# The Trickster's Legacy: The Folklore Roots of Corruption

**Jean-Baptiste Marigo** [#]
University of Strasbourg

**Laurent Weill** [*]
EM Strasbourg Business School, University of Strasbourg

## Abstract

This study examines the role of oral traditions in shaping corruption. Using firm-level data on a large sample of 82,922 firms covering 285 cultural societies in 125 countries, we analyze how the representation of antisocial behavior in folklore influences corruption. Our findings reveal that societies with folklore portraying antisocial behaviors as successful tend to exhibit higher levels of corruption, while those with oral traditions emphasizing the negative outcomes of such behaviors experience lower corruption. By distinguishing demand- and supply-side corruption channels, we demonstrate folklore's pervasive influence on bribe-related decisions, affecting firms as both initiators and targets. Thus, our work contributes to the understanding of deep-rooted cultural determinants of corruption.

**JEL Codes**: D73, O17, Z10, Z13

**Keywords**: culture, folklore, corruption.

[#]Sciences Po Strasbourg, 7 rue de l'Ecarlate, CS 20024, 67082 Strasbourg, France. Email: jb.marigo@unistra.fr

[*]Corresponding author. Sciences Po Strasbourg, 7 rue de l'Ecarlate, CS 20024, 67082 Strasbourg, France. Email: laurent.weill@unistra.fr

# 1. Introduction

Corruption remains one of the most pervasive and damaging challenges facing societies worldwide. The United Nations Development Programme has targeted corruption as an obstacle to its Sustainable Development Goals, asserting that corruption is both a development issue "and most importantly, a trust issue of citizens in their government institutions".[1] Indeed, a large body of literature has shown that corruption exerts a detrimental impact on economic growth (Mauro, 1995; Fisman and Svensson, 2007; D'Agostino et al., 2016; Gründler and Potrafke, 2019; Uberti, 2022) and erodes trust in institutions (Anderson and Tverdova, 2003, Hakhverdian and Mayne, 2012).

Fighting corruption requires policy recommendations enlightened by accurate insights on its consequences, as well as also on its causes. A wide range of determinants of corruption have been identified, including economic factors such as economic development, income inequality, and natural resources (Treisman, 2000), and institutional factors such as the form of government and the democratic characteristics of a state (Serra, 2006). Research has also emphasized the importance of underlying cultural institutions such as dominant religious affiliation, colonial history, and trust (Svensson, 2005; Treisman, 2007; Dimant and Tosato, 2018). Uslaner (2004, 2012) establishes a negative relationship between trust and corruption. He notes that "corruption does not depend upon venal leaders. Rather, it is ingrained into the political culture of a society" (Uslaner, 2004, p.9). In this sense, he emphasizes how deep the roots of corruption can run, and concludes how slowly the corruption, or the culture of corruption changes, if at all.

Our aim in this paper is to investigate the role of folklore on corruption. We build upon the recent research by Michalopoulos and Xue (2021), which examines the role of folklore in shaping economic outcomes and attitudes in contemporary societies. Folklore is defined as the "collection of traditional beliefs, customs, and stories of a community passed through the generations by word of mouth" (Michalopoulos and Xue, 2021, p. 1994). Based on the original anthropological data collected by Berezkin (2015), they utilize a collection of oral traditions covering 958 cultural regions for which the anthropologist identifies 2,564 motifs, i.e., "a combination of images, episodes, or structural elements found in two or more texts, including sacred and profane ones" (Michalopoulos & Xue, 2021, p.1995). They conduct country-level evidence that oral traditions explain contemporary variations in trust, gender bias, and risk-

[1]https://www.undp.org/governance/inclusive-and-future-smart-public-goods-and-services/anti-corruption#:~:text=UNDP%20integrates%20anti%2Dcorruption%20solutions,%2Dcorruption%2C%20and%20leading%20anti%2D

taking across societies. In the case of trust, they consider motifs associated with either punishment or no punishment of antisocial behavior in oral traditions. They perform country-level estimations showing that the degree of relative punishment of antisocial behavior has a positive effect on trust measures. Consequently, they show that individuals are more trustworthy if they were raised on stories in which antisocial behavior is punished.

We extend their work by focusing on the way antisocial behaviors are depicted in folklore following the idea that these are deep-rooted cultural factors that are related to trust. Oral traditions with antisocial behavior feature characters violating social norms like theft or deceit, or exhibiting a disregard for the well-being of others like selfishness. We test the hypothesis that the way antisocial behavior is depicted in folklore affects corruption. We assume that positive descriptions of antisocial behavior in oral traditions promote corruption through their weakening effect on trust. Narratives in which antisocial behavior leads to positive outcomes for the trickster may subtly endorse the acceptability of such actions, potentially incentivizing unethical behavior in real-world contexts. Conversely, stories that highlight the adverse consequences of such behavior are likely to strengthen norms of accountability and trust, and thus reduce corruption.

We conduct an empirical investigation at the firm level. We combine firm-level data from the World Bank Enterprise Survey (WBES) with the measures of folklore from Michalopoulos and Xue (2021). We construct a final sample of 82,922 firms from 287 cultural groups located in 125 countries over the period 2005-2022. This novel dataset allows us to analyze the relationship between folklore and corruption, distinguishing between demand- and supply-side channels of bribery. We maintain the scale of the cultural regions as originally identified by Berezkin (2015), rather than aggregating data at the country level like Michalopoulos and Xue (2021). Thus, our approach preserves the heterogeneity of cultural influences at the ethnolinguistic level, rather than following borders.

In our main estimations, we test the effect of folklore on corruption using measures of corruption that take into account the different dimensions of corruption. Our main measure of corruption is a perception measure of the extent to which respondents see corruption as an obstacle for the operations of their firm. We additionally use a measure of the reported amount of bribes in percentage of total annual sales that establishments like the respondent's one pay in informal payments or gifts to public officials. We find evidence that folklore affects corruption. We show that firms located in cultural societies with higher (lower) sanctions of antisocial behaviors in narratives are less (more) plagued with corruption.

We further investigate the channels through which folklore affects corruption. Following Gauthier et al. (2021), we distinguish between the demand side and the supply side of corruption. The demand side of corruption occurs when the respondent has accepted a bribe request from a corrupt official. However, corruption can also be initiated by the respondent in order to obtain public contracts or circumvent regulations. Thus, the supply side of corruption occurs when bribe is at the giver's initiative. We therefore examine the impact of folklore on corruption by considering these different bribe-paying behaviors, whether the firm is the target or originator of the bribe. We find that folklore has effects on both a firm's propensity to pay a bribe and a firm's propensity to be the target of a bribe.

We additionally investigate the influence of firm characteristics on the relationship between folklore and corruption. We find no evidence that firm size, foreign ownership, and exporting activities shape the influence of folklore on corruption at the firm-level. Overall, we conclude that folklore still affects corruption consistently, despite variations in the characteristics of the firm. We conduct robustness tests to address potential biases we are concerned about. First, we drop observations from establishments located in cultural regions that belong to the lowest decile in terms of the number of motifs, i.e., folklore materials, collected. In this way, we hope to rule out the possibility that our data could be unrepresentative of the population studied. Second, we use an instrumental variable approach to account for a potential omitted variable bias. Our results remain unchanged.

This paper contributes to the literature in two ways. First, we contribute to the literature on corruption by investigating the influence of deep-rooted cultural institutions. In this area, religion has been particularly scrutinized as a potential determinant of corruption (Treisman, 2000; Guiso, Sapienza and Zingales, 2003; North, Orman and Gwin, 2013). We extend this literature by showing the influence of folklore. Second, we contribute to the literature on the influence of culture by adding to Michalopoulos and Xue (2021) for the effects of folklore. Our paper adds to this recent work by providing firm-level evidence at the level of cultural groups to show the influence of folklore on different dimensions of corruption.

The paper is organized as follows. Section 2 provides the background of the research question. Section 3 describes the data and methodology. Section 4 presents the empirical findings. Section 5 concludes.

# 2. Background

## 2.1 Folklore: a catalog of motifs

Michalopoulos and Xue (2021) rely on the work of anthropologist and folklorist Yuri Berezkin, who comprehensively studied oral traditions and mythology on all continents. Berezkin (2015, p.58) refers to folklore as "all kinds of traditional stories and tales, long and short, sacred and profane" that are present in at least two ethnolinguistic communities. These tales are transmitted vertically, from one generation to another, or horizontally, through contact between cultural groups. Other works such as by Spolaore and Wacziarg (2013) explore these transmission mechanisms. Indeed, the authors show that the direct parents' education and indirect influence of symbolic school and religious systems pass on cultural traits vertically. As a result, these traits form deep-rooted sets of values and beliefs that affect contemporary economic outcomes. However, they also argue in favor of the existence of horizontally transmitted traits, as they describe how certain characteristics spread across individuals and cultural groups, even if they are less directly related.

Berezkin provides two major contributions. First, he identifies cultural groups all around the word, for which he defines a geographical center called "centroid". These are characterized by a latitude and longitude, which allows to locate them. For example, the *Wales* centroid corresponds to the following set of coordinates: latitude 52, and longitude -4. Overall, Berezkin identifies 958 cultural societies across 199 countries from all continents.

Second, Berezkin collects thousands of "motifs" from all cultural areas. These are text elements, episodes, or images "retold or described in narratives that are registered in at least two […] different traditions" (Berezkin, 2015, p.61). They are narrative units that make the content of a tale or myth. For example, a widespread motif titled "Tasks of the in-laws" in Berezkin's catalog is associated with the following description: "Father or other kinsmen of hero's wife or bride try to kill or test him and/or suggest to him difficult tasks". Such an episode can be found in many oral traditions across different ethnolinguistic communities. In total, Berezkin collects 2,564 motifs, the median motif spanning 18 oral traditions, and the median cultural group comprising 62 motifs. Using these data, it is possible to identify the motifs that are present in any folklore, along with their content.

## 2.2 Underlying cultural factors

The idea is that folklore carries cultural representations that are characteristic of the corresponding ethnolinguistic community. We should therefore find these representations at the individual-level, which means that they could shape the behavior of economic agents. The temporality of the data on folklore supports such an argument. Berezkin uses different written sources such as books and articles from thousands of authors and publishing houses, which mainly appeared during the early XX$^{th}$ century. All of these sources list many oral traditions on different folklore groups, which stem from ancestral stories and tales. Furthermore, it is possible that these oral transmissions are still being transmitted vertically and horizontally nowadays, which means that they may continue to affect individuals nowadays with respect to their values and beliefs. Unfortunately, this folklore dataset does not provide us with any information about the popularity of a motif in a cultural society: it simply tells us whether it exists in a group.

Why should we believe that oral traditions have ancestral roots? A strand of the anthropology literature supports the idea that contemporary tales find their origin thousands of years ago. This means that their influence on the set of values and beliefs of a group should hold over long periods. For example, Silva and Tehrani (2016, p.8) use phylogenetic methods combined with autologistic modelling to show that the Beauty and the Beast fairy tale "can be securely tracked back to […] between 2500 and 6000 years ago". Additionally, Michalopoulos and Xue (2021) study the relationship between folklore and certain outcomes such as innovation, entrepreneurship, and trust. In the end, they show a cultural effect that persists over second-generation immigrants. This would correspond to the vertical transmission of cultural traits through the education of parents, despite the change of cultural region. The new set of cultural traits would therefore be partially inherited through the influence of other vertical institutions such as the education or religious systems. It could even be due to a horizontal transmission through contact with individuals from the new group. Such a result further supports the idea that the sets of values and beliefs depicted in a folklore produce long-lasting effects on individuals, even despite migrations.

Interestingly, we note that Akerlof and Snower (2016) study the role of narratives in shaping a society's economic behaviors, using the example of the telling of the Bolshevik and Soviet revolution. They show that narratives carry what they define as "categories", i.e., mental images that are imperfect reflections of reality, which work as simplified processes for perception and decision-making. These categories should be reminiscent of Berezkin's motifs in folklore. Indeed, Akerlof and Snower also show that their categories shape the mental representations of individuals.

Overall, it appears that the content of oral traditions can hold across millennia, and that their effect on the perceptions and behaviors of individuals persist despite generations and migrations. Thus, we should find representations, values and beliefs of individuals in the motifs of the ethnolinguistic community to which they belong. To this extent, these motifs should still be characteristic of members of a cultural group nowadays, such that "images and episodes in folklore appear to endure and possibly still shape how individuals perceive the world today" (Michalopoulos and Xue, 2021, p.2041). This is why data on folklore can help explain contemporary measures of economic outcomes.

**2.3 Classifying motifs**

To create a dataset on folklore suitable for a quantitative approach, Michalopoulos and Xue (2021) categorize motifs based on their theme and outcome. They study three cultural components: challenge and competition (risk-taking), antisocial behaviors (trust), and various representations of men and women (gender norms). They choose to aggregate folklore data at the country-level. However, we will not replicate this choice, as we managed to keep Berezkin's data at the centroid-level, i.e., following cultural geography rather than national borders.

First, they use a semantic network, *ConceptNet,* to find a list of words related to their concept of interest. As we are interested in antisocial behavior (absurd, obscene, or deceiving), we will focus on this theme throughout our explanation. To generate its output, the network requires a seed word, so that it can establish a list of the top-50 related terms. Here, Michalopoulos and Xue used the words *cheat*, *deceive*, and *trick* as seeds for different iterations related to antisocial behaviors. The authors then match this list with the motifs from the folklore dataset. For example, multiple motifs relate to the theft of fire, which is a recurring theme in folkloristics. The tale depicts how trickster figures managed to steal fire from a supernatural guardian for the benefit of humanity. Such figures are often animals, such as possums, swallows, beavers, dears, parrots, etc. In the original folklore dataset, motif *d4h* reads: "Swallow steals fire for people". Following the *ConceptNet* approach, the authors thus matched this motif with at least one of the lists associated with their three seed words on antisocial behaviors.

Second, they include a human-based assessment in their methodology. To do so, they hired on average nine Amazon Mechanical Turks (MTurks) to characterize the outcome of the motifs matched as related to antisocial behaviors. MTurks were tasked with categorizing the motifs based on whether the outcome was positive or negative for the trickster. More precisely, they consider that a trickster can be punished or simply unsuccessful (negative outcome), or

can get away with his behavior (positive outcome). Alternatively, it is also possible that the outcome of a motif cannot be characterized with certainty as positive or negative. Therefore, they can classify the outcome as either *Successful*, *Unsuccessful*, or *Unclear*. Otherwise, they also have the possibility to state that the motif does not depict an antisocial behavior, effectively contradicting *ConceptNet*. In these cases, these motifs are excluded from our folklore dataset. With our example, the MTurks mainly describe the *dh4* motif on the theft of fire as *Successful*, which means that the trickster gets away with his behavior. On the contrary, should the workers have chosen to describe the outcome as *Unsuccessful,* this would have meant that the trickster is punished in the end. It is important to note at this stage that only the content present in a motif is characterized. In this way, motif *dh4* on the theft of fire does not tell us anything about how the whole story of the oral traditions in which it appears.

When applying this methodology to the folklore dataset, we get the number of motifs in each cultural group that depicts all possible outcomes with respect to antisocial behaviors. It is therefore possible to compute the shares of *Successful* or *Unsuccessful* motifs for each culture, which opens the way to quantitative comparisons.

### 2.4 Folklore and corruption

An extensive literature has scrutinized the determinants of corruption with notably political institutions (Treisman, 2000; Svensson, 2005; La Porta et al., 2008; Chatterjee and Ray, 2012; Dimant and Tosato, 2018) but also cultural factors with notably the influence of religious affiliation or colonial history (Mauro, 1995; Svensson, 2005; Treisman, 2007; Dimant and Tosato, 2018).

Further arguments motivate the potential effect of culture on corruption, relying on the relationship between trust and corruption on the one hand, and antisocial behaviors in folklore on the other hand.

First, Uslaner (2004) shows that beyond the relationship between these two variables, it is possible to establish a causal linkage. Thus, when comparing different parameters of corruption, he writes “the strongest determinant of change in corruption is change in trust” (Uslaner, 2014, p.18). Moreover, he adds that the roots of trust lie in cultural factors. For example, he uses the concept of *raccomodazione*, a phenomenon of corruption deep-rooted in the Italian culture, such as highlighted by Husted (1999). Uslaner argues that changing the institutional framework is not enough to reduce corruption, as the latter is ingrained into the political culture of a society. Chatterjee and Ray (2012) also report that trust in the judicial system should reduce business corruption, while Clausen, Kraay and Nyiri (2011) show a

negative correlation between confidence in the institutions and corruption. Both measures of interpersonal trust or trust in the institutions should yield the same results, as supported by Rothstein (2000).

Second, Michalopoulos and Xue (2021) prove that folklore that emphasizes successful outcomes in antisocial behaviors is negatively associated with multiple measures of trust nowadays, such as from the Integrated Values Survey and the Global Preferences Survey. To this extent, we can consider that the representations of successful or unsuccessful behaviors in folklore are a way to measure the underlying cultural factors related to trust such as mentioned by Uslaner (2004, p.22), when he talks about a "culture of corruption". Through representations of antisocial behavior, we therefore get informative insights on trust in each cultural community. For these reasons, we expect a causal relationship between folklore and corruption.

# 3. Data and methodology

## 3.1 Data

We combine the folklore data from Michalopoulos and Xue (2021) with the World Bank Enterprise Survey (WBES) dataset. The latter provides us with the various measures of corruption we need, as well as much firm-level information, which we use to construct controls. We assign firms to a cultural society using the shortest distance between the location of each establishment and the centroids, i.e. the geographical centers of cultural groups.

We obtain a final sample of up to 82,922 firms and 287 cultural groups located in 125 countries over the 2005-2022 period. From the initial dataset of 195,824 firms, we lose observations for multiple reasons. First, we drop all observations for which the reliability of the respondents' answers to the WBES questionnaire is not marked as truthful.[2] Although we lose 74,000 observations, we therefore avoid introducing bias in our final sample. Then, we drop 32,000 firms due to one or multiple lacking or incorrect values for our firm-level control variables. This leaves us with large sample of more than 89,000 firms, from which we further lose around 4,000 observations when adding the country-level control variables. Then, we exclude about 3,000 companies for which we do not have any information regarding corruption.

[2] For each questionnaire, the perception of the questions regarding opinions and perceptions is categorized as either « truthful », « somewhat truthful », or « not truthful » by the screener

### 3.2 Measuring folklore

We use the anthropological data collected by Berezkin (2015), and more specifically the information on the representations of antisocial behaviors in folklore such as defined by Michalopoulos and Xue (2021) to explain corruption at the firm-level.

Following the explanations in the background section, we use the shares of motifs as a relative measure of the over- and underrepresentation of successful or unsuccessful antisocial behaviors in each cultural group. We compute a share by dividing the number of motifs characterized by a particular outcome by the total number of collected motifs, whether or not they represent antisocial behaviors. In total, we create five different cultural predictors based on the outcome of the motif.

If a motif depicts a situation in which one or multiple characters undertake antisocial behaviors and get away with their misconduct, then the label is *Success*. However, if the character is punished for her/his behavior, the outcome of the motif is *Not success*. If it is not possible to determine with certainty whether the outcome is successful or not, the classification of the motif is *Unclear*. Using all motifs for antisocial behaviors regardless of their outcome, we also create an *All* variable. The latter then indicates the percentage of total collected motifs in a cultural group that focus on this theme. Finally, following Michalopoulos and Xue (2021)'s methodology, we also provide a *Relatively unsuccessful* predictor by subtracting the shares of successful motifs from the percentage of unsuccessful ones. The coefficient estimate for this variable thus provides us with a direct and singular interpretation of the effect of antisocial representations in folklore on corruption.

Because folklore variables can be highly correlated, we use three specifications of folklore variables in the estimations. First, we only include the *All* variable, as to reflect how much of a concern antisocial behaviors are for a cultural group, without any information at this stage on how there are portrayed. Second, we use the *Success*, *Not success*, and *Unclear* variables, with a particular focus on the two first ones. Indeed, they provide us with information on whether characters who engage in antisocial behaviors get away with their misdeeds, or are sanctioned. In other words, we can see whether a folklore describes antisocial behaviors more or less positively and negatively. Third, we include only *Relatively unsuccessful* as a synthetic measure of antisocial representations.

For example, we can see that the Kara Kalpak cultural group in Uzbekistan is one of the groups in our final sample for which antisocial behaviors in folklore are most often successful. 16 of its 136 collected motifs depict such an outcome, which gives a share of 11.76% for the *Success* variable. On the opposite, the Saho and Afar folklore in Djibouti is characterized by an

*Not success* variable of 11.32%, as 6 out of its 53 total motifs tell the story of antisocial behaviors that end up being punished or unsuccessful.

In our final sample, the median number of total collected motifs is 72, with the lowest and highest percentiles being respectively equal to 4 and 565 motifs. Thus, we know that the shares of successful or unsuccessful motifs might be highly sensitive to the number of listed motifs. To this extent, we winsorize the shares of motifs for which the outcome is successful at the one percent level for the highest values. In this way, we hope to avoid introducing potentially overinflated values in our regression. Furthermore, we later provide a robustness test by re-estimating the effects of our folklore variables after removing all firms that are located in cultural areas that belong to the lowest decile of total collected motifs.

### 3.3 Measuring corruption

Corruption is inherently difficult to measure. In the literature, two main methods based on questionnaires are used to obtain data on corruption, which vary with the type of respondent. On the one hand, there are methods relying on the evaluations of experts from each country who provide a national assessment of corruption, such as the Corruption Perceptions Index from Transparency International. On the other hand, there are methods based on data from individual respondents, whether they are citizens or managers with insider knowledge of the prevalence of corruption in their economic activity or sector, such as data from the WBES dataset. Chatterjee and Ray (2012) specifically examine whether the evidence varies depending on the type of corruption data used. They find that all types of corruption measures are highly correlated.

When asked to individuals, questions about corruption are always at least partially indirect. The idea is that the respondent should not feel threatened by the survey. Indeed, the respondent could believe that her/his answers could be provided to the government non-anonymously. In the WBES dataset, one question (J.7) therefore takes the following form: “It is said that establishments are sometimes required to make gifts or informal payments to public officials to “get things done” regarding customs, taxes, licenses, regulations, services etc. On average, what percentage of total annual sales […] do establishments like this one pay in informal payments or gifts to public officials for this purpose?” One might question the nature of the insight that such a question provides. However, Gauthier et al. (2021) show that firms answer to these types of inquiries based on their own experience of corruption. In this respect, we believe that these data are informative, especially since the WBES dataset provides us with information on the reliability of the responses.

We use the answers of top managers and owners of businesses to various WBES questions on corruption, which allows us to obtain alternative measures of corruption. First, we use data on the degree to which the respondents perceive corruption as an obstacle for the operations of the establishment. The answer varies between 0 (*No obstacle*), 1 (*Minor obstacle*), 2 (*Moderate obstacle*), 3 (*Major obstacle*), and 4 (*Very severe obstacle*). We note that the manager can also choose to answer that she/he does not know, or that the question does not apply. Overall, and for all WBES variables, we drop all observations for which we have these two last answers. Using this question, we construct two measures of corruption. We create a dummy variable *Corruption Obstacle_D,* equal to one if the respondent answers 3 or 4, meaning corruption is perceived as a major or very severe obstacle for the operations of the establishment, and takes the value zero otherwise. We also use *Corruption Obstacle,* which is an ordinal variable equal to the answer from 0 to 4.

Second, we use the answers to the following question "On average, what percentage of total annual sales [...] do establishments like this one pay in informal payments or gifts to public officials […]?" In addition, the questionnaire first points out that "it is said that establishments are sometimes required to make gifts or informal payments to public officials to "get things done" regarding customs, taxes, licenses, regulations, services etc". Although this question is not directly related to the situation of the surveyed firm, Gauthier et al. (2021) show that the answers reflect the actual amounts spent by the firm in terms of bribes. Thus, such a formulation maximizes the probability that a respondent will answer this question on corruption without fear of repercussions. This continuous variable is named *Bribe Amount*.

### 3.4 Methodology

To analyze the impact of folklore on corruption, we run regressions with the following specification:

$$Corruption_{ik} = \beta_0 + \beta_1 Folklore_k + \gamma Controls_{ik} + \varepsilon_{jk}$$

Where $i$ is the firm, $k$ the cultural society**,** $Corruption_{ik}$ the corruption variable, $Folklore_k$ the set of folklore variables, $Controls_{ik}$ the set of controls, and $\varepsilon_{ik}$ the error term. We include year fixed effects to capture any exogenous shock for any year. As explained above, we consider three specifications for the set of folklore variables: only the *All* variable; the variables *Success*, *Not success*, and *Unclear*; or *Relatively unsuccessful*.

We use three types of models for our estimations, depending on the nature of the corruption variable. We use a probit model for the binary corruption measures (*Corruption Obstacle_D*, *Demand*, *Supply, Incorruptible,* or *Non-payers*). We estimate an ordered logit model when explaining the ordinal measure of corruption (*Corruption Obstacle*). Finally, we use an OLS model when explaining the continuous measure (*Bribe Amount)*.

One potential concern is that our results may be influenced by endogeneity. However, the nature of our data rules out the possibility of reverse causality: data on folklore come from written sources from the early XX$^{th}$ century. Moreover, the content of these sources is made of even older folklore material. In comparison, data on corruption cover the past two decades. To address omitted variable bias, we include a comprehensive set of variables at the firm, centroid, and country levels, which should account for many unobserved characteristics. Still, we later provide an instrumental variable approach for a robustness check to test against endogeneity and bring further evidence that our estimates hold.

We follow previous works on corruption to determine our set of controls (Treisman, 2007; Rand and Tarp, 2012; Dimant and Tosato, 2018). We include a set of firm-level controls. We use the *Age* of the firm. We add the *Experience* of the manager using her/his number of years in the industry. We include the *Size* of the firm measured by the natural logarithm of the number of full-time permanent workers. Further, we use the reported sales of a company on the last available three-year period to compute its *Sales growth*. Doing so, we observe that some values are far beyond a 100% growth. Therefore, we winsorize the two percent highest values in the right tail of the distribution

We include dummy variables if the firm is a sole proprietorship (*Sole proprietorship*), is privately held (*Private or non-traded*) or publicly traded (*Publicly traded*), is owned by foreign investors (*Foreign-owned*). We also include dummy variables if the firm is an exporter (*Exporter*), the firm's financial statements are audited (*Audited*), and if the firm belongs to a larger group (*Subsidiary*). We also take into account the industry of the firm with dummy variables *Manufacturing* and *Retail*.

We create three subnational control variables at the centroid level to include cross-regional differences within countries that can influence corruption, following Osei-Tutu and Weill (2024). We compute these variables by using questions from the WBES dataset to create a group mean for each cultural society. First, we select the winsorized three-year sales growth values at the firm-level to create the *Growth centroid* predictor. To do so, we calculate a mean rate by cultural area. We also create the *Electricity centroid* variable to account for the quality of the electricity grid and infrastructures at the regional level. To do this, we use the responses

to the WBES survey on the difficulty of access to electricity as perceived by businesses, the latter being an ordinal variable comprised between 0 ("no obstacle") and 4 ("very severe obstacle"). Thus, we calculate the median by group and round it down to the nearest unit, following a conservative approach. Finally, we build a third control (*Exporter centroid*) by computing the mean of the percentage of sales as direct exports of companies in each cultural group. These predictors then help us capture regional effects in terms of economic growth, development, and openness to trade, all of which are important in determining corruption.

Finally, we include some country-level controls. We use the natural logarithm of income per capita (*GDP per capita*) as a measure of economic development. We use the *Fuel exports* variable as the share of fuel exports in merchandise exports to control for the effect of a potential resource curse on corruption. Both predictors come from the World Development Indicators database. We include *Rule of law*, extracted from the Worldwide Governance Indicators, to capture the quality of institutions. We rescale this indicator from an initial range of -2.5 to 2.5 to a range of 0 to 10, to allow for a better reading of the results. We use data from the World Religion Project to obtain the mean percentage of protestant population per country over the 1945-2010 period (*Protestants*), following the existing literature on corruption. Finally, we include three dummy variables on the legal origin of the institutions of each country, using the La Porta et al. (2008)'s dataset. We provide descriptive statistics for all variables in Table 1. All variables are defined in the Appendix.

## 4. Results

### 4.1 Main estimations

We report the estimations for the impact of folklore on corruption following our three corruption measures: *Corruption Obstacle_D* in Table 2, *Corruption Obstacle* in Tables 3-4, and *Bribe Amount* in Table 5.

To analyze the relationship between folklore and corruption, we consider the three specifications adopted by Michalopoulos and Xue (2021) for their folklore variables in each each of our tables. In column (1), we include *All* to take into account all motifs that depict antisocial behaviors. In column (2), we consider *Success*, *Not success*, and *Unclear* to distinguish between the different outcomes. We include the *Unclear* variable because it provides information about folklore that we can use to disentangle the effects between our

different cultural treatment variables. However, the estimates for this variable are not easily interpretable with respect to our key hypotheses. Therefore, we consider it as a folklore control variable. In column (3), we use *Relatively unsuccessful*, which is a single measure of the outcome of representations of antisocial behaviors. Thus, positive values indicate that the folklore of a cultural group provides depictions of sanctions more frequently than positive outcomes in the case of antisocial behavior.

We start our analysis with the first specification. We find that *All* is significantly negative when explaining *Corruption Obstacle_D* and *Corruption Obstacle,* but significantly positive when explaining *Bribe Amount*. These results suggest that the mere focus of a folklore on antisocial behavior is associated with lower levels of corruption in terms of perception but higher levels of corruption in terms of amount of bribes. They are ambiguous to interpret, but it must be emphasized that the *All* variable does not account for the outcome of tales. Therefore, we focus on the interpretation of our second and third model specifications.

We can then question whether the effect of folklore on corruption is conditional on the outcome of tales. A striking result is the significantly positive coefficient for *Success* and the significantly negative coefficient for *Not success* with all corruption variables. These results imply that when a folklore is associated with a more positive portrayal of antisocial behaviors, firms face higher corruption. Reversely, a more negative portrayal of antisocial behaviors helps reduce corruption faced by firms.

We observe differences across estimations for the *Unclear* variable. It is significantly negative with *Corruption Obstacle_D* and *Corruption Obstacle,* but significantly positive with *Bribe Amount*. Given that we use *Unclear* as a control variable, being not easily interpretable for our key hypothesis, we cannot conclude anything about these results.

The results of the third specification confirm those of the second one. We find a significantly negative coefficient for *Relatively unsuccessful* with all corruption variables. This result shows that a higher difference between unsuccessful and successful outcomes of antisocial behaviors increases corruption.

In terms of economic magnitude, we further observe that these results are economically significant. Looking at the *Corruption Obstacle_D* variable from Table 2, we observe that a 10 percentage points increase in the share of unsuccessful (successful) motifs is this time associated with a 2 percentage points decrease (increase) in the probability that a firm sees corruption as a significant obstacle.

Interestingly, for all three corruption measures, we observe that the order of magnitude is similar for motifs with successful outcomes and for those with unsuccessful outcomes. In other words, we do not observe a greater sensitivity of corruption to either type of tales.

Thus, our main conclusion is that folklore has an impact on corruption. The outcome of oral traditions in tales related to antisocial behavior influences the prevalence of corruption among the firms in a cultural society. More successful antisocial actions lead to more prevalent corruption, in line with the view that they hinder trust in the society. Conversely, more unsuccessful antisocial actions reduce corruption for symmetric reasons.

### 4.2 Exploring the channels of corruption

We have shown that folklore affects corruption. We can now investigate the channels through which folklore exerts an impact on corruption. Literature generally concentrates on the demand side of corruption, through which bribes results from requests by public officials. The firm then agrees to pay the bribe not to be punished by the administration or not to be excluded from trade. However, corruption can also come from the supply side in the sense that bribes can be at the initiative of firms. Firms may indeed initiate bribes to get some benefits, such as avoiding a burdensome regulation or obtaining public contracts or privileges.

Gauthier et al. (2021) introduce a distinction between the demand and supply sides of corruption using the WBES data to compare both channels. Following their approach, we create a *Corruption demand* dummy variable, equal to one if the respondent reports having accepted at least one type of bribe request from a corrupt official. Similarly, we define a *Corruption supply* dummy variable that equals one if the firm paid a bribe without any request from the official. In this way, we distinguish between the demand and supply sides of corruption.

In addition to these bribe-paying behaviors, we follow Gauthier et al. (2021) to categorize non-bribe paying firms with two dummy variables. *Incorruptible* is equal to one when there was at least one corruption request from a corrupt official without any bribe payment from the firm. *Non-payers* is equal to one if there was neither a bribe request nor a bribe payment. Both categories provide information about the channels of corruption by providing information about firms that do not face or accept to pay bribes.

We first examine whether folklore affects the demand-side of corruption. Table 6 reports the estimations using *Corruption demand* as the dependent variable. We find that *All* is significantly positive. This means that a greater focus on antisocial behavior in folklore increases the probability that firms accept at least one corruption request from corrupt officials.

We therefore turn to the specification that differentiates between the outcomes of motifs. On the one hand, *Success* is significantly positive, supporting the view that firms in cultural groups where misbehaviors are more often depicted as successful are more likely to accept bribe requests. On the other hand, *Not success* is not significant, implying that misbehaviors portrayed as not successful do not affect corruption demand. Finally, we comment on the last specification, which shows a significantly negative coefficient for *Relatively unsuccessful*. This accords with the view that folklore with relatively less successful than unsuccessful misbehaviors reduces corruption demand.

We then turn to the analysis of the effect of folklore on the supply-side of corruption. Table 7 displays the estimations by using *Corruption supply* as the dependent variable. We find similar results as for the demand-side of corruption. While *All* is significantly positive, we observe that *Success* is significantly positive and *Not success* is not significant, while *Relatively unsuccessful* is significantly negative. Here again, the latter result is consistent with the findings for *Success* and *Not success*.

We complete this analysis by considering the two additional categories of firms in the taxonomy proposed by Gauthier et al. (2021). We examine the impact of folklore on *Incorruptible* and *Non-payers* respectively in Tables 8 and 9. This brings information about the influence of tales on the probability of being an incorruptible firm, meaning a firm that does not respond positively to a bribe request from a corrupt official, or on the probability not to be concerned by corruption.

Both tables of estimations yield similar results. *All* is significantly negative in both tables, meaning that a folklore that focuses more on antisocial behavior increases the probability to be incorruptible or non-payer. We find that *Success* is significantly negative and *Not success* is not significant in both estimations. The finding for *Success* is consistent with the view that firms located in cultural regions where folklore more often depicts misbehaviors as successful have a lower probability of refusing a request for a bribe, or of not being a target. The finding for *Not success* again shows that tales with unsuccessful misbehaviors do not affect corruption.

Overall, these findings support the view that folklore plays a similar impact on the demand-side and the supply-side of corruption, as well as on the likelihood of being an incorruptible or a non-payer firm. They also show that the channels of corruption are influenced only by the depictions of positive outcomes in antisocial behaviors.

### 4.3 Firm heterogeneity

The results reported so far show the influence of folklore on corruption. We have, however, performed our investigation for all types of firms without considering if this impact differs among them. We thus examine the effect of three firm-level characteristics.

First, we investigate whether the impact of folklore on corruption varies with firm size. A lower influence of folklore on large firms can be attributed to several factors. First, large firms typically exhibit a higher degree of rationalization in their financial decision-making processes, which minimizes the impact of cultural influences. Second, large firms often have a more global presence, with numerous subsidiaries and a workforce recruited from diverse regions, reducing their connection to local cultural environments. In contrast, small firms are more likely to be deeply embedded in their local environment, making them more susceptible to influences such as folklore.

Second, we examine whether foreign ownership influences the impact of folklore on corruption. We expect that folklore should have a lower influence on foreign-owned firms, because the foreign influence from their ownership should partially immune them to oral traditions from the location of the firm.

Third, we investigate whether being an exporter affects the effect of folklore on corruption. Exporting firms could be less affected by folklore than non-exporting firms. These firms are more likely to employ staff from different cultural backgrounds or regions, diluting the influence of the local traditions. Broadly speaking, their global orientation through international clients and partners can reduce their sensitivity to local folklore.

We focus on the *Relatively unsuccessful* variable, as it provides a synthetic measure of both the *Success* and *Not success* outcomes. We perform all estimations with our main corruption variable, *Corruption Obstacle_D.* We redo our estimations by adding now alternatively the interaction term between *Relatively unsuccessful* on the one hand, and *Size, Exporter,* or *Foreign-owned* on the other hand. Because the dependent variable, *Corruption Obstacle_D*, is a binary variable, we estimate a probit model and use the approach proposed by Ai and Norton (2003) and Norton et al. (2004) to interpret the interaction terms.

We provide figures to represent the sign and significance of the interaction terms. Figures 1 to 3 display the results for the three firm characteristics. Overall, we do not find any significant effects for the interaction of our folklore variable and the three firm characteristics. Our observations fall between the two 90% confidence lines, which indicates that the interaction is not significant at this level. Still, for a few cases for which there are very low predicted probabilities that corruption is indeed an important obstacle for the firm, we have statistically

significant effects. In these cases, folklore seems to matter more for larger and foreign-owned firms in Figures 1 and 2. On the contrary, the effect of folklore is reduced for firms with exporting activities, as shown in Figure 3. Nevertheless, the effects of all three interactions are very low, especially for low probabilities that *Corruption Obstacle_D* is equal to one.

Thus, we conclude that there are no statistically or economically significant effects between folklore and the three tested firm characteristics. Therefore, folklore does not have a differential influence on corruption for larger, foreign-owned firms, or exporting firms.

**4.4 Robustness checks**

We test the robustness of our results in two ways. In all robustness checks, we use our main corruption variable, *Corruption Obstacle_D*.

First, we drop all observations that come from establishments located in cultural regions that belong to the 10% lowest decile of total collected motifs. In this way, we hope to rule out the possibility of unrepresentativeness of our data, which could stem from the little information we have about the folklore of some cultural groups. The results are reported in Table 10. We obtain the same results than in the main estimations. While *All* is significantly negative, we observe a significantly positive coefficient for *Success* and significantly negative coefficients for *Not success* and *Relatively unsuccessful*. Thus, our findings are robust to the exclusion of the observations most likely to be unrepresentative of our population.

Second, we address endogeneity concerns. We use a two-step instrumental variable probit using the procedure from Newey (1987). In the first step, we run a first-stage linear regression with our endogenous variables as our explained variables, and our instruments and controls as our predictors. In a second step, we use the fitted values from our first stage as the values for the potentially endogenous predictors, and then run our probit estimations.

With respect to our instruments, we create a "macro neighbour" for each folklore, based on all cultural groups located in a radius around its centroid. To perform our two-step ivprobit, we must therefore select a radius to create our instrumental variable. We eventually choose a 1000 km radius to best maximize three criteria. First, we want our instrument to be relevant. Therefore, we select a distance that yields the highest possible correlation between our instrumented variable and our instrument. Similarly, we want the lowest association between our instrumental variable and the corruption dependent variable. Finally, we prefer to use a radius that allows us to match all cultural groups in the final sample with at least one neighboring folklore. In this way, we avoid losing observations non-randomly.

We select our radius according to these criteria for our three instruments simultaneously. Thus, we create the *Neighbor Success*, *Neighbor Not success*, and *Neighbor Unclear* instruments. We calculate them using the number of motifs among all neighboring cultural areas that depict a specific outcome, and by dividing it by the total number of collected motifs in the folklore of this "macro neighbor". In other words, we follow the same approach as when we first create our *Success, Not success,* and *Unclear* treatment variables, while considering the created "macro" neighboring folklores as our new cultural groups.

We report the results in Table 11. We obtain statistically significant results that are consistent with our original estimates in terms of the signs of the coefficients. We still find that positive representations of antisocial behaviors in folklore increase the probability that a firm located in the cultural region reports higher amounts of gifts or informal payments. Similarly, the opposite conclusion still holds, as we again find that folklores that more often describe sanctions for misconduct are associated with lower levels of bribes. Overall, our results are robust to potential endogeneity concerns.

## 5. Conclusion

This study delves into the relationship between folklore and corruption, exploring the deep-rooted cultural narratives that shape societal norms, and their implications for economic and institutional outcomes nowadays. Using a dataset that merges firm-level data with subnational data on folklore, we provide evidence that the portrayal of antisocial behavior in oral traditions has a significant influence on corruption.

Our findings reveal that folklore plays a dual role in shaping corruption. Societies where oral traditions depict antisocial behaviors as successful tend to experience higher levels of corruption. Conversely, when folklore emphasizes the negative outcomes for the perpetrators of such behaviors, corruption is less pervasive. These results align with the hypothesis that cultural narratives influence trust and normative behaviors, subsequently shaping the propensity to engage in or resist corrupt practices.

By distinguishing between the demand and supply sides of corruption, we highlight how folklore's influence pervades different dimensions of bribery. Whether firms are the initiators or targets of corruption, the narratives embedded in their cultural milieu significantly affect their actions. Notably, our results show a consistent pattern where the relative success or failure

of antisocial behavior in folklore serves as a predictor for corruption outcomes. Further analysis of firm characteristics, such as size, foreign ownership, and export orientation, yielded no evidence of heterogeneity in folklore's impact. The pervasive effect of cultural narratives remains broadly consistent across firm types.

The implications of our findings are twofold. First, they underscore the importance of integrating cultural narratives into the design and implementation of anti-corruption policies. Addressing corruption requires acknowledging its cultural underpinnings and leveraging narratives that promote trust and accountability. Second, our results shed light on the enduring impact of deep-rooted cultural elements, suggesting that policy interventions must consider the slow-changing nature of these influences to achieve sustainable development goals. Further research could try to understand how modern communication channels and globalization influence the persistence or evolution of traditional narratives could provide new insights into combating corruption in an increasingly interconnected world. To conclude, this study contributes to the growing literature on the role of cultural institutions in economic outcomes, emphasizing the necessity of holistic approaches to combat corruption. By uncovering the links between folklore and corruption, we hope to inspire nuanced strategies that address the cultural dimensions of governance and institutional trust.

## References


Ai, C., Norton, E. (2003). Interaction terms in logit and probit model. Economic Letters, 80, 123-129.

Akerlof, G., Snower, F. (2016). Bread and bullets. Journal of Economic Behavior and Organization, 126: 58–71.

Anderson, C., Tverdova, Y. (2003). Corruption, political allegiances, and attitudes toward government in contemporary democracies. American Journal of Political Science, 47, 91-109.

Berezkin, Y. (2015). Folklore and mythology catalogue: its lay-out and potential for research. The Retrospective Methods Network Newsletter, 10: 58–70.

Chatterjee, I., 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, 3215-3227.

Clausen, B., Kray, A., Nyiri, Z. (2011). Corruption and Confidence in Public Institutions: Evidence from a Global Survey. The World Bank Economic Review, 25, 212-249.

D'Agostino, G., Dunne, J.P., Pieroni, L. (2016). Government Spending, Corruption and Economic Growth. World Development, 84, 190-205.

Dimant, E., Tosato, G. (2017). Causes and Effects of Corruption: What has Past Decade's Empirical Research Taught Us? A Survey. Journal of Economic Surveys, 32, 335-356.

Fisman, R., Svensson, J. (2007). Are corruption and taxation really harmful to growth? Firm level evidence. Journal of Development Economics, 83, 63-75.

Gauthier, B., Goyette, J., Kouamé, W. A.K. (2021). Why do firms pay bribes? Evidence on the demand and supply sides of corruption in developing countries. Journal of Economic Behavior and Organization, 190, 463-479.

Gründler, K., Potrafke, N. (2019). Corruption and economic growth: new empirical evidence. European Journal of Political Economy, 60, 101810.

Guiso, L. Sapienza, P., Zingales, L. (2003). People's Opium? Religion and Economic Attitudes. Journal of Monetary Economics, 50:1, 225-282.

Hakhverdian, A., Mayne, Q. (2012). Institutional trust, education and corruption: A micro-macro interactive approach. Journal of Politics, 74 (3) (2012), 729-750.

Husted, B. (1999). Wealth, Culture, and Corruption. Journal of International Business Studies, 30:2, 339-359.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. (2008). The Economic Consequences of Legal Origins. Journal of Economic Literature, 46:2, 285-332.

Mauro, P. (1995). Corruption and Growth. The Quarterly Journal of Economics, 110:3, 681-712.

Newey, W. (1987). Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables. Journal of Econometrics, 36, 231-250.

North, C. M., Orman, W. H., Gwin, C. R. (2013). Religion, Corruption and the Rule of Law. Journal of Money, Credit and Banking, 45:5, 757-779.

Norton, E., Wang, H., Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. The Stata Journal, 4:2, 154-167.

Osei-Tutu, F., Weill, L. (2024). Regional Favoritism and Access to Credit. Public Choice 200: 1-2, 293-321.

Rand, J., Tarp, F. (2012). Firm-Level Corruption in Vietnam. Economic Development and Cultural Change, 60:3, 571-595.

Rothstein, B. (2000). Trust, Social Dilemmas, and Collective Memories: On the Rise and Decline of the Swedish Model. Journal of Theoretical Politics, 12:4, 477-499.

Serra, D. (2006). Empirical Determinants of Corruption: A Sensitivity Analysis. Public Choice, 126:1, 225-256.

Silva, S.G., Tehrani, J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. Royal Society open science, 3:1, 1-11.

Spolaore, E., Wacziarg, R. (2013). How Deep Are the Roots of Economic Development? Journal of Economic Literature, 51:2, 325-369.

Svensson, J. (2005). Eight Questions about Corruption (2005). Journal of Economic Perspectives, 19:3, 19-42.

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

Treisman, D. (2007). What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research? Annual Review of Political Science, 10, 211-244.

Uberti, L. J. (2022). Corruption and growth: Historical evidence, 1790–2010. Journal of Comparative Economics, 50, 321-349.

Uslaner, E. M. (2004). Trust and Corruption. In: The New Institutional Economics of Corruption, 1-33.

Uslaner, E. M. (2012). Trust and Corruption Revisited: How and Why Trust and Corruption Shape Each Other. Quality & Quantity, 47, 3603-3608.

## Table 1.

## Summary statistics.

This table reports the descriptive statistics for the variables employed in the analysis. All variables are defined in Table A.1 in the Appendix.

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Corruption Obstacle_D | 82922 | 0.2874 | 0.4526 | 0 | 1 |
| Corruption Obstacle | 82922 | 1.4589 | 1.4628 | 1 | 5 |
| Bribe Amount | 62849 | 1.1542 | 5.6586 | 0 | 100 |
| Corruption demand | 5389 | 0.5482 | 0.4977 | 0 | 1 |
| Corruption supply | 32582 | 0.1134 | 0.3171 | 0 | 1 |
| Incorruptible | 5389 | 0.4518 | 0.4977 | 0 | 1 |
| Non-payers | 32582 | 0.8866 | 0.3171 | 0 | 1 |
| All (%) | 82922 | 10.9633 | 5.4149 | 0 | 45.0549 |
| Success (%) | 82922 | 4.4272 | 2.9096 | 0 | 14.2857 |
| Not success (%) | 82922 | 4.1954 | 2.4210 | 0 | 25 |
| Unclear (%) | 82922 | 2.3407 | 2.5375 | 0 | 22.2222 |
| Relatively unsuccessful (%) | 82922 | -0.2317 | 2.9485 | -14.2857 | 17.8571 |
| Age | 82922 | 20.6222 | 18.3236 | 1 | 220 |
| Size (ln) | 82922 | 3.3469 | 1.3695 | 0 | 11.0666 |
| Sole proprietorship | 82922 | 0.2940 | 0.4556 | 0 | 1 |
| Private or non-traded | 82922 | 0.4542 | 0.4979 | 0 | 1 |
| Publicly traded | 82922 | 0.0522 | 0.2223 | 0 | 1 |
| Exporter | 82922 | 0.1821 | 0.3859 | 0 | 1 |
| Foreign-owned | 82922 | 0.0753 | 0.2639 | 0 | 1 |
| Subsidiary | 82922 | 0.1753 | 0.3802 | 0 | 1 |
| Audited | 82922 | 0.5636 | 0.4959 | 0 | 1 |
| Sales growth (%) | 82922 | 60.8092 | 195.9495 | -100 | 1219.973 |
| Manufacturing | 82922 | 0.5465 | 0.4978 | 0 | 1 |
| Retail | 82922 | 0.1733 | 0.3785 | 0 | 1 |
| Experience | 82922 | 19.3333 | 11.3506 | 0 | 75 |
| Growth centroid (%) | 82922 | 62.3599 | 45.0257 | -40.9874 | 414.0373 |
| Electricity centroid | 82922 | 1.2459 | 0.9815 | 0 | 4 |
| Informal credit centroid | 82922 | 8.5369 | 5.9823 | 0 | 26 |
| GDP per capita (ln) | 82922 | 9.2852 | 0.9586 | 6.6797 | 11.6487 |
| Rule of law | 82922 | 4.7719 | 1.5379 | 1.3247 | 9.0693 |
| Fuel exports (%) | 82922 | 17.0911 | 22.4573 | 0 | 99.70458 |
| Protestant (%) | 82922 | 7.9876 | 16.5316 | 0 | 91.8247 |
| Legal origin (UK) | 82922 | 0.3197 | 0.4664 | 0 | 1 |
| Legal origin (Fr) | 82922 | 0.4923 | 0.4999 | 0 | 1 |
| Legal origin (Ger) | 82922 | 0.1592 | 0.3658 | 0 | 1 |

## Table 2.

## Main estimations : Explaining *Corruption Obstacle_D.*

This table reports the results of probit regressions. The dependent variable is *Corruption Obstacle_D*. The variable is equal to 1 if the establishment reports that corruption is either a *Major* or a *Very severe* obstacle to its operations, and 0 otherwise. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | -0.0026*** | | |
| | (0.0003) | | |
| Success (%) | | 0.0021*** | |
| | | (0.0006) | |
| Not success (%) | | -0.0022*** | |
| | | (0.0007) | |
| Unclear (%) | | -0.0098*** | |
| | | (0.0007) | |
| Relatively unsuccessful (%) | | | -0.0020*** |
| | | | (0.0005) |
| Age | 0.0004*** | 0.0004*** | 0.0004*** |
| | (0.0001) | (0.0001) | (0.0001) |
| Size (ln) | -0.0082*** | -0.0084*** | -0.0081*** |
| | (0.0013) | (0.0013) | (0.0013) |
| Sole proprietorship | -0.0300*** | -0.0291*** | -0.0305*** |
| | (0.0044) | (0.0044) | (0.0044) |
| Private or non-traded | -0.0124*** | -0.0131*** | -0.0131*** |
| | (0.0043) | (0.0043) | (0.0043) |
| Publicly traded | -0.0542*** | -0.0561*** | -0.0575*** |
| | (0.0078) | (0.0078) | (0.0078) |
| Exporter | -0.0025 | -0.0027 | -0.0034 |
| | (0.0044) | (0.0044) | (0.0044) |
| Foreign-owned | -0.0266*** | -0.0247*** | -0.0275*** |
| | (0.0061) | (0.0061) | (0.0061) |
| Subsidiary | -0.0042 | -0.0033 | -0.0045 |
| | (0.0041) | (0.0041) | (0.0041) |
| Audited | 0.0292*** | 0.0292*** | 0.0303*** |
| | (0.0033) | (0.0033) | (0.0033) |
| Sales growth (%) | 0.0000* | 0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) |
| Manufacturing | -0.0030 | -0.0028 | -0.0014 |
| | (0.0036) | (0.0036) | (0.0036) |
| Retail | -0.0153*** | -0.0149*** | -0.0149*** |
| | (0.0046) | (0.0046) | (0.0046) |
| Experience | 0.0013*** | 0.0012*** | 0.0013*** |
| | (0.0001) | (0.0001) | (0.0001) |
| Growth centroid (%) | -0.0006*** | -0.0006*** | -0.0006*** |
| | (0.0000) | (0.0000) | (0.0000) |
| Electricity centroid | 0.0473*** | 0.0472*** | 0.0457*** |
| | (0.0019) | (0.0019) | (0.0019) |
| Informal credit centroid | -0.0011*** | -0.0014*** | -0.0013*** |
| | (0.0003) | (0.0003) | (0.0003) |
| GDP per capita (ln) | 0.0241*** | 0.0246*** | 0.0332*** |
| | (0.0032) | (0.0032) | (0.0031) |
| Rule of law | -0.0575*** | -0.0596*** | -0.0606*** |
| | (0.0020) | (0.0020) | (0.0020) |

| | | | |
|---|---|---|---|
| Fuel exports (%) | -0.0003*** | -0.0004*** | -0.0004*** |
| | (0.0001) | (0.0001) | (0.0001) |
| Protestant (%) | -0.0001 | 0.0003 | -0.0004* |
| | (0.0002) | (0.0002) | (0.0002) |
| Legal origin (UK) | 0.1911*** | 0.2176*** | 0.1640*** |
| | (0.0288) | (0.0290) | (0.0289) |
| Legal origin (Fr) | 0.2270*** | 0.2558*** | 0.2001*** |
| | (0.0294) | (0.0296) | (0.0296) |
| Legal origin (Ger) | 0.1382*** | 0.1597*** | 0.1072*** |
| | (0.0289) | (0.0291) | (0.0291) |
| Observations | 82,922 | 82,922 | 82,922 |
| Pseudo $R^2$ | 0.0631 | 0.0815 | 0.1013 |
| Log likelihood | -44699.157 | -44613.254 | -44725.987 |
| Year FE | Yes | Yes | Yes |

## Table 3.

## Main estimations : Explaining *Corruption Obstacle.*

This table reports the results of ordered logit regressions. The dependent variable is *Corruption Obstacle*. All variables are defined in the Appendix. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | -0.0168*** | | |
| | (0.0014) | | |
| Success (%) | | 0.0068*** | |
| | | (0.0026) | |
| Not success (%) | | -0.0097*** | |
| | | (0.0031) | |
| Unclear (%) | | -0.0614*** | |
| | | (0.0030) | |
| Relatively unsuccessful (%) | | | -0.0076*** |
| | | | (0.0022) |
| Age | 0.0024*** | 0.0023*** | 0.0024*** |
| | (0.0004) | (0.0004) | (0.0004) |
| Size (ln) | -0.0081 | -0.0096* | -0.0073 |
| | (0.0057) | (0.0057) | (0.0057) |
| Sole proprietorship | -0.1577*** | -0.1510*** | -0.1579*** |
| | (0.0191) | (0.0191) | (0.0191) |
| Private or non-traded | -0.1224*** | -0.1246*** | -0.1249*** |
| | (0.0189) | (0.0189) | (0.0189) |
| Publicly traded | -0.2793*** | -0.2875*** | -0.2960*** |
| | (0.0331) | (0.0332) | (0.0331) |
| Exporter | 0.0154 | 0.0147 | 0.0098 |
| | (0.0188) | (0.0188) | (0.0188) |
| Foreign-owned | -0.1558*** | -0.1459*** | -0.1619*** |
| | (0.0262) | (0.0262) | (0.0262) |
| Subsidiary | -0.0112 | -0.0080 | -0.0141 |
| | (0.0179) | (0.0179) | (0.0179) |
| Audited | 0.2061*** | 0.2075*** | 0.2135*** |
| | (0.0144) | (0.0144) | (0.0144) |
| Sales growth (%) | 0.0000 | 0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) |
| Manufacturing | -0.0304* | -0.0292* | -0.0198 |
| | (0.0157) | (0.0157) | (0.0157) |
| Retail | -0.0963*** | -0.0943*** | -0.0924*** |
| | (0.0202) | (0.0202) | (0.0202) |
| Experience | 0.0046*** | 0.0047*** | 0.0047*** |
| | (0.0006) | (0.0006) | (0.0006) |
| Growth centroid (%) | -0.0039*** | -0.0038*** | -0.0041*** |
| | (0.0002) | (0.0002) | (0.0002) |
| Electricity centroid | 0.2674*** | 0.2688*** | 0.2581*** |
| | (0.0086) | (0.0087) | (0.0086) |
| Informal credit centroid | -0.0096*** | -0.0113*** | -0.0107*** |
| | (0.0015) | (0.0015) | (0.0015) |
| GDP per capita (ln) | 0.1201*** | 0.1217*** | 0.1763*** |
| | (0.0143) | (0.0144) | (0.0136) |
| Rule of law | -0.3088*** | -0.3207*** | -0.3265*** |
| | (0.0089) | (0.0090) | (0.0088) |
| Fuel exports (%) | -0.0020*** | -0.0021*** | -0.0024*** |
| | (0.0004) | (0.0004) | (0.0004) |

| | | | |
|---|---|---|---|
| Protestant (%) | 0.0006 | 0.0032*** | -0.0009 |
| | (0.0009) | (0.0010) | (0.0009) |
| Legal origin (UK) | 1.1893*** | 1.3818*** | 1.0457*** |
| | (0.1083) | (0.1096) | (0.1087) |
| Legal origin (Fr) | 1.3233*** | 1.5282*** | 1.1817*** |
| | (0.1112) | (0.1125) | (0.1115) |
| Legal origin (Ger) | 1.0781*** | 1.2377*** | 0.9139*** |
| | (0.1084) | (0.1095) | (0.1087) |
| Observations | 82,922 | 82,922 | 82,922 |
| Pseudo $R^2$ | 0.0631 | 0.0643 | 0.0625 |
| Log likelihood | -116827.67 | -116679.87 | -116896.74 |
| Year FE | Yes | Yes | Yes |

## Table 4.

## Marginal effects: Explaining *Corruption Obstacle*.

This table provides the marginal effects for the results of ordered logit regressions. The dependent variable is *Corruption Obstacle*. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| Corruption outcome | No | Minor | Moderate | Major | Very severe |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| (a) | | | | | |
| All (%) | 0.0034*** | 0.0002*** | -0.0005*** | -0.0013*** | -0.0018*** |
| | (0.0003) | (0.0000) | (0.000) | (0.0001) | (0.0001) |
| (b) | | | | | |
| Success (%) | -0.0014*** | -0.0001*** | 0.0002*** | 0.0005*** | 0.0007*** |
| | (0.0005) | (0.0000) | (0.0001) | (0.0002) | (0.0003) |
| Not success (%) | 0.0020*** | 0.0001*** | -0.0003*** | -0.0007*** | -0.0010*** |
| | (0.0006) | (0.0000) | (0.0001) | (0.0002) | (0.0003) |
| Unclear (%) | 0.0125*** | 0.0006*** | -0.0019*** | -000047*** | -0.0065*** |
| | (0.0006) | (0.0000) | (0.0001) | (0.0002) | (0.0003) |
| (c) | | | | | |
| Relatively unsuccessful (%) | 0.0016*** | 0.0001*** | -0.0002*** | -0.0006*** | -0.0008*** |
| | (0.0005) | (0.0000) | (0.0001) | (0.0002) | (0.0002) |
| Firm-level controls | Yes | Yes | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |

## Table 5.

## Main estimations: Explaining *Bribe Amount*.

This table reports the results of OLS regressions. The dependent variable is *Bribe Amount*. The variable is the percentage of total annual sales that an establishment similar to that of the respondent pays in informal payments or gifts to public officials "to get things done", according to the respondent. All variables are defined in the Appendix. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | 0.0187*** | | |
| | (0.0048) | | |
| Success (%) | | 0.0815*** | |
| | | (0.0093) | |
| Not success (%) | | -0.0797*** | |
| | | (0.0108) | |
| Unclear (%) | | 0.0434*** | |
| | | (0.0104) | |
| Relatively unsuccessful (%) | | | -0.0811*** |
| | | | (0.0080) |
| Age | -0.0022 | -0.0020 | -0.0021 |
| | (0.0014) | (0.0014) | (0.0014) |
| Size (ln) | -0.0332* | -0.0331* | -0.0347* |
| | (0.0198) | (0.0198) | (0.0198) |
| Sole proprietorship | -0.0588 | -0.0959 | -0.0897 |
| | (0.0669) | (0.0669) | (0.0669) |
| Private or non-traded | 0.0502 | 0.0014 | 0.0060 |
| | (0.0640) | (0.0642) | (0.0642) |
| Publicly traded | 0.9637*** | 0.9025*** | 0.9145*** |
| | (0.1156) | (0.1157) | (0.1156) |
| Exporter | 0.2724*** | 0.2695*** | 0.2710*** |
| | (0.0640) | (0.0640) | (0.0640) |
| Foreign-owned | -0.2715*** | -0.2836*** | -0.2688*** |
| | (0.0899) | (0.0899) | (0.0898) |
| Subsidiary | 0.2474*** | 0.2589*** | 0.2623*** |
| | (0.0616) | (0.0616) | (0.0616) |
| Audited | -0.1976*** | -0.2029*** | -0.2091*** |
| | (0.0496) | (0.0496) | (0.0495) |
| Sales growth (%) | 0.0010*** | 0.0010*** | 0.0010*** |
| | (0.0001) | (0.0001) | (0.0001) |
| Manufacturing | -0.3529*** | -0.3525*** | -0.3577*** |
| | (0.0537) | (0.0536) | (0.0536) |
| Retail | -0.3691*** | -0.3873*** | -0.3889*** |
| | (0.0697) | (0.0697) | (0.0696) |
| Experience | -0.0012 | -0.0013 | -0.0015 |
| | (0.0022) | (0.0022) | (0.0022) |
| Growth centroid (%) | 0.0024*** | 0.0022*** | 0.0023*** |
| | (0.0007) | (0.0007) | (0.0007) |
| Electricity centroid | 0.2306*** | 0.2000*** | 0.2093*** |
| | (0.0300) | (0.0301) | (0.0300) |
| Informal credit centroid | 0.0240*** | 0.0267*** | 0.0269*** |
| | (0.0055) | (0.0055) | (0.0055) |
| GDP per capita (ln) | -0.2437*** | -0.2626*** | -0.2994*** |
| | (0.0494) | (0.0494) | (0.0466) |
| Rule of law | -0.1277*** | -0.1173*** | -0.1136*** |
| | (0.0308) | (0.0309) | (0.0304) |

| | | | |
|---|---|---|---|
| Fuel exports (%) | 0.0071*** | 0.0072*** | 0.0074*** |
| | (0.0013) | (0.0013) | (0.0013) |
| Protestant (%) | 0.0060* | 0.0020 | 0.0048 |
| | (0.0031) | (0.0031) | (0.0031) |
| Legal origin (UK) | 0.6858** | 0.2795 | 0.5168* |
| | (0.2822) | (0.2879) | (0.2820) |
| Legal origin (Fr) | 0.2996 | -0.1076 | 0.1323 |
| | (0.2899) | (0.2958) | (0.2899) |
| Legal origin (Ger) | 0.4031 | 0.0008 | 0.2179 |
| | (0.2816) | (0.2863) | -0.0021 |
| Constant | 2.6837*** | 3.3406*** | 3.5518*** |
| | (0.6397) | (0.6434) | (0.6229) |
| | | | |
| Observations | 62,849 | 62,849 | 62,849 |
| $R^2$ | 0.0339 | 0.0355 | 0.0352 |
| Year FE | Yes | Yes | Yes |

## Table 6.

## Explaining *Corruption demand*.

This table reports the results of probit regressions. The dependent variable is *Corruption demand*. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | 0.0046*** | | |
| | (0.0011) | | |
| Success (%) | | 0.0092*** | |
| | | (0.0026) | |
| Not success (%) | | -0.0030 | |
| | | (0.0027) | |
| Unclear (%) | | 0.0070*** | |
| | | (0.0023) | |
| Relatively unsuccessful (%) | | | -0.0066*** |
| | | | (0.0022) |
| Observations | 5,389 | 5,389 | 5,389 |
| Pseudo $R^2$ | 0.1416 | 0.1429 | 0.1407 |
| Log likelihood | -3185.0498 | -3180.1593 | -3188.3925 |
| Firm-level controls | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |

## Table 7.

## Explaining *Corruption supply*.

This table reports the results of probit regressions. The dependent variable is *Corruption supply*. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | 0.0027*** | | |
| | (0.0003) | | |
| Success (%) | | 0.0051*** | |
| | | (0.0007) | |
| Not success (%) | | -0.0003 | |
| | | (0.0008) | |
| Unclear (%) | | 0.0027*** | |
| | | (0.0007) | |
| Relatively unsuccessful (%) | | | -0.0033*** |
| | | | (0.0006) |
| Observations | 32,582 | 32,582 | 32,582 |
| Pseudo $R^2$ | 0.1232 | 0.1242 | 0.1218 |
| Log likelihood | -10100.408 | -10089.741 | -10117.329 |
| Firm-level controls | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |

## Table 8.

## Explaining Incorruptible.

This table reports the results of probit regressions. The dependent variable is *Incorruptible*. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | -0.0046*** | | |
| | (0.0011) | | |
| Success (%) | | -0.0092*** | |
| | | (0.0026) | |
| Not success (%) | | 0.0030 | |
| | | (0.0027) | |
| Unclear (%) | | -0.0070*** | |
| | | (0.0023) | |
| Relatively unsuccessful (%) | | | 0.0066*** |
| | | | (0.0022) |
| Observations | 5,389 | 5,389 | 5,389 |
| Pseudo $R^2$ | 0.1416 | 0.1429 | 0.1407 |
| Log likelihood | -3185.0498 | -3180.1593 | -3188.3925 |
| Firm-level controls | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |

## Table 9.

## Explaining *Non-payers*.

This table reports the results of probit regressions. The dependent variable is *Non-payers*. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | -0.0027*** | | |
| | (0.0003) | | |
| Success (%) | | -0.0051*** | |
| | | (0.0007) | |
| Not success (%) | | 0.0003 | |
| | | (0.0008) | |
| Unclear (%) | | -0.0027*** | |
| | | (0.0007) | |
| Relatively unsuccessful (%) | | | 0.0033*** |
| | | | (0.0006) |
| Observations | 32,582 | 32,582 | 32,582 |
| Pseudo $R^2$ | 0.1232 | 0.1242 | 0.1218 |
| Log likelihood | -10100.408 | -10089.741 | -10117.329 |
| Firm-level controls | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |

**Table 10.**
**Robustness check: estimations without the first decile of folklores (total number of motifs).**

This table reports the results of probit regressions excluding the 10% lowest decile of total collected motifs. The dependent variable is *Corruption Obstacle_D*. All variables are defined in the Appendix. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | -0.0030*** | | |
| | (0.0004) | | |
| Success (%) | | 0.0027*** | |
| | | (0.0007) | |
| Not success (%) | | -0.0020** | |
| | | (0.0008) | |
| Unclear (%) | | -0.0161*** | |
| | | (0.0009) | |
| Relatively unsuccessful (%) | | | -0.0018*** |
| | | | (0.0006) |
| Observations | 78,999 | 78,999 | 78,999 |
| Pseudo $R^2$ | 0.0989 | 0.1019 | 0.0982 |
| Firm-level controls | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |

## Table 11.

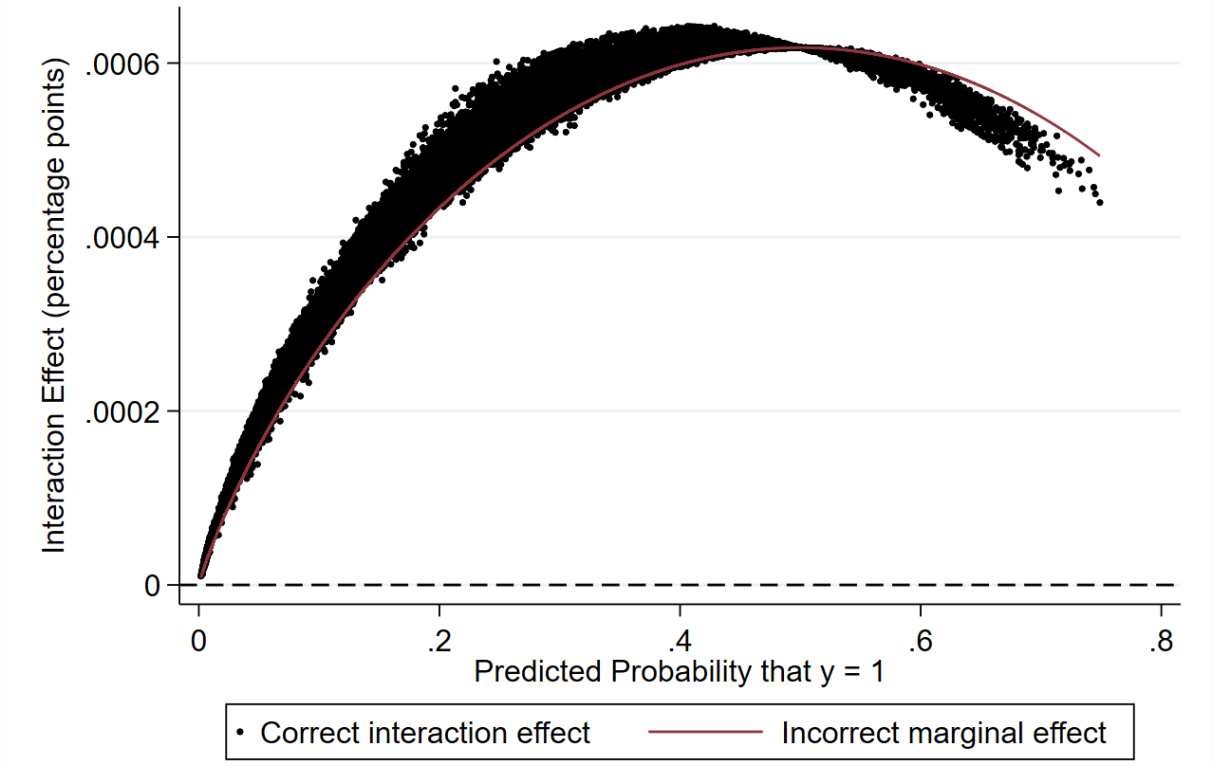
## Robustness check: testing for endogeneity.

This table reports the results of probit regressions with instrumental variables (IV) following Newey's procedure for the *Corruption Obstacle_D* dependent variable. The instruments are constructed using data from the neighboring folklores in a radius of 1000 km. The Wald Test compares the instrumented model and non-instrumented model. Under the null hypothesis, both models provide similar results. Estimated marginal effects are reported and standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

| | (1) | (2) | (3) |
|---|---|---|---|
| All (%) | -0.0313*** | | |
| | (0.0024) | | |
| Success (%) | | 0.0427** | |
| | | (0.0207) | |
| Not success (%) | | -0.0557*** | |
| | | (0.0212) | |
| Unclear (%) | | -0.1821*** | |
| | | (0.0081) | |
| Relatively unsuccessful (%) | | | 0.2691*** |
| | | | (0.0194) |
| Observations | 82,922 | 82,922 | 82,922 |
| Firm-level controls | Yes | Yes | Yes |
| Centroid-level controls | Yes | Yes | Yes |
| Country-level controls | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| Wald test | 108.83*** | 562.15*** | 271.92*** |

**Figure 1.**
**Interaction effects (Size).**

These graphics show correct interaction effects after probit, as well as the corresponding z-statistics and red reference lines for statistical significance, for different values of the dependent variable *Corruption Obstacle_D*. The interacted variables are *Relatively unsuccessful* and *Size*. All variables are defined in the Appendix.



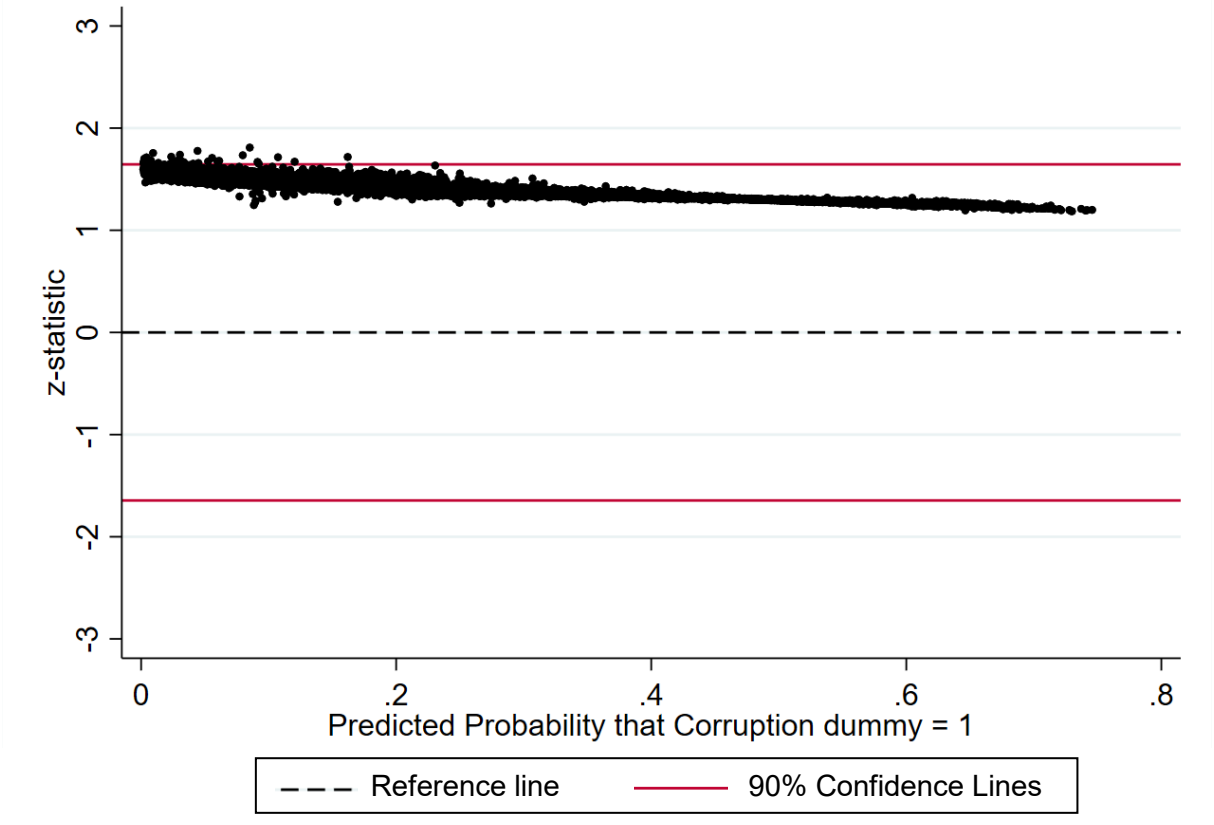
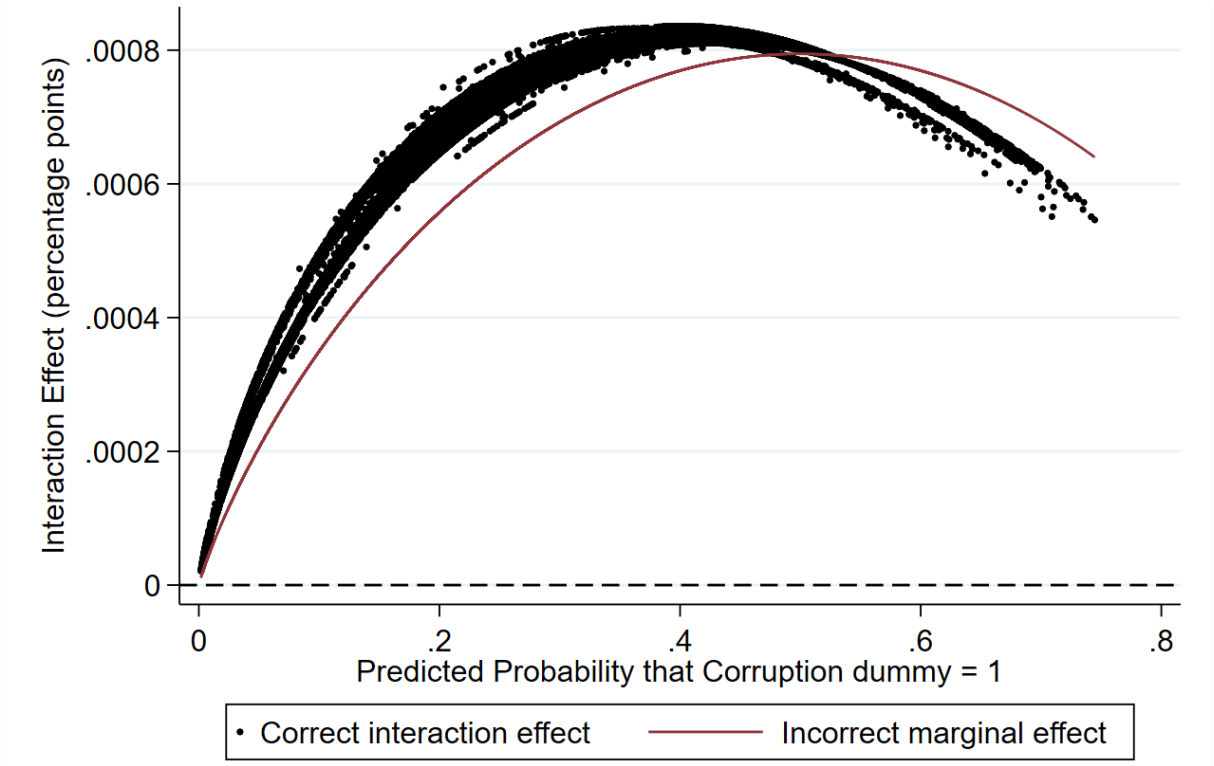



**Figure 2.**
**Interaction effects (Foreign owned).**

These graphics show correct interaction effects after probit, as well as the corresponding z-statistics and red reference lines for statistical significance, for different values of the dependent variable *Corruption Obstacle_D*. The interacted variables are *Relatively unsuccessful* and *Foreign owned*. All variables are defined in the Appendix.



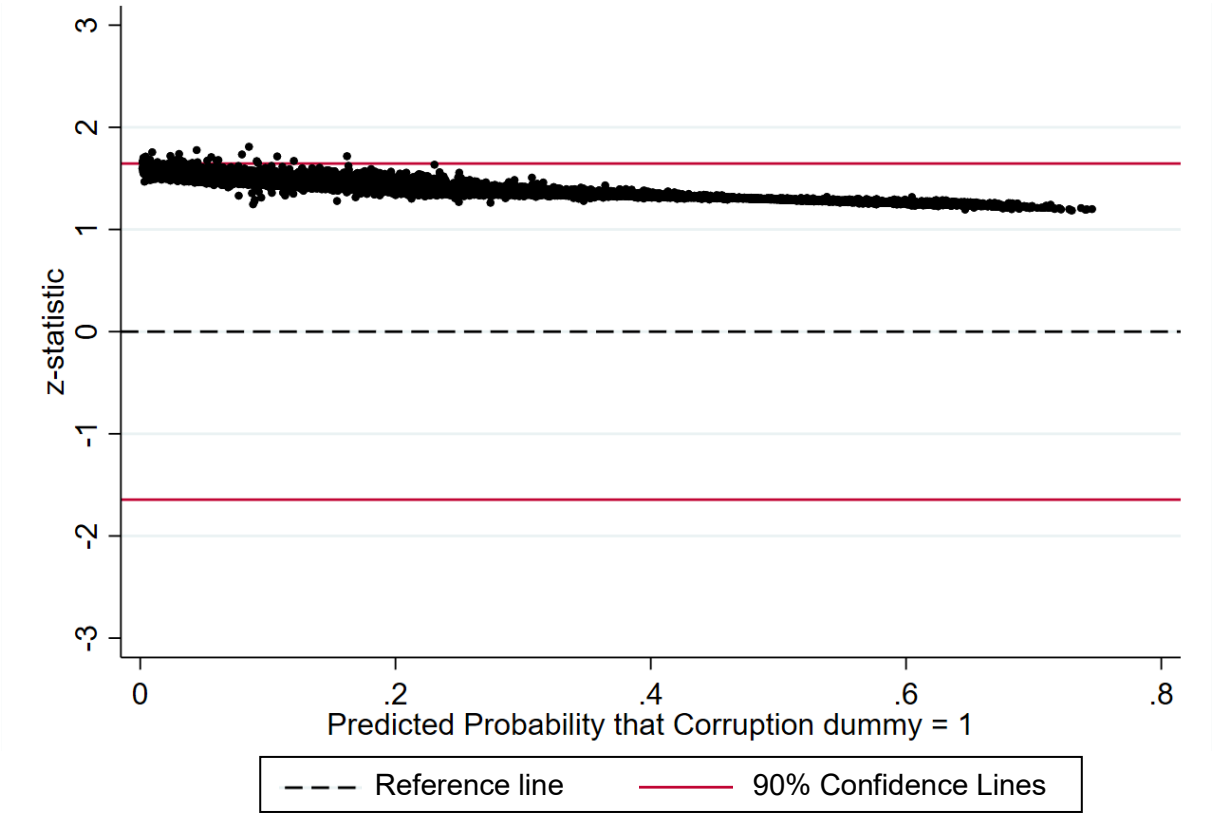
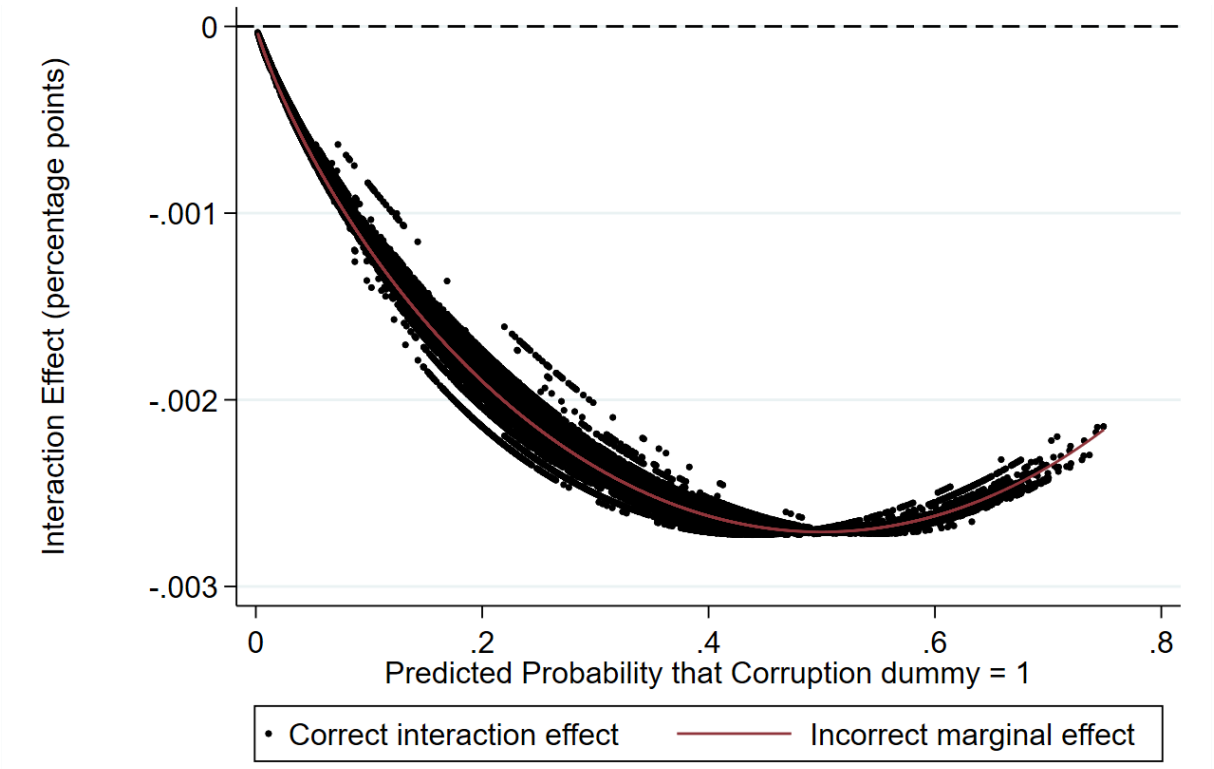
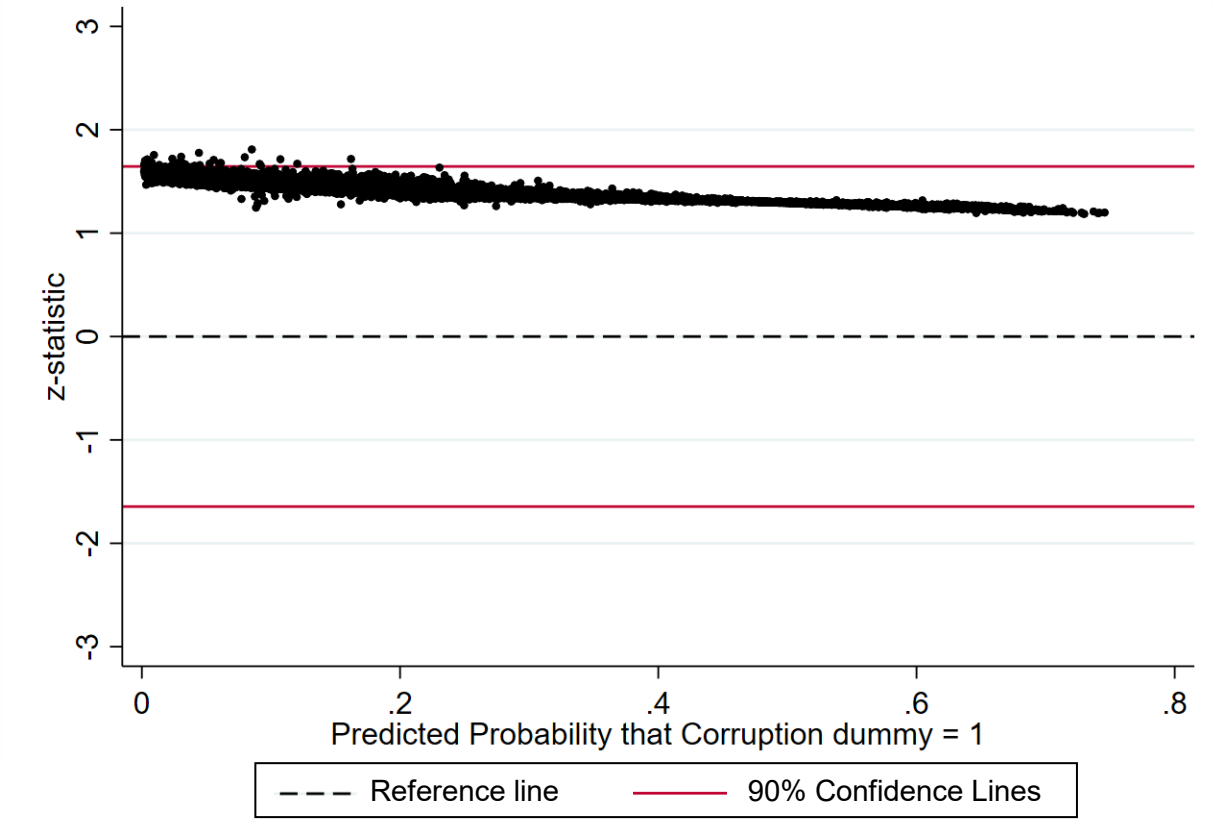



**Figure 3.**
**Interaction effects (Exporter).**

These graphics show correct interaction effects after probit, as well as the corresponding z-statistics and red reference lines for statistical significance, for different values of the dependent variable *Corruption Obstacle_D*. The interacted variables are *Relatively successful*, and *Exporter*. All variables are defined in the Appendix.



# Appendix

**Table A.1**

**Definitions and sources of variables.**

| Variable | Definition and source |
|---|---|
| *Dependent variables* | |
| Corruption Obstacle_D | Dummy=1 if the establishment sees corruption as *No* or a *Minor* obstacle, =0 otherwise. Source: WBES. |
| Corruption Obstacle | Ordinal=1-5 if the establishment sees corruption as *No* (1), a *Minor* (2), *Moderate* (3), *Major* (4), *Very severe* (5) obstacle. Source: WBES. |
| Bribe Amount | Percentage of total annual sales that similar establishments pay in informal payments or gifts to public officials, according to the respondent. Source: WBES. |
| Corruption demand | Dummy=1 if the establishment of the respondent has accepted at least one type of bribe request, =0 otherwise. Source: WBES. |
| Corruption supply | Dummy=1 if the establishment of the respondent has paid at least one type of bribe without request, =0 otherwise. Source: WBES. |
| Incorruptible | Dummy=1 if the respondent has refused all bribe requests, =0 otherwise. Source: WBES. |
| Non-payers | Dummy=1 if the respondent has not paid any type of bribe, and has not been a target for a corruption request, =0 otherwise. Source: WBES. |
| *Folklore variables* | |
| All | Share (%) of motifs that depict an antisocial behavior. Source: Berezkin (2015). |
| Success | Share (%) of motifs that depict a successful outcome for an antisocial behavior (winsorized). Source: Berezkin (2015). |
| Not success | Share (%) of motifs that depict an unsuccessful outcome for an antisocial behavior. Source: Berezkin (2015). |
| Unclear | Share (%) of motifs for which the outcome is unclear. Source: Berezkin (2015). |
| Relatively unsuccessful | Relative unsuccessfulness (%) in a culture, computed as Relative = Not success – Success. Source: Berezkin (2015). |
| *Firm variables* | |
| Age | Age of the firm (number of years since incorporation). Source: WBES. |
| Size (ln) | Natural logarithm of the number of full-time permanent employees. Source: WBES. |
| Sole proprietorship | Dummy=1 if the firm is a sole proprietorship, =0 otherwise. Source: WBES. |
| Private or non-traded | Dummy=1 if is private or non-traded, =0 otherwise. Source: WBES. |
| Publicly traded | Dummy=1 if is publicly traded, =0 otherwise. Source: WBES. |
| Audited | Dummy=1 if the firm's financial statements were checked and certified by an external auditor, =0 zero otherwise. Source: WBES. |
| Experience | Top manager's number of years of experience in the sector. Source: WBES. |
| Foreign-owned | Dummy=1 if at least 50% of the firm's ownership is held by foreigners, =0 otherwise. Source: WBES. |
| Exporter | Dummy =1 if at least 10% of the firm's annual sales is derived from direct exports, =0 otherwise. Source: WBES. |
| Subsidiary | Dummy=1 if a firm is part of a large group, =0 otherwise. Source: WBES. |
| Sales growth | Sales growth rate of the firm over a three-year period (winsorized). Source: WBES. |
| Manufacturing | Dummy=1 if the firm's sector is manufacturing. Source: WBES. |

| | |
|---|---|
| Retail | Dummy=1 if the firm's sector is retail. Source: WBES. |
| *Centroid variables* | |
| Growth centroid | Mean of the sales growth rates of firms over a three-year period (winsorized). Source: WBES. |
| Electricity centroid | Median of the categorical variable of the degree to which firms perceive access to electricity as an obstacle, ranging from 0 ("no obstacle") to 4 ("very severe obstacle"). Source: WBES. |
| Informal credit centroid | Mean share (%) of firms whose share of working capital or fixed assets funded by moneylenders, friends, or relatives is greater than 0 (winsorized). Source: WBES. |
| *Country variables* | |
| Fuel exports | Share (%) of merchandise exports. Source: World Bank. |
| GDP per capita (ln) | Natural logarithm of GDP per capita, PPP (constant 2017 international dollar). Source: World Bank. |
| Legal origin (UK) | Dummy=1, if the legal origin of the country is British, =0 otherwise. Source: La Porta et al. (2008). |
| Legal origin (Fr) | Dummy=1, if the legal origin of the country is French, =0 otherwise. Source: La Porta et al. (2008). |
| Legal origin (Ger) | Dummy=1, if the legal origin of the country is German, =0 otherwise. Source: La Porta et al. (2008). |
| Protestants | Mean share (%) of protestant population over the 1945-2010 period. Source: World Religion Project. |
| Rule of law | Score on the degree to which agents have confidence in and abide by the rules of society, rescaled to a 0 to 10 range, from the worst to the best state of rule of law Source: World Governance Indicators. |