear.

We address the problem of potential size distortions by invoking a method proposed by Chernozhukov and Hansen (2005), which allows us to construct confidence intervals with the correct size regardless of the strength of instruments.[50] Moreover, Chernozhukov and Hansen show how these size-adjusted confidence intervals can be made robust to non i.i.d. errors. Finally, the procedure has good power properties.

Panel B reports these size-adjusted confidence intervals, which are also heteroskedasicity robust. As is evident from the table, size-adjusted confidence intervals deem $DINTERNET$ above zero in all columns. Hence, qualitatively, weak-instrument robust confidence intervals add nothing new to the 2SLS and OLS story. In addition, the midpoints of the intervals are all very close to 2SLS. In column 1, for instance, the midpoint is roughly 0.045, which is above both OLS (0.016) and 2SLS (0.031) estimates. This pattern is repeated in the remaining columns.

In sum, using OLS to address the economic impact of Internet use on corruption is not misleading in the cross-country sample. However, OLS will likely provide conservative estimates as both 2SLS and the midpoints of size-adjusted confidence intervals are larger than the OLS estimate.

5.4.2 U.S. states

In the sample of U.S. states, the three different instruments relied upon are: INSTRUMENT 1, the cloud-to-ground lightning density based on U.S. National Lightning Detection Network; INSTRUMENT 2, the MBR-based lightning density of a state; and INSTRUMENT 3, the lightning density at the midpoint of the state MBR.

Table 7 provides results. Inspecting Panel A, we see that the logic of the identification strategy also holds in the U.S. sample: The lightning density is strongly and negatively related to the diffusion of the Internet across U.S. states. We again note that the choice of instrument is of little importance; results are essentially unaltered moving from column 1 to column 3. Column 1, relying on ground-based lightning, is two observations short, cf. Section 4.3. Columns 4 and 5 therefore reestimate columns 2 and 3 on this reduced sample, with the alternative instruments, but with little change in results.

While Tables 6 and 7 are not directly comparable (due to differences in corruption data), it is nevertheless worth observing that the point estimate for lightning is numerically smaller in the U.S. sample, compared to the cross-country sample. Taken at face value, this suggests that lightning has a smaller impact on Internet diffusion in the U.S., compared to across countries. The reason may be the following: In Section 3

[50] These confidence intervals are weak-identification robust since information about the (partial) correlation between instruments and the endogenous variable is not used.

we noted that lightning in theory may impact on both marginal costs and marginal benefits of IT capital. However, since electrical outages and disturbances undoubtedly are of much shorter duration in the U.S., the marginal benefit channel is quite possibly of secondary importance here. In poor countries, however, it may be of substantial importance. This could explain the smaller marginal first-stage impact of lightning on $DINTERNET$ in the U.S. sample, compared to the cross country sample.

- Table 7 about here -

We cannot reject the null of homoskedasticity. The rule-of-thumb is therefore informative, and since F-values are always above ten, instruments are strong in all columns.

As in the cross-country sample, 2SLS yield results that are broadly consistent with OLS. $DINTERNET$ remains significant at conventional levels. Testing the equality of OLS and 2SLS via the Hausman test does not allow rejection of the null that $DINTERNET$ is exogenous. We are thus led to conclude that valid inferences can be drawn from OLS in the U.S. sample as well. Observe that in the U.S. sample OLS also tends to produce smaller point estimates compared with 2SLS; OLS results therefore represent conservative estimates of the impact from the spread of the Internet on changes in corruption.

5.5 Economic significance

How large is the effect of an increase in the number of Internet users on corruption? We will address this question using our OLS estimates to remain conservative.

5.5.1 Cross country

In the cross-country sample we rely on the levels specification (2), i.e.

$$CCI_{2006} = \alpha_0 + \alpha_1 DINTERNET + (\alpha_2 + 1)\, CCI_{1991}.$$

Using the results from OLS reported in column 2 of Table 2, the effect of an increase in $INTERNET_{2005}$ *holding the initial level of Internet users constant* will be considered. In this case, we have that

$$\Delta CCI_{2006} \simeq \alpha_1 \Delta INTERNET_{2005}. \tag{8}$$

The "typical" country in the cross-country sample is Morocco ($CCI_{2006} = -0.060$ and $INTERNET_{2005} = 15.25$). How would Morocco's corruption ranking change if we increase the number of Internet users in 2005 from the median to the third quartile of the world distribution? This amounts to increasing $INTERNET$ from 15 in one hundred to 35 in one hundred (i.e. to the level of Spain). Inserting in (8) gives $\Delta CCI_{2006} \simeq (0.016) \cdot 20 = 0.32$. That is, it would increase the CCI_{2006} score for Morocco from -0.060 to 0.26. Put differently, moving from the median to the third quartile in terms of Internet users would move the corruption score from the median to roughly the 62nd percentile, which happens to be the level of Italy. While being economically important, the effect does not seem implausible large.

5.5.2 U.S. states

In the U.S. sample we will also rely on the levels specification associated with (2), i.e.

$$\log(1 + CC_{2006}) = \alpha_0 + \alpha_1 DINTERNET + (\alpha_2 + 1)\log(1 + CC_{1991}).$$

If we linearize this levels specification, treating CC_{1991} as a constant, the following simple approximation emerges:

$$\Delta CC_{2006} \simeq \alpha_1 (1 + CC_{2006}) INTERNET_{2003}, \tag{9}$$

where we have used that for the U.S., $DINTERNET = INTERNET_{2003}$, cf. Section 4.2.

We will consider the impact of changes in Internet use on corruption for the "typical" U.S. state. The typical U.S. state in terms of corruption convictions and Internet use is Missouri ($CC = 20$ and $INTERNET_{2003} = 60.45$). In a similar fashion as before, we may perform an experiment by which Missouri is moved to the third quartile in the distribution of Internet users in 2003; this is equivalent to an increase of exactly 4 Internet users per 100 people. Using the OLS results reported in column 2 in equation (9) we find $\Delta CC_{2006} \simeq (1 + 20) \cdot (-0.06) \cdot (4) = -5.04$. That is, it would reduce the number of corruption convictions (i.e. reduce the level of corruption) by roughly 5 yearly convictions. Put differently, moving up eight places in the ranking of Internet users (i.e. from the median to the third quartile) would move Missouri down three places in the U.S. state corruption convictions ranking (i.e. from the 65th percentile to the 59th percentile of the corruption convictions distribution in 2006).

5.6 Robustness

In this section we examine robustness of the cross-country results using an alternative corruption variable. The ICRG corruption indicator (from the International Country Risk Guide) is intended to capture the likelihood that high ranked officials in government will demand special payments, and in addition the extent to which illegal payments are expected throughout government tiers (see Svensson, 2005). This indicator allows us to go back to 1991, the launch date of the WWW. The ICRG corruption variable goes from zero to 6; a higher score means *less* corruption.

Table 8 provides results. Note first that the instrument is very strong, c.f. columns 3 and 4. In addition, the Hausman test cannot detect any difference between OLS and 2SLS results. Hence, qualitatively, no results are altered using the ICRG corruption measure. Quantitatively, results also stay roughly the same using ICRG as compared to using CCI. The typical country in terms of Internet use and corruption in the large sample associated with column 1 is South Africa ($INTERNET_{2005} = 10.88$ and $ICRG_{2005} = 2.17$). Increasing the number of Internet users in South Africa to the level of Malta, i.e. moving from the median to the third quartile in the $INTERNET_{2005}$ distribution, would move South Africa to the 68th percentile in the corruption distribution (the level of Hungary), where we recall that higher scores means less corruption. Consequently, cross-country results are robust to changes in corruption measure.

- Table 8 about here -

6 Concluding remarks

The Internet is a powerful technology in the struggle to reduce corruption: it facilitates the dissemination of information about corrupt behavior, thus making it more risky for bureaucrats and politicians to take bribes, and it obviates the need for potentially corrupt officials to serve as middlemen between the government and the public. In this paper, we provide evidence of this hypothesis by documenting that the spread of the Internet has reduced corruption during its time in operation.

The correlation between changes in Internet penetration and changes in corruption, conditional on the initial level of corruption, is strong and robust. It holds across the world at large, and across different sub-samples of the world, using best available corruption perception indices as dependent variables. It also holds across U.S. states, using corruption convictions. Moreover, in the U.S. sample the risk of omitted variable bias tainting the OLS results is somewhat reduced, as this sample can be viewed as reasonable homogenous in many respects.

Nevertheless, in order to examine whether Internet penetration has had a causal impact on corruption, the paper develops a new instrument for the user cost of digital equipment such as computers, and thereby for Internet diffusion. Digital equipment is highly sensitive to power disruptions; and, to a considerable extent, lightning activity causes power disruption around the world. Indeed, according to some calculations, lightning causes some $17,000$ computers around the world to crash each second (Yeager and Stahlkopf, 2000).

Using global satellite data developed by NASA, and never used before in economic analysis, we construct country and state level measures for average lightning density. Using cross-country data, we document that lightning is highly correlated with electricity outages, and that the latter is associated with a slower pace of Internet penetration. This motivates the use of lightning density as an instrument for Internet diffusion. Empirically, lightning density turns out to be a strong instrument for Internet diffusion, especially in the U.S. sample. Our 2SLS estimates confirm that the spread of the Internet has lowered corruption, both in the U.S. and worldwide.

We are unable to reject that the spread of the Internet is exogenous in our specifications, though OLS estimates are considerably lower than the 2SLS counterparts. However, despite remaining conservative, by using OLS estimates in our subsequent calculations, we find the impact of the Internet on corruption to be economically substantial. Accordingly, to the extent that corruption affects economic growth, these findings provide one mechanism by which the Internet may work to spur growth. In this way, our findings provide a new perspective on the importance of the well-documented global "digital divide".

The identification strategy developed in this paper may prove useful in future research. By providing a strong instrument for Internet use, researchers may be able to make new progress on the impact of computers and the Internet on other outcomes, like the return to education or productivity growth more broadly.

References

[1] Acemoglu, D. and J. Robinson, 2001. Political Losers As a Barrier to Economic Development. *American Economic Review*, 90, 126-130

[2] Ades, A. and R. di Tella, 1999. Rents, Competition and Corruption. *American Economic Review*, 89, 982-993

[3] Aidt T., 2003, Economic Analysis of Corruption: A Survey. *Economic Journal,* 113, 632-651

[4] Bardhan, P., 1997. Corruption and Development: A Review of Issues. *Journal of Economic Literature,* 35, 1320-1346

[5] Beilock, R. and D.V. Dimitrova, 2003. An exploratory model of inter-country Internet diffusion. *Telecommunications Policy*, 27, 237-52.

[6] Besley, T. and R. Burgess, 2002. The Political Economy of Government Responsiveness: Theory and Evidence from India. *Quarterly Journal of Economics*, CXVII, 1415–1451

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

[8] Brunetti, A. and B. Weder, 2003. A free press is bad news for corruption. *Journal of Public Economics*, 87*,* 1801-1824.

[9] Caldwell, D., 2005. Unlocking the Mysteries of the Bounding Box. *Coordinates*: Online Journal of the Map and Geography Round Table, American Library Association, Series A. no. 2. Available online at <http://www.sunysb.edu/libmap/coordinates/seriesa/no2/a2.htm>

[10] Caselli, F. and W.J. Coleman III, 2001. Cross-Country Technology Diffusion: The Case of Computers. *American Economic Review*, 91, 328-335.

[11] Chernozhukov, V. and C. Hansen, 2005. The Reduced Form: A Simple Approach to Inference with Weak Instruments. Mimeo (Massachusetts Institute of Technology)

[12] Chinn, M.D. and R.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

[13] Chisholm, W., 2000. Lightning Protection. Chapter 4.10 in: *The Electrical Power Engineering Handbook*. Edited by Grigsby, L., IEEE Press.

[14] Chisholm, W., and Cummins, K., 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>

[15] Conroy, M., 1999. Power Quality and Your Computer. *Innovations*, March. Available online at: <http://www.copper.org/innovations/1999/03/power.html>

[16] Eisensee, T. and D. Stömberg, 2007. New Droughts, New Floods and U.S. Disaster Relief. Mimeo (Institute for International Economic Studies, Stockholm University). Forthcoming *Quarterly Journal of Economics.*

[17] 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>

[18] Ferraz, C. and F. Finan, 2008. Exposing Corrupt Politicians: The Effect of Brazil's Publicly Released Audits on Electoral Outcomes, with Frederico Finan. Forthcoming, *Quarterly Journal of Economics* (Available as IZA Discussion Paper No. 2836, June 2007).

[19] Glaeser, E. and R. Saks, 2006. Corruption in America. *Journal of Public Economics*, 90, 1053-1072

[20] Fredriksson, P. H. Vollebergh and E. Dijkgraaf, 2004. Corruption and energy efficiency in OECD countries: theory and evidence. *Journal of Environmental Economics and Management*, 47, 207-231

[21] Hall, R.E. and D. Jorgenson, 1967. Tax Policy and Investment Behavior. *American Economic Review*, 57, 391-414

[22] Hargittai, E., 1999. Weaving the Western Web: explaining differences in Internet connectivity among OECD countries. *Telecommunications Policy*, 23, 701-718

[23] 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

[24] Kaufmann, D., A. Kraay, and M. Mastruzzi, 2007. Governance Matters IV: Governance Indicators for 1996-2006. Available online at <http://www.worldbank.org/wbi/governance/wp-governance.html>

[25] Kressel, H., 2007. *Competing for the Future: How Digital Innovations are Changing the World*. Cambridge University Press

[26] Licht, A. N., C. Goldschmidt, and S. H. Schwartz, 2007. Culture Rules: The foundations of the rule of law and other norms of governance. *Journal of Comparative Economics*, 35, 659-88

[27] Mauro, P. 1995. Corruption and Growth. *Quarterly Journal of Economics* 110, 681-712

[28] McMillan, J. and P. Zoido, 2004. How to Subvert Democracy: Montesinos in Peru. *Journal of Economic Perspectives*, 18, 69-92

[29] Mokyr, J., 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press

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

[31] 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>

[32] Olken, B., 2006. Corruption Perceptions vs. Corruption Reality. *NBER Working Paper* #12428

[33] Olken, B., 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy,* 115, 200-249

[34] Oyelaran-Oyeyinka, B. and K. Lal, 2005. Internet diffusion in sub-Saharan Africa: A cross-country analysis. *Telecommunications Policy*, 29, 507-527

[35] Parente, S. and E. Prescott, 1999. Monopoly Rights: A Barrier to Riches. *American Economic Review*, 1216-1233

[36] Persson, T., G. Tabellini and F. Trebbi, 2003. Electoral rules and corruption. *Journal of the European Economic Association,* 1, 958-989.

[37] Reinikka, R. and J. Svensson, 2004. The power of information: Evidence from a newspaper campaign to reduce capture. Policy Research Working Paper Series 3239, World Bank

[38] Shim, J.K., A Qureshi and J.G. Sieqel, 2000. *The International Handbook of Computer Security.* Glenlake Publishing Company Ltd.

[39] Staiger, D., Stock, J., 1997. Instrumental variables with Weak Instruments. *Econometrica*, 65, 557-586

[40] Strömberg, D., 2001. Mass Media and Public Policy, *European Economic Review,* 45, 652-663

[41] Strömberg, D., 2004. Radio's Impact on Public Spending, *Quarterly Journal of Economics,* 119, 189-221

[42] Svensson, J., 2005. Eight Questions about Corruption. *Journal of Economic Perspectives,* 19, 19-42

[43] Treisman, D., 2000. The causes of corruption: a cross-national study. *Journal of Public Economics*, 76. p 399-457

[44] Yeager, K., Stahlkopf, K., 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>

[45] Wescott, C., 2003. E-government to combat corruption in the Asia Pacific Region. Available online at <http://www.adb.org/Governance/egovernment_corruption.pdf>

[46] WDI, 2007. *World Development Indicators 2007.* World Bank

[47] West, D., 2005. *Digital Government.* Princeton University Press

Table 1: Summary statistics

	Obs.	Mean	Median	Std. dev.	Min	Max
	Panel A: Cross-country sample					
CCI_{2006}	113	0.105	-0.251	1.080	-1.767	2.573
CCI_{1996}	113	0.112	-0.169	1.057	-2.084	2.297
$INTERNET_{2005}$	113	22.56	15.24	21.66	0.208	86.94
$INTERNET_{1996}$	113	1.541	0.196	3.216	0	18.26
GYCAP	106	0.272	0.191	0.294	-0.351	1.416
Average lightning density	113	8.530	5.309	8.328	0.021	40.13
Outages	93	27.5	9.6	46.6	0.04	248.96
	Panel B: U.S. sample					
CC_{2006}	50	19.50	9	21.28	0	83
CC_{1991}	50	12.94	5	20.85	0	108
$INTERNET_{2003}$	50	60.10	59.80	6.055	42.60	71.55
GYCAP	50	0.356	0.354	0.064	0.170	0.580
Average lightning density	50	10.13	8.436	7.449	0.126	27.34

Notes: Average lightning density is the average flash rate density, calculated using the Minimum Bounding Rectangle of the Optical Transient Detector-based 5-year total lightning variable.

Table 2: Ordinary Least Squares on cross-country sample

Dependent variable: DCCI	(1)	(2)	(3)
DINTERNET		0.016***	0.016***
		(0.004)	(0.004)
CCI_{1996}	-0.070**	-0.287***	-0.291***
	(0.028)	(0.065)	(0.066)
GYCAP			0.171
			(0.141)
CONSTANT	0.001	-0.307***	-0.364***
	(0.043)	(0.087)	(0.091)
Observations	113	113	106
R-squared	0.03	0.22	0.28

Notes: Dependent variable is the change in Control of Corruption, DCCI, over the period 1996 to 2006. The explanatory variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***,**,* indicate significance at the 1, 5, and 10% level, respectively.

Table 3: Ordinary Least Squares on cross-country subsamples

Dependent variable: DCCI	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DINTERNET	0.013***	0.016***	0.016***	0.015***	0.019***	0.016***	0.016**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)
CCI_{1996}	-0.235***	-0.285***	-0.286***	-0.275***	-0.339***	-0.301***	-0.291***
	(0.064)	(0.066)	(0.065)	(0.065)	(0.063)	(0.069)	(0.106)
CONSTANT	-0.221**	-0.305***	-0.306***	-0.334***	-0.334***	-0.325***	-0.317**
	(0.085)	(0.088)	(0.090)	(0.090)	(0.090)	(0.099)	(0.122)
Observations	91	112	110	100	100	90	75
Excluded region	SSA	NA	SOA	MENA	EAP	LAC	ECA
R-squared	0.17	0.22	0.22	0.24	0.28	0.26	0.16
F-test	0.389	0.995	0.990	0.877	0.385	0.895	0.984
[H0: DINTERNET = 0.016, p-value]							

Notes: Dependent variable is the change in Control of Corruption, DCCI, over the period 1996 to 2006.The explanatory variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***, **, * indicate significance at the 1, 5, and 10% level, respectively. The regional key is as follows: Sub-Saharan Africa (SSA), North America (NA), South Asia (SOA), Middle East & North Africa (MENA), East Asia & Pacific (EAP), Latin America & the Caribbean (LAC), Europe & Central Asia (ECA). The F-test tests the null that the slope estimate of DINTERNET associated with each subsample is equal to the slope estimate associated with the full sample, i.e. column 2 in Table 2.

Table 4: Ordinary Least Squares on U.S. states

Dependent variable: DCC	(1)	(2)	(3)
DINTERNET		-0.061**	-0.060**
		(0.025)	(0.025)
$\log(1+\mathrm{CC}_{1991})$	-0.388***	-0.453***	-0.464***
	(0.089)	(0.088)	(0.087)
GYCAP			-1.131
			(1.965)
CONSTANT	1.257***	5.019***	5.424***
	(0.252)	(1.541)	(1.604)
Observations	50	50	50
R-squared	0.21	0.31	0.32

Notes: Dependent variable is the change in log(1+number of corruption convictions), DCC, over the period 1991 to 2006. The explanatory variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***,**,* indicate significance at the 1, 5, and 10% level, respectively.

Table 5: Zeroth Stage Regressions

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	log(OUTAGES)	log(OUTAGES)	DINTERNET	DINTERNET	DINTERNET
log(OUTAGES)			-6.219***	-5.284***	-5.221***
			(0.768)	(1.007)	(1.021)
LIGHTNING	0.537***	0.385**			-0.481
	(0.162)	(0.163)			(1.330)
CCI_{1996}		-0.701***		5.001***	4.759***
		(0.193)		(1.744)	(1.998)
CONSTANT	1.162***	1.126***	28.04***	26.79***	27.48***
	(0.358)	(0.324)	(2.139)	(2.277)	(3.090)
Observations	93	86	70	67	67
R-Squared	0.13	0.25	0.57	0.61	0.61

Notes: Dependent variable in columns (1) and (2) is the log of outages; the dependent variable in columns (3) to (5) is the change in Internet use over the period 1996 to 2006. All variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***,**,* indicate significance at the 1, 5, and 10% level, respectively.

Table 6: Two-Stage Least Squares on cross-country sample

	(1)	(2)	(3)
Dependent variable: DINTERNET	Panel A: First-stage regression		
CCI_{1996}	11.402***	12.072***	12.090***
	(1.375)	(1.307)	(1.229)
INSTRUMENT 1	-3.353***		
	(1.222)		
INSTRUMENT 2		-3.599***	
		(1.218)	
INSTRUMENT 3			-3.633***
			(1.276)
CONSTANT	24.912***	26.180***	26.437***
	(2.469)	(2.760)	(2.892)
Breusch-Pagan/Cook-Weisberg test	0.018	0.018	0.019
[H0: constant error variance, p-value]			
F (first-stage) value	7.530	8.730	8.110
[H0: INSTRUMENT = 0]			
Dependent variable: DCCI	Panel B: Second-stage regression		
DINTERNET	0.031**	0.032**	0.029**
	(0.013)	(0.013)	(0.011)
CCI_{1996}	-0.496***	-0.504***	-0.469***
	(0.184)	(0.188)	(0.165)
CONSTANT	-0.602**	-0.613**	-0.564**
	(0.244)	(0.249)	(0.226)
Observations	113	113	113
R-squared	0.04	0.03	0.08
Pagan-Hall test	0.864	0.612	0.913
[H0: constant error variance, p-value]			
95% Chernozhukov-Hansen confidence intervals for DINTERNET	[0.006, 0.084]	[0.007, 0.080]	[0.005, 0.071]
Hausman test	0.317	0.338	0.411
[H0: DINTERNET is exogeneous, p-value]			

Notes: Dependent variable is the change in Control of Corruption, DCCI, over the period 1996 to 2006. The explanatory variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***,**,* indicate significance at the 1, 5, and 10% level, respectively. Note: INSTRUMENT 1 is the MBR-based average lightning density for a country, INSTRUMENT 2 is the lightning density in the capital of a country, and INSTRUMENT 3 is the lightning density at the midpoint of the country MBR

Table 7: Two-Stage Least Squares on U.S. sample

	(1)	(2)	(3)	(4)	(5)
Dependent variable: DINTERNET	Panel A: First-stage regression				
$\log(1+CC_{1991})$	-0.351	-0.566	-0.772	-0.569	-0.770
	(0.605)	(0.533)	(0.487)	(0.569)	(0.510)
INSTRUMENT 1	-2.844***				
	(0.851)				
INSTRUMENT 2		-2.340***		-3.201***	
		(0.845)		(0.991)	
INSTRUMENT 3			-3.486***		-4.195***
			(1.199)		(1.127)
CONSTANT	65.487***	65.588***	68.452***	67.576***	70.071***
	(1.600)	(1.786)	(2.554)	(2.065)	(2.311)
Breusch-Pagan/Cook-Weisberg test	0.848	0.675	0.890	0.805	0.681
[H0: constant error variance, p-value]					
F (first-stage) value	13.53	10.89	12.82	13.51	15.76
[H0: INSTRUMENT = 0]					
Dependent variable: DCC	Panel B: Second-stage regression				
DINTERNET	-0.121***	-0.085*	-0.085*	-0.103**	-0.090**
	(0.043)	(0.050)	(0.045)	(0.049)	(0.044)
$\log(1+CC_{1991})$	-0.494***	-0.480***	-0.480***	-0.478***	-0.467***
	(0.096)	(0.097)	(0.096)	(0.096)	(0.095)
CONSTANT	8.700***	6.550**	6.547**	7.590**	6.825**
	(2.688)	(3.083)	(2.818)	(3.044)	(2.723)
Observations	48	50	50	48	48
R-squared	0.23	0.29	0.29	0.28	0.30
Pagan-Hall test	0.532	0.443	0.449	0.511	0.479
[H0: constant error variance, p-value]					
Hausman test	0.416	0.875	0.817	0.942	0.716
[H0: DINTERNET is exogeneous, p-value]					

Notes: Dependent variable is the change in Control of Corruption, DCCI, over the period 1996 to 2006. The explanatory variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***,**,* indicate significance at the 1, 5, and 10% level, respectively. Note: INSTRUMENT 1, the cloud-to-ground lightning density based on U.S. National Lightning Detection Network; INSTRUMENT 2, the MBR-based lightning density of a state; and INSTRUMENT 3, the lightning density at the midpoint of the state MBR.

Table 8: Robustness analysis on cross-country sample

Dependent variable: DICRG	(1)	(2)	(3)	(4)
	OLS		2SLS	
$INTERNET_{2005}$	0.036***	0.031***	0.043***	0.042***
	(0.006)	(0.005)	(0.009)	(0.010)
$ICRG_{1991}$	-0.747***	-0.654***	-0.811***	-0.780***
	(0.088)	(0.083)	(0.109)	(0.128)
GYCAP		-0.038		-0.028
		(0.216)		(0.216)
CONSTANT	0.981***	0.750***	1.066***	0.950***
	(0.231)	(0.204)	(0.239)	(0.261)
Observations	102	94	102	94
R-squared	0.57	0.48	0.56	0.45
F (first-stage) value			20.94	18.25
[H0: INSTRUMENT = 0]				
Hausman test			0.685	0.801
[H0: INTERNET is exogeneous, p-value]				

Notes: The dependent variable is the change in the ICRG's index over the period 1991 to 2005, DICRG. The explanatory variables are described in the main text. Robust standard errors are reported in parentheses. Asterisks ***, **, * indicate significance at the 1, 5, and 10% level, respectively. The instrument used is the log of the MBR based average lightning density.

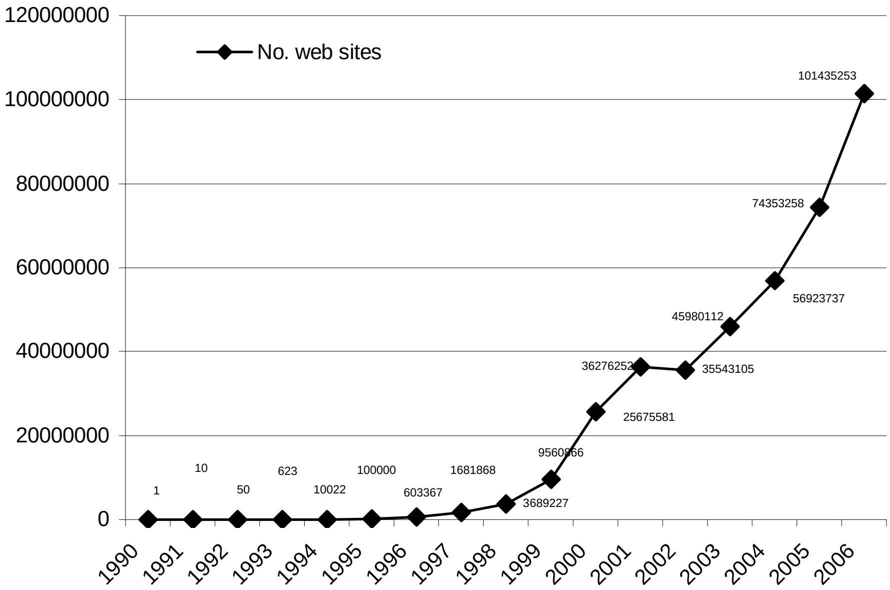

Figure 1: The figure shows the total number of Web sites on the Internet, 1990-2006. Generally, the numbers refer to December 1st in the individual year. *Source*: Hobbes' Internet Timeline v8.2 <http://www.zakon.org/robert/internet/timeline/>.

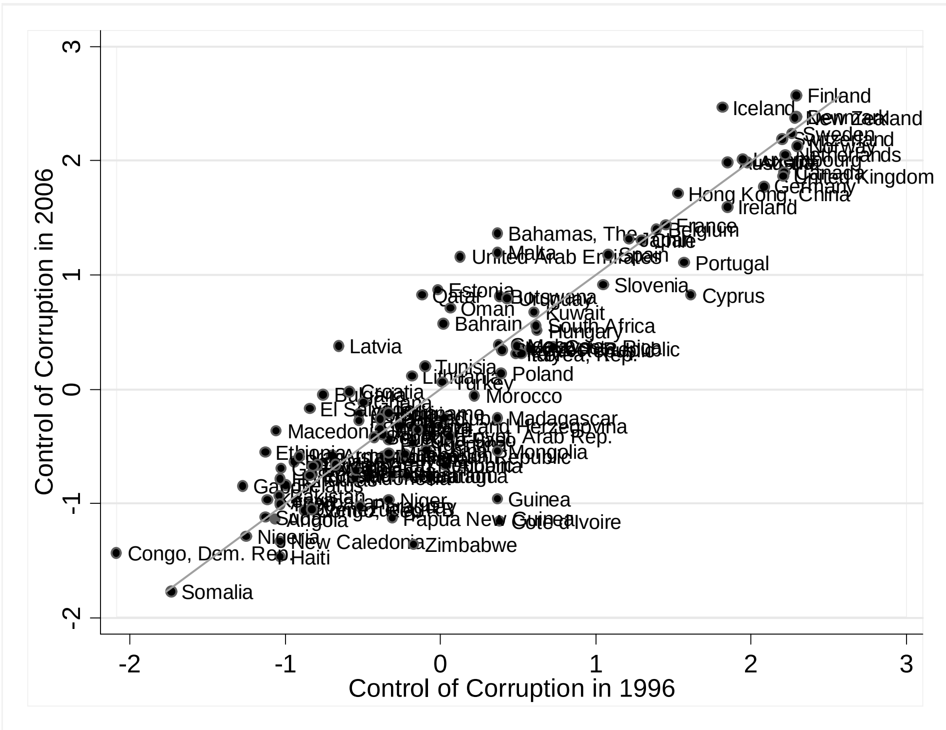

Figure 2: The figure shows Control of Corruption in 1996 (horizontal axis) versus Control of Corruption in 2006 (vertical axis). The full line is the 45-degree line. The sample used is the largest estimation sample used in the empirical analysis below. The number of observations is 117.

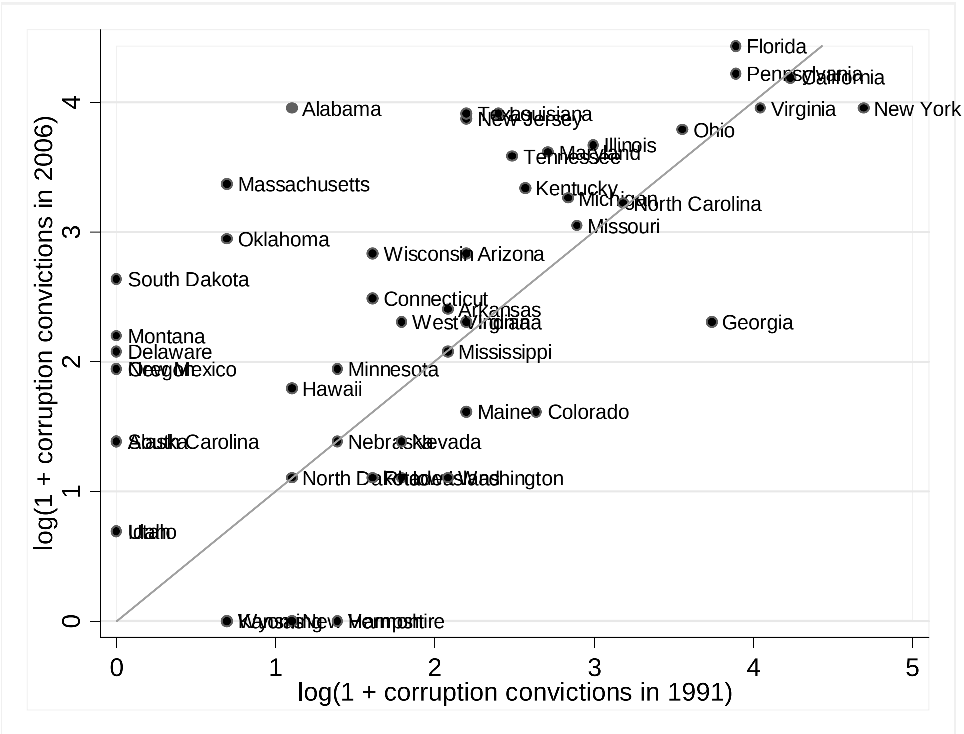

Figure 3: The figure shows a scatter plot of log(1 + corruption convictions in 1991) (horizontal axis) versus log(1 + corruption convictions in 2006) (vertical axis). The full line is the 45-degree line. Number of observations is 50.

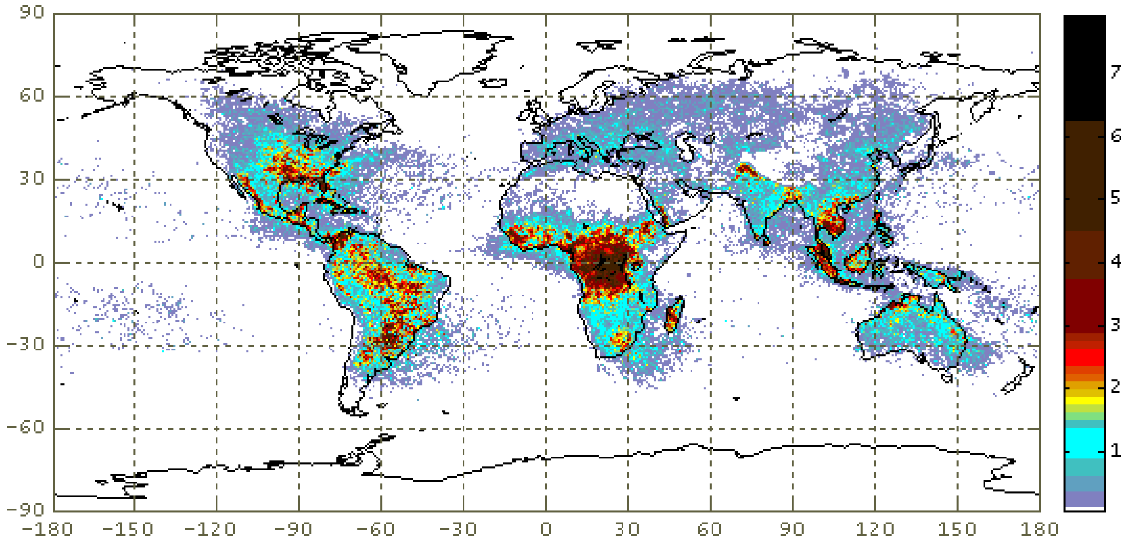

Figure 4: Average flash density (flashes per year per km^2). The figure is constructed using the OTD Global Lightning Distributions for the period April 12, 1995 to December 31, 1999.

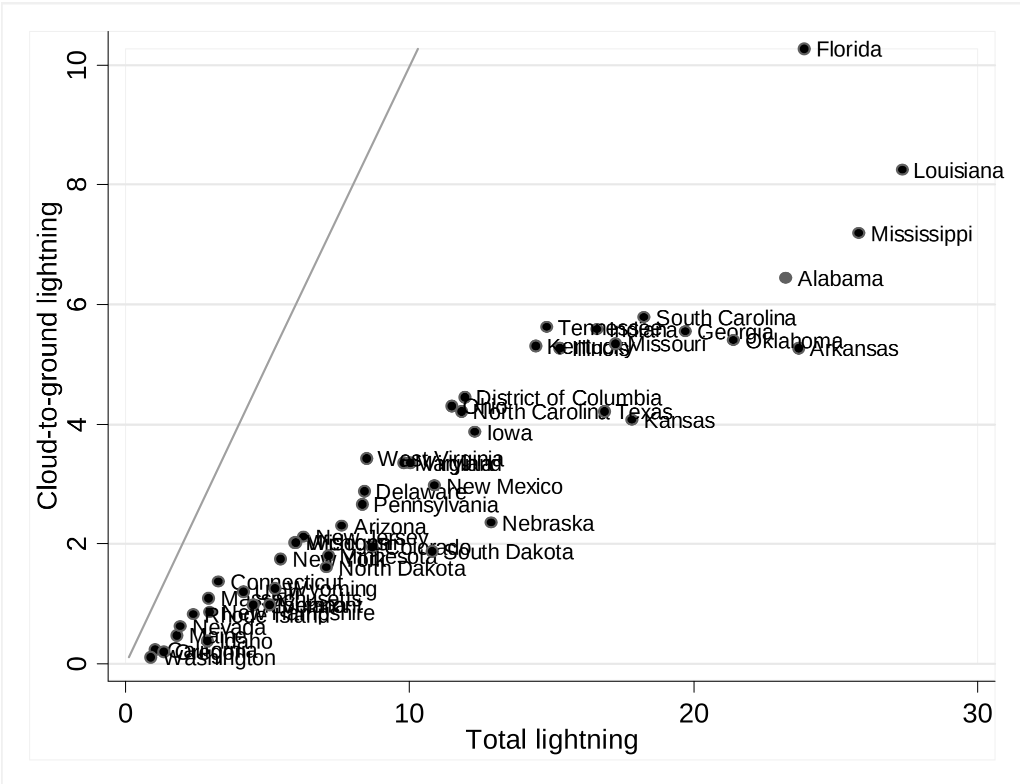

Figure 5: The figure shows a scatter plot of total lightning flash rate density (flashes per km^2 per year) based on the Optical Transient Detector high resolution data (horizontal axis) versus cloud-to-ground flash rate density (flashes per km^2 per year) based on the U.S. National Lightning Detection Network (vertical axis). The full line is the 45-degree line. Number of observations is 48.

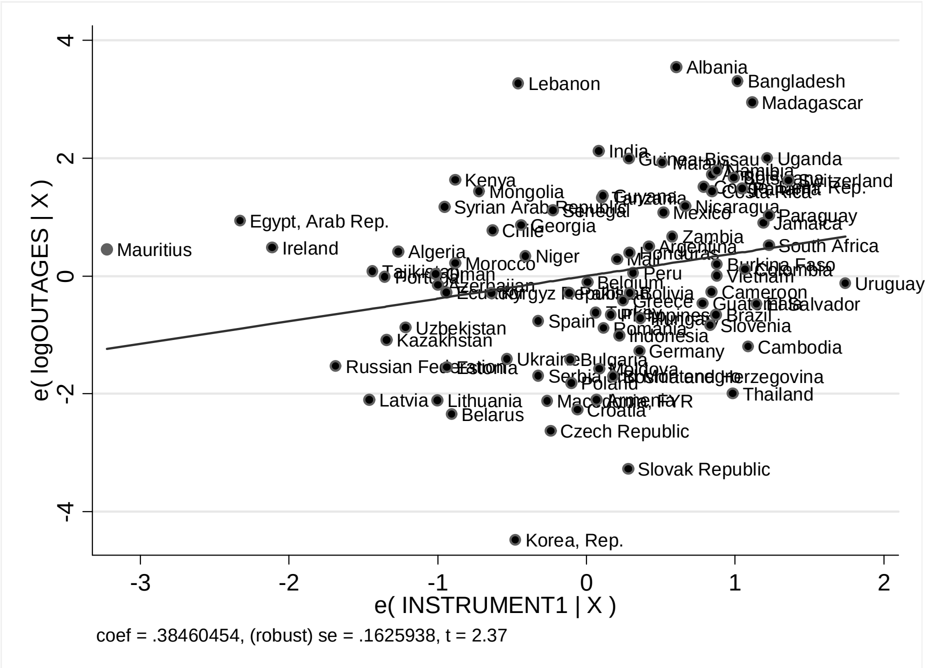

Figure 6: The figure shows the relationship between average lightning density and electrical outages once the influence of corruption in 1996 is partialled out.

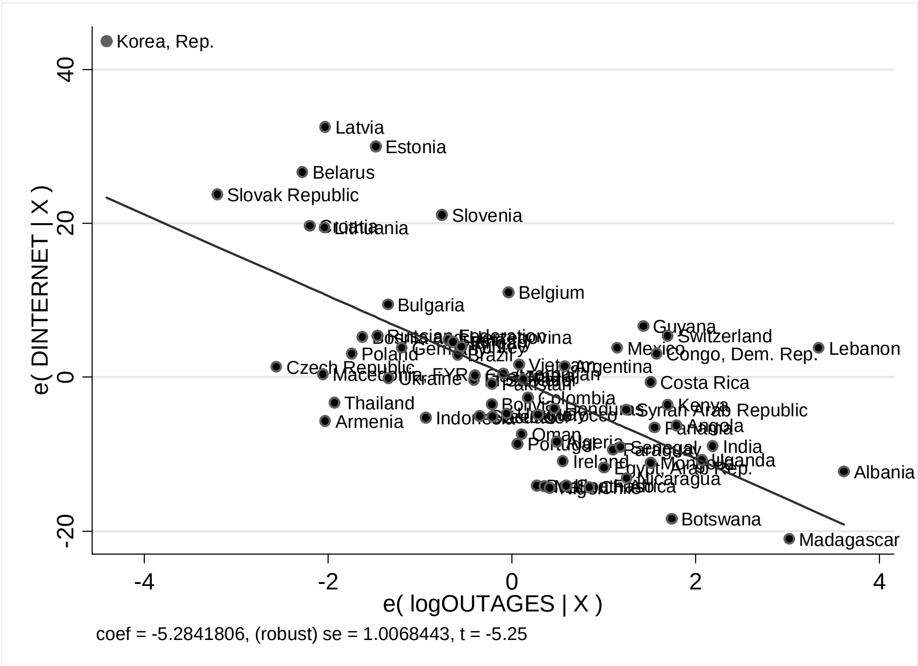

Figure 7: The figure shows the relationship between electrical outages and changes in Internet use once the influence of corruption in 1996 is partialled out.