E. J. Ourso College of Business
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Social Media, Internet and Corruption

Chandan Jha
Louisiana State University

Sudipta Sarangi
DIW Berlin & Louisiana State University

Working Paper 2014-03
http://bus.lsu.edu/McMillin/Working_Papers/pap14_03.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

Social Media, Internet and Corruption*

Chandan K. Jha[1] and Sudipta Sarangi[2]

Department of Economics, Louisiana State University, Baton Rouge, LA 70803.

[1]e-mail: cjha1@lsu.edu, [2]e-mail: sarangi@lsu.edu

January 2014

Abstract

In this paper we study the relationship between multi-way means of communication and corruption. Unlike traditional platforms like TV or print media, which only provide a one-way channel of communication, the internet and social media platforms provide for two-way flow of information. Using Facebook as a proxy for social media, we show that Facebook penetration and corruption are negatively associated. The same holds for internet penetration. We then exploit variations in cross-country technological adoption in the field of communication in 1500 AD to address endogeneity concerns. We show that internet penetration and Facebook penetration have a causal and negative impact on corruption. Our results also suggest that these effects are sizable making them effective tools against corruption.

Keywords: Corruption, Transparency of Information, Facebook, Internet, Social Media

JEL Classification Codes: D73, D83, O1, H0

*We would like to thank Areendam Chanda, Cary Deck, Ozkan Eren, Gautam Hazarika, Carter Hill, Munechika Katayama, Douglas McMillin, Naci Mocan, Bibhudutta Panda and Bulent Unel for their helpful and constructive comments and suggestions.

1 Introduction

The role of "liberation technology"[1] such as internet, mobile phones and social media has been recognized in empowering individuals, increasing their participation in political process, facilitating communication and mobilization on social issues, and strengthening an emergent civil society (Diamond, 2010; Saleh, 2012). It has been shown that greater access to information is negatively associated with the level of corruption in a country (DiRienzo et al., 2007). Brunetti and Weder (2003) argue that a free press reduces the cost of fighting corruption and show that the countries, where the press enjoys greater freedom, are less corrupt. Traditional media (print and broadcast media) provides only one-way communication and has often been subject to censure and control by the authoritarian regimes, typically by either monopolizing or regulating the print and broadcast media. Clamping down or controlling digital media is much more difficult since they allow for multi-way communication of information. Indeed, the spread of internet challenged the monopoly of authoritarian governments on information by making it easily available to the public and even leading to a change in regimes in some instances.[2] Inspired by these facts, this paper explores the possibility that internet penetration and increasing use of social media may have a negative impact on corruption.

Studies exploring the effect of communication technology (more specifically internet penetration and social media usage) on corruption are scarce. There are two important factors behind the scarcity of these studies – first, the lack of variation in corruption indices over time; and second, concerns for endogeneity and the lack of a valid instrument. The former makes a panel analysis of studies regarding corruption almost impossible, and hence, exacerbates the concerns for endogeneity. On the other hand, internet penetration and social media can be endogenous to corruption – not only because of the possibility of omitted variable bias, but also, because of concerns of reverse causality. For example, a corrupt government might discourage investments in the adoption of technologies that might have an adverse impact on corruption.

[1] The term liberation technology comes from Diamond (2010) who defines it as "any form of information and communication technology (ICT) that can expand political, social and economic freedom".

[2] Often cases of corruption, human rights violations and police brutality have been taboo subjects for citizens and have been censored in a number of countries such as China, Malaysia and Iran.

We bridge this gap in the literature by using a newly constructed variable on historical technological adoption from the Cross-country Historical Adoption of Technology (henceforth, CHAT) dataset (Comin and Hobijn; 2009). There is considerable cross-country variation in technology adoption in 1500 AD. Comin et al. (2010) compute indices of technology adoption in 1500 AD in five different sectors – agriculture, transportation, military, industry and communication. They present robust evidence of a positive and significant association between these indices and technology adoption today. We exploit the variation in cross-country communication technology adoption in 1500 AD, and its association with technology adoption today to identify the impact of internet penetration and Facebook penetration on corruption. Specifically, we use technology adoption in communication in 1500 AD as an instrument for internet penetration (technology adoption) today.

Our results show that there is a causal and negative impact of internet penetration and social media usage on corruption. We show that our instrument is strong even when we control for a number of institutional, cultural, historical and contemporary variables that might potentially be correlated with past and present levels of corruption, and past and current levels of technology adoption. Moreover, we perform a variety of robustness exercise that may render the validity of our instrument questionable. Potential concerns about the validity of instruments have been discussed at length in instrumental variables (IV) analysis section. We now briefly review the existing literature to make clear what this paper contributes to it.

Our paper follows a rich existing literature studying corruption across countries. In his seminal work, Daniel Treisman (2000) identifies several factors that determine the level of (perceived) corruption in a country. Using the annual indexes for $1996 - 1998$ of Corruption Perception Index (CPI) published by Transparency International (TI), Treisman empirically estimates the explanatory power of the theoretically plausible determinants of corruption. The findings of his paper suggest that countries with Protestant traditions, long democratic exposure and British colonial histories are less corrupt. In another significant paper, Brunetti and Weder (2003) investigate the impact of the press freedom on corruption. They classify press freedom as an external mechanism to control corruption – a control exercised by individuals or organizations that are outsiders to the network of corruption, *i.e.* the bureaucratic system. They argue that press freedom puts a check

on corruption by reducing the cost of fighting extortive corruption as well as collusive corruption. According to Brunetti and Weder a free press is especially more effective in fighting collusive corruption where the internal controls of corruption – the agencies that control corruption within bureaucracy – are likely to be less effective.[3]

In this paper, we argue that the internet and the social media too act as external controls of corruption and help reduce the cost of fighting corruption in several ways. First, larger internet penetration and the spread of social media would mean a larger audience for the victims of extortive corruption who wish to share the incident of corruption. Second, the internet and social media provide cheap and speedy means of sharing information and reaching a larger audience to organize public protests against the corrupt activities of the government officials and politicians.[4] Third, internet can be used to provide electronic-government, or in short, e-government services which eliminate the need of direct interaction between the citizens and public officials, reducing the scope of bribe demand (Bhatnagar, 2003; Andersen, 2009). Based on the above facts, we hypothesize the following:

H1: Corruption will be lower in countries with a greater Facebook penetration.

H2: Corruption will be lower in countries with a greater internet penetration.

While there are a few studies that have investigated the direct link between internet penetration and corruption (Andersen et al., 2011), and also the impact of internet facilitated services such as e-government on corruption (Andersen, 2009), to the best of our knowledge, no one has studied the impact of social media on corruption. This paper is the first to investigate and quantify the impact of social media on corruption. We also re-visit the association between internet penetration and corruption by using an exogenous instrument for the internet penetration.

Internet and the social media play a complementary role and augment the effect of a free press

[3]See Brunetti and Weder (2003) for an excellent discussion of the two kinds of corruption and how a free press is an effective tool against these kinds of corruption.

[4] A victim of extortive corruption may share the corruption incident on social networking sites to mobilize support for the fight against corruption. For instance, an Indian non-government organization, *"Janaagraha"* runs a website where people can share their detailed experience regarding corruption, and uses the information "to argue for improving governance systems and procedures, tightening law enforcement and regulation and thereby reduce the scope for corruption in obtaining services from the government". The website address is: http://www.ipaidabribe.com/. A Facebook page called *"India Against Corruption"* was used by social activists in India to mobilize protests against corruption.

(traditional print and broadcast media) on corruption in several ways. Greater internet penetration and a larger proportion of population using social media would result in the news of a free press reaching a larger proportion of the population. In addition, it takes much longer for news to reach to the public via print media while on-line news becomes instantly available to the public. Next, internet and social media provide a platform for everyone to share their experiences via the use of blogs, and social media websites such as Facebook, Twitter, Google Plus and YouTube where the marginal use of these resources is costless. Finally, interaction in social media platforms is often among friends and family and the personal touch to information from such sources may give it more credibility. Individuals might feel more compelled to act on such information just to show solidarity with their near and dear ones. All of these factors taken together make it important to study the impact of these multi-way flow channels of communication on corruption.

The rest of the paper is organized as follows. In the next section, we briefly describe our data sources. Section 3 outlines the empirical strategy. In section 4, we present OLS results. Section 5 deals with the endogeneity concerns and presents IV results. Section 6 concludes.

2 Data

This paper uses the 2011 Control of Corruption Index (CCI) published by the World Bank as the primary measure of corruption (*Corruption*).[5] Kaufmann et al. (2010) describe the objective of the CCI as – "capturing perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests." The CCI takes values in the range of -2.5 to 2.5, with -2.5 representing the most severe corruption and 2.5 representing the lowest level of corruption.

The data source for our one of the main variables of interest – Facebook penetration (*Facebook*) is *'Quintly',* a social media benchmarking and analytics solution company. The data was accessed from its website (http://www.quintly.com/facebook-country-statistics?period=1year) on May 10, 2013. *'Quintly'* uses the Facebook advertising tool in order to collect data on Facebook

[5] The results are robust to the use of an alternative measure of corruption – the 2011 Corruption Perception Index published by Transparency International that defines corruption as "the misuse of public power for private benefit".

users in different countries. Even though the advertising tool belongs to Facebook, *Quintly* data has to be interpreted carefully since Facebook claims that these data are not completely accurate. The official number of Facebook users may be slightly different from the number of Facebook users indicated by this advertising tool. For example, Facebook reported 1.11 billion users at the end of March, 2013 while the number of users reported by the advertising data was 965 million, or a difference of roughly 13 percent indicating that "[this] advertising data can be seen as rough estimates for the Facebook country statistics. No more, no less." (Nierhof, 2013). Data for the other variable of interest – internet penetration ($Internet$) comes from the World Bank.

The data source for our instrument is Comin et al. (2010) who use a number of historical sources of information to compute an index of cross-country technology adoption in 1000 BC, 0 AD, and 1500 AD. They note that 1500 AD data is more precise because of a large number of sources documenting the technology adoption patterns. Our instrument is technology adoption in communication following the fact that internet is a part of communication industry. The construction of communication index uses four variables – 'the use of movable block printing', 'the use of woodblock printing', 'the use of books' and 'the use of paper'. The index is created on the basis of the extensive margin of technology adoption and not the extent to which these technologies were used (i.e. intensive margin). The index takes a value between 0 and 1.[6]

We use Gross Domestic Product (GDP) as a measure of income. The data for GDP per capita ($\log(GDPPC)$) has been taken from the World Bank. Freedom House publishes data on political rights ($Pol\,Rights$) and press freedom ($Press$). The political rights index can take a value from 1 through 7, with a lower value representing better political rights. A very high rating such as 7 would imply that political rights of the citizens are severely compromised. Such countries may be characterized by severe government oppression, absence of a functioning central government and widespread extreme violence. Each country is given a score from 0 to 100 in the press freedom index, with a lower score implying a freer press. The 2011 press freedom index for a country

[6]The construction of this data takes into account the fragmentation and unification of countries subsequent to the period to which the data belongs. Also, note that the technology adoption data in 1500 AD are estimated before colonization began, and hence, does not incorporate technology transferred by Europeans to the rest of the world. Interested readers may refer to Comin and Hobijn (2009) and Comin et al. (2010) for the details about the methodology of constructing the index and the sources of information that have been used to construct these indices.

is determined on the basis of its performance in three broad categories: the legal, political and economic environment. The legal environment takes into account the extent to which freedom of expression of individuals including journalists and bloggers are protected by the law, and the extent of freedom granted by law to media regulatory bodies. The political environment category ranks a country on the basis of the following criteria – editorial independence, official and self-control of state-owned and private media, public access to media coverage, and the ability of local and foreign journalists to publish news freely and without harassment from agents of the state or others. Finally, the economic environment consists of the cost associated with establishing media agencies and government's control over media, transparency and concentration of media ownership, and the extent to which journalists and bloggers are influenced by economic incentives from private or public sources.

Population ($log\,(Population)$) data comes from the World Bank. The data for the proportion of population belonging to Christian ($Christian$) and Muslim ($Muslim$) faiths in the total population is available from the Association of Religion Data Archive.[7] The World Bank also provides data for the number of cellphone subscribers ($Cell$) per capita. The data on urban population ($Urban$) comes from the World Bank and refers to the percentage of population living in the areas which have been defined as 'urban' by the country's national statistical office. Average years of schooling ($Education$) has been used a measure of educational attainment for which the data source is Barro and Lee (2013). The share of imports in GDP is used as a measure of openness ($Openness$) using data from the World Bank. Summary statistics for the variables are presented in Table 1.

[7] http://thearda.com. Following are the principal investigators for this data set: Jaime Harris, Robert R. Martin, Sarah Montminy, and Roger Finke of The Association of Religion Data Archive.

3 Empirical Strategy

In order to assess the impact of internet penetration on corruption we estimate the following equation using ordinary least squares (OLS),

$$\begin{aligned} Corruption_i = \zeta + \alpha\, Internet_i + \gamma_1\, \log(GDPPC_i) + \gamma_2\, Pol\, Right_i \\ +\gamma_3\, \log(Population_i) + \gamma_4\, Christian_i + \gamma_5\, Muslim_i + \gamma_6\, Urban_i + \epsilon_i \end{aligned} \tag{1}$$

Next, we add the Facebook penetration in the model to investigate the impact of Facebook penetration on corruption as below

$$\begin{aligned} Corruption_i = \eta + \beta\, Facebook_i + \delta\, Internet_i + \gamma_1\, \log(GDPPC_i) + \gamma_2\, Pol\, Right_i \\ +\gamma_3\, \log(Population_i) + \gamma_4\, Christian_i + \gamma_5\, Muslim_i + \gamma_6\, Urban_i + \varepsilon_i \end{aligned} \tag{2}$$

where the subscript i denotes country i. Note that we use the negative of the corruption index so that a higher value implies a higher corruption. Hence, the expected signs of α, β and δ are negative.

We control for (log of) GDPPC since rich countries can afford to have better institutions and therefore may be able to control corruption more effectively (Treisman, 2000). As a result the coefficient of $\log(GDPPC)$ is expected to be negative. Corruption may be less in countries with long democratic history (Treisman, 2000). In countries where people enjoy higher political rights, the press is free and the judiciary is independent, corruption is likely to be lower. Therefore, we control for the political rights as a measure of democratic and political institutions. We use the negative of the political rights in all our specifications so that a higher number indicates higher political rights. Thus, expected sign of the coefficient of the *Pol Rights* is negative. In countries with very large population the cost of controlling corruption may be high. Consequently, we also control for the (log of) population. We control for the proportion of population belonging to Christian and Muslim faiths in order to capture cultural aspects of corruption (Swami et al., 2001). It is often argued that internet penetration would be higher in urban areas or places with high

population density. Also, urbanization in an area may potentially be correlated with corruption. Hence, we control for the proportion of population living in urban areas.

In order to investigate the impact of Facebook penetration on corruption we estimate the regression specification given by equation 2. Note that in equation 2 we also control for internet penetration in order to separate the social media effect on corruption from the total impact of internet penetration on corruption. If we omit internet penetration from the model, the estimates of Facebook penetration on corruption will be biased upward.

4 Results

4.1 OLS estimates

Table 2 presents the OLS estimates of the specification given in equation 1. The coefficient of internet penetration is highly significant and has the expected sign in the baseline specification (column 1). In order to minimize the omitted variable bias concern, we control for a number of variables in the next four columns. First, we replace the political rights index with the press freedom index in the baseline specification and present the results in column 2. The reason for doing so is that higher internet penetration may be capturing a part of the effect of press freedom on corruption when press freedom is omitted from the model. We use negative of the press freedom index and therefore expect the coefficient of this variable to be negative. Note that while the coefficient of internet penetration is smaller now, it is still highly significant and has the expected negative sign. In column 3, we control for cellphone penetration as they are also used as means of communication. This does not seem to have much of an impact on the coefficient of internet penetration. It has been argued that countries with higher human capital are less corrupt. Ades and Di Tella (1999) find that more open countries are less corrupt. We find that our results are robust to the inclusion of years of schooling (column 4) and openness to trade (column 5).

In Table 3 we present the results of the specification given by equation 2 to investigate the impact of Facebook penetration on corruption. In all the columns, the coefficient of the Facebook penetration is significant at conventional levels and has the negative sign suggesting a negative

association between the number of Facebook users and corruption. As expected the coefficient of internet penetration in each column of Table 3 is smaller than that of Table 2 since the impact of Facebook penetration is captured by internet penetration when the former is not controlled for.[8] Clearly, this is because internet can affect corruption in several other ways than just communication of information through social media.

The OLS estimates may, however, be biased because of potential endogeneity concerns. First, in a cross-country specification the possibility of omitted variable bias can not entirely be ruled out. Second, a corrupt government might restrict the use of internet making it endogenous to the model. Hence, an instrument for internet penetration is required to establish the causal impact of internet penetration and Facebook penetration on corruption. We address this issue in the next section.

5 Potential Endogeneity and IV Analysis

Different countries exert control on the internet in different ways. While some may curb internet usage, others may restrict content, and yet others may allow internet usage only to monitor the activities of its citizens on the internet. A number of countries have begun regulating and/or censoring the content that can be published and shared on the internet using different mechanisms. For example, countries like China and Iran use a very sophisticated and multi-layered censoring system in order to restrict the content that can be accessed by citizens (Freedom House, 2009). Another example is the Tunisian government creating focal points of control in order to censure the internet prior to 2011 (Wagner, 2012). In less oppressive regimes such as Egypt, Russia and Malaysia where the internet was of paramount importance in creating a free speech environment, vague and flexible security laws are used in order to intimidate bloggers and thereby suppress anti-government ideas from spreading over the internet (Freedom House, 2009).

In order to deal with the issue of endogeneity and establish causality, an exogenous instrument

[8] We also perform median and robust regressions in order to ensure that our estimates are not driven by outlying observations. As additional robustness exercise, we check the sensitivity of our results to the use of an alternative measure of corruption – Corruption Perception Index published by Transparency International. These results are consistent with the results reported above and have been omitted for the sake of brevity. Interested readers may refer to the Appendix of the working paper version of the paper.

is required. A valid instrument should be highly correlated with internet penetration today, and at the same time, should not have an impact on corruption via any channel other than internet penetration.

Comin et al. (2010) show that technological differences in 1500 AD are persistent even today. There can be several reasons for why this may be the case. These include but are not limited to the lower cost of and higher benefits from the adoption of new technologies resulting from the complementarity between the existing and new technology, cross-sectoral technological spillovers, lower cost of innovation and adoption, economies of scale, economies of scope of general purpose technologies, learning by doing and using a combination of old technologies to make new ones (See Comin at al., 2010 and references therein). Thus, we use the measure of technology adoption in communication in 1500 AD created by Comin et al. (2010) as an instrument for internet penetration today.

Table 4 reports the results of the IV analysis. Column 1 in panel 1, shows that communication technology in 1500 AD is a significant predictor of internet penetration today. In panel 2, we present the IV estimates. The coefficient of internet penetration is highly significant and has a negative sign. It suggests that a one percentage point increase in the internet penetration is associated with an improvement in the corruption index by about 0.027 points, which is sizable given that mean of the index is 0.04 for the sample of 106 countries included in this regression. Next, we address potential concerns regarding the validity of our instrument and reliability of the coefficients.

First, there may be concerns that technology in 1500 AD might be correlated with institutions in 1500 AD, which might be correlated with institutions today. If this is true, omission of an institutional variable may bias our estimates. We address this concern by controlling for the colonial status of a country (column 2) as a proxy to institutional quality. Following previous literature (Treisman, 2000 and Swamy et al., 2001) we include dummies for former British colonies and countries that were never colonized. Our results are, however, robust to the inclusion of dummies for French colonies, Spanish and Portuguese colonies, and other colonies. Inclusion of these dummies have little impact on the coefficient of internet penetration.[9] The coefficient of

[9] Acemoglu et al. (2001) argue that colonial status is a good proxy for institutions. We take the colonial status data from Treisman (2007).

internet penetration is still highly significant and negative. Also, the negative coefficient of former British colonies suggests that they are less corrupt which is consistent with the findings of the previous studies (Treisman, 2000). Column 3 controls for cultural variables which are arguably exogenous to corruption (Treisman, 2000). Neither magnitude nor significance of the coefficient of internet penetration is affected when we control for the proportion of Christians and Muslims in the total population.

There may still be a concern, however, that the countries that had better technology in 1500 AD might have also had higher incomes. Those countries may be able to constrain corruption more effectively in the beginning which may persist over time. If this is true, then we need to control for the income levels. Though there are no data available for income level of countries in 1500 AD, there are very good proxies available for the income levels such as urbanization and population density in that period (Acemoglu et al., 2002). We use the data from Acemoglu et al. (2002) for urbanization in 1500 AD and population density in 1500 AD. Technology adoption in 1500 AD remains a significant predictor of internet penetration today even after controlling for urbanization in 1500 AD (panel 1 column 4).[10] The correlation between the urbanization and technology adoption in communication is 0.345 with a $p-$value less than 0.005 indicating that technological adoption is positively correlated with the economic prosperity in 1500 AD. In column 4 of panel 2, the coefficient of internet penetration remains highly significant with expected sign. Also, note that the consistency of the coefficient in each column. Inclusion of various controls has little impact on the coefficient of internet penetration. Finally, in column 5, we control for a number of contemporaneous variables including GDP per capita. The coefficient of internet penetration appears with the expected sign and is significant at conventional levels. These IV coefficients are somewhat larger than the OLS estimates suggesting that OLS estimates are biased downwards because of endogeneity. Furthermore, the $F-$statistic at the bottom of the table indicates that the instrument is strong.[11]

[10] Acemoglu et al. (2002) show that income per capita is associated with both urbanization and population density. However, they note that theoretically, an association between population density and income per capita is more complex. Hence, we report our results with the urbanization rate as a proxy for prosperity or income levels in 1500 AD. The results are, however, robust to the use of alternative proxy for economic development, *i.e.*, population density.

[11] In addition, we also performed an informal instrument falsification test. We found that 1500 AD communication

Table 5 presents IV results with Facebook penetration included in the model. Each column in Table 5 has the same specification as the corresponding column in Table 4 plus the Facebook penetration term. The implicit assumption here is, that Facebook penetration is exogenous to the level of corruption once internet penetration has been instrumented. The coefficient of Facebook penetration as well as internet penetration have the expected sign in each column and are significant at conventional levels.

The results presented in this section are robust and show that Facebook and internet penetration have a causal and negative impact on corruption. Moreover, these effects are sizable. Next, the results of a falsification test for the association between Facebook penetration and corruption are reported. In addition, we also perform a number of robustness checks.

5.1 Falsification Test for the Facebook-Corruption Association

In this section, we perform a falsification test on the relationship between Facebook penetration and corruption. If our results are driven by omitted variables, then we should observe a significant relationship between Facebook penetration and corruption even before Facebook came into existence in 2003. In column 1 of Table 6 our dependent variable is the 1996 Control of Corruption Index. As we can see, the coefficient of Facebook penetration is not significant conditional on internet penetration and other control variables for the same year (see notes below the table). The coefficient of Facebook penetration is significant at 10% level when the dependent variable is the 2005 Control of Corruption Index in column 2. Since Facebook was launched in 2003 this was only two years later. Hence we find that the coefficient of Facebook penetration though small is significant and has the expected sign. Finally, when the dependent variable is the 2011 Control of Corruption Index, the coefficient of Facebook penetration is not only significant but about 50 percent larger than the coefficient in column 2. These estimates indicate that the results are not driven by some omitted variables and Facebook penetration has a causal impact on corruption.

technology has a significant negative impact on corruption if internet penetration is not controlled for. However, once we control for internet penetration, 1500 communication technology does not have any impact on corruption. This suggests that our instrument is valid and affects corruption only via internet penetration.

6 Conclusion

We investigate how digital media that provides for the possibility of two-way communication impacts corruption using data for over 150 countries. We find that the internet penetration has a significant and negative impact on corruption which is consistent with the findings of the previous studies (Andersen et al., 2011). We also find a sizable and statistically significant impact of social media (proxied by Facebook penetration) on corruption. We then use an instrument from the CHAT dataset to address endogeneity concerns and show that our results are robust.

Our analysis, however, uses unofficial Facebook data and we only have data on Facebook penetration. Ideally it would be better to have data on the total number of social media users in every country. Future research should be targeted at assessing the impact of social media on corruption more broadly. We believe that that as social media use continues to expand rapidly, more research on this topic, especially incorporating other social media platforms is necessary.

In addition, the privacy settings of the Facebook and legal system may also undermine the impact of social media on corruption. Clearly, the degree of anonymity of users sharing their bribing experiences may play a crucial role in the realization of the effect of social media on corruption. We would expect that the countries where there is asymmetric liability for petty corruption, the social media would be more effective in reducing corruption. Future research should also investigate this hypothesis.

7 References

- Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review,* 91:1369-1401.
- Acemoglu, Daron, Simon Johnson and James A. Robinson. 2002. "Reversal of fortune: Geography and institutions in the making of the modern world income distribution." *Quarterly Journal of Economics,* 117:1231-1294.
- Andersen, T. B. 2009. "E-Government as an Anti-corruption Strategy." *Information Economics and Policy,* 21:201-210.

- Andersen, T. B., Bentzen, J., Dalgaard, Carl-John and Selaya, P. 2011. "Does the Internet Reduce Corruption? Evidence from U.S. States and Across Countries." *World Bank Economic Review,* 25:387-417.
- Barro, Robert J., and Jong-Wha Lee. 2013. "A new data set of educational attainment in the world, 19502010." *Journal of Development Economics,* 104:184-198.
- Bhatnagar, S. 2003. "E-government and Access to Information." Global Corruption Report, 2003. *Transparency International.*
- Brunetti, Aymo and Beatrice Weder. 2003. "A Free Press is Bad News for Corruption." *Journal of Public economics,* 87:1801-1824.
- Comin, Diego A. and Bart Hobijn. 2009. "The CHAT dataset." *National Bureau of Economic Research,* Working Paper No. 15319.
- Comin, Diego, William Easterly, and Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?." *American Economic Journal: Macroeconomics,* 2:65-97.
- Diamond, Larry. 2010. "Liberation Technology." *Journal of Democracy,* 21:69-83.
- DiRienzo, C. E., Das, J., Cort, K. T. and Burbridge, J. 2007. "Corruption and the Role of Information." *Journal of International Business Studies,* 38:320-332.
- Freedom House. 2009. "Freedom on the Net: A Global Assessment of Internet and Digital Media." Available at http://www.freedomhouse.org/
- Nierhoff Maximilian H. 2013. "Facebook Country Statistics May 2013 – Rough Estimates Or Something?" Retrieved from http://www.quintly.com/blog/2013/05/facebook-country-statistics-may-2013/ on July 10, 2013.
- Saleh, Nivien. 2012. "Egypt's Digital Activism and the Dictator's Dilemma: An Evaluation." *Telecommunications Policy,* 36:476-483.
- Swamy, Anand, Stephen Knack, Young Lee and Omar Azfar. 2001. "Gender and Corruption." *Journal of Development Economics,* 64:25-55.
- Treisman, Daniel. 2000. "The Causes of Corruption: A Cross-National Study." *Journal of Public Economics,* 76:399-457.
- Treisman, Daniel. 2007. "What Have We Learned About the Causes of Corruption From Ten Years of Cross-National Empirical Research?." *Annual Review of Political Science,* 10:211-244.
- Wagner, Ben. 2012. "Push-Button-Autocracy in Tunisia: Analysing the Role of Internet Infrastructure, Institutions and International Markets in Creating a Tunisian Censorship Regime." *Telecommunications Policy,* 36:484-492.

Tables

Table 1: Summary statistics

Variable	Mean	Std. Dev.	N
Control of Corruption Index	0.009	0.991	164
Facebook penetration (2012)	20.469	18.5	164
Internet penetration	37.675	28.373	164
PCGDP	15194.849	16631.693	164
Political Rights	-3.299	2.082	164
Press Freedom	-46.791	22.843	163
Population	15.76	1.994	164
Christian Proportion (2005)	57.952	37.613	164
Muslim Proportion (2005)	22.676	33.915	164
Cellphone penetration	97.751	40.389	163
Urbanization	55.416	23.325	164
Years of Schooling (2010)	8.207	2.651	128
Openness	50.824	24.34	150
Technology adoption Index (1500 AD)	0.443	.401	106

All the variables belong to year 2011, unless otherwise indicated in brackets next to them. Whenever possible we use the control variable for the same year as the internet penetration. The data for Facebook penetration for year 2011 was available for a limited number of countries, hence, we use Facebook penetration data for year 2012.

Table 2: Internet and Corruption: OLS estimates.

Dependent Variable: Control of Corruption Index.

	(1)	(2)	(3)	(4)	(5)
Internet penetration	-0.0213***	-0.0188***	-0.0214***	-0.0236***	-0.0205***
	(0.00367)	(0.00353)	(0.00395)	(0.00479)	(0.00389)
ln (Per Capita GDP)	0.00132	-0.0117	-0.0897	-0.116	-0.0302
	(0.0810)	(0.0783)	(0.0899)	(0.108)	(0.0818)
Political Rights	-0.131***		-0.128***	-0.114***	-0.150***
	(0.0257)		(0.0257)	(0.0290)	(0.0271)
Christian Proportion	0.00364*	0.00380*	0.00427**	0.00292	0.00433**
	(0.00191)	(0.00197)	(0.00186)	(0.00177)	(0.00185)
Muslim Proportion	0.00456**	0.00395**	0.00467***	0.00457**	0.00594***
	(0.00185)	(0.00185)	(0.00178)	(0.00176)	(0.00187)
Population	0.0739***	0.0564***	0.0799***	0.0693***	0.0795***
	(0.0188)	(0.0192)	(0.0179)	(0.0235)	(0.0232)
Urbanization	-0.00157	-0.00236	-0.00259	-0.00300	0.000256
	(0.00304)	(0.00309)	(0.00313)	(0.00380)	(0.00300)
Press Freedom		-0.0156***			
		(0.00269)			
Cellphone penetration			0.00436***		
			(0.00149)		
Years of Schooling				0.0862**	
				(0.0342)	
Openness					0.000725
					(0.00331)
Constant	-1.026	-0.972	-0.696	-0.332	-1.125
	(0.680)	(0.649)	(0.705)	(0.810)	(0.853)
Observations	164	163	163	128	150
Adjusted R^2	0.704	0.729	0.717	0.732	0.714

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the negative of the control of corruption index such that a higher value implies higher corruption.

Table 3: Facebook and Corruption: OLS estimates.

Dependent Variable: Control of Corruption Index.

	(1)	(2)	(3)	(4)	(5)
Facebook penetration	-0.0121***	-0.0114***	-0.0111**	-0.00895*	-0.0129***
	(0.00453)	(0.00435)	(0.00449)	(0.00478)	(0.00489)
Internet penetration	-0.0173***	-0.0151***	-0.0178***	-0.0211***	-0.0165***
	(0.00392)	(0.00371)	(0.00413)	(0.00498)	(0.00418)
ln (Per Capita GDP)	0.0288	0.0124	-0.0547	-0.0807	-0.00241
	(0.0808)	(0.0782)	(0.0875)	(0.112)	(0.0804)
Political Rights	-0.114***		-0.113***	-0.103***	-0.132***
	(0.0260)		(0.0260)	(0.0295)	(0.0279)
Population	0.0611***	0.0455**	0.0673***	0.0548**	0.0626***
	(0.0187)	(0.0187)	(0.0179)	(0.0250)	(0.0232)
Christian Proportion	0.00426**	0.00442**	0.00479**	0.00340*	0.00495**
	(0.00195)	(0.00199)	(0.00188)	(0.00182)	(0.00193)
Muslim Proportion	0.00532***	0.00465**	0.00536***	0.00505***	0.00679***
	(0.00186)	(0.00187)	(0.00178)	(0.00178)	(0.00193)
Urbanization	-0.000117	-0.000907	-0.00122	-0.00177	0.00221
	(0.00309)	(0.00316)	(0.00312)	(0.00390)	(0.00285)
Press Freedom		-0.0143***			
		(0.00271)			
Cellphone penetration			0.00396***		
			(0.00143)		
Years of Schooling				0.0794**	
				(0.0328)	
Openness					0.000540
					(0.00310)
Constant	-1.052	-0.999	-0.751	-0.335	-1.084
	(0.657)	(0.632)	(0.675)	(0.805)	(0.814)
Observations	164	163	163	128	150
Adjusted R^2	0.715	0.739	0.726	0.737	0.726

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the negative of the control of corruption index such that a higher value implies higher corruption.

Table 4: Internet and Corruption: IV estimates. Instrument: Technology adoption in communication in 1500 AD. Instrumented: Internet Penetration.

	First-stage Regression. Dependent variable: Internet Penetration.				
	(1)	(2)	(3)	(4)	(5)
Technology adoption in Communication in 1500 AD	37.71***	32.92***	35.70***	30.18***	19.22***
	(5.742)	(6.099)	(5.573)	(7.063)	(4.554)
	Second-stage Regression. Dependent variable: Control of Corruption Index.				
Internet penetration	-0.0265***	-0.0252***	-0.0252***	-0.0265***	-0.0236**
	(0.00398)	(0.00464)	(0.00413)	(0.00650)	(0.0104)
Former British Colony		-0.359**	-0.336**		-0.394***
		(0.149)	(0.141)		(0.150)
Never Colonized		-0.324	-0.295		-0.370**
		(0.198)	(0.185)		(0.183)
Urbanization in 1500 AD				-0.00894	
				(0.0134)	
Constant	0.995***	1.099***	1.010***	0.955**	-0.639
	(0.156)	(0.149)	(0.164)	(0.403)	(2.016)
Cultural Controls	No	No	Yes	Yes	Yes
Ccontemporaneous Controls	No	No	No	No	Yes
Observations	106	104	103	67	95
$F-$statistic$^{\#}$	43.135	29.143	41.044	18.264	17.815

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. A higher value of corruption index implies higher corruption. Cultural controls include the proportion of Christians and proportion of Muslims in the total population. Other contemporaneous controls include *ln(GDPPC), ln(population), urbanization, cellphone penetration and openness to trade* $^{\#}$ The Cragg-Donald F-stat is not valid when the assumption of independently and identically distributed errors are dropped. The reported F-statistic are Kleibergen-Paap rk Wald F statistic (as reported by STATA 13) which are valid when i.i.d. assumption is dropped and "robust" option is invoked. R^2 not reported since it does not have any statistical meaning in the context of IV regression (for details, see http://www.stata.com/support/faqs/statistics/two-stage-least-squares/). .

Table 5: Facebook, Internet and Corruption: IV estimates. Instrument: Technology adoption in communication in 1500 AD. Instrumented: Internet Penetration.

	First-stage Regression. Dependent variable: Internet Penetration.				
	(1)	(2)	(3)	(4)	(5)
Technology adoption in 1500 AD	20.05***	17.41***	19.47***	26.36***	16.54***
	(3.780)	(3.554)	(3.703)	(4.023)	(3.805)
	Second-stage Regression. Dependent variable: Control of Corruption Index.				
Facebook penetration	-0.0240**	-0.0247**	-0.0235**	-0.0271***	-0.0192**
	(0.0119)	(0.0125)	(0.0115)	(0.00913)	(0.00913)
Internet penetration	-0.0175**	-0.0156*	-0.0161**	-0.0238***	-0.0193*
	(0.00799)	(0.00899)	(0.00809)	(0.00592)	(0.0114)
Former British Colony		-0.358***	-0.336**		-0.391***
		(0.135)	(0.131)		(0.152)
Never Colonized		-0.370*	-0.316*		-0.411**
		(0.201)	(0.185)		(0.187)
Urbanization in 1500 AD				-0.0124	
				(0.0103)	
Constant	1.150***	1.253***	0.974***	1.249***	-1.371
	(0.0979)	(0.0902)	(0.164)	(0.234)	(1.706)
Cultural Controls	No	No	Yes	Yes	Yes
Ccontemporaneous Controls	No	No	No	No	Yes
Observations	106	104	103	67	95
$F-$statistic$^{\#}$	28.131	24.006	27.639	42.913	18.907

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. A higher value of corruption index implies higher corruption. Cultural controls include the proportion of Christians and proportion of Muslims in the total population. Other contemporaneous controls include *ln(GDPPC), ln(population), urbanization, cellphone penetration and openness to trade* $^{\#}$ The Cragg-Donald F-stat is not valid when the assumption of independently and identically distributed errors are dropped. The reported F-statistic are Kleibergen-Paap rk Wald F statistic (as reported by STATA 13) which are valid when i.i.d. assumption is dropped and "robust" option is invoked. Dependent variable is negative of control of corruption index such that a higher value implies more corruption.

Table 6: Facebook and Corruption: Falsification Test.

	Dependent Variable: Control of Corruption Index for year		
	1996	2005	2011
Facebook penetration	-0.00741	-0.00801*	-0.0121***
	(0.00580)	(0.00429)	(0.00453)
Internet penetration	-0.110***	-0.0220***	-0.0173***
	(0.0220)	(0.00307)	(0.00392)
Control Variables#	Yes	Yes	Yes
Constant	1.589**	-0.169	-1.052
	(0.782)	(0.618)	(0.657)
Observations	141	167	164
Adjusted R^2	0.735	0.801	0.715

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the negative of the control of corruption index such that a higher value implies higher corruption. All the control variables are taken from the same year as the dependent variable except proportions of Christians and Muslims in the total population which are taken from year 2005. Facebook penetration in each column is from year 2012. # Control variables include $\log(Per\,Capita\,GDP)$, $Political\,Rights$, $Christian\,Proportion$, $Muslim\,Proportion$, $\log(Population)$, and $Urbanization$.