2017

Three Essays on Corruption

Sumi Sharma

Three Essays on Corruption

Sumi Sharma

Dissertation submitted
to the College of Business and Economics
at West Virginia University

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy in
Economics

Shuichiro Nishioka, Ph.D., Chair
Brian Cushing, Ph.D.
Stratford Douglas, Ph.D.
Eugene Bempong Nyantakyi, Ph.D.
Daniel Berkowtiz, Ph.D.

Department of Economics

Morgantown, West Virginia
2017

Keywords: markup, competition, corruption, income inequality

Abstract

Three Essays on Corruption

Sumi Sharma

This dissertation examines the link between market competition and corruption for developing countries and top income inequality and corruption in the US. The first two chapters explore the link between firm-level markup and corruption for a global dataset. I test the hypothesis that high-markup firms are less likely to engage in corruption. To investigate this relationship, I use firm-level data from World Bank Enterprise Survey (WBES) for 95 developing countries from 10 manufacturing industries. I find that high markup firms that operate in less competitive environments are less likely to bribe. These results are robust across three other measures of competition and two measures of corruption. I also look at the response rate of these firms to bribe-related question from survey data. I find that higher markup firms are more likely to be in contact with public officials, less likely to engage in bribes, and more likely to not answer bribe-related questions. These results highlight the importance of sample selection bias on the measure of competition and reveals that high markup rms and government-owned enterprises can determine the likelihood of responses to corruption-related questions. The third chapter discusses the issue of top income inequality and corruption within the US. I find a positive correlation between top income inequality measured by the top 1% and the top 0.1% income share and state-level corruption. Further, the results are magnified and continue to hold when three instrumental variables are used to exploit the exogenous variation in income inequality. These results suggest a policy focused on redistributive income as a means to tackle political corruption in the US.

Dedication

To my parents for their love, trust, and support

Acknowledgements

I would like to thank my committee members for their guidance and support throughout the entire process of my Ph.D. career. I am thankful to Dr. Shuichiro Nishioka, my dissertation advisor and chair, for his continuous guidance, insightful comments, and encouragement at each step of my dissertation. His mentorship throughout this process was critical in shaping this dissertation. A very special thanks to Dr. Stratford Douglas, Dr. Brian Cushing, Dr. Daniel Berkowtiz, and Dr. Eugene Bempong Nyantakyi for serving on my committee and providing excellent comments and suggestions on my work. I am also grateful for Dr. Feng Yao for his helpful comments.

The Ph.D. in Economics community at WVU has played a critical role in bringing this dissertation to completion. I thank the economics department for providing financial support. Several faculty members, staff members, and fellow students have helped me on this journey. I am grateful to them for their advice, encouragement, and support. I want to specially thank my friends at the Darien library for guidance on my writing.

I am forever indebted to my parents, Dr. Shiva Sharma and Mira Sharma, for encouraging me to follow my dreams. Thank you for believing in me. I thank with love my husband, Prajwal Kharel, for his encouragement and support that helped me get through hard times in this journey. Special thanks to my daughter, Ariya Kharel, for understanding why her mother is stuck at the library all the time and bringing great happiness to my life. Thanks to my brother, Sumit Sharma, for always being there for me. Finally, special thanks to my parent in-laws, Dr. Mohan Kharel and Shyam kala Kharel, and brother in-laws family, Prasant Kharel, Bebina Kharel, and Ava Kharel for their love and support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Markup and Corruption: Evidence of Firm-Level Data from Developing Countries

1.1 Introduction

Corruption is one of the main obstacles to economic growth and development of developing countries. The World Bank estimates the loss caused by corruption to be around 5% of global GDP, which amounts to $2.6 trillion, and over $1 trillion of that amount paid in bribes each year.[1] Corruption can have a number of negative effects on a firm's operation and relationship with government. Therefore, understanding ways to cure corruption is critical to facilitate fair competition, increase foreign direct investments, and economic growth. Many economists and policy makers have related market competition as an approach to dealing with corruption, but the question on the direction and sign is far from settled. This paper clarifies this relation by investigating market competition by firm-level markup and exploring the link between corruption. The main findings show that the high-markup firms operating in less competitive environments tend to decrease the amount paid as bribery.

There are two broad categories of economic theory that suggest competition may be important for understanding and curing bribery. The first category of research emphasizes

[1]World Bank, 2005

competition at the government official level where the officials are responsible for providing goods and services to citizens or economic agents.[2] These studies argue that competition at the government official level works similarly to a firm facing Bertrand competition, which eventually drives down prices. If no individual official has monopoly power, economic agents can freely choose an official to work with in obtaining permits or licenses. This drives the equilibrium price for bribes down to zero (Shleifer and Vishny, 1993). The other category of research focuses on competition at the level of economic agents who are seeking goods and services, such as licenses or permits, from the government official.[3] These studies indicate that increased competition among economic agents drives profit to zero, which leaves little surplus for extortion by the corrupt official. This paper focuses on the second strand of literature by looking at whether a firm's market power (or level of competition) has an effect on corruption.

Most of the empirical research at the cross-country level find a negative association between the degree of competition and corruption. The most common approach is to use perception-based corruption indexes (e.g., Transparency International Corruption Index and Political Risk Service's International Risk Guide Indicators) and indirect measures of competition (e.g., ratio of imports of GDP to total GDP and Economic Freedom indexes of the Heritage Foundation) (For instance, see Ades and Di Tella (1999), Bliss and Di Tella (1997), Emerson (2006), Treisman (2000a)). There are several limitations for these cross-country studies that rely on broad aggregate indicators. First, this type of data is not adequately suited for within country, industry, and firm-level analysis and has limited coverage. In particular, there are many economic differences across industries and firms that determine the incidence of bribes and level of competition, which cannot be controlled for with cross-country data. Second, the Corruption Perception Indexes are based on expert opinion that might not accurately portray the true corruption level. By contrast, data obtained through surveys reflect the firm's actual experiences on corruption and provide valuable details on other firm-level determinants of corruption. It is, therefore, interesting to assess whether firm's operating environment can shape the firm's decision to bribe after these firm-level

[2]see Rose-Ackerman (1978), Shleifer and Vishny (1993)

[3]see Ades and Di Tella (1999), Bliss and Di Tella (1997), Emerson (2006), Treisman (2000a)

characteristics are controlled in the empirical model.

More recent studies using firm-level survey data find a positive association between market competition and the amount of bribes paid. My research shows only two other published papers on this topic by Alexeev and Song (2013) and Diaby and Sylwester (2015). Both papers measure market competition by several measures, including number of competitors and markup. My research is closely related to these papers but differs in the following ways. In contrast to these studies that use profit-to-sales markup (i.e., profit margin), this paper estimates firm-level markup as a ratio of the output elasticities of intermediate inputs to the intermediate input's expenditure share. One of the more recent and leading paper by De Loecker and Warzynski (2012) shows the precise identification strategy to generate firm-specific markups for a cost-minimizing firm.

More specifically, markup can be identified following three basic steps (details section 3.2.2) which can be summarized as follows. First, this paper estimates several specifications of production functions and generates the beta coefficients (or output elasticities) for intermediate inputs. Second, the revenue shares of intermediate inputs are calculated directly from the survey data.[4] Third, markup is then estimated as a ratio of output elasticities in intermediate inputs (step 1) to revenue shares of the intermediate inputs (step 2). Thus, this markup ratio can be used to interpret the pricing power of the firm where a high markup means stronger market power or the firms faces less market competition.

An advantage to estimating markup this way is that, unlike the profit-to-sales markup that requires information on profitability and operating costs, this approach only needs information on at least one freely adjustable input which is readily available in the data. This, in turn, makes the markup ratio estimate less noisy because it is directly estimated and does not rely on price information (Cassiman and Vanormelingen (2013)). In addition, the assumption that firms are profit maximizers for profit-to-sales markup may not be valid because the sample covers developing countries where firms are less motivated to maximize profits and more likely to hire excess labor (Azmat et al. (2012)). The optimal input demand,

[4] De Loecker and Warzynski (2012) use the revenue share and the expenditure share interchangeably. This paper calculates expenditure share as the ratio of total cost of intermediate inputs to total revenue (or total sales).

therefore, as required by profit maximization condition may not be satisfied.

This paper uses a rich dataset from the World Bank's Enterprise Survey (henceforth, WBES) that covers 10 main manufacturing sectors from 95 developing countries. Corruption is measured by a sales-based bribe measure and a contract-based bribe measure. The sales-based bribe measure is defined as the firms payment of bribes to a government official in order "to get things done", while the contract-based bribe measure is defined as the firms payment to obtain a government contract. To clearly understand the competition-corruption link at the firm-level, the models control for the firm's age, ownership status, location (capital), and export status. Since almost 75% of the firms report zero bribe payment for both variables, I use a tobit model for the main estimation. However, probit, logit, and linear probability models are also used for robustness. The main results can be summarized as follows. Based on an unbalanced pooled cross-section of 22,000 observations from 2006-2016, the tobit models suggest there is a negative and statistically significant relationship between low levels of market competition (high markups) and corruption. For the sales-based bribe measure, a 10 percent increase in markup decreases the amount of bribes paid by 0.5 percent. Similarly, for the contract-based bribe measure, a 10 percent increase in markup decreases the amount of bribes paid by 0.9 percent. The positive (negative) relationship between a stronger (weaker) product market competition and corruption continues to hold when alternative measures of competition are tested. Both the firm's reported number of competitors and the firm's informal competition status show a positive and statistically significant relationship with corruption. Overall, the results reconfirm the positive (negative) link between stronger (weaker) market competition and corruption that previous firm-level empirical research has shown.

The paper proceeds as follows. Section 2 discusses past theory and empirical research on the relationship between competition and corruption. Section 3 outlines the WBES data set with summary statistics. Section 3 also explores markup and the different bribe variables in detail. Section 4 provides an empirical analysis for the tobit, probit, and logit estimates. Section 5 presents the results and section 6 concludes with a brief policy implication.

1.2 Literature Review

1.2.1 Theory Background

Corruption is defined as the act whereby government officials extract rents from individuals and businesses for service provided. Corruption increases transaction costs for firms (Rose-Ackerman, 1978) and becomes more expensive than a government tax due to the need for secrecy (Shleifer and Vishny, 1993). As a consequence, some firms are reluctant to invest in highly corrupt countries, which reduces foreign direct investments and economic growth in the long-run.[5] Therefore, understanding the root causes of corruption are important to policy makers.

The theory between product market competition and corruption remains ambiguous with respect to its sign and direction. There are mainly two strands of literature that suggest that competition is in fact important for addressing corruption. The first strand models competition at the government official level who are responsible for providing goods and services to economic agents (Rose-Ackerman (1978), Shleifer and Vishny (1993)). Rose-Ackerman (1978) was the first to suggest competition at the official level as a way of reducing corruption. Subsequently, other models where the government official remains in-charge of providing access to the market, license, or permits etc., emerged (Shleifer and Vishny (1993)). The main idea of these models is that increasing the number of government officials who are in-charge of providing goods and services reduces the amount of bribes demanded. The models work similar to a firm facing competition that observes a reduction in prices. Shleifer and Vishny (1993) provide an example on how this works, which eventually eliminates bribe payments (p. 607):

"A citizen can obtain a U.S. passport without paying a bribe. The likely reason for this is that if an official asks him for a bribe, he will go to another window or another city. Because collusion between several agents is difficult, bribe competition between the providers will drive the level of bribes down to zero."

Alternatively, Bliss and Di Tella (1997) argue that it is the corrupt officials who have an incentive to drive less efficient firms out of market, and therefore the official's main problem

[5]For instance, see Wei (2000), Svensson (2005), Knack and Keefer (1995), Mauro (1995)

is to maximize the expected bribe revenue per firm. More specifically, the authors consider the number of firms endogenous to the model and show that greater competition increases corruption as firms seek to gain advantages over their competitors. They use lower overhead costs to proxy for one of the "deep competition" parameters and show a significant reduction in number of firms in equilibrium.

The other strand identifies competition faced by the economic agents who have to deal with corrupt government officials regularly. Economic agents might need to engage in corruption (pay bribes) to obtaining a license, permit, etc (see, for instance, Ades and Di Tella (1999), Bliss and Di Tella (1997)). Increased competition for economic agents, in this case, eliminates excess profits which reduces corruption. In other words, perfect competition aids to control corruption since bribes are harder to extract when profits are zero. Alternatively, Ades and Di Tella (1999) provide a scenario where this relationship might be ambiguous since less competition, measured by the number of firms, increases the economic rent enjoyed by the firm. At the same time, the public is keener and more likely to spend resources to monitor the officials which, in turn, results in less corruption.

Emerson (2006) argues that a government agent acts alone to demand graft from firms in order to limit the number of firms in the market. However, the government official is subject to a "detection technology" that increases with number of firms and bribe payments. Emerson obtains two stable equilibria: high corruption and low competition, and low corruption and high competition, and concludes that "competition is antithetical to corruption". Other studies argue that officials can restrict entry to a market and extract rent accordingly (For instance, see Campos et al. (2010), Dutta and Mishra (2004),Aidt and Dutta (2008)). These models treat the number of competitors and degree of corruption as jointly determined and focus on the causality from corruption to competition.

1.2.2 Empirical Background

Most of the empirical research finds a negative relationship between competition and corruption, although the direction of causality is unclear. Ades and Di Tella (1999) use a cross-country study to show that level of rents and market structure determines the level of corruption. They show that countries with less market competition (or higher rents) tend have higher corruption. They proxy for competition by share of imports to total GDP, share of fuel and mineral exports in total exports, and distance to the world's major exporters in the 2SLS model to deal with endogeneity.

Alternatively, for a cross-country setting, Emerson (2006) show a negative link between corruption (measured by bribes) and industrial competition. The direction is from corruption to competition in this case. Competition is measured by two indexes: rankings based on business leaders collected by Economic Freedom indexes of the Heritage Foundation and an index based on trade policy, foreign direct investment, property rights, etc., obtained from the World Economic Forum's Global Competitiveness rankings.

The aforementioned papers rely on a small country-level sample with perception-based corruption indexes. For example, Ades and Di Tella (1999) uses corruption data from Business International and World Competitiveness Reports and Emerson (2006) uses corruption data from Transparency International and World Audit Organization. By using a country-level indicator for corruption, these studies assume homogeneous levels of corruption for different firms and industries, which may or may not be accurate. In addition, these indexes assume only one type of corruption and it is impossible to identify the type of corruption, such as bribery by government officials, which the firm engages in. Corrupt officials may demand a bribe or the firm may pay for a bribe to get things done, both of which will not show up in the aggregate data. Likewise, country-level indicators for competition may not accurately portray the firm's competitive environment in a given country or industry. There could be many firm-level differences across industries within the same country that determine the level of bribes and competition, which need to be controlled in analyzing the competition-corruption link.

This paper complements the more recent empirical development in this literature that relies on firm-level survey data. My research shows only two other published papers on this topic Alexeev and Song (2013) and Diaby and Sylwester (2015); both papers find a positive association between market competition and the amount of bribes paid. Alexeev and Song (2013) use five direct measures of competition: number of competitors; firm-level markup; an index to customer reaction of price increase; national and local market share; and the

Herfindahl-Heirschman Index (HHI). They instrument the degree of competition by U.S. capital-labor ratios to control for endogeneity.

Overall, they find a positive association between bribes and different measures of competition. However, the results mostly hold for number of competitors but are not robust to other measures of competition. For example, markup is only significant after controlling for endogeneity in the 2SLS model; other measures such as the local and national shares and the HHI are not significant at all, although show the correct signs.[6]

Alexeev and Song (2013) discuss drawbacks to using each measures of competition in their paper ; however, I briefly outline the reasons for markups. First, they estimate price markup as the "difference between total market value of production and the firm's operating costs to the firm's total market value of production" (p.163). This measure may be inconsistent as firms might not include the cost of bribes in total costs. Furthermore, the sample covers developing countries where firms are less motivated to maximize profits and more likely to hire excess labor as Azmat et al. (2012) point out in their paper. This raise questions about their approach for generating markups.

Recently, a closely related paper by Diaby and Sylwester (2015) also finds a positive relationship between market competition and corruption for post-communist countries. They include the local number of competitors as a proxy for competition in the main model. However, they also include other measures of market competition such as: national competitors; markup; competitive pressure from imports; and the hypothetical question: "what would happen to firm sale should the firm raise its price by 10%?".[7]

Both these papers include markup as a measure of competition but are unable to find a statistically significant relationship between markup and bribes. An important difference in this paper is that markup is calculated as the ratio of the output elasticity of intermediate

[6]The authors deal with reverse causality with a 2SLS and second stage GMM model. For the 2SLS, only markup is significant at the 5% level and number of competitors is significant at the 15% level. For the second stage GMM model, only number of competitors is significant at the 5% level and markup is significant at the 15% level.

[7]Diaby and Sylwester (2015) instrument the number of competitors with: questions such as "what if suppliers raises the price, would firm still purchase from them?"; sources of attracting new customers; and whether the firm is a trade association. They also measure competition by anti-competitive trade practices, domestic competition in firm's decision to innovate, competition from foreign firms, extent of domestic competition in firm's decision to cut production, and competition from foreign firms.

inputs from the production function to share of expenditure of the intermediate inputs. This markup estimate is a good indicator of competition because it is not dependent on firm's operating costs as used to calculate profitability. More importantly, the instrumental variable used by Alexeev and Song (2013), which is the U.S. industry capital- labor ratio, is considered the same across all countries. This assumption might not be accurate because firm-level fixed effects parameters can differ by country. In addition, the firms competitiveness can also be determined by institutions in a given country.

According to Sequeira and Djankov (2010), firms engage in two forms of corruption when seeking a service. First, the collusive (or cost-reducing) corruption that "emerges when public officials and private agents collude to share rents generated by the illicit transactions" (p.12). Second, coercive corruption (or cost-increasing) that emerges when a public bureaucrat demands a fee from a private agent to gain access to public services. Coercive corruption increases the price of goods and services (above the official price) as firms have to pay both a bribe and the cost.[8] Alexeev and Song (2013) argue that cost-reducing corruption is more likely to happen with an increase in market competition, where firms are willing to pay a bribe to lower fixed or variable costs.

1.3 Data

The data used for this study comes from the Enterprise Surveys of the World Bank (WBES). The WBES contains general firm-level information on degree of competition, business-government relations, corruption, finance, labor, and productivity. The WBES uses a stratified sampling procedure and covers different sectors from various countries, sectors, and years. For the manufacturing industry, establishments with five or more employees located in major metropolitan areas of the country are surveyed.

The major advantage of the WBES data is that it provides specific information about representative firms that operate in a particular country. In other words, as opposed to

[8] Shleifer and Vishny (1993) define collusive corruption as corruption without theft and coercive corruption (also known as extortionary corruption) as corruption with theft. In the case of coercive corruption, citizens pay bribes on goods and services they entitled too, as compared to collusive corruption where bribes are paid on goods and services that are illegal.

the indexes based purely on expert perception, the surveys are administered face-to-face with managing directors, business owners, accountants, and other relevant staff members who have firsthand experience on issues such as corruption and competition. Although corruption data from the WBES have been extensively used in empirical literature, the measure still faces several criticisms. A major concern with the survey data is the reliability of self-reporting values and the amount of non-responses from firms. Both of these issues arise because corruption and business-government relationships remain a sensitive issue in many countries.

The World Bank acknowledges these issues and takes appropriate steps to ensure confidentiality and accuracy in the data. The government is not directly involved in gathering the data, but rather the World Bank coordinates with other private and local contractors to conduct the surveys.[9] In addition, respondents are not required to provide any information that could identify them or the firm. Despite these criticisms, the micro-level data is clearly of interest since most firms interact with public officials at this level, which can tell us exactly how firms' competitive environment affects the likelihood to bribe. The data also provides detailed information on various firm-level characteristics which aids in controlling for unobserved heterogeneity on the corruption measure. Furthermore, several papers attest the importance and accuracy of the WBES data. Fisman and Svensson (2007) cite "with appropriate survey methods and interview techniques" (p.68) firm managers are able to provide a detail and accurate response to corruption related questions. Similarly, Olken and Pande (2012) cite "since survey-based data on bribes can be easily replicated, it is one of the only areas where consistent measurement is now being carried out across countries and over time" (p. 483).

The data covered in this study includes 95 developing countries and covers 10 main manufacturing industries from 2006 to 2016.[10] For most cases, I report results based on 3,722 to 22,482 observations. The difference arises because the sample varies according to the measurement of competition and bribe variables. For markup, the total observations is 22,482. The manufacturing industries have been classified according to the major 2 digit ISIC code.

[9] For more information on methodology visit www.enterprisesurveys.org/methodology

[10] Sample updated September 2016.

In particular, the industries are: food and beverages; textiles, leathers, garments; products, leather, and footwear; wood and furniture; chemicals and pharmaceuticals; non-metallic and plastics; metals, machinery and equipment; electronics; auto and auto components; retail; and other manufacturing.

Table 1.1 provides a summary statistics for key variables in the WBES 2006-2016 panel data. The average informal payment calculated from the sales-based bribe variable is 1.61. Almost 90 % of the firms have some kind of private ownership. The average age of the firm is 19 years and about 50 % of the firms are located in the capital city.

1.3.1 Bribes

Since the primary focus is on how firm-level markup affects corruption in developing countries, defining corruption variables are important. Corruption is measured by bribes as percentage of total sales (*bribes_sales*) and bribes as percentage of a contract value (*bribes_contract*). These measures approximate the firm's behavior and government's rent-seeking behavior where public officials expect informal payments "to get this done" with regards to custom, taxes, licenses, regulations, services, etc. Specifically, the data for these indexes comes from the two questions respectively, "on average, what percentage of total annual sales or total estimated value of bribe payment, do establishments like this one pay in informal payments and gifts to public officials for this purpose?" , and"when establishments like this one do business with government, what percent of the contract value would be paid in informal payments or gifts to secure the contract?".

For the sales-based bribe variable, data can be obtained from two responses: the percentage a firm pays as an informal gift and the total annual amount of bribes paid informally. If the respondent reports the annual amount rather than percentages, the LCU amount is divided by total sales (*100) to convert it to percentage. Two of these indexes are merged to create one sales-based bribe (bribes_sales) index which reports the maximum value from these two responses. This increases the total observations from 18,741 to 22,482.[11]

I also construct two dummy variables: sales_dummy and contract_dummy. Both these

[11] In most cases, I use the merged value for sales-based bribe index but the results still hold for just the individual measure of *bribe_sales*.

variables take a value one to represent firms who responded with a positive value for sales-based bribe and contract-based bribe index, respectively, and zero otherwise. Answers that state “do not know” or “refuse to answer” are changed to missing values. Answers “does not apply” or “not applicable” is replaced with zero.

The average firm pays about 1.61% of their total sales as informal payments or gifts to government officials. Madagascar has the highest amount of bribe payment at 9.40 % of total sales while Israel has the lowest bribe payment at 0.20 % of total sales. Similarly, Israel has the lowest percentage of bribe of contract, valued at 0.41% while Philippines has the highest at 11.80%. The average firm pays about 2.94 % bribe as a percentage of contract value.

To validate the use of these measures of corruption, I consider a country-level measure of Control of Corruption(CC) from the World Governance Indicator (WGI). The WGI indexes are based on numerous individual data sources obtained from citizens, surveys from public and private non government organizations, and assessment of expert's opinions.[12] The indices capture experts opinions “on 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.” This index ranges from -2.5 (more corruption) to 2.5 (less corruption) and for the countries sampled the average CC is -1.52. I generate a cross-country average for the WGI data from 2006 through 2016 and check its correlation with an average of bribes across all firms in each country. The correlation between average *bribe_sales* and average WGI is -0.45 at 1% level of significance. The correlation between *bribe_contract* and average WGI is - 0.42 at 10% level of significance. These results suggest that our data of firm-level responses is highly correlated with other cross country measures.

1.3.2 Markup

This section introduces the methodology used to calculate markups from production function and the survey data. The first step describes the process to obtain coefficients from the production function and the second step demonstrates the calculation for markups (μ_{it}).

[12]For more detailed information, visit http://info.worldbank.org/governance/wgi/

Hall (1988) used data on total inputs and total output to calculate sector-level markups from production functions. In his seminal paper, Hall stated that under perfect competition input revenue share is equal to input cost share.[13] The difference between the two can be identified as firm-level markup. However, this methodology faces problems with identifying total costs and marginal costs for the markup estimation (Cassiman and Vanormelingen (2013)).

The solution recently suggested by De Loecker and Warzynski (2012) is to include an assumption that firms are cost-minimizers and choose at least one freely adjustable input. They calculate markup as the ratio of output elasticity of an input(s) to the total expenditure share of the input(s) and relate these to firm level export status. An advantage to this approach is that is relatively easy to estimate firm specific markup without requiring any information on the market structure and the firm's input demand (De Loecker and Warzynski, 2012). One difficulty, however, with this approach is addressing the unobserved productivity shocks of the production function. I do not investigate that in this paper because the WBES does not provide a dynamic panel data structure needed to calculate total factor productivities.

To estimate markups, I follow the recent work of De Loecker and Warzynski (2012). Consider the following cost minimization problem faced by firm i located in country c at time t:

$$\begin{aligned} \underset{N_{it},M_{it},K_{it}}{\text{minimize}} \quad & w_{it}N_{it} + r_{it}K_{it} + p^m_{it}M_{it} \\ \text{subject to} \quad & F_{it}(N_{it}, M_{it}, K_{it}) \geq Q_{it}, \end{aligned} \tag{1.1}$$

where N_{it}, M_{it}, K_{it} represents labor, intermediate inputs, and capital for firm i in period t respectively and $w_{it},\ p^m_{it},\ r_{it}$ denote the wage rate, the price of intermediate inputs, and the rental price of capital respectively. F(.) denotes the production function which is continuous and twice differentiable with respect to all of its arguments. Q_{it} is the total output of the firm. The Lagrange function associated with equation (1.1) can be written as:

$$\mathcal{L}(N_{it}, K_{it}, M_{it}, \lambda_{it}) = w_{it}N_{it} + r_{it}K_{it} + p^m_{it}M_{it} + \lambda_{it}[Q_{it} - F_{it}(N_{it}, M_{it}, K_{it})] \tag{1.2}$$

[13]Input revenue share is calculated as the ratio of total cost of the input to the total revenue. The input cost share is calculated as the ratio of total cost of an input to the marginal cost times the total output. For details see Cassiman and Vanormelingen (2013)

The first order condition for intermediate inputs is denoted by:

$$\frac{\partial \mathcal{L}}{\partial M_{it}} = p^m_{it} - \lambda_{it}\frac{\partial F_{it}}{\partial M_{it}} = 0 \tag{1.3}$$

After rearranging equation (1.3) and multiplying both sides by $\frac{M_{it}}{Q_{it}}$, the equation takes the following form:

$$p^m_{it}\frac{1}{\lambda_{it}}\frac{M_{it}}{Q_{it}} = \frac{\partial F_{it} Mit}{\partial M_{it} Q_{it}} \tag{1.4}$$

The final step is to consider P_{it} as the price of the final product sold and $\lambda_{it} = \frac{\partial \mathcal{L}}{\partial Q_{it}}$ as the marginal cost (mc_{it}) for a given level of output. Given markup is defined as the ratio of price over marginal cost, i.e. $\mu_{it} = \frac{P_{it}}{mc_{it}} = \frac{P_{it}}{\lambda_{it}}$, we can rewrite equation (1.4) as:

$$\frac{P_{it}}{P_{it}}p^m_{it}\frac{1}{\lambda_{it}}\frac{M_{it}}{Q_{it}} = \frac{\partial F_{it} Mit}{\partial M_{it} Q_{it}} \tag{1.5}$$

or markup (μ_{it}) can be denoted as:

$$\mu_{it} = \theta^m_{it}(\alpha^m_{it})^{-1} \tag{1.6}$$

where $\theta^m_{it} = \frac{\partial F_{it}}{\partial M_{it}}\frac{M_{it}}{Q_{it}}$ is the output elasticity of intermediate input and $\alpha^m_{it} = \frac{p^m_{it} M_{it}}{P_{it} Q_{it}}$ is the expenditure share of intermediate input in total sales.

The estimation of firm-level markup relies on two factors. First, it is important to consider a cost-minimizing firm and choose an input that is free of any adjustment cost. Therefore, it is critical to correctly estimate the output elasticities for intermediate inputs from the production function. Output elasticities can be calculated for both labor and capital; however, I choose to use intermediate inputs as labor is not freely adjusted in developing countries. This is mainly true in state-owned enterprises and in the presence of unions ((Azmat et al., 2012), Shleifer and Vishny (1994)). Firms in these countries flexibly optimize the purchase of intermediate inputs rather than labor and capital. In addition, depending on the country, the WBES provides at maximum 2-3 years of data and firms are not consistently linked across time through an unique firm id. This makes it difficult to calculate output elasticities for capital which is considered a dynamic input in the literature. Second, it is important to collect data on expenditure share for intermediate inputs and the total sales (or revenue) for the firm. This is readily available at the WBES.

The econometric procedure to generate markups using production function consists of two steps which I outline in the following sections.

1.3.3 Identify output elasticities

I start by assuming a Cobb-Douglas production function of the following form:

$$Y_{it} = A_{it} N_{it}^{\beta_n} K_{it}^{\beta_k} M_{it}^{\beta_m} \tag{1.7}$$

where Y_{it} is total real output (sales), A_{it} is total factor productivity, N_{it} is human capital, K_{it} is capital, and M_{it} is intermediate inputs for firm i at time t. The WBES defines total sales as the value of all annual sales, including manufactured goods and goods the establishment buys for trading. Capital is constructed from balance sheet information and defined as net book value, which is the sum of machinery and equipment (including transportation and installation costs) minus depreciation accumulated since the date of purchase. An alternative estimator for capital-the answer from the manager's evaluation for the firm's equipment, land and building if sold on the market - is used when necessary. Labor or manpower costs is measured as labor adjusted by human capital and is defined as the total annual wages and all annual benefits, including food, transport, and social security (i.e. pensions, medical insurance, and unemployment insurance). Intermediate inputs are the sum of annual cost of electricity, communications services, raw materials, intermediate goods used in production, fuel, transportation of goods -excluding fuel, water, and other cost of production. Since the aforementioned variables are in local currency, all variables have been exchanged to U.S. Dollars using World Development Indicators.[14] The data are then deflated using GDP price deflator for the United States with 2005 as the base year.[15]

Estimating equation (1.7) with OLS leads to biased results if the inputs in the production

[14]WDI indicator code: PA.NUS.FCRF
[15]World Bank indicator

function are endogenous to the model. Marschak and Andrews (1944) noted that inputs in the production function are not independently chosen, but are determined by the characteristics of firms. The endogeneity problem arises because of productive factor unobservable to the econometrician, but observable to the firm which affects the input demand. More recent literature suggests using control function approaches (see Ackerberg et al. (2015) for detail). These literatures suggest that under profit maximization, observed investment (Olley and Pakes (1992)) and intermediate materials (Levinsohn and Petrin (2003)) can be inverted and used as a proxy to solve for the correlation between unobserved productivity shock and input levels. However, due to the lack of good instrumental variables and dynamic panel, I rely on fixed effects to get consistent coefficients on labor, capital, and intermediate inputs.

To obtain the estimates of output elasticities, I rely on Cobb-Douglas (CD), Constant Returns to Scale (CRTS), and Translog production functions. For each case I use a gross output (revenue) production function with two variable inputs without adjustment costs. Markup can be obtained from either labor or intermediate inputs. However, I use intermediate inputs since there is evidence of excess employment and wages in state-owned firms for social stability (e.g., Shleifer and Vishny (1994)).

Based on the discussion above, I estimate the following production functions using Ordinary Least Sqaures (OLS) for the full sample:

Model 1. Cobb-Douglas (CD) Production Function

$$y_{it} = \beta_n n_{it} + \beta_k k_{it} + \beta_m m_{it} + \delta_c + \delta_j + \delta_t + \epsilon_{it} \tag{1.8}$$

where y_{it} is the (log) real total sales for firm i at time t. $n_{it}, k_{it},\ and\ m_{it}$ represent the (log) inputs of human capital adjusted for labor (wage bill), (log) real capital, and (log) real intermediate inputs respectively for firm i at time t. δ_c, δ_c, δ_t is the country, industry, and year fixed-effects that captures productivity and ϵ_{it} accounts for random errors.

Model 2 Constant Returns to Scale (CRTS) Production Function

With the CRTS production function, I examine whether a proportional change in input (constant factor increase) leads to a change in output. I impose the following restriction to equation (1.8)

$$\beta_n n_{it} + \beta_k k_{it} + \beta_m m_{it} = 1 \tag{1.9}$$

Model 3. Translog Production Function

I use a translog form for labor and capital which allows me to diverge from just having a linear term to having both a linear and a quadratic term.

$$y_{it} = \beta_n n_{it} + \beta_k k_{it} + \beta_m m_{it} + \beta_{nn} n_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{nk} n_{it} k_{it} + \delta_c + \delta_j + \delta_t + \epsilon_{it} \tag{1.10}$$

where $\beta_n, \beta_k, \beta_m$ are the first derivatives; β_{nn}, β_{kk} are own second derivatives and β_{nk} is the cross second derivatives. Other variables remain the same as in equation (1.8).

Table 1.2 provides summary statistics for the variables included in the production function. Before estimating the production function, I have eliminated questionable outliers for the main production parameters: output, capital, intermediate inputs, and labor. Firms with large absolute values after the log transformations and observations that result in zero, negative, or missing values for the production parameters are eliminated from the sample. In addition, the top 1% and bottom 1% have been dropped from the sample.

Table 1.3 shows the results of estimating the CRTS production function (Model 1) for 10 main manufacturing industries, including retail and wholesale trade, with country and year fixed-effects. The manufacturing industries have been classified according to the 2 digit ISIC code. To increase observations per group, smaller industries are merged with larger industries (for e.g., leather is merged with garments and textiles). To allow industry differences in the production parameters, I run regressions for each industry separately and include country and year fixed-effects. All of the estimated parameters are significant at a 1% level. The coefficients on intermediate inputs are mostly stable between industries and fluctuate from 0.65 for metals, machineries, and electronics to 0.59 for textiles, garments, and auto industry.

Table 1.4 shows the coefficients of the production function for Model 1-3 with country and year fixed- effects.[16] All of the estimated parameters are significant at a 1% level and the estimated factor coefficients are close to the known input shares. The results confirm that the manufacturing industry is labor intensive in developing countries. For the whole sample, the output elasticity of labor (human capital) is 0.313 for CD production function and 0.318 for CRTS production function. The output elasticity of capital is 0.06 for both the CD and CRTS model. For the CD production function, a 10% increase in capital is associated with an increase of 0.6% in output. The output elasticity of intermediate inputs is 0.62 for CD and CRTS production function. This implies that, for the CD production function, an increase of 10% in intermediate inputs is associated with an increase of 6.2% in output. To sum, approximately 32% of the production output is allocated to human capital, 62% to intermediate inputs, and 6% to capital. The estimated parameters are similar in columns (3) and (4) which represent the results from estimating the Translog and Kmenta production function.

1.3.4 Identify markups

The next step is to calculate markup as $\mu_{it} = \frac{\beta_m}{\alpha_{it}}$ from the models estimated above. B_m is the scalar coefficient for intermediate inputs obtained in Model 1-3 and α_{it} is the share of expenditure for intermediate inputs in total sales. The interpretation of a high markup of firm is that the firm has high market power or faces a less market competition. It also suggests that the firm is able to charge a higher markup compared with a lower-markup firm that faces several competitors in the market. For statistical estimation of markups, I have eliminated high leverage points that is, for log markup, values below -0.39 and higher than 5 have been eliminated from the sample. In order to ensure the sample of the firms are representative of the true population, I re-estimated the sample by replacing values less than -0.39 with -0.39 and values higher than 5 with 5. The results are comparable. The mean markup is 2.03 from the 3 models and the median is around 1.27.

Table 1.5 shows the differences between the markup values across industries. Both the mean and the median values fall in the reasonable 1-2 range. The three industries with

[16]Results are similar when industry fixed-effects are added to the model.

higher markups are wood and furniture (1.38), textiles, garments, and leather (1.33), and electronics (1.24). Conversely, the three industries with lowest markups (at the level) are the auto and its components (0.97), non-metallic and plastic materials (1.21), and chemicals and pharmaceuticals (1.22). Note: for empirical analysis I use markup estimates from the industry-level CRTS model.

To ensure validity and robustness to the markup estimates, I regress the firm's reported intervals for number of competitors on *markup*. Table 1.6 reports the results using an OLS regression with country, industry, and year fixed-effects. This regression will shed light on whether an increase in the number of competitors leads to lower markups. If the results hold, the markup estimate is consistent because higher markup indicates less competition in the market. The results confirm the hypothesis that firms facing competitors greater than two, but less than five, and greater than five see a decrease in markup. These findings are statistically significant at the 1% level.

1.3.5 Other Controls

Following previous firm-level studies, I control for various firm-level characteristics that can influence the relationship between competition and corruption. Batra et al. (2003) find that private firms are more likely to bribe, pay a higher revenue share as bribes, and more likely to consider bribe as an obstacle. Similarly, firms with large private, foreign, or government shares could face different bribe environments in dealing with public officials. Hence, I include a dummy variable equal to one for firms with more than 50% of percentage of ownership by private domestic individuals, companies or organizations ($prishare$), equal to two for percentage of firm owned by private foreign domestic individuals, companies or organizations ($forshare$), and zero for more than 50% ownership by government or state ($govshare$). In my sample, 93% of the firms have some degree of *prishare*, 12% of firms with some degree of *forshare* and 2% of firms with some degree of *govshare*.

Age is calculated as the logarithmic difference between the survey year and and the year in which the establishment started its operation. The suggested direction of *age* is mixed in the literature. On the one hand, young firms could pay more bribes relative to older firms to enter the market. Alternatively, older firms could pay more bribes compared to young firms to remain in the market.

I also include a variable on export status, where *exporter* is a dummy variable representing a value one if the sum of indirect exports (sold domestically to third party that exports products) and direct exports is greater than 50%, with national sales as the benchmark. Studies have shown that firms that export internationally, as compared to domestic sales, might be more prone to the government rent extraction to avoid customs or taxes. Lastly, I control for the capital city since most government offices are located in the city center. *capital* is a dummy variable with a value of one for firm's located in the capital city. All else equal, I expect a positive relation between *capital* and bribes which indicates the firm's influence on local government officials.

1.4 Methodology

The literature on the effects of competition on corruption has generated concerns as the cross-country studies do not address potential omitted biases. To overcome these problems, I estimate the following probit model using firm-level controls that could determine corruption. An advantage of using the firm-level data is that it sheds light on how a firm's operating environment, within a specific industry or country, can effect the firm's decision to bribe government official. The probit model is estimated with maximum likelihood methods and takes the following form:

$$P(Bribe_{it} = 1) = P(\beta_1 markup_{it} + \beta_2 X_{it} + \delta_c + \delta_t + \delta_j + \epsilon_{it} > 0) \tag{1.11}$$

where $Bribe_{it}$ represents dummy variables for the two corruption indicators: sale-based bribe measure ($bribe\_sales$) and contract-based bribe measure ($contract\_sales$). More specifically, the dummy variable for the sales-based bribe measure (dummy_sales) represent a value of one if firm i in country c at time t reported a positive value for bribe indicated as percentage of total sales, while the dummy variable for the contract-based (dummy_contract) measure represents a value of one if firm i in country c at time t reported a positive value for bribes indicated as percentage of contract value. The main independent variable is *markup* which is estimated as the ratio of output elasticity of intermediate inputs to the total share of expenditure on intermediate inputs. The model also controls for firm-level characteristics (X_{it}) that could determine bribe incidences. These firm-level controls include firm's age, export status, ownership status, and capital. In addition, the equations include country, industry, and year-fixed effects (denoted as $\delta_c,\ \delta_j, and\ \delta_t$, respectively).

Since the sample includes 95 different countries from four continents, the inclusion of country-fixed effects controls for countrywide factors- country's legal system, legal origin, rule of law, and regulation of various economic activities that could influence corruption (Treisman (2000a)). The inclusion of industry-fixed effects is to control for the exogenous variation in firm productivity that can influence the relationship between markups and bribe payments. In the same vein, the inclusion of time fixed-effects is to capture any macroeconomic trend or policies of a country that could potentially influence the firm's intention to bribe. All three models use robust standard errors clustered at the country-industry level. This is to allow errors to be correlated within the same country and industry. Equation (1.11) is also estimated using logit and linear probability model for robustness.

Given that almost 75 % of the firms report zero informal payments for the corruption indexes, the Ordinary Least Squares (OLS) results can lead to biased estimates. Therefore, I also estimate the following Tobit mode with a lower bound set to zero:

$$bribe_{it} = \beta_1 markup_{it} + \beta_2 X_{it} + \delta_c + \delta_t + \delta_j + \epsilon_{it} \tag{1.12}$$

In equation 1.12, the dependent variable represents the percentage of total annual sales or total estimated value of bribe payment (sales_based_bribe) and the percent of the contract value paid in informal payments to secure the contract (contract_based_bribe). All other right hand side variables remain the same as in equation 1.11.

1.5 Results

This section illustrates the econometric analysis of the relationship between firm-level markup and corruption, while controlling for firm's age, location, ownership status, and export status. First, I estimate whether high markup firms operating in less competitive environments are less likely to engage in corruption using the probit, logit, and linear probability model in table 1.7. If the firm's competition determines the amount of bribes paid, high markup firms (less competition) should be negatively correlated with bribes. More specifically, the results of equation 1.11 are reported and include country, industry, and year fixed-effects. All of these estimations include standard errors clustered at country-industry levels. The dependent variable in columns 1-3 are a binary assigned a value of one if the firm responded with a positive value for the sales-based bribe measure, while the dependent variable in columns 4-6 is a binary with a value of one if the firm responded with a positive value for contract-based bribe measure. Columns 1 and 4 report the marginal effects of the probit model. Columns 2 and 5 report the point estimate of the logit model, while columns 3 and 6 represent the coefficient estimates of the linear probability model. As expected, the coefficients of markup are negative and statistically significant at the 5% levels in all columns. The results show that high markup firms, compared to less markup firms, are less likely to bribe as measured by both the sales-based bribe and contract-based bribe measure.

Second, in table 1.8, I look at the results for the tobit model with country-industry clustered standard errors. Specifically, equation 1.12 is estimated with firm-level controls and fixed effects. The estimates in column 1 and 3 include markup as the only independent variable with country, industry, and year fixed-effects. Columns 2 and 4 add other firm-level controls. The dependent variable in columns 1-2 is bribes measured as percentages of total sales ($bribe\_sales$) and the dependent variable in column 3-4 is bribe measured as percentages of contract value. These variables are in logged terms. The results show that high markup firms decrease the incidence of bribes paid. For the sales-based bribe measure, a 10% increase in markup decreases the incidence to bribe by 0.5 percent, which is statistically significant at the 1% level. In other words, the elasticity of the sales-based bribe measure with respect to markup is -0.05. The results are similar for the contract-based bribe measure in columns

3-4. A 10% increase in markup decreases the incidence to bribe for the contract-based bribe measure by 0.9 percent. The result is statistically significant at 1 % level. As Beck and Maher (1986) explain, the model for contract-bribes can be compared to a competitive bidding model where the expected rent increases with the number of bidders. The results from the contract-based bribe measure, therefore, suggest that firms with high market power might (less bidders) likely already have established networks and are less likely to engage in bribes to get things done.

Next, I look at the firm-level characteristics in each model where several features are worth noting. Age enters the equation positively which supports the hypothesis that older firms, compared to younger firms, are more likely to bribe. Capital is positive and significant suggesting firm's operating in the capital city are more likely to bribe. The positive sign may be due to most government offices also being located in capital cities. Likewise, the firm's export status (=1 if exporter) enters the equation positively implying exporters are more likely to engage in bribes. Ownership variables are negative and significant. Compared to firms with private firm owners, foreign and government owners are less likely to bribe.

These results as consistent with previous literature in which stronger (weaker) market competition is associated with higher (lower) bribes. However, it should be noted that the markup results are different from the previous two studies. In this paper, markup is calculated as the ratio of output elasticity of intermediate inputs to the revenue share of intermediate inputs. Both previous papers are unable to find a statistically significant relationship between markup and bribes, although the signs are correct. Thus, the results suggest that the relationship between competition at the firm-level is economically meaningful and highly robust to various firm-level controls on corruption.

Finally, I provide a series of checks by replacing markup with other competition measures to see if the results continue to hold. The first measure is an indicator on whether the firm faces informal competition in the market. The WBES specifically asks "Does the firm face informal competition against informal or unregistered firms?" Based on responses to this question, a binary variable *(informal_comp)* is created where responses "yes" are coded as one and zero otherwise. The level of competition a firm faces, based on whether there is informal competition in the market, is difficult to interpret as the exact number of competitors are unknown. However, as La Porta and Shleifer (2014) show that informal firms can account to about 50 % of economic activity in developing countries, including this measure is important. The second measure is the logarithm of one plus the number of competitors reported by the firm (*Num_Comp*). This is different from the interval range listed below because it asks the firm to provide a value for the exact number of competitors. This measure, however, has its concerns. A large firm potentially could have several small size firms as competitors which might not impact a large firm's operation. On the other hand, a large firm could potentially have several large size firms as competitors which might have a big impact on a large firm's operation. In addition, the sample size is relatively small (about 7,000 observations) for this measure and including country, industry, and year fixed effects could lead to a loss of degrees of freedom. The third measure is range of number of competitors that denotes the intensity for competition. More specifically, the survey question asks "for the main market in which this establishment sold its main product, how many competitors did this establishment's main product/product line face?" The survey only provides four categories for a response and the data is coded as follows: no competition (coded 0, used as the benchmark); competitor=1 (coded 1); competitors between 2-5 (coded 2); Competitors >5 (coded 3).

Table 1.9 reports the results of estimating a series of different measures of competition on corruption with the tobit model 1.12. Overall, the estimated result confirms that greater market competition increases the incidence to bribes. The dependent variable in columns 1-3 is the sales-based bribe measure and the dependent variable in columns 4-6 is the contract-based bribe measure. In column 1, the number of competitors is insignificant but has the correct sign. This could be due to less observations and the loss of degrees of freedom due to inclusion of fixed effects. Column 2 indicates that firms facing no informal competitors are less likely to bribe. The sales-based bribe measure increases by 0.44 units for firms facing informal competition, which is statistically significance at the 1% level. For the interval of competitors, firms facing one competitor (compared to none) are less likely to bribe. In other words, firms facing competition (with one firm) decreases the incidence of bribes for sales-based bribe measure by 0.276 units. This relationship is statistically significant at the 10% level. Similarly, firms that face competition with more than 5 firms (compared to none), the incidence to bribes, measured by the sales-based bribe indicator, increases by 0.245 units. This result is statistically significant at the 5% level. For the contract-based measure in column 4, the coefficient on the number of competitors is significant at the 1 % level. In column 5, the coefficient of informal competition is positive and significant at the 1% level. The results indicate that there is about a 0.94 units increase in contract-based bribe measure for firms facing informal competition. The last column includes the intervals of number of competitors. Firms facing 5 or more competitors are more likely to engage in bribes as measured by the contract-based bribe measure.

Two points are noteworthy. First, the results presented in this paper are consistent with previous research that have utilized firm-level data to analyze the competition-corruption link. Despite differences in our data sample and measurement of market competition, this paper arrives at the same conclusion that higher (lower) market competition levels increases (decreases) the incidence to bribe. Second, the results are consistent across alternative measures of market competition when firm-level controls are employed in the models.

1.6 Conclusion and Limitations

This paper revisits the link between market competition and corruption using the WBES data for 95 developing countries from 2006-2016. For my analysis, I examine market competition as firm-level markup which is calculated as the ratio of output elasticity of intermediate inputs to expenditure share of intermediate inputs. Previous empirical research at the firm-level have calculated markup as the profit-sales ratio; however, this method might not be accurate because the firms are less likely to maximize profits and more likely to hire excess labor in developing countries. In addition, these studies do not find a statistically significant relationship between markup and corruption. This paper, therefore, contributes to the literature by finding a statistically significant negative correlation between markup and corruption. This result is consistent across other measures of competition including number of competitors and informal competition. Thus, I conclude that firm-level competition does indeed matter in determining the levels of corruption.

The markup estimated in this paper does have limitations. One can argue that the endogeneity associated in production function estimation is ignored. However, given the lack of dynamic panel data, the best alternative method to generate the input coefficients is with the inclusion of country, industry, and year fixed-effects. Another downside is the significant reduction in observations due to focus on the manufacturing sector. The future version of the paper will explore the service sector and find an appropriate instrument to test the exogenous variation in markup.

My findings have implications for designing policies related to fair competition in developing countries. It should be noted that this paper does not recommend a decrease in market competition to decrease the level of corruption. But, it is of importance to figure out the extent to which a firm has market power and if there are any costs associated with the high market power. If the costs associated with the monopoly power are greater than the costs associated bribery, then it might be important to look deeper at any policies that lead to fairer competition. In addition, it is important to explore the extent to which firms cluster together for the positive relationship between competition and corruption to hold. Do the competitors have to be located in the same city or region? How much of monopoly power does the firm possess in a specific city or region? I hope to address these questions in future works.

Table 1.1: Summary statistics for main variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
bribe_sales*	0.25	0.66	0	4.615	26928
bribe_contract*	0.596	1.049	0	4.615	8339
sales_dummy	0.194	0.395	0	1	26929
contract_dummy	0.199	0.399	0	1	11702
markup*	0.393	0.624	-0.387	2.892	36301
Competitors*	1.996	1.232	0	9.210	9916
Interval_Comp	3.464	0.807	1	4	24862
age*	2.683	0.813	0	5.088	53675
exporter	0.261	0.439	0	1	54370
capital	0.511	0.5	0	1	44126
prishare	90.028	27.725	0	100	53646
forshare	7.698	24.756	0	100	53592
govshare	0.721	6.708	0	100	53613

Notes: * denotes the variable is entered in log form. bribe_sales is the log of one plus bribe measured as the percentage of total sales. bribe_contract is the log of one plus bribe measured as percentage of contract value. Markup is the log of ratio of the output elasticity from the production function to share of expenditure of on intermediate inputs. Competitors is logarithm of one plus the number of competitors reported by the firm. Interval_comp is the intervals for number of competitors (0; Competitor=1; Competitors between 2-5; Competitors >5). Age is the log difference between the survey year and and the year in which the establishment started its operation. Exporter is a dummy variable where if indirect exports - sold domestically to third party that exports products- and direct exports > 5% is coded as one and zero otherwise. Capital is a dummy that takes on 1 if the firm is located in the capital city, 0 otherwise. For ownership, prishare, forshare, govshare represent the percentage of ownership by private, foreign, government parties.

Table 1.2: Summary statistics for Production Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
ln_real_R	17.32	3.127	7.576	30.405
ln_N	15.326	2.992	2.708	25.767
ln_real_K	15.68	3.245	0.694	27.455
ln_real_M	16.42	3.194	4.249	27.745

Notes: All variables represent log transformation. R is total sales of the firm, K is the net book value, N is the wage bill, M is intermediate inputs. All values are in 2005 U.S. dollars.

Table 1.3: Production Function Estimates By Industry with Fixed Effects

Industry	ln_N	s.e.	ln_real_K	s.e.	ln_real_M	s.e.	Constant	s.e.	Observations
Food and Beverages	0.313***	(0.010)	0.072***	(0.008)	0.615***	(0.011)	1.089***	(0.038)	7,316
Textiles, Leathers, and Garments	0.347***	(0.010)	0.063***	(0.006)	0.590***	(0.009)	1.190***	(0.070)	7,887
Wood and Furniture	0.295***	(0.034)	0.072***	(0.017)	0.633***	(0.031)	1.364***	(0.523)	1,201
Chemicals and Pharmaceuticals	0.333***	(0.014)	0.063***	(0.010)	0.604***	(0.015)	1.497***	(0.016)	3,270
Non-metallic and Plastic materials	0.312***	(0.016)	0.066***	(0.009)	0.622***	(0.016)	1.009***	(0.065)	4,016
Metals and Machinery	0.292***	(0.012)	0.060***	(0.007)	0.649***	(0.012)	4.986***	(1.819)	5,763
Electronics	0.309***	(0.025)	0.044***	(0.014)	0.647***	(0.023)	1.433***	(0.240)	1,140
Auto and Auto Industry	0.352***	(0.033)	0.055***	(0.016)	0.593***	(0.034)	0.775***	(0.015)	662
Other Manufacturing	0.311***	(0.014)	0.038***	(0.007)	0.651***	(0.014)	1.142***	(0.037)	4,092
Retail and Whole Sale Trade	0.423***	(0.058)	-0.039	(0.038)	0.616***	(0.050)	2.218***	(0.089)	298

Notes: OLS regression for each industry is run separately with country and year fixed-effects. All variables represent log transformation $ln(x+1)$.
R is total sales of the firm, K is the net book value, N is the wage bill, M is intermediate inputs. All values are in 2005 U.S. dollars..
Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 1.4: Production Function Estimates with Fixed Effects

VARIABLES	(1) CD	(2) CRS-CD	(3) Translog	(4) Kmenta
ln_N	0.313***	0.318***	0.316***	0.332***
	(0.006)	(0.005)	(0.013)	(0.005)
ln_real_K	0.060***	0.061***	0.079***	0.055***
	(0.003)	(0.003)	(0.010)	(0.003)
ln_real_M	0.620***	0.621***	0.614***	0.614***
	(0.005)	(0.005)	(0.005)	(0.005)
ln_N_sq			0.019***	0.017***
			(0.002)	(0.001)
ln_K_sq			0.016***	0.017***
			(0.001)	(0.001)
ln_N_K			-0.036***	-0.035***
			(0.003)	(0.003)
Constant	1.250***	1.220***	1.168***	1.183***
	(0.058)	(0.057)	(0.065)	(0.056)
Observations	33,313	33,313	33,313	33,313
R-squared	0.940		0.941	
ES				3.939

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: OLS regression for each specification includes year and country fixed effects. Results are similar when industry fixed effects are used. All variables represent log transformation $ln(x+1)$. R is total sales of the firm, K is the net book value, N is the wage bill, M is intermediate inputs. All values are in 2005 U.S. dollars.. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 1.5: Level Markups by Sector

industry_n	**median**	**mean**
Food & Beverages	1.235	1.993
Textiles , Leather , Garments	1.338	2.122
Wood and furniture	1.389	2.083
Chemicals and pharmaceuticals	1.228	2.086
Non-metallic and plastic materials	1.217	1.996
Metals and Machinery	1.248	2.017
Electronics	1.255	2.150
Auto and auto components	0.975	1.648
Other manufacturing	1.373	1.992
Retail	1.117	2.070
Total	1.273	2.034

Table 1.6: OLS regression: markup and the number of competitors

VARIABLES	(1) markup	(2) markup
1 Competitor	0.012	
	(0.036)	
2-5 Competitors	-0.089***	
	(0.023)	
> 5 Competitors	-0.093***	
	(0.022)	
Num_comp		-0.006
		(0.007)
Observations	15,861	5,167
R-squared	0.058	0.188

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The dependent variable is the natural logarithm of markup which is calculated as the ratio of the output elasticity from the production function to share of expenditure of on intermediate inputs. Column (2) measures the interval for number of competitors and coded as (no competitors= 0; Competitor=1; Competitors between 2-5; Competitors >5. In column (2)) Num_comp measures the natural logarithm of one plus number of competitors reported by the firm. Both columns include country and year fixed effects.

Table 1.7: Probit, Logit and OLS regression for markup and bribes

Dep Var:	sales-based bribe				contract-based bribe	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Probit	Logit	OLS	Probit	Logit	OLS
ln_markup	-0.008***	-0.008***	-0.009**	-0.013***	-0.013**	-0.017**
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.007)
ln_age	0.000	0.000	0.003	0.003	0.024	0.001
	(0.000)	(0.000)	(0.004)	(0.004)	(0.059)	(0.005)
exporter	0.008	0.010	0.009	-0.003	-0.002	-0.005
	(0.013)	(0.014)	(0.007)	(0.017)	(0.018)	(0.011)
capital	0.039	0.032	0.013	0.04	0.04	0.019
	(0.010)	(0.010)	(0.009)	(0.013)	(0.12)	(0.016)
forshare	-0.049***	-0.045***	-0.027***	-0.006	-0.006	-0.005
	(0.025)	(0.082)	(0.023)	(0.018)	(0.018)	(0.016)
govshare	-0.090***	-0.076***	-0.038*	-0.059	-0.61	-0.027
	(0.02)	(0.199)	(0.024)	(0.045)	(0.044)	(0.047)
Country FE	yes	yes	yes	yes	yes	yes
Sector FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
Observations	22,284	22,284	22,486	3,722	3,722	3,722
R-squared			0.113			0.172

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Column 1 & 4 report estimated marginal effects from the probit model. Column 2 & 5 report values from the logit. Column 3 & 6 report values from linear probability model. All columns report robust standard errors clustered at the country-sector level in brackets. In column 1-3, the dependent variable is a binary variable that indicates whether firm has paid bribe, measured as a percentage of total sales. In column 4-6, the dependent variable is a binary variable that indicates whether firm has paid bribe, measured as a percentage of government contract. All columns control for firm-level characters: age, export, capital, ownership.

Table 1.8: Tobit model for markup and bribes

Dep Var:	sale-based bribe		contract-based bribe	
VARIABLES	(1)	(2)	(3)	(4)
ln_markup	-0.034	-0.056*	-0.046***	-0.096***
	(0.027)	(0.033)	(0.009)	(0.013)
ln_age		0.022***		-0.128***
		(0.06)		(0.012)
exporter		0.057		-0.213***
		(0.062)		(0.034)
capital		0.118*		0.093***
		(0.068)		(0.035)
forshare		-0.300***		-0.105***
		(0.087)		(0.030)
govshare		-0.419*		-0.120**
		(0.215)		(0.053)
Observations	29,518	22,485	9,307	3,722

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Reported values are estimated from the tobit model with country, industry, and year fixed effects. Clustered standard errors at the country-sector level in brackets. The dependent variable in column (1) & (2) is bribe measured as percentage of total sales. The dependent variable in column (3) & (4) is bribe measured as percentage of contract value. Column (2) & (4) include firm-level controls, such as, age, export, capital, ownership.

Table 1.9: Tobit model for competition and corruption (graft_sales)

Dep Var:	sale-based bribe			contract-based bribe		
	(1)	(2)	(3)	(4)	(5)	(5)
VARIABLES	bribe_sales	bribe_sales	bribe_sales	sales_contract	sales_contract	sales_contract
Num_comp	0.072			0.194***		
	(0.067)			(0.024)		
Informal_Comp		0.444***			0.944***	
		(0.061)			(0.029)	
1 Competitor			-0.276*			-0.213
			(0.149)			(0.189)
2-5 Competitors			0.057			-0.151
			(0.110)			(0.189)
> 5 Competitors			0.245**			0.352*
			(0.124)			(0.183)
ln_age	0.089	0.043	-0.074**	0.093***	-0.135***	-0.140***
	(0.066)	(0.032)	(0.033)	(0.023)	(0.012)	(0.051)
exporter	0.199	0.103	0.083	0.695***	-0.083**	-0.069
	(0.137)	(0.067)	(0.074)	(0.067)	(0.033)	(0.105)
capital	-0.148	0.076	0.269**	0.249***	0.045	0.226
	(0.108)	(0.069)	(0.107)	(0.055)	(0.032)	(0.221)
prishare	-0.071	-0.250***	-0.397***	-1.334***	0.041	-0.168
	(0.257)	(0.095)	(0.094)	(0.047)	(0.029)	(0.129)
govshare	-0.068	-0.189	-0.270	-2.929***	-0.869***	-0.574
	(0.483)	(0.217)	(0.252)	(0.096)	(0.043)	(0.487)
Observations	7,794	30,792	15,422	1,692	5,375	8,310

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Reported values are estimated from the Tobit model with country, industry, and year fixed effects. Clustered standard errors at the country-sector level in brackets. The dependent variable in column 1-3 is bribe as a percentage of total sales. The dependent variable in column 4-6 is bribe as percentage of contract value. All columns include age, export, capital, ownership as firm-level control.

Chapter 2

Competition, Corruption, and Nonresponses at the Firm-level

2.1 Introduction

In this paper I investigate the relationship between firm-level competition, business constraints, managerial traits, and missing corruption observations using the World Bank Enterprise Survey (henceforth, WBES). The paper focuses on three important questions: Do high-markup firms have more contact with public officials for a service? Are high-markup firms more likely to make a payment for bribes? For high-markup firms, what are the response rates on the payment question? The motivation for these questions can be summarized as follows. First, despite the unanimity about the effects of market competition on corruption at the firm-level, the effects of the nonresponse to the corruption-related questions has not been measured in this literature. The general model that links competition to corruption hinges on truthful responses to corruption-related questions that are close to actual estimates. This implies that to understand corruption, it is important to first look at firm-level characteristics that determine the exposure, incidence, and nonresponses to corruption (bribery). Yet, most studies that have utilized firm-level data consider the amount stated by the respondent accurate, and ignore the implications of nonresponses and false responses in the data.

Second, Mendez and Sepulveda (2010) theoretically show corruption equilibria differs in regards to behavior and incentives; therefore, different measures of corruption can lead to very different estimates. Before availability of firm-level data, early research focused on country-level indicators based on expert opinions. More recently, studies have used firm-level data that measures the firms' experience in dealing with corruption; however, researchers still continue to differ in what defines corruption.[1] For instance, Svensson (2003) use data on incidences (frequency) to bribes, while Clarke and Xu (2004), Alexeev and Song (2013) uses data on amounts paid by private firms (or percentage of total sales). Similarly, Olken (2007) use differences in total reported expenditures and total expenses expenditures in Indonesia. Therefore, by providing a comprehensive sequence of events to derive the results, I attempt to find consistency in firm-level data. I focus on looking at different situations in which a firm has to have contact with a public official i.e. to apply for government contracts and licenses.

The main results of the paper can be summarized as follows. My research first matches the exposure to bribery to informal payments to show that high-markup firms are more likely to be self-selected, which creates a biased sample population. The paper identifies four different categories of corruption[2] which includes: dealing with tax officials, securing a government contract, obtaining import licenses, and obtaining operating licenses. Second, the paper determines the probability of a firm being requested (or expected) to pay at each of these contacts. The results show that there is a systematic self-selection in the data generation process. High-markup firms, operating in less competitive environments, are more likely to have contact with public officials, but less likely to actually make a payment. It is important to note that I do not address the nonresponses in the initial phase of questions. Respondents that state "no" on whether a service was applied are assumed to have paid no bribes. My research then matches the nonresponses to the corruption-related question

[1]See Bardhan (2006) for a similar description.

[2]This paper uses government officials, public officials, and higher authority interchangeably. Likewise, there is a distinction between payments made and whether a firm came in contact with a public official. If the firm applied for a service or is visited by a tax official, I collective term this as exposure (contact) with the public official. Similarly, the paper also defines expected payments or payments made to a public official for a service as corruption. The answer to this question depends on how one chooses to measure corruption. For instance, to secure a government contract, the question is framed as: "what percentage of the contract value is paid in bribes?", while to obtain an import license, the question is framed as: "was the firm expected to make a payment?". In both cases, I consider a payment made if the firm stated a positive value in the former and an affirmative response in the later.

at each incidence to show that the high-markup firms are more likely to not respond to the question on payments. These results, therefore, imply that the negative correlation between weaker competition and corruption, generated in the second phase and by other research papers, might not be representative. It also suggests that competition (determined by markups) is important to determine the response rate to corruption-related questions.

This paper makes two important contributions to the literature. To my knowledge, this paper is the first to exploit different firm-level controls to assess whether the competition-corruption relationship continues to hold at each sequences (i.e. exposure, incidence, and nonresponses). Moreover, understanding which firm-level controls influence nonresponses on corruption-related questions is crucial in order to accurately evaluate the impact of markup. While the amount of nonresponses has been considered a problem in several cross-national studies, little research has focused on the data generation process and solutions to address the systematic self-selection to the corruption related questions.[3] Second, the research highlights the importance of sample selection bias on the measure of competition and reveals that high-up firms and government-owned enterprises determine the likelihood of responses to corruption-related questions.

The rest of the paper proceeds as follows. Section 2 discusses previous firm-level literature between competition and corruption, and highlighting the importance of nonresponses to corruption-related questions. Section 3 provides data and summary statistics. Section 4 provides three different empirical series and section5 provides a brief conclusion.

2.2 Literature Review

Before exploring the literature on incidences to bribes, the paper highlights research that considers market competition as an approach to dealing corruption.[4] This paper, therefore, complements the more recent empirical development in this literature that relies on firm-level survey data. My research shows only two other published papers on this topic by Alexeev

[3]For instance, see Jensen et al. (2010) and Berinsky and Tucker (2006)

[4]Please refer to chapter 1 for a complete discussion on the theoretical side of the competition-corruption. In addition, the focus in this chapter is more on research that have utilized firm-level data. A detailed literature review on papers that have utilized country-level data has been carefully crafted in Chapter 1.

and Song (2013) and Diaby and Sylwester (2015); both papers find a positive association between market competition and the amount of bribes paid. However, it is important to note that the aforementioned papers find no statistical significant relationship between firm-level markup and corruption. In addition, both papers drop the missing bribe observations which can create a systematic bias in the data. For instance, Alexeev and Song (2013) consider missing observations 'white noise' and exclude them from the sample. This is noteworthy since ignoring almost half of the sample to make inference on the competition-corruption relation could result in incorrect inferences. If firms that face competition are self-selected by different government officials, the estimated results on corruption might not be accurate due to biased sample.

The focus of this paper is, therefore, to determine the response rates to corruption-related questions. Despite all of the work on the causes and consequences of corruption,[5] there has been little work on the nonresponses to corruption-related questions, especially at the firm-level. Among other things, the amount of nonresponses has been attributed to the national political environment in a country where the firm operates. Using data from the WBES, Jensen et al. (2010) show that both at the national and firm-levels, countries with lower levels of political freedom are more likely to not answer bribe-related question and are more likely to provide a false response. Similarly, the authors find firms report corruption as being less severe than those reported by the Kaufmann et al. (2007) corruption index. Another paper by Svensson (2003) addresses the issue of nonresponses when estimating the incidences (or exposure) to bribes. Using the Ugandan enterprise survey, they find considerable variation in reported graft across firms. They, however, find no differences in the firm-level characteristic for the firms that refused to respond and firms that respond to corruption-related question.[6] A closely related paper by Rand and Tarp (2012) also looks

[5]See Dimant and Tosato (2017) for detailed survey of the corruption literature. With the availability of different datasets, there has been a large increase in the literature on the causes and consequences of corruption. For instance, more earlier studies on the causes and consequences were conducted by Rose-Ackerman (1978), Shleifer and Vishny (1993), and Bardhan (1997). More recent studies look at the effect of corruption by focusing on foreign direct investments and economic growth in the long-run. For instance, see Wei (2000), Svensson (2005), Knack and Keefer (1995), Mauro (1995).

[6]Svensson (2003) use a dummy variable (=1) for missing bribe responses as the dependent variable and firm size, profit, capital stock, and capital as the main independent variables. More specifically, they estimate whether the coefficients from this regression are different from a model where the dependent variable is whether firm reported a positive bribe (=1).

at the incidence to bribe in Vietnam and find that the mechanisms for the incidences and amount of bribes paid vary significantly. They also find that bribes is significantly correlated when dealing with government official with regards to a contract; however, this result does not hold when dealing with custom officials. Therefore, my study aims to take this empirical strategy one-step further by also generating probabilities on refusal to answer key corruption-related question.

2.3 Data

The data used for this study comes from the Enterprise Surveys of the World Bank (WBES). The data includes 141 countries and covers 14 main manufacturing industries from 2006 to 2016.[7] The manufacturing industries have been classified according to the major 2 digit ISIC code.[8]

2.3.1 Corruption Variables

The paper identifies four different categories of corruption which includes exposure or payments to different situations: dealing with tax officials, securing a government contract, obtaining import licenses, and obtaining operating licenses. For each of these four situations, three sequences of estimation are performed. First, to determine the probability of being self-selected, I look at responses to the question on whether the firm has exposure (or contact) to a public official in the aforementioned scenarios. For example, I look at whether high-markup firms are more likely to apply for an import license or be visited by a tax official. Table 2.1 summarizes the key corruption variables used in the analysis. 59.42% of firms reported being visited by a tax official in the previous 12 months. Similarly, 17.58% of firms attempted to secure a government contract in the previous 12 months. Likewise, 12.62% (24.44%) of firms applied for an importing license (operating license) in the past 12 months.

Second, to determine the probability of whether an informal payment or gift is made or

[7]Sample updated September 2016. Refer to table 2.10 for a list of countries in the sample.

[8]In particular, the industries are: food and beverages; textiles; leathers; garments; products, leather, and footwear; wood and furniture; chemicals and pharmaceuticals; non-metallic and plastics; metals, machinery and equipment; electronics; auto and auto components; retail; and other manufacturing.

expected, responses conditional on exposure, are analyzed. For example, given the responses provided in the first question, I look at whether high-markup firms are more likely to make a payment. 14.57% of firms reported that a gift or informal payment was expected or requested when visited by a tax official. Similarly, 20.0% of firms reported paying bribes to secure a government contract. Likewise, 13.20% (16.14%) of firms reported that a gift or informal payment was expected or requested when applying for an importing license (operating license).

Finally, to determine the probability of a firm's nonresponse to payments, I look at missing observations in regards to the payment question contingent on the firm having a contact (exposure) with a public official. More specifically, this research intends to evaluate survey responses "do not know" and "refuse to answer" given the firm was in contact with public officials. 5.2% of firms refused to answer whether a gift or informal payment was expected or requested when visited by a tax official. Similarly, 85.4% of firms refused to answer whether bribes were paid to secure a government contract. Likewise, 6.1% (5.2%) of firms refused to answer on whether a gift or informal payment was expected or requested when applying for an importing license (operating license).

2.3.2 Markup

The main interest is to figure out whether the high-markup firms are self-selected in the data generation process. Therefore, the main independent variable is the level of competition or markups. The solution recently suggested by De Loecker and Warzynski (2012) is to include an assumption that firms are cost-minimizers and choose at least one freely adjustable input. They calculate markup as the ratio of output elasticity of an input(s) to the total expenditure share of the input(s) and relate these to firm level export status. An advantage to this approach is that it is relatively easy to estimate firm specific markup without requiring any information on the market structure and the firm's input demand (De Loecker and Warzynski, 2012). I, therefore, follow this simple intuitive idea that relies the price=marginal cost condition for firms with a perfectly elastic demand curve. Consequently, at this point the total labor expenditure is equal to the total revenue (or total sales) and any difference between the two constitutes the markup.

To calculate markups, first I estimate the labor expenditure share using data on labor costs and total revenue. The WBES defines labor costs as the total annual wages and all annual benefits, including food, transport, and social security (i.e. pensions, medical insurance, and unemployment insurance) and total sales as the value of all annual sales, including manufactured goods and goods the establishment buys for trading. Second, I regress country, sector, and ownership fixed-effects on labor expenditure share to retrieve country and sector specific coefficients. After that, the linear predicted values on the labor expenditure share are estimated. For simplicity, the predicted values for each firm i in industry j located in country c will be the same i.e. the predicted values will differ by each country and each industry. The final step is to calculate the differences between the (observed) labor expenditure share and the predicted values. If the firm is operating in a perfectly competitive setting, the difference between the two values should be zero. Likewise, if the firm is operating in an imperfect setting, the difference between the two values will be the markups.

Table 2.2 provides the summary statistics for markups by sector averages. Before estimating these results, I have eliminated questionable outliers; therefore, firms with large absolute values after the log transformations on markups that result in zero, negative, or missing values are eliminated from the sample. This includes trimming the the top and bottom 1% from the sample. The mean of markup is 2.02 with a standard deviation of 4.5. There is considerable differences for markups across industries. The three sectors with higher markups are leather (2.25), retail (2.29), and electronics (2.11). Conversely, the three sectors with lowest markups (at the level) are the auto and its components (1.73), garments and textiles (1.78 and 1.84, respectively), and metals and machinery (1.87). Similarly, in table 2.5 , I look at different firm-level traits and markups. Firms that export have a high-markup (2.24) compared to non-exporters (1.96). This result is consistent with De Loecker and Warzynski (2012) who find that exporters have higher markups due to the productivity premiums. Firms with male owners have a higher markup (2.03) compared to female owners (1.98); however, the difference is minimal. Similarly, firms with a quality-controlled certificate have higher markups implying that there is some degree of power associated with obtaining a certificate approval. Foreign and government firms also have higher markups.

Before presenting the main empirical results, I first look at firm-level characteristics that determine markup. Table 2.5 shows the results for the Ordinary Least Squares (OLS) regressions where the dependent variable is log of markup. The main independent variables include the firm's export and ownership status, age, certificate on quality, and the respondents opinion on whether access to finance and transportation are an obstacle to the current operation and growth of the firm. Columns 2, 3, and 4 controls, respectively, for sector, country, and time fixed-effects, while column 5 includes all three fixed-effects. Standards errors are clustered at country-sector-year levels in all columns.

The results can be summarized as follows. There is a positive and statistically significant correlation between firms that export, compared to non-exporters, and markups implying that firms that export have some degree of monopoly power. Similarly, the coefficient on certificate is also positively correlated with markup implying that the internationally-recognized certificate allows firm to have higher markups. More established firms are also likely to exhibit some degree of market power. These results are statistically significant at the 1% level and are robust to the introduction of different fixed-effects. Similarly, there is a positive correlation between the respondent that considers transportation an obstacle to the current operation and markups implying that the high-markup firms are geographically protected. On the contrary, there is a negative correlation between the respondent that considers access to finance an obstacle to the current operation and markups implying that readily available finance can facilitate competition, thereby reducing monopoly power of a firm.

2.3.3 Firm-level determinants to bribery

Following previous firm-level studies, I control for firm-level characteristics that can influence the relationship between competition and the incidences, payments, and amount of nonresponses to corruption. Batra et al. (2003) find that private firms are more likely to bribe, pay a higher revenue share as bribes, and more likely to consider bribe as an obstacle. Similarly, firms with large private, foreign, or government shares could face different bribe environments in dealing with public officials. Hence, I include a dummy variable equal to one for firms with more than 50% of ownership by private foreign individuals, companies, or organizations ($forshare$), and two for more than 50% ownership by government or state ($govshare$), while considering the percentage of shares owned by the private domestic individuals, companies, or organizations as a benchmark. I also include a variable on export status, which is a dummy variable representing a value one if the sum of the indirect exports[9] and direct exports (denoted as a percentage of the establishment's sales) is greater than 5%, zero otherwise (i.e., national sales is coded zero). Studies have shown that firms that export internationally, as compared to domestic sales, might be more prone to government rent extraction and therefore, are more likely to bribe to continue operation. Certificate is a dummy variable where one represents the firm has an internationally-recognized quality certification, 0 otherwise. Approximately 11.03% of the firms have some degree of foreign share and 2% of firms have some degree of government share.[10] About 27% of firms have the internationally recognized quality certificate and 20% of the firms are exporters. Summary statistics reported in table 2.4 and table 2.5.

Age is calculated as the logarithmic difference between the survey year and and the year in which the establishment started its operation. The suggested direction of *age* is mixed in the literature. It is possible that old firms are more established and have networks with government officials and, therefore, need to bribe less to remain in the market. On the other hand, young firms may be more likely to bribe if the market is competitive. The same may be expected for the manager's experience level as a relatively more experienced manager might be able to report the requested bribes to a higher authority because of established networks. Gender of the owner can also determine the response rate and incidence to bribes since males and females could have different response behavior in each scenario. These latter two set of controls are linked to the managerial traits of the firm that can influence corruption. The mean level of manager's experience is 16.3 years with a standard deviation of 11.3 years. Likewise, the average age of the firm is 17.5 years with a standard deviation of 14.56 years. Similarly, 29% of the firms have a female owner.

Other barriers to the firms current operations are also considered: crimes, theft, disorder;

[9]Indirect exports include the percentage of sales sold domestically to third party that exports products.
[10]For this estimation, I look at foreign and government ownership greater than 1 %.

electricity; transportation of goods, supplies, and inputs; access to financing; and tax rates. For each obstacle, a binary variable is create that takes on a value of zero if the respondent states “no”, “minor”, and “moderate” on whether the aforementioned business constraints affect current operation of the firm. Responses “major” and “very severe” are coded as one. These questions are tied to the perception of the respondent; however, including them in the model reveals how business constraints can affect the corruption incidences and amount of nonresponses in the data. About 19%, 35%, and 8% of the respondents consider crime, power, and the interpretation of rules by public officials an obstacle to current operation. All variables and descriptions are carefully outlined in 2.9.

2.4 Empirical Analysis

The goal of the paper is to employ a sequence of estimations to determine whether the high-markup firms are indeed more likely to not respond to the corruption-related questions. Therefore, I carefully outline each sequence in the following sections.

2.4.1 Do high-markup firms have more exposure to public officials?

The empirical approach first consists of estimating an establishment's probability of being self-selected in four different scenarios: visit by a tax official, attempt to secure a government contract, submit an application for an import license, and operating license. The objective is to the test whether high-markup firms are more likely to be in a situation where extraction of bribes is possible. More specifically, the following probit model is estimated:

$$P(exposure_{it} = 1) = \Phi(\beta_1 markup_{it} + \beta_2 X_{it} + \beta_3 obstacles_{it} + \delta_c + \delta_t + \delta_j + \epsilon_{it} > 0) \quad (2.1)$$

The dependent variable is a binary that takes on a value of one if the respondent answers yes to the question on whether the firm applied for a services (government contract, import, operating licenses) or was visited by a tax official, and zero if the respondent answers no.

The subscript notation is described as follows: a firm i is located in country c at a given time t. Note that the empirical analysis only focuses on situations where the firm comes in contact with a public official with the assumption that firms that do not come in contact are not likely to make bribe payments.

All specifications include firm-level controls (X_{it}), business obstacles ($obstacles_{it}$), and country, industry, and year fixed effects (denoted as δ_c, δ_j, and δ_t respectively). Since the sample includes 141 different countries from four continents, the inclusion of country fixed effects controls for countrywide factors: country's legal system, legal origin, rule of law, and regulation of various economic activities that could influence corruption (Treisman (2000a)). The inclusion of industry fixed effects is to control for the exogenous variation in different industries that can influence the relationship between markups and bribe payments. In the same vein, the inclusion of time fixed effects is to capture any macroeconomic trend or policies of a country that could potentially influence the firm's intention to apply for a services (or be visited). All estimations use robust standard errors clustered at the country-sector-year level to allow errors to be correlated within the same country, industry, and year. Column 1 (2) regresses markup on exposure to visit from tax officials (with fixed-effects), column 3 (4) regresses markup on government procurement (with fixed-effects), column 5 (6) and 7 (8) regresses markup on applying for import license (with fixed-effects) and operating license (with fixed-effects), respectively.

The coefficients from estimating the probit model are presented in table 2.6. The effect of markups on exposure to bribery is positive and statistically significant at the 1% level in all columns. For interpretation, the marginal effect of markup in column 2 is 0.01 (not reported) implying that an infinitesimal increase in markup enhances the probability of being visited by tax official by 0.1 percentage point. In all cases, high-markup firms are self-selected and are more likely to be in contact with public officials that could potentially increase the probability of bribing. Therefore, in this set-up, firms that do not meet with public officials do not bribe, compared to firms that frequently need to visit different public officials.

Several firm-level features are worth noting. The firm's foreign ownership status is positively correlated with each exposure to bribery. The probability of a visit by a tax official, securing a government contract (not significant), applying for an operating license, and applying for an import license increases with foreign ownership. The latter two cases are likely associated to the operating regulations of each country that the foreign firm needs to apply to start or remain in business. Similarly, it could be that foreign firms have less networks (or association) in the host country of business, which increases the probability of exposure to different public officials.[11]

Other surprising results are that the probability of submitting an import and operating license, and securing a government contract increases with having a female owner (compared to male owner.) Similarly, for attempts to secure a government contract and import license, the probability of exposure to bribery decreases with an increase in the manager's level of experience. This supports the hypothesis that experienced managers already have established networks and are less likely to be exposed to public officials. The results for firm's age is mixed. Excluding the application for operating licenses, an increase in the firm's age enhances the probability of being exposed to bribery.

Next, I look at different business constraints that determine the exposure to bribes. Firms that consider crime, electricity, interpretation of the rules and laws, and tax rates as a major and severe obstacle to its current operation are also likely to enhance the probability of being in contact with different public officials. Overall, the results strongly suggest that high-markup firms are likely to be self-selected by public officials in all four categories. This result also holds models with fixed-effects. Having contact with public officials, therefore, could increase bribe extraction for the high-markup firms. As a consequence, the next empirical strategy is to test whether high-markup firms made a bribe payment at the aforementioned scenarios.

[11]This result is consistent with Chatterjee and Ray (2012) who show that there is a positive association between exports and foreign ownership and bribe demands.

2.4.2 Are high-markup firms more likely to pay bribes?

The second empirical strategy is to estimate the probability of an informal payment[12] made given that an exposure was made. The goal is to compare this outcome to the probability of no extortions (no requests/expectation) conditional on the probability of being exposed (or having an incidence to bribery). As mentioned earlier, firms that do not have contact with public officials do not bribe i.e. to bribe, the firm has to be in one of situation where bribe extraction is possible. To look at this, the following probit model is estimated:

$$P(pay_{it} = 1|expo_{it} = 1) = \Phi(\beta_1 mu_{it} + \beta_2 X_{it} + \beta_3 obst_{it} + \delta_c + \delta_t + \delta_j + \epsilon_{it} > 0) \qquad (2.2)$$

The dependent variable is a binary that takes on a value of one if the respondent answers yes to the question on whether a gift or informal payment was requested or made, and zero if the respondent answers no. All other firm-level controls and business obstacles remain same as equation 2.1.

Table 2.7 shows the results for the probit model with clustered standard errors at the country, industry, and year level. In all columns, the coefficient on markups is insignificant and negative. Thus, while the probability of being exposed increases for high-markup firms, at each event, the probability of an informal payment decreases. These results are not surprising as stronger (weaker) market competition has been associated with greater (less) corruption in more recent firm-level studies;[13] however, both studies do not show that the probability of payment decreases and this is contingent on high-markup firms being self-selected in the first place.

Next I look at managerial skills and traits that determine the informal payments. The probability of having a payment requested decreases with the manager's experience level. This could be due to the manager's established networks or ability to report the bribes

[12]Note that informal payments, expected payments, and payments made all refer to the same issue that bribery occurred. The difference in the response lies in the way the question is posed and on how one chooses to measure corruption. For instance, to secure a government contract, the question asks:" what percentage of the contract value is paid in bribes?", while to obtain an import license, the question asks:" was the firm expected to make a payment?". In both cases, I consider a payment made if the firm stated a positive value in the former and an affirmative response in the latter.

[13]For instance, see Alexeev and Song (2013) and Diaby and Sylwester (2015).

demanded to a higher authority. Also related to managerial skills, the coefficient on female is mostly negative implying that, compared to male counterparts, female owners are less likely to be requested an informal payment or gift. As Swamy et al. (2001) and Mocan (2008) explain, this maybe due to the lower labor force participation of females in many developing countries, and therefore females are less likely to be in contact with public officials.

Surprisingly, although foreign firms are more likely to have contact with public officials, the probability of having a payment requested at these incidents actually decreases. It is also interesting to see firms that report tax rates, crimes, and power as moderate and severe obstacles to the current operation are more likely to make a payment. Thus, while high-markup firms are in more contact with public officials, the expected payment made or requested is actually not significant for the high-markup firms. This result is puzzling as high-markup firms are self-selected by the public officials, yet the amount of bribes paid (expected) is negative. Next, the paper explores whether the high-markup firms are choosing to not respond questions on payments.

2.4.3 Are high-markup firms more likely to not respond to corruption related questions?

The third sequence is to estimate the probability of the nonresponses to payment contingent on firm's contact with different public officials. Therefore, the paper compares this aforementioned probability to the benchmark category, where firms respond to whether a payment was made (either yes or no) conditional on the firm's contact with a public official. More specifically, the following probit model is estimated:

$$P(ref_{it} = 1 | exp_{it} = 1 \& pay_{it} = miss) = \Phi(\beta_1 markup_{it} + \beta_2 X_{it} + \beta_3 obstacles_{it} + \delta_c + \delta_t + \delta_j + \epsilon_{it} > 0) \tag{2.3}$$

Once again the dependent variable is a binary that takes on a value of one if the firm refuses to answer the question on payments but made contact with a public official and zero otherwise. All other firm-level controls and business obstacles remain same as equation 2.1.

Table 2.8 presents the probit coefficients from estimating equation 2.3. The first row of the table shows that an increase in markup, increases the probability of not responding to the payment question. This positive correlation holds for payments made to tax officials and operating licenses which are both significant at the 10% level. However, there is a negative correlation between high-markup firms and attempts to secure government contracts. Since the number of observations for government contract is relatively small, the results might not be representative. Similarly, in column 5, the probit coefficient for payments on import licenses is significant at the 10% level, but the significance disappears once fixed-effects are added.

Other noteworthy variables include the manager's level of experience which is mostly negative and significant implying that the probability of not responding the payment question decreases with an increase in the manager's experience level. This result is interesting because an increase in the manager's experience also decreases the probability of a bribe payment, and increases the probability of being in contact with a public official. Likewise, firms that are predominately owned by the government (more than 50% ownership) are also more likely to not respond to the payment question. This result is surprising because, as shown in the previous section, state-owned enterprises (SOE) are less likely to make payments to obtain one of the services. It also implies that information on corruption obtained from SOEs could be underestimated and not accurate. Similarly, firms that report government official's interpretation of rules and laws as consistent are more likely to provide responses. This shows the country's legal and political institutions could be important in obtaining responses to corruption-related questions. This is consistent with Jensen et al. (2010) who show the country's political environment determines the amount of nonresponses in firm-level data.

2.5 Conclusion

Beyond understanding the probability of exposure to a public officials and whether a bribe is paid, my research explored the relationship between the high-markup firms and nonresponses to corruption. I find strong empirical support that high-markup firms are more likely to be exposed to corruption. For instance, high-markup firms are more likely to be visited by tax officials and more likely to secure a government contract. I also find that high-markup firms are less likely to (expected to) pay bribes; however, the results also show that the high-markup firms are more likely to not respond to the corruption-related question. The results hold instances when a firm is visited by a tax official and applies for an import and operating license. In contrast, the amount of nonresponses to payments on government contract is negative implying that the exposure process can also determine the response rate to corruption. My research sheds light on the potential biases caused by nonresponse in the competition- corruption link. Further analysis needs to be done to look at the trade theories behind the results.

Table 2.1: Summary statistics for Corruption Variables

Variable	**Mean**	**Std. Dev.**	**N**
Contact or Exposure to Bribe			
tax_inc	0.594	0.491	122771
gov_inc	0.176	0.381	102178
imp_inc	0.126	0.43	122543
ope_inc	0.244	0.429	122548
Informal payment or gift requested or expected			
tax_pay	0.146	0.353	68883
gov_pay	0.200	0.400	14423
imp_pay	0.132	0.339	14459
ope_pay	0.161	0.368	28152
Nonresponse to Bribe			
tax_ref	0.052	0.222	72675
gov_ref	0.854	0.353	4137
imp_ref	0.061	0.24	15404
ope_ref	0.052	0.222	29689

Notes: All variables are described in table 2.9.

Table 2.2: Summary Statistics: Level Markups by Sector

sector	**mean**	**p25**	**p50**	**p75**	**sd**
Textiles	1.781	0.526	0.900	1.604	3.755
Leather	2.255	0.423	0.750	1.699	5.432
Garments	1.849	0.507	0.846	1.573	4.242
Food	2.038	0.467	0.823	1.689	4.557
Metals and machinery	1.871	0.495	0.820	1.545	3.967
Electronics	2.113	0.496	0.827	1.655	4.788
Chemicals and pharmaceuticals	1.911	0.499	0.857	1.660	4.256
Wood and furniture	1.872	0.446	0.744	1.471	4.534
Non-metallic and	1.870	0.478	0.839	1.608	4.178
Auto and auto components	1.731	0.578	0.859	1.674	4.022
Other manufacture	1.727	0.524	0.858	1.567	3.568
Retail and whole	2.294	0.426	0.861	1.904	5.256
Hotels and restaurant	1.761	0.496	0.820	1.470	3.993
Other services	2.291	0.405	0.771	1.778	5.338
Other: Construction	2.136	0.430	0.805	1.705	4.852
Total	2.024	0.468	0.834	1.672	4.593

Table 2.3: Summary Statistics: Level Markups and firm-level determinants

Variables	ln_markup	
exporter?	Mean	Obs
no	1.969482	79032
yes	2.245828	19380
female owner?		
no	2.039468	68680
yes	1.987944	29732
certificate?		
no	1.931493	76674
yes	2.349844	21738
ownership?	mean	
private	2.009957	90213
foreign	2.186867	7202
SOE	2.108523	997

Notes: Exporter is a dummy variable (=yes) if firm is an exporter (percentage of the establishment's direct and indirect sales > 5), 0 otherwise. Certificate is a dummy variable (=yes) if the firm has an internationally-recognized quality certification, 0 otherwise. Female is a dummy variable (=yes) if firm is owned by female, 0 otherwise. Dummy variable (foreign) if firm foreign owned (more than 50 % ownership by foreign individuals, firms, or corporations); Dummy variable (SOE) if firm SOE (more than 50 % ownership by government); Benchmark is the private owned firms.

Table 2.4: Summary statistics for firm-level controls

Variable	Mean	Std. Dev.	N
female	0.296	0.457	124351
ln_exp	2.679	0.687	120597
exp	16.336	11.365	123266
age	17.526	14.566	122201
ln_age	2.656	0.738	122201
ownership	0.092	0.321	124351
exporter	0.201	0.401	124351
tax_rates	0.311	0.463	124351
crime	0.192	0.394	124351
power	0.352	0.478	124351
gov_off	0.089	0.285	124351
informal	0.251	0.433	124351

Notes: All variables are described in Appendix table 1.

Table 2.5: OLS regression of markup on firm-level determinants

	(1)	(2)	(3)	(4)	(5)
VARIABLES	ln_markup	ln_markup	ln_markup	ln_markup	ln_markup
exporter	0.066***	0.069***	0.070***	0.066***	0.070***
	(0.018)	(0.018)	(0.016)	(0.017)	(0.016)
certificate	0.151***	0.161***	0.154***	0.148***	0.162***
	(0.017)	(0.016)	(0.015)	(0.016)	(0.014)
trans	0.053***	0.052***	0.063***	0.051***	0.062***
	(0.013)	(0.012)	(0.011)	(0.012)	(0.011)
finance	-0.076***	-0.077***	-0.099***	-0.081***	-0.103***
	(0.012)	(0.012)	(0.010)	(0.011)	(0.009)
ln_age	0.043***	0.046***	0.042***	0.043***	0.041***
	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)
countryFE	no	no	yes	no	yes
sectorFE	no	yes	no	no	yes
timeFE	no	no	no	yes	yes
Observations	97,296	97,296	97,296	97,296	97,296
R-squared	0.008	0.009	0.018	0.010	0.021

Notes: ***, **, * denotes statistical significance at 1, 5 and 10 percent respectively. Constants are included in the model but are not reported. Reported values are estimated from OLS model with clustered standard errors at the country-sector-year level in brackets. The dependent variable is log of markup. All columns include various firm-level controls: ln(age), export status, binary on certificate for quality, binary on whether access to finances and transportation are considered an obstacle.

Table 2.6: Probit model for Contact (Exposure) to Bribery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	tax_inc	tax_inc	gov_inc	gov_inc	imp_inc	imp_inc	ope_inc	ope_inc
ln_markup	0.0434***	0.0510***	0.0302***	0.0349***	0.0896***	0.102***	0.0219***	0.0299***
	(0.00706)	(0.00584)	(0.00763)	(0.00768)	(0.00745)	(0.00718)	(0.00729)	(0.00701)
forshare	0.221***	0.194***	0.0236	-0.0237	0.583***	0.486***	0.184***	0.0451**
	(0.0260)	(0.0226)	(0.0270)	(0.0257)	(0.0307)	(0.0237)	(0.0333)	(0.0224)
govshare	0.0250	-0.0444	0.279***	0.246***	-0.0282	0.147*	0.0224	0.101
	(0.0715)	(0.0894)	(0.0829)	(0.0779)	(0.0843)	(0.0789)	(0.0763)	(0.0709)
ln_exp	-0.0340**	0.0148	0.0775***	0.0649***	0.0723***	0.0344***	-0.104***	0.000830
	(0.0147)	(0.00952)	(0.0122)	(0.0110)	(0.0160)	(0.0117)	(0.0151)	(0.0114)
female	-0.0111	0.0130	0.0448***	0.0416***	0.101***	0.0415**	0.174***	0.0930***
	(0.0195)	(0.0146)	(0.0170)	(0.0157)	(0.0216)	(0.0169)	(0.0246)	(0.0164)
tax_rates	0.129***	0.137***	0.126***	0.0765***	0.0142	0.0204	0.0631***	0.0637***
	(0.0191)	(0.0164)	(0.0221)	(0.0194)	(0.0212)	(0.0157)	(0.0216)	(0.0159)
ln_age	0.00592	0.0955***	0.0689***	0.121***	0.0896***	0.0866***	-0.0495***	-0.0594***
	(0.0129)	(0.0115)	(0.0125)	(0.0107)	(0.0189)	(0.0121)	(0.0170)	(0.0113)
crime	0.0266	0.0285*	0.0452**	0.0299*	0.0946***	0.00859	0.0672***	0.0671***
	(0.0196)	(0.0159)	(0.0197)	(0.0177)	(0.0233)	(0.0171)	(0.0228)	(0.0175)
power	0.167***	0.0356**	-0.0342**	0.00987	0.0400	0.0231	0.0974***	0.0442***
	(0.0204)	(0.0161)	(0.0170)	(0.0162)	(0.0260)	(0.0154)	(0.0287)	(0.0147)
gov_off	-0.0665	0.00380	-0.154***	-0.0329	0.0354	-0.0496*	-0.140***	-0.0533*
	(0.0418)	(0.0217)	(0.0464)	(0.0377)	(0.0382)	(0.0258)	(0.0434)	(0.0272)
countryFE	no	yes	no	yes	no	yes	no	yes
sector FE	no	yes	no	yes	no	yes	no	yes
timeFE	no	yes	no	yes	no	yes	no	yes
Observations	94,857	94,857	78,233	78,233	94,759	94,759	94,933	94,933
r-square	0.00804	0.114	0.00751	0.0800	0.0261	0.140	0.0107	0.188
#_depvar	12	179	12	173	12	179	12	179
%_ones	0.594	0.594	0.176	0.176	0.126	0.126	0.244	0.244
conv	1	1	1	1	1	1	1	1

Notes: ***, **, * denotes statistical significance at 1, 5 and 10 percent respectively. Constants are included in the model but are not reported. Reported values are estimated from the probit model with clustered standard errors at the country-sector-year level in brackets. The dependent variable is a binary variable that takes on a value of one if the respondent answers yes to the question on whether the firm applied for a services (government contract, import, operating licenses) or was visited by a tax official, and zero if the respondent answers no. All columns include firm's ownership status, age, manager's experience, gender of owner, business obstacles- crime, tax rates, power, government interpretation of law as firm-level controls.

Table 2.7: Probit model for whether a gift or informal payment is expected or requested

VARIABLES	(1) tax_pay	(2) tax_pay	(3) gov_pay	(4) gov_pay	(5) imp_pay	(6) imp_pay	(7) ope_pay	(8) ope_pay
ln_markup	0.00163	0.00841	-0.0101	-0.0163	-0.00296	-0.0104	-0.00604	-0.000439
	(0.00857)	(0.00836)	(0.0143)	(0.0146)	(0.0187)	(0.0178)	(0.0122)	(0.0118)
forshare	-0.110***	-0.0138	-0.188***	-0.177***	-0.0601	-0.0496	-0.152***	-0.0424
	(0.0358)	(0.0321)	(0.0591)	(0.0650)	(0.0519)	(0.0566)	(0.0531)	(0.0440)
govshare	-0.146	-0.0988	-0.204	-0.325**	-0.455*	-0.606**	-0.501***	-0.365**
	(0.120)	(0.114)	(0.151)	(0.163)	(0.254)	(0.276)	(0.166)	(0.178)
ln_exp	-0.123***	-0.00732	-0.0689***	-0.0191	-0.229***	-0.0341	-0.143***	-0.0399**
	(0.0166)	(0.0142)	(0.0250)	(0.0251)	(0.0374)	(0.0298)	(0.0240)	(0.0194)
female	-0.0587**	0.0340*	-0.00943	0.00204	0.0209	0.00426	-0.112***	0.00220
	(0.0228)	(0.0204)	(0.0351)	(0.0327)	(0.0406)	(0.0407)	(0.0314)	(0.0305)
tax_rates	0.181***	0.257***	0.169***	0.234***	0.109**	0.259***	0.167***	0.259***
	(0.0222)	(0.0189)	(0.0351)	(0.0355)	(0.0476)	(0.0390)	(0.0299)	(0.0277)
crime	0.0413	0.119***	0.184***	0.130***	0.0153	0.153***	0.0581*	0.171***
	(0.0270)	(0.0221)	(0.0418)	(0.0434)	(0.0436)	(0.0514)	(0.0304)	(0.0326)
power	0.250***	0.0728***	0.161***	0.0456	0.262***	0.0786	0.200***	0.0476
	(0.0269)	(0.0196)	(0.0363)	(0.0381)	(0.0650)	(0.0524)	(0.0405)	(0.0327)
gov_off	-0.276***	-0.231***	0.226***	-0.111	-0.440***	-0.116	-0.408***	-0.226***
	(0.0461)	(0.0333)	(0.0814)	(0.0895)	(0.0787)	(0.0843)	(0.0566)	(0.0579)
countryFE	no	yes	no	yes	no	yes	no	yes
sector FE	no	yes	no	yes	no	yes	no	yes
timeFE	no	yes	no	yes	no	yes	no	yes
Observations	54,366	53,810	11,591	11,269	11,455	10,389	22,798	21,775
r-square	0.0203	0.161	0.0157	0.159	0.031	0.211	0.0232	0.136
#_depvar	11	178	11	172	11	178	11	178
%_ones	0.146	0.146	0.2	0.2	0.132	0.132	0.161	0.161
conv	1	1	1	1	1	1	1	1

Notes: ***, **, * denotes statistical significance at 1, 5 and 10 percent respectively. Constants are included in the model but are not reported. Reported values are estimated from the probit model with clustered standard errors at the country-sector-year level in brackets. The dependent variable is a binary variable that takes on a value of one if the respondent answers yes to the question on whether a gift or informal payment was requested or expected, and zero if the respondent answers no. All columns include firm's ownership status, age, manager's experience, gender of owner, business obstacles- crime, tax rates, power, government interpretation of law as firm-level controls.

Table 2.8: Probit model for nonresponses to payment-related questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	tax_ref	tax_ref	gov_ref	gov_ref	imp_ref	imp_ref	ope_ref	ope_ref
ln_markup	0.0242**	0.0208*	-0.0474**	-0.0495*	0.0377*	0.0331	0.0427***	0.0265*
	(0.0108)	(0.0107)	(0.0241)	(0.0288)	(0.0220)	(0.0224)	(0.0149)	(0.0155)
prishare	0.0299	0.161***	0.0548	0.262*	-0.0494	-0.000248	-0.0417	0.0674
	(0.0384)	(0.0396)	(0.134)	(0.151)	(0.0635)	(0.0661)	(0.0600)	(0.0641)
govshare	0.294**	0.0695	-0.150	-0.00177	0.509**	0.442*	0.359*	0.0972
	(0.117)	(0.133)	(0.271)	(0.273)	(0.228)	(0.258)	(0.208)	(0.207)
ln_exp	-0.0612***	-0.0231	-0.0531	-0.0143	-0.0905**	-0.0406	-0.0760***	-0.0578**
	(0.0183)	(0.0184)	(0.0515)	(0.0567)	(0.0357)	(0.0405)	(0.0249)	(0.0256)
female	-0.0434	-0.0461	-0.0647	-0.140**	-0.00714	-0.0411	-0.0734**	-0.0943**
	(0.0265)	(0.0295)	(0.0598)	(0.0647)	(0.0499)	(0.0521)	(0.0335)	(0.0380)
ln_age	0.0168	0.0146	-0.0933*	-0.00165	-0.0153	0.0321	0.00648	-0.0153
	(0.0195)	(0.0189)	(0.0489)	(0.0561)	(0.0310)	(0.0356)	(0.0267)	(0.0270)
tax_rates	-0.0143	0.0298	0.172***	0.141**	-0.0405	0.00205	0.0326	0.0233
	(0.0274)	(0.0268)	(0.0625)	(0.0699)	(0.0505)	(0.0560)	(0.0372)	(0.0404)
crime	-0.0741**	-0.00230	-0.163*	-0.0716	-0.0794	-0.0295	-0.0646	-0.0526
	(0.0340)	(0.0333)	(0.0875)	(0.101)	(0.0565)	(0.0629)	(0.0435)	(0.0479)
power	-0.151***	-0.0270	-0.0737	-0.111	-0.0889*	0.0429	-0.157***	-0.0388
	(0.0297)	(0.0267)	(0.0673)	(0.0762)	(0.0463)	(0.0543)	(0.0424)	(0.0444)
gov_off	-0.670***	-0.219***	-0.487**	-0.126	-0.501***	-0.234**	-0.676***	-0.293***
	(0.0586)	(0.0642)	(0.191)	(0.247)	(0.104)	(0.118)	(0.0918)	(0.106)
countryFE	no	yes	no	yes	no	yes	no	yes
sector FE	no	yes	no	yes	no	yes	no	yes
timeFE	no	yes	no	yes	no	yes	no	yes
Observations	56,899	54,355	3,038	2,812	12,053	10,799	23,807	22,087
r-square	0.0181	0.118	0.00876	0.201	0.0174	0.145	0.0192	0.134
#_depvar	11	178	11	166	11	178	11	178
%_ones	0.0522	0.0522	0.854	0.854	0.0613	0.0613	0.0518	0.0518
conv	1	1	1	1	1	1	1	1

Notes: ***, **, * denotes statistical significance at 1, 5 and 10 percent respectively. Constants are included in the model but are not reported.Reported values are estimated from the probit model with clustered standard errors at the country-sector-year level in brackets. The dependent variable is a binary variable that takes on a value of one if the respondent refuses to answer the question on payments contingent on having contact (or exposure) to bribery. All columns include firm's ownership status, age, manager's experience, gender of owner, business obstacles- crime, tax rates, power, government interpretation of law as firm-level controls.

Table 2.9: Definition of variables from WBES

Incidence Variables	
tax_inc	Over the last 12 months, was this establishment visited and or inspected by tax officials?
	Dummy variable (=1) if the respondent answers yes, 0 if the respondent answers no
Gov_inc	Over the last 12 months, has this establishment secured or attempted to secure a contract with the government?
	Dummy variable (=1) if the respondent answers yes, 0 if the respondent answers no
Imp_inc	Over the last two years, did this establishment submit an application to obtain an import license?
	Dummy variable (=1) if the respondent answers yes, 0 if the respondent answers no
Op_inc	Over the last two years, did this establishment submit an application to obtain an operating license?
	Dummy variable (=1) if the respondent answers yes, 0 if the respondent answers no
Payment expected or requested	
Tax_pay	In any of these inspections or meetings was a gift or informal payment expected or requested?
	Dummy variable (=1) if the respondent answers yes and answers yes (=1) to tax_inc,
	0 if the respondent answers no and answers yes (=1) to tax_inc
Gov_pay	what percent of the contract value would be typically paid in informal payments or gifts to secure the contract?
	Dummy variable (=1) if the respondent answers yes and answers yes (=1) to gov_inc,
	0 if the respondent answers no and answers yes (=1) to gov_inc
Imp_pay	In reference to that application for an import license, was an informal gift or payment expected or requested?
	Dummy variable (=1) if the respondent answers yes and answers yes (=1) to imp_inc,
	0 if the respondent answers no and answers yes (=1) to imp_inc
Op_pay	In reference to that application for an operating license, was an informal gift or payment expected or requested?
	Dummy variable (=1) if the respondent answers yes and answers yes (=1) to imp_inc,
	0 if the respondent answers no and answers yes (=1) to imp_inc
Refused to answer (Nonresponses)	
Tax_ref	Dummy variable (=1) if there is a missing observation for tax_pay and tax_inc =1,
	0 if a response is provided to tax_pay and tax_inc =1
Gov_ref	Dummy variable (=1) if there is a missing observation for gov_pay and gov_inc =1,
	0 if a response is provided to gov_pay and gov_inc =1
Imp_ref	Dummy variable (=1) if there is a missing observation for imp_pay and imp_inc =1,
	0 if a response is provided to imp_pay and imp_inc =1
Op_ref	Dummy variable (=1) if there is a missing observation for op_pay and op_inc =1,
	0 if a response is provided to op_pay and op_inc =1
Firm Characteristics	
Ln_age	Log of survey year minus the established year
exporter	Dummy (=1) if firm is an exporter (% of the establishments direct and indirect sales ¿ 5), 0 otherwise
forshare	Dummy (=1) if firm foreign owned (more than 50 % ownership by foreign individuals, firms, or government)
govshare	Dummy (=2) if firm foreign owned (more than 50% ownership by foreign individuals, firms, or government)
Ln_exp	Experience of the top manager in the establishments sector
Female	Dummy variable (=1) if firm is owned by female, 0 otherwise.
certificate	Dummy variable (=1) if the firm has an internationally-recognized quality certification, 0 otherwise
Total sales	total annual sales in the previous year
Wage bill	Total annual cost of labor (including wages, salaries, bonuses, social payments) in the previous year
Business constraints	
crime	Dummy variable (=1) if the respondent answers major obstacle, very severe to
	the severity of crime, theft and disorder being an obstacle to the current operations of the establishment.
	Dummy variable (=0) if the respondent answers no obstacle, minor obstacle, and moderate obstacle.
power	Dummy variable (=1) if the respondent answers major obstacle, very severe to
	the severity of electricity being an obstacle to the current operations of the establishment.
	Dummy variable (=0) if the respondent answers no obstacle, minor obstacle, and moderate obstacle.
transport	Dummy variable (=1) if the respondent answers major obstacle, very severe to
	the severity of transportation of goods, supplies, and inputs being an obstacle.
	Dummy variable (=0) if the respondent answers no obstacle, minor obstacle, and moderate obstacle.
Tax_rates	Dummy variable (=1) if the respondent answers major obstacle, very severe to the severity of tax rates.
	Dummy variable (=0) if the respondent answers no obstacle, minor obstacle, and moderate obstacle.
Gov_off	Dummy variable (=1) if the respondent answers tend to agree, strongly agree to whether
	the government officials interpretations of the laws and regulations affecting this establishment.
	Dummy variable (=0) if the respondent answers strongly disagree, tend to disagree.
informal	Dummy variable (=1) if the respondent answers major obstacle, very severe to the
	severity of informal competition being an obstacle to the current operations of the establishment.
	Dummy variable (=0) if the respondent answers no obstacle, minor obstacle, and moderate obstacle.
finance	Dummy variable (=1) if the respondent answers major obstacle, very severe to the
	severity of access to finance being an obstacle to the current operations of the establishment.
	Dummy variable (=0) if the respondent answers no obstacle, minor obstacle, and moderate obstacle.

Table 2.10: List of Countries in the sample

Country	Obs	Percent	Cpercent	Country	Obs	Percent	Cper	Country	Obs	Per	Cper
Afghanistan	943	0.76	0.76	Gabon	179	0.14	31.05	Panama	969	0.78	68.72
Albania	664	0.53	1.29	Gambia	174	0.14	31.19	PapuaNewGuinea	63	0.05	68.77
Angola	785	0.63	1.92	Georgia	733	0.59	31.78	Paraguay	974	0.78	69.55
Antiguaandbarbuda	151	0.12	2.05	Ghana	1,214	0.98	32.76	Peru	1,632	1.31	70.86
Argentina	2,117	1.7	3.75	Grenada	153	0.12	32.88	Philippines	2,628	2.11	72.98
Armenia	734	0.59	4.34	Guatemala	1,112	0.89	33.77	Poland	997	0.8	73.78
Azerbaijan	770	0.62	4.96	Guinea	223	0.18	33.95	Romania	1,081	0.87	74.65
Bahamas	150	0.12	5.08	GuineaBissau	159	0.13	34.08	Russia	5,224	4.2	78.85
Bangladesh	2,946	2.37	7.45	Guyana	165	0.13	34.21	Rwanda	453	0.36	79.21
Barbados	150	0.12	7.57	Honduras	796	0.64	34.85	Samoa	102	0.08	79.29
Belarus	633	0.51	8.08	Hungary	601	0.48	35.34	Senegal	1,064	0.86	80.15
Belize	150	0.12	8.2	India	9,210	7.41	42.74	Serbia	747	0.6	80.75
Benin	150	0.12	8.32	Indonesia	2,743	2.21	44.95	Sierra Leone	150	0.12	80.87
Bhutan	476	0.38	8.7	Iraq	756	0.61	45.56	Slovak Republic	543	0.44	81.31
Bolivia	975	0.78	9.48	Israel	483	0.39	45.94	Slovenia	546	0.44	81.75
Bosnia and Herzegovina	718	0.58	10.06	Jamaica	376	0.3	46.25	Solomon Islands	148	0.12	81.87
Botswana	610	0.49	10.55	Jordan	565	0.45	46.7	SouthAfrica	937	0.75	82.62
Brazil	1,802	1.45	12	Kazakhstan	1,144	0.92	47.62	Southsudan	733	0.59	83.21
Bulgaria	1,594	1.28	13.28	Kenya	1,437	1.16	48.78	SriLanka	610	0.49	83.7
BurkinaFaso	394	0.32	13.6	Kosovo	468	0.38	49.15	StKittsandNevis	150	0.12	83.82
Burundi	424	0.34	13.94	Kyrgyz Republic	505	0.41	49.56	StLucia	150	0.12	83.94
Cambodia	793	0.64	14.58	LaoPDR	992	0.8	50.36	StVincentandGrenadines	154	0.12	84.07
Cameroon	363	0.29	14.87	Latvia	607	0.49	50.85	Sudan	647	0.52	84.59
CapeVerde	156	0.13	15	Lebanon	530	0.43	51.27	Suriname	152	0.12	84.71
Centralafricanrepublic	150	0.12	15.12	Lesotho	151	0.12	51.39	Swaziland	307	0.25	84.95
Chad	150	0.12	15.24	Liberia	150	0.12	51.51	Sweden	600	0.48	85.44
Chile	2,048	1.65	16.88	Lithuania	546	0.44	51.95	Tajikistan	719	0.58	86.02
China	2,700	2.17	19.06	Madagascar	960	0.77	52.72	Tanzania	1,232	0.99	87.01
Colombia	1,942	1.56	20.62	Malawi	661	0.53	53.26	Thailand	951	0.76	87.77
Congo	151	0.12	20.74	Malaysia	989	0.8	54.05	Timor Leste	150	0.12	87.89
Costarica	538	0.43	21.17	Mali	850	0.68	54.74	Timor-Leste	126	0.1	87.99
Croatia	992	0.8	21.97	Mauritania	358	0.29	55.02	Togo	155	0.12	88.12
Czech Republic	504	0.41	22.37	Mauritius	398	0.32	55.34	Tonga	150	0.12	88.24
Cte dIvoire	526	0.42	22.8	Mexico	2,960	2.38	57.72	TrinidadandTobago	370	0.3	88.54
DRC	1,228	0.99	23.79	Micronesia	67	0.05	57.78	Tunisia	592	0.48	89.01
Djibouti	266	0.21	24	Moldova	723	0.58	58.36	Turkey	2,496	2.01	91.02
Dominica	150	0.12	24.12	Mongolia	722	0.58	58.94	Uganda	1,325	1.07	92.08
DominicanRepublic	360	0.29	24.41	Montenegro	266	0.21	59.15	Ukraine	1,853	1.49	93.57
Ecuador	1,024	0.82	25.23	Morocco	406	0.33	59.48	Uruguay	1,228	0.99	94.56
Egypt	2,897	2.33	27.56	Mozambique	479	0.39	59.87	Uzbekistan	756	0.61	95.17
ElSalvador	693	0.56	28.12	Myanmar	632	0.51	60.37	Vanuatu	128	0.1	95.27
Elsalvador	360	0.29	28.41	Namibia	891	0.72	61.09	Venezuela	820	0.66	95.93
Eritrea	179	0.14	28.55	Nepal	850	0.68	61.77	Vietnam	2,045	1.64	97.58
Estonia	546	0.44	28.99	Nicaragua	814	0.65	62.43	West Bank And Gaza	434	0.35	97.93
Ethiopia	1,492	1.2	30.19	Niger	150	0.12	62.55	Yemen	776	0.62	98.55
Fiji	162	0.13	30.32	Nigeria	4,537	3.65	66.2	Zambia	1,204	0.97	99.52
Fyr Macedonia	726	0.58	30.91	Pakistan	2,164	1.74	67.94	Zimbabwe	599	0.48	100

Notes: List of country in the sample with number of observations (obs), percentages (percent), and cumulative percentages (cpercent).

Chapter 3

Income Inequality, Ethnic Diversity, and Corruption in the U.S. States

3.1 Introduction

According to the U.S. Department of Justice (henceforth, DOJ), between 1984-2005, a total of 18,029 public officials (at the local, state, and federal level) were prosecuted for corruption-related crimes. Relative to other Organization for Economic Cooperation and Development (OECD) countries such as Germany, U.K., and Canada, the U.S. as a whole is perceived to be more corrupt,[1] although corruption levels vary widely across states. For example, according to the DOJ between 1985-2005, there were 0.709 and 0.706 convictions of public officials for corruption-related crimes per 100,000 population in South Dakota and Louisiana respectively, while there were only 0.102 and 0.099 convictions of public officials for corruption-related crimes per 100,000 population in New Hampshire and Oregon respectively (see complete list in table 3.1). Why is it that states vary in the levels of corruption despite the U.S. having strong economic and political institutions?

Of the studies that have been written on the subject, research shows income inequality as a major source of corruption because as a society becomes more divided, the high-income individuals have the resources to engage in corruption to maintain their elite position (Glaeser and Saks, 2006). However, other recent studies show that income inequality is a direct con-

[1]For example, according to Transparency International's Corruption Perception Index (CPI) of 2016, the U.S. ranked 18th out of 168; whereas Canada, U.K., Germany ranked 9th, 10th, 10th out of 176 respectively.

sequence of corruption because low-income individuals pay a higher share of income in bribes compared to high-income individuals (Gupta et al., 2002). Therefore, the question I pose is whether income inequality is a cause for or a consequence of corruption. Understanding this relationship is important, not only to policy makers, but to Americans who face growing income disparity and are directly affected by the consequences of corruption.

In the past 3 decades, the U.S. has seen a dramatic rise in the top income earnings and several economists contend that this trend is likely to continue in the future (Piketty and Saez, 2013). The gap between the rich and the poor has been continually widening, with the top 1% earnerd 22% of income share in 2012 as compared to only 17 % in 2002. This figure is much higher compared to the average of 9.6 times more for OECD countries. Similarly, the Gini Index was 0.64 in 2012 compared to 0.58 in 2002. My research shows that income inequality, more specifically increases in the top income share, is one of the key ingredients that has exacerbated the variation in state level corruption.

My empirical approach adapts You and Khagram (2005) cross-country study that examines the bidirectional relationship between income inequality and corruption. They illustrate two potential channels through which income inequality increases corruption. First, in more democratic countries, wealthy individuals have the resources to engage in corruption and produce policy outcomes closely favorable to them at the cost of the poor. Second, the rise in income inequality affects individual perception and widens the norm for the acceptability of corruption. My hypothesis is that, for a democratic country like the U.S., income inequality across states can explain the rise in corruption, which implies that income inequality and corruption measures should be positively correlated.

I examine income inequality using the top 1% and top 0.1 % income share, and Gini coefficient. I choose to include the top 1% income shares since most of the changes observed in U.S. income inequality throughout the twentieth century are predominately observed at this end of distribution (Piketty and Saez, 2013). In addition, the top income earners are likely to have the resources and political connections to engage in corruption. With the traditional broad based indicators (namely the Gini and Atkinson Indexes), it is difficult to breakdown the components to determine which subgroup is contributing to overall inequality. Similarly, in a more recent paper, Aghion et al. (2015) also highlight the importance of using top income shares as a measure of income inequality rather than broad based indicators. They find increases in the top 1% income share is likely associated with increases in growth from innovation; however, the result is insignificant for broad income inequality measures.

Corruption is measured by a conviction-based measure from the DOJ and a perception-based measure from Boylan and Long (2003). The conviction-based measure has panel data on all 48 mainland states from 1984-2002, while the perception-based measure has information on only 45 states for the year 1998. Finally, I explicitly control for ethnic diversity to clearly understand the relation between income inequality and corruption levels. Studies have shown that when a society is more diverse, politicians tend to allocate resources that favor their own ethnic groups (or voters) . In return, the diverse group (or voters) will continue to support politicians of their own ethnicity, despite knowing their corrupt behavior (Glaeser and Saks, 2006).

The main results can be summarized as follows. The Ordinary Least Square (OLS) results suggest there is a positive and statistically significant correlation between top income earners and state level corruption. A 10% increase in income share for the top 1% and top 0.1% are associated with an increase of 1.47% and 0.75%, respectively, in the rates of corruption. Similarly, a 10% increase in Gini is associated with an increase in rates of corruption by about 7.5%. Several studies have shown that the causal link runs from corruption to income inequality; therefore, it is plausible that the baseline empirical model is endogenous. If this is true, the OLS results will be biased. To mitigate these concerns and to further ensure robustness of the results, I include three instruments: state slave share in 1860, mature age cohort size, and median household income in 1970. The instruments have no direct relation with corruption but are correlated with the income inequality variables. The instruments pass the validity and exogeneity tests in the 2SLS model. The results from the IV 2SLS are analogous to the OLS results but with a higher economic magnitude. This finding confirms that the OLS results are biased downwards due to endogeneity. More specifically, a 10% increase in the top 1% income share increases the rates of corruption by 3.25% and a 10% increase in the top 0.1% income share increases the rates of corruption by 2.10%. The results confirm the hypothesis that an increase in income inequality increases the rates of corruption as measured by the conviction-based index. However, the results are not consistent for the perception-based corruption index; therefore, the paper mostly focuses on the state-level conviction-based index that has been extensively used in the literature.[2]

This paper is the first, to my knowledge, to analyze the effect of income inequality as measured by the income share held by the top 1% and top 0.1% on state-level corruption in the U.S. States. Much of the research for the U.S. consists of examining the effect of corruption on the Gini and the Atkinson Indexes (Dincer and Gunalp (2012), Apergis et al. (2010)) or the effect of the Gini (in 1970) on state-level corruption (Glaeser and Saks (2006)). However, these papers do not discuss the importance of the higher income earners. As stated previously, it is likely the income held by the top 1% (ultra-rich) matters the most as they have the resources to engage in corruption. The Gini index, by contrast, is non-decomposable which makes it difficult to look at different income subgroups (Frank, 2009).

The paper proceeds as follows: section 2 discusses previous literature between income inequality and corruption, section 3 provides an overview of the data, and section 4 provides an empirical analysis with the OLS and IV results. Lastly, section 5 highlights the main conclusions.

3.2 Corruption and Income Inequality

Income inequality and corruption are closely related and recent studies find that corruption has a positive effect on income inequality in the United States (Dincer and Gunalp (2012), Apergis et al. (2010)) and in Africa (Gyimah-Brempong (2002)), but a negative effect in Latin America (Andres and Ramlogan-Dobson (2011)). Dincer and Gunalp (2012) use Current Population Survey (CPS) data and test whether state-level corruption has an effect on income inequality, measured by the Gini coefficient and Atkinson indexes, for 1981-1997. They find evidence that an increase in corruption leads to an increase in income inequality in the U.S. States. The results are consistent when the authors control for endogeneity in a Generalized Methods of Moments (GMM) model. Similarly, Apergis et al. (2010) investigate the Granger causality from corruption to income inequality (Gini) using a panel vector

[2]For instance, see Glaeser and Saks (2006), Goel and Nelson (1998), Leeson and Sobel (2008), and Johnson et al. (2010).

error correction mode. For the U.S. states between 1980-2004, the authors find a positive bidirectional relationship between the income inequality and corruption in the short-run and the long-run.

Alternatively, Glaeser and Saks (2006) test the effect of median household income and Gini coefficient, both for 1970, on average corruption from 1976-2002 for the U.S. States. They use income from 1940 and a quadratic function of latitude and longitude as instruments for the median household income[3] while subjecting their empirical testing to include a wide set of controls such as racial and ethnic fractionalization, urban population, college degree, and share of government employment. Glaeser and Saks conclude that states with higher income and education are less likely to be corrupt; however, they also find that states with higher income inequality and racial heterogeneity are more corrupt.

Therefore, it is not only possible that state-level corruption increases income inequality, but as income inequality increases, the state-level corruption can increase as well. According to Uslaner (2008), this vicious cycle is known as the "inequality traps" and leads to more income inequality and corruption in a nation. In a similar vein, You and Khagram (2005) find evidence that an increase in income inequality, especially in democracies, increases corruption. They use data on Corruption Perception Indexes and Gini Coefficient from 129 countries. According to Meltzer and Richard (1981), for regions that are more economically divided, poorer citizens will pressure government for redistributive policies more favorable towards them (aimed towards the lower end of the income distribution). For example, this could be done by pressuring the state for a tax increase on the rich which will decline their income share. As a result, the wealthy have a greater incentive to engage in political and bureaucratic corruption to avoid paying higher taxes and to evade tax payments altogether as stated by You and Khagram (2005). Furthermore, poor individuals are also likely to engage in "petty corruption" to ensure these public services are providedto them (You and Khagram, 2005).The paper concludes by stating that the effect of income inequality on corruption is magnified in more democratic societies compared to dictatorships. The main reason is that the rich have the means to engage and rely on corruption to direct favorable policy outcomes since "repression of the mass" is not easy in a democracy.

[3]Note: For the Gini Coefficient, the authors only report the OLS result.

Several studies at the national level have examined the effects of ethnic diversity on levels of corruption. Why might ethnic diversity play a role in determining the rates of corruption? Mauro (1995) claims that regions with high ethnic-linguistic fractionalization reduce the likelihood of holding a corrupt official accountable. Likewise, other studies show corruption more pervasive in countries that are ethnically fragmented (Treisman (2000b), Alesina et al. (2003), LaPorta et al. (1999)). For the U.S., Glaeser and Saks (2006) argue that politicians tend to allocate resources toward their own ethnicity and these ethnic groups continue to support the same politicians despite knowing about their corrupt behavior. Dincer (2008) finds a positive and linear relationship between ethnic and religious diversity and corruption within the U.S. States.[4] Hence, by including a control for ethnic diversity, I aim to clearly elucidate the income inequality and corruption link. All else equal, the hypothesis is that states with greater ethnic heterogeneity are more likely to be corrupt.

Based on the aforementioned research, this paper will address the effects of income inequality on corruption in the U.S. States. The focus of this paper is on the U.S. because, despite being one of the strongest democracies in the world, the disparity in income has been continually widening in the past 3 decades. In addition, the U.S. has seen a growing disparity in politics where the top 1% have substantial power to influence public policies favorable to them. This paper advances the literature by focusing on the importance of the top 1% and top 0.1% income share in determining the levels of corruption.

3.3 Data

3.3.1 Corruption

First, the number of federal convictions for corruption-related crimes are obtained from *Report to Congress on the Activities and Operations of Public Integrity Section* (U.S. Department of Justice). The focus of the PIN is to investigate public corruption, ranging from election crimes and conflicts of interest crimes to campaign finance violations (Section,

[4] Glaeser and Saks (2006) use data from the 1980 Census for ethnic diversity. They run a cross-section OLS with 50 observations. Dincer (2008) use 10 year averages from 1980-1989 and 1990-1999 for ethnic diversity. They run a pooled OLS with 96 observations.

2011). According to the DOJ, a local case is handled at the federal level for four different scenarios: to ensure fairness if the government official being prosecuted has strong ties to the government, cases falling under multiple jurisdictions, cases referred directly from federal agencies, and cases requiring extra resources or shared responsibilities. For the main measure of corruption, I use the annual rate of corruption (CONVICTION) which is defined as the number of federal-convictions of public officials per 100,000 state population. Based on averages from 1984-2005, South Dakota, Louisiana, and Alaska were the top three corrupt states, while Oregon, New Hampshire, and Washington were three least corrupt states. This type of data was originally published in Glaeser and Saks (2006) where the rate of corruption for each state is defined as the number of federal convictions per state divided by the average state population. Similar measures have been used in the literature by Goel and Nelson (1998), Leeson and Sobel (2008), and Johnson et al. (2010) to measure corruption across states

Second, I use a time-series average of state level corruption from 1976-2002 (STATE_AVG). This data is obtained from Glaeser and Saks (2006). The authors cite the use of time-series average to mitigate the year-to-year fluctuation in the corruption index. Therefore, by using this measure, I ensure comparability of my results with other papers that have utilized this time-frame. For the time-series average from 1976-2002, the top three corrupt states are Alaska, Mississippi, and Louisiana and the three least corrupt states are Colorado, Wisconsin, and Nebraska.

Third, I include a perception-based (PERCEPTION) measure introduced by Boylan and Long (2003). The authors survey 300 state reporters that deal with state politics to determine the perception of public corruption. The reporters were asked to rate their state in terms of level of corruption of all government employees (elected officials, political appointees, and civil servants) on a scale from one (least corrupt) to seven (more corrupt). The reporter's average response for each state is included as the perception-based index. Their study identifies the top three corrupt states as Rhode Island, Louisiana, and New Mexico and three least states as Colorado, North Dakota, and South Dakota.

Although the measures differ on how to estimate corruption, the conviction-based measure is likely superior and I refer to it closely in this paper. The fundamental difference is the DOJ measure uses actual federal conviction rather than relying on the perception of individuals.[5] The use of the conviction-based measure provides an estimate of the actual level of state-corruption and provides evidence of "culture of corruption" in a state.

3.3.2 Income Inequality

The data for share of income of the top 1% to top 0.01% for the U.S. States comes from The World Wealth and Income Database. The data was originally published at the US State-Level Income Inequality Database by Frank (2009). The Database contains information on total income earned (both pre-tax and pre-transfer) by individuals at the top tail of the income distribution. The data is based on the Internal Revenue Service (IRS) individual tax returns and includes wages, entrepreneurial income, and capital gains.[6] A major advantage of the data is its availability from 1913 onwards for all the U.S. states. The IRS, however, does not report data below a certain threshold of income which makes it difficult, although not impossible, to estimate the lower income shares.

The U.S. State-Level Income Inequality Database also provides data on the Gini coefficient (Gini), Atkinson Index (0.5), and Theil Index. According to Frank (2009), Gini is defined as "representing the average distance between all pairs of proportional income in the population" (p.256). Theoretically, Gini is bounded between zero (perfect equality) and one (perfect inequality). A problem associated with using Gini is that it is non-decomposable which makes it difficult to look at inequality within different income subgroups. It is, therefore, possible for subgroups within a population to exhibit higher rates of inequality, while the aggregate Gini would show an overall decrease (Frank, 2009). Figure 3.2 plots the average values of the Gini coefficient from 1984-2004 and figure 3.1 plots the top income share for different categories from 1920 onwards.

[5]For detailed discussion, see Glaeser and Saks (2006) for a discussion on the advantages of using the conviction-based measure. The authors point out that the DOJ measure is more objective and independent of the state's judicial law. The authors also argue that this mitigates the problem associated with a corrupt state not prosecuting a corrupt official. The federal judicial system is independent of the local corruption and "... should treat people similarly across space"(p. 1054).

[6]Note that capital gains include dividends, interests, rents, and royalties. Notable exclusions are the state and federal transfers, interest on state and local bonds. For more information, see Frank (2009).

3.3.3 Ethnic diversity

The data for ethnic diversity (ETHNIC) comes from the Census Bureau for the year 2000. Following the literature, ETHNIC denotes the probability of two randomly selected individuals belonging to two different ethnic groups. More specifically, ETHNIC is calculated as:

$$ethnic = 1 - \sum_{e=1}^{n} (s_{ie})^2 \tag{3.1}$$

where s_{ie} is the population share of ethnic group e in each state i and n is the total ethnic groups. In my sample, e = White, Asian, Hispanic, and Black for the year 2000. ETHNIC ranges between zero and one, where zero a state with complete homogeneity and one signifies a state with complete ethnic heterogeneity. The former signifies that everyone belongs to the same ethnic group.

3.3.4 Other Controls

I also control for different state-level characteristics that can potentially effect the corruption rates. Following Glaeser and Saks (2006), I control for the percentage of population with a college education (EDU, year 1984-2002), population (POP, year 1984-2002), and share of government employment relative to state population (GOVSH, year 1984-2002). The data for EDU comes from the Census, POP and GOVSH from the Bureau of Economic Analysis. Further, I control for four geographic regions to be consistent with the Census Bureau categories: Northeast, South, Midwest, and West.

Table 3.2 provides summary statistics for the variables used in this paper. The average rate of conviction is 0.31 per 100,000 state population with a standard deviation of 0.27. The average perception score for the perception-based measure is 3.48 with a standard deviation of 1.15. A detailed summary on the states that are included in each corruption measure are outlined in table 3.5. On average, the top 10% income share is about 40 % while the top 1% income share is about 14 %. About 23% of the population has a college degree or higher. The mean of ethnic diversity is 0.316 and the share of government employment is about 15%.

3.4 Econometric Analysis

3.4.1 Baseline Specification

In order to estimate the effect of income inequality on state-level corruption, I estimate the following regression equation:

$$corruption_{it} = \beta_1(income\_inequality_{it}) + \beta_2(ethnic_{i2000}) + \beta_3(X_{it}) + \eta_i + \tau_t + \epsilon_{it} \tag{3.2}$$

where subscript i denotes the 48 mainland states from 1984-2002. Income_inequality includes the Gini coefficient, the top 10%, top 1%, top 0.1%, top 0.01% income share for the time period 1984-2002. Corruption denotes the number of federal convictions per 100,000 state population for each year (CONVICTION), the time-series average corruption (STATE_AVG) from Glaeser and Saks (2006), and a perception-based index (PERCEPTION) from Boylan and Long (2003). Both income inequality and corruption variables are measured in logs.[7] Ethnic diversity is denoted as ETHNIC and is calculated for the year 2000. X_{it} is a set of control variables that include percentage of population with a college education (EDU, 1984-2002) and share of government employment relative to population (GOVSH, 1984-2002). η_j and τ_t represent region and time-fixed effects, respectively.

3.4.2 OLS Results

The first result reported (in table 3.4) is for the OLS regressions where the dependent variable is the number of federal convictions of public officials per 100,000 state population (CONVICTION). The independent variables include the top 10%, top 1%, top 0.5%, top 0.1%, top 0.01% income share, and the Gini Coefficient. All columns include log transformations of ethnic diversity (ETHNIC), share of government employment (GOVSH), and college education (EDU). Further, all columns include region and time fixed-effects. The final balanced panel includes 864 observations for the 48 mainland states from 1984-2002.

[7] The log transformation takes the form of ln(x+1) to keep observations that might include 0.

Table 3.4 shows that the effect of income inequality on rates of corruption is always positive and significant. For the top income shares (Column 1-5), the coefficients are positive and statistically significant at the 5% level and 1% level, respectively. A 10% increase in the income share of the top 1% and top 0.1 %are associated with an increase in the rates of corruption by 1.47% and 0.75%, respectively. The point estimates show that a one standard deviation (0.1) increase in the top 0.1% income share is associated with a 0.021 point increase in CONVICTION. This amounts to 11.3 % of a standard deviation. For the top 1% income share, a one standard deviation (0.211) increase is associated with an increase of 0.0310 in CONVICTION. This amounts to approximately 17% of a standard deviation in CONVICTION. Column 6 shows a 10% increase in Gini is associated with a 7.5% increase in CONVICTION, which is statistically significant at the 1% level. In other words, a one standard deviation (0.06) increase is associated with an increase in CONVICTION of about 24% of a standard deviation.

It is often argued that corruption increases with the share of government employees (GOVSH), level of ethnic diversity (ETHNIC), and decreases with the education levels (EDU). After including these variables as controls, I find that ETHNIC and GOVSH are positively and significantly associated with CONVICTION, while higher EDU is negatively and significantly associated with CONVICTION. The results for ETHNIC suggest that a 10% increase in ETHNIC (or greater ethnic diversity) increases the rates of corruption by 0.6%. The magnitude on ETHNIC is similar and statistically significant at the 1% level in all columns. These results are consistent with the previous research that find a positive effect of ethnic diversity on corruption levels in the U.S. states (see Dincer (2008) and Glaeser and Saks (2006))

In table 3.5, I include the state time-series average (STATE_AVG) for corruption from 1976-2002 as the main dependent variable. The data was originally published in Glaeser and Saks (2006). The results in all columns are positive and statistical significant at the 1% level implying a strong correlation between the top income inequality and corruption across the states. Next, in table 3.6, I look at two things. First, the coefficients in columns 1-3 are estimated using a panel OLS with region and time fixed-effects, while the coefficients in columns 4-6 are estimated with a cross-section OLS, with averages from 1976-2002, including region fixed-effects. The results are similar to those obtained in table 3 but with much lower magnitudes on the independent variables. A 10% increase in the top 10% income share increases STATE_AVG corruption by 1.6% (compared to 2.15% in table 2, column 1). Similarly, a 10% increase in the top 1% income share and the Gini are associated with an increase of 1.1% and 5.2% (compared to 1.4% and 7.5% in table 3, column 2 and 3) in STATE_AVG, respectively.

Finally, I include the perception-based corruption measure (PERCEPTION) on the right hand side of equation 2 (3.7). The coefficients in columns 1-3 are estimated using a panel OLS with region and time fixed-effects, while the coefficients in columns 4-6 are estimated with a cross-section OLS and includes region fixed-effects (in table 6). The sample size is much smaller in table 3.7 compared to table 3.6 3.5 because we only have information on 45 states.[8] The results provide mixed evidence. The first column shows that a 10% increase for the top 10% income share is associated with a 2.5% increases in the perception-based corruption measure, which is statistically significant at the 1% level. Column 2 shows that, by contrast, the point estimate of the top 1% income share is actually negative and does not enter the equation significantly. Similarly, column 3 shows that the effect of Gini on PERCEPTION is negative and significant at the 1% level.

The results presented above support the hypothesis that the top level income inequality is likely an important factor for state-level corruption. The results, however, cannot be interpreted as causal since corruption and income inequality are likely interdependent. I introduce three different instruments to test the exogenous variation in income inequality, in the next section.

3.4.3 Instrumental Variables and 2SLS Results

A problem arises with the OLS estimation because corruption is likely endogenous in the model. As a consequence, there is the possibility that corruption may be endogenous to income inequality. Gupta et al. (2002) and Dincer and Gunalp (2012) provide an overview of why corruption can lead to income inequality, which can be summarized in three main

[8]Refer to Appendix 1 for a list of states in the sample.

reasons: first, corruption generates a tax system that suits the rich and can lead to tax evasion, lower tax revenues, and reduced welfare programs that benefit the poor. Secondly, when there is a high concentration of asset ownership within a small interest group, these groups can use their wealth to influence public policy. For example, pushing for tax policies favorable to the rich or increasing spending on programs that will increase returns on their assets. Thirdly, corruption can reduce resources available for social programs which are put in place for the low-income individuals. Corruption, therefore, diverts resources from infrastructure building programs to programs where bribes can be easily extracted.

To address endogeneity, and to follow the literature, I have identified three potential instruments that are correlated with different income inequality variables (the endogenous variable) and uncorrelated with the rates of corruption. First, I include state slave share from 1860. Bertocchi and Dimico (2014) show that U.S. counties with a higher share of slaves relative to total population in 1860 are more unequal today. In particular, a one percent increase in slave share is associated with a statistically significant 0.045 increase in overall income inequality. Bertocchi and Dimico (2014) discuss two main reasons for the persistence of long-term income inequality seen today. The first is the differences in human capital attainment (or educational inequality) of blacks relative to whites, which accumulated overtime resulting in overall income inequality. The second reason is the racial discrimination mechanism, which prevented blacks from acquiring skills and depressed wages resulting in income inequality.

The use of slave share, however, has its concerns. There is information on only 40 of the 48 mainland states which reduces the number of observation from 864 to 720. For the remaining 40 states, there is a large variation in slave shares between 0 and 55 %. Nonetheless, I exploit the difference in slave shares across states to understand the impact of income inequality on rates of corruption today. Slave share is likely to have no relationship between rates of corruption, the dependent variable, which makes it a relevant instrument.

Second, I include mature cohort size defined as the proportion of adult population (age 16-59) who are between the age of 40-59. Higgins and Williamson (1999) show that "large working-age cohorts are associated with a lower income inequality, whereas large young-age cohorts are associated with a higher income inequality" (p. 4). While COHORT is likely to be correlated with the income inequality measures, it is unlikely to be correlated with the conviction-based corruption measure. Both Leigh (2006) and You and Khagram (2005) have instrumented income inequality with the mature cohort size.

Third, I consider median household income in 1970 as an instrument for income inequality today. Starting out with a high household median income in 1970 can be a strong predictor of the top income share.

Table 3.8 presents the IV results using two staged-least squares (2SLS) based on the econometric specification analogous to those reported in table 3-5. I perform several diagnostic tests to assess the reliability of the IV 2SLS model. First, I employ the Hansen-J statistics which tests the joint validity of instruments when there are more instruments than endogenous variables. The results in all four columns show that the null hypothesis of the validity of instruments cannot be rejected. Second, I report the F-test (the Kleibergen-Paap rk Wald F-statistic) from the first stage regressions. The first-stage F-test range between 46-53 which show that the instrumental variables are relevant and reject the joint insignificance at a 1% level. Third, I report the Shea partial-R^2 that considers the intercorrelations between instruments. The Shea partial-R^2 ranges from 0.068-0.219 implying that the instruments are valid.

After confirming the relevance and validity of the instruments, I now look at the coefficients on the income inequality variables. First, I report the 2SLS for cases where the dependent variable is the conviction-based corruption (CONVICTION). The columns report the result for the top 1% to top 0.01% income share for only the 40 states that are listed in the slave share data from Bertocchi and Dimico (2014). The three instruments include: slave share relative to state population (SLAVE), mature cohort size (COHORT), and median income from 1970 (INC1970).

The effect of income inequality on corruption is positive and statistically significant in all columns. The economic magnitudes of the results are quiet large as well. A 10% increase in the top 1% income share increases the conviction-based measure by 3.25% and a 10% increase in the top 0.1% income share increases the conviction-based corruption measure by 2.10%. To put this in perspective, a one standard deviation increase in top 1% income share (0.1) is associated with a 0.0325 point increase or about 17 % of an increase in the standard deviation of CONVICTION. Similarly, a one standard deviation in top 0.1% income share (0.211) is associated with a 0.084 point increase or almost a 45.40% of a standard deviation increase in CONVICTION. The results continue to hold for the top 0.01 and top 0.5% as well. Moreover, the estimated effects in all cases are much larger for the 2SLS compared to the OLS model. Also note that the estimated coefficients on ETHNIC, GOVSH and EDU retains their signs and are statistically significant at the 1% level.

Next, I look at the 2SLS for the average time-series rates of corruption (STATE_AVG) in table 3.9. Once again, the instruments pass several diagnostic tests for validity and exogeneity. The F-test in all columns are strong (33.09, 25.80, 54.79, respectively) and easily reject the null hypothesis of the joint insignificance of the instruments. The Hansen J-stat for columns (2) and (3) are strong and the validity of the null hypothesis cannot be rejected (0.37 and 4.2, respectively). However, the Hansen J-Stat for column (1) is about 5.8 with a χ_2 p-value of 0.054 which makes it possible to reject the validity of instruments at the 10% level. Keeping these results in mind, I now interpret the 2SLS results with the instrumented coefficient on income inequality. In the first column, the results show that a 10% increase in the top 10% income share increases the state time-series average corruption (STATE_AVG) by 3%. In the second column, a 10% increase in the top 1% increases STATE_AVG by 1.8%. Likewise, in the third column, a 10% increase in Gini increases STATE_AVG by 3.1%. In all columns, the results are statistically significant at the 1 % level. The estimated coefficients on income inequality are much higher compared to the OLS model reported in table 5.

Lastly, in table 3.10 I look at the 2SLS for CONVICTION after controlling for different sectors share: financial, mining, and oil and gas extraction in total state GDP. The motivation of this exercise is to test whether the relationship between top income inequality and corruption across states is driven by higher state shares in these sectors. Therefore, I control for the share of finance, mining, and oil and gas extration in total state GDP to see whether the results continue to hold. After controlling for the sector shares, the results continue to remain significant and the coefficients are similar to table 3.8.

3.5 Conclusion

This paper shows that income inequality measured by the top income inequality and the Gini coefficient have a substantial positive and significant effect on corruption across the U.S. States. Moreover, the effect on corruption is magnified when the three instruments: state slave share in 1860, cohort size, and income from 1970 are used in an IV 2SLS model. In particular, the results support the hypothesis by You and Khagram (2005) which states income inequality is an important ingredient for the causes of corruption in a democratic country. Therefore, reducing income inequality would significantly reduce corruption at the state-level. Although previous research have tested the relation between the Gini Index and corruption for the U.S. States; to my knowledge, this is the first evidence that highlights the importance of the top income inequality (more specificially, the top 1%, top 0.1%) as a cause of corruption. Given that corruption and unequal distribution of income are detrimental to a nation's long-term economic growth, these results recommend a redistributive policy reform in the U.S. to tackle corruption.

Figure 3.1: Income Share in the United States 1912-2013. Source: The World Wealth and Income Database

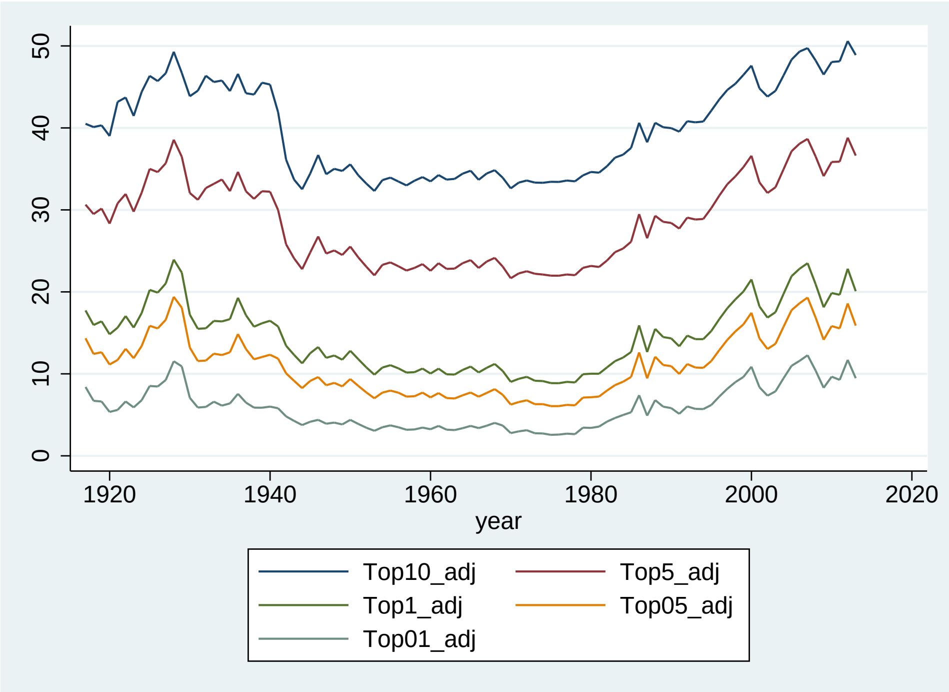

Figure 3.2: Gini in the United States 1912-2013: Source: Frank (2009)

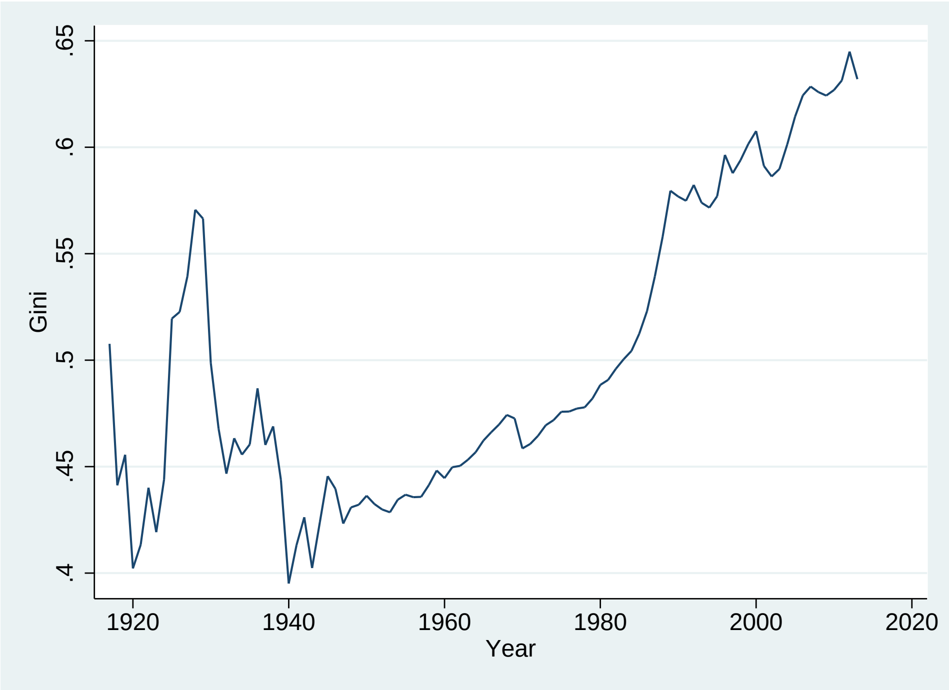

Table 3.1: Conviction Rates (average yearly convictions divided average pop from 1985-2005)

State	Corruption Rate	State	Corruption Rate
South Dakota	.709	Rhode Island	.295
Louisiana	.7063	Maine	.2859
Alaska	.701	South Carolina	.2683
Mississippi	.6844	Texas	.2584
North Dakota	.6298	Connecticut	.2511
Montana	.5438	California	.2509
Kentucky	.5037	Arizona	.2454
Illinois	.5024	Arkansas	.2433
Virginia	.4695	New Mexico	.2306
Alabama	.4501	Indiana	.228
Tennessee	.449	Idaho	.2269
New York	.4336	Michigan	.2244
Ohio	.4202	North Carolina	.2048
Florida	.4047	Vermont	.1996
Pennsylvania	.3854	Nevada	.1887
West Virginia	.3745	Kansas	.1831
New Jersey	.3637	Wisconsin	.1778
Delaware	.3533	Iowa	.1775
Georgia	.3441	Colorado	.1664
Hawaii	.34417	Nebraska	.1611
Wyoming	.3431	Minnesota	.1461
Maryland	.3408	Utah	.1334
Oklahoma	.3269	Washington	.1259
Missouri	.3211	New Hampshire	.102
Massachusetts	.3069	Oregon	.0994

Table 3.2: Summary statistics

Variable	Mean	Std. Dev.	N
Corruption			
log(conviction)	0.252	0.185	912
conviction	0.31	0.274	912
log(perception)	1.465	0.274	855
perception	3.487	1.152	855
log(avg_state)	0.233	0.096	912
avg_state	0.268	0.124	912
Income Inequality			
Top10_adj	39.604	4.195	912
Top1_adj	14.214	3.599	912
Top0.5_adj	9.648	3.981	912
Top0.1_adj	4.909	2.867	912
Top0.01_adj	1.862	1.544	912
gini(*100)	55.637	3.374	931
Controls			
edu	23.706	4.310	912
govsh	15.348	2.709	864
log(pop)	15.039	0.986	912
ethnic	0.316	0.142	912
Instruments			
slave share	0.114	0.176	760
ln(cohort)	0.392	0.018	912
ln(income1970)	8.233	0.155	912

Notes: Conviction is the number of convictions per 100,000 state population. Avg_state is the state time-series average corruption per 100,000 population from 1976-2002 obtained from Glaeser and Saks (2006). Perception is the perception-based corruption indicator based on reporters survey for the 1999 obtained from Boylan and Long (2003). Top 10%, top 1%, top 0.5%, top 0.1%, top 0.01% are the income share for different groups. Gini is the measure of income inequality where 100 represents perfect equality.Control variables include ethnic diversity, share of government employment relative to total population, percentage of population with a college degree, state population. Instruments include slave share in 1860 obtained fromBertocchi and Dimico (2014), mature cohort size as a ratio $\frac{40-59}{15-69}$ age group, log income in 1970.

Table 3.4: OLS Regression of Corruption and Income Inequality from 1984-2002

	(1)	(2)	(3)	(4)	(5)	(6)
DEP VARIABLE			CONVICTION			
Top 10%	0.216**					
	(0.098)					
Top 1%		0.147***				
		(0.047)				
Top 0.5%			0.107**			
			(0.042)			
Top 0.1%				0.075**		
				(0.034)		
Top 0.01%					0.059**	
					(0.030)	
Gini						0.750***
						(0.182)
ln_edu	-0.227***	-0.228***	-0.234***	-0.235***	-0.234***	-0.232***
	(0.039)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
ln_ethnic	0.064***	0.060***	0.063***	0.066***	0.069***	0.058***
	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)
ln_govemp_sh	0.300***	0.317***	0.313***	0.303***	0.295***	0.241***
	(0.051)	(0.053)	(0.053)	(0.053)	(0.052)	(0.047)
Year_FE	yes	yes	yes	yes	yes	yes
Region_FE	yes	yes	yes	yes	yes	yes
Observations	864	864	864	864	864	864
R-squared	0.187	0.191	0.190	0.188	0.187	0.206

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable (in logs) is the average corruption per 100,000 population. The independent variables (in logs) are top 10% , top 1% , top 0.5%, top 0.1%, top 0.01% income share, and Gini respectively. Control variables include ethnic diversity, share of government employment, and percentage of population with a college degree.

Table 3.5: OLS Regression of STATE_AVG and Income Inequality

VARIABLES	(1) STATE_AVG	(2) STATE_AVG	(3) STATE_AVG	(4) STATE_AVG
Top 1%	0.113***			
	(0.015)			
Top 0.5%		0.063***		
		(0.015)		
Top 0.1%			0.046***	
			(0.012)	
Top 0.01%				0.036***
				(0.010)
ln_edu	-0.227***	-0.231***	-0.231***	-0.231***
	(0.013)	(0.012)	(0.012)	(0.012)
ln_ethnic	0.066***	0.068***	0.070***	0.072***
	(0.004)	(0.004)	(0.004)	(0.004)
ln_govemp_sh	0.275***	0.270***	0.265***	0.260***
	(0.018)	(0.018)	(0.018)	(0.018)
Year_FE	yes	yes	yes	yes
Region_FE	yes	yes	yes	yes
Observations	864	864	864	864
R-squared	0.516	0.513	0.512	0.510

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: $***$,$**$, $*$ denote statistical significance at the 1, 5, and 10 percent levels, respectively. Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed-effects. The dependent variable (in logs) in all columns is the state time -series average corruption per 100,000 population from 1976-2002. The independent variables (in logs)are top 1% , top 0.5%, top 0.1%, and top 0.01% income share from 1984-2002 respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree.

Table 3.6: OLS Regression of STATE_AVG and Income Inequality

	(1)	(2)	(3)	(4)	(5)	(6)
DEP VARIABLE	Panel 1976-2002			Cross-section 1976-2002		
VARIABLES	state_avg	state_avg	state_avg	state_avg	state_avg	state_avg
Top 10%	0.168***			0.418		
	(0.030)			(0.264)		
Top 1%		0.113***			0.214**	
		(0.015)			(0.096)	
Gini			0.527***			0.864**
			(0.049)			(0.385)
ln_edu	-0.227***	-0.226***	-0.234***	-0.215***	-0.220***	-0.235***
	(0.011)	(0.011)	(0.010)	(0.065)	(0.061)	(0.054)
ln_ethnic	0.068***	0.070***	0.072***	0.068***	0.058***	0.055***
	(0.018)	(0.018)	(0.018)	(0.112)	(0.095)	(0.103)
ln_govemp_sh	0.236***	0.259***	0.206***	0.300***	0.330***	0.238***
	(0.015)	(0.016)	(0.013)	(0.094)	(0.091)	(0.076)
Year_FE	yes	yes	yes	no	no	no
Region_FE	yes	yes	yes	yes	yes	yes
Observations	1,296	1,296	1,296	48	48	48
R-squared	0.489	0.497	0.527	0.512	0.522	0.569

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: *** ,** , * denote statistical significance at the 1, 5, and 10 percent levels, respectively. Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable (in logs) in all columns is the state time-series average corruption per 100,000 population from 1976-2002. The independent variables are top 10% income share, top 1% income share, and Gini from 1984-2002 respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree. Column 1-3 is a panel OLS and column 4-6 is a cross-section OLS for the time period 1976-2002.

Table 3.7: OLS Regression of PERCEPTION on Income Inequality from 1976-2002

	(1)	(2)	(3)	(4)	(5)	(6)
DEP VARIABLE		Panel 1976-2002			Cross-section 1976-2002	
VARIABLES	avg_percep	avg_percep	avg_percep	avg_percep	avg_percep	avg_percep
Top 10%	0.254**			0.864		
	(0.101)			(0.843)		
Top 1%		-0.002			0.074	
		(0.047)			(0.334)	
Gini			-1.523***			-2.527**
			(0.139)			(0.981)
ln_edu	-0.617***	-0.632***	-0.644***	-0.571**	-0.616**	-0.623***
	(0.041)	(0.041)	(0.039)	(0.245)	(0.242)	(0.218)
ln_ethnic	1.008***	1.054***	1.230***	0.921**	1.030**	1.362***
	(0.076)	(0.077)	(0.071)	(0.430)	(0.455)	(0.407)
ln_govemp_sh	-0.127**	-0.166***	-0.150***	-0.087	-0.166	-0.186
	(0.056)	(0.057)	(0.052)	(0.335)	(0.347)	(0.318)
Year_FE	yes	yes	yes	no	no	no
Region_FE	yes	yes	yes	yes	yes	yes
Observations	1,215	1,215	1,215	45	45	45
R-squared	0.437	0.434	0.484	0.447	0.436	0.514

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable in all columns is the perception-based corruption measure from Boylan and Long (2003) for the year 1999. The independent variables are top 10% income share, top 1% income share, and Gini respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree.

Table 3.8: IV 2SLS of Corruption on Top Income Inequality 1984-2002

	(1)	(2)	(3)	(4)
DEP VARIABLE		CONVICTION		
Top 1%	0.325***			
	(0.121)			
Top 0.5%		0.272***		
		(0.103)		
Top 0.1%			0.210**	
			(0.0818)	
Top 0.01%				0.173**
				(0.0698)
ln_govemp_sh	0.291***	0.324***	0.310***	0.293***
	(0.0840)	(0.0838)	(0.0807)	(0.0765)
ln_ethnic	0.0518**	0.0566**	0.0638***	0.0703***
	(0.0237)	(0.0224)	(0.0205)	(0.0189)
ln_edu	-0.227**	-0.224***	-0.221***	-0.217***
	(0.0402)	(0.0409)	(0.0409)	(0.0409)
Observations	720	720	720	720
R-squared	0.247	0.248	0.244	0.241
Hansen J statistic	0.592	2.252	2.757	3.153
χ^2 p-value	0.7437	0.320	0.2519	0.2067
Shea Partial -R square	0.0873	0.1999	0.211	0.219
F-test (first-stage) of excluded instruments	46.038	45.381	49.80	53.83

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: *** ,** , * denote statistical significance at the 1, 5, and 10 percent levels, respectively. Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable in column 1-3 is the conviction-based corruption measure 100,000 population. The independent variables are top 1%, top 0.5%, top 0.1%, top 0.01% income share respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree. Instruments include slave share in 1860, mature cohort size as a ratio $\frac{40-59}{15-69}$ age group, and log income in 1970. F-test (Kleibergen-Paap rk Wald in STATA) tests the joint significance of the instruments from the first- stage regression. Shea- partial R-square denotes the intercorrelation among instruments. Hansen J-stat is an overidentification test.

Table 3.9: IV 2SLS of Corruption on Gini, top 10%

VARIABLES	(1) STATE_AVG	(2) STATE_AVG	(3) STATE_AVG
Top 10%	0.303***		
	(0.0654)		
Top 1%		0.182***	
		(0.0344)	
Gini			0.316***
			(0.0716)
ln_pop	0.0125***	0.0139***	0.0140***
	(0.00214)	(0.00234)	(0.00212)
ln_ethnic	0.250***	0.228***	0.296***
	(0.0273)	(0.0280)	(0.0193)
ln_govemp_sh	0.230***	0.294***	0.176***
	(0.0298)	(0.0395)	(0.0182)
ln_edu	-0.231***	-0.233***	-0.236***
	(0.0114)	(0.0119)	(0.0117)
Observations	1,040	1,040	1,040
R-squared	0.615	0.577	0.598
Hansen J statistic	5.818	0.371	4.249
χ^2 p-value	0.0545	0.8306	0.1195
Shea Partial -R square	0.0622	0.046	0.0891
F-test (first-stage) of excluded instruments	33.094	25.809	54.793

Notes: *** ,** , * denote statistical significance at the 1, 5, and 10 percent levels, respectively. Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable in all columns is the state time-series average corruption per 100,000 population from 1976-2002. Data obtained from Glaeser and Saks (2006). The independent variables are top 10% , top 1%, and Gini respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree, and population. Instruments include slave share in 1860, mature cohort size as a ratio $\frac{40-59}{15-69}$ age group, log income in 1970. F-test (Kleibergen-Paap rk Wald in STATA) tests the joint significance of the instruments from the first-stage regression. Shea- partial R-square denotes the intercorrelation among instruments. Hansen J-stat is an overidentification test.

Table 3.10: IV 2SLS of Corruption on top income inequality 1984-2002

	(1)	(2)	(3)	(4)	(5)	(6)
DEP VARIABLE			CONVICTION			
Top 1%	0.318***		0.340***		0.304**	
	(0.121)		(0.125)		(0.124)	
Top 0.1%		0.207**		0.208**		0.194**
		(0.0817)		(0.0860)		(0.0830)
ln_govemp_sh	0.342***	0.326***	0.354***	0.331***	0.331***	0.313***
	(0.0872)	(0.0834)	(0.0814)	(0.0793)	(0.0879)	(0.0839)
ln_ethnic	0.0552**	0.0663***	0.0455**	0.0601***	0.0548**	0.0662***
	(0.0235)	(0.0204)	(0.0231)	(0.0202)	(0.0244)	(0.0212)
ln_edu	-0.230***	-0.236***	-0.238***	-0.248***	-0.216***	-0.223***
	(0.0424)	(0.0426)	(0.0505)	(0.0497)	(0.0420)	(0.0421)
Mining	-0.294	-0.255				
	(0.229)	(0.227)				
Finance			0.174	0.187		
			(0.236)	(0.233)		
Oil & Gas					-0.194	-0.162
					(0.247)	(0.245)
Observations	720	720	720	720	663	663
R-squared	0.251	0.247	0.246	0.246	0.256	0.250

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: *** ,** , * denote statistical significance at the 1, 5, and 10 percent levels, respectively. Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable in all columns is the state time-series average corruption per 100,000 population from 1976-2002. Data obtained from Glaeser and Saks (2006). The independent variables are top 1% , top 0.1%, and Gini respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree, and population. Instruments include slave share in 1860, mature cohort size as a ratio $\frac{40-59}{15-69}$ age group, log income in 1970. F-test (Kleibergen-Paap rk Wald in STATA) tests the joint significance of the instruments from the first-stage regression. Shea- partial R-square denotes the intercorrelation among instruments. Hansen J-stat is an overidentification test.

Table 3.11: IV 2SLS of Corruption on Gini 1984-2002

VARIABLES	(1) CONVICTION	(2) CONVICTION
Top 10%	0.935***	
	(0.337)	
Gini		1.386***
		(0.526)
ln_govemp_sh	0.349***	0.200***
	(0.0930)	(0.0537)
ln_ethnic	0.0231	0.0556**
	(0.0327)	(0.0219)
ln_edu	-0.201***	-0.214***
	(0.0417)	(0.0407)
Observations	720	720
R-squared	0.223	0.231
Robust standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		
Hansen J statistic	1.038	2.045
χ^2 p-value	0.5952	0.3597
Shea Partial -R square	0.1871	0.0686
F-test (first-stage) of excluded instruments	22.04	21.83

Notes: *** ,** , * denote statistical significance at the 1, 5, and 10 percent levels, respectively. Constants are included in the model but not reported. Robust standard errors in brackets. All columns include region and time fixed- effects. The dependent variable in column 1-3 is the conviction-based corruption measure 100,000 population. The independent variables are top 10% income share, top 1% income share, and Gini respectively. Control variables include ethnic diversity, share of government employment, percentage of population with a college degree, population. Instruments include slave share in 1860, mature cohort size as a ratio $\frac{40-59}{15-69}$ age group, and log income in 1970. F-test (Kleibergen-Paap rk Wald in STATA) tests the joint significance of the instruments from the first- stage regression. Shea- partial R-square denotes the intercorrelation among instruments. Hansen J-stat is an overidentification test.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Ades, A. and Di Tella, R. (1999). Rents, competition, and corruption. *American Economic Review*, 89(4):982–993.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., and Hmous, D. (2015). Innovation and top income inequality. Working Paper 21247, National Bureau of Economic Research.

Aidt, T. S. and Dutta, J. (2008). Policy comprimises: corruption and regulation in a democracy. *Economics & Politics*, 20(3):335–360.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2):155–194.

Alexeev, M. and Song, Y. (2013). Corruption and product market competition: An empirical investigation. *Journal of Development Economics*, 103(C):154–166.

Andres, A. R. and Ramlogan-Dobson, D. C. (2011). Is corruption really bad for inequality? evidence from latin america. *The Journal of Development Studies*, 47(7):959–976.

Apergis, N., Dincer, O. C., and Payne, J. E. (2010). The relationship between corruption and income inequality in u.s. states: evidence from a panel cointegration and error correction model. *Public Choice*, 145(1):125–135.

Azmat, G., Manning, A., and Reenen, J. V. (2012). Privatization and the decline of labour's share: International evidence from network industries. *Economica*, 79(315):470–492.

Bardhan, P. (1997). Corruption and development: A review of issues. *Journal of Economic Literature*, 35(3):1320–1346.

Bardhan, P. (2006). The economists approach to the problem of corruption. *World Development*, 34(2):341 – 348.

Batra, G., Kaufmann, D., and Stone, A. H. W. (2003). *Investment Climate Around the World : Voices of the Firms from the World Business Environment Survey*. Number 15143 in World Bank Publications. The World Bank.

Beck, P. J. and Maher, M. W. (1986). A comparison of bribery and bidding in thin markets. *Economics Letters*, 20(1):1–5.

Berinsky, A. J. and Tucker, J. A. (2006). don't knows and public opinion towards economic reform: Evidence from russia. *Communist and Post-Communist Studies*, 39(1):73 – 99.

Bertocchi, G. and Dimico, A. (2014). Slavery, education, and inequality. *European Economic Review*, 70:197 – 209.

Bliss, C. and Di Tella, R. (1997). Does competition kill corruption? *Journal of Political Economy*, 105(5):1001–1023.

Boylan, R. T. and Long, C. X. (2003). Measuring public corruption in the american states: A survey of state house reporters. *State Politics & Policy Quarterly*, 3(4):420–438.

Campos, N. F., Estrin, S., and Proto, E. (2010). Corruption as a barrier to entry: Theory and evidence. IZA Discussion Papers 5243, Institute for the Study of Labor (IZA).

Cassiman, B. and Vanormelingen, S. (2013). Profiting from innovation: Firm level evidence on markups.

Chatterjee, I. and Ray, R. (2012). Does the evidence on corruption depend on how it is measured? results from a cross-country study on microdata sets. *Applied Economics*, 44(25):3215–3227.

Clarke, G. R. and Xu, L. C. (2004). Privatization, competition, and corruption: how characteristics of bribe takers and payers affect bribes to utilities. *Journal of Public Economics*, 88(910):2067 – 2097.

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–71.

Diaby, A. and Sylwester, K. (2015). Corruption and market competition: Evidence from post-communist countries. *World Development*, 66(0):487 – 499.

Dimant, E. and Tosato, G. (2017). Causes and effects of corruption: What has past decade's empirical research taught us? a survey. *Journal of Economic Surveys*.

Dincer, O. C. (2008). Ethnic and religious diversity and corruption. *Economics Letters*, 99(1):98 – 102.

Dincer, O. C. and Gunalp, B. (2012). Corruption and income inequality in the united states. *Contemporary Economic Policy*, 30(2):283–292.

Dutta, I. and Mishra, A. (2004). Corruption and Competition in the Presence of Inequality and Market Imperfections. Working papers, Centre for Development Economics, Delhi School of Economics.

Emerson, P. (2006). Corruption, competition and democracy. *Journal of Development Economics*, pages 193–212.

Fisman, R. and Svensson, J. (2007). Are corruption and taxation really harmful to growth? firm level evidence. *Journal of Development Economics*, 83(1):63 – 75.

Frank, M. (2009). Inequality and growth in the united states: Evidence from a new state-level pandel of income inequality measure. *Ecnomic Inquiry*, 47.

Glaeser, E. L. and Saks, R. E. (2006). Corruption in america. *Journal of Public Economics*, 90(67):1053 – 1072.

Goel, R. K. and Nelson, M. A. (1998). Corruption and government size: A disaggregated analysis. *Public Choice*, 97(1-2):107–20.

Gupta, S., Davoodi, H., and Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Economics of Governance*, 3(1):23–45.

Gyimah-Brempong, K. (2002). Corruption, economic growth, and income inequality in africa. *Economics of Governance*, 3(3):183–209.

Hall, R. (1988). The relation between price and marginal cost in the u.s. industry. *Journal of Political Economy*, 96(5):921–947.

Higgins, M. and Williamson, J. G. (1999). Explaining inequality the world round: Cohort size, kuznets curves, and openness. Working Paper 7224, National Bureau of Economic Research.

Jensen, N., Li, Q., and Rahman, A. (2010). Understanding corruption and firm responses in cross-national firm-level surveys. *Journal of International Business Studies*, 41(9):1481–1504.

Johnson, N. D., LaFountain, C. L., and Yamarik, S. (2010). Corruption is bad for growth (even in the united states). *Public Choice*, 147(3):377–393.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2007). Governance Matters VI: Governance Indicators, 1996-2005. Policy Research Working Paper Series 4280, The World Bank.

Knack, S. and Keefer, P. (1995). Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics & Politics*, 7(3):207–227.

La Porta, R. and Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives*, 28(3):109–26.

LaPorta, R., de Silanes, F. L., Shleifer, A., and Vishny, R. (1999). The quality of government. *Journal of Law, Economics and Organization*, 15(1):222–279.

Leeson, P. T. and Sobel, R. S. (2008). Weathering corruption. *The Journal of Law and Economics*, 51(4):667–681.

Leigh, A. (2006). Does equality lead to fraternity? *Economics Letters*, 93(1):121 – 125.

Levinsohn, J. and Petrin, A. (2003). Estimating production function using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Marschak, J. and Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, 12:143–205.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3):681–712.

Meltzer, A. H. and Richard, S. F. (1981). A rational theory of the size of government. *Journal of Political Economy*, 89(5):914–927.

Mendez, F. and Sepulveda, F. (2010). What do we talk about when we talk about corruption? *The Journal of Law, Economics, and Organization*, 26(3):493.

Mocan, N. (2008). What determines corruption? internationall evidence from microdata. *Economic Inquiry*, 46(4):493–510.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of Political Economy*, 115(2):200–249.

Olken, B. A. and Pande, R. (2012). Corruption in developing countries. *Annual Review of Economics*, 4(1):479–509.

Olley, G. S. and Pakes, A. (1992). The dynamics of productivity in the telecommunications equipment industry. Working Paper 3977, National Bureau of Economic Research.

Piketty, T. and Saez, E. (2013). Top Incomes and the Great Recession: Recent Evolutions and Policy Implications. *IMF Economic Review*, 61(3):456–478.

Rand, J. and Tarp, F. (2012). Firm-level corruption in vietnam. *Economic Development and Cultural Change*, 60(3):571–595.

Rose-Ackerman, S. (1978). *Corruption: A study in political economy.* Academic Press, New York. NY.

Section, P. (2011). Report to congress on the activities and operations of the public integrity section. Technical report.

Sequeira, S. and Djankov, S. (2010). An Empirical Study of Corruption in Ports. Technical report.

Shleifer, A. and Vishny, R. W. (1993). Corruption. *The Quarterly Journal of Economics*, 108(3):599–617.

Shleifer, A. and Vishny, R. W. (1994). Politicians and firms. *The Quarterly Journal of Economics*, 109(4):995–1025.

Svensson, J. (2003). Who must pay bribes and how much? evidence from a cross section of firms*. *The Quarterly Journal of Economics*, 118(1):207.

Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives*, 19(3):19–42.

Swamy, A., Knack, S., Lee, Y., and Azfar, O. (2001). Gender and corruption. *Journal of Development Economics*, 64(1):25 – 55.

Treisman, D. (2000a). The causes of corruption: a cross-national study. *Journal of Public Economics*, 76(3):399–457.

Treisman, D. (2000b). The causes of corruption: a cross-national study. *Journal of Public Economics*, 76(3):399 – 457.

Uslaner, E. (2008). *Corruption, Inequality, and the Rule of Law: The Bulging Pocket Makes the Easy Life*. Cambridge University Press.

Wei, S.-J. (2000). How taxing is corruption on international investors? *Review of Economics and Statistics*, 82(1):1–11.

You, J.-s. and Khagram, S. (2005). A comparative study of inequality and corruption. *American Sociological Review*, 70(1):136–157.