BANCA D'ITALIA

EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Corruption red flags in public procurement: new evidence from Italian calls for tenders

by Francesco Decarolis and Cristina Giorgiantonio

February 2020

Number 544

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Corruption red flags in public procurement: new evidence from Italian calls for tenders

by Francesco Decarolis and Cristina Giorgiantonio

Number 544 – February 2020

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

CORRUPTION RED FLAGS IN PUBLIC PROCUREMENT: NEW EVIDENCE FROM ITALIAN CALLS FOR TENDERS

by Francesco Decarolis* and Cristina Giorgiantonio†

Abstract

This paper contributes to the analysis of quantitative indicators (i.e., red flags or screens) to detect corruption in public procurement. Expanding the set of commonly discussed indicators in the literature to new ones derived from the operating practices of police forces and the judiciary, this paper verifies the presence of these red flags in a sample of Italian awarding procedures for roadwork contracts in the period 2009-2015. Then, it validates the efficacy of the indicators through measures of direct corruption risks (judiciary cases and police investigations for corruption-related crimes) and indirect corruption risks (delays and cost overruns). From a policy perspective, our analysis shows that the most effective red flags in detecting corruption risks are those related to discretionary mechanisms for selecting private contractors (such as the most economically advantageous offer or negotiated procedures), compliance with the minimum time limit for the submission of tenders and subcontracting. Moreover, our analysis suggests that greater standardization in the call for tender documents can contribute to reducing corruption risks. From a methodological point of view, the paper highlights the relevance of prediction approaches based on machine learning methods (especially the random forests algorithm) for validating a large set of indicators.

JEL Classification: D44, D47, H57, R42.
Keywords: public procurement, corruption, red flags.
DOI: 10.32057/0.QEF.2020.544

Contents

* Bocconi University, Department of Economics and IGIER.
† Bank of Italy, Consumer Protection and Anti-money Laundering Directorate, Financial Education Division.

1 Introduction*

Corruption is commonly defined as abuse of public power in order to obtain private benefits. It is widely believed to entail large economic and social costs. Its importance for economic growth has been of policy interest to governments, entrepreneurs, and investors around the world, with the IMF estimating that corruption costs exceed 2% of world's GDP IMF (2016).

The economic literature has so far explored several channels through which corruption may affect economic growth and allocative efficiency. Some authors argue that corruption acts as a sand in the wheel and hampers economic growth, through channels such as barriers to entrepreneurship and firm investment, limited access to finance, and higher transaction costs (Shleifer and Vishny (1993), Mauro (1995), Svensson (2005)), resulting in resource misallocation across firms Hsieh and Klenow (2009) and within firms (Murphy, Shleifer and Vishny (1991), Dal Bó and Rossi (2007), Colonnelli and Prem (2017)). Others highlight its effects in terms of distortion of human capital accumulation Mo (2001). Others focus on the activities of the public sector, documenting relationships between corruption and inefficiencies in the composition of government expenditure Mauro (1998), lower productivity of public investments Del Monte and Papagni (2001), higher shares of goods and services procured by the public administration on noncompetitive markets Hessami (2014), worse selection and misallocation of public employees Mocetti and Orlando (2017).[1]

A particularly critical area for corruption is public procurement (Golden and Picci (2005), ANAC (2015)): from construction to education and healthcare to innovation, nearly all activities where the public sector is involved imply the need to procure goods, services or works, but the disconnection between who procures these contracts and who pays for them creates scope for corruption. The vulnerability of this key area – representing 15% of the EU-wide GDP – to corruption is a key motive behind continuous efforts to monitor, measure and fight corruption in public procurement. Moreover, in public as well as in private procurement, corruption might be the necessary evil that comes together with empowering agents to use their discretion: curbing corruption through rigid procurement rules might impose significant efficiency losses, even higher than those resulting from corruption itself.[2]

In this study, we analyze how different features of public procurement are associated with the risk of corruption. Crucially, we do not seek to evaluate the trade-offs created by, for instance, bolstering flexibility in the choice of the most reliable contractors relative to the

*We are grateful for the comments received by Silvia Giacomelli, Sauro Mocetti, Tommaso Orlando and Paolo Sestito, and by seminar participants at the Bank of Italy and Bocconi University where preliminary versions of this paper were presented. The opinions expressed in this research work remain, in all cases, the exclusive responsibility of the authors and do not reflect those of their respective Institutions. Decarolis is grateful for financial support of the European Research Council (ERC-2015-StG - 679217-REPCOR).

[1]However, it should be noted that these findings are not univocal. It is also argued that in presence of red-tape, corruption may act as a grease in the wheel of bureaucracy, reducing costs of doing business and improving performance (Leff (1964), Huntington (2006), Dreher and Gassebner (2013), Bologna, Ross et al. (2015)), and potentially leading to efficient allocation as the most productive firms may have the highest willingness to pay (Lui (1985), Shleifer and Vishny (1994), Weaver (2016)).

[2]This is a well known theoretical fact, (Manelli and Vincent, 1995) and (Calzolari and Spagnolo, 2009), which has been shown by recent work of Decarolis et al. (2019) to characterize Italian public procurement.

risk of abusing discretion. We, instead, focus exclusively on the risk of corruption. Even more specifically, we take an ex ante perspective and ask which features of the tendering design are best capable of predicting the risk of follow on corruption at the contract awarding stage. But corruption by its inherent nature is difficult to measure. It is a secretive act, even more so than most crimes as it does not have a clearly identified victim with an interest in reporting it. Detailed data are, hence, often unavailable and researchers have mostly relied on survey, voluntary disclosure reports or other indexes, which could be biased. For these reasons corruption measurement remains an area of active theoretical and empirical debate.

Using novel data concerning the procurement of public works[3], in this paper, we contribute to this debate by providing new evidence on so-called "red flags", i.e. indicators of potential corruption risk.[4] These are identifiable features of the calls for tenders that are plausibly associated with corrupt practices. The first part of this study presents these indicators, some of which are new to the literature. In fact, some of them are based on operating practices (e.g., investigations by the sector Authority (Italian Anticorruption Authorithy - ANAC) or judgments). We organize their discussion on the basis of the type of activity that they involve and whether they are directly available to the Anticorruption Authority or not. This entity monitors corruption risk, but does not systematically collect information on all indicators. We thus indicate as *oblivious indicators*, those that can be derived from looking at the individual calls for tender, as we do, but that are not otherwise available to the Anticorruption Authority.

Next, we introduce our outcome measures of corruption risk. Reliable measures of corruption are clearly hard to get, but crucial for indicators' validation. We use a novel measure of corruption developed by Decarolis et al. (2019) that is based on police investigations. As explained in greater detail in their study, it exploits uniquely detailed data on firm-level corruption risk: an indicator variable measuring for each firm winning a contract whether any of its owners or top managers have been object of a police investigation for corruption-related crimes. We show quantitatively the usefulness of this new measure by comparing it to four alternative measures that we also collect and that are more typically used in the literature: two based on judiciary cases and two based on economic outcomes (delays and cost overruns).

We then assess the prediction capability of the various indicators using common ML algorithms: LASSO, ridge regression and random forest. We analyze their performance both in a large dataset where, however, only a small set of indicators is available and in a different, partially overlapping, sample where many more indicators are observed. The first sample comes from the Anticorruption Authority and contains the universe of contract awards that this Authority monitors (13 thousand contracts in our sample period). For this sample, the data allows us to observe 12 main red flags. The other sample is smaller, containing nearly 3.5 thousand contract awards. But for each of this contracts we got access to the call for tenders documents. Through a combination of both human and machine

[3]Specifically, contracts for the building and maintenance of roads and highways in Italy, where corruption is considered a very relevant phenomenon (GRECO (2012), PCM (2010), European Comission (2014)).

[4]See Fazekas, Toth and King (2016) for a recent review of the use of red flags to detect corruption in public procurement.

textual analysis of the call for tenders, we were obtain a larger set of indicators: the previous 12 indicators, plus 20 indicators more. These includes the oblivious indicators, known by participants to be not systematically monitored.

Our main findings are as follows. First, when we use the smaller set of red flags available to the Anticorruption Authority, we find a systematic association with corruption risk of some, highly policy relevant tools, like the use of an awarding criterion based on multiple parameters (i.e., scoring auction or MEAT). This is the awarding criterion that, starting with the EU Procurement Directive 24/2014 has become the default system in the EU, replacing the previous system which had price competition as the default mechanism.[5] Second, some indicators that the common wisdom suggests to be positively associated with corruption are clearly negatively associated with it. This is, for instance, the case of the call for tender invoking special procedures due to an "urgency" situation or a variable counting the number of days in which the call for tenders is open to firms to submit their bids. We argue that is precisely for its known corruption-related risks that such an obvious and scrutinized tool is not adopted by corrupted agents that are aware of being monitored. This line of reasoning leads us to explore the contribution to prediction of the oblivious indicators. Our third result is that, once the broader set of indicators is included, the model's accuracy - as measured by the mean squared error criterion - improves, but this is improvement is mostly limited to the random forest model and not to the LASSO or Ridge models. We argue that, as found in other studies, this likely to the greater functional form flexibility of the random forest relative to the other two alternative methods considered. Fourth, under the random forest model and exploiting the larger set of indicators, the ranking of the most relevant red flags somewhat differs from what found with the alternative methods and specifications: for instance, while the scoring rule remains an important predictor, features like the completeness of the documentation (measured in terms of the total number of document pages and words) that are only marginally relevant under LASSO and RIDGE become among the most important predictors. We conclude by noting that a larger set of indicators is not only important because some of them could be individually relevant red flags, but also because they allow the prediction approaches of ML, and especially the random forest algorithm, to deliver its potential.

The remainder of the paper is organized as follows: section 2 reviews the literature; section 3 describes the institutional framework and presents our red flags; section 4 illustrates the samples of public work contracts and provides descriptive evidence on the indicators; section 5 discusses the outcome measures and the strategy for their empirical analysis; section 6 presents the findings; section 7 concludes, highlighting some policy considerations.

[5] Clearly, scoring auctions are very valuable tools to enhance a proper trade-off of the multiple cost and quality elements that can make a procurement process cost effective. But, as stated earlier, this study only looks at corruption predictors, without attempting to evaluate the complex trade-off. See Decarolis et al. (2019) for an empirical exploration of these trade-offs posed by scoring rule auctions and, more broadly, discretion in procurement. See also Coviello, Guglielmo and Spagnolo (2017) on the analogous trade-offs in the case of restricted vs. open auctions.

2 Literature

Although understanding the effects of corruption on the (mis)allocation of resources is at the heart of the economic and political debate, answering the question of how we can measure the extent of corruption presents a major challenge.

A first strand of the literature to which this study contributes is the vast literature on corruption measurement. This is well known to be a complex and elusive task.[6] We present a new measure on firms at risk of corruption from a related study of Decarolis et al. (2019) and compare it with alternative measures that we collect on both judicial and economic outcomes.

Differently from subjective measures often used in the corruption literature such as corruption indexes[7], our measure (like judicial data) has the merit of being an objective way for measuring corruption. However, it is superior to judicial data measures as they only measure the emerged corruption (e.g., convictions include only acts of corruption in which the individual was actually caught and convicted).

The extent to which corruption crimes are successfully prosecuted by the judiciary depends on many factors such as the enforcement level. These considerations explain why they can be rarely used as a measure of corruption.[8] In the light of the limitations of the judiciary data, the most recent economic literature has moved towards developing new and more objective tools to assess the extent of corruption. Some have made use of direct measurements of outcomes (Di Tella and Schargrodsky (2003), Bandiera, Prat and Valletti (2009), Golden and Picci (2005)). But we show that, within our context, the most typically used indirect measures of corruption (delays and cost overruns) are not usfeul proxies for corruption.[9]

Another part of the literature on corruption to which our study is closely connected is that on red flags. Especially in public procurement, researchers have discussed indicators or red flags that point to corruption, which may eventually be used to develop an indicator-based risk assessment (Di Nicola and McCallister (2006)). The potential usefulness of red flags is that corrupt activities require certain forms of economic behaviour (e.g., low bid participation

[6]See Rizzica and Tonello (2015) and Brodi, Occhilupo and Tonello (2012) for a comprehensive review of the main problems.

[7]Most times, such indexes are based on perceptions of the phenomena; other indexes are based on descriptions of broad aspects of governance, and as such are more tenuously linked with the corruption phenomenon. Indicators of this type range from The Economist Intelligence Unit's Business International Indicators, to the World Bank Governance Indicators, the Transparency International Corruption Perceptions Index, the Global Corruption Barometer, and the European Commission Eurobarometer. See Fazekas, Toth and King (2016) and Rizzica and Tonello (2015) for an extensive review of these indicators and their use.

[8]See, for example, Mitchell and Campbell (2009) who have used federal corruption conviction rates in the U.S. as a measure of corruption. For another recent example see Schulze, Sjahrir and Zakharov (2016).

[9]More recently, others scholars have used data from random audits of governmental processes in Brazil in order to construct new measures of political corruption in local governments (Ferraz and Finan (2011), Brollo et al. (2013), Brollo and Troiano (2016)) and corruption in public procurement contracts Colonnelli and Prem (2017). While many of these studies have proposed more accurate and reliable methods of measuring corruption, their implementability remains limited because they are generally very costly and difficult to replicate across countries.

rates, inexplicably rich public officials, poorly negotiated public procurement contracts) and that this behaviour leaves traces (Kenny and Musatova (2010)). Consequently, red flags are accumulations of traces that may point to the presence of corrupt activities. They are primarily aimed at helping practitioners, investigators and policy makers in estimating the probability of corruption of a certain procurement case and lay the foundation of a new evidence-based approach to fighting corruption.[10] Our paper contributes to this strand of the literature in multiple ways: by proposing new indicators, by validating them and by quantifying the marginal contribution of oblivious indicators.

Finally, from a methodological perspective, our use of ML algorithms is in the spirit of Kleinberg et al. (2015)'s "prediction policy problems." These are policy problems involving a prediction component and, for them, ML techniques are likely to dominate other statistical methodologies. The use of ML algorithms can prove to be particularly effective also when researchers need to model complex relationships without having "a priori" knowledge on the exact structure of the problem. Furthermore, in presence of data availability constraints, gains in predictive accuracy due to funcitonal flexibility might outweight those coming from additional data. Closest to the spirit of our work is the paper by Sestito et al. (2019), offering a clear example of the gains in accuracy that can be achieved through ML flexibility. The authors show how significantly more accurate school rankings can be obtained by applying random forests algorithms to estimate schools' Value Added. By comparing standard OLS and RF-based estimates predictive performance using two different sets of explanatory variables, they adopt a very similar approach to ours and compare the gains from flexibility to the ones of additional data.

In particular, we use off-the-shelf methods and find great improvements when using random forests to select functional forms flexibly. For assessing red flags, ML methods are useful not only because they are expressly meant to deal with the tradeoff between expressiveness of the model (e.g., more covariates included in a linear regression) and risk of over-fitting (e.g., too many covariates relative to the sample size).[11] But also because few red flags have a ground truth causal effect on corruption. Most red flags look at mere tools for corrupt arrangements. But as these tools can be easily substituted with others, so the usefulness of the red flags is closely connected with how easily such modification in corruption practices can take place. Thus, the prediction exercise is both appropriate to study red flags and a first step in the search for indicators having causal effects on corruption.

[10] In 2010, the World Bank issued a guide on the top ten most common red flags of fraud and corruption in procurement for bank financed projects. At the European level, various policy projects have investigated the use of red flags to detect corruption and other illegal behaviors: for example, the European Commission Anticorruption Report (EU Commission, 2014, Report from the commission to the council and the European parliament-EU anticorruption report), which aimed at launching a debate involving the European Commission, Member States and other stakeholders, to assist the anticorruption efforts and to identify policy in which the European dimension can help. Within the EU Horizon 2020 framework, DIGIWHIST (http://digiwhist.eu/), which brings together six European research institutes, with the aim of empowering society to combat public sector corruption through the systematic collection and analysis of information on public procurement in EU countries.

[11] See Mullainathan and Spiess (2017) and Athey (2019).

3 Institutional Framework and Corruption Indicators

In the period of our analysis (2009-2015), the regulations in place for the procurement of public works entail an highly decentralized system.[12] Local authorities (municipalities, counties and regions) hold the vast majority of public tenders and spend about half of the total resources allocated every year to public infrastructures, about €25 billion. In this highly fragmented system, about twelve thousand different purchasing authorities (PAs) were estimated to be active as of 2018. These PAs are heterogeneous in their tasks, capacities and, clearly, in their risk of being involved in corruption episodes.

Despite such heterogeneity, there is mostly a uniform set of rules that these PAs must follow to award public contracts, based on the provisions of European Directives on public procurement. Contracts involving higher amounts (€5million or more) must be procured using competitive procedures (open or restricted), where all qualified firms can participate and the winner is selected either solely on the basis of the price offered or using a scoring formula that combines together points earned for the price and the technical components of the bid.[13] This latter criterion to select the winner is known as most economically advantageous tender (MEAT). Below the €5million threshold, PAs have more discretion to pick not only between the price-only or MEAT criteria, but also to use alternative procedures to the open one. In particular, the smaller is the economic value of the contract up for tender, the more the PAs can restrict competition running either competitive procedures only to selected bidders or conducting a direct negotiation with one or a few bidders[14].

In addition to the awarding criteria and procedures, PAs have discretion over other

[12] The Italian public procurement sector of goods, services and works is mostly shaped around the European Union wide regulations laid down in the EU Public Procurement Directives. In the period that we analyze these are Directives 2004/17/EC and 2004/18/EC, and the Italian law implementing these regulations is the Legislative Decree 12/4/2006, n. 163. See Decarolis, Giorgiantonio and Giovanniello (2011) for an in depth discussion of the national regulations and Decarolis and Giorgiantonio (2017) for the local regulations. Relevant regulatory changes have been introduced in Italy after 2016, when the Legislative Decree 18 April 2016, n. 50 implemented current EU procurement Directives (24/2014 and 25/2014). Among other things, the Legislative Decree of 2016 provided for a reorganization of the functions of contracting authorities through a greater centralization and the introduction of a special qualification system, in order to increase the professional specialization in the public contracts sector. However, this reorganization has not been implemented yet. It should be noted that, starting from the end of December 2015, the ability of smaller municipalities (those that are not provincial capitals) to purchase goods and services over €40,000 and works over €150,000 was reduced. Above these thresholds, these municipalities can merge their public procurement offices with those of either other municipalities or with their province procurement office or, for some purchases, relying on regional or national central purchasing bodies (see the Law Decree 24 April 2014, n. 66 and the Budget Law 2016). However, these measures were suspended by the Law 14 June 2019, n. 55 until the end of 2020.

[13] E.g. the quality of the work or the time for completion. It should be noted that the 2014 European Directives on public procurement provided for the MEAT criterion as ordinary criterion for awarding public contracts.

[14] Negotiated procedures, marked by significant discretionary powers for the administration, are those where the PAs consult their chosen economic operators and negotiate the conditions of the contract with one or more of them. Insofar as these procedures represent a derogation to the general ban on renegotiating offers, they are basically exceptional, being admissible (except for small amount contracts) only when specific conditions apply (chiefly those related to urgency or lack of appropriate offers or applicants).

features of the call for tenders that are likely to affect the corruption risk. Beyond some minimal requirements prescribed by the law, PAs can influence two main aspects. The first is the transparency of the process. They decide both how widely to advertise the call for tenders (for instance by advertising it online and over traditional media) and how detailed are the job descriptions disclosed to potential bidders. The second is the degree of the obstacles to participation that they can erect. While the national (and European) regulations try to curtail this margin of discretion, under certain conditions the PAs can restrict participation to lists of trusted bidders or impose more subtle, but nevertheless effective, barriers. For instance, they specify ad hoc rules for subcontracting, restricting the amount of work that can be subcontracted or to whom (for instance excluding those firms bidding in the call for tenders). Furthermore, PAs can make compulsory for bidders to inspect the detailed project specs or the worksite (or both) and, simultaneously, restrict how and when these inspections can take place: nothing in the law prevents a PA from making the compulsory worksite inspection available for just a tiny window of time. These margins of discretion can serve an important role to help the PAs to achieve publicly desirable goals, but can also trigger corruption phenomena[15]. .

In the light of these considerations, we incorporate these and other elements of the call for tenders into a wide set of corruption indicators. In Table 1, we present our list of indicators along with three different dimensions that we use to classify them. The first column reports the 18 indicators, some of which have sub indicators.[16] The following three columns subdivide the indicators along three dimensions: type of activity that they pertain to, their accessibility to the Supervising Authority (Anticorruption Authority – ANAC) and their source being the literature or operating practices[17].

i) *Source*. Bringing into the literature some indicators previously used in the operating practice of the fight against corruption, but absent from the literature, is indeed a first contribution of this study. The ample existing literature reviewed earlier provides us with a broad set of indicators that have been either used in practice or just derived as implications of models of corruption in public contracting. To this literature, we contribute by adding a few additional indicators that we indicate as originating from operating practices (OP) in Table 1. By this we refer to the fact that our extensive review of the sentences of the judicial authority for the universe of cases involving corruption in public auctions (discussed in the next section) allowed us to identify certain indicators that characterized the actions of agents involved in known corruption cases, but that have not been discussed in the literature. As an example, in a large corruption scandal in the area of Naples, the judge identified that

[15]As a robustness check, we also tried to include the (log) of total corruption crimes committed in a given municipality in the years 2004-2008, thus before our period of analysis begins. Unfortunately, this type of indicator is available only for a subsample of procurement authorities, namely municipalities, and thus create important attrition issues. When repeating our estimates on this restricted sample, the relative ranking of the variables selected by the random forest algorithm differs but results are qualitatively similar to the ones shown in the paper.

[16]An example of indicators is whether the solicitation procedure is negotiated (as opposed to a competitive auction), and one of its sub-indicators is whether the procedure involves or not the publication of a call for tenders.

[17]E.g. investigations by the Sector Authority (Anticorruption Authority – ANAC) or judgments.

Table 1: Corruption Indicators: The Eighteen Red Flags

Indicator	Sub-indicator	Activity	Accessibility	Source
1. Absence of tender call		Information completeness	No	OP
2. call for tenders: page and word number		″	No	OP
3. ANAC info available		″	No	OP
4. Negotiated procedures	*4.1 Negotiated procedure*	Awarding procedures	Yes	L/OP
	4.2 Urgency	″	Yes	L/OP
	4.3 No tender	″	Yes	L/OP
	4.4 No t/n	″	Yes	L/OP
5. Legality protocols		″	Yes	OP
6. Local regulations		″	No	OP
7. Design-Build		″	Yes	OP
8. Scoring rule (MEAT)	*8.1 MEAT*	Awarding criteria	Yes	L/OP
	8.2 MEAT – Tech Score	″	Yes	L/OP
	8.3 MEAT – Qual. Score	″	No	L/OP
9. Price Only - w. ABA		″	No	L/OP
10. No possibility of single source award		″	Yes	L
11. Preferred firm indications	*11.1 Firm list preference*	Obstacles to participation	No	L/OP
	11.2 Firm other preference	″	No	L/OP
12. Open tender days (ODT)	*12.1 OTD*	″	Yes	OP
	12.2 OTD violation	″	Yes	OP
13. Document verification (DV)	*13.1 DV*	″	No	OP
	13.2 DV – Specific dates	″	No	OP
	13.3 DV – Hours share	″	No	OP
	13.4 DV – Hours total	″	No	OP
14. Worksite verification (WV)	*14.1 WV*	″	No	OP
	14.2 WV – Specific dates	″	No	OP
	14.3 WV – Hours share	″	No	OP
	14.4 WV – Hours total	″	No	OP
15. Ad hoc rules for subcontracting	*15.1 Ad hoc rules*	″	No	L/OP
	15.2 No subcontracting	″	Yes	L/OP
16. Prohibition of pooling agreements		″	No	OP
17. Multiple contact points		″	No	OP
18. External contact points		″	No	OP

a distorted use of a provision in the call for tenders was key for the corruption scheme[18]: the visit to the worksite entailed the interaction with a specific individual in the PA. But this individual was using his knowledge about firms interested in the job to inform of their identity the Camorra local clan (the Casalesi), who could then dissuade these firms from bidding. Hence, the provision of a compulsory worksite visit and the details of its working allowed the Casalesi to have full control of the public works administered by the corrupt public agent. In our analysis of calls for tender we, thus, collect a few indicators pertaining to the worksite visit. The same we do for several other indicators in the group denoted by OP in Table 1.

ii) *Type of activity.* A second way in which we classify our indicators is by the type of activity they pertain to. We can distinguish four groups in which the red flags are organized: *a*) information completeness; *b*) awarding procedures; *c*) awarding criteria; *d*) obstacles to participation. To the first group belong indicators involving the transparency and publicity of the call, like the availability and completeness of the call for tenders. To the second belong those indicators specifying how the awarding procedure differs from the default open auction system, like a negotiated procedure (with or without a public call for tenders). To

[18]See Tribunal of Santa Maria Capua Vetere 26 March 2014, II Criminal Section, Picardi, President; Giovanniello, Extensor Judge. This sentence was confirmed by the Supreme Court on February 2019.

the third belong indicators for the awarding criterion used, which at the most aggregate level can be either a price-only criterion or a scoring rule one, weighing together price and other quantitative or qualitative technical features. The fourth group contains a large set of obstacles to participation that PAs can erect by directly limiting firms' participation or, more indirectly, make harder through various requirements on behaviors to take ex ante (like visiting the worksite) or ex post (like limiting subcontracting). A detailed discussion of each individual indicator is contained in the Appendix.

iii) *Accessibility.* The last type of classification is by the indicators' accessibility. Here we take the point of view of how readily available is the measurement of the indicator for the Anticorruption Authority. Systematic surveillance over a certain indicator requires that this indicator is among the fields that PAs have to fill in the online forms that feed the database maintained by the Anticorruption Authority.[19] If the indicator is communicated, we consider it as accessible, otherwise not. While not discussed in the literature, we consider this type of partitioning of the indicators particularly interesting as it allows us to discuss the well known phenomenon of the elusion of monitoring efforts: when agents are aware of being monitored, they might intentionally behave in ways aimed at not raising suspects. In section 7, we return to this distinction to contrast the effectiveness of corruption detection with accessible and oblivious indicators [20].

Interestingly, many of these indicators are common across the public procurement sectors of many countries. Firstly, our accessible indicators are based on fundamental elements present in all public procurement legal framework (e.g., the distinction between price-only and MEAT criteria or competitive and negotiated procedures): not only in European Member States due to the harmonization of EU Directives and Regulations, but also in other non-European countries such as United States, Canada, Australia or Latin America. Moreover, regarding oblivious indicators, they are mainly related to ordinary activities in awarding public works contracts. So, for instance, subcontracts, document verifications or worksite visits are typically provided not only in European countries (e.g., France, Germany or Spain), but also in non-European ones such as United States or Canada.[21] Thus, the relevance of these new indicators has the potential to be rather broad.

[19] In 2000 the Sector Authority established a database on public works contracts. According to the Italian legislation, contracting authorities have to communicate a set of information related to each contract they award (e.g., value, awarding procedure and criterion, open tender days) to the Sector Authority (for more details see Decarolis and Giorgiantonio (2015)). Our corruption indicators include both data that have to be communicated by contracting authorities and data that we collected by a combination of human and machine learning analysis of the tender documentation (see paragraph 4): so, this last type of data is not directly available for the Sector Authority (currently, Anticorruption Authority).

[20] A further potentially relevant dimension of classification could be on firms' and the public authority's awareness of monitoring of specific indicators. One could think that indicators that are known to be monitored by the Authority should have less predictive power. However, based on extensive conversations with those experts in the field of investigation that provided us with the data, we were told that, contrary to what outsiders might expect, there are no indicators that are specifically monitored and targeted by investigators.

[21] In particular, in US Federal contracts for construction works worksite visits are typically required.

4 Main and Verification Data: Descriptive Statistics

We verify the presence of our red flags in two different datasets: Main and Verification data. Our Main data contain 12,786 contracts awarded by counties and municipalities between January 2009 and November 2015. Contracts involved the procurement of simple roadwork jobs (mostly paving jobs and other maintenance works on roads, highways and bridges) and were held in seven regions: three in the North (Lombardy, Piedmont, Veneto), two in Center (Lazio and Umbria) and two in the South (Campania and Sicily).

Our Verification data includes 3,553 contracts for which we obtained both the call for tenders and award notice documentation. The call for tenders is the document with which the PA announces publicly that a tendering procedure is ongoing, while the award notice describes the outcomes of this procedure in terms of who wins, at what price and, possibly, who else participated and with which bid. These contracts involve the same time period, type of jobs and geographical regions of the contracts in the Main data. But the two datasets originate from two different sources and cover a slightly different set of contracts: the Main data are from the Italian Anticorruption Authority (ANAC), which is the public body in charge of the supervision over Italian public procurement system. The Verification data are from a private company (Telemat) that collects and resells to potential bidders detailed tender documentation. About 60% of the contracts in the Verification are also part of the Main data, the remaining 40% being contracts either below the threshold under which the Anticorruption Authority does collect data (€40 thousand) or, in a few cases, contracts whose data have been withheld by the Anticorruption Authority for further inspections due to data incompleteness.

A. Descriptive Evidence on the two datasets. Table 3 presents summary statistics separately for the two datasets. In panel A, we present some basic tender characteristics: the reserve price (i.e., the publicly announced highest price that the PAs is willing to pay), the winning discount (bids are rebates over the reserve price) and the number of bidders (both overall and for the subset of bids clearing admissibility checks; the last row reports the number of invited bidders, as some tenders are by invitation only). Comparing the statistics for the two sets of data reveals several differences. The contracts in the Verification data have a reserve prices that is both higher on average and substantially more dispersed. They also have an higher number of bidders, both invited and effective.

B. Descriptive evidence on the indicators. In panels B and C, we compare the two datasets along our red flags. As mentioned in section 3, we refer to the set of variables included in panel B as the Accessible Indicators, because they can be readily computed and used by the Anticorruption Authority which maintains the sector' supervision through the dataset from which the Main data has been extracted. In panel C, instead, we report additional characteristics that are not currently part of the data collection effort of the Anticorruption Authority[22], and refer to them as Oblivious Indicators.

[22] We collected these additional information by a combination of human and machine learning analysis of the tender documentation. The latter type of methods was applied whenever feasible (for instance, to count the number of pages and words in the call for tenders). When we had to resort to human inspection, each document was scored by two persons and any conflict resolved through expert legal advise. Although any

Table 2: Summary Statistics for the Main and Verification Data

A. Basic Tender Characteristics

	Main Data			Verification Data		
	Mean	S.D.	N	Mean	S.D.	N
Reserve Price (000)	266.38	370.87	12,786	455.17	718.55	3,200
Winner Discount	18.80	13.54	12,500	23.07	12.96	3,439
No. Bidders	17.27	41.67	12,822	46.77	65.34	3,486
No. Accepted Bidders	16.20	39.31	12,822	44.52	62.01	3,486
No. Invited Bidders	5.14	12.68	12,822	14.18	21.97	1,089

B. Accessible Indicators

	Main Data			Verification Data		
Design-Build	0.00	0.03	12,823	0.02	0.13	3,155
Urgency	0.02	0.14	12,823	0.01	0.07	2,812
Negotiated	0.78	0.41	12,814	0.54	0.50	2,697
Negotiated-No Tender	0.96	0.21	10,010	0.32	0.47	1,454
Price Only - w. ABA	0.26	0.44	9,780	0.63	0.48	3,154
Scoring Rule (MEAT)	0.10	0.30	9,753	0.08	0.27	3,314
Open Tender Days	22.07	11.54	12,420	32.06	15.60	2,436
Open Tender Day V.	0.39	0.49	12,420	0.11	0.32	2,420

C. Oblivious Indicators

	Verification Data		
Tender Call Absence	0.33	0.47	3,553
Page Count	25.54	16.55	2,384
Word Count (000)	9.17	8.90	2,384
Legality Protocols	0.32	0.47	2,402
Local Regulations	0.33	0.47	2,449
Negotiated-No T/N	0.09	0.29	3,474
Sole Source Forbidden	0.03	0.16	2,408
Average Qualit. Score	3.84	15.92	3,553
Firm List Preference	0.04	0.18	2,397
Firm Other Preference	0.28	0.45	2,396
Documents Verificat	0.54	0.50	2,392
Worksite Verificat	0.51	0.50	2,392
Ad Hoc Subcontract	0.21	0.41	2,391
No Subcontr to Bid	0.21	0.41	2,391
Contact Points Out	0.30	0.46	1,439
DV-Hours Share	0.77	0.32	2,682
WV-Hours Share	0.95	0.19	2,687

Accessible indicators are available for both datasets (panel B) and oblivious indicators only for the Validation data, for which they were specially collected (panel C). In terms of the

entity supervising the sector take steps analogous to the ones we undertook to collect the same data, the fact that these type of data are not readily available implies that market participants are aware of the non systematic monitoring of these indicators.

partitioning by the type of activity, we have that for the indicators involving Information completeness the call for tenders is missing in 33% of the cases, while information which contracting authorities have to communicate to ANAC is not available for 42% of procedures. On average, calls for tender are 25 pages long and contain about 9.000 words. With respect to the red flags related to the category of awarding procedures, it should be noted that the cases in which negotiated procedures are used when specific conditions related to urgency apply represent only 1% of our sample.[23] In the remaining cases (41% of our sample) they are used according to a discretionary choice of contracting authority. Negotiated procedures without the publication of a call for tenders represent 13% of our sample, while negotiated procedures without the publication of any other notice are 9%. Legality protocols apply in 32% of cases, while local regulations are present in 33%. Design-build project delivery method is used for 2% of contracts.

As regards awarding criteria indicators, most contracts (63%) are awarded using the lowest price criterion and the automatic exclusion of abnormal tenders (ABA). The MEAT criterion is used only in 8% of procedures: in these cases, the incidence of the technical score is largely predominant compared to the qualitative one. Single source awards are allowed in 97% of cases.

Finally, for indicators of obstacles to participation, their presence in the Verification data is summarized by the statistics in panel C. For the two preferred firm indications, a firm register is used in 4% of cases, while other indications of preferred firms are present in 28%. Given an average of 32 days to submit a tender (see open tender days in panel B), the instances in which the number of days provided by the call for tenders is less than the minimum required by law occur in 11% of the cases. Document verification is mandatory in 54% of procedures, worksite verification in 51%. Ad hoc rules for subcontracting are present in 42% of cases: 21% of calls for tender provide for a clause which prohibits the use of subcontracting, while 21% establish rules beyond those provided by law. Multiple contact points for economic operators are present in 5% of cases, external contact points in 30%.

5 Indicators' Validation

The literature and operating practices provide no shortage of red flags. But which of them are truly important to detect corruption? There is little systematic evidence to answer this fundamental question and the reason lies in the scarcity of reliable outcome measures of corruption. Both direct measures of judicial cases and indirect measures involving the price/quality ratio of what procured face the problems discussed earlier.

A second contribution of this paper is that of validating the proposed indicators. We do so through novel measures of firm-level corruption risk. In particular, our main measure is based on police investigation records collected by Decarolis et al. (2019): it allows us to observe for each firm winning a contract an indicator of whether any of its owners or top managers

[23] I.e., the cases for which they were originally provided, given that they should be basically exceptional.

have been object of a police investigation for corruption-related crimes.[24] More precisely, the three types of crimes considered are: (i) corruption, malfeasance and embezzlement, (ii) abuse of power and undue influence and (iii) violations in public auctions. The indicator variable, *corruption risk*, thus takes the value of one whenever the firm has at least one of its owners or top managers who was ever investigated by any Italian police force (civil or military) for at least one crime of the types mentioned above. The usage of an indicator variable rather than the number of crimes (or crimes per person) limits the danger of merely capturing a proxy for firms' size. Although the opening of an investigation is by no means a proof of corruption, given the difficulty of capturing the phenomenon of interest we consider this approach as appropriate to identify firms that are at risk. Furthermore, as explained in greater detail in Decarolis et al. (2019), the typical situation of a flagged firm in our record involves another firm making allegations of corruption, the police conducting for roughly a couple of weeks preliminary investigations to assess the reasonableness of these allegations and, only then, formally opening the investigation that is at the basis of our measure. Thus, while false positive are certainly present, our measure is not a mere list of allegations.

In Table 3, we report summary statistics for both datasets for five different corruption outcome measures. Our new measure, *corruption risk*, appears in the first row, followed by four alternative corruption measures: two based direct judicial evidence (*convicted* and *debarred*) or indirect economic outcomes (*extra cost* and *extra time*). The incidence of *corruption risk* firms across the two datasets is very consistent: 15% of contracts are won by corruption risk firms.[25]

Not surprisingly, the extent of corruption appears much smaller and nearly negligible if measured through judicial data. In particular, to build the *convicted* measure, we reviewed the universe of conviction sentences for corruption cases involving public procurement by the highest court (*Corte di Cassazione*) in the period 1995-2015. We then traced back the whole set of firms involved in the case (by reviewing the first two degrees of judgment preceding the one in front of the highest court). But when matched with our datasets, only 2% of the contracts won in the Main data and 1% of those in the Verification data are awarded to convicted firms. This confirms the limited possibility to use judicial data as a measure of corruption (see paragraph 2) and is in line with what legal scholars and policymakers have lamented about the Italian legal framework to combat corruption, which appears not capable to use convictions as a deterrent.[26]

[24] Decarolis et al. (2019) obtained this information for all firms ever involved in public works in the period between 2000 and 2017. They obtained it not only for firms winning the contracts, but also for participants at the auctions and subcontractors, as well as for the public administrations handling the contracts. The collection of such data was possibile within a framework agreement between the agency for the internal intelligence and security under the Presidency of the Council of Ministers and Bocconi University.

[25] Interestingly, this share of corruption risk winners is roughly constant independently of how we split the sample. Regardless of whether we use coarse measures of the location of the PA (i.e., North vs. South) or more sophisticated measures of the PA's human capital from Baltrunaite et al. (2018), among either its bureaucrats or its ruling local politicians, the proportion of corruption risk winners is always very similar and close to 15%.

[26] Corruption cases are generally complex, and convictions relatively rare. This is particularly true in Italy, where the trial must go through three levels of judgment (*Primo grado*, *Appello*, and *Cassazione*) within a relatively short statutes of limitation (between 6 and 12 years) considering the length of criminal proceedings.

The other judiciary measure, *debarred*, measures a peculiar tool meant to combat criminal (especially mafia) infiltrations in public contracts. Even without conviction sentences, firms can be excluded (i.e., debarred) from the awarding of public contracts if the local police forces signal that, on the basis of the available evidence, they present serious risks of criminal infiltration. But this measure can be appealed in court, this is why we consider it as a judicial measure too. In our data, the instances of contracts awarded to firms that were ever subject to at least one debarment is once again very limited with just 1% of the cases.[27]

Table 3: Summary Statistics: Outcomes

	Main Data			Verification Data		
	Mean	S.D.	N	Mean	S.D.	N
Criminal	0.15	0.36	11,752	0.15	0.36	3,195
Convicted	0.02	0.12	11,752	0.01	0.11	3,195
Debarred	0.01	0.09	11,752	0.01	0.12	3,195
Extra Cost	0.07	0.15	5,122	0.10	0.17	715
Extra Time	0.65	0.77	3,576	0.48	0.68	703

While the judicial measures suffer from underestimating the corruption phenomenon, the two alternative measures based on economic outcomes suffer from measuring it very imprecisely. In both datasets, the average contract experiences a substantial delay in its execution: an *extra time* of 50%, which is about the average, indicates that the execution of the work took one time and half what was originally established at the time of the contract awarding. The cost overruns measured by *extra cost* are also non negligible, albeit less striking. But there are two main problems with these variables. First, the data is incomplete and, most likely, selected: in both datasets, the information is available for less than half of the contracts. The second problem is that, even if the data were complete, it would not be immediate to associate poor contract performance and corruption. As Bandiera, Prat and Valletti (2009), clearly showed for the procurement of standardised goods, the presence of bureaucratic inefficiency might lead to overestimate corruption. Furthermore, for the procurement of public work, as renegotiation might be an optimal strategy for complex contracts Herweg and Schwarz (2018).

Table 4 offers clear evidence on the limits of using *extra cost* and *extra time* as indirect corruption outcomes. This matrix of correlations among the five outcome measures shows that both measures are uncorrelated with *corruption risk*. On the contrary, despite the very limited variation in the judicial variables, we observe that they have a positive and significant correlation with *corruption risk*. Overall, we consider the evidence as strongly indicative of

Only recently, the Law 9 January 2019, n. 3 provided for a lengthening of statutes of limitation, which is entered into force from January 1st, 2020. It should be noted that also the novel measures of firm-level corruption risk that we use can be influenced by the enforcement level, but with a much more limited extent given that – *inter alia* – in these cases statutes of limitations are not provided for.

[27] The data on debarments is not publicly available, but we obtained them through the related project of Decarolis et al. (2019) previously described. The data contains each instance of *interdittiva*, *informativa* and *white list* denial.

Table 4: Outcomes' Correlation Matrix

	Criminal	Convicted	Debarred	Extra Cost	Extra Time
Criminal	1				
Convicted	0.107***	1			
Debarred	0.053***	-0.001	1		
Extra Cost	0.001	0.047***	0.035**	1	
Extra Time	-0.013	-0.016	-0.009	0.095***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the greater merits of *corruption risk* as a measure of corruption than the four other alternative measures collected. This conclusion is also supported when we disaggregate the data at the regional level, as reported in Table 5. The rest of the analysis thus uses *corruption risk* as the outcome of the regression and prediction models presented next.

Table 5: Summary Statistics: Outcomes, By Region

Regione	Criminal	Convicted	Debarred	Extra Cost	Extra Time
Campania	.14	0	.03	.12	.78
Lazio	.20	.01	.03	.06	.29
Lombardia	.19	.01	.01	.11	.65
Piemonte	.22	.06	.01	.12	.53
Sicilia	.17	.003	.04	.10	.52
Umbria	.21	0	0	.37	.73
Veneto	.07	0	0	.08	.59

6 Empirical strategy

The objective of our empirical analysis is to determine whether red flags help predicting corruption. Hence, we are not seeking the estimation of the causal effect of one (or more) of these indicators, but we are interested in how red flags obtainable from tender notices serve to correctly predict that a contract is awarded to a corruption risk firm. This interest in model selection is more typical of the ML literature than of economics. Nevertheless, we see our problem as one of those economic questions considered well suited for ML methods.[28] This is for at least three reasons.

First, in the typical economic study, model selection happens through knowledge of the market forces. But here all indicators are, at least in principle, fully plausible on the basis of existing theories or operating practices. This underscores the elusiveness of the corruption problem that we analyze. Our goal is not to test one (or more) of these theories and heuristics,

[28]"Prediction policy problems", see Kleinberg et al. (2015).

but to let the data drive the model selection stage. Indeed, a novelty of our contribution is precisely to propose new indicators and to validate them. Second, several of the proposed indicators cannot have a ground truth *causal* role, but are nevertheless interesting from a policy perspective. For instance, take the case of the number of days a call for tenders is open for bidding: finding that shorter periods are associated with more corruption is policy relevant if corruption is societally costly and decisions have to be made on what contracts to investigate.[29] But it is unlikely that corruption is *caused* by shorter bidding periods. Allowing bureaucrats discretion over the length of this period can facilitate corruption, but it is unlikely to be a deep driver of the phenomenon.[30] Third, the ML emphasis on model fit is particularly appropriate given the nature of the oblivious indicators. It offers a way to assess the usefulness of investing in learning these indicators beyond our Verification data.

Therefore, in the spirit of Kleinberg et al. (2015), our strategy to use ML tools in economics entails using off-the-shelf ML methods. Within the vast and growing ML literature, our analysis lives within the context of "supervised learning" and, hence, we focus on three workhorse algorithms: LASSO, ridge regression and random forests.[31] The first two are regularization methods aimed at reducing the dimensionality of the model specification, by either dropping (LASSO) or shrinking (ridge) some of the covariates. Both algorithms are well known in economics, being in several ways the tools in ML closest to an OLS.[32] But contrary to an OLS, these methods are algorithms requiring the user make some choices when applying them to the data at hand. This issue is even more pronounced with random forests. Despite this algorithm inherits the simplicity and intuitiveness of the three-based classification approaches, it also requires some adaptations.[33] To minimise the arbitrari-

[29] We see model selection as both useful by itself (in the spirit of designing and using statistical screens, Abrantes-Metz and Bajari (2009)).

[30] This logic does not apply to all indicators, some of which might indeed have deeper effects on the behavior and incentives of the agents involved. We return to this issue after having presented the results.

[31] Following Athey (2019), ML methods are better described as algorithms that might estimate many alternative models and then select among them to maximize a criterion. There is a plethora of ML methods for supervised learning. The most common ones include regularized regression (LASSO, ridge and elastic net), regression trees, ensemble methods (random forest), neural networks, matrix factorization, support vector machines and many others. See Varian (2014), Mullainathan and Spiess (2017) and Athey (2019) for an excellent overview of many of these methods and their applications to economics.

[32] See Ng (2018), Chernozhukov et al. (2017) and Primiceri, Giannone and Lenza (2018). Sparse-models select a relatively small set of regressors that maximize predictive power. The most popular method among sparse modeling techniques is the LASSO (Tibshirani (1996), Tibshirani, Wainwright and Hastie (2015), Belloni, Chernozhukov and Wang (2011)). On other side, dense models aim at keeping all possible regressors in the model, but solve the over-fitting problems by reducing the size of the parameter estimates whenever the sample information reaches a lower bound. These methods are usually known as "shrinkage" or "regularizing" methods. Ridge regression is the most common example.

[33] Random Forest is an ensemble tree-based learning algorithm, meaning that the final prediction model is an average of the predictions obtained growing many individual trees. The algorithm uses bootstrap aggregating, also known as bagging, to limit over-fitting and improve out-of-sample accuracy. Bagging implies fitting each tree on a bootstrap sub-sample, rather than on the original full training sample. The method consists of the following steps. First, a given number of random sub-samples are drawn from the training sample. Second, a random subset of variables among the total set of predictors is selected. This subset of variables is used to determine each subsequent split in a tree. Each internal node is split until a certain predictor optimizes the splitting criterion. Then, several decision trees are grown, one for each randomly-drawn sample. Each decision tree is built up to its maximum size (no pruning occurs). This leads

ness of our choices and ensure replicability, we implemented all three algorithms through commonly used statistical packages.[34]

Our data structure has features both in common and different from the typical machine learning exercise. As typical in ML, we have a large number of potential predictors, while observing a relatively small set of contracts. In this context, standard techniques such as OLS are known to perform poorly and to be inferior to alternatives proposed by the ML literature. In this sense, looking at LASSO and ridge regression is a natural starting point as these methods have been developed to address this type of problem. But the results below will also clearly point toward the usefulness of the random forests algorithm. This is likely due to its greater functional form flexibly, combined with the dense nature of our data. For all methods, we will report measures of their prediction accuracy. But where our data departs from the typical ML setting. We acknowledge that our two datasets have potentially different distributions of the relevant variables and, hence, our two samples shall not be confused with the training and validation data to which the ML literature refers to. Furthermore, while we observe the outcome variable in both datasets, it is the set of indicators that differs between the two. Verification data can be analyzed through either a large model with all indicators or a small one using only a subset of them. But only the latter, small model is feasible for the Main data. Our interest is in learning how this differences limits the ability of a few, standard algorithms to accurately predict the outcome.

7 Results

We begin the presentation of the results by contrasting OLS estimates with those obtained with the two ML workhorse algorithms: LASSO and ridge regression. After having discussed the results for both the Main and Verification data, we introduce the findings from the random forests and, then, conclude by comparing all four methods to evaluate the contribution of the oblivious indicators.

A. Main data findings The three columns of Table 6 report the estimated coefficients obtained through OLS, LASSO and ridge regression. The model specification includes all the indicators available in the Main data, as well as year and region fixed effects. This is a small set of indicators and the problems of curse of dimensionality are unlikely to bite, but for consistency with the analysis that follows we apply ML algorithms with this restricted set of indicators. These indicators are those directly observed by the Anticorruption Authority and, with regard to the classification by type of activity involved, it entails mostly indicators of the awarding procedures and criteria groups. The OLS estimates indicate that the model has a low explanatory power with an adjusted R^2 of less than 4% and a MSE of 0.35. Among the individual coefficients, the only one that is statistically significant is that on the MEAT criterion. Contracts awarded using this multi-criteria approach are positively associated with

to a very dense forest. Finally, the trees are combined, averaged.

[34] We report in the text the results using Stata15 routines of Townsend (2017) for LASSO and Ridge and of Schonlau (2019) for the random forest. Both Stata and Python codes will be made available on the authors' web site.

corruption risk winners. This indication is in line with what the theory would suggest and it is an interesting finding given the widespread usage of this type of criterion[35].

Table 6: Estimates for the Small Model - Main Data

	OLS	LASSO	Ridge
	Corruption risk	Corruption risk	Corruption risk
Design-Build	0.003	0.003	0.003
	[0.004]		
Urgency	-0.005	-0.004	-0.004
	[0.003]		
Negotiated	0.004	0.002	0.002
	[0.005]		
Negotiated-No Tender	0.003	0.002	0.003
	[0.004]		
ABA	-0.003	-0.003	-0.003
	[0.004]		
MEAT	0.006*	0.008	0.009
	[0.004]		
Open Tender Days	-0.003	-0.001	-0.001
	[0.005]		
Open Tender Day V.	0.000	0.001	0.001
	[0.004]		
Observations	12,623	12,623	12,623
Adj R2	0.036		
MSE	0.355	0.126	0.126
False Positive	3,323	3,350	3,335
False Negative	2,159	2,161	2,172

$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$. All specifications include year and region fixed effects. Robust standard errors in parentheses for OLS estimates. MSE is equal to the root mean squared error for OLS, and to the minimal cross-validation mean squared error for LASSO and Ridge regressions. False Positive indicates the number of cases in which a non-corrupt firm is classified as corrupt by the model. False Negative indicates the number of cases in which a corrupt firm is classified as non-corrupt by the model.

LASSO and ridge regression confirm this indication as for both algorithms the MEAT is the indicator with the largest coefficient. As standard in this literature, the comparability of all the coefficients across the three columns is ensured by the usage of both outcomes and covariates that are centered and standardized. Thus, all three methods indicate a magnitude of the MEAT coefficient that is about twice as that of the next best indicator. The three methods also agree on what this second best indicator is: urgency. Contrary to a naive view that the greater flexibility allowed to bureaucrats when they award contracts under the faster procedures allowed by invoking an urgency, urgency is negatively associated with corruption risk winners. An analogous surprising negative effect is found for the open tender days indicator. But possibly both indicators are high on the list of the usual suspect for corrupt behavior of the Anticorruption Authority, so that this evidence is compatible with actions aimed at avoiding detection by the monitoring entity.[36] Also negative is the sign on

[35]For instance, as we mentioned before (section 3), the 2014 European Directives on public procurement provide for the MEAT criterion as ordinary criterion for awarding public contracts

[36]It should be noted that, in the period of our analysis (2009-2015), the threshold whithin which negotiated

ABA. But this is in line with our expectations: this indicator marks contracts awarded with a lottery-style mechanism that is prone to collusion by bidders coordinating their bids, but very hard to pilot for a corrupt bureaucrat. All other indicators are positively associated with corruption risk winners, as expected from the literature: design-and-build contracts (as opposed to build only contracts), negotiated procedures (and negotiated without prior publication of the call for tenders) and violations in the minimum number of days during which the call for tenders is published. Interestingly, the LASSO does not exclude any of the indicators considered, dropping only some of the fixed effects, and this points to our problem being dense.[37]

While the estimates reported in the three columns are remarkably similar, the overall performance of the models is disappointing. Although the MSE is halved in the ML methods if compared to the OLS one, the prediction is highly inaccurate for all three methods. This is clearly showed by the high fraction of both type I error (false positive) and type II error (false negative) reported at the bottom of the table.[38] For all three models, the former are about 26% of the cases, while the latter are 17% of the cases. Random forests will allow substantial improvements in this classification accuracy, but before discussing that we briefly discuss the findings for the Verification data.

B. Verification data findings Estimates for the Verification data are reported in Table 7. The algorithms are the same discussed above, but now we feed the algorithms with two different sets of indicators. For each algorithm, the first column considers a small model with the same set of indicators used for the Main data. The second column uses a large model that includes all available indicators. Furthermore, for the large model we also include 20 dummy variables to account for all instances in which some of the contracts have indicators that cannot be (unambiguously) assessed.[39]

There are a number of interesting results on the individual indicators that we can learn from the large model. First, for the MEAT criterion it appears that the association with corruption is stronger the more the scoring rule assigns points for qualitative (as opposed to quantitative) parameters. In contrast to the previous findings, all three methods indicate

procedures can be used for awarding public works contracts even if specific urgency conditions do not apply, was gradually raised from €200,000 to €1,000,000. These regulatory changes can have reduced the necessity of invoking an urgency to use faster procedures and have a greater flexibility.

[37] This indication is also confirmed when using a recent approach proposed by Primiceri, Giannone and Lenza (2018) to test for the sparse vs. dense nature of the data.

[38] It is important to emphasize that the ML literature does not frame itself as solving estimation problems so estimating $Pr(Y = k|X = x)$ is not the primary goal. Instead, the goal is to achieve goodness of fit in an independent test set by minimizing deviations between actual outcomes and predicted outcomes. In applied econometrics, we often wish to understand an object like $Pr(Y = k|X = x)$ in order to perform exercises like evaluate the impact of changing one covariate while holding others constant. This is not an explicit aim of ML modeling. LASSO and ridge regression are the only exception.

[39] This is mainly due to a) incomplete communications by the contracting authorities in the tender documents, or b) problems in the match with the complementary ANAC data on some outcomes and controls. While it would be impossible to run estimation on a subsample of tenders for which we have no missing across all variables (as we would end up with less observations than variables), we proceed by filling in the missing values in the following way. For each indicator, we replace the missing value with the median value across the sample. We then create a dummy variable tracking the filling procedure for that particular variable. We thus include 20 extra dummy variables in our regressions.

Table 7: Estimates for the Small and Large Models - Verification Data

	OLS		LASSO		Ridge	
	Corruption risk	Corruption risk	Corruption risk	Corruption risk	Corruption risk	Corruption risk
Design-Build	-0.068	-0.071	-0.041	0	0.005	0.010
	[0.065]	[0.066]				
Urgency	-0.074	-0.063	-0.054	-0.017	-0.040	-0.024
	[0.076]	[0.074]				
Negotiated	0.000	0.000	0	0	0.000	0.000
	[0.000]	[0.000]				
Negotiated-No Tender	0.020	0.022	0.016	0	0.008	0.002
	[0.028]	[0.037]				
Price Only - w. ABA	0.025	0.021	0	0	-0.012	-0.012
	[0.020]	[0.020]				
Scoring Rule (MEAT)	0.042	-0.019	0.037	0	0.026	0.013
	[0.033]	[0.047]				
Open Tender Days	0.001	0.001	0.001	0.001	0.001	0.000
	[0.001]	[0.001]				
Open Tender Day V.	0.077**	0.079**	0.082	0.064	0.046	0.025
	[0.031]	[0.031]				
Tender Call Absence		-0.162		0		-0.003
		[0.100]				
Page Count		0.000		0		0.000
		[0.001]				
Word Count		0.001		0.001		0.001
		[0.001]				
Legality Protocols		-0.008		0		-0.005
		[0.026]				
Local Regulations		-0.035		-0.018		-0.012
		[0.027]				
Negotiated-No T/N		-0.011		0		0.002
		[0.050]				
Sole Source		-0.070*		-0.053		-0.028
Forbidden		[0.040]				
MEAT-Qual. Score		0.0015*		0.001		0.000
		[0.0008]				
Firm List Preference		0.043		0.020		0.014
		[0.047]				
Firm Other		-0.012		0		0.002
Preference		[0.022]				
Documents Verificat		-0.017		-0.006		-0.003
		[0.037]				
Worksite Verificat		0.024		0		-0.001
		[0.032]				
Ad Hoc Subcontract		-0.190***		-0.005		-0.008
		[0.035]				
No Subcontr to Bid		0.196***		0		-0.005
		[0.040]				
Contact Points Out		0.023		0.014		0.009
		[0.027]				
DV-Hours Share		0.025		0.021		0.011
		[0.038]				
WV-Hours Share		0.069		0.058		0.029
		[0.043]				
Observations	3,195	3,195	3,195	3,195	3,195	3,195
Adj R2	0.028	0.029				
MSE	0.356	0.356	0.128	0.128	0.129	0.128
False Positive	649	643	709	693	769	748
False Negative	630	623	626	620	622	623

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$. All specifications include year and region fixed effects. All the choices involving standard error and MSE calculations are identical to those reported in the note to Table 6.

that violations in the minimum number of days for which the call for tenders is open is an important predictor. The large model also allows to discover the relevance of several indicators, mostly belonging to the group that we described as obstacles to participation. In particular, we observe that features related to the existence and characteristics of the obligations involving both access to the tender documentation and the worksite inspection matter. The easier it is satisfying these requirements (in terms of allowing a larger share of time during the bidding period in which these obligations can be satisfied), the less likely the winner is a corruption risk firm. The estimates in the table reveal that among the broad set of indicators there are both positively and negatively associated indicators with corruption. There are also indicators that the LASSO completely drops and that the ridge regression shrinks to nearly zero. Most of the indicators in the group covering information completeness are of this kind, with the only exception of the number of words in the call for tenders, whose coefficient is – however – rather small in magnitude.

In terms of the overall model fit, the low predictive ability discussed earlier for the Main data also applies to the Verification data. Interestingly, the fit measure improves little when moving from the small to the large model. Along all three dimensions of MSE and type I and II errors the findings are quantitatively very close to those reported for the Main data.

To improve on these measures, we introduce next a random forests approach. The random forest algorithm gives a more accurate estimate of the error rate, as compared with standard decision trees (Breiman (2001)). Error rate is measured by the out-of-bag error during the training process. In each tree of the random forest, the out-of-bag error is calculated based on predictions for observations that were not in the bootstrap sub-sample for that tree. Parameter tuning is done in order to minimize out-of-bag error. After training the random forest algorithm, it is possible to get estimates of the relative importance of each of the covariates in terms of predictive power.[40]

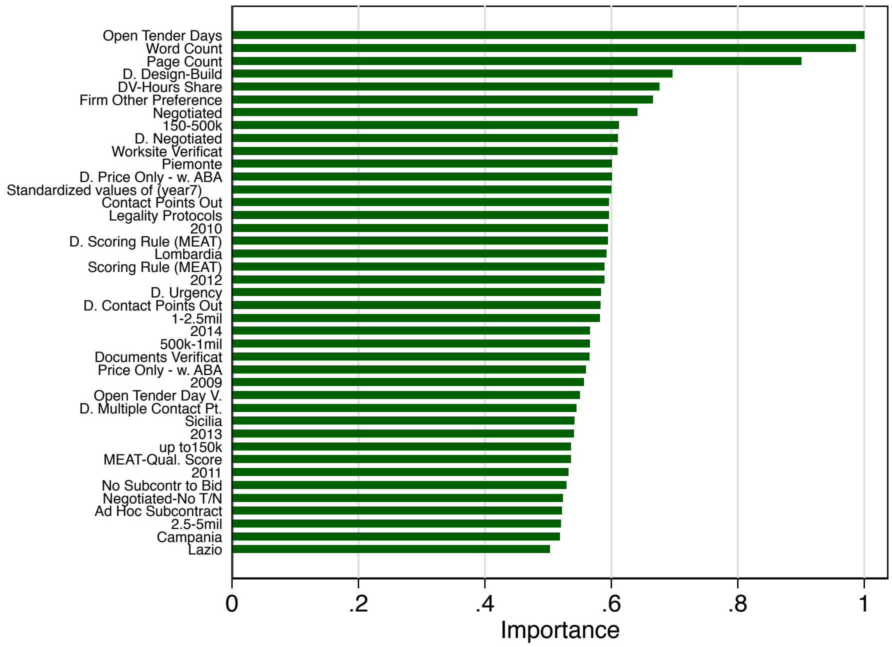
Contrary to regression models, there is no simple way to fully represent the results of a random forest. Figure 1 reports the importance of the indicators, separating high (top panel) and low (bottom panel) importance indicators. This type of visualization describes how much each indicator contributes to the average decrease in impurity over trees. Although routinely used to summarize, for a given model, those features which are most important in explaining the target variable, it has well known biases. We therefore comment it only briefly and then move on to the discussion of the entire model in terms of MSE.

In Figure 1, we immediately see that the random forests agrees with the other methods for some variables (like the MEAT), but not for all. Indeed, many of the information completeness indicators are found to be highly relevant: the total number of both words and pages in the call for tenders are among the top red flags in terms of importance, while being only marginally relevant (albeit never fully excluded) under the alternative prediction models. Through Figure 1, we can also offer a visual representation of the various additional controls - fixed effects and dummy variables for missing red flag data - that were included in the earlier models, but not reported for readability.

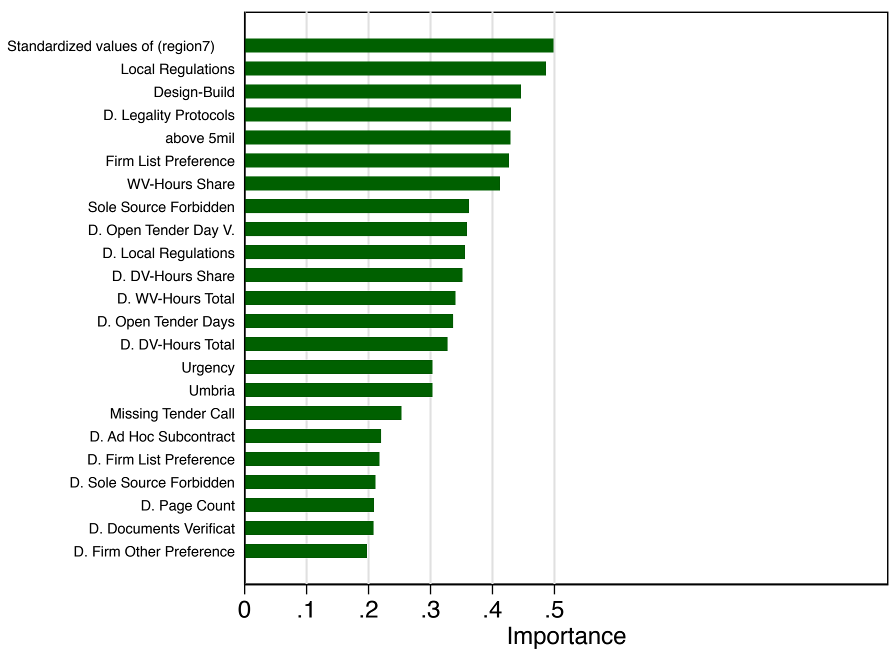
[40]This might be useful in order to select only a subset of variables to use in a standard regression model, or – in general – to have a sense of which are the most important drivers of the observed outcomes.

Figure 1: Random Forest

(a) High Importance



(b) Low Importance



Interestingly, indicators for those regions often considered more at risk of corruption do not show up prominently. Also relevant is that the indicators for missing record are mostly concentrated - albeit with some exceptions - among the lesser relevant indicators, thus suggesting that the incompleteness of the call for tenders is not a major driver in the findings. But this is also, at least in part, attributable to type of results representation used, which over-represents the importance of continuous features and high-cardinality categorical variables. For this reason, to better assess the usefulness of the random forest model relative to the other models, especially with respect to the inclusion of te oblivious indicators, we now turn to a MSE comparison of these models.

Table 8: Predictive Accuracy Across Samples and Models

A. Verification Data

	Small Model			Large Model		
Model:	MSE	False Positive	False Negative	MSE	False Positive	False Negative
OLS	0.356	0.199	0.197	0.356	0.201	0.197
Lasso	0.128	0.227	0.196	0.128	0.231	0.194
Ridge	0.128	0.224	0.194	0.128	0.214	0.195
Random Forest	0.188	0.001	0.049	0.162	0.000	0.059

B. Main Data

	Small Model		
Model:	MSE	False Positive	False Negative
OLS	0.355	0.263	0.171
Lasso	0.126	0.265	0.171
Ridge	0.126	0.264	0.172
Random Forest	0.126	0.000	0.082

C. Oblivious indicators contribution In Table 8 we report the prediction accuracy for all four methods, across all datasets and models. The results for OLS, LASSO and ridge regression are those discussed above. They are reported in this table only to ease their comparison with the random forests ones. Beginning with the top panel where results for the Validation data are reported, the random forests appears to perform well in terms of both MSE and classification errors. Its MSE is close to those of both LASSO and Ridge, while its classification errors are substantially lower. While the direction of this effect is to be effected, since the random forests algorithm tends to fit very well in sample, its magnitude is rather interesting as to false positives drop to essentially zero and false negatives to a low share.

A second feature worth noticing how the MSE changes for the small model between the Main and Validation data. For all the algorithms, moving to the substantially larger Main data leads to a drop in the MSE, but this drop is stronger for the random forest model.

Its MSE indeed become identical in the Main data to that of the LASSO and Ridge, while preserving the lower rate of classification errors.

Finally, it is interesting to discuss how the different models respond to the inclusion of the oblivious indicators in the large model. The random forest model is the one whose performance in terms of MSE improves the most, while also retaining the lowest classification errors: passing from the small to the large model reduces the MSE by 0,026, or 14% at its baseline value in the small sample. This model, likely do to the greater flexibility of its functional form, is better able to exploit the additional information provided by the inclusion of more red flags. Importantly, however, as the ranking of indicatos in Figure 1 show, adding indicators is not only important because some of them are individually relevant red flags. On the contrary, it is their overall contribution to the model performance that improves thanks to their inclusion.

8 Conclusions

In this study, we exploited new contract-level data collected both directly by us in the call for tenders documents and through data warehouse of the public entity monitoring corruption risk in Italian public procurement. We use these data to measure a large set of red flags for corruption, some novel to the literature, but part of operating practices of the Italian judiciary. We then combined these red flags with detailed firm-level corruption measures allowing us to obtain a measure of corruption superior to most of the alternatives in the literature. Finally, using ML tools we explored the usefulness of the red flags to predict contract-level corruption.

We succeed in determining that some indicators have a clear predictive power by comparing different methods and samples. We also show that, among ML methods, the random forests algorithm provides the most accurate prediction and, more crucially, that if this algorithm is used large prediction improvements are attained by including those indicators that we directly collect, but are not monitored by the supervising entity. Overall, these results represent the first systematic evidence on the predictive contribution of a large number of red flags for corruption. Given the high perceived costs for society of corruption, our results offer a way to think about the benefits in investing in the collection of red flags for corruption, especially considering that many of these indicators can be common across the public procurement sectors of many countries.[41] Statistical tests are by no means a sufficient element for conviction, but can be fundamental to direct in the right direction the scarce resources of the monitoring authorities.[42]

[41] While we are not aware of cases for corruption detection via red flags, there are known cases of success stories regarding collusion. For instance, Imhof, Rutz and Karagk (2018) discuss a case of a cartel of firms bidding on roadwork procurement auctions in Switzerland that was initially detected through statistical screens and, later, convicted by court once hard evidence on this cartel emerged.

[42] Furthermore, this is an interesting instance of the practice anticipating the theory as very recent by Chassang et al. (2018) has offered theoretical foundations and further empirical verification precisely for the main red flags that triggered the Swiss case.

In terms of our approach in this study, the choice of adopting a ML approach is well targeted to the fact that only some indicators have the potential to truly have causal effects. Some of the indicators are mere tools to achieve corruption and they are likely to be highly fungible with other tools. However, it would clearly be of interest to explore the causal effects of those indicators that have the potential of ground truth causal effects. For instance, it is a nearly universal feature of procurement systems that the awarding criterion used is either price or MEAT. If one of these two systems must be chosen and they are systematically different in the scope and incentives they create for corruption, it would be reasonable and useful to assess causal impacts on corruption. In this respect, we see our results a first step for a causal analysis. Athey (2019) foresees a steady increase in the connection between ML and economic methods that involves using the ML methods to conduct initial model selection a phase that is currently often done in an informal and undocumented way, but that might refute the validity of the estimates via problems of specification search and multiple testing. In this sense, our work can be seen as indeed part of a broader research project that is continued in other works where we analyze through conventional economic methods the effects of more discretionary awarding procedures and criteria on political connections (Baltrunaite et al. (2018)) and corruption (Decarolis et al. (2019)).

From a policy point of view, our results highlight several relevant aspects. Firstly, discretionary mechanisms for selecting private contractors (in particular, MEAT criterion and negotiated procedures) should be appropriately used. Discretion plays a crucial role for effective procurement, especially in the case of complex contracts. Indeed, when the technical characteristics can be differentiated in advance according to a quality scale and graduated by degree of desirability in view of the administrations objectives, the MEAT criterion – which typically include a balance of price and quality parameters – is preferable. However, given its high vulnerability to corruption risks, a careful selection of quality features according to which public contracts will be awarded is crucial. In this respect, contracting authorities should clearly define the objectives pursued in the call for tenders and prefer "measurable" parameters[43], that can be less easily manipulated. Regarding negotiated procedures, they show some advantages over competitive procedures, representing a faster and more flexible instrument for selecting private contractors. However, the provision of transparency requirements is essential, in particular limiting the use of a negotiated procedure without the publication of any notice. Moreover, our analysis shows the relevance of monitoring compliance with the minimum time limit for submission of tenders and providing adequate controls on subcontracting, which seems to represent an area very vulnerable to corruption risks.

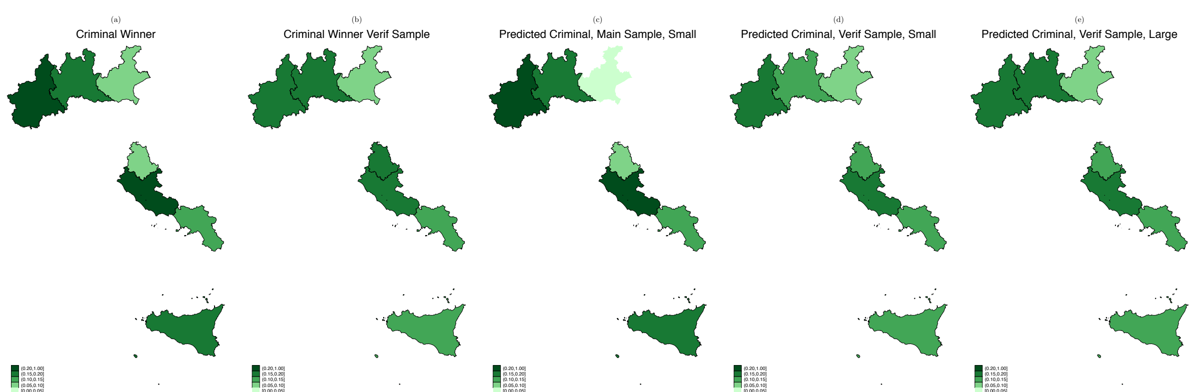
At a more general level, our analysis suggests that a higher standardization of call for tenders documents can contribute to reduce corruption risks. For this purpose, sector authorities or specialised public bodies can play a crucial role. In addition to diffusing best practices, these structures may contribute to the harmonization of standards, increasing the degree of certainty of interpretation in a highly complex regulatory context. Moreover, an adequate centralization and professionalisation of contracting authorities (*inter alia*, in

[43] For example, square meters for green spaces or the number of windows instead of more generic formulas such as "architectural merit" or "aesthetic value of the work".

terms of specialized technical skills and project management capability) should be ensured in order to properly select private contractors, also mitigating corruption risks. For instance, prerequisites in terms of competence and integrity can be provided for public procurers.

In terms of the avenues for future research, a dimension that we do not explore in our analysis but that might be interesting to explore in the future is that of exploiting geographical differences among contracting authorities. For instance, Figure 2 offers a different way to represent the red flags' contribution to the description of criminal infiltrations in public contracts. It reports five heat maps of the Italian regions part of our analysis. This can be related to the descriptive evidence in Table 5 earlier. Darker shades of green indicate an higher share of contracts in that region awarded to a corruption risk winner. In plot a, using the Main sample corruption risk winners are observed through the corruption risk firm variable. The same is true in panel b for the verification sample. The next three panels, instead, report the prediction of which contracts are awarded to corruption risk firms based on the various random forest models presented above: Main sample and small model (panel c), Verification sample and small model (panel d) and Verification sample and large model (panel e). Panels a and b are interesting as they indicate how the geographical dispersion of the phenomenon is not a simple North/South divide, but is characterised by wider variation. The Veneto region in the North-West appears to have a systematically lower exposure to the risk of corruption risk penetration in the contracts. At the opposite end of the spectrum there is Lazio, in the Center, where there is the highest incidence of corruption risk winners. The predictive model based on the Main sample tracks closely the penetration of corruption risk firms observed for this sample in panel a. Between the two predictive models for the Verification data, we see that the large model allows to improve the classification especially for the Lombardy region, in the North-Center, although across all regions the prediction gets closer to the observed evidence of panel b.

Figure 2: Corruption Risk Penetration across Sample Regions



References

Abrantes-Metz, Rosa, and Patrick Bajari. 2009. "Screen for conspiracies and their multiple applications." *Antitrust*, 24: 66.

ANAC. 2015. "Aggiornamento 2015 al Piano Nazionale Anticorruzione. Determinazione n. 12 del 28 ottobre 2015."

Athey, Susan. 2019. "The Impact of Machine Learning on Economics." In *The Economics of Artificial Intelligence: An Agenda.* , ed. Ajay K. Agrawal, Joshua Gans and Avi Goldfarb. Chicago:University of Chicago Press.

Baltrunaite, Audinga, Cristina Giorgiantonio, Sauro Mocetti, and Tommaso Orlando. 2018. "Discretion and Supplier Selection in Public Procurement." *Bank of Italy Working Paper No.1178.*

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and passive waste in government spending: evidence from a policy experiment." *American Economic Review*, 99(4): 1278–1308.

Belloni, Alexandre, Victor Chernozhukov, and Lie Wang. 2011. "Square-root lasso: pivotal recovery of sparse signals via conic programming." *Biometrika*, 98(4): 791–806.

Bologna, Jamie, Amanda Ross, et al. 2015. *Corruption and entrepreneurship: Evidence from a random audit program.* West Virginia Univ., Department of Economics.

Breiman, Leo. 2001. "Random forests." *Machine learning*, 45(1): 5–32.

Brodi, E., R. Occhilupo, and Marco Tonello. 2012. "La corruzione in Italia e le proposte normative di contrasto al fenomeno." *Bank of Italy Temi di Discussione (Working Paper).*

Brollo, Fernanda, and Ugo Troiano. 2016. "What happens when a woman wins an election? Evidence from close races in Brazil." *Journal of Development Economics*, 122: 28–45.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. 2013. "The political resource curse." *American Economic Review*, 103(5): 1759–96.

Calzolari, Giacomo, and Giancarlo Spagnolo. 2009. "Relational Contracts and Competitive Screening." *CEPR Discussion Paper No. DP7434.*

Chassang, Sylvain, Kei Kawai, Jun Nakabayashi, and Juan Ortner. 2018. "Data Driven Regulation : Theory and Application to Missing Bids." *Working Paper.*

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. 2017. "Double/debiased/neyman machine learning of treatment effects." *American Economic Review*, 107(5): 261–65.

Colonnelli, Emanuele, and Mounu Prem. 2017. "Corruption and Firms: Evidence from Randomized Audits in Brazil."

Coviello, Decio, Andrea Guglielmo, and Giancarlo Spagnolo. 2017. "The effect of discretion on procurement performance." *Management Science*, 64(2): 715–738.

Dal Bó, Ernesto, and Martín A Rossi. 2007. "Corruption and inefficiency: Theory and evidence from electric utilities." *Journal of Public Economics*, 91(5-6): 939–962.

Decarolis, Francesco, and Cristina Giorgiantonio. 2014.

Decarolis, Francesco, and Cristina Giorgiantonio. 2017. "Local Public Procurement Regulations: The Case of Italy." *International Review of Law and Economics*, 43: 209–226.

Decarolis, Francesco, Cristina Giorgiantonio, and Valentina Giovanniello. 2011. "The awarding of public works in Italy: an analysis of the mechanisms for the selection of contractors." *Mercato Concorrenza Regole*, 13(2): 235–272.

Decarolis, Francesco, Ray Fisman, Paolo Pinotti, and Silvia Vannutelli. 2019. "Corruption in Procurement: New facts from Italian Government Contracting." *Working Paper.*

Del Monte, Alfredo, and Erasmo Papagni. 2001. "Public expenditure, corruption, and economic growth: the case of Italy." *European journal of political economy*, 17(1): 1–16.

Di Nicola, Andrea, and Andrew McCallister. 2006. "Existing experiences of risk assessment." *European Journal on Criminal Policy and Research*, 12(3-4): 179–187.

Di Tella, Rafael, and Ernesto Schargrodsky. 2003. "The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires." *The Journal of Law and Economics*, 46(1): 269–292.

Dreher, Axel, and Martin Gassebner. 2013. "Greasing the wheels? The impact of regulations and corruption on firm entry." *Public Choice*, 155(3-4): 413–432.

European Comission. 2014. "Report from the Commission to the Council and the European Parliament. EU Anti-corruption Report."

Fazekas, Mihaly, Istvan Janos Toth, and Lawrence Peter King. 2016. "An objective corruption risk index using public procurement data." *European Journal on Criminal Policy and Research*, 22(3).

Ferraz, Claudio, and Frederico Finan. 2011. "Electoral accountability and corruption: Evidence from the audits of local governments." *American Economic Review*, 101(4): 1274–1311.

Golden, Miriam A, and Lucio Picci. 2005. "Proposal for a new measure of corruption, illustrated with Italian data." *Economics & Politics*, 17(1): 37–75.

GRECO. 2012. "Evaluation Report on Italy Incriminations (ETS 173 and 191, GPC 2)."

Herweg, Fabian, and Marco A Schwarz. 2018. "Optimal Cost Overruns: Procurement Auctions with Renegotiation." *International Economic Review*, 59(4): 1995–2021.

Hessami, Zohal. 2014. "Political corruption, public procurement, and budget composition: Theory and evidence from OECD countries." *European Journal of political economy*, 34: 372–389.

Hsieh, Chang-Tai, and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448.

Huntington, Samuel P. 2006. *Political order in changing societies.* Yale University Press.

IMF. 2016. "Corruption : Costs and Mitigating Strategies."

Imhof, David, Samuel Rutz, and Yavuz Karagk. 2018. "Screening for bid rigging - does it work?" *Journal of Competition Law & Economics*, 14(2): 235–261.

Kenny, Charles, and Maria Musatova. 2010. *Red flags of corruption in world bank projects: an analysis of infrastructure contracts.* The World Bank.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction policy problems." *American Economic Review*, 105(5): 491–95.

Leff, Nathaniel H. 1964. "Economic development through bureaucratic corruption." *American behavioral scientist*, 8(3): 8–14.

Lui, Francis T. 1985. "An equilibrium queuing model of bribery." *Journal of political economy*, 93(4): 760–781.

Manelli, Alejandro M., and Daniel R. Vincent. 1995. "Optimal Procurement Mechanisms." *Econometrica*, 63(3): 591–620.

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

Mauro, Paolo. 1998. "Corruption and the composition of government expenditure." *Journal of Public Economics*, 69(2): 263–279.

Mitchell, David T, and Noel D Campbell. 2009. "Corruption's effect on business venturing within the United States." *American Journal of Economics and Sociology*, 68(5): 1135–1152.

Mocetti, Sauro, and Tommaso Orlando. 2017. "Corruption and personnel selection and allocation in the public sector." *Bank of Italy Occasional Paper*, , (402).

Mo, Pak Hung. 2001. "Corruption and economic growth." *Journal of Comparative Economics*, 29(1): 66–79.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives*, 31(2): 87–106.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny. 1991. "The allocation of talent: Implications for growth." *The Quarterly Journal of Economics*, 106(2): 503–530.

Ng. 2018. “Variable Selection in Predictive Regressions.” *Handbook of Economic Forecasting*, 2: 752–789.

PCM. 2010. “La corruzione in Italia. Per una politica di prevenzione. Analisi del fenomeno, profili internazionali e proposte di riforma, Rapporto della Commissione per lo studio e l?elaborazione di proposte in tema di trasparenza e prevenzione della corruzione nella pubblica amministrazione.”

Primiceri, Giorgio, Domenico Giannone, and Michele Lenza. 2018. “Economic Predictions with Big Data: The Illusion of Sparsity.” *Working Paper*.

Rizzica, Lucia, and Marco Tonello. 2015. “Exposure to media and corruption perceptions.” *Bank of Italy Temi di Discussione (Working Paper) No*, 1043.

Schonlau, Matthias. 2019. “RANDOMFOREST: Stata module to implement Random Forest algorithm.” *Statistical Software Components, Boston College Department of Economics*.

Schulze, Gnther G., Bambang Suharnoko Sjahrir, and Nikita Zakharov. 2016. “Corruption in Russia.” *The Journal of Law and Economics*, 59(1): 135–171.

Sestito, Paolo, Fritz Schiltz, Tommaso Agasisti, and Kristof De Witte. 2019. “The Added Value of More Accurate Predictions for School Rankings.” *SSRN Electronic Journal*.

Shleifer, Andrei, and Robert W Vishny. 1993. “Corruption.” *The quarterly journal of economics*, 108(3): 599–617.

Shleifer, Andrei, and Robert W Vishny. 1994. “Politicians and firms.” *The Quarterly Journal of Economics*, 109(4): 995–1025.

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

Tibshirani, Robert. 1996. “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.

Tibshirani, Robert, Martin Wainwright, and Trevor Hastie. 2015. *Statistical learning with sparsity: the lasso and generalizations.* Chapman and Hall/CRC.

Townsend, Wilbur. 2017. “ELASTICREGRESS: Stata module to perform elastic net regression, lasso regression, ridge regression.” *Statistical Software Components, Boston College Department of Economics*.

Varian, Hal R. 2014. “Big data: New tricks for econometrics.” *Journal of Economic Perspectives*, 28(2): 3–28.

Weaver, Jeff. 2016. “Jobs for sale: Corruption and misallocation in hiring.” Tech. rep., Mimeo, Yale University. B. Treatment. 1.