

THE MEKONG ECONOMIC RESEARCH NETWORK (MERN)

**THE DYNAMICS OF THE INFORMAL SECTOR:
EVIDENCE FROM VIETNAM**

Nguyen Dinh Chuc

Tran Thi Bich

La Hai Anh

Pham Van Long

Hanoi, October 2014

ACKNOWLEDGEMENTS

This study is conducted with the research grant and technical support from the Mekong Economic Research Network (MERN) - a research initiative managed by the Centre for Analysis and Forecasting (CAF) of the Vietnam Academy of Social Sciences (VASS) with financial support from the International Development Research Centre (IDRC), Canada. The authors are grateful for helpful comments and suggestions by Dr. Xavier Oudin, Axel Demenet, Dr. Nguyen Thang and participants at the final networking meeting of MERN, Phnom Penh, Cambodia 8-9 September 2014.

ABSTRACT

This paper investigates the formalisation process of small and medium enterprises in Vietnam using data from SME surveys in 2009 and 2011. To capture well the nature of informality, Multiple Correspondence Analysis is applied to create an informality index. The paper, then, uses the Cluster Analysis to segment firms into different clusters to identify factors associating with the transition of firms from informal to formal status and vice versa. The results find that informal status is the matter of no choice for small and vulnerable businesses. Furthermore, firms moving from formal to informal conditions are either weak businesses that have no potential to expand or strong enterprises which want to escape from government regulations. Contrarily, formal firms and those moving from informal to formal status are strong and younger businesses that achieve the highest technical and scale efficiencies. Formalisation associates with the burden of regulation interventions. Findings from the paper imply that the Vietnamese government should provide more assistance to weak firms and release regulation interventions to promote the formalisation.

Key words: Informality, Technical Efficiency, Multiple Correspondence Analysis, Data Envelopment Analysis, Cluster Analysis

I. INTRODUCTION

Informality is by no mean a new phenomenon. However, the existence of the sector has divided scholars. Moser (1984) argues the appearance of the informal sector as a failure of the macro-economy. At the national level, informality means the loss of taxes for social welfare and mis-allocation of resources that hinder the economic development (see Loayza 1996; Dabla-Norris and Feltenstein 2005). At firm level, informality impedes firms to access to secure property rights, formal contract mechanism, financial services (Levenson and Maloney, 1998, Maloney et al. 2011; Rand and Torm, 2012), and trade across the country border (Tenev et al., 2003). At the individual level, working in the informal sector implies having no access to social insurance (USAid, 2005; Rand and Torm, 2012). Thus, the informal sector needs to be formalised and there is no need for policy interventions. If any, that should be designed to clear the sector in the course of development.

Contrarily, other economists consider the informal sector as the issue of selection. Harris and Todaro (1970) see informal jobs as the matter of no choice where people lose or are unable to find jobs within the formal sector. On the other hand, Hart (1972) and Maloney (1999) argue that informal enterprises select themselves into the sector and workers sometimes prefer self-employment to salaried jobs. Although not identifying informal jobs as the matter of selection, Cling et al. (2013) confirm the role of the informal sector as the absorption of labour redundancies from the formal economy during the global economic crisis, thus contributing to poverty reductions in developing countries. Therefore, the existence of the sector is necessary and deserves to policy considerations.

Maloney (2004) argues that different views on the informality can be compromised by taking into account the heterogeneity of the informal sector. In addition, we find that different results from empirical studies might come from various definitions of the sector. Often, traditional studies use definitions in terms of statistical measures.

For instance, formality is defined as having tax code and/or paying social insurances for employees or and/or having workers greater than a certain threshold. These definitions sometimes over- or under-estimate the informal sector.

This *first-ever* paper creates the informality index to capture well the nature of informality using the Multiple Correspondence Analysis (MCA). By using Cluster Analysis to segment firms into distinct clusters and investigate factors associated with each cluster, the paper contributes to a small number of studies focusing on high heterogeneity of informal enterprises in Vietnam.

The study of Vietnam is of particular interest because Vietnam is considered as one of the country with high level of informality. This high level of informality continues to remain together with recently increasing debates on whether the sector needs policy interventions. Therefore, in-depth insights to each cluster of informal firms will provide evidence for policy design.

The paper is structured as followed. Section 2 gives an overview of the informal sector in Vietnam. Methodology and data description are provided in Section 3. Section 4 discusses the dynamics of the sector and factors associated with the transformation process and Section 5 concludes.

II. THE INFORMAL SECTOR IN VIETNAM

Traditionally, the concept of 'informality' refers to the part of the economy that is not accordance with prescribed regulations (Portes et al., 1989; ILO, 2002; Becker, 2004; Oviedo, 2009). As such, informality may take many forms from small-scale companies, unregistered business activities, the street vendors to large companies using a workforce without labour contracts but does not include illegal activities such as crimes and drug trafficking. Given this broad definition, empirical studies on the informal sector use diverse applicable definitions leading to different results.

Perhaps, the definition of the 15th International Conference of Labour Statisticians (ICLS)¹ starts to set up a clear distinction between the formal and informal sector. To distinguish informal enterprises, the 15th ICLS recommends using the following three criteria: non-registration of the enterprise; small size in terms of employment; and non-registration of the employees.

Since the 15th ICLS lets the cut-point of firm size to countries, various thresholds are applied in different countries such as five laborers in Central American (Funkhouser, 1996), six for Bolivia, Mexico and Peru (Pradhan and van Soest, 1995; Marcoullier et al., 1997, Pradhan and van Soest, 1997, Maloney, 1999), ten for Kenya and Nigeria (Livingstone, 1991; Arimah, 2001), and twenty for Sudan (Cohen and House, 1996). These studies are based on the assumption that the majority of firms under the above clarified sizes are likely not to comply with government regulations.

Studies on informality in terms of employees also do not converge to a unanimous agreement. According to Duval-Hernández (2006), informality should be measured according to the worker's legal status such as labour contract. However, such a measure has no relevance for the case of self-employed workers in practice since self-employees cannot contract with themselves. Therefore, the alternative

¹ International Labour Office (1993). 15th International Conference of Labour Statisticians: Highlights of the Conference and text of the three resolutions adopted. *Bulletin of Labour Statistics* 1993-2, IXXXIV. Geneva.

indicator of informality is social security status (Merrick, 1976), which is measured by no social protection or non-payment of social security taxes (Portes, Blitzner, and Curtis, 1986; Marcoullier et al., 1997, Maloney, 1999; Saavedra and Chong, 1999).

Some studies compare the sensitivity of informality rates to firm-size and non-payment of social security taxes. The results show that the informal sector measured by the latter criteria is larger than that estimated by the former (Marcoullier et al., 1997, Saavedra and Chong, 1999), Pisani and Pagán, 2004). This suggests that different measures may behave diversely.

Similar to other countries, there has been no consensus on the definition of the informal sector in Vietnam leading to controversy measure of the sector's size. Until 2007, Razafinfrakoto, Cling and Roubaud introduced international standards to measure the informal sector and informal employment in Vietnam (Razafindrakoto et al. 2008). In accordance to international measures, the informal sector is defined as 'all private unincorporated enterprises that produce at least some of their goods and services for sale or barter do not register and engage in non-agricultural activities' (Cling et al. 2011: 5).

Since then, the above definition has been applied in the annual Labour Force Survey (LFS) in Vietnam. Results from LFS show that the informal economy is predominant in Vietnam. The share of informal employment is around 70% during the period from 2007 to 2013 (Table 1).

Table 1. Proportion of informal employments in Vietnam, 2007-09 (%)

	2007	2009	2010	2011	2012	2013
Informal sector	74.4	70.96	74.57	74.48	70.91	68.53
Formal sector	25.6	29.04	25.43	25.52	29.09	31.47
Total	100	100	100	100	100	100

Source: Authors' calculation using LFS 2007-2013

The characteristics of the Vietnamese informal sector are similar to those observed in other developing countries. For instance, informal enterprises have low productivity and income, lack of funds and investments with low level of integration into the economy. In terms of employment, informal workers have to work in unstable operation and working conditions.

As informality is partly seen as a weakness in economic development, formalising the sector is a desirable goal in policy designs. However, the degree of transition depends on several factors. Thiam (2007) states that the transition from informal to formal status is triggered through incentives and enabling environment reforms such as access to credit, trade facilitation, formalization of business linkages, making costs of formalization worthwhile. Moreover, (Tenev et al., 2003; Dabla-Norris et al., 2005, 2008) find that both regulation burden and legal quality are important determinants of informality. Other researchers shows that formality relates to tax burden and/or costs of complying with regulatory requirements (Marcouiller and Young 1995, Cebula 1997, Friedman et al. 2000, Azuma and Grosman 2002, Giles and Tedds 2002, Straub, 2005), entry costs (Auriol and Warlters, 2005), labor (Friedman et al., 2000, Johnson et al., 1997, 1998, 2000; Botero et al., 2004), and financial development (Straub, 2005).

In Vietnam, the interactions of the government with business activities, level of taxation, business environment, and the level of access to resources affect the transition from informal to formal status (Tenev et al., 2003). Cling et al. (2010) find that businesses' size, income, and professional premises are positively and significantly correlated with the registration decision. In addition, the education level of entrepreneurs influences their behaviour of working under regulations. Nevertheless, the number of years in business apparently has no impact on registration.

III. ANALYTICAL FRAMEWORK AND DATA

3.1 Analytical framework

Because informality is a multi-dimensional concept and measured by several criteria, empirical studies struggle to match the definition of informality in terms of statistical measures with those proposed by ICLS and ILO. Therefore, to better capture the nature of informality, the paper firstly uses MCA to construct the formality index. Then, it applies the nonparametric Data Envelopment Analysis (DEA) to measure a firm technical and scale efficiency. Those efficiencies together with entrepreneurs and firms' characteristics, and business environment are used as input variables for the Cluster Analysis (CA) which is used to classify firms into different clusters and identify factors associated with each cluster.

Constructing the formality index

Multiple Correspondence Analysis is a data reduction technique which is similar to Principal Component Analysis (PCA) but applied for categorical data. To compress data, the method measures associations among variables by calculating the Chi-square distance between different categories of the variables and between the observations (Le Roux and Rouanet, 2004). Homogeneity between observations and variables are then maximized to find out the underlying dimensions which are best able to describe the central oppositions in the data.

The method is briefly explained as follows. We have an $n \times m$ data matrix with rows corresponding to objects and columns to variables. Assuming variable j that has different k values (categories) and define G_j as the $n \times k_j$ indicator matrix corresponding to this variable. MCA determines the vector y_j which quantifies the categories of each of the variables such that homogeneity is maximised. Let $G_j y_j$ represents a single quantification of the n objects induced by variable j . MCA works with p simultaneously dimensional quantifications. Let call these $k_j \times p$ matrices Y_j the multiple nominal quantifications of variable j . Then the matrices $G_j Y_j$ induce m multiple quantifications of the objects. MCA minimises the loss of homogeneity

with loss function which is determined in terms of squared deviations and written as follows:

$$\min \sigma(X, Y) = \sum_{j=1}^m SSQ(X - G_j Y_j) \quad (1)$$

Solutions of this minimisation problem produce principal components or latent dimensions which reflect as much as possible original information. The number of retained components is determined by modified eigen values. As in PCA, the first axis is the most important dimension in terms of the amount of variance accounted for.

Measuring a firm performance

One of the contributions of this paper is to explore the relationship between the formalization process and firms' technical and scale efficiencies. Efficiencies are estimated using a popular non-parametric DEA approach. The method firstly sets up a nonparametric frontier which is the production level of firms dominating other enterprises in the same industry. As in Daraio and Simar (2007), we set up the DEA production frontier as:

$$\hat{\psi}_{DEA} = \left\{ \begin{array}{l} (x, y) \in \mathfrak{R}_+^{p+q} \mid y \leq \sum_{i=1}^n \gamma_i Y_i; x \geq \sum_{i=1}^n \gamma_i X_i, \text{ for } (\gamma_1, \dots, \gamma_n) \\ \text{s.t.} \quad \sum_{i=1}^n \gamma_i = 1; \gamma_i \geq 0, i = 1, \dots, n \end{array} \right\} \quad (2)$$

where (X_i, Y_i) are observations in a convex hull of $\mathcal{X} = \{(X_i, Y_i), i = 1, \dots, n\}$ covering unit (x, y) .

Formula (2) allows the variant returns to scale production technology, where outputs under efficient production change by a different proportional to the change in inputs. Different types of returns to scale can be achieved by changing the constraint $\sum_{i=1}^n \gamma_i = 1$. If $\sum_{i=1}^n \gamma_i = 1$ is dropped from the formula we will have a constant

returns to scale technology. If $\sum_{i=1}^n \gamma_i \geq 1$ or ≤ 1 we allow non-decreasing or non-increasing returns to scale, respectively.

Technical efficiency can be measured using either the input or output orientation. Since firms in our study are micro and small businesses which are price takers, the appropriate approach is the input-oriented technical efficiency.

Under the assumption of variable returns to scale production technology, the input-oriented technical efficiency score for a firm operating at the level (x_0, y_0) will be:

$$\begin{aligned} \hat{\lambda}_{DEA}(x_0, y_0) = \min \left\{ \lambda \mid y_0 \leq \sum_{i=1}^n \gamma_i Y_i; \lambda x_0 \geq \sum_{i=1}^n \gamma_i X_i; \lambda \geq 0; \right. \\ \left. \sum_{i=1}^n \gamma_i = 1; \gamma_i \geq 0; i = 1, \dots, n \right\} \end{aligned} \quad (3)$$

Banker et al. (1984) combine definitions of technical efficiency and returns to scale technology to split technical efficiency under constant returns to scale (TE_{CRS}) into scale efficiency (SE) and pure efficiency in variant returns to scale (TE_{VRS}) as follows:

$$TE_{CRS} = TE_{VRS} * SE \quad (4)$$

When a firm operating at TE_{CRS} , it achieves both technical and scale efficiency and thus has the highest productivity level.

Analysing firm dynamics

Given the high heterogeneity of the informal sector, the paper uses the Cluster Analysis to investigate the transformation of a firm from informal to formal status

and factors associating with the transition. The objective of CA is to classify firms into groups or clusters in such a manner that all observations within a group are similar, while observations in different groups are differential based on the Chebychev distance. In this research, we use the hierarchical method since the number of clusters is unknown before the clustering procedure. We also choose Ward's linkage to partition firms since this algorithm merges observations while increasing the overall within-cluster variance to the smallest degree (Mooi and Sarstedt, 2011).

3.2 Data and variable measurement

3.2.1. Data

The paper investigates the dynamics of the informal sector in Vietnam by using the data sets from the surveys on Small and Medium Enterprises (SMEs) in Vietnam in 2009 and 2011. Both surveys cover approximate 2,500 enterprises and are implemented in 10 provinces which are distributed equally across regions and rural/urban areas of Vietnam. The provinces include Hanoi, Phu Tho, Ha Tay, and Hai Phong in the north, Nghe An, Quang Nam, Khanh Hoa in the centre, and Lam Dong, Ho Chi Minh City, and Long An in the south.

The sample is stratified by ownership forms based on the list of formal enterprises and an on-site random selection of informal businesses (For more information on the survey, see Rand and Torm, 2012). Information collected by the surveys include firms and entrepreneurs' characteristics, firms' production and performance, and business environment. Table 1 presents the distribution of the sample by ownership types.

Table 2 Distribution of the sample by ownership types

	2009	2011
- Household Businesses	1,734	1,640
- Private company	214	203

	2009	2011
- Partnership or Collective	83	71
- Limited company	531	532
- Other ownership	97	106
Total	2,659	2,552

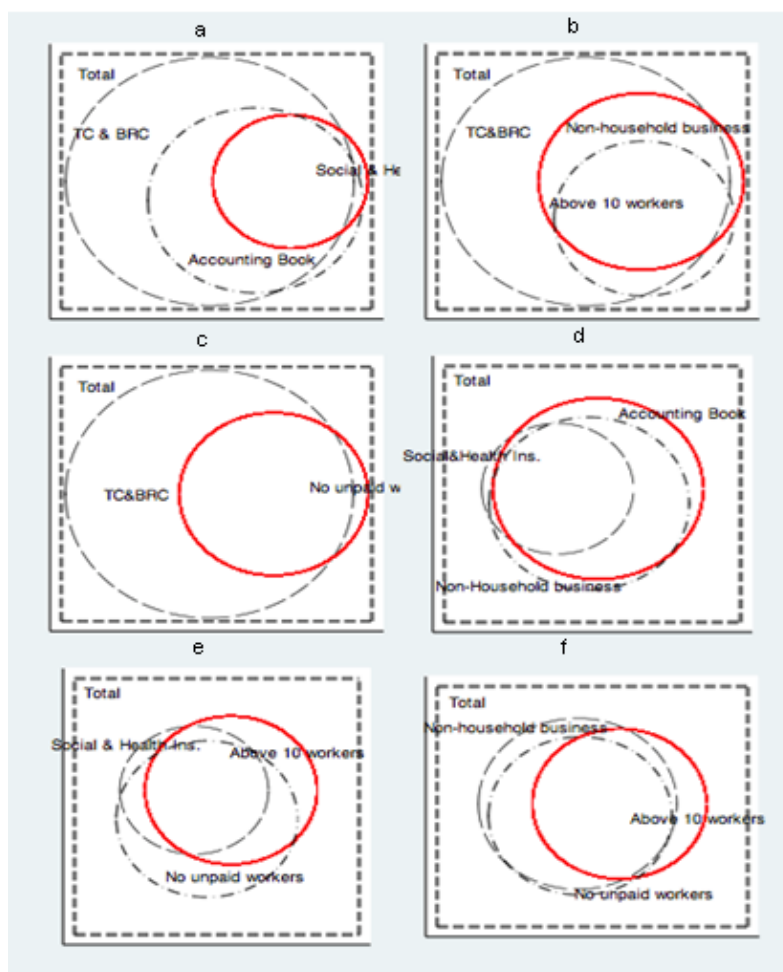
Source: Authors' calculation using SMEs data in 2009 and 2011.

3.2.2. Variable measurement

Informality index

More often, empirical studies use a basic statistical definition to define an informal firm as *not having business registration certificate (BRC)* and sometimes without *paying social and health insurance*. In any case, the informal sector may be over or under-estimated. For instance, in SME survey sample, if having BRC and Tax Code (TC) is used to define a formal firm, other criteria of formality such as paying social and health insurance and having accounting book are partially violated as indicated in Figure 1a. Among firms holding BRC and TC, only a quarter of them pay social and health insurance and a half have accounting books. In this case, informality is under-estimated in comparison with the definition proposed by ICLS and ILO. Contrarily if combining both 'having BRC and TC' and 'paying social and health insurance', informality is over-estimated because formality is bounded by the smallest circle in this Figure.

Figure 1. Venn Diagrams showing correlation between criteria



Source: Authors' calculation using SME data set 2011

Note: The area of squares represent all SMEs while the area of circles describe number of SMEs meeting one of the following conditions (i) holding BRC and TC; (ii) paying social and health insurance; (iii) having accounting book; (iv) being non-household businesses; (v) having above 10 workers; (vi) having no unpaid workers.

By investigating the rest of Figure 1, we find that formality criteria including paying social and health insurance, employing more than 10 workers, having accounting book, being non-household business and having no unpaid workers are mainly coincided each other (Figures 1d, e, f). This visualisation suggests using the informality index which is a combination of all formality criteria proposed by ICLS and ILO. Therefore, we proceed to calculate the informality index using MCA.

The index is set as the score on the first principal axis derived from many firms' characteristics associated with different aspects of the informality. They include (i) incompliance with government regulations (no BRC, no TC, employees without social insurance, employees without social and health insurance, no accounting books), (ii) firm's resource endowment (number of employees, unpaid workers) and (iii) type of ownership (household businesses or not). These variables can be a mixture of nominal and categorical data, which could be integrated via the MCA method. Table 2 presents the description of variables used to define the nature of informality.

Table 3. Description of variables to define informality index

Percentage (%) of enterprises	2009	2011
Have Tax Code	65.63	70.92
Have Business Registration Code	67.32	73.00
Pay Social Insurance	19.37	20.49
Pay Health Insurance	19.33	20.96
Have Accounting Book	41.52	39.30
Ownership		
- <i>Household Businesses</i>	65.21	64.26
- <i>Private company</i>	8.05	7.95
- <i>Partnership or Collective</i>	3.12	2.78
- <i>Limited company</i>	19.97	20.85
- <i>Other ownership</i>	3.65	4.15
Full-time workers		
- <i>Under 5 workers</i>	42.16	45.42
- <i>5 to 10 workers</i>	28.21	27.08
- <i>Above 10 workers</i>	29.64	27.51
Unpaid workers		
- <i>No unpaid worker</i>	29.26	30.53
- <i>1 unpaid worker</i>	21.66	22.02
- <i>2 unpaid workers</i>	35.43	35.70
- <i>3 unpaid workers</i>	8.42	7.29
- <i>4 unpaid workers</i>	3.84	3.17

Percentage (%) of enterprises	2009	2011
- 5 unpaid workers	0.79	0.82
- From 6 unpaid workers	0.60	0.47
Number of enterprises	2,659	2,552

Source: Authors' calculation using the 2009, 2011 SME surveys

The Vietnamese regulations require a formal firm to have a BRC and TC.² However, in fact, many firms operate with both BRC and TC, while others have BRC but without TC³ (Rand and Torm, 2012). Therefore, the percentage of firms holding BRC is greater than that of firms having TC (Table 2). The sample is dominant by household businesses (65 per cent). Moreover, unpaid workers⁴ seem to be popular among SMEs. Regarding social and health insurance, only 19 per cent of businesses in 2009 pay social and health insurances for their employees and this ratio increases to 20 per cent in 2011. Out of the sample, 39-42 per cent of enterprises have accounting books.

The variances (eigenvalues) of the first five axes and their inertia rates are given in Table 3. The MCA result shows that the first dimension explains 91-92 per cent of the total variance (Table 3). Therefore, the informality index is defined as scores of the first dimension of MCA.

Table 4. Dimensions in MCA

	Dim1	Dim2	Dim3	Dim4	Dim5
2009					
Eigen values	0.34	0.01	0.00	0.00	0.00
% inertia (variance)	92.12	1.77	0.03	0.00	0.00
2011					
Eigen values	0.34	0.01	0.00	0.00	0.00
% inertia (variance)	91.16	2.22	0.03	0.01	0.00

Source: Authors' calculation using SME data sets in 2009, 2011

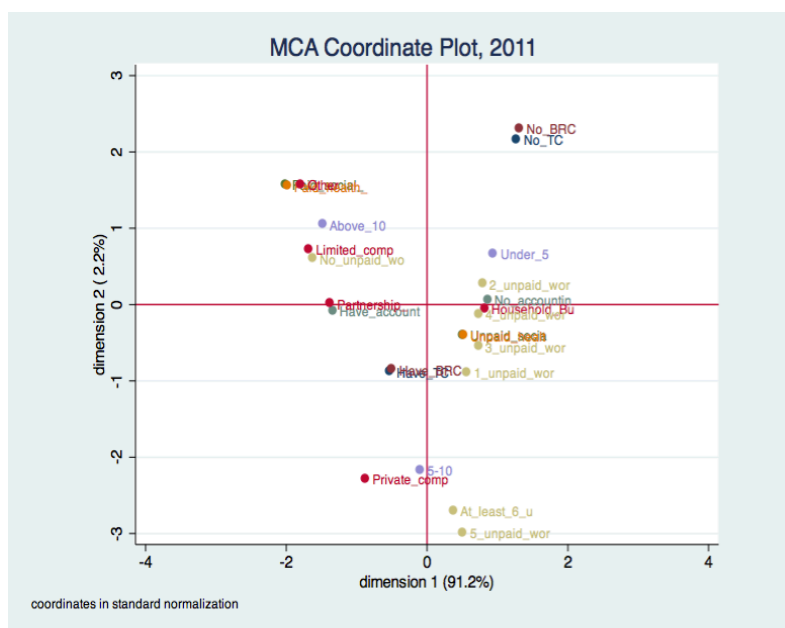
² The Government Decree No. 88/2006/ND-CP dated August 29, 2006.

³ This indicates that government officials would come to collect (usually on a monthly basis) a lump-sum tax/fee.

⁴ Family members who tend to work for household businesses are usually unpaid.

To test new informality index, we plot the principal coordinates of nominal and categorical variables of the first two dimensions using the 2011 SME survey (see Figure 3). This figure shows that all informality criteria including no BRC and TC, no social and health insurance, under five employees, at least one unpaid worker, no accounting book, and household businesses are located on the right of the first dimension with positive values. In contrast, the negative values of the first principal axis represent well all conditions of formality: paying social and health insurance, employing above 10 employees (with noting that the category “5-10 employees” located close to zero), no unpaid workers, having account books, and non-household businesses. This visualisation suggests taking positive values of the first principal axis as informality and negative values as formality. The results are consistent and robust for both 2009 and 2011 (see Appendix 1 for MCA coordinate plot for 2009).

Figure 2. MCA Coordinate Plot



***Source:** Authors' calculation using the 2011 SME survey*

Firm performance: technical and scale efficiency

The non-parametric DEA is applied to estimate technical and scale efficiencies of the sampled firms in 2009 and 2011. The reason to use technical and scale

efficiencies as indicators of a firm performance rather than total factor productivity (TFP) is that when a firm operates optimally at both technical and scale efficiencies, it definitely achieves the highest level of productivity and thus TFP (Coelli et al., 1998). Moreover, scale efficiency is a good measure for a firm performance when the firm operates under variant returns to scale technology.

By imposing the restriction on $\sum_1^n \gamma$, the non-parametric DEA calculates a firm technical efficiency under both constant and variant return to scales and splits its into technical and scale efficiencies using equation (4). Summaries of technical and scale efficiencies of formal and informal firms are presented in Table 4.

Table 5. Technical and scale efficiencies of formal and informal firms

	2009		2011	
	Technical efficiency	Scale efficiency	Technical efficiency	Scale efficiency
Formal firms	0.54	0.86	0.61	0.86
Informal firms	0.57	0.68	0.55	0.69
Total	0.56	0.75	0.57	0.76

Source: Authors' calculation using data set from SME survey in 2009&11.

As can be seen from Table 4, while technical efficiency is not much different between formal and informal businesses, scale efficiency of formal firms is much higher than that of informal counterparts. The results imply that formal enterprises are more optimal at scale operation than informal businesses.

IV. DYNAMICS OF THE INFORMAL SECTOR

Using the balanced panel of 1,774 firms in each year, we start our discussion of the main variable of interest, i.e., the formality incidence. Based on results from MCA, we define a formal firm if its informality index is negative and informal if the index is positive. Table 6 documents the formal-informal status and dynamics of firms during the period 2009-11.

Table 6. Informal-formal status during 2009-11

2009	2011		Total
	Formal	Informal	
Formal	620 (90) (95)	72 (10) (6)	692 (39)
Informal	31 (3) (5)	1,051 (97) (94)	1,082 (61)
Total	651 (37)	1,123 (63)	1,774

Source: Authors' calculation using SME data 2009-11

Note: Entries are the numbers of enterprises (percentage are in parentheses)

As can be seen from Table 6, informal businesses are over-represented in both years and experience a slightly increasing trend during the period 2009-11 (61% and 63% in 2009 and 2011, respectively). This period also witnesses a dominant transition from formal to informal status (10% versus 3% that move in the opposite direction).

To find out the reasons why informality has an increasing tendency while business environment is more improved overtime, we use the cluster analysis method. As suggested in the literature, characteristics of entrepreneurs and firms, firm performance and policy environment are among the determinants of informality level. They are selected as input variables for CA. Table 7 presents summary statistics of variables included in CA.

Table 7. Summary statistics of variables used in cluster analysis

Variables	Summary	Variables	Summary
1. Characteristics of entrepreneurs		4. Policy environment	
Sex (number)		Constraints in 2009	
Male	1,108	No constraint	300
Female	666	Capital constraint	554
Age (mean)	46.60	Labour constraint	46
Education (number)		Technical constraint	72
Not finished primary	31	Market constraint	509
Finished secondary	650	Outside service constraint	101
Finished high school	1,093	Land constraint	135
Technical skill (number)		Policy constraint	45
Unskilled	121	Constraints in 2011	
Elementary worker	92	No constraint	303
Technical worker	1,121	Capital constraint	661
College and above	440	Labour constraint	85
Social capital	1,269	Technical constraint	63
2. Characteristics of firms		Market constraint	439
Firm age (mean)	14.11	Outside service constraint	73
Have electronic access (%)	35.79	Land constraint	83
3. Firm performance and dynamics		Policy constraint	41
Expansion and innovation in 2009 (%)	57.72	5. Government assistance and regulations (%)	
Expansion and innovation in 2011 (%)	59.25	Having financial assistance in 2009	29.98
Efficiency (mean)		Having financial assistance in 2009	10.59
Technical efficiency in 2009	0.56	Having technical assistance in 2011	2.93
Technical efficiency in 2011	0.57	Having technical assistance in 2011	3.55
Scale efficiency in 2009	0.75	% of management time dealing with government regulations in 2009	1.11
Scale efficiency in 2011	0.76	% of management time dealing with government regulations in 2011	2.35

Source: Authors' calculation

To cluster firms, the Ward's linkage cluster analysis assigns observations into more homogenous groups based on selected variables for the analysis. To decide the number of clusters, we refer to dendrogram and Duda_Hart index. While the dendrogram gives a somewhat arbitrary solution (Mooi and Sarstedt, 2011), Duda_Hart rule provides a more scientific selection with large $Je(2)/Je(1)$ indexes and smaller pseudo_T_squared values indicating more distinct clustering (Duda and Hart, 1973). The dendrogram and Duda_Hart indexes give the solution of four clusters for the studied sample (see Appendix 2 for more explanations of the solution).

Validation tests are also provided to check whether clusters are distinct from one another. Table 8 documents Sidak post-hoc for pair-group comparison after one-way ANOVA for quantitative variables while Table 9 reflects Chi-squared test for categorical variables.

Table 8. ANOVA and Sidak post-hoc for pair-group comparison

	ANOVA	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Entrepreneur age	***	***			2
		***	***		3
		***	***	***	4
Firm age	***	***			2
		***	***		3
		***	***	***	4
Technical efficiency 2009					2
					3
					4
Scale efficiency 2009	***	**			2
			*		3
			***		4
Technical efficiency 2011	***				2
			**		3
		***		***	4
Scale efficiency 2011	***	***			2
			***		3

	ANOVA	Cluster 1	Cluster 2	Cluster 3	Cluster
		***		***	4
% of management time dealing with regulations 2009	***	***			2
			**		3
		*	*		4
% of management time dealing with regulations 2011	***	***			2
			***		3
		***	***		4

Note: '***' '**' '*' are statistically significant at 1%, 5%, and 10%, respectively.

Table 9. Chi-squared test for categorical variables

	S.L ⁺		S.L
1. Characteristics of entrepreneurs		4. Policy environment	
Sex	***	Constraints in 2009	***
Education	***	Constraints in 2011	***
Technical skill	***	Financial assistance in 2009	
Social capital	***	Financial assistance in 2011	
2. Characteristics of firms		Technical assistance in 2009	*
Electronic access	***	Technical assistance in 2011	***
3. Firm performance and dynamics		Being inspected in 2009	***
Informal-formal transition	***	Being inspected in 2011	**
Expansion and innovation in 2009	***		
Expansion and innovation in 2011	***		

Note: ⁺ S.L: significant level.

'***' '**' '*' are statistically significant at 1%, 5%, and 10%, respectively.

Results from ANOVA and Chi-squared tests show that all variables included in CA are meaningful and 4 clusters are distinct from one another. We, then, proceed to the interpretation of clusters. Table 10 provides streamlined information on cluster

characteristics.⁵It should be noted that the number of observations reduces to 1,765 because some missing values are removed from the sample.

Table 10. Characteristics of clusters

	Group 1	Group 2	Group 3	Group 4	Total
Number of observation	441	392	97	835	1,765
Percentage	25%	22%	5%	47%	100%
1. Characteristics of entrepreneurs and firms					
Age of entrepreneurs (mean)	56.44	50.88	59.32	37.90	46.59
Age of firms (mean)	18.50	9.02	40.84	10.97	14.06
3. Firm performance and dynamics					
Informal-formal transition					
Stayed as informal	72.8%	48.5%	70.1%	55.5%	59.0%
Formal	21.3%	43.9%	27.8%	39.2%	35.1%
Moving from informal to formal	1.6%	1.5%	0.0%	2.2%	1.8%
Moving from formal to informal	4.3%	6.1%	2.1%	3.2%	4.1%
Efficiency (mean)					
Technical efficiency 2009	0.55	0.56	0.55	0.55	0.55
Technical efficiency 2011	0.55	0.58	0.50	0.59	0.57
Scale efficiency 2009	0.74	0.80	0.72	0.74	0.75
Scale efficiency 2011	0.72	0.79	0.67	0.78	0.76
4. Policy environment					
Constraints in 2009					
No constraint	25%	23%	8%	43%	100%
Capital constraint	20%	20%	4%	56%	100%
Labour constraint	24%	20%	2%	54%	100%
Technical constraint	28%	11%	7%	54%	100%
Market constraint	30%	25%	6%	40%	100%
Outside service constraint	29%	25%	4%	43%	100%
Land constraint	19%	28%	3%	49%	100%
Policy constraint	40%	11%	13%	36%	100%
Constraints in 2011					
No constraint	37%	14%	10%	39%	100%
Capital constraint	20%	21%	4%	54%	100%
Labour constraint	32%	22%	4%	42%	100%
Technical constraint	30%	21%	3%	46%	100%

⁵A full table of Cluster Analysis is provided upond request.

	Group 1	Group 2	Group 3	Group 4	Total
Number of observation	441	392	97	835	1,765
Percentage	25%	22%	5%	47%	100%
Market constraint	24%	27%	4%	45%	100%
Outside service constraint	26%	25%	5%	44%	100%
Land constraint	17%	34%	6%	43%	100%
Policy constraint	20%	17%	10%	54%	100%
% of management time dealing with government regulations in 2009 (mean)	0.92%	1.35%	0.92%	1.13%	1.12%
% of management time dealing with government regulations in 2011 (mean)	1.75%	3.15%	1.82%	2.36%	2.36%

Source: Authors' calculation

Cluster 1–Informal and moving from formal to informal status. Compared to the average level, this group is over-represented by informal businesses (72.8%) and those moving from formal to informal status (4.3%). Firms in this group run by old entrepreneurs aged of 56 on average. This might be in line with the literature that old people prefer self-employment than salaried work (Marcouiller et al., 1997). Moreover, runners of informal businesses have lower education, technical skills, and social capital than their counterparts in the formal sector. This may support for the argument that old and low educated people have trouble finding a waged job (Cunningham and Maloney, 2001). Firms in this group are among the second lowest technical and scale efficiency level. The group has little opportunity to expand and innovate. The most constraint faced by this group is policy constraint followed by technical and market constraints. Firms in this group report the least intervention by regulations (0.92% and 1.75% of management working time in 2009 and 2011, respectively).

Cluster 2–Formal and moving from formal to informal. This group is representative by both formal firms (43.9%) and businesses moving from formal to informal status (6.1%). Although run by old people with average age of 51 years old, this group includes the youngest businesses aged 9 years old. Firms in this group achieve the

highest technical and scale efficiency compared to the sample average level. Land and market constraints are the most obstacles facing by this group. Technical constraint is not cited because firms in this group often get technical assistance from the government (35% and 38% in 2009 and 2011, respectively). The hypothesis that formal firms suffer regulation burdens and the desire to avoid this burden by moving to informal status seem to be correct for the case of Vietnam as firms in this cluster are the most interfered by government regulations (1.35% and 3.15% of management working time in 2009 and 2011, respectively).

Cluster 3 – Informal and vulnerable. This group is the smallest one, comprising 5% of the whole sample. Firms in this group are old, run by the oldest and low educated people. The group has the lowest efficiency level. Furthermore, the policy constraint is cited as the most important obstacle by firms in this group. It seems that this cluster is composed of people who have no choice rather than being self-employees. The group is ignored by the government as it receives no government assistance as well as interventions.

Cluster 4 – Formal and moving from informal to formal. This group includes formal firms and those which move from informal to formal. Firms in this group are young and run by the youngest entrepreneurs with high level of education and technical skills. This might be because the business environment in Vietnam is less constrained to the new-enters. The group is characterised by modern enterprises with high electronic access. Firms in this cluster achieve high level of efficiency and have more opportunities to grow. Therefore, it is reasonable when firms in this group often cite capital, labour, and technical as the most constraints. Being formal is accompanied with regulation interventions because 52% of firms in this group state that they are inspected in 2011 compared to 47% of the sample average.

V. POLICY DISCUSSION AND CONCLUSION

Informality is a complex concept which is unable to be well captured by some criteria for the purpose of statistical measure. Therefore, this paper calculates the informality index by using the multiple correspondence analysis to combine a set of variables indicating the informal status in the data set of SMEs in Vietnam in 2009 and 2011. Non-parametric data envelopment analysis is, then, applied to estimate a firm technical and scale efficiencies that are investigated as factors associating the transition from informal to formal status and vice versa.

By clustering firms into different segments, hypotheses about informality seem to be correct for the case of Vietnam. Firstly, we find that informal status is the only option for small and vulnerable businesses because they have no capacity to grow and move to formal condition. Government policy nearly ignores this group. With the objective of helping people to get out of poverty sustainably, the government should release policies which constrain the existence and growth of firms and pay attention on technical assistance for firms in this group.

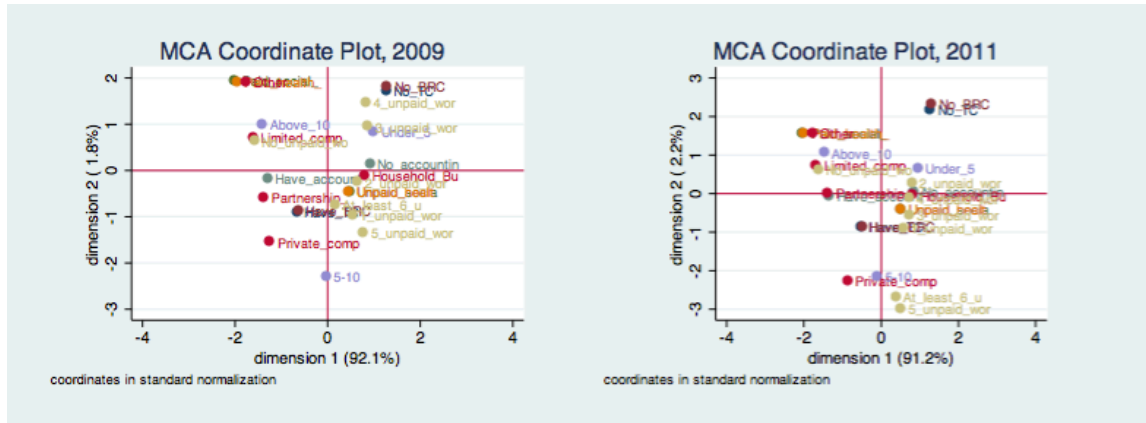
The paper also finds that firms moving from formal to informal status fall into two categories. They may be weak firms that have no potential to expand and thus choose informal status to escape formal cost burdens (cluster 1) or strong firms want to escape formal regulations (cluster 2). Therefore, to impede the movement from formal to informal condition, government policies should remove labour, market, and technical constraints as well as policy interventions.

Findings from the paper reveal that formal firms and those moving from informal to formal status are young and stronger enterprises run by capable entrepreneurs. Benefits from formality such as providing more opportunities on receiving government assistance encourage firms moving to formal status but government regulations seem to interfere too much with this group. The expansion of firms in this group needs more human and capital resources as well as technical assistance

from the government. Therefore, pro-formalisation policies should focus on loan program and training skilled workers and technical know-how for firms.

APPENDIX

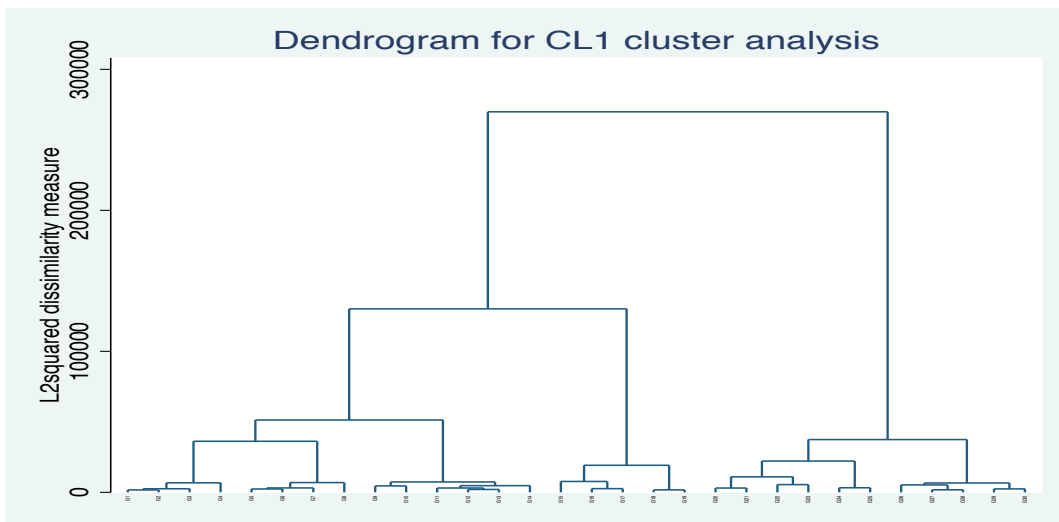
APPENDIX 1. MCA COORDINATE PLOTS



Source: SME 2009, 2011

APPENDIX 2. SOLUTION OF CLUSTER ANALYSIS

A2a. Dendrogram



Source: Authors' calculation

A2b.Duda_Hart indexes for SME clustering

# of clusters	Je(2)/Je(1)	pseudo T-squared
1	0.6492	952.50
2	0.6296	545.95
3	0.7164	328.97
4	0.7474	281.58
5	0.5808	316.92
6	0.7133	214.26
7	0.5245	86.13
8	0.7044	114.13
9	0.5281	60.77
10	0.8294	80.25
11	0.7434	91.81
12	0.7086	70.32
13	0.8012	73.94
14	0.6861	98.81
15	0.6705	59.95

Source: Authors' calculation

The dendrogram provides a solution of either 3, 4, or 6 clusters. We, therefore, base on the Duda_Hart index. Combining both Je(2)/Je(1) index and pseudo_T_squared value, the solution of 4 clusters seem to overweight the other two solutions as Je(2)/Je(1) index is the largest while pseudo_T_squared is the smallest ones.

REFERENCES

1. Anderson, J.H. (1998), The Size, Origins, and Character of Mongolia's Informal Sector during Transition, World Bank Policy Research Working Paper 1916, Washington, D.C.: The World Bank.
2. Arimah, B.C. (2001), "Nature and Determinants of Linkages between Formal and Informal Sector in Nigeria", in African Development Bank, Vol. --, 114-144.
3. Auriol, E., and M. Warlters, 2005, "Taxation Base in Developing Countries," Journal of Public Economics, Vol. 89, pp. 625-46.
4. Azuma, Y., and H. Grossman, 2002, "A Theory of the Informal Sector," NBER Working Paper No. 8823 (Cambridge, Massachusetts: National Bureau of Economic Research).
5. Banker, R.D., A. Charnes, and W.W. Cooper. (1984). "Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis." Management Science Sept. 1078-1092.
6. Blau, P.M. and Scott, W.R, 1963. Formal Organizations: A Comparative Approach, London, Routledge and Kegan Paul.
7. Botero, J., S. Djankov, R. La Porta, F. Lopez-de-Silanes, and A. Shleifer, 2004, "The Regulation of Labor," Quarterly Journal of Economics, Vol. 119, No. 4, pp. 1339-82.
8. Burt, C. (1950). The factorial analysis of qualitative data. British Journal of Psychology, 3, 166-185;
9. Cling, J-P., Lagree, S., Razafindrakoto, M., and Roubaud, F., (2013). Thách thức lớn về phát triển: nâng cao hiểu biết về kinh tế phi chính thức nhằm đưa ra các chính sách phù hợp, in J-P, Cling, Đỗ Hoài Nam, Lagree, S., Razafindrakoto, M., and Roubaud, F., (Eds) Kinh tế phi chính thức tại các nước đang phát triển, pp.11-26. Nhà xuất bản tri thức, Viện Hàn lâm Khoa Học Xã Hội Việt Nam.
10. Coelli, T., Rao, D.P. and Battese, G., (1998). An Introduction to Efficiency and Productivity Analysis, Kluwer Academic Publishers, Boston.
11. Cohen, B. and House, W.J. (1996), "Labor market choices, earnings and informal networks in Khartoum, Sudan", Economic Development and Cultural Change, 44(3): 589-618.
12. Cunningham, W.V., and W.F. Maloney, 2001. Heterogeneity among Mexico's Microenterprises: an application of factor and cluster analysis. Economic Development and Culture Change, Vol. 50, Iss. 1, pp. 131-156.
13. Dabla-Norris et al., 2005. What Causes Firms to Hide Outputs? The Determinants of Informality. IMF Working Paper, WP/05/160.
14. Duda, R. O., and P. E. Hart. 1973. Pattern Classification and Scene Analysis. New York: Wiley.
15. Duval-Hernández, Robert. 2006. Informality, segmentation and earnings in urban Mexico. Mimeo. Center for U.S.-Mexican Studies. San Diego, CA, University of California.
16. Fisher, R. A. (1940). The precision of discriminant functions. Annals of Eugenics, 10, 422-429;
17. Friedman, E., S. Johnson, D. Kaufman, and P. Zoido-Lobaton, 2000, "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries," Journal of Public Economics, Vol. 76, pp. 459-92.
18. Funkhouser, E. (1996). The urban Informal Sector in Central America: Household Survey Evidence, World Development, 24(11), 1731-1751.
19. Giles, D., and L. Tedds, 2002, "Taxes and the Canadian Underground Economy," Canadian Tax Paper No. 106 (Toronto: Canadian Tax Foundation).
20. Girt, A. (1981). Nonlinear multivariate analysis. Leiden, The Netherlands: University of Leiden, Afdeling Data-theorie;
21. Gouldner, A.W., 1954. Patterns of Industrial Bureaucracy, New York, Free Press.
22. Greenacre, M. J. (1984). Theory and applications of correspondence analysis. London: Academic Press;

23. Greenacre, Michael (2007). *Correspondence Analysis in Practice*, Second Edition. London: Chapman & Hall/CRC.
24. Greenacre, Michael and Blasius, Jörg (editors) (2006). *Multiple Correspondence Analysis and Related Methods*. London: Chapman & Hall/CRC.
25. Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In Horst et al. (Eds.), *The prediction of personal adjustment*. (pp. 319-348). New York: Social Science Research Council;
26. Hart, K. (1970) Small-scale Entrepreneurs in Ghana and Development Planning. *Journal of Development Studies*, 6 (4): 104-120.
27. Hart, K., 1972. *Employment, Income and Inequality: a strategy for increasing productive employment in Kenya*. Geneva: ILO.
28. Hart, K., 1973. "Informal Income Opportunities and Urban Employment in Ghana", *Journal of Modern African Studies*, 11, 1, pp 61-89.
29. Hayashi, C. (1950). On the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 2 (No. 1), 35-47.
30. Hirshfield, H. O. (1935). A connection between correlation and contingency. *Cambridge Philosophical Society Proceedings*, 31, 520-524;
31. Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology*, 6, 369-374;
32. International Labour Office (1972). *Employment, Income and Equality: A Strategy for Increasing Productive Employment in Kenya*. International Labour Office, Geneva.
33. International Labour Office (1993). 15th International Conference of Labour Statisticians: Highlights of the Conference and text of the three resolutions adopted. *Bulletin of Labour Statistics* 1993-2, IXXXIV. Geneva.
34. International Labour Office (2002). Effect to be given to resolutions adopted by the International Labour Conference at its 90th Session (2002), (b) Resolution concerning decent work and the informal economy. Governing Body, 285th Session, Seventh item on the agenda; Geneva, November 2002 (doc. GB.285/7/2)
35. Rand, John and Nina Torm (2012). The benefits of formalization: Evidence from Vietnamese Manufacturing SMEs. *World Development*, 40(5): 983-998, May 2012.
36. Le Roux, B. and H. Rouanet (2004), *Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis*, Dordrecht. Kluwer: p.179.
37. Le Roux, B. and H. Rouanet (2004). *Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis*. Dordrecht. Kluwer: p.180.
38. Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive analysis: Correspondence analysis and related techniques for large matrices*. New York: Wiley-Interscience;
39. Levenson, A. R. , and W. F. Maloney, 1998. *The Informal Sector, Firm Dynamics and Institutional Participation*. Working Paper no. 1988. International Bank for Re- construction and Development, Washington, D.C.
40. Livingstone, I. (1991), "A reassessment of Kenya's rural and urban informal sector", *World Development*, 19(6): 651-670.
41. Maloney, W.F. (1999), "Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico", *World Bank Economic Review*, 13(2): 275-302.
42. Maloney, W.F. (1999), "Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico", *World Bank Economic Review*, 13(2): 275-302.
43. Marcouiller, D., and L. Young, 1995, "The Black Hole of Graft: The Predatory State and the Informal Economy," *American Economic Review*, Vol. 85, pp. 630-46.
44. Marcouiller, D., V. Ruiz de Castilla, and C. Woodruff, 1997. Formal Measures of the Informal Sector Wage Gap in Mexico, El Salvador, and Peru. *Economic Development and Culture Change*, Vol. 45, Iss. 2, pp. 367-392.

45. Marcoullier, D., Ruiz de Casilla, V. and Woodruff, C. (1997), "Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, 45(2): 367-92.
46. Marcoullier, D., Ruiz de Casilla, V. and Woodruff, C. (1997), "Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, 45(2): 367-92.
47. Marcoullier, D., Ruiz de Casilla, V. and Woodruff, C. (1997), "Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, 45(2): 367-92.
48. Merrick, T.W. (1976), "Employment and earnings in the informal sector in Brazil: the case of Belo Horizonte", *Journal of Developing Areas*, 10(3): 337-353.
49. Mlinga, R.S. and Wells, J. (2002), "Collaboration Between Informal and Formal Enterprises in the Construction Sector in Tanzania", *Habitat International*, Vol. 26, 269-280.
50. Mooi, E., and M. Sarstedt, 2011. *A Concise Guide to Market Research*. 1st Edition, Springer-Verlag Berlin Heidelberg.
51. Nishisato, S. (1980). *Analysis of categorical data: Dual Scaling and its applications*. Toronto: University of Toronto Press;
52. Pisani, M.J. and Pagan, J.A. (2004), "Self-employment in the era of the New Economic Model in Latin America: a case study from Nicaragua", *Entrepreneurship and Regional Development*, 16: 335-350.
53. Portes, A., Blitzner, S. and Curtis, J. (1986), "The urban informal sector in Uruguay: its internal structure, characteristics, and effects", *World Development*, 14(6): 727-741.
54. Pradhan, M. and van Soest, A. (1995), "Formal and informal sector employment in urban areas of Bolivia", *Labor Economics*, 2: 275-297.
55. Pradhan, M. and van Soest, A. (1995), "Formal and informal sector employment in urban areas of Bolivia", *Labor Economics*, 2: 275-297.
56. Richardson, M., & Kuder, G. F. (1933). Making a rating scale that measures. *Personnel Journal*, 12, 36-40;
57. Saavedra, J. and Chong, A. (1999), "Structural reform, institutions and earnings: Evidence from the formal and informal sectors in urban Peru", *Journal of Development Studies*, 35(4):95-116.
58. Saavedra, J. and Chong, A. (1999), "Structural reform, institutions and earnings: Evidence from the formal and informal sectors in urban Peru", *Journal of Development Studies*, 35(4):95-116.
59. Straub, S., 2005. Informal sector: the credit market channel. *Journal of Development Economics*, Vol. 78, Iss. 2, pp. 299-321.
60. Tenev et al., 2003. *Informality and The Playing Field in Vietnam's Business Sector*. IFC, World Bank, MPDF. Washington, D.C.