Identifying corruption through latent class models: evidence from transition economies

Pieroni, Luca and d'Agostino, Giorgio and Bartolucci, Francesco

University of Perugia

13 January 2013

Identifying corruption through latent class models: evidence from transition economies

*Luca Pieroni, ** Giorgio d'Agostino, *Francesco Bartolucci

* Department of Economics, Finance and Statistics, University of Perugia, Italy.

**Department of Economics, University of Rome III, Italy.

11th January 2013

Abstract

Evaluation of corrupt activities is incrementally based on administration of questionnaires to firms in business, and generally involves a large number of items. Data collected by questionnaires of this type can be analyzed by Latent Class (LC) models in order to classify firms into homogeneous groups according to the perception of corruption. In this paper, we propose a multidimensional framework, based on an LC model, to identify various types of corruption. By using a dataset for transition economies, we identify four classes of corrupt activities, which go beyond the usual classification into administrative and political types of corruption; we then validate our estimates by using a direct administrative corruption index from the same dataset and by comparing, at country level, corruption perception rankings published by Transparency International. The potential of the proposed approach is illustrated through an application to the relationship between firms' competitiveness and the identified latent corruption classes, with evident heterogeneity in the interpretation of results regarding policy implications.

Keywords: Latent class models, multidimensional item response theory, corruption, transition economies.

JEL Classification: C51 C52 D22 D73

1 Introduction

Corruption is one of the most important, and yet complex, issues facing a large number of countries throughout the world[1]. Several papers have paid increased attention to the underlying causes and consequences of corruption, as well as to assessment of political action in the pursuit of its reduction, which affects the economic development of a country (see, for a review, Harstad and Svensson 2008; Lambsdorff 2007). However, little attention has been given to the methodological question of how corruption is estimated and whether any such results can bias empirical facts and policy suggestions[2]. The lack of research is surprising, as the answer to this question is critical to an understanding of why corruption persists at all, and why it appears to be increasing over time in some parts of the world.

This paper analyses corruption, with particular focus on identifying the latent traits of this multidimensional phenomenon. Corruption mostly occurs in a latent manner. According to Svensson (2005), corruption is linked with a country's legal, economic, cultural and political institutions, and is a response to either beneficial or harmful regulations. Corrupt activities may arise in response to benevolent regulations when individuals pay bribes to avoid penalties for harmful conduct or when monitoring is insufficient[3]. In any case, corruption is a crime which goes unnoticed by its victims, and offenders who have no incentive to make their corrupt transactions public, conducting them clandestinely (Graeff 2005).

We exploit item responses from the Business Environment Survey 2002 (BEEPS), potentially linked with corrupt practices, and examine whether a heterogeneous framework of corruption arises across countries of Eastern Europe and Central Asia. We examine whether the dimensional structure of corruption can help to identify latent classes and whether they can explain the weight of different types of corruption at country level. We use this sample of microdata because it includes countries which have known heterogeneous corruption levels and which was experiencing transition toward freer markets.

Our paper is in line with the criticism of the excessively narrow current definition of

[1]Rose-Ackerman (2004) estimates that total bribes in a year are about 3% of the world GDP.

[2]For a discussion on the problematic interpretation of the corruption index, see Dreher and Schneider (2006).

[3]Conversely, Djankov *et al.* (1977) identify the probable increase of corruption in bad policies or inefficient institutions.

corruption, characterized by the misuse of public power for private benefit[4]. Svensson (2005) argues that this definition leaves open so many conceptual ambiguities. For example, the term "private benefit" describes receiving money or assets of value, but may also involve an increase in political power or status, some kind of advantage from receiving promises for future favors, or benefits for close persons (nepotism or favoritism). In addition, public power is exercised by both bureaucrats and politicians in a large variety of sectors and behaviors (Lambsdorff, 2007).

While economic literature in recent years has assessed the effects of corruption, the existing evidence on exactly what corruption measures (or how it is measured) is sparse. There are two lines of research: one states that corruption is a distinct part of the quality of governance. By using different outcomes of the Worldwide Governance Indicators (WGI), Licht *et al.* (2007) interpret separate regressions as additional evidence that the quality of governance differs from the concept of corruption. The second line focuses on the components of corruption. Beyond mere corruption indicators, at least two substantial categories of public corruption exist, defined as *political corruption* and *administrative corruption* (Bardhan, 1997, 2006; Warren, 2004). However, the difficulty in building these indicators - for example, the corruption perception index (CPI) of Transparency International (TI) - is that their aggregation from numerous sources is essentially a summary of row-specific indices linked with corruption, with no exploration of the dimensionality of the underlying data (Thomas, 2007)[5]. These criticisms may also be extended to the argument of Mocan (2008), who stated that corruption is strictly a latent phenomenon and that this feature indicates the need for a statistical identification. As a novelty in this literature, Neumann and Graeff (2010) showed the existence of multi-traits in corruption, so that the endogeneity issues in empirical studies could be reinterpreted in regression specifications as a question of dimensionality.

Based on latent class (LC) analysis (Goodman, 1974, 1978) and Item Response Theory (IRT) (Hambleton, 1996), in this work we use the statistical methodology proposed by Bartolucci (2007) to characterize the role played by the dimensionality of the items in the questionnaire and to identify the heterogeneous components of corrupt activities. This meth-

[4]For a discussion of various definitions of corruption, see Lambsdorff (2007).

[5]See the discussion of Andersson and Heywood (2009) on the use and abuse of Transparency International's approach to measuring corruption.

odology is called multidimensional IRT (M-IRT). The statistical model is used to extract the dimensional structure correlated to the latent phenomenon (corruption) from a set of items; this framework is then used to identify latent classes and determine corruption quantitatively. The M-IRT model includes the well-known multidimensional Rasch model (Rasch, 1961) as a restricted case.

There are alternative techniques for inferring a measurement scale of corruption from a list of pre-ordered (or pre-classified) items of the questionnaire. One basic approach is the sum score technique, which consists simply of a weighed - or not weighed - summing up of the indicators. Factor analysis techniques are also widely used to check whether a set of items fits a unidimensional measurement scale. Factor analyses are performed by examining the pattern of correlations (or covariance) between the observed measures. Measures which are highly correlated (either positively or negatively) are probably influenced by the same factors, whereas those which are relatively uncorrelated are probably influenced by different factors[6]. However, as a main advantage, the M-IRT model allows us to generate a consistent measurement index jointly with estimating its determinants. In particular, the items used for generating the index are selected on the basis of their reliability and their ability to describe predominant traits.

The IRT model has already been applied to other research areas. It was initially used in psychometrics and educational testing to investigate the latent traits of ability/achievement tests (see Bock, 1997). However, our paper also draws on literature increasingly being applied to social and economic measures containing items, which are scored in dichotomous or polytomous fashion. Cappellari and Jenkins (2007) applied the IRT model to account for several issues concerning the derivation of deprivation scales of widely used sum-score deprivation indices. Similar methods were used by Kuklys (2004) to analyze housing and health functioning and Faye *et al.* (2009) used a set of food insecurity indicators to derive a food deprivation scale. We also based our model on the public health applications of Bartolucci *et al.* (2010, 2012), who used M-IRT to identify the health status latent traits of elderly patients who currently receive healthcare assistance in Italy.

[6]A principal components analysis of political systems of corruption related to democratization levels of American state is investigated, for example, in Hill (2003).

Using a sample of more than 6600 enterprises in 26 countries of Eastern Europe and Central Asia taken from the BEEPS survey, we assess whether the latent classes of corruption are useful in restricting the field under investigation to more specific aspects of corruption, which characterize the endemic level in a specific country context. We focus particularly on the assignment of each firm at country level, because our (four) latent classes express a growing quantitative contribution to the corruption level of a country. Our findings also indicate the robust association of our identified latent classes of administrative and political corruption with the indexes of (mainly) administrative corruption derived from BEEPS and TI, respectively. As an illustrative example, we use the magnitude of the firm scores to estimate the effects of firms' competitiveness on the probabilistic response of corruption yielded by the identified latent classes.

The paper is organized as follows. Section 2 describes the evolution of corruption in transition economies and illustrates the dataset we use. Section 3 discusses our empirical framework for quantifying the number of latent classes and presents our main results. Section 4 provides the robustness of these results, comparing the aggregate indices of the firms' components of corruption with those extracted directly from BEEPS and TI. An illustrative example is reported in Section 5, in which we review the relationship between firms' competitiveness and corruption using the identified corruption components. Some conclusions are then drawn in Section 6.

2 Some basic facts and data used to explore corruption in transition economies

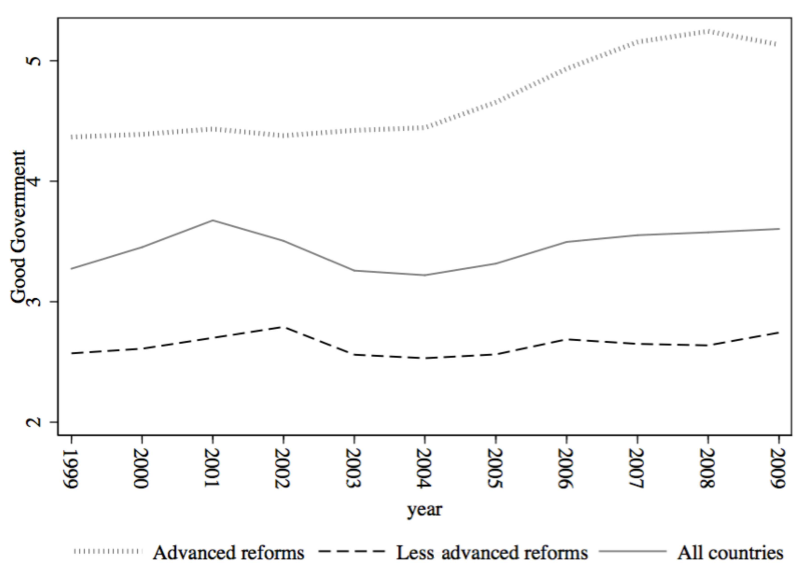
In this section, we first describe the patterns of corruption in transition economies by using the country indices derived from the relevant literature and provided by international organizations. We then discuss the items of the survey, making a distinction between *administrative corruption* and *political corruption.*

2.1 Preliminaries: corrupt attitudes in transition economies

Since 1990, a large number of post-communist countries of Eastern Europe have undergone the transition to a market economy (i.e., transition economies). These countries were characterized by a propensity toward economic reform, focused on macroeconomic stabilization, price and trade liberalization, privatization, and the establishment of the legal foundations of a market economy (Anderson and Gray, 2006).

Ex-post economic assessments of these reforms have been contrasting. For example, Falcetti *et al.* (2006) found a positive cross-country correlation between market-oriented reforms and the economic growth rate during the transition process. Conversely, Hodgson (2006) showed that many countries underwent economic recession in the first few years of transition, due to the difficulties in changing institutions quickly. More interestingly for our analysis, Shleifer and Vishny (1993), Johnson *et al.* (1999) and Rose-Ackerman (1999) observed previously a unheard-of increase in corruption in the first decade of transition, correlated with a negative economic cycle, partly explained by an unchanged political class which persisted in following old habits in newly reformed market institutions.

Figure 1: Patterns of corruption in transition economies



Notes: Corruption is inversely expressed in terms of good government and ranges from 0 to 10, where 10 represents highest level of good government.

Figure 1 (solid line) shows the patterns of good government from 1999 to 2009 in a sample of 27 transition economies, which are inversely measured with respect to the scale of corruption perception indices. Good government ranges from 0 to 10, where 10 represents the highest level for this index and is widely used for practical applications at country level (Bardhan, 1997; Nowak, 2001; Svensson, 2005). We exploit the country classification of the institutional reforms in 2000 by Davoodi and Abed (2000) to cluster countries with advanced institutional reforms (dotted line in Figure 1) with respect to countries with less advanced reforms (dashed line in Figure 1)[7].

As expected, Figure 1 shows higher good government (i.e., smaller perceived corruption) in those countries, which underwent an advanced institutional reform season although, at least until 2009, these patterns are stable and do not show particular improvements. They are in line with the argument of Nowak (2001) who explained the persistence of corruption in transition economies according to their Communist roots: that is, although the effects of institutional reforms led to evident improvements in the good government of transition countries, they generally seem to be arrested in the last years.

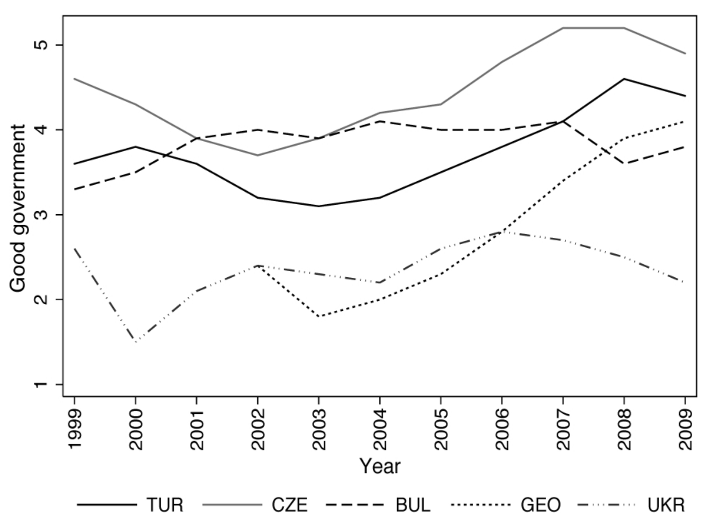
More interestingly, we would know whether good government - or inversely corruption patterns - are associated with the anti-corruption policies, which were applied and largely publicized in some countries. For example, Bulgaria changed important laws to reform the public administration, including a Civil Service Act (Open Society Institute, 2002). The Czech Republic also addressed this problem in its *Combating Corruption Program*, based on three principal anti-corruption measures: *(i)* an educational program to increase awareness of corruption and the ability of civil servants to fight it; *(ii)* an extensive public administration anti-corruption plan, involving the majority of public offices (e.g., police, health services); *(iii)* extensive reform of political parties. We also include Turkey in our sample, even though it is not an economy geographically in transition, because it underwent a transition to the market economy similar to countries in Eastern Europe. In fact, Turkey was obliged to adopt anti-corruption policies as part of the actions within the Emergency Action Plan (Memisolu

[7] We follow Davoodi and Abed (2000) including in the cluster of countries with an advanced institutional reforming season Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovak Republic and Slovenia; all the other transition economies are included in the less advanced cluster of countries. The threshold was chosen on the basis of a BEEPS structural reform index. In particular, their inclusion is based on eight indicators of reforms: i) large-scale privatizations, ii) small-scale privatizations, iii) enterprise restructuring, iv) price liberalization, v) trade and exchange rate systems, vi) competition policies, vii) banking reform, viii) securities markets and non-bank financial institutions.

and Durguin, 2008)[8] to achieve the objectives required for pre-accession to the European Union.

We observe that, with the exception of the Czech Republic, the implemented reforms have mainly focused on *administrative* corruption, and virtually little progress has been made towards fighting *political* corruption at the level of government, Parliament (National Assembly) and political parties in countries subjected to this type of corruption. Note that, in this context, one way of fighting political corruption is through the voting process, replacing the corrupt government or "revolutions"under the flag of legality. In the latter case, interesting examples of corrupt governments in transition economies removed after accusations of corruption have been the "Rose Revolution" in Georgia and the "Orange Revolution" in Ukraine.

Figure 2: Patterns of corruption in countries which applied anti-corruption policies



Notes: Corruption is inversely expressed in terms of good government ranges from 0 to 10, where 10 represents highest level of good government. Before 2002 no data were available for Georgia (Source, Transparency International, various years).

Figure 2 refers explicitly to those countries which accepted anti-corruption reforms, with data based on CPI. The figure shows that improvements in good government were only mod-

[8] The structure of Emergency Action Plan is based on: i) the creation of new institutions to establish the principle of the state of law and to protect individual rights, ii) enforcement of access to information and documents held by the administration in public management, and iii) the rise in ethical standards in public administration, through the prevention of gifts or advantages offered, information supply, transparency, participation, accountability of managers, and declaration of property.

erate. In particular, the reforms were totally ineffective in combating corruption in Bulgaria whereas, in a comparison between Georgia and Ukraine, Figure 2 shows an improvement in good government in Georgia, and a persistence of the pre-communist situation in the level of corruption in Ukraine.

This preliminary evidence emphasizes the idea of corruption as a complex and latent phenomenon, which involves government institutions at different stages of public duties. This limits the effectiveness of anti-corruption policies which can identify and intervene in the various sources of corruption.

2.2 Data by firm survey

Our analysis mainly uses data on corrupt attitudes perceived by firms, although we refer to CPI data on the occurrences. As anticipated, the data come from BEEPS, jointly implemented by the European Bank for Reconstruction and Development (EBRD) and the World Bank. Surveys were carried out yearly from 1999, focusing on several substantial issues affecting firms in Eastern Europe and Central Asia. For example, the surveys included questions about firms' sales and their investment, innovations, and access to financing. In addition, several questions regarded law and business regulation, taxation, and the qualitative perception of the business environment.

A section of BEEPS also addressed government policies and practices and includes items of perceived corruption. Note that one of the most interesting features of the survey is the sample design used to collect the data. Based on the perception of managers and related to the line of business in which they operate, these subjective measures of corrupt practices covered the direct experience of the interviewees, limiting measurement bias which depends on unverified knowledge (e.g., hearsay)[9]. The survey also contains firms' characteristics, such as sector and size in terms of employees.

To verify the existence of various types of corruption, we used data from the survey carried out in 2002, because the questionnaire for that year does not contain many missing in items linked with the problem of corruption. Our dataset concerns 6667 firms in 26 countries (see, Table 1).

[9]For a discussion of this topic, see Fries *et al.* (2003).

Table 1: List of countries and firms

Country	Number of firms	Percentage of firms
Yugoslavia	250	3.75
FYROM	170	2.55
Albania	170	2.55
Croatia	187	2.80
Turkey	514	7.71
Bosnia and Herzegovina	182	2.73
Slovenia	188	2.82
Poland	500	7.50
Ukraine	463	6.94
Belarus	250	3.75
Hungary	250	3.75
Czech Republic	268	4.02
Slovak Republic	170	2.55
Romania	255	3.82
Bulgaria	250	3.75
Moldova	174	2.61
Latvia	176	2.64
Lithuania	200	3.00
Estonia	170	2.55
Georgia	174	2.61
Armenia	171	2.56
Kazakhstan	250	3.75
Azerbaijan	170	2.55
Uzbekistan	260	3.90
Russia	506	7.59
Tajikistan	176	2.64
Kyrgyz Republic	173	2.59
Total	6,667	100.00

Because our aim was to characterize the various dimensions of corruption from the questionnaire items, we exploited the potential of this questionnaire, which distinguishes *a priori* two types of corruption linked with *administrative* and *political* futures, as proposed in the economic literature starting from Scott (1972) and developed by Bardhan (1997, 2006) and Warren (2004). As described in Fries *et al.* (2003), political corruption describes how firms influence the content and application of specific laws and regulations to the benefit of a narrow private interest, rather than the broad public interest; administrative corruption is often associated with the arbitrary application of existing laws and regulations.

The questionnaire items classified by Fries *et al.* (2003) is listed in Table 2. Looking at the definition of administrative corruption, we select ten items covering a wide range of futures related with corrupt practices. Two items account for corruption in public services and permits, which are closely related with the fixed costs of doing business. A second set of items is linked with bribes, with the aim of weakening the activities of public inspections within a firm or related to inspections of buildings, health and environmental safety. The remaining sub-set of items describes informal payments required to deal with the public administration, imports and customs, courts and tax collection.

As regards political corruption, we present a set of six items[10]. In particular, managers responded concerning private payments or gifts, made with the aim of affecting the votes of parliamentarians and government officials on specific laws, the contents of government decrees, or decisions of elected officials. In addition, private payments or gifts were considered to influence the decisions of criminal and civilian courts, together with benefits to central bank officials in influencing central bank policies and decisions.

Table 2: Description of dataset

How often would they make payments/gifts for the following purposes? (proxies of **administrative corruption**):		
item01	(1)	to get connected to and maintain public services (e.g., electricity and telephone lines)
item02	(2)	to obtain business licenses and permits
item03	(3)	to obtain government contracts
item04	(4)	to deal with occupational health and safety inspections
item05	(5)	to deal with fire and building inspections
item06	(6)	to deal with environmental inspections
item07	(7)	to deal with taxes and tax collection
item08	(8)	to deal with customs/imports
item09	(9)	to deal with courts
item10	(10)	to influence the contents of new legislation rules, decrees, etc.
To what extent have the following practices had a direct impact on your business? (proxies of **political corruption**):		
item11	(11)	private payments/gifts or other benefits to parliamentarians to affect their votes
item12	(12)	private payments/gifts or other benefits to government officials to affect the content of government decrees
item13	(13)	private payments/gifts or other benefits to judges to affect the decisions of criminal court cases
item14	(14)	private payments/gifts or other benefits to judges to affect the decisions in commercial cases
item15	(15)	private payments/gifts or other benefits to central bank officials to affect central bank policies and decisions
item16	(16)	illegal contributions to political parties and/or election campaigns to affect the decisions of elected officials

Notes: Data come from Business Environment Survey (BEEPS), jointly implemented by data from the European Bank for Reconstruction and Development (EBRD) and the World Bank.

3 Empirical framework

In this section, we discuss an LC model which accounts for the dimensionality issue of corruption. Based on the constrained version of the LC model, we view corruption as a

[10]See also Svensson (2005) for a discussion on the definition of political corruption.

latent phenomenon identified by the conditional probabilities of success of each item, as the respondent is in a certain latent class. Statistical information criteria are mainly used in the choice of the number of latent classes, and the estimates are based on the BEEPS dataset.

3.1 The model

We describe the M-IRT model proposed by Bartolucci (2007). Let y_j denote the response variable for the $j-th$ item of the questionnaire, which corresponds to the case of a binary item ($y_j = 0, 1$). We denote by n the number of firms in the sample and assume that they respond to J items which measure D different latent traits and that every item measures only one latent trait[11]. This model allows us to investigate the correlation between the latent dimensions of corruption. Here, $n = 4610$ and $J = 16$, and on the basis of the items defined in the dataset, we expect at least two different dimensions, so that $D \geq 2$.

The model we adopt is based on the following parameterization of the conditional probabilities of success $\lambda_{j|c}$ as a logit function:

$$\text{logit}(\lambda_{j|c}) = \gamma_j \left(\sum_d \delta_{jd}\theta_{cd} - \beta_j \right), \qquad j = 1, \ldots, J, \tag{1}$$

where, with reference to item j, $\left|\lambda_{j|c}\right| = |(p(y_j)|\ firm\ is\ in\ class\ c)|$ and δ_{jd} is a dummy variable equal to 1 if $j \in \mathcal{J}_d$ and to 0 otherwise, $\mathcal{J}_d$ being the subset of items measuring dimension d; θ_{cd} is a measure of the latent trait (dimension d) for subjects in latent class c, and β_j indicates the *difficulty* parameter as the overall tendency to respond 0 to item j (for a discussion, see Bartolucci 2007). Parameters γ_j may be set equal to a fixed value following one-parameter logistic parameterization (1PL) or left unconstrained and thus estimated as in two-parameter logistic parameterization (2PL). In the second case, the model allows for a different sensitivity of the item measuring the latent trait. The relative importance of the difference between a firm's trait level and item threshold is therefore determined by the magnitude of the *discriminatory* power of the item.

[11]When the number of partitions J_d is equal to the number of used items (to define the latent trait), the IRT model is equivalent to an LC model.

Under the assumption of *local independence*[12], the distribution of $\boldsymbol{Y}$ for subjects in the $c-th$ latent class is:

$$p(\boldsymbol{y}|c) = p(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\Theta} = \boldsymbol{\theta}_c) = \prod_j \lambda_{j|c}^{y_j}(1-\lambda_{j|c})^{1-y_j}, \tag{2}$$

where $p(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\Theta} = \boldsymbol{\theta}_c)$ is the conditional probability that a subject with latent vector $\boldsymbol{\theta}$ provides to response configuration $\boldsymbol{y}$. Through a finite mixture, we can express the distribution of $\boldsymbol{Y}$ as:

$$p(\boldsymbol{y}) = \sum_c p(\boldsymbol{y}|c)\pi_c, \tag{3}$$

where $\pi_c = p(\boldsymbol{\Theta} = \boldsymbol{\theta}_c)$ are the weights corresponding to each latent class.

The log-likelihood function, which is used to estimate the above multidimensional LC Rash model, is thus:

$$\ell(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}} n(\boldsymbol{y})\log(p(\boldsymbol{y})). \tag{4}$$

where $\boldsymbol{\theta}$ is the vector containing all identifiable parameters of the model and $p(\boldsymbol{y})$ is computed as a function of $\boldsymbol{\theta}$. In particular, to make the parameters identifiable, we use the constraint $\beta_j = 0,\ j \in \mathcal{D}$, when the parameterization is of 1PL type, and $\beta_j = 0, \gamma_j = 1,\ j \in \mathcal{J}_d$ when it is of 2PL type, where $D = j_1, ..., j_d$, and j_d denotes a specific element of $\mathcal{J}_d$.

The maximization of $\ell(\boldsymbol{\theta})$ may be performed by the EM algorithm (Dempster *et al.*, 1977). See also Bartolucci (2007) and Bartolucci *et al.* (2012), for details. Let us assume that we know frequencies $m(\boldsymbol{y}, c)$ of the contingency table in which these subjects are cross-classified according to response configuration $(\boldsymbol{y})$ and latent class (c). We can thus estimate $\boldsymbol{\theta}$ by maximizing the so-called *complete* log-likelihood as:

$$\ell^*(\boldsymbol{\theta}) = \sum_{\boldsymbol{y}}\sum_c m(\boldsymbol{y}, c)\log p(c, \boldsymbol{y}). \tag{5}$$

Therefore, at the E-step, we compute the expected value of $\hat{m}(\boldsymbol{y}, c)$ for each $\boldsymbol{y}$ and c, given $n(\boldsymbol{y})$, and at the M-step we maximize the *complete* log-likelihood, in which every frequency

[12] Following the *local independence* assumption, the response variables are conditionally independent, given the latent variables.

$m(\boldsymbol{y}, c)$ is replaced by the corresponding expected value $\hat{m}(\boldsymbol{y}, c)$. The EM alternates these two steps until convergence.

The estimated parameters from the EM algorithm are then used to compare (two) multidimensional models. The hypothesis test is of type:

$$H_0 : \boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0} \tag{6}$$

where $\boldsymbol{g}(\boldsymbol{\theta})$ is a vector-valued function and $\mathbf{0}$ denotes a column vector of zeros of suitable dimensions. For testing the best model for the data at hand, we use the likelihood ratio (LR) test statistic,

$$LR = 2 \sum_{\boldsymbol{y}} n(\boldsymbol{y}) \log \left[\frac{\hat{p}_D(\boldsymbol{y})}{\hat{p}_{D-1}(\boldsymbol{y})} \right] \tag{7}$$

where $\hat{p}_D(\boldsymbol{y})$ refers to the assumed model with D different dimensions and $\hat{p}_{D-1}(\boldsymbol{y})$ to the constrained model with $D-1$ dimensions. When the response probability is modeled by an M-IRT model, the resulting LR statistic has a χ^2 distribution with a number of degrees of freedom equal to the difference between the number of parameters of the full multidimensional and restricted models, where we merge two distinct dimensions.

A crucial point for the identification of latent traits of corruption is the strength of the correlation between two distinct dimensions. We compute this correlation as:

$$\hat{\rho}_{d_1,d_2} = \sum_{c=1}^{C} \hat{\theta}_{c,d_1} \hat{\theta}_{c,d_2} \pi_c \tag{8}$$

where θ_{c,d_1} and θ_{c,d_2} are the standardized estimates of the latent trait referring to two specific dimensions, d_1 and d_2, for subjects in latent class c (e.g., $\hat{\pi}_c$ is the estimated weight of this class with $c = 1, ..., C$). After identifying the latent classes on the basis of $\hat{\theta}_{cd}$ and $\hat{\rho}_{d_1,d_2}$, as a final step, we compute the firm's trait scores by using the expected value of frequencies $\hat{m}(\boldsymbol{y}, c)$ estimated at the E-step, defined as:

$$\hat{m}_{\boldsymbol{y}|c} = n(\boldsymbol{y}) \frac{p(\boldsymbol{y}|c)\pi_c}{\sum_h p(\boldsymbol{y}|h)\pi_h}. \tag{9}$$

We then build a binary variable, which identifies latent classes and attributes 1 to the firm with the highest expected value of frequencies in that class, and 0 otherwise.

3.2 Model selection

The choice of the number of classes in the multidimensional LC Rash model is crucial for model identification. Although selection is mainly based on information criteria, we complement this choice with indices measuring goodness of fit. The most frequently used index is the Bayesian information criterion (BIC) (Schwarz, 1978), which is based on the index:

$$BIC = -2\hat{\ell} + m\log(n), \tag{10}$$

where $\hat{\ell}$ is the maximum value of log-likelihood test statistics and m is the number of free parameters. Another widely used criterion is the Akaike information criterion (AIC), (Akaike, 1973), which is based on the index:

$$AIC = -2\hat{\ell} + 2\log(n), . \tag{11}$$

Two extended versions of the AIC index were proposed by Andrews and Currim (2003), which include different weights in estimating the log-likelihood function. In the first case, AIC is penalized with a factor 3 instead of 2 ($AIC3$); penalties in AIC are included in models with a larger number of parameters to define a consistent Akaike information criteria (e.g., $CAIC$).

We also use measures based on the capacity of the model to fit the data. These measures are based on $R^2_{entropy}$ and variance $R^2_{variance}$ (Magidson and Vermunt, 2001), the estimated proportion of classification errors (E)[13], and the Average Weight of Evidence (AWE). In particular, the last index is built by adding a third dimension to the BIC index that, makes the performance of individual classification within groups more efficient, as argued in Banfield and Raftery (1993). The formal specification is:

$$AWE = -2\hat{\ell}^c + 2m\left[\frac{2}{3} + \log(n)\right] \tag{12}$$

[13]Unlike the other measure, E proves model accuracy when the values are close to zero.

where $\hat{\ell}^c$ is the log-likelihood classification (Biernacki and Govaert, 1998). Table 3 lists the log-likelihood statistics and classification statistics for a number of predetermined latent classes, from 2 to 8.

Table 3: Latent class selection

	Log-likelihood statistics				
	Log-likelihood (log LL)	BIC (log LL)	AIC (log LL)	AIC3 (log LL)	CAIC (log LL)
Class1	-46461.658	93058.2930	92955.3173	92971.3173	93074.2930
Class2	-40417.9682	81114.3238	80901.9364	80934.9364	81147.3238
Class3	-38985.0967	78391.9925	78070.1933	78120.1933	78441.9925
Class4	-382.66.8741	77242.3707	76701.7482	76785.7482	77326.3707
Class5	-38204.7666	77261.5674	76611.5331	76712.5331	77362.5674
Class6	-38404.3510	77373.9129	76942.7020	77009.7020	77440.9129
Class7	-38171.2934	77338.0329	76578.5869	76696.5869	77456.0329
Class8	-38148.8201	77436.4980	76567.6402	76702.6402	77571.4980
	Goodness of LC model fit				
	E	$R^2_{entropy}$	Standard $R^2_{variance}$	AWE l^c	AWE index
Class1	0.0000	10.000	10.000	-46461.6586	93241.2687
Class2	0.0374	0.8712	0.8922	-40829.3882	82314.5512
Class3	0.0532	0.8630	0.8754	-39674.2046	80342.0076
Class4	0.0613	0.8593	0.8633	-39214.4170	79760.2558
Class5	0.1283	0.7791	0.7547	-39787.7568	81244.7587
Class6	0.1432	0.7623	0.7295	-39945.4201	81897.9089
Class7	0.1845	0.7214	0.6728	-40429.0881	83203.0681
Class8	0.1974	0.7070	0.6541	-40630.1816	83943.0787

Notes: Upper part of table lists Log-likelihood statistics and estimates Log-likelihood function. Proposed Log-likelihood statistics are: Bayesian Information Criteria (BIC), Akaike information criterion (AIC), Akaike information criterion with 3 as penalizing factor (AIC3), and Consistent Akaike information criterion (CAIC). Bottom part of table lists several measures and statistics based on ability of model to fit data, and include: estimated proportion of classification errors (E), two pseudo R-square measures ($R^2_{entropy}$) and ($R^2_{variance}$), and Average Weight of Evidence (AWE).

This analysis strongly suggests that the optimal number of latent classes is 4, as shown by the BIC index and CAIC. This result is also confirmed by the measures of goodness of fit, reported at the end of the table. We note that the estimated proportion of classified error E doubles passing from 4 to 5 latent classes, whereas the two pseudo R^2 rapidly decrease from 4 to 5 classes. Consistently, the AWE statistic reports a minimum value at 4 classes.

3.3 Results

We now summarize the main results of the multidimensional approach to corruption. As described in the previous section, the procedure consists of running a sequence of nested models which begins with an LC model, in which the number of dimensions is equal to the number of selected items (unconstrained model). An analytical overview of the results from the hierarchical clustering analysis is given in Table 4. The third column lists the combinations of items for each of the sequential steps; the fourth and fifth columns list the deviance from the initial LC model and from the previous model, respectively. Lastly, the $p-$value of LR test statistics appears in the last column.

Table 4: Output of hierarchical cluster analysis

Seq.	Dim.	Clusters	Deviance	LR-test	P-value
1	15	1.3.4.5.6.7.9.10.11.12.13.14.15.16.(2.8)	0.194	0.194	0.908
2	14	1.3.4.5.6.7.11.12.13.14.15.16.(2.8).(9.10)	0.500	0.306	0.858
3	13	1.3.4.7.11.12.13.14.15.16.(2.8).(9.10).(5.6)	1.109	0.609	0.737
4	12	1.3.4.7.11.12.14.16.(2.8).(9.10).(5.6).(13.15)	1.789	0.680	0.712
5	11	3.4.11.12.14.16.(2.8).(9.10).(5.6).(13.15).(1.7)	2.690	0.901	0.637
6	10	3.4.11.12.14.(2.8).(9.10).(5.6).(1.7).(13.15.16)	3.939	1.249	0.535
7	9	3.4.11.12.14.(9.10).(5.6).(13.15.16).(1.2.7.8)	6.185	2.246	0.325
8	8	3.4.14.(9.10).(5.6).(13.15.16).(1.2.7.8).(11.12)	9.666	3.481	0.175
9	7	3.4.(9.10).(5.6).(1.2.7.8).(11.12).(13.14.15.16)	13.296	3.630	0.163
10	6	3.(9.10).(4.5.6).(1.2.7.8).(11.12).(13.14.15.16)	18.088	4.792	0.091
11	5	(9.10).(4.5.6).(1.2.3.7.8).(11.12).(13.14.15.16)	23.668	5.580	0.061
12	4	(9.10).(4.5.6).(1.2.3.7.8).(11.12.13.14.15.16)	42.936	19.268	0.000
13	3	(4.5.6).(1.2.3.7.8.9.10).(11.12.13.14.15.16)	84.831	41.895	0.000
14	2	(1.2.3.4.5.6.7.8.9.10).(11.12.13.14.15.16)	207.149	122.318	0.000
15	1	(1.2.3.4.5.6.7.8.9.10.11.12.13.14.15.16)	3059.953	2852.805	0.000

Notes: Column (1) lists the sequence of steps (seq), (2) the number of dimensions for each step (dim), and (3) combinations of items in each dimension for each sequential step (Clusters). Deviance for each step (Deviance), Likelihood ratio test (LR-test), and $p-values$ are presented in last three columns. For classifications and definitions of items, see Table 2.

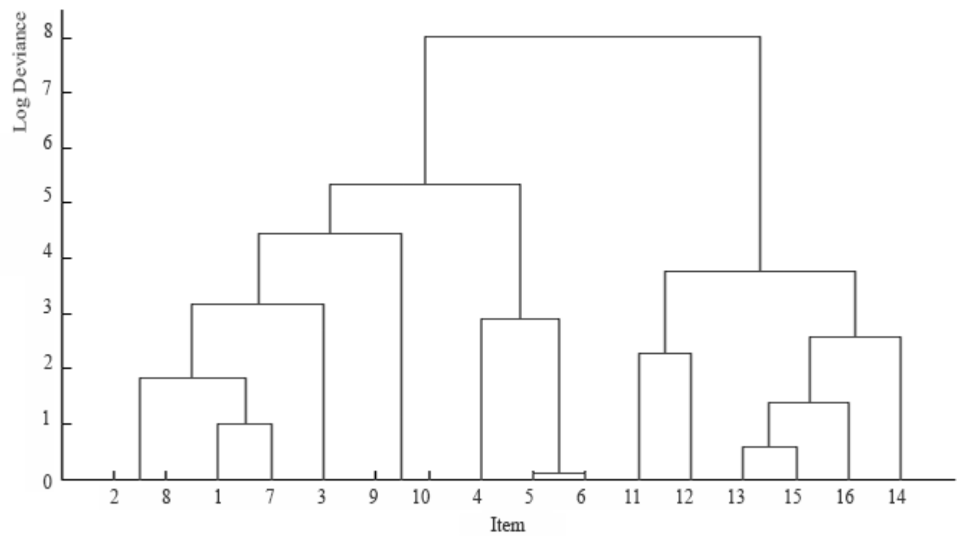
Table 4 shows the presence of multidimensionality, keeping five significant dimensions, as also confirmed by the hierarchical clustering analysis in Figure 3. Differences within types of corruption arise from the items of administrative corruption which characterize dimensions 1, 2 and 3; items of political corruption are those of dimensions 4 and 5.

Looking at the item definitions (Table 2), we can progressively identify the dimensions of administrative corruption as:

- bureaucrats' need to influence legislation rugulations and the timing of applications of law;
- unofficial payments for inspections in occupational health and safety, fire and buildings, and environmental works;
- application of government contracts, business licenses and contracts, public services and general issues of taxation.

The items that cluster political corruption are:

Figure 3: Dendrogram



Notes: Classifications and definitions of items are listed in Table 2.

- corruption by firms to influence the contents of specific laws and regulations;
- corruption to influence decisions of criminal courts and commercial cases, and decisions or policies of the officials of the central bank.

Table 5: Estimation of conditional probabilities and discriminatory power of items of selected model

Dim.	Item	Class1	Class2	Class3	Class4	γ
3	item1	0.044	0.194	0.433	0.777	1.000
3	item2	0.102	0.417	0.723	0.934	1.118
3	item3	0.055	0.248	0.528	0.848	1.057
2	item4	0.031	0.144	0.644	0.881	1.000
2	item5	0.046	0.205	0.735	0.919	1.001
2	item6	0.012	0.080	0.596	0.887	1.192
3	item7	0.075	0.371	0.705	0.937	1.208
3	item8	0.047	0.226	0.507	0.842	1.086
1	item9	0.011	0.206	0.371	0.877	1.000
1	item10	0.011	0.139	0.246	0.730	0.850
4	item11	0.005	0.657	0.035	0.731	1.000
4	item12	0.008	0.738	0.055	0.799	0.979
5	item13	0.002	0.649	0.038	0.822	1.000
5	item14	0.008	0.800	0.113	0.900	0.895
5	item15	0.002	0.499	0.035	0.686	0.861
5	item16	0.008	0.613	0.081	0.759	0.752

Notes: Column (1) lists the dimension for each item (dimension), (2) item code (item). Conditional probabilities and the discriminatory power (γ) of items of selected model. For classifications and definitions of the items, see Table 2.

Table 5 lists the estimated conditional probabilities of the LC model for each selected item, together with the estimates of γ. It shows that there is a dissimilar order between the first 10 items (administrative corruption) and the last 6 (political corruption). For a clearer interpretation, we refer to the components of *breadth* and *depth* of corruption as defined by Mocan (2008), in which the former describes the extent to which corruption is widespread in a country, and the latter describes the extent of each LC component affecting corruption. The results indicate that administrative corruption items follow an increasing order across the four latent classes, and the last six political corruption items show higher conditional probabilities, regarding latent class 2 and, partly, latent class 4. In addition, we find a lower level of the conditional probabilities in latent class 1, which achieves a (modest) influence in items 2 and 7 (see Table 5), associated with administrative corruption. This implies that the estimated level of corruption in latent class 1, mainly identified by tax evasion or improving access to public services, does not seem to be an obstacle for the business environment. That is, although we postulate that administrative corruption imposes a burden on firms in terms, for example, of tax bribes, estimates by conditional probability of a low corruption enables us to identify this latent class as referring to unconstrained *physiological level of administrative corruption.*

Thus, as emerges from this discussion, a crucial point of the present study is the interpretation of latent classes. Useful suggestions for this come jointly from the estimate of dimensions on different classes and from the correlation measures between identified dimensions (Table 6).

Table 6: Estimation of θ and correlation among dimensions

	Dim1	Dim2	Dim3	Dim4	Dim5	Weight
Impact of dimensions on classes						
Class1	-3.072	-3.027	-4.498	-6.372	-5.326	0.439
Class2	-1.427	-1.352	-1.350	0.615	0.651	0.108
Class3	-0.269	1.020	-0.526	-3.233	-3.311	0.313
Class4	1.247	2.425	1.961	1.527	1.000	0.140
Correlation among dimensions						
	Dim1	Dim2	Dim3	Dim4	Dim5	
Dim1	1.000					
Dim2	0.989	1.000				
Dim3	0.990	0.959	1.000			
Dim4	0.747	0.645	0.834	1.000		
Dim5	0.676	0.563	0.773	0.994	1.000	

Notes: First part of table lists standardized estimate of latent trait level ($\theta_{c,d}$) for each class c and dimension d, with estimated weight in each class (π_c). The second part lists correlation among dimensions.

The analysis identifies three main results. First, estimated parameter θ_{cd} in equation (1) formally identifies the four latent classes according to their different dimensions. As shown in the top part of Table 6, within the four latent classes, we confirm that all five dimensions are increasingly ordered, with the lowest values in class 1 and the highest in class 4. This also means that we can order corruption according to *depth* of the latent classes.

Second, as a brief index of how widespread corruption is in firms, we observe that most firms (43.9%) belong to class 1, followed by class 3 (31.3%), whereas the smallest class is class 2, with only 10.8%. The remaining firms are in class 4 (14%).

Third, looking at the values of the dimensions in each class, we identify class 3 as the one with issues in administrative corruption, and class 2 with characteristics of political corruption. In more detail, and according to the results shown in Table 5, since dimension 2 prevails in class 3, we can identify this LC as administrative corruption, as more linked with public inspections, other than taxes, licenses and permits. In addition, although the items of political corruption may partly affect the results of class 4, the prevalence in that class of the effects of administrative corruption dimensions 1, 2 and 3 shows that the highest level of corruption is more associated with general administrative corruption. Figure 4, summarizes the identification process derived from the empirical framework.

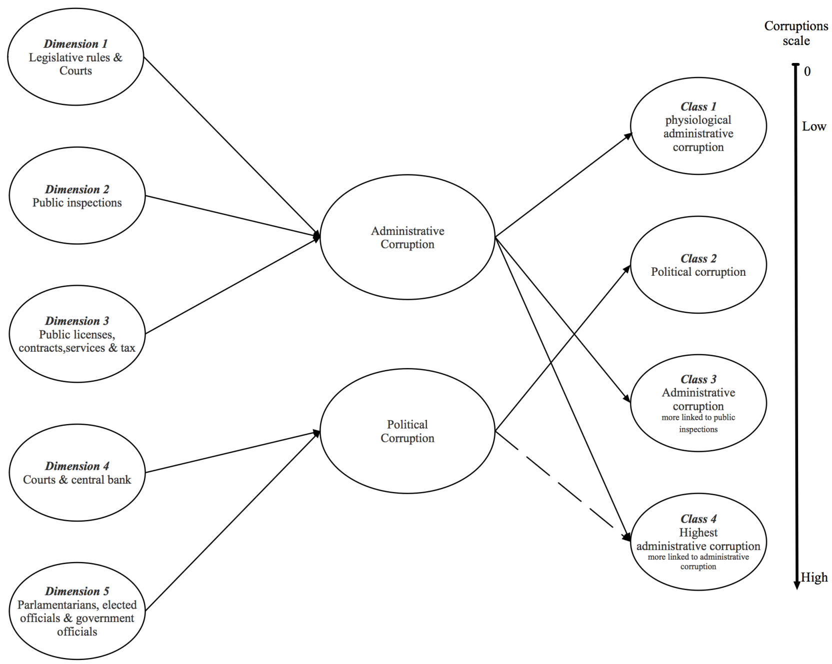
4 Confirmatory analysis

In this section, we ascertained the robustness of the results by comparing our country corruption rankings with those obtained from perception measure of administrative corruption extracted from a direct item of the BEEPS and CPI ranking.

We use the expected value of posterior probabilities, as estimated by the *E-step*, to define a binary variable, where each firm is attributed 1 in the latent class for the highest level of the expected value of probability, and zero otherwise. This new variable identifies the perceived prevalence of corrupt practices for a firm classified in one of the LCs.

This analysis not only allows us to identify different types of corruption and their quantitative influence but also, in line with the definition of breadth of corruption by Mocan (2008), to ascertain how widespread corruption is in a country, and thus to evaluate its importance in Eastern European countries. For this aim, we use our previous firms' binary variable to

Figure 4: Identification of corruption indices



Notes: Analytic description of latent classes and dimensions, according to results of Table 6. Continuous vertical arrow (right) measures scale of corruption, from 0 (absence of corruption) to high.

build a corruption index for each LC at country level.

4.1 Comparison with synthetic BEEPS corruption index

We compare our estimates with the perceived index of corruption, directly obtained from the BEEPS dataset; in this item, firms respond to the following question: *Is it common for firms in your line of business to have to make irregular additional payments or gifts to get things done as regards customs, taxes, licenses, regulation services etc.?*. The answer includes six modalities on a growing scale of perceived corruption (*Never*, *Seldom*, *Sometimes*, *Frequently*, *Usually*, *Always*). However, for empirical purpose, we rescale it into four modalities, considering both the responses *Seldom* and *Sometimes* and *Frequently* and *Usually*; the new corruption index is thus mentioned as *absence*, *low*, *medium* and *high*, and increases the quantitative perception of corruption in line with the estimated LC. Note that this synthetic variable is a measure of administrative corruption, as we expect a strong association with administrative LCs (1, 3 and 4), and a non-significant association with the class identified as political corruption.

Table 7 lists the results of the LR test statistic which assumes an independence hypothesis between each estimated LC and that concerning administrative corruption, directly obtained from the BEEPS for each country.

Our estimates of LCs 1 and 4 appear to be very well associated with the directly perceived administrative corruption index. We show the novelty of the proposed corruption measure characterizing other latent classes identified. In the class 3, the data reject the independence hypothesis at 5% level between our firm score variable and that obtained using *medium corruption*, extracted directly from the BEEPS. Only in a few countries, we can note a non-significant association. This result strengthens our findings, indicating that, although the index built with a direct item of the perceived administrative corruption in BEEPS, its specificity generates the clear-cut effect of public inspections and tax evasion in some countries, with respect to general administrative corruption (e.g., Bulgaria, Croatia, Hungary, Latvia, Uzbekistan and Yugoslavia). Finally, column 2 lists the result of a "false experiment" specification in which *low administrative corruption*, as extracted from the BEEPS, is expected independent from political corruption variable estimated in LC 2. We find confirmation that administrative corruption and political corruption are not related, and that, independently from the level of corruption, these two components affect country corruption differently.

4.2 Comparison with CPI

There is substantial debate regarding the use of corruption indices, because they have stimulated many cross-country studies and come up with interesting findings concerning the causes and consequences of corrupt activities (for a recent discussion, see Neumann and Graeff 2010). Among the most prominent indices measuring corruption at country level, the CPI - along with the corruption indicator by Kaufmann *et al.* (2003) - is the most frequently cited in the relative literature. In contrast with its widespread diffusion, it is questionable the general validity of such indices, because the economic literature has shown that the use of this index of corruption often reaches very different conclusions (Dreher and Schneider, 2006).

Table 7: Tests for independence between identified latent classes and direct BEEPS measures of corruption

	Physiological administrative corruption VS BEEPS index of corruption absence		Political corruption VS BEEPS index of corruption low		Administrative corruption (inspections) VS BEEPS index of corruption medium		Administrative Corruption VS BEEPS index of corruption high	
Albania	213.708	*	0.811		57.638	*	101.677	*
	(0.000)		(0.368)		(0.016)		(0.001)	
Armenia	172.723	*	0.064		49.305	*	58.879	*
	(0.000)		(0.801)		(0.026)		(0.015)	
Azerbaijan	370.101	*	0.055		113.989	*	49.467	*
	(0.000)		(0.815)		(0.001)		(0.026)	
Belarus	321.844	*	15.058		126.818	*	159.203	*
	(0.000)		(0.220)		(0.000)		(0.000)	
Bosnia & Herzegovina	688.497	*	12.560		429.223	*	211.593	*
	(0.000)		(0.262)		(0.000)		(0.000)	
Bulgaria	412.088	*	22.689		0.156		226.317	*
	(0.000)		(0.132)		(0.693)		(0.000)	
Croatia	202.493	*	82.793	*	0.554		119.937	*
	(0.000)		(0.004)		(0.457)		(0.001)	
Czech Republic	664.122	*	10.352		392.161	*	137.012	*
	(0.000)		(0.309)		(0.000)		(0.000)	
Estonia	599.359	*	35.750		373.577	*	116.024	*
	(0.000)		(0.059)		(0.000)		(0.001)	
FYROM	600.870	*	0.756		365.622	*	80.978	*
	(0.000)		(0.385)		(0.000)		(0.004)	
Georgia	496.622	*	0.004		265.603	*	122.376	*
	(0.000)		(0.949)		(0.000)		(0.000)	
Hungary	258.069	*	50.323	*	12.937		152.746	*
	(0.000)		(0.025)		(0.255)		(0.000)	
Kazakhstan	255.949	*	29.099		45.286	*	71.213	*
	(0.000)		(0.088)		(0.033)		(0.008)	
Kyrgyzstan	268.411	*	0.169		92.779	*	91.552	*
	(0.000)		(0.681)		(0.002)		(0.002)	
Latvia	428.327	*	54.366	*	16.125		75.011	*
	(0.000)		(0.020)		(0.204)		(0.006)	
Lithuania	323.478	*	0.445		143.077	*	65.670	*
	(0.000)		(0.505)		(0.000)		(0.010)	
Moldova	170.909	*	0.546		52.514	*	59.833	*
	(0.000)		(0.460)		(0.022)		(0.014)	
Poland	417.191	*	27.214		161.722	*	120.334	*
	(0.000)		(0.099)		(0.000)		(0.001)	
Romania	385.822	*	20.116		182.290	*	55.487	*
	(0.000)		(0.156)		(0.000)		(0.018)	
Russia	399.378	*	23.117		42.141	*	114.458	*
	(0.000)		(0.128)		(0.040)		(0.001)	
Slovak Republic	81.053	*	0.107		113.735	*	20.003	
	(0.004)		(0.744)		(0.001)		(0.157)	
Slovenia	448.439	*	0.612		301.775	*	42.201	*
	(0.000)		(0.434)		(0.000)		(0.040)	
Tajikistan	389.076	*	11.664		199.459	*	54.259	*
	(0.000)		(0.280)		(0.000)		(0.020)	
Turkey	525.004	*	59.588	*	123.636	*	190.284	*
	(0.000)		(0.015)		(0.000)		(0.000)	
Ukraine	941.732	*	0.866		812.055	*	52.655	*
	(0.000)		(0.352)		(0.000)		(0.022)	
Uzbekistan	479.520	*	0.295		10.885		149.778	*
	(0.000)		(0.587)		(0.297)		(0.000)	
Yugoslavia	292.500	*	0.849		34.337		62.623	*
	(0.000)		(0.357)		(0.064)		(0.012)	

Notes: Independence test is likelihood ratio statistic test at the country level. Asterisks: 5% significance level. All tests have as null hypothesis independence between our estimated country score indicator of corruption and those corresponding to BEEPS.

To validate our estimates, we aggregated firms classified in one latent class of the binary variable at country level and obtained a score index of corruption for each country (Table 8). Note that, whereas in CPI the maximum value of the index represents the best government performance, our scores are measured in reverse. We then use these scores to obtain a new corruption ranking of countries in each latent class, and compare this index with CPI. This comparison is particularly important, because it not only suggests the external validity of our country ranking, but may also indicate the dimension of the bias with a synthetic index and the validity of the tool for measuring corruption.

Table 8: Country corruption scores and rankings

Country	Physiological administrative corruption		Political corruption		Administrative corruption (inspections)		Administrative corruption		CPI (2002) index	
	Score	Ranking	Score	Ranking	Score	Ranking	Score	Ranking	Score	Ranking
Albania	29.1	(21)	43.3	(12)	100.0	(21)	42.5	(17)	41.6	(15)
Croatia	70.4	(5)	23.1	(3)	40.8	(3)	49.8	(18)	63.3	(8)
Turkey	53.5	(14)	83.6	(20)	54.1	(7)	31.5	(13)	53.3	(11)
Bosnia & Herzegovina	37.7	(18)	23.5	(4)	43.1	(4)	100.0	(21)	36.6	(19)
Slovenia	100.0	(1)	60.2	(17)	37.4	(2)	15.5	(5)	100.0	(1)
Poland	77.4	(3)	30.3	(9)	58.4	(9)	17.5	(6)	66.6	(6)
Ukraine	49.4	(16)	29.5	(8)	90.4	(19)	27.0	(9)	40.0	(16)
Belarus	66.8	(8)	26.3	(7)	77.2	(15)	14.7	(3)	80.0	(4)
Czech Republic	65.6	(9)	32.8	(10)	60.1	(10)	31.8	(14)	61.6	(9)
Slovak Republic	56.8	(12)	50.1	(14)	64.9	(12)	31.9	(15)	81.6	(3)
Romania	50.3	(15)	26.2	(6)	84.8	(18)	33.2	(16)	43.3	(14)
Bulgaria	31.2	(20)	100.0	(21)	61.0	(11)	51.7	(19)	66.6	(7)
Moldova	36.6	(19)	77.4	(19)	84.2	(17)	29.6	(12)	35.0	(20)
Latvia	70.0	(6)	48.1	(13)	48.4	(5)	29.4	(11)	61.6	(10)
Lithuania	75.1	(4)	54.9	(15)	50.4	(6)	15.4	(4)	80.0	(5)
Estonia	65.2	(10)	16.9	(1)	73.5	(14)	26.3	(8)	93.3	(2)
Georgia	43.9	(17)	41.4	(11)	67.5	(13)	54.4	(20)	40.0	(17)
Kazakhstan	67.1	(7)	25.3	(5)	79.8	(16)	11.9	(1)	38.3	(18)
Azerbaijan	65.1	(11)	61.8	(18)	55.2	(8)	22.8	(7)	33.3	(21)
Uzbekistan	78.8	(2)	60.2	(16)	35.2	(1)	27.1	(10)	48.3	(12)
Russia	56.3	(13)	20.4	(2)	98.8	(20)	13.4	(2)	45.0	(13)

Notes: We exclude from the comparison Yugoslavia, FYROM, Hungary, Tajikistan and Kyrgyzstan, since Transparency International does not report their country measures for 2002. Ranking for each country is shown in round parenthesis. First and fifth columns are ranked in terms of good government (100=best degree of government institutions); columns 2 to 4 are ranked in terms of highest corruption level (100=most corrupt countries).

The hypothesis proposed is that, if our score estimates for each latent class are correct, physiological (administrative) corruption should partly reproduce the CPI country classification. We find that rankings are in line with this hypothesis for half countries. However, there are countries which show a very different position with respect to CPI. The explanation suggested here, which is generally the criticism in the literature, is that the scores obtained in other latent classes are means, among each other, at country level.

Some examples serve to clarify the idea. In Kazakhstan, the high estimated score in public inspections leads to an increase in the mean level of corruption perception, as measured by CPI. In Azerbaijan and Uzbekistan, the difference between our estimated index of physiological corruption and CPI can be extracted by a relatively greater perception of political corruption. Instead, for Estonia, CPI estimates excellent ranking, and we can say that it is associated with political corruption. More distant are the positions obtained by the Slovak Republic and Bulgaria across the corruption index. CPI indicates that these two countries are in positions 3 and 7 respectively. As also noted in Section 2, Bulgaria is indeed the country with the highest political corruption greatly constrained by the inefficiencies, or failure, of the legal system (see Delavallade 2011).

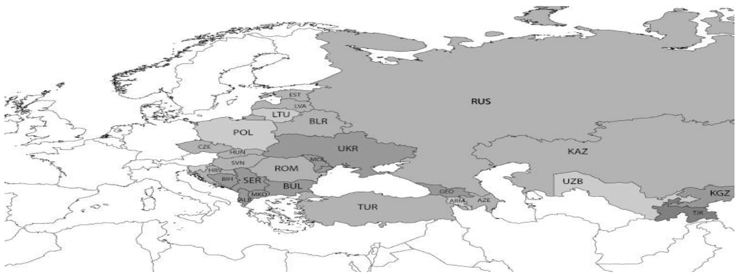
Country rankings for administrative corruption in latent classes 3 and 4 is particularly affected by the level of taxation, permits and inspections, and government contracts, the latter linked with the regulatory quality and enforcement of property rights. As argued by Tanzi (1998), the monopolistic and discretionary power of officials in charge of authorizations or inspections subject to regulations or taxation gives bureaucrats greater opportunities of using their public power to extract bribes. Restrictive regulation and taxation, associated with insufficient enforcement of property rights, are strong determinants of ongoing corruption in Russia, Ukraine, Moldova and Albania.

However, the determinants of corruption are slightly different from what happened in Bosnia with respect to the former USSR countries, in which administrative corruption is shown to rise with weak enforcement of property rights, revealing the incapacity of the courts to implement the law (Johnson *et al.*, 1999). As an example of the highest levels of administrative corruption in Bosnia, Nowak (2001) attributed to the space for corrupt activities the post-war context, which complicated power structures and fragmented administration. Figure 5 shows the ranking of components of corruption in the Eastern Europe countries.

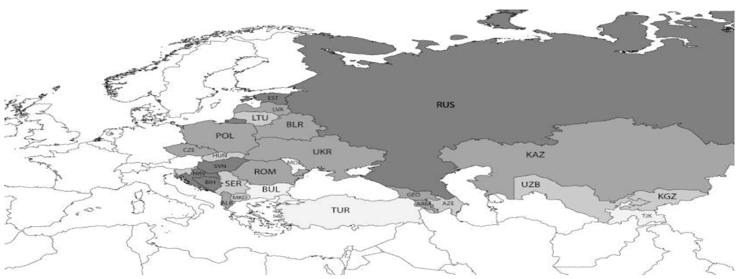
5 An illustrative example of the relationship between firm competitiveness and corruption

Many papers have been written about the determinants of corruption, with particular emphasis on the role of market competition (Ades and Di Tella, 1999; Herzfeld and Weiss,

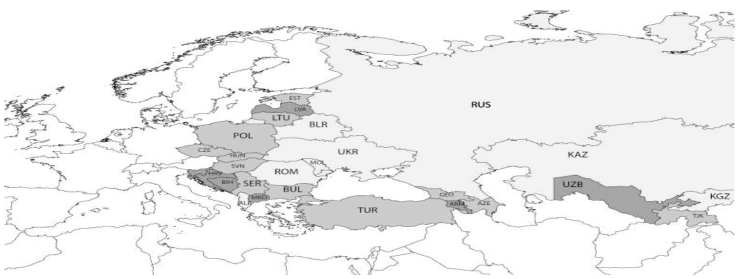
Figure 5: Representation of types of corruption, by country



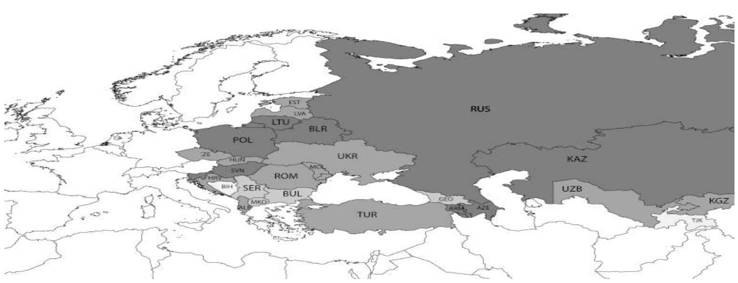
(a) Physiological administrative corruption



(b) Political corruption



(c) Administrative corruption (inspections)



(d) Administrative corruption (public contracts, licensing, taxes and regulations)

Notes: Representation of corruption by country is constructed with corruption scores of Table 8. Score intervals are: i) minimum value of scores of each index (min), ii) scores around the mean of each index (medium), and iii) maximum value of scores of each index (max).

2003; Dreher *et al.*, 2007). In the traditional view, the competitiveness of firms is presumed to lead to more corruption, because they can spend part of their profits on bribes to public officials. In particular, the model of Bliss and Di Tella (1997) finds that firm corruption increases with the firm profitability, so that less competitive firms leave the market; in turn, higher profitability increases bribes, consolidating a non-transparent system of bureaucrats' behavior.

Ades and Di Tella (1997) and Clarke and Xu (2002) empirically examine the hypothesis of a positive correlation between firm competitiveness with corruption in East European and Central Asian countries. The results support the hypothesized mechanism that the most competitive firms have much more money to pay bribes to bureaucrats and increase corruption. These findings are in line with the estimates in the same relationship carried out by Svensson (2003) and Reinikka and Svensson (2003) in the case study of Uganda. However, this economic literature has also proposed and tested an alternative hypothesis: if a comparative disadvantage in competitiveness exists, a firm invests in bribes in order to stay in the market. This hypothesis was supposed by Gauthier and Reinikka (2001) and empirically confirmed for a sample of North African countries by Delavallade (2011).

To illustrate the role of firm competitiveness as a determinant of corruption, the above literature generally distinguishes two components, administrative and state capture indices (in our case proxied by political corruption). One shortcoming of this approach is that one must test whether the aggregate index of corruption can be pooled with either administrative or state capture. It is certainly possible that using perceived corruption indicates that an overlap between these indices exists. Our M-IRT model identifies four monotonically increasing LCs associated with: i) *physiological (administrative) corruption*; ii) *political corruption*; iii) *administrative corruption associated with inspections*; and iv) *administrative corruption linked with increasing issues of public contracts, licensing taxes ad legislative regulations.* Our model allows some firms in these latent classes to be similar to others, but they comprise statistically distinct groups and cannot be pooled. With our model, we can test whether, and how, firm competitiveness affects the identified components of corruption by LC.

In order to re-examine the competitiveness of firms on the specific latent classes of corruption, we set up an empirical framework based on logit models, in which the dependent variables are the firms' dummy variables derived from the estimates of the expected value of frequencies (see Section 4). Competitiveness is also a qualitative variable in three modalities, in which the highest score indicates greater competitiveness and is proxied by the increase in sales over the previous two years (e.g., sales in 1999). We extend the concept of competitiveness by also including firms which, at least, did not suffer reduced sales. This allows us to avoid the general criticism concerning firm competitiveness linked with increases sales (e.g. to maximize firm profits) and to substitute it with a more general aim of firm survival in a competitive market.

The model specification includes the size of the firm (*size*) and of the sector in which it operates (*sector*). In particular, following the results in transition economies of Hellman and Kaufmann (2000), we assume that small firms tend to engage in administrative corruption rather than political corruption, likely because the former is less costly. We include the variable *size* dichotomized, based on full-time employees of firms calculated as "small" (e.g., $2-49$, reference modality), "medium" (e.g., $50-249$), or "large" (e.g., > 250), expecting the size of the firm to affect administrative corruption negatively. In addition, this specific literature finds conflicting evidence that corruption may differ between sectors. Industrial sectors, in which projects involve large amounts of money or high rent-generating public procurements, may be more open to corruption, particularly to political corruption. Thus, we consider a dummy variable according to whether the sector of a firm is "manufacturing" or "services", the latter being the reference modality.

The coefficient estimates of the logit model are listed in Table 5. Column 2 shows the results in terms of the odds ratio of the effect of competitiveness on the components of corruption, as described in Figure 4. We point out three important results. Firstly, firm competitiveness has a positive impact on countries with a physiological level of corruption (e.g., latent class 1), but a negative impact on the highest levels of inspections in administrative corruption. Long-term survival strategy is characterized by the firms which are stable or increase their competitiveness indicating that a representative firm has less need to bribe bureaucrats to obtain, for example, public procurements in its line of business. We estimate that countries with "physiological" corruption do not constrain economic development, finding that competitive (stable) firms increases the probability of having only physiological

Table 9: Firm competitiveness and corruption: estimates

		Physiological administrative corruption		Political corruption		Administrative corruption: inspections		Administrative corruption: public contracts and licensing, taxes and legislative regulations	
Size	*(Ref. small)*								
	Medium	1.124		0.776	*	0.922		1.085	
		(0.091)		(0.104)		(0.078)		(0.122)	
	Large	1.418	***	0.926		0.771	***	0.805	
		(0.128)		(0.134)		(0.076)		(0.109)	
Sector	*(Ref. services)*								
	Manufacturing	1.073		0.924		1.038		0.846	
		(0.095)		(0.128)		(0.097)		(0.103)	
Competitiveness	*(Ref. decreasing)*								
	Stable	1.396	***	0.968		0.786	**	0.775	*
		(0.133)		(0.141)		(0.080)		(0.111)	
	Increasing	1.077		0.933		0.909		1.094	
		(0.084)		(0.112)		(0.074)		(0.120)	
Country	*(Ref. Georgia)*								
	Yugoslavia	1.044		1.580		0.766		1.008	
		(0.289)		(0.640)		(0.221)		(0.307)	
	FYROM	1.128		1.609		0.471	**	1.407	
		(0.335)		(0.690)		(0.161)		(0.443)	
	Albania	0.540	**	1.100		2.029	**	0.723	
		(0.167)		(0.490)		(0.559)		(0.237)	
	Croatia	2.207	***	0.578		0.542	**	0.819	
		(0.596)		(0.297)		(0.164)		(0.256)	
	Turkey	1.347		2.308	***	0.723		0.524	***
		(0.286)		(0.736)		(0.157)		(0.129)	
	Bosnia and Herzegovina	0.697		0.579		0.579	*	2.708	***
		(0.214)		(0.316)		(0.185)		(0.800)	
	Slovenia	3.088	***	1.555		0.470	***	0.230	***
		(0.756)		(0.572)		(0.128)		(0.083)	
	Poland	2.755	***	0.704		0.818		0.269	***
		(0.591)		(0.252)		(0.179)		(0.075)	
	Ukraine	1.150		0.714		1.666	**	0.425	***
		(0.254)		(0.262)		(0.363)		(0.113)	
	Belarus	1.942	***	0.636		1.262		0.224	***
		(0.459)		(0.267)		(0.301)		(0.079)	
	Hungary	2.364	***	0.250	**	1.045		0.345	***
		(0.559)		(0.135)		(0.252)		(0.108)	
	Czech Republic	1.804	**	0.804		0.895		0.537	**
		(0.421)		(0.316)		(0.215)		(0.152)	
	Slovak Republic	1.442		1.298		0.977		0.512	**
		(0.389)		(0.541)		(0.270)		(0.173)	
	Romania	1.203		0.640		1.474		0.527	**
		(0.293)		(0.274)		(0.354)		(0.154)	
	Bulgaria	0.623	*	2.980	***	0.827		0.965	
		(0.162)		(1.023)		(0.208)		(0.262)	
	Moldova	0.766		2.134	**	1.440		0.458	**
		(0.211)		(0.798)		(0.374)		(0.152)	
	Latvia	2.176	***	1.231		0.639		0.468	**
		(0.603)		(0.535)		(0.193)		(0.169)	
	Lithuania	2.538	***	1.349		0.685		0.235	***
		(0.616)		(0.505)		(0.176)		(0.085)	
	Estonia	1.887	**	0.395		1.174		0.403	**
		(0.518)		(0.232)		(0.328)		(0.149)	
	Armenia	5.748	***	1.077		0.344	***	0.050	***
		(1.585)		(0.446)		(0.108)		(0.037)	
	Kazakhstan	1.940	***	0.604		1.348		0.180	***
		(0.463)		(0.259)		(0.323)		(0.069)	
	Azerbaijan	1.712	*	1.636		0.807		0.386	**
		(0.477)		(0.676)		(0.239)		(0.152)	
	Uzbekistan	2.819	***	1.586		0.435	***	0.418	***
		(0.694)		(0.588)		(0.121)		(0.131)	
	Russia	1.424		0.490	*	1.970	***	0.195	***
		(0.306)		(0.187)		(0.420)		(0.058)	
	Tajikistan	0.397	***	2.168	**	0.907		1.623	*
		(0.121)		(0.810)		(0.244)		(0.448)	
	Kyrgyzstan	0.619	*	1.791		1.430		0.760	
		(0.179)		(0.698)		(0.379)		(0.235)	
	Constant	0.391	***	0.124	***	0.541	***	0.379	***
		(0.080)		(0.040)		(0.110)		(0.087)	
	pseudo R^2	0.053		0.050		0.036		0.070	
	LR χ^2	331.513	***	156.335	***	206.960	***	260.869	***
	Log lk	-2976		-1481		-2780		-1720	
	N	4590		4590		4590		4590	

Notes: Estimates reported as odds ratios. In brackets: standard errors of estimated parameters; asterisks: significant p-value, $* \ p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$. A LR test statistic for country fixed effects is shown at bottom of table.

corruption by about 40% (e.g., 39.6%), a practice known in the corruption literature as the "grease the wheels" hypothesis under excessive and inefficient regulation (see Dreher and Gassebner, 2011). However, the effects of competitiveness change for components characterized by the highest levels of corruption. Competitive firms, which occupy a stable position in their markets, have a significantly lower probability than less competitive ones of being involved in corruption. This may imply that the less competitive firms resort to unofficial payments to compensate for their competitive position, distorting the rules of competition for the countries mainly affected by a high level of corrupt activities.

Secondly, including dummies for the specific effects of *size*, we find significant coefficients in the latent classes identified as the highest levels of corruption. In particular, large enterprises reduce the propensity (e.g., odds ratio=0.771 and 0.805, for the components of corruption 3 and 4, respectively) to seek influence in inspections, public contracts and licensing taxes etc., matching previous literature on transition economies of a positive relationship between administrative corruption and small firm size. By contrast, we do not find significant differences between the sectors of the major lines of business in each component of corruption.

Thirdly, we provide support for the findings of the above country classification on corruption. By exploiting CPI in 2002, we use Georgia as a reference country, because it was the most corrupt in the context of transition economies. With few exceptions, almost all the countries show a very low probability of incurring corruption compared with Georgia. Not in contrast with the classification of Figure 5, we estimate that Turkey and Bulgaria have the highest propensity for political corruption.

6 Conclusions

Bartolucci (2007) showed how the M-IRT model can be used to characterize the dimensionality of items in identifying latent classes for a certain latent phenomenon. This paper extends that work to the estimation of corruption, a latent phenomenon often measured by synthesis of observed indices assumed to be correlated with corruption. Following this approach, the constrained version of the M-IRT model is used here to search for hidden types of corruption among the perceptions of firms in a large sample of transition economies. We reject the hypothesis that items in a multidimensional perspective reproduce the distinction between administrative and political corruption, and find that there were four LCs, with increasing level of corruption. We identified of corruption and validated the results according to a direct corruption index derived from the BEEPS and CPI rankings of TI. An illustrative example of the relationship between firm competitiveness and the four identified LCs shows the evident heterogeneity to interpret of the results addressed to policy implications. We confirm for transition economies that weakly competitive firms are more tempted to resort to bribery and bias the rules of competition, and this propensity is estimated greater in small firms associated with administrative corruption.

The model we present should be helpful to empirical researchers in several respects. First, in the growing body of the empirical literature of corruption, contrasting evidence appears to be associated with the bias of an aggregate indicator, partly as an explanation of economic or social determinants. The M-IRT model allows us to disentangle these relationships more appropriately and in the illustration presented here, enables at least to test whether, as in the case of firm competitiveness, this variable is associated with a particular component of corruption. Secondly, the model can be used in social and economic contexts, where the use of items from questionnaires can be statistically summarized, keeping the substantial heterogeneous behaviors. In these situations, the M-IRT model can help to reduce the bias associated with analysis which incorrectly classifies these phenomena.

References

Ades, A. and Di Tella, R., "The new economics of corruption: a survey and some new results," *Political Studies*, 45 (1997), 496–515.

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

Akaike, H., "*Second International Symposium on Information Theory*", in Akademiai Kiado, eds., Information Theory and an Extension of the Maximum Likelihood Principle (Budapest, 1973), 267–281.

Anderson, J. H. and Gray, C. W., "*Anticorruption in Transition 3: Who is Succeeding ... and Why?*" (World Bank, Washington D.C., 2006).

Andersson, S. and Heywood, P. M. "The politics of perception: use and abuse of Transparency international's approach to measuring corruption", *Political Studies*, 57 (2009), 746–767.

Andrews, R. L. and Currim, I. S., "A comparison of segment retention criteria for finite mixture logit models", *Journal of Marketing Research*, 40 (2003), 235–243.

Banfield, J. D. and Raftery, A. E., "Model-based Gaussian and non-Gaussian clustering", *Biometrics*, 49 (1993), 803–821.

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

Bardhan, P., "The economist's approach to the problem of corruption", *World Development*, 34 (2006), 341–348.

Bartolucci, F. "A class of multidimensional IRT models for testing unidimensionality and clustering items", *Psychometrika*, 72 (2007), 141–157.

Bartolucci, F., d'Agostino, G. and Montanari, G. E., "An investigation of the discriminant power and dimensionality of items used for assessing health condition of elderly people", Discussion Paper, arXiv:1008.3268v1 [stat.AP], 2010.

Bartolucci, F., Montanari, G. E. and Pandolfi, S., "Item Selection by an Extended Latent Class Model: An Application to Nursing Homes Evaluation", Discussion Paper, SSRN, 2012.

Biernacki, C. and Govaert, G., "Choosing Models in Model-based Clustering and Discriminant Analysis", Discussion Paper, Institut National de Recherche en Informatique et en Automatique, 1998.

Bliss, C. and Di Tella, R., "Does competition kill corruption?", *Journal of Political Economy*, 105 (1997), 1001–23.

Bock, R.D., "A brief history of item response theory", *Educational Measurement: Issues and Practice*, 16 (1997), 21–23.

Cappellari, L. and Jenkins, SP., "Summarizing multiple deprivation indicators", In: Jenkins S, Micklewright J, eds. *Inequality and Poverty: Re-examined.* (Oxford: Oxford University Press, 2007), 166-184.

Clarke, G. R. G. and Xu, L. C., "Ownership, competition, and corruption: bribe takers versus bribe payers", Discussion Paper No 2783, The World Bank, 2002.

Davoodi, H. R. and Abed, G. T., "Corruption, Structural Reforms, and Economic Performance in the Transition Economies", Discussion No 00/132, International Monetary Fund, 2000.

Delavallade, C., "What Drives Corruption? Evidence from North African Firms", Discussion Paper No 244, Economic Research Southern Africa, 2011.

Dempster, A. P., Laird, N. M. and Rubin, D. B., "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society. Series B (Methodological)*, 39 (1977), 1–38.

Djankov, S., Glaeser, E., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A., "The new comparative economics. *Journal of Comparative Economics*", 31 (2003), 595–619.

Dreher, A. and Gassebner, M., "Greasing the wheels? The impact of regulations and corruption on firm entry", *Public Choice*, (2011), DOI 10.1007/s11127-011-9871-2.

Dreher, A., Kotsogiannis, C. and McCorriston, S., "Corruption around the world: evidence from a structural model", *Journal of Comparative Economics*, 35 (2007), 443–466.

Dreher, A. and Schneider, F., "Corruption and the Shadow Economy: An Empirical Analysis", *Public Choice*, 144 (2006), 215–238.

Falcetti, E., Lysenko, T. and Sanfey, P., "Reforms and growth in transition: Re-examining the evidence", *Journal of Comparative Economics*, 34 (2006), 421–445.

Faye, O., de Laat, J. and Zulu, E., *Poverty Dynamics and Mobility in Nairobis Informal Settlements* (APHRC, 2009).

Fries, S., Lysenko, T. and Polanec, S., "The 2002 Business Environment and Enterprise Performance Survey: results from a survey of 6,100 firms" Discussion Paper No. 83, European Bank for Reconstruction and Development, 2003.

Gauthier, B. and Reinikka, R., "Shifting tax burdens through exemptions and evasion - an empirical investigation of Uganda", Discussion Paper No. 2735, The World Bank, 2001.

Goodman, L. A., "Exploratory latent structure analysis using both identifiable and unidentifiable models", *Biometrika*, 61 (1974), 215–231.

Goodman, L. A., *Analyzing qualitative/categorical data log-linear models and latent-structure analysis*, (Addison-Wesley Pub. Co, 1978).

Graeff, P., "Why should one trust in corruption? The linkage between corruption, trust, norms, and social capital", in Graf Lambsdorff, J., Taube, M., Schramm, M., eds., *The New Institutional Economics of Corruption* (Routledge, London, 2005), 21–42.

Hambleton, R. K. (ed.) *Handbook of Modern Item Response Theory* (Springer, 1996).

Harstad, B. and Svensson, J., "From Corruption to Lobbying and Economic Growth", mimeo, Northwestern University, 2008.

Hellman, J. G., Joel S. and Kaufmann, D., "Seize the State, Seize the Day: State Capture, Corruption and Influence in Transition", Discussion Paper No. 2444, World Bank, 2004. Paper.

Herzfeld, T. and Weiss, C., "Corruption and legal (in)effectiveness: an empirical investigation", *European Journal of Political Economy*, 19 (2003), 621–632.

Hodgson, G. M., "Institutions, recessions and recovery in the transitional economies", *Journal of Economic Issues*, 90 (2006), 1043–1061.

Johnson, S., Kaufmann, D. and Zoido-Lobaton, P., "Corruption, public finances, and the unofficial economy", Discussion Paper No. 2169, The World Bank, 1999.

Kaufmann, D., Kraay, A. and Mastruzzi, M., "Governance Matters III: Governance Indicators for 1996-2002", Discussion Paper No. 3106, World Bank, 2003 .

Kuklys, W., "Measuring standards of living in the UK an application of Sens functioning approach using structural equation models", Discussion Paper on Strategic Interaction, Max Planck Institute, 2004.

Lambsdorff, J. G., *The Institutional Economics of Corruption and Reform: Theory, Evidence and Policy* (Cambridge University Press, 2007).

Licht, A. N., Goldschmidt, C. and Schwartz, S. H., "Culture rules: the foundations of the rule of law and other norms of governance", *Journal of Comparative Economics*,35 (2007), 659–688.

Magidson, J. and Vermunt, J. K., "Latent class factor and cluster models, bi-plots, and related graphical displays", *Sociological Methodology*, 31 (2001), 223–264.

Memisolu, D. and Durguin, A. a., "Public administration reforms and corruption in Turkey", Discussion Paper, Suleyman Demirel University, 2008.

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

Neumann, R. and Graeff, P., "A multitrait-multimethod approach to pinpoint the validity of aggregated governance indicators", *Quality & Quantity: International Journal of Methodology*, 44 (2010), 849–864.

Nowak, R., "Corruption and transition economies", Presented to the preparatory seminar for the 9th OCSE Economic Forum, Bucharest, 2001.

Open Society Institute, "Monitoring the EU Accession Process: Corruption and Anti-Corruption Policy", Discussion Paper.

Rasch, G., "On general laws and the meaning of measurement in psychology", *Berkeley Symposium on Mathematical Statistics and Probability*, 4 (1961), 321–333.

Reinikka, R. and Svensson, J., "Survey techniques to measure and explain corruption", Discussion Paper No. 3071, The World Bank, 2003.

Rose-Ackerman, S., *Corruption and Government: Causes, Consequences, and Reform* (Cambridge University Press, 1999).

Rose-Ackerman, S., *Governance and Corruption* (Cambridge University Press, 2004), 301–344.

Schwarz, G., "Estimating the dimension of a model", *The Annals of Statistics*, 6 (1978), 461–464.

Scott, J. C., "*Comparative political corruption.*" (Prentice-Hall Englewood Cliffs, N.J, 1972).

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

Svensson, J., "Who must pay bribes and how much? Evidence from a cross section of firms", *The Quarterly Journal of Economics*, 118 (2003), 207–230.

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

Tanzi, V., "Corruption Around the World - Causes, Consequences, Scope, and Cures", Discussion Paper No. 98/63, International Monetary Fund, 1998.

Thomas, M., "What do the Worldwide Governance Indicators Measure?", Discussion Paper, Social Science Research Network (SSRN), 2007.

Transparency International, "Global Corruption Report", Tech. rep, 2008.

Warren, E. M., "What does corruption mean in a democracy?", *American Journal of Political Science*, .48 (2004), 328–343.